Computational Techniques Toward 1 nm Super-resolution Microscopy

COMPUTATIONAL TECHNIQUES TOWARD 1 NM SUPER-RESOLUTION MICROSCOPY

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology
by the authority of the Rector Magnificus, prof. dr. ir. T.H.J.J. van der Hagen,
chair of the Board for Doctorates
to be defended publicly on
Wednesday 24th, April, 2024 at 15:00 o'clock

by

WENXIU WANG

Master of Science in Photonics,
Friedrich-Schiller-University Jena, Germany
born in Nanjing, China

This dissertation has been approved by the promotors.

Composition of the doctoral committee:

Rector Magnificus,	chairperson
Prof. dr. B. Rieger,	Delft University of Technology, *promotor*
Prof. dr. S. Stallinga,	Delft University of Technology, *promotor*
Independent members:	
Prof. dr. I. Testa	KTH Royal Institute of Technology, Sweden
Prof. dr. U. Endesfelder	University of Bonn, Germany
Dr. C. S. Smith	Delft University of Technology
Prof. dr. L. Kuipers	Delft University of Technology
Reserve member:	
Prof. dr. S. Gröblacher	Delft University of Technology

This work is part of the research programme supported by the Dutch Research Council (NWO), VICI grant no. 17046.

Keywords: single-molecule localization microscopy, particle fusion, biological structure analysis, physical optics modeling

Printed by: Ipskamp Printing

Cover: Inspired by Kandinsky's "Sky Blue", this cover abstracts and reinterprets essential elements present in the thesis' illustrations. It was conceptualized and handcrafted by Wenxiu Wang and Yi Zhang.

ISBN 978-94-6473-0465-2

An electronic copy of this dissertation is available at
https://repository.tudelft.nl/.

CONTENTS

SUMMARY

This thesis explores advanced computational techniques in super-resolution microscopy (SRM), with the primary goal of pushing the limits of achievable resolution towards the 1 nm scale. It includes developments in particle fusion algorithms, data analysis of complex biological structures, and exploration of the impact of molecular dipole orientation on MINFLUX localization accuracy and precision.

In the first part, we present a novel fast particle fusion method tailored to single molecule localization microscopy (SMLM). This method first registers particles based on Joint Registration of Multiple Point Clouds (JRMPC) and then classifies and reconnects misaligned locally optimally clustered sets of particles. This approach significantly reduces computational cost compared to earlier template free methods in particular for a large number of particles.. This advancement enables more detailed and accurate reconstructions of super-particles, enhancing the capabilities of SMLM.

The second part of the dissertation deals with a data analysis of nuclear pore complexes (NPCs) reconstructed by the earlier developed particle fusion technique. By fusing thousands of NPCs labeled at nucleoporin Nup96 and analyzing the high-resolution reconstructions, we reveal intricate details of the NPC structure, in particular the unit structure of Nup96. This analysis showcases the potential of SRM in combination with advanced data analysis to contribute to structural biology on the length scale below 10 nm.

The third part focuses on the influence of the dipole orientation on the localization accuracy and precision of MINFLUX. We simulate the imaging process with a physically realistic vector diffraction Point Spread Function (PSF) model and then localize the emitters based on the simplified Gaussian doughnut PSF model used in MINFLUX so far. Our study, including dipoles with free and fixed orientations and key simulation parameters, reveals the need for more refined modeling to overcome the bias, especially for fixed dipole orientations and background fluorescence. This investigation helps to understand the limitations of MINFLUX in its current form and paves the way for future improvements of the technique.

Finally, we discuss potential future directions for improving SRM techniques. These include refining the fast particle fusion method by incorporating localization uncertainties and prior knowledge, optimizing experimental parameters in MINFLUX, and developing advanced localization strategies to improve accuracy and efficiency. By addressing these future challenges, SRM technologies can move closer to the goal of 1 nm resolution in super-resolution imaging.

SAMENVATTING

Dit proefschrift onderzoekt geavanceerde computationele technieken in superresolutiemicroscopie (SRM), met als primair doel het verleggen van de grenzen van de haalbare resolutie naar de schaal van 1 nm. Het omvat ontwikkelingen op het gebied van deeltjesfusiealgoritmen, data-analyse van complexe biologische structuren en onderzoek naar de invloed van moleculaire dipooloriëntatie op de nauwkeurigheid en precisie van MINFLUX-lokalisatie.

In het eerste deel presenteren we een nieuwe snelle deeltjesfusiemethode, afgestemd op enkel molecuul lokalisatiemicroscopie (SMLM). Deze methode registreert eerst deeltjes op basis van Joint Registration of Multiple Point Clouds (JRMPC) en classificeert en verbindt vervolgens verkeerd uitgelijnde, lokaal optimaal geclusterde sets deeltjes. Deze aanpak verlaagt de rekenkosten aanzienlijk in vergelijking met eerdere sjabloonvrije methoden, met name voor een groot aantal deeltjes. Deze vooruitgang maakt gedetailleerdere en nauwkeurigere reconstructies van superdeeltjes mogelijk, waardoor de toepassingsmogelijkheden van SMLM worden vergroot.

Het tweede deel van het proefschrift behandelt een data-analyse van celkernporie complexen (NPC's) gereconstrueerd met de eerder ontwikkelde deeltjesfusietechniek. Door duizenden NPC's gelabeld met het nucleoporine Nup96 te fuseren en de daaruit gemaakte reconstructies met hoge resolutie te analyseren, leggen we structurele details van de ligging van Nup96 in het NPC bloot, in het bijzonder de eenheidsstructuur van Nup96. Deze analyse toont de potentie van SRM in combinatie met geavanceerde data-analyse aan om bij te dragen aan de structurele biologie op de lengteschaal onder 10 nm.

Het derde deel richt zich op de invloed van de dipooloriëntatie op de lokalisatienauwkeurigheid en precisie van MINFLUX. We simuleren het beeldvormingsproces met een fysisch realistisch vectordiffractie puntspreidingsfunctie (PSF) model en lokaliseren vervolgens de fluorescente moleculen op basis van het vereenvoudigde Gaussische 'doughnut' PSF-model dat tot nu toe in MINFLUX wordt gebruikt. Ons onderzoek, inclusief dipolen met vrije en vaste oriëntaties en verschillende relevante simulatieparameters, geeft aan dat er een noodzaak is voor meer verfijnde modellering om de systematische fout te elimineren, vooral voor vaste dipooloriëntaties en achtergrondfluorescentie. Dit onderzoek helpt de beperkingen van MINFLUX in zijn huidige vorm te begrijpen en maakt de weg vrij voor toekomstige verbeteringen van de techniek.

Ten slotte bespreken we mogelijke toekomstige richtingen voor het verbeteren van SRM-technieken. Deze omvatten het verfijnen van de snelle deeltjesfusiemethode door lokalisatie-onzekerheden en voorkennis op te nemen, het optimaliseren van experimentele parameters in MINFLUX en het ontwikkelen van geavanceerde lokalisatiestrategieën om de nauwkeurigheid en efficiëntie te verbeteren. Door deze toekomstige uitdagingen op te pakken, kunnen SRM-technologieën dichter bij het doel komen van 1 nm superresolutie.

1

INTRODUCTION

1

1.1. FROM FLUORESCENCE MICROSCOPY TO SUPER-RESOLUTION MICROSCOPY

1.1.1. FLUORESCENCE MICROSCOPY

A fluorescence microscope is an optical microscope which utilizes fluorescence to observe samples. Fluorescence refers to the emission of light from certain substances (fluorophores) after absorbing the energy of excitation beams [1, 2]. In fluorescence microscopy, the sample is illuminated with an excitation beam with a specific bandwidth of wavelengths. Fluorophores in the specimen absorb the excitation energy and emit light after typically a few nanoseconds. The emission light has a longer wavelength compared to the excitation light. This phenomenon is termed Stokes shift [3]. In fluorescence microscopy, most of the excitation beam passes through the sample while another part is scattered back into the objective lens [4]. Compared to the backscattered part of the excitation beam, the intensity of the emission light is much lower. Therefore, it is crucial to separate the excitation and emission beams properly. Owing to the difference between the wavelengths, the excitation and emission light can be separated by a dichroic mirror and an emission filter to ensure the detector only receives emitted fluorescence.

As any optical microscope, the resolution of a traditional fluorescence microscope is limited by the diffraction of light. A point light source will be imaged by the optical system as a spot with a significantly larger size because of diffraction. The resolution limit describes the smallest feature size of the object that is transmitted by the optical system to the image. In 1873, Ernst Abbe found the diffraction limit d as the inverse of the spatial frequency cutoff of the optical imaging system [5]

$$d = \frac{\lambda}{2NA}, \tag{1.1}$$

where λ is the wavelength of the fluorescence and NA is the numerical aperture of the microscope as illustrated in Fig. 1.1. The NA is a crucial parameter as a higher NA will lead to a better resolution. A conventional fluorescence microscope can only achieve a resolution of approximately 200 nm due to the Abbe limit.

In recent decades, both the development of fluorescent probes and microscopes have extended the boundary of fluorescence microscopy. Fluorescent probes with desired features such as high brightness, low toxicity, specific spectra and switchability enable researchers to study more cellular structures and processes with higher selectivity and increased sensitivity [6–8]. Simultaneously, progress in cutting-edge microscopy techniques has lifted up the achievable resolution beyond the diffraction limit. Many modalities of fluorescence microscopy such as stimulated emission depletion (STED) [9], stochastic optical reconstruction microscopy (STORM) [10], photo-activated localization microscopy (PALM) [11] and minimal photon flux (MINFLUX) [12] enable acquisition of super-resolution images at the nanometer scale by various strategies. This set of technological innovations allow humans to reveal unprecedented details and obtain a deeper understanding of the biological world.

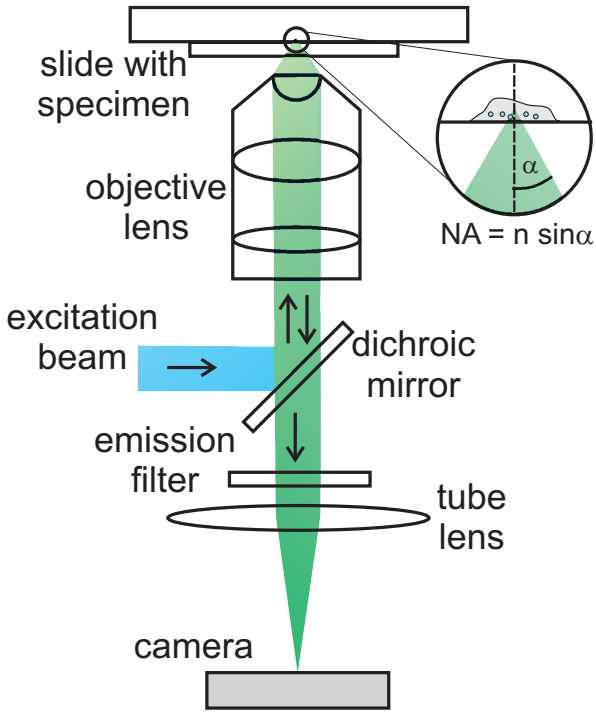


Figure 1.1: The schematics of traditional fluorescence microscopy. Fluorescence microscopy illuminates the specimen with an excitation beam of a certain wavelength to activate fluorescent molecules. Once activated, the fluorescent molecules emit light of a longer wavelength that follows the same path as the excitation beam until it reaches the dichroic mirror. The dichroic mirror separates the emission light from the excitation light, allowing the emission light to pass through an emission filter and to be focused by the tube lens and collected by a camera. The objective lens focuses the excitation beam and collects the emission light. The numerical aperture $NA = n \sin\alpha$ is defined by the refractive index n of the immersion medium of the objective and the half maximum angle α of the light that can enter the objective lens. Image from ref. [13].

1.1.2. SUPER-RESOLUTION MICROSCOPY

Super-resolution microscopy (SRM) comprises a series of advanced techniques that bypass the diffraction limit and significantly improve the resolution limit of the optical microscope from hundreds of nanometers to a few nanometers. These techniques empower researchers to study biological systems in unprecedented detail and resolution. Various super-resolution microscopes apply distinct strategies to overcome the diffraction limit. For instance, stimulated emission depletion microscopy (STED) illuminates the sample with a depletion beam and an excitation beam [9]. The excitation beam 'activates' the fluorophores, while the doughnut-shaped depletion beam 'deactivates' them, resulting in a much narrower excitation area. Thus, the super-resolution image can be obtained from fluorescence emitted from the center of the excitation focus. Expansion microscopy

(ExM) achieves a higher-resolution image by expanding the sample physically [14]. This is achieved by embedding the sample in a polymer gel matrix and swelling the gel. This technique is compatible with many other microscopy modalities. Structured illumination microscopy (SIM) employs a spatially periodic illumination light pattern to obtain high-frequency information which cannot be acquired by standard uniform illumination [15]. Then, computational methods are used to reconstruct the signal to achieve super-resolution images. Both stochastic optical reconstruction microscopy (STORM) [10] and photo-activated localization microscopy (PALM) [11] activate part of the fluorophores at time frames and reconstruct images with the sub-diffraction-limit spatial resolution from precise localizations of fluorophores.

Even though electron microscopy (EM) [16] achieves sub-nanometer resolution by using electron beams which have a much smaller wavelength than light beams, SRM emerges as a significant complementary technique to study biological structures due to its advantages over EM. Firstly, the high specificity of SRM enables researchers to label and visualize the parts of interest from an entire specimen, e.g. certain protein species. Moreover, SRM is more accessible compared to EM as it needs less sample preparation and is more affordable. With so many advantages, SRM has been widely applied in many fields such as cell biology [17, 18], neuroscience [19, 20] and virology [21].

As a rapidly developing field, SRM has been expanding its capabilities [22]. Progress includes innovations on fluorophores, computational techniques and optical setups to shift the obtainable resolution of an optical microscope to approach 1 nanometer.

1.2. SINGLE-MOLECULE LOCALIZATION MICROSCOPE (SMLM)

Single-molecule localization microscope (SMLM) is an important subset of super-resolution microscopy. In order to achieve a resolution beyond the diffraction limit, SMLM combines biochemical techniques, optical instrumentation and processing algorithms to localize single molecules precisely. We will introduce how to acquire and process SMLM data in this section.

1.2.1. DATA ACQUISITION

To perform SMLM, biological samples must be labeled with specific fluorescent molecules. These fluorescent molecules can be activated to an 'on-state' (fluorescent) and deactivated to an 'off-state' (non-fluorescent) by manipulating the environmental conditions such as changing the excitation light parameters and chemical buffer composition. During the imaging process, samples are exposed to excitation beams of specific wavelengths and intensities to ensure that only a small fraction of fluorophores are in the 'on-state' at any time frame. In ideal cases, adjacent fluorophores within a distance smaller than the Abbe limit will not be fluorescent simultaneously, allowing for localizing the position of the fluorophore precisely from the measured spot generated from a single point light source. The average or expected shape of this spot is called the Point Spread Function (PSF). To avoid the overlap of several spots, experimental conditions should be adjusted to optimize the 'blinking' performance so that fluorophores are fluorescent sparsely in the spatial domain while most fluorophores are collected in the temporal domain. A single image from a time frame includes isolated spots originating from single fluorescent

molecules. Multiple (about 10 to 100 thousand) raw images are recorded over time so that most of the fluorophores are captured. The detector in SMLM should have high sensitivity and low noise to capture raw images from weak signals. These raw images are then processed to acquire accurate localizations and reconstruct the super-resolution image.

There are three prominent SMLM techniques including photo-activated localization microscopy (PALM) [11], stochastic optical reconstruction microscopy (STORM) [10], and points accumulation for imaging in nanoscale topography (PAINT) [23]. They achieve single molecule localization by different strategies. PALM illuminates photo-switchable fluorescent proteins [24] or photoactivatable fluorescent proteins [25] with an activation light to achieve the blinking of fluorophores. STORM, on the other hand, stimulates fluorescent organic dyes [26] with the combination of an activation laser and specialized oxygen-scavenging imaging buffers [27]. In contrast, the PAINT technique labels samples with antibodies conjugated with docking strands. Imager strands create 'blinking events' by transiently binding with docking strands.

Despite various strategies enabling the precise single-molecule localization, there remain some limitations and challenges associated with the data acquisition of SMLM such as photobleaching, the density of labelling and sample drift [28, 29]. Photobleaching [30] refers to the loss of fluorescence ability of fluorophores caused by repeated activation of fluorophores, which ultimately decreases the SNR and quality of raw images acquired by SMLM. Managing labelling density is crucial to achieving high-quality SMLM images. Under-labelling, for example, results in insufficient information. During the data acquisition time, samples may drift due to the instability of the environment, negatively influencing localization precision. Continuous efforts are made by researchers to overcome these limits and challenges to improve the data acquisition of SMLM [31, 32]. Advanced computational techniques in the processing pipeline are also helpful to compensate for the limitations in the data acquisition process.

1.2.2. PROCESSING PIPELINE

Processing after data acquisition is essential for transferring tens of thousands of frames to one super-resolved visualization. The processing pipeline is depicted in Fig. 1.2 and will be explained step by step in this section.

SEGMENTATION

Each raw camera frame contains not only spots of fluorescent molecules in the 'on-state' but also background noise. The segmentation step extracts multiple regions of interest (ROIs) from the entire image, with each ROI containing a single fluorescent molecule's spot, thereby distinguishing valuable spots from the noisy background. Various segmentation algorithms employ diverse approaches including local maximum detection [10], wavelet theory [33] and machine learning [34].

LOCALIZATION

Following the segmentation, the localization step derives coordinates of molecules with sub-pixel resolution from the ROIs. In real experiments, ROIs exhibit various intensity distributions due to factors such as the molecule orientation, aberrations and axial position [35]. The localization step aims to precisely estimate the single molecule's positions. Both maximum likelihood estimation (MLE) [36] and least squares (LS) fitting[37]

are widely used to fit intensity distributions in ROIs to PSF models, enabling the localization of molecule positions.

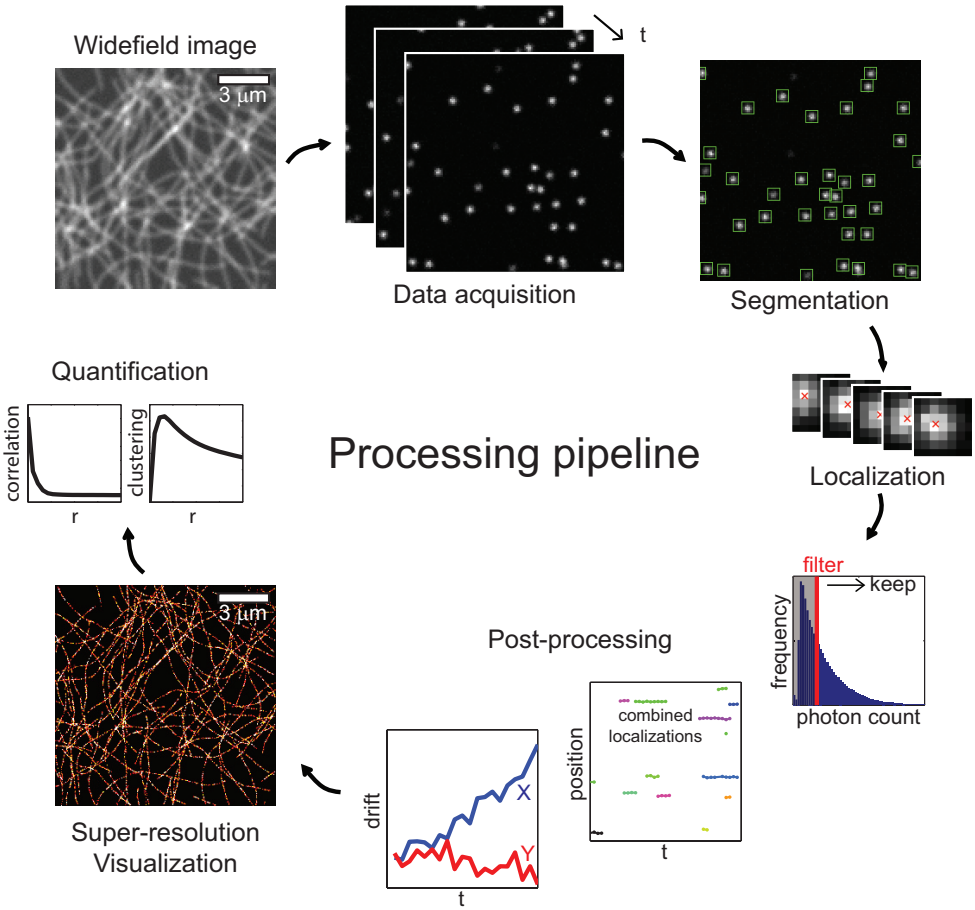


Figure 1.2: The complete image processing pipeline of SMLM to generate a super-resolution visualization. Steps include acquiring raw images, segmenting ROIs, localising molecules, filtering out 'poor' localizations, combining localizations, correcting drifting, visualizing super-resolved images and quantifying the visualization. Image from ref. [13].

FILTERING

The filtering step removes inaccurate localizations. This step reduces the number of coordinates and leads to an overall better resolution. For experimental data, factors like overlapping molecules, aberrations and out-of-focus will result in inaccurate localizations. These 'poor' localizations are typically filtered out based on estimation parameters such as photon count, localization precision and the width of the PSF [13].

POST-PROCESSING

Post-processing includes frame connection and drift correction. In the frame connection step, localizations in close proximity within both spatial (within a few localization uncertainties) and temporal (across several frames) domains are considered to originate from the same fluorescent molecule and are combined into a single localization [13]. The data acquisition of SMLM lasts from several minutes to hours. During the acquisition time, the relative position between the sample and the detector may shift, leading to a reduced quality of the final visualization. This drift can be corrected by fixing the relative positions of hardware components, by using fiducial markers as references in the raw image [10], or from correlations in the dataset itself [38, 39].

VISUALIZATION

The visualization step renders the super-resolution image from a list of localization coordinates. This can be achieved by several approaches. Gaussian rendering, for example, represents each localization as a Gaussian distribution. The center of the Gaussian distribution is placed at the localization and its standard deviation is proportional to the precision of the localization [40]. In contrast, histogram-based rendering divides the field of view into square pixel bins and assigns each localization to its nearest pixel. The intensity value of the pixel is the count of localizations within it.

QUANTIFICATION

Quantification is crucial analysis step for SMLM as it facilitates an in-depth interpretation of the super-resolution reconstruction. This step involves quantifying various aspects of the final result such as resolution measurement and cluster analysis. Fourier Ring Correlation (FRC) [41] is a practical way to measure the resolution of SMLM as it takes all the factors that influence image resolution into account and needs no prior knowledge. Cluster analysis investigates the spatial arrangement of the SMLM image, making it possible to extract meaningful information from the super-resolution visualization [42]. The pair correlation function [43] and techniques based on Ripley's K-function [44] are commonly used methods for cluster analysis.

1.3. MINFLUX SUPER-RESOLUTION MICROSCOPY

MINFLUX is a super-resolution technique that combines modulated illumination with single molecule localization to achieve a resolution of several nanometers [12]. Nowadays, as a powerful fluorescent microscopy technique, MINFLUX has many flavours [45, 46] and derived methods such as SIMFLUX [47, 48] and is used to resolve biological structures [49], to track molecule movements [12] and to observe dynamic movements of molecular machines [50]. In this section, we will introduce the principle of MINFLUX in detail.

MINFLUX uses doughnut-shaped excitation beams to accurately locate emitters. The position and size of the doughnut-shaped beams can be precisely controlled. The intensity of the doughnut-shaped beam varies with the distance from the center. As a result, the molecule located at different positions within the excitation beam would emit different numbers of photons. Notably, if a molecule is located at the doughnut's zero point, no photons would be emitted. However, the exact position of the molecule is unknown, making it almost impossible to align the molecule with the doughnut zero.

In two-dimensional MINFLUX, a fluorescent molecule is exposed sequentially to a doughnut-shaped beam with four different centers at selected positions over an area of size L, as shown in Fig. 1.3. Three positions of the doughnut beam centers are arranged as an equilateral triangle, while the fourth position is located at the triangle's center. The single molecule yields a set of photon counts from these four illuminations. Given the known intensity distribution and locations of the doughnut-shaped illuminations, the localization of the molecule can be estimated from a set of photon counts received by the detector.

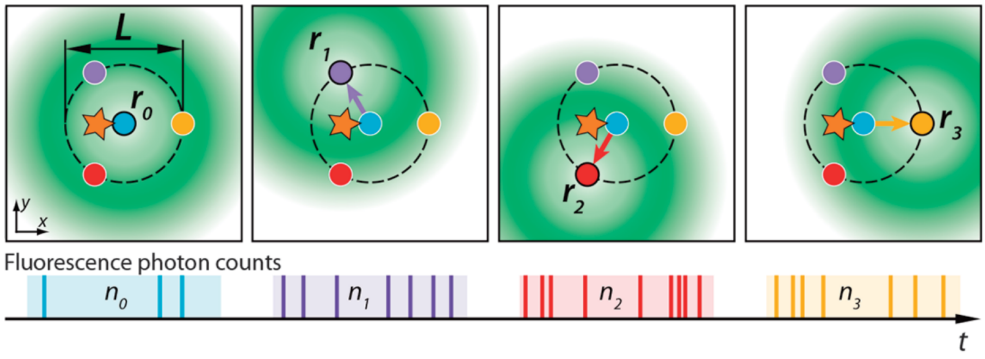


Figure 1.3: An illustration of the four doughnut beam illuminations displayed in the focal plane for two-dimensional MINFLUX. The centers of the beams ($\mathbf{r}_0$, $\mathbf{r}_1$, $\mathbf{r}_2$, $\mathbf{r}_3$) are shown as coloured dots within an area of size L. The single molecule, shown as an orange star, generates fluorescence photon counts (n_0, n_1, n_2, n_3) in response to each beam. The doughnut-shaped beams illuminate the molecule individually at different time intervals. Image taken from ref. [12].

MINFLUX employs Maximum Likelihood Estimation (MLE) to localize molecules with high precision. MLE is a statistical method for estimating the parameters of a model. It determines parameters by maximizing the likelihood function for the observed data to fit the statistical model. In MINFLUX, the estimated parameter is the fluorescent molecule's localization ($\mathbf{r}_m$). The fluorescent molecule is exposed to K excitation beams with centers at different positions ($\mathbf{r}_0, \mathbf{r}_1, ..., \mathbf{r}_{K-1}$). The observed data is the set of detected photon counts ($\bar{n} = \{n_0, n_1, ..., n_{K-1}\}$). For two-dimensional MINFLUX, the intensity distribution of the doughnut-shaped beam $I_{doughnut}$ is simplified as

$$I_{doughnut}(\mathbf{r}) = A_0 4e\ln(2)\frac{r^2}{FWHM^2}e^{-4\ln(2)\frac{r^2}{FWHM^2}}, \tag{1.2}$$

where A_0 is the peak intensity of the doughnut beam, $FWHM$ is the Full Width Half Maximum of the distribution and $r = |\mathbf{r}|$ is the distance from the doughnut zero. In the statistical model, after neglecting background and detector dark counts, the number of photons collected is generated as a function of illumination intensity according to Poissonian

statistics. After dropping the multiplicative factors, a simplified log-likelihood function is

$$l(\mathbf{r}_m|\bar{n}) = \sum_{i=1}^{K-1} n_i \ln p_i \tag{1.3}$$

where $K = 4$ is the number of exposures a molecule encounters, and where the p_i are a set of parameters to describe the probabilities for measuring the collection of photons, defined by

$$p_i = \frac{I_{doughnut}(\mathbf{r}_m - \mathbf{r}_i)}{\sum_{j=1}^{K-1} I_{doughnut}(\mathbf{r}_m - \mathbf{r}_j)}, i \in (0, ..., K-1) \tag{1.4}$$

The estimated molecule position $\mathbf{r}_m$ is the coordinate that gives the maximum $l(\mathbf{r}_m|\bar{n})$.

1.4. PARTICLE FUSION IN SMLM

The resolution of SMLM is limited by several factors including a finite photon budget, giving rise to a limited localization precision, and low labelling density. Particle fusion is an effective computational technique to improve the resolution and signal-to-noise ratio of structures by averaging chemically identical copies (particles) [51].

SMLM experiments are capable of capturing thousands of particles, each in different poses within the field of view. The image of a single particle obtained by SMLM is usually incomplete as shown in Fig. 1.4a-b). The incompleteness is due to the inevitable imperfect Density Of Labelling (DOL). The typical DOL is about 30% to 70%, which means that more than half of the binding sites can very well be unlabeled. The number of localizations and the missing parts between particles vary due to the stochastic nature of the labelling process and 'blinking' events. Merging particles by properly rotating and translating them can result in a super-particle (Fig. 1.4c) of better quality since particles share the same underlying structure, despite the limited quality and stochastic variations of the particles.

The idea of particle fusion originates from single-particle analysis (SPA) [52] within the field of electron microscopy (EM), which uses this technique to determine the average structures of macromolecular complexes. Although SPA for EM is applicable to SMLM [53], it is advantageous to develop specific particle fusion methods tailored to SMLM. This is due to the distinct properties of SMLM such as coordinate-based imaging, incomplete fluorescence labelling, and statistical variations in labelling and repeated localizations per binding site [51]. Some particle fusion methods [54, 55] use a template to register particles, which can lead to a template-biased reconstruction [56]. Template-free methods have also been developed based on different strategies including pyramid registration [55] and all-to-all registration [51, 57]. Template-free methods can reconstruct subcellular structures without prior knowledge and have the potential to reveal unknown structures from SMLM datasets.

Particle fusion is a crucial technique in SMLM that addresses the challenge of underlabelling. It enhances the structural information by merging particles, allowing for a better resolution of the reconstruction. This approach reduces the resolution gap between the EM and SMLM and improves the capabilities of SMLM in structural biology.

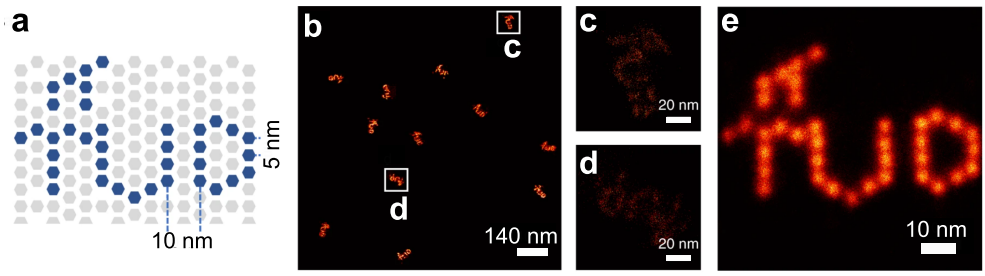


Figure 1.4: Schematic representation of particle fusion for experimental 2D TUD-logo DNA-origami particles with DOL=80%. a) The template of the DNA origami model with 37 DNA-PAINT docking sites. b) A field of view with many particles. c-d) Two single particles with different poses and incomplete labelling as highlighted in panel b. e) The super-particle reconstructed by fusing 383 individual particles. Image adapted from ref. [51].

1.5. IMAGING OF NUCLEAR PROTEIN COMPLEX (NPC)

1.5.1. INTRODUCTION

The nuclear pore complex (NPC) is a protein assembly embedded in the nuclear envelope that connects the nucleus to the cytoplasm [58–60]. It functions as a crucial molecular machine involved in many cellular processes [61]. Given the complexity of its structure and the importance of its function, the imaging and study of the NPC have attracted considerable attention in the scientific community for many years.

As one of the largest protein complexes in the eukaryotic cell, the NPC has a doughnut-shaped structure that contains three scaffold rings; the inner ring (IR), the cytoplasmic ring (CR) and the nuclear ring (NR), as shown in Fig. 1.5. The CR and NR are outer rings situated on the cytoplasmic and nucleoplasmic sides, respectively, exhibiting a twofold symmetry across the nuclear membrane. The IR is located within the nuclear envelope and sandwiched between the CR and NR [62]. These rings are built from multiple copies of ~30 distinct proteins, termed nucleoporins (Nups). These Nups form multiple subcomplexes, the so-called Y-complex being the most prominent one [63]. The outer rings are formed by 16 copies of the Y-complexes and show an eight-fold symmetry in each ring [64]. The size and mass of the NPC can vary depending on the cell type and functional state. The human NPC has a molecular mass of about 109 MDa [65] while the yeast NPC is approximately 52 MDa [66].

The unique structure of the NPC enables its multifaceted roles in various cellular processes. Serving as a gateway, the NPC mediates the exchange of molecules between the nucleus and the cytoplasm. Molecules smaller than ~40 kDa can pass through the nuclear envelope freely, while larger macromolecules are selectively imported and exported by the NPC. This activity functions as a diffusion barrier, helping to separate and preserve the distinct environments of the nuclear compartment and the cytoplasm. Through actively transporting macromolecules such as mRNA, tRNA, viruses, proteins and ribosomes [67–69], the NPC plays significant roles in gene expression [67, 70], viral replication and infection [71], and chromatin organization [72].

The study of the NPC contributes significantly to understanding various biological

processes, which may lead to potential treatments for a range of diseases such as cancer, viral infections and neurodegenerative disorder [73]. Despite the extensive research about NPCs, there are still many unsolved questions about their structure and function, making NPC a continuously popular topic to study in the field of biology.

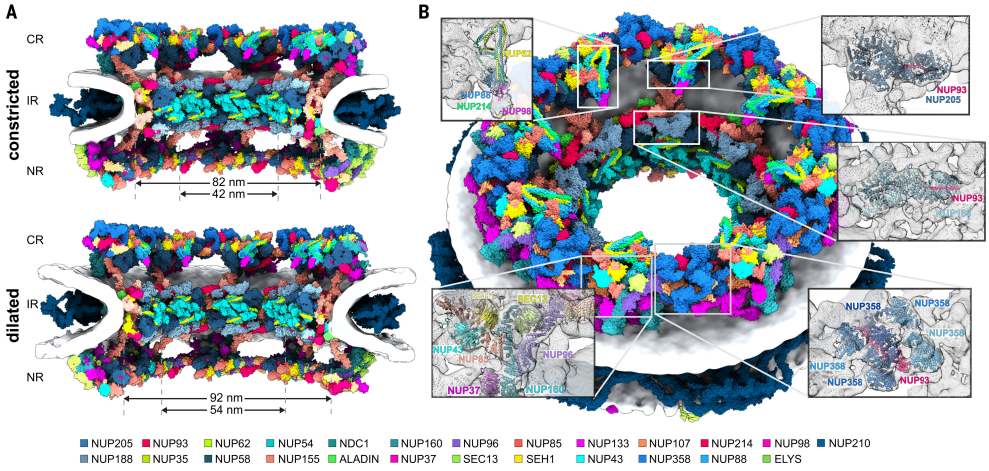


Figure 1.5: The near-complete structural model of the human NPC. The grey isosurface is the nuclear envelope while the inner ring (IR), the cytoplasmic ring (CR) and the nuclear ring (NR) are colourful. Distinct components are colour-coded for clarity. A)Cut-away views illustrate the constricted and dilated state of the human NPC. B) Oblique view of the constricted state from the cytoplasmic side. Insets provide the details of specific substructures of the NPC with Nups overlaid onto the isosurface-rendered cryo-ET map. Image taken from ref. [63].

1.5.2. NPC IMAGING WITH SRM

Cryogenic electron microscopy (cryo-EM) and cryo-electron tomography (cryo-ET) have contributed significantly to our understanding of the NPC by resolving the structure of the NPC in high resolution [59, 60, 74, 75]. However, SRM is emerging as a powerful complementary technique for NPC imaging. While EM offers a better resolution, SRM brings its unique advantages to NPC studies. Primarily, SRM offers high specificity, allowing a specific Nup from an entire NPC to be labeled and visualized. This ability to study individual Nups improves our understanding of their specific roles and interactions within the NPC.

In recent years, substantial efforts have been made in the field of NPC imaging. For instance, Lőschberger et al. in 2012 used direct stochastic optical reconstruction microscopy (dSTORM) to visualize gp210 proteins in NPCs on isolated Xenopus laevis oocyte nuclear envelopes [76]. By merging hundreds of particles, they demonstrated high SNR reconstructions of the eight-fold rotational symmetric ring of gp210 in the NPC. A year later, Szymborska et al. analyzed the NPC scaffold structure by SRM and particle averaging [77]. They measured the average positions of fluorescent molecular labels by averaging thousands of particles and achieved a precision in radius better than 1 nanometer. In 2021, Heydarian et al. reconstructed the 3D structure of Nup107 and Nup96 using a template-

free particle fusion method [57]. The eight-fold symmetry and the phase shift between two rings of the NPC were revealed from these 3D reconstructions. MINFLUX has been used on labeled Nup96 in NPCs for obtaining high-resolution 3D reconstructions that resolved the eight-fold symmetry of Nup96 in individual NPCs [78]. Most recently, in 2023, Reinhardt et al. imaged Nup96s of NPCs with an unprecedented spatial localization precision using a DNA-barcoding method called resolution enhancement by sequential imaging (RESI) [79]. By applying a model-free average [80] to over a thousand particles, individual Nup96 proteins can be discerned in the resulting reconstruction.

Updated SRM techniques and advanced computational techniques have led to an improvement in the resolution achievable, thereby enhancing the ability of SRM to image biological structures. Further advances can be made using SRM to observe subcellular structures including the NPC.

1.6. MOTIVATION AND OUTLINE

1.6.1. MOTIVATION

The motivation of this thesis is to push the boundaries of achievable resolution in super-resolution microscopy towards 1 nm through the development and application of advanced computational techniques.

The advent of super-resolution microscopy offers significant advances in biological research. The 'diffraction-unlimited' resolution provided by SRM enables the visualization of subcellular structures with unprecedented detail down to the nanometer scale. Despite these advancements, SRM faces challenges that limit its potential in terms of obtainable resolution. On the one hand, experiments are continuously striving for better data. On the other hand, the development of advanced computational techniques helps to overcome these challenges.

This thesis includes three interconnected chapters, each focusing on a particular aspect of SRM computational techniques: the particle fusion algorithm, data analysis for high-resolution reconstruction, and physical optics based simulation for MINFLUX. While the subjects might seem distinct, they represent a concerted effort to improve the capability of SRM in the computational aspect and each chapter of this thesis is motivated by the same goal of shifting the obtainable resolution in SRM toward 1 nm. Advanced computational techniques are showcased as powerful tools, helping SRM to overcome existing limitations and achieve superior resolution.

1.6.2. OUTLINE

JOINT REGISTRATION OF MULTIPLE POINT CLOUDS FOR FAST PARTICLE FUSION IN LOCALIZATION MICROSCOPY

In Chapter 2, we describe a fast particle fusion method based on the joint registration of multiple point clouds (JRMPC) tailored for localization microscopy. The incorporation of template-free particle fusion methods allows for the registration of particles, producing a super-particle without reliance on prior knowledge. This approach effectively compensates for the low density of labelling in SMLM. While existing cutting-edge template-free particle fusion methods are proven to work well, their computational cost increases unfavourably with the number of particles. Our method, based on JRMPC, overcomes this

problem, partially achieving a linear scaling of computational cost with the number of particles. In our method, we register input particles to a Gaussian Mixtures Model (GMM). Subsequently, we compute the probabilities for each particle to fit within the GMM. Iterative updates are then implemented for both input particles and the GMM, optimizing likelihood estimations. However, the results of the direct application of JRMPC converge to a local optimum, encompassing several groups of particles with varied poses. We then apply a classification method to separate these clusters. These separated clusters are then connected in a subsequent step, resulting in an accurate and comprehensive reconstruction. The fast particle fusion method significantly reduces the time and computational resources required, providing greater efficiency, especially for large-scale or time-sensitive studies. The ability to register a large number of particles enhances the resolution of the super-particle, providing a more detailed reconstruction than that obtained from fewer particles.

PARTICLE FUSION OF SUPER-RESOLUTION DATA REVEALS THE UNIT STRUCTURE OF NUP96 IN NUCLEAR PORE COMPLEX

In Chapter 3, we conduct a careful data analysis of high-resolution particle fusion reconstructions comprising thousands of Nup96 labeled NPCs. Despite advancements in EM revealing the positions of the Nup96 copies within the NPC, localizing the Nup96s remains a challenge in the field of SRM due to limitations in achievable resolution. In previous studies, every pair of adjacent Nup96s are shown as a single blob in the eight-fold symmetric ring of the NPC. By applying fast particle fusion to more than four thousand particles, we obtained detailed three-dimensional reconstructions of Nup96s within the NPCs with high resolution. Contrary to previous spherical representations, the blobs in our reconstructions display a distinct ellipsoidal form. These ellipsoidal blobs suggest potential locations for pairs of Nup proteins. We further refine these reconstructions by fitting them to anisotropic Gaussian models, followed by a careful position analysis derived from this fitting process. The positions of the two Nup96 per unit, as analyzed from our SRM reconstructions, align accurately with the reference standard established from EM data. This study underscores the promising applications of SRM in biological structure research, demonstrating its potential to provide insights into the complex organization of cellular components.

IMPACT OF MOLECULAR DIPOLE ORIENTATION ON THE ACCURACY OF MINFLUX

In Chapter 4, we investigate the impact of the orientation of the fluorophore's dipole moment on the localization accuracy of MINFLUX. MINFLUX is based on measuring the energy absorbed by fluorescent molecules without considering the effect of dipole orientation, claiming that the unknown orientation of the molecular orientation has no influence on the solution [12]. However, based on physical principles, distinct dipole orientations could result in different energy absorption from an identically polarized excitation beam. Differences between experiments and MLE calculations regarding energy absorbed by the dipole could so introduce a bias in the localization. We therefore investigate how dipole orientation influences the accuracy of localization in MINFLUX.

To this end, we first establish a physical optics based forward model to generate a vectorial doughnut-shaped PSF. Subsequently, we calculate the energy absorbed by a dipole with a random orientation. Based on the energy absorbed by the dipole when exposed to

various electric fields, we generate a photon count that follows Poissonian statistics. Next, we use the MLE method and the scalar PSF model as described in the MINFLUX papers [12] to estimate the emitter position based on these photon counts.

Chapter 5 concludes the thesis by highlighting the critical findings from each chapter and how they contribute towards achieving 1 nm resolution in SRM. It also discusses potential future improvements in computational techniques to achieve 1 nm resolution in SRM.

REFERENCES

[1] J. R. Lakowicz and J. R. Lakowicz, *Introduction to fluorescence,* Principles of Fluorescence Spectroscopy (1999), pp. 1–23, doi: 10.1007/978-0-387-46312-4.

[2] B. Valeur and M. N. Berberan-Santos, *Molecular fluorescence: principles and applications* (John Wiley & Sons, 2012) doi: 10.1002/9783527650002.

[3] P. Abbyad, W. Childs, X. Shi, and S. G. Boxer, *Dynamic stokes shift in green fluorescent protein variants,* Proceedings of the National Academy of Sciences **104** (2007), pp. 20189–20194, doi: 10.1073/pnas.0706185104.

[4] J. Vangindertael, R. Camacho, W. Sempels, H. Mizuno, P. Dedecker, and K. Janssen, *An introduction to optical super-resolution microscopy for the adventurous biologist,* Methods and Applications in Fluorescence **6** (2018), p. 022003, doi: 10.1088/2050-6120/aaae0c.

[5] E. Abbe, *Beiträge zur Theorie des Mikroskops und der mikroskopischen Wahrnehmung,* Archiv für mikroskopische Anatomie **9** (1873), pp. 413–468, doi: 10.1007/BF02956173.

[6] M. Vendrell, D. Zhai, J. C. Er, and Y.-T. Chang, *Combinatorial strategies in fluorescent probe development,* Chemical Reviews **112** (2012), pp. 4391–4420, doi: 10.1021/cr200355j.

[7] J. Zhang, R. E. Campbell, A. Y. Ting, and R. Y. Tsien, *Creating new fluorescent probes for cell biology,* Nature Reviews Molecular Cell Biology **3** (2002), pp. 906–918, doi: 10.1038/nrm976.

[8] R. Y. Tsien, *The green fluorescent protein,* Annual Review of Biochemistry **67** (1998), pp. 509–544, doi: 10.1146/annurev.biochem.67.1.509.

[9] S. W. Hell and J. Wichmann, *Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy,* Optics Letters **19** (1994), pp. 780–782, doi: 10.1364/ol.19.000780.

[10] M. J. Rust, M. Bates, and X. Zhuang, *Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM),* Nature Methods **3** (2006), pp. 793–796, doi: 10.1038/nmeth929.

[11] S. T. Hess, T. P. Girirajan, and M. D. Mason, *Ultra-high resolution imaging by fluorescence photoactivation localization microscopy,* Biophysical Journal **91** (2006), pp. 4258–4272, doi: 10.1529/biophysj.106.091116.

[12] F. Balzarotti, Y. Eilers, K. C. Gwosch, A. H. Gynnå, V. Westphal, F. D. Stefani, J. Elf, and S. W. Hell, *Nanometer resolution imaging and tracking of fluorescent molecules with minimal photon fluxes,* Science **355** (2017), pp. 606–612, doi: 10.1126/science.aak9913.

[13] B. Rieger, R. Nieuwenhuizen, and S. Stallinga, *Image processing and analysis for single-molecule localization microscopy: Computation for nanoscale imaging,* IEEE Signal Processing Magazine **32** (2014), pp. 49–57, doi: 10.1109/MSP.2014.2354094.

[14] F. Chen, P. W. Tillberg, and E. S. Boyden, *Expansion microscopy,* Science **347** (2015), pp. 543–548, doi: 10.1126/science.1260088.

[15] M. G. Gustafsson, *Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy,* Journal of Microscopy **198** (2000), pp. 82–87, doi: 10.1046/j.1365-2818.2000.00710.x.

[16] J. J. Bozzola and L. D. Russell, *Electron microscopy: principles and techniques for biologists* (Jones & Bartlett Learning, 1999).

[17] K. Xu, G. Zhong, and X. Zhuang, *Actin, spectrin, and associated proteins form a periodic cytoskeletal structure in axons,* Science **339** (2013), pp. 452–456, doi: 10.1126/science.1232251.

[18] S.-H. Shim, C. Xia, G. Zhong, H. P. Babcock, J. C. Vaughan, B. Huang, X. Wang, C. Xu, G.-Q. Bi, and X. Zhuang, *Super-resolution fluorescence imaging of organelles in live cells with photoswitchable membrane probes,* Proceedings of the National Academy of Sciences **109** (2012), pp. 13978–13983, doi: 10.1073/pnas.1201882109.

[19] B. A. Wilt, L. D. Burns, E. T. Wei Ho, K. K. Ghosh, E. A. Mukamel, and M. J. Schnitzer, *Advances in light microscopy for neuroscience,* Annual Review of Neuroscience **32** (2009), pp. 435–506, doi: 10.1146/annurev.neuro.051508.135540.

[20] M. Maglione and S. J. Sigrist, *Seeing the forest tree by tree: super-resolution light microscopy meets the neurosciences,* Nature Neuroscience **16** (2013), pp. 790–797, doi: 10.1038/nn.3403.

[21] W. Muranyi, S. Malkusch, B. Müller, M. Heilemann, and H.-G. Kräusslich, *Super-resolution microscopy reveals specific recruitment of HIV-1 envelope proteins to viral assembly sites dependent on the envelope C-terminal tail,* PLoS Pathogens **9** (2013), p. e1003198, doi: 10.1371/journal.ppat.1003198.

[22] K. Prakash, B. Diederich, R. Heintzmann, and L. Schermelleh, *Super-resolution microscopy: a brief history and new avenues,* Philosophical Transactions of the Royal Society A **380** (2022), p. 20210110, doi: 10.1098/rsta.2021.0110.

[23] J. Schnitzbauer, M. T. Strauss, T. Schlichthaerle, F. Schueder, and R. Jungmann, *Super-resolution microscopy with DNA-PAINT,* Nature Protocols **12** (2017), pp. 1198–1228, doi: 10.1038/nprot.2017.024.

[24] M. Bates, B. Huang, and X. Zhuang, *Super-resolution microscopy by nanoscale localization of photo-switchable fluorescent probes,* Current Opinion in Chemical Biology **12** (2008), pp. 505–514, doi: 10.1016/j.cbpa.2008.08.008.

[25] K. A. Lukyanov, D. M. Chudakov, S. Lukyanov, and V. V. Verkhusha, *Photoactivatable fluorescent proteins,* Nature Reviews Molecular Cell Biology **6** (2005), pp. 885–890, doi: 10.1038/nrm1741.

[26] U. Resch-Genger, M. Grabolle, S. Cavaliere-Jaricot, R. Nitschke, and T. Nann, *Quantum dots versus organic dyes as fluorescent labels,* Nature Methods **5** (2008), pp. 763–775, doi: 10.1038/nmeth.1248.

[27] N. Olivier, D. Keller, P. Gönczy, and S. Manley, *Resolution doubling in 3D-STORM imaging through improved buffers,* PloS One **8** (2013), p. e69004, doi: 10.1371/journal.pone.0069004.

[28] M. Lelek, M. T. Gyparaki, G. Beliu, F. Schueder, J. Griffié, S. Manley, R. Jungmann, M. Sauer, M. Lakadamyali, and C. Zimmer, *Single-molecule localization microscopy,* Nature Reviews Methods Primers **1** (2021), p. 39, doi: 10.1038/s43586-021-00038-x.

[29] A. Shivanandan, H. Deschout, M. Scarselli, and A. Radenovic, *Challenges in quantitative single molecule localization microscopy,* FEBS Letters **588** (2014), pp. 3595–3602, doi: 10.1016/j.febslet.2014.06.014.

[30] C. Eggeling, J. Widengren, L. Brand, J. Schaffer, S. Felekyan, and C. A. Seidel, *Analysis of photobleaching in single-molecule multicolor excitation and Förster resonance energy transfer measurements,* The Journal of Physical Chemistry A **110** (2006), pp. 2979–2995, doi: 10.1021/jp054581w.

[31] Y. Tang, X. Wang, X. Zhang, J. Li, and L. Dai, *Sub-nanometer drift correction for super-resolution imaging,* Optics Letters **39** (2014), pp. 5685–5688, doi: 10.1364/ol.39.005685.

[32] A. O. Khan, V. A. Simms, J. A. Pike, S. G. Thomas, and N. V. Morgan, *CRISPR-Cas9 mediated labelling allows for single molecule imaging and resolution,* Scientific Reports **7** (2017), p. 8450, doi: 10.1038/s41598-017-08493-x.

[33] I. Izeddin, J. Boulanger, V. Racine, C. Specht, A. Kechkar, D. Nair, A. Triller, D. Choquet, M. Dahan, and J. Sibarita, *Wavelet analysis for single molecule localization microscopy,* Optics Express **20** (2012), pp. 2081–2095, doi: 10.1364/oe.20.002081.

[34] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, *Pointnet: Deep learning on point sets for 3D classification and segmentation,* in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2017) pp. 652–660, doi: 10.1109/CVPR.2017.16.

[35] C. N. Hulleman, R. Ø. Thorsen, E. Kim, C. Dekker, S. Stallinga, and B. Rieger, *Simultaneous orientation and 3D localization microscopy with a Vortex point spread function,* Nature Communications **12** (2021), p. 5934, doi: 10.1038/s41467-021-26228-5.

[36] C. S. Smith, N. Joseph, B. Rieger, and K. A. Lidke, *Fast, single-molecule localization that achieves theoretically minimum uncertainty,* Nature Methods **7** (2010), pp. 373–375, doi: 10.1038/nmeth.1449.

[37] A. V. Abraham, S. Ram, J. Chao, E. Ward, and R. J. Ober, *Quantitative study of single molecule location estimation techniques,* Optics Express **17** (2009), pp. 23352–23373, doi: 10.1364/oe.17.023352.

[38] M. J. Mlodzianoski, J. M. Schreiner, S. P. Callahan, K. Smolková, A. Dlasková, J. Šantorová, P. Ježek, and J. Bewersdorf, *Sample drift correction in 3D fluorescence photoactivation localization microscopy,* Optics Express **19** (2011), pp. 15009–15019, doi: 10.1364/oe.19.015009.

[39] Y. Wang, J. Schnitzbauer, Z. Hu, X. Li, Y. Cheng, Z.-L. Huang, and B. Huang, *Localization events-based sample drift correction for localization microscopy with redundant cross-correlation algorithm,* Optics Express **22** (2014), pp. 15982–15991, doi: 10.1364/oe.22.015982.

[40] E. Betzig, G. H. Patterson, R. Sougrat, O. W. Lindwasser, S. Olenych, J. S. Bonifacino, M. W. Davidson, J. Lippincott-Schwartz, and H. F. Hess, *Imaging intracellular fluorescent proteins at nanometer resolution,* science **313** (2006), pp. 1642–1645, doi: 10.1126/science.1127344.

[41] R. P. J. Nieuwenhuizen, K. A. Lidke, M. Bates, D. Leyton Puig, D. Grünwald, S. Stallinga, and B. Rieger, *Measuring image resolution in optical nanoscopy,* Nature Methods **10** (2013), pp. 557–562, doi: 10.1038/nmeth.2448.

[42] D. J. Nieves, J. A. Pike, F. Levet, D. J. Williamson, M. Baragilly, S. Oloketuyi, A. De Marco, J. Griffié, D. Sage, E. A. Cohen, *et al.*, *A framework for evaluating the performance of SMLM cluster analysis algorithms,* Nature Methods **20** (2023), pp. 259–267, doi: 10.1038/s41592-022-01750-6.

[43] S. L. Veatch, B. B. Machta, S. A. Shelby, E. N. Chiang, D. A. Holowka, and B. A. Baird, *Correlation functions quantify super-resolution images and estimate apparent clustering due to over-counting,* PloS One **7** (2012), p. e31457, doi: 10.1371/journal.pone.0031457.

[44] B. D. Ripley, *Modelling spatial patterns,* Journal of the Royal Statistical Society: Series B (Methodological) **39** (1977), pp. 172–192, doi: 10.1111/j.2517-6161.1977.tb01615.x.

[45] L. A. Masullo, F. Steiner, J. Zähringer, L. F. Lopez, J. Bohlen, L. Richter, F. Cole, P. Tinnefeld, and F. D. Stefani, *Pulsed interleaved MINFLUX,* Nano letters **21** (2020), pp. 840–846, doi: 10.1021/acs.nanolett.0c04600.

[46] K. Zhao, X. Xu, W. Ren, D. Jin, and P. Xi, *Two-photon MINFLUX with doubled localization precision,* eLight **2** (2022), pp. 1–10, doi: 10.1186/s43593-021-00011-x.

[47] J. Cnossen, T. Hinsdale, R. Ø. Thorsen, M. Siemons, F. Schueder, R. Jungmann, C. S. Smith, B. Rieger, and S. Stallinga, *Localization microscopy at doubled precision with patterned illumination,* Nature Methods **17** (2020), pp. 59–63, doi: 10.1038/s41592-019-0657-7.

[48] P. Jouchet, C. Cabriel, N. Bourg, M. Bardou, C. Poüs, E. Fort, and S. Lévêque-Fort, *Nanometric axial localization of single fluorescent molecules with modulated excitation,* Nature Photonics **15** (2021), pp. 297–304, doi: 10.1038/s41566-020-00749-9.

[49] C. P. Grabner, I. Jansen, J. Neef, T. Weihs, R. Schmidt, D. Riedel, C. A. Wurm, and T. Moser, *Resolving the molecular architecture of the photoreceptor active zone with 3D-MINFLUX,* Science Advances **8** (2022), p. eabl7560, doi: 10.1126/sciadv.abl7560.

[50] T. Deguchi, M. K. Iwanski, E.-M. Schentarra, C. Heidebrecht, L. Schmidt, J. Heck, T. Weihs, S. Schnorrenberg, P. Hoess, S. Liu, *et al.*, *Direct observation of motor protein stepping in living cells using MINFLUX,* Science **379** (2023), pp. 1010–1015, doi: 10.1126/science.ade2676.

[51] H. Heydarian, F. Schueder, M. T. Strauss, B. Van Werkhoven, M. Fazel, K. A. Lidke, R. Jungmann, S. Stallinga, and B. Rieger, *Template-free 2D particle fusion in localization microscopy,* Nature Methods **15** (2018), pp. 781–784, doi: 10.1038/s41592-018-0136-6.

[52] G. Tang, L. Peng, P. R. Baldwin, D. S. Mann, W. Jiang, I. Rees, and S. J. Ludtke, *EMAN2: an extensible image processing suite for electron microscopy,* Journal of Structural Biology **157** (2007), pp. 38–46, doi: 10.1016/j.jsb.2006.05.009.

[53] C. Sieben, N. Banterle, K. M. Douglass, P. Gönczy, and S. Manley, *Multicolor single-particle reconstruction of protein complexes,* Nature Methods **15** (2018), pp. 777–780, doi: 10.1038/s41592-018-0140-x.

[54] R. D. Gray, C. Beerli, P. M. Pereira, K. M. Scherer, J. Samolej, C. K. E. Bleck, J. Mercer, and R. Henriques, *Virusmapper: open-source nanoscale mapping of viral architecture through super-resolution microscopy,* Sci. Rep. **6** (2016), pp. 1–8, doi: 10.1038/srep2913.

[55] J. Broeken, H. Johnson, D. S. Lidke, S. Liu, R. P. Nieuwenhuizen, S. Stallinga, K. A. Lidke, and B. Rieger, *Resolution improvement by 3D particle averaging in localization microscopy,* Methods and Applications in Fluorescence **3** (2015), p. 014003, doi: 10.1088/2050-6120/3/1/014003.

[56] R. Henderson, *Avoiding the pitfalls of single particle cryo-electron microscopy: Einstein from noise,* Proceedings of the National Academy of Sciences **110** (2013), pp. 18037–18041, doi: 10.1073/pnas.1314449110.

[57] H. Heydarian, M. Joosten, A. Przybylski, F. Schueder, R. Jungmann, B. v. Werkhoven, J. Keller-Findeisen, J. Ries, S. Stallinga, M. Bates, and B. Rieger, *3D particle averaging and detection of macromolecular symmetry in localization microscopy,* Nature Communications **12** (2021), pp. 1–9, doi: 10.1038/s41467-021-22006-5.

[58] B. Hampoelz, A. Andres-Pons, P. Kastritis, and M. Beck, *Structure and assembly of the nuclear pore complex,* Annu. Rev. Biophys. **48** (2019), pp. 515–536, doi: 10.1146/annurev-biophys-052118-115308.

[59] D. H. Lin and A. Hoelz, *The structure of the nuclear pore complex (an update),* Annu. Rev. Biochem. **88** (2019), pp. 725–783, doi: 10.1146/annurev-biochem-062917-011901.

[60] S. Petrovic, G. W. Mobbs, C. J. Bley, S. Nie, A. Patke, and A. Hoelz, *Structure and function of the nuclear pore complex,* Cold Spring Harb. Perspect. Biol. **14** (2022), p. a041264, doi: 10.1101/cshperspect.a041264.

[61] G. Kabachinski and T. U. Schwartz, *The nuclear pore complex–structure and function at a glance,* Journal of Cell Science **128** (2015), pp. 423–429, doi: 10.1242/jcs.083246.

[62] J. Kosinski, S. Mosalaganti, A. Von Appen, R. Teimer, A. L. DiGuilio, W. Wan, K. H. Bui, W. J. Hagen, J. A. Briggs, J. S. Glavy, *et al.*, *Molecular architecture of the inner ring scaffold of the human nuclear pore complex,* Science **352** (2016), pp. 363–365, doi: 10.1126/science.aaf0643 .

[63] S. Mosalaganti, A. Obarska-Kosinska, M. Siggel, R. Taniguchi, B. Turoňová, C. E. Zimmerli, K. Buczak, F. H. Schmidt, E. Margiotta, M.-T. Mackmull, *et al.*, *AI-based structure prediction empowers integrative structural analysis of human nuclear pores,* Science **376** (2022), p. eabm9506, doi: 0.1126/science.abm9506.

[64] E. Hurt and M. Beck, *Towards understanding nuclear pore complex architecture and dynamics in the age of integrative structural analysis,* Current Opinion in Cell Biology **34** (2015), pp. 31–38, doi: 10.1016/j.ceb.2015.04.009.

[65] A. Ori, N. Banterle, M. Iskar, A. Andrés-Pons, C. Escher, H. Khanh Bui, L. Sparks, V. Solis-Mezarino, O. Rinner, P. Bork, *et al.*, *Cell type-specific nuclear pores: a case in point for context-dependent stoichiometry of molecular machines,* Mol. Syst. Biol. **9** (2013), p. 648, doi: 10.1038/msb.2013.4.

[66] S. J. Kim, J. Fernandez-Martinez, I. Nudelman, Y. Shi, W. Zhang, B. Raveh, T. Herricks, B. D. Slaughter, J. A. Hogan, P. Upla, *et al.*, *Integrative structure and functional anatomy of a nuclear pore complex,* Nature **555** (2018), pp. 475–482, doi: 10.1038/nature26003.

[67] V. Zila, E. Margiotta, B. Turoňová, T. G. Müller, C. E. Zimmerli, S. Mattei, M. Allegretti, K. Börner, J. Rada, B. Müller, *et al.*, *Cone-shaped hiv-1 capsids are transported through intact nuclear pores,* Cell **184** (2021), pp. 1032–1046, doi: 10.1016/j.cell.2021.01.025 .

[68] M. Stewart, *Polyadenylation and nuclear export of mRNAs,* Journal of Biological Chemistry **294** (2019), pp. 2977–2987, doi: 10.1074/jbc.REV118.005594.

[69] C. E. Wing, H. Y. J. Fung, and Y. M. Chook, *Karyopherin-mediated nucleocytoplasmic transport,* Nature Reviews Molecular Cell Biology **23** (2022), pp. 307–328, doi: 10.1038/s41580-021-00446-7.

[70] M. A. D'Angelo, *Nuclear pore complexes as hubs for gene regulation,* Nucleus **9** (2018), pp. 142–148, doi: 10.1080/19491034.2017.1395542.

[71] V. Le Sage and A. J. Mouland, *Viral subversion of the nuclear pore complex,* Viruses **5** (2013), pp. 2019–2042, doi: 10.3390/v5082019.

[72] O. Kerscher, P. Hieter, M. Winey, and M. A. Basrai, *Novel role for a Saccharomyces cerevisiae nucleoporin, Nup170p, in chromosome segregation,* Genetics **157** (2001), pp. 1543–1553, doi: 10.1093/genetics/157.4.1543.

[73] S. Sakuma and M. A. D'Angelo, *The roles of the nuclear pore complex in cellular dysfunction, aging and disease,* in *Seminars in Cell & Developmental Biology*, Vol. 68 (Elsevier, 2017) pp. 72–84, doi: 10.1016/j.semcdb.2017.05.006.

[74] A. P. Schuller, M. Wojtynek, D. Mankus, M. Tatli, R. Kronenberg-Tenga, S. G. Regmi, P. V. Dip, A. K. Lytton-Jean, E. J. Brignole, M. Dasso, *et al.*, *The cellular environment shapes the nuclear pore complex architecture,* Nature **598** (2021), pp. 667–671, doi: 10.1038/s41586-021-03985-3.

[75] S. Petrovic, D. Samanta, T. Perriches, C. J. Bley, K. Thierbach, B. Brown, S. Nie, G. W. Mobbs, T. A. Stevens, X. Liu, *et al.*, *Architecture of the linker-scaffold in the nuclear pore,* Science **376** (2022), p. eabm9798, doi: 10.1126/science.abm9798.

[76] A. Löschberger, S. van de Linde, M.-C. Dabauvalle, B. Rieger, M. Heilemann, G. Krohne, and M. Sauer, *Super-resolution imaging visualizes the eightfold symmetry of gp210 proteins around the nuclear pore complex and resolves the central channel with nanometer resolution,* Journal of Cell Science **125** (2012), pp. 570–575, doi: 10.1242/jcs.098822.

[77] A. Szymborska, A. De Marco, N. Daigle, V. C. Cordes, J. A. Briggs, and J. Ellenberg, *Nuclear pore scaffold structure analyzed by super-resolution microscopy and particle averaging,* Science **341** (2013), pp. 655–658, doi: 10.1126/science.1240672.

[78] R. Schmidt, T. Weihs, C. A. Wurm, I. Jansen, J. Rehman, S. J. Sahl, and S. W. Hell, *MINFLUX nanometer-scale 3D imaging and microsecond-range tracking on a common fluorescence microscope,* Nature Communications **12** (2021), p. 1478, doi: 10.1038/s41467-021-21652-z.

[79] S. C. Reinhardt, L. A. Masullo, I. Baudrexel, P. R. Steen, R. Kowalewski, A. S. Eklund, S. Strauss, E. M. Unterauer, T. Schlichthaerle, M. T. Strauss, *et al.*, *Ångström-resolution fluorescence microscopy,* Nature **617** (2023), pp. 711–716, doi: 10.1038/s41586-023-05925-9.

[80] Y.-L. Wu, P. Hoess, A. Tschanz, U. Matti, M. Mund, and J. Ries, *Maximum-likelihood model fitting for quantitative analysis of SMLM data,* Nature Methods **20** (2023), pp. 139–148, doi: 10.1038/s41592-022-01676-z.

2

JOINT REGISTRATION OF MULTIPLE POINT CLOUDS FOR FAST PARTICLE FUSION IN LOCALIZATION MICROSCOPY

[1]

We present a fast particle fusion method for particles imaged with single-molecule localization microscopy. The state-of-the-art approach based on all-to-all registration has proven to work well but its computational cost scales unfavourably with the number of particles N, namely as N^2. Our method overcomes this problem and achieves a linear scaling of computational cost with N by making use of the Joint Registration of Multiple Point Clouds (JRMPC) method. Straightforward application of JRMPC fails as mostly locally optimal solutions are found. These usually contain several overlapping clusters, that each consist of well-aligned particles, but that have different poses. We solve this issue by repeated runs of JRMPC for different initial conditions, followed by a classification step to identify the clusters, and a connection step to link the different clusters obtained for different initializations. In this way a single well-aligned structure is obtained containing the majority of the particles.

[1] This chapter has been published as: **Wenxiu Wang**, Hamidreza Heydarian, Teun A.P.M. Huijben, Sjoerd Stallinga, and Bernd Rieger, *Joint registration of multiple point clouds for fast particle fusion in localization microscopy*, Bioinformatics 38, no. 12 (2022): 3281-3287, doi: 10.1093/bioinformatics/btac320

2.1. INTRODUCTION

The diffraction of light limits the resolution of conventional microscopy to about 200 nm. Several super-resolution microscopy techniques enable "diffraction unlimited" resolution [1–3]. Single-molecule localization microscopy (SMLM) is a widely used member of the family of super-resolution techniques, and obtains super-resolved images by localizing single fluorescent emitters. The resolution of these super-resolved images is not infinite, but in practice restricted to about 20 nm due to the incomplete fluorescent labelling and a limited number of collected photons per localization event [4]. In recent years, significant improvements have been made to increase the photon count per localization [5]. Increasing the density of labelling (*DOL*) using biochemical means is difficult, where *DOL* values of around 50% are typically achieved. In addition, a high local *DOL* can lead to an increased rate of mislocalizations [6] which is detrimental for the quality of the imaging process. If the sample includes many chemically identical bio-complexes (called particles in the following), the limitation imposed by a low *DOL* can be lifted by fusion of all these particle into one single reconstruction, the so-called super-particle, leading to a much better resolution and signal-to-noise ratio (SNR) [7, 8]. This approach by particle fusion, of course, ignores potential heterogeneity in the underlying biology within the collection of particles. Template-driven particle fusion methods have been used [7–10], but have a substantial risk of resulting in a biased reconstructed structure. Heydarian *et al.* proposed a template-free particle fusion method based on an all-to-all registration (all-to-all method in short), which is robust against underlabelling and misregistration [11, 12]. The all-to-all method has proven to work well and produces reconstruction resolutions down to a few nanometers. Despite this success, computational times of around a day for a number of particles N exceeding about 1000 are not uncommon and are only feasible with the use of GPU acceleration. The root cause lies within the unfavourable scaling of computational cost with N^2, because each particle is registered to all other particles, resulting in $N(N-1)/2$ registration pairs. The all-to-all method has another drawback, the so-called "hot-spot" problem. For symmetric structures, random variations in the localization data with binding site are amplified by the pair-wise optimal registration process. Heydarian *et al.* solved this problem by first detecting the present symmetry and then imposing it on the data in a post-processing step. Thus, a particle fusion algorithm that is fast and which avoids the hot-spot artifact is desired.

An alternative to the all-to-all method is based on the Joint Registration of Multiple Point Clouds (JRMPC) method [13]. In the JRMPC method, particles are iteratively rotated and translated to fit to a Gaussian Mixtures Model (GMM), which is updated itself in each iteration round. The key advantage of the JRMPC method is that the computational complexity scales linearly with the number of particles N, which makes it inherently faster than the all-to-all method if N grows large. In addition, hot-spot artifacts in symmetric structures are avoided without imposing (a-priori) symmetry information, because the joint registration treats each particle equally. There are, however, major drawbacks to the JRMPC method. First, the outcome of the JRMPC turns out to be highly susceptible to the initialization of the GMM (number of Gaussians, center positions and widths). Different initial settings of the GMM parameters lead to different sets of final estimated particle rotations and translations. Second, the final outcome usually consists of several clusters, where the particles within the clusters are well-registered, but where the clusters have dif-

ferent poses. We attribute these issues with robustness of the algorithm to trapping in local optima of the iterative optimization (outlined in section 2.2.1 in detail).

The goal of the work presented in this paper is to overcome the robustness problems of the JRMPC method while maintaining the inherent speed advantage. To this end, we propose a processing pipeline in which we combine JRMPC registration outcomes obtained with different GMM initializations using cluster analysis tools. The cluster analysis uses our recent unsupervised classification framework [14], which is based on the Bhattacharya distance metric [9] together with multi-dimensional scaling (MDS) [15] and k-means clustering [16, 17]. The process of JRMPC and classification is repeated several times for different GMM initializations. Pairs of clusters from different initializations may share particles. The relative poses of such particles in different clusters is used in a final step to combine the different clusters into a single well-aligned structure.

2.2. METHOD

Our proposed algorithm has three main steps, illustrated in Figure 2.1. The steps are (1) alignment of particles with JRMPC using multiple initializations, (2) classification of JRMPC registered particles into clusters, and (3) connection of the identified clusters into a single final reconstruction.

The input data is a union of particles $\mathbf{A} = \{\mathbf{A}_j\}_{j=1}^{N}$, with N the number of particles. Each particle is characterized by a set of localization coordinates $\mathbf{V}_j$ and attendant localization uncertainties Δ_j as $\mathbf{A}_j = \{\mathbf{V}_j; \Delta_j\}$. The coordinates of particle j represent M_j localizations:

$$\mathbf{V}_j = [\mathbf{v}_{j1} \dots \mathbf{v}_{ji} \dots \mathbf{v}_{jM_j}] \in \mathbb{R}^{d \times M_j},$$

where the $\mathbf{v}_{ji}$ are vectors with elements equal to the d coordinates of the i-th localization in particle j. Depending on the data, the dimensionality d can be 2 or 3. In general, the localization uncertainties of the M_j localization events in particle j are:

$$\Delta_j = [\Sigma_{j1}, \dots \Sigma_{ji} \dots \Sigma_{jM_j}] \in \mathbb{R}^{d \times d \times M_j},$$

where the Σ_{ji} are $d \times d$ matrices equal to the covariance matrices of the i-th localization in particle j. Often a more simple description of the localization uncertainty is possible. For 2D data for example, the uncertainties are isotropic, and Δ_j can be written as:

$$\Delta_j = \left[\delta_{j1}, \dots \delta_{ji} \dots \delta_{jM_j}\right],$$

where the δ_{ji} are now scalar values that represent the localization uncertainty in the xy plane for the i-th localization in particle j. For most 3D data, Δ_j is represented as:

$$\Delta_j = \left[\delta_{j1}, \tau_{j1}; \dots, \delta_{ji}, \tau_{ji}; \dots \delta_{jM_j}, \tau_{jM_j}\right],$$

where now τ_{ji} is the localization uncertainty along the z-axis for the i-th localization in particle j. This axial localization uncertainty is typically larger than the uncertainty in the xy plane [18].

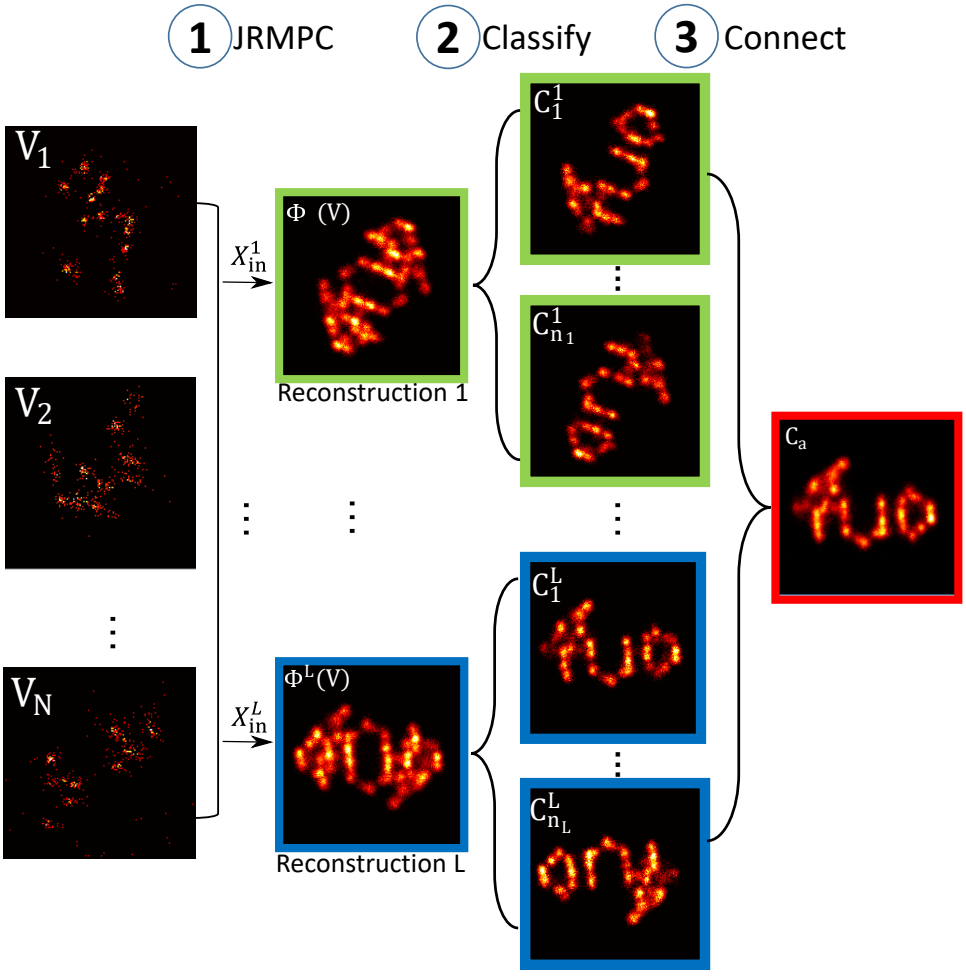


Figure 2.1: The three main steps of the proposed particle fusion algorithm. Step 1: Use JRMPC [13] to initially align N input particles $\mathbf{V} = \{\mathbf{V}_j\}_{j=1}^{N}$ *with L random initializations of the GMM* $\{\mathbf{X}_{\text{in}}^{l}\}_{l=1}^{L}$ *leading to L different reconstructions* $\{\Phi^{l}(V)\}_{l=1}^{L}$. *Step 2: Apply the unsupervised classification method of Huijben et al. [14] to classify each reconstruction* $\Phi^{l}(V)$ *into* n_l *clusters* $\{C_n^l\}_{n=1}^{n_l}$ *separating different overlapping poses in the reconstructed particles. Step 3: Connect particles from different clusters into the final super-particle reconstruction* C_a, *such that each input particle is present at most once.*

2.2.1. ALIGNMENT

The structure of the reconstruction is characterized in the JRMPC method by a GMM with parameters $\mathbf{G} = \{\mathbf{G}_k\}_{k=1}^{K}$, where each of the K Gaussians components $\mathbf{G}_k = [p_k, \boldsymbol{\mu}_k, \sigma_k]$ has a mixing coefficient (weight) p_k, a set of d coordinates $\boldsymbol{\mu}_k$ that represent the mean of the Gaussian, and a standard deviation σ_k (an isotropic covariance matrix $\sigma_k^2 \mathbf{I}_d$ is taken). The GMM parameters have an initial setting $\mathbf{G}_{\text{in}}$, described in section 2.3.2. The parameters that are updated during the iterative JRMPC algorithm are:

$$\Theta = \left\{ \{\mathbf{G}_k\}_{k=1}^{K}, \{\mathbf{R}_j, \mathbf{t}_j\}_{j=1}^{N} \right\}, \tag{2.1}$$

where $R_j \in \mathbb{R}^{d\times d}$ is the rotation applied to particle j and where $t_j \in \mathbb{R}^{d\times 1}$ is the translation applied to particle j. The coordinates of the reconstruction are then:

$$\Phi(\mathbf{V}) = \{\mathbf{R}_j \mathbf{V}_j + \mathbf{t}_j\}_{j=1}^{N}, \tag{2.2}$$

which thus contains the coordinates of all localization events in all particles. It is noted that the localization uncertainties are not taken into account in the JRMPC method. Further details on the steps in each iteration round of the JRMPC are given in Appendix 2.7.4.

The outcome of the JRMPC depends on the choice of the initial GMM centers in $\mathbf{G}_{\text{in}}$. Our algorithm uses L differently initialized GMMs $\{\mathbf{G}_{\text{in}}^{l}\}_{l=1}^{L}$, leading to L different JRMPC alignments $\Phi(\mathbf{V}) = \{\Phi^{l}(\mathbf{V})\}_{l=1}^{L}$ of the same union of particles with coordinates $\mathbf{V}$.

2.2.2. CLASSIFICATION

The JRMPC algorithm can end up in a local optimum, resulting in multiple groups of particles (clusters) with different overlapping poses in the reconstruction. To separate these clusters, we use an unsupervised classification method recently proposed by our group [14]. This method enables the analysis of structural heterogeneity in localization datasets arising from e.g. naturally occurring biological variations. Here, we use this pipeline to decompose the L different JRMPC outcomes into clusters of particles, where the particles within each cluster are well-aligned. First, we compute the normalized Bhattacharya cost function between every transformed particle $\Phi_a^l(\mathbf{V}_a)$ and every other transformed particle $\Phi_b^l(\mathbf{V}_b)$ within the JRMPC registration for each initialization $l = 1, 2, \ldots, L$. This one time computation gives an upper triangular matrix with $N(N-1)/2$ cost function values S. The normalized Bhattacharya cost is in general given by the sum over the M_a localizations of particle a and M_b localizations of particle b as:

$$S(a,b) = \frac{1}{M_a M_b} \sum_{q=1}^{M_a} \sum_{r=1}^{M_b} \frac{1}{\sqrt{\det \Omega_{qr}^{ab}}} \quad \exp\left(-\frac{1}{2} \delta\phi_{qr}^{ab\,T} \Omega_{qr}^{ab} \delta\phi_{qr}^{ab} \right). \tag{2.3}$$

Here $\delta\phi_{qr}^{ab} = \phi(\mathbf{v}_{aq}) - \phi(\mathbf{v}_{br})$ is the difference in transformed (rotated and translated) coordinates of localization q of particle a and localization r of particle b, and Ω_{qr}^{ab} is defined in terms of the uncertainty covariance matrices of the localizations as:

$$\Omega_{qr}^{ab} = \Sigma_{aq} \left(\Sigma_{aq} + \Sigma_{br} \right)^{-1} \Sigma_{br}. \tag{2.4}$$

For example, for 2D-data with isotropic localization uncertainties, this reduces to:

$$S(a,b) = \frac{1}{M_a M_b} \sum_{q=1}^{M_a} \sum_{r=1}^{M_b} \frac{1}{(\delta_{aq}^2 + \delta_{br}^2)} \exp\Big(- \frac{(\phi(\mathbf{v}_{aq}) - \phi(\mathbf{v}_{br}))^2}{2\left(\delta_{aq}^2 + \delta_{br}^2\right)} \Big). \tag{2.5}$$

The normalization of the cost function with the numbers of localizations per particle reduces the impact of the variations in these number, which makes it a better descriptor of the similarity between the structure of the particles. The next step is to transfer dissimilarity values:

$$D(a,b) = \max(S) - S(a,b) \tag{2.6}$$

to spatial coordinates in a multidimensional space suitable for classification using multidimensional scaling (MDS) [15]. The transformed particles will then be partitioned into clusters by k-means clustering [16, 17] in this multidimensional space. Parameter settings for the classification step are given in section 2.3.2. This process is repeated for the JRMPC reconstructions $l = 1,2,\ldots,L$ leading to $n = 1,2,\ldots,n_l$ clusters that are denoted as C_n^l (see Figure 2.1).

2.2.3. CONNECTION

As we repeat the JRMPC reconstruction L times, pairs of clusters from different initializations may share different particles. Therefore, we need to combine the different clusters into a single well-aligned structure. In a first step we discard clusters with less than ϑ particles. This threshold helps to filter out poorly aligned clusters as well as clusters with particles of poor quality, as these tend to accumulate in clusters with low number of particles.

Next, the cluster with the largest number of particles is selected as initial estimate of the super-particle reconstruction $\mathbf{C}_a$. This main cluster, $\mathbf{C}_m^{l_m}$, is used as the target for a pairwise comparison of clusters. A loop over all clusters $\mathbf{C}_n^l$ for $l \neq l_m$ is done, and clusters $\mathbf{C}_n^l$ and $\mathbf{C}_m^{l_m}$ are compared to check for particles that are in both clusters. If there exists at least one common particle c with coordinates $\mathbf{V}_c \in \mathbf{V}$ then the clusters $\mathbf{C}_n^l$ can be added to the super-particle reconstruction estimate $\mathbf{C}_a$ following:
Step 1: apply the inverse transformation of particle c in cluster $\mathbf{C}_n^l$ to transform all particles in the cluster $\mathbf{C}_n^l$ to the original position and pose of $\mathbf{V}_c$:

$$\mathbf{C}_n^l|_{\mathbf{V}_c} = \{\mathbf{R}_c^l\}^{-1}\mathbf{C}_n^l - \mathbf{t}_c^l. \tag{2.7}$$

Step 2: apply the transformation of particle c in the main cluster $\mathbf{C}_m^{l_m}$ to all particles in the cluster $\mathbf{C}_n^l|_{\mathbf{V}_c}$ to the position and orientation of $\mathbf{C}_m^{l_m}$:

$$\mathbf{C}_n^l|_{\mathbf{C}_m^{l_m}} = \mathbf{R}_c^{l_m}\mathbf{C}_n^l|_{\mathbf{V}_c} + \mathbf{t}_c^{l_m}. \tag{2.8}$$

Now that the cluster $\mathbf{C}_n^l$ is aligned with the pose of the main cluster $\mathbf{C}_m^{l_m}$ the particles of $\mathbf{C}_n^l$ can be added to the super-particle reconstruction estimate $\mathbf{C}_a$. In this way more and

more particles accumulate in the final reconstruction, yielding the final outcome of our proposed algorithm.

Care must be exercised for two subtleties. First, it can happen that there is more than one common particle between the two clusters $\mathbf{C}_n^l$ and $\mathbf{C}_m^{l_m}$. Then, if there exists more than one common particles between two clusters, we will calculate all the common particles' translation matrices and rotation matrices from the cluster $\mathbf{C}_a^{l_1}$ to the cluster $\mathbf{C}_a^{l_1}$,

$$\mathbf{t}_c|_{\mathbf{C}_a^{l_1}\rightarrow\mathbf{C}_b^{l_2}} = \mathbf{t}_c^{l_2} - \mathbf{R}_c^{l_2}\{\mathbf{R}_c^{l_1}\}^{-1}\mathbf{t}_c^{l_1}\,, \tag{2.9}$$

$$\mathbf{R}_c|_{\mathbf{C}_a^{l_1}\rightarrow\mathbf{C}_b^{l_2}} = \mathbf{R}_c^{l_2}\{\mathbf{R}_c^{l_1}\}^{-1}\,, \tag{2.10}$$

then we compare all the $\mathbf{t}_c$ and $\mathbf{R}_c$ and use the common particle with rotation and translation matrix that are closest to the median of all translation and rotation matrices of all the common particles. Second, we must check if the particles of cluster $\mathbf{C}_n^l$ are not already in the reconstruction estimate $\mathbf{C}_a$. Only the unique particles that are not already contained in the reconstruction are added to $\mathbf{C}_a$. In Appendix 2.7.2 we give pseudo code for this connection pipeline.

2.3. EXPERIMENTS

2.3.1. EXPERIMENTAL DATA

We applied our method to four different localization microscopy experiments described here:

DNA origami TUD-logo: We tested three different 2D TUD-logo DNA origami datasets [11] with *DOL* of 30%, 50% and 80%. We compared the results of the currently proposed method and the all-to-all method [11] in Figure 2.2. The data is available online [19].

2D nuclear pore complex: We further applied our method to 2D Nuclear Pore Complex (NPC) data which were previously described in ref. [7]. In Figure 2.4, we show our reconstruction of NPCs together with the reconstruction of the all-to-all method [11] to compare the methods' capabilities in the reconstruction of symmetrical structures.

3D nuclear pore complex: We applied our algorithm to 3D NUP107 NPC data [12] acquired by two different localization microscopy techniques. The data is available online [20]. The poses of the NPCs are experimentally constrained as they are all embedded in the nuclear envelope which is imaged as flat as possible on the cover glass. The lower and upper ring of all particles are therefore roughly perpendicular to the optical axis of the microscope [12].

DNA origami Digits data: The so-called nanoTRON datasets [21] consist of DNA origami structures in the shape of the digits 1, 2, and 3 and in the shape of a 3×4 rectangular grid. The data is available online [22] and contains on the order of a few thousand particles. These datasets are used to showcase the processing speed advantages of our method.

Simulation data Simulation data of the DNA-origami TUD-logo was generated as described in ref. [11].

2.3.2. PARAMETER SETTINGS

A number of parameters in the three algorithmic steps of alignment, classification and connection must be set. The default values given in the table in the Appendix 2.7.3 are

suitable for most of the cases.
We estimate the number of initial GMM centers K by applying the mean-shift method [16, 23] to the outcome of ζ randomly selected input particles coarsely transformed by JRMPC with K_0 randomly generated GMM centers. We set $K_0 = \min\left((\Sigma_{j=1}^{N} M_j)/N, 100\right)$, i.e. equal to the average number of localizations of all input particles with a minimum of 100. We choose $\zeta = 20$, if the number of input particles $N < 20$ then $\zeta = N$. The value of K estimated in this way is approximately equal to the number of binding sites in most cases. All initial values for the prior probabilities of the K Gaussians are set uniformly to $p_k = 1/K$. The initial values of the center positions μ_k^0 are generated randomly within a rectangular bounding box containing all the localizations. We initialize the transformation as $\mathbf{R}_j^0 = \mathbf{I}_d$ and $\mathbf{t}_j^0 = \bar{\boldsymbol{\mu}}^0 - \bar{\mathbf{v}}_j$, where $\bar{\boldsymbol{\mu}}^0$ is the average of the K GMM centers. The diagonal of the bounding box containing all the input particles after applying the initial translation is set as the initial value of all Gaussian standard deviations σ_k^0. We set the default value for the number of clusters n_l to 2 in the classification step because the registration of JRMPC usually only contains two flipped structures. The threshold ϑ for a cluster to be used in the connection step is set as $N/(n_l + 1)$. The default number of repetitions L for the JRMPC initializations is 2.

We use the default parameter settings throughout with two exceptions. The reconstruction of the nanoTRON 3×4 grid (Figure 2.3(d) uses non-default parameters with a larger number of clusters ($n_l = 8$) to guarantee clusters that contain well-aligned particles. The reconstruction of the 3D NPC particles (Figure 2.5) uses a non-default value for the initial Gaussian standard deviation (we use $\sqrt{1000}$, much smaller than the default value) to better fit with the limited range of initial poses of the NPCs. An inferior alignment is observed with the default value. In general we find that the quality of the individual clusters can be improved by increasing n_l or ϑ. A larger number of JRMPC initializations L can help to increase the number of particles in the final reconstruction after the connection step.

2.3.3. BENCHMARK ALGORITHMS AND EVALUATION METRICS

We compare our proposed method with the all-to-all method [11] [12]. We use the Fourier Ring Correlation (FRC) [4] to measure the resolution of the super-particle reconstructions. We form two independent input image subsets from the super-particle reconstruction to perform the FRC analysis. The first subset is the main cluster $C_m^{l_m}$ and the second subset consists of all other particles in the reconstruction. These two subsets can be used as statistically independent image subsets that are the necessary inputs for the FRC measurement because each subset contains a similar number of different particles from different independent experiments. We cross-checked the outcomes of this FRC computation with the standard method of independently processing two subsets of the total set of input particles and found outcomes within the uncertainty margin of the FRC estimation. In addition, we calculate the localization distribution over the azimuthal angles to analyze the reconstruction symmetry for symmetrical structures. For the 3D NPC data, we also visualize and compare the distributon of z positions of the localizations, the radius of each of the two rings, and in a rose plot the localization distribution over azimuthal angles. In the simulations, we compute the root mean square distance between the localizations after particle fusion and the attendant binding sites to quantify the quality of the fusion

process [12].

2.4. RESULTS

2.4.1. COMPUTATIONAL COST

Compared to the all-to-all method, which has an unfavorable computational cost scaling as N^2, our method is much faster as it is linear with N. Figure 2.2(a-b) shows the reconstructions of 383 experimental TUD-logo particles with DOL=80% and 788,875 localizations obtained with the all-to-all method and our method. We repeated our method on the 80% DOL TUD-logo particles 30 times in order to assess the uncertainty in FRC-resolution and computation time. Both methods achieve a similar reconstruction quality, consistent with near equal FRC resolutions (3.3±0.3 nm for the single instance of the all-to-all, 3.6±0.3 nm for the 30 runs for our method). The computational time of the all-to-all method, however, is almost 12 times longer than for our method. More importantly, our computational time of 9.6±0.6 minutes was performed on a simple CPU (40 core Xeon E5-2670v3), opposed to the GPU-implementation of the all-to-all registration (K40c Tesla GPU). The all-to-all method is practically impossible on a CPU when having more than 100 particles. The estimated number of Gaussian centers K is 40±3, which is close to the actual number of binding sites (37). The random initializations of the JRMPC usually result in a final GMM that is similar to the combination of two inverted TUD-logos, which can be classified appropriately in only two clusters.
Our method can effectively handle large amounts of particles because of the favorable reconstruction speed. To show the capability of our method to handle this large data we applied it to the nanoTRON datasets, which contain an order of magnitude more particles than the TUD-logo datasets. We achieved clear structures of the digits 1, 2, and 3 and of the 3×4 grid in only 1.1 h, 1.3 h, 45 min. and 4.8 h, respectively, in CPU compared to a computational time of multiple days for the GPU-accelerated all-to-all method. It would have taken several days to resolve the full dataset with the all-to-all method. Due to this speed limitation we only used part of the data in the all-to-all method. The FRC resolution obtained by the all-to-all registration for these four datasets (digits 1, 2, and 3 and of the 3×4 grid) containing 1219, 1309, 1278 and 1194 particles are 3.69 ± 0.02 nm, 4.40 ± 0.19 nm, 3.98 ± 0.22 nm and 3.59 ± 0.15 nm, respectively [14]. Our reconstructions include 4155, 4943, 2541 and 5961 particles for these four datasets and the FRC resolutions are 2.76 ± 0.92 nm, 2.80 ± 0.54 nm, 3.21 ± 0.33 nm and 3.51 ± 0.28 nm, respectively. These numbers are smaller as we are able to assemble more particles in the final reconstruction compared to the all-to-all method. For the digits 1, 2, and 3, the estimated K (25, 23, 34) is close to the actual number of binding sites (18, 23, 25). For the 3×4 grid particles, our K-estimation algorithm estimates $K = 42$ which is much more than the 12 binding sites. For that reason the JRMPC reconstructions have more clusters and we need a larger $n_l = 8$ to separate them correctly.

2.4.2. 2D NPC DATA: INFLUENCE OF SYMMETRY

Our method also overcomes the second disadvantage of the all-to-all method, the hot-spot problem occurring for symmetrical structures. In Figure 2.4, we compare reconstructions of 2D NPC particles with eight-fold rotational symmetry. The reconstruction

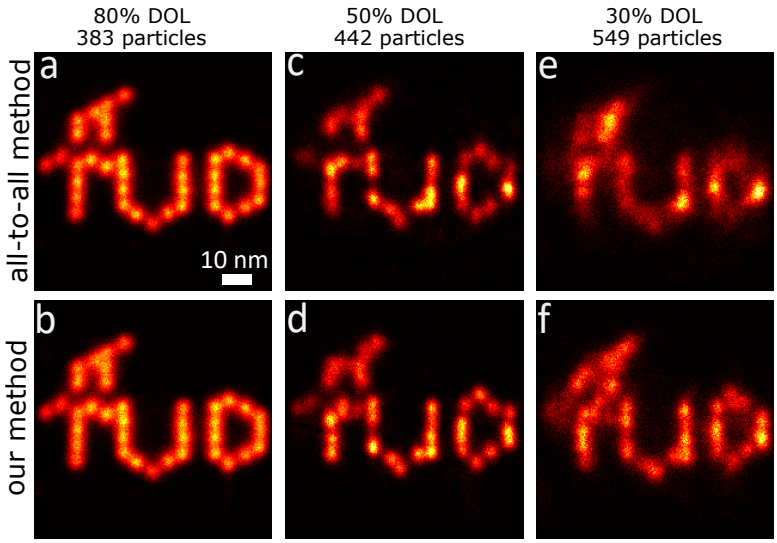


Figure 2.2: Comparison of the particle fusion performance with our method and the all-to-all method on experimental 2D TUD-logo DNA origami particles. (a,c,e) Reconstruction by all-to-all registration (FRC resolution of 3.3, 3.5, 5.0 nm for 80%, 50% and 30% DOL respectively. Computational time for (a) is about 2 hours (GPU)). (b,d,f) Reconstruction by our method (FRC resolution of 3.2, 3.1, 3.3 nm for 80%, 50% and 30% DOL respectively. Computational time for (b) is about 9.5 minutes (CPU)). Scale bar of (a) applies to (b-f).

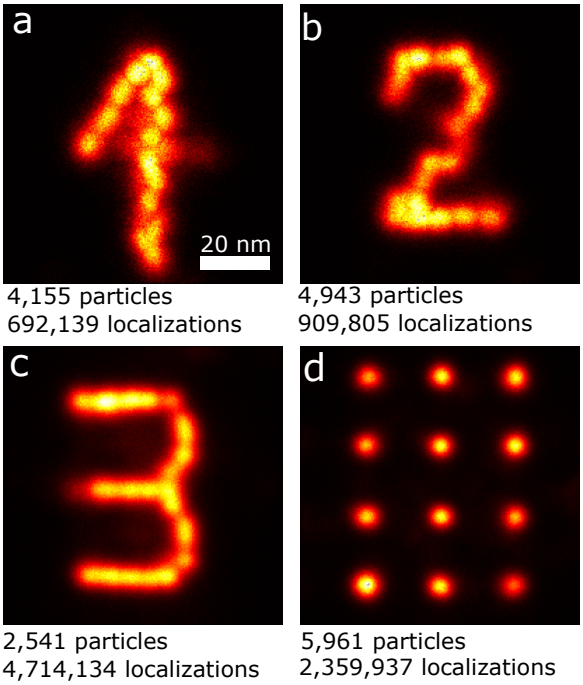


Figure 2.3: Particle fusion speed for experimental 2D DNA-origami with a large number of particles. (a) Reconstruction of digit 1, computational time 1.1 h (CPU). (b) Reconstruction of digit 2, computational time 1.3 h (CPU). (c) Reconstruction of digit 3, computational time 48 min (CPU). (d) Reconstruction of 3×4 *grid, computational time 4.8 h (CPU). The number of particles and localizations in each reconstruction are indicated below the figures. Scale bar of (a) applies all sub-images.*

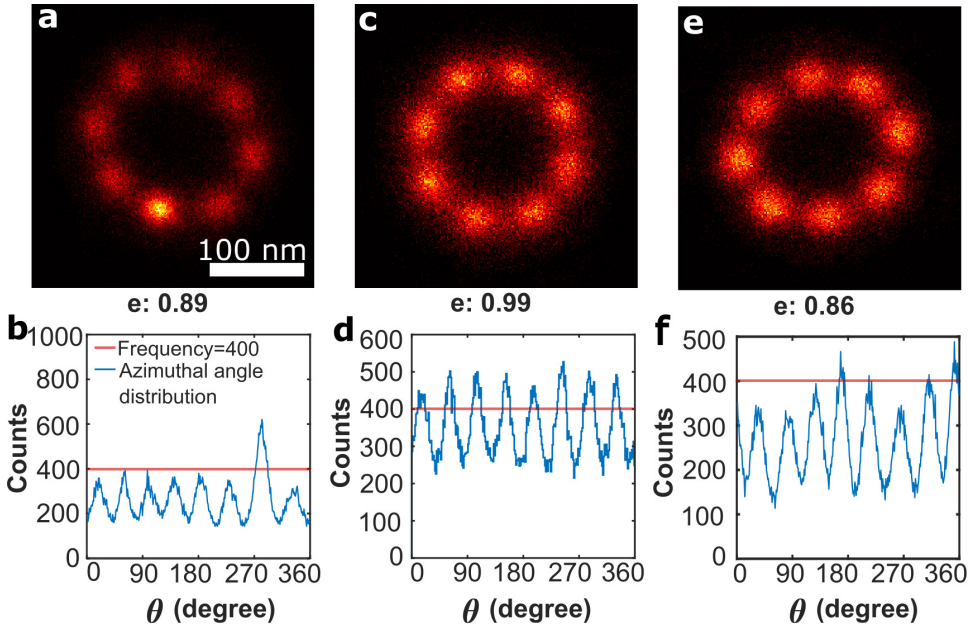


Figure 2.4: Comparison of particle fusion performance between our method and the all-to-all method on 304 experimental 2D nuclear pore complex particles. (a) Reconstruction with the all-to-all method without prior knowledge. A "hot-spot" is visible due the enhancement by pair-wise registration. Fitted ellipticities e to the reconstruction are shown below. (b,d,f) Histogram of the azimuthal angles of the localizations in (a,c,e) respectively; for comparison, a red line indicates 400 counts. (c) Reconstruction with the all-to-all method after explicitly imposing eight-fold symmetry. (e) Reconstruction with our method without prior knowledge. Even without imposing symmetry no hot-spot occurs. Scale bar of (a) applies to (c,e).

of the all-to-all method without prior knowledge (Figure 2.4(a)) and (b)) shows one apparent "hot-spot" with more than 600 localizations compared to other blobs with around 400 localizations. After imposing eight-fold rotational symmetry the hot-spot disappears (Figure 2.4(c)). Imposing this symmetry changes the ellipticity of the reconstructed NPC ring from the earlier 0.89 to 0.99. So, symmetry has been restored, but at the expense of a shape that changed from an ellipse to a circle. Our method applied to the same NPC particles does not result in a hot-spot (Figure 2.4(e)), quantified by a more uniform distribution of localizations over the 8 peaks (compare (b) and (f)). The ellipticity of our reconstruction is 0.86 which matches reasonably well with the all-to-all value of 0.89. The number of Gaussian components K in the GMM is estimated by our algorithm to be 8 which is obviously equal to the number of visible binding sites in the 2D NPC.

2.4.3. LOW LABELLING 2D DNA ORIGAMI DATA

A major accomplishment of the all-to-all method is its ability to handle poorly labelled data. It appears our method outperforms the all-to-all method even in this respect. Figure 2.2(c-f) shows a comparison of reconstructions of hundreds of TUD-logos with low *DOL* values equal to 50% and 30%. Our method results in a visually better reconstruction quality, especially for the worst quality *DOL*=30% dataset (compare Figure 2.2(e) and (f)). Nearly all binding sites on the origami at a distance of about 5 nm are resolved in (f) where in (e) especially the edges are washed out and localizations are concentrated to a few binding sites. This is consistent with the FRC resolutions of 3.1 nm and 3.3 nm for the 50% and 30% *DOL* datasets, respectively, which compares favourably with the FRC resolutions for the all-to-all method equal to 3.5 nm and 5.0 nm for the 50% and 30% *DOL* datasets, respectively. The mean-shift method estimates $K = 46$ for the data with 30% *DOL* and $K = 37$ for 50% *DOL*. These two K values are very close to the actual number of 37 binding sites of the origami design. The initial Gaussian standard deviation is quite large (~ 100 nm) at first. Most of the Gaussian components shrink to a small size (less than 3 nm) eventually, and only a few to a medium size (~ 10 nm). Most of the initially randomly generated GMM centers $\boldsymbol{\mu_k}$ are finally positioned near the binding sites of the TUD-logo.

2.4.4. 3D NPC DATA

Another major achievement of the all-to-all method is the ability to reconstruct 3D data [12]. Our method shows a comparably good performance on 3D datasets. Figure 2.5 shows a comparison of 3D Nup107 NPC structures imaged with both PAINT and STORM. Our method shows reconstructions of similar quality as the all-to-all method (compare Figure 2.5(a) and (k) and compare Figure 2.5(f) and (p)). Here, the all-to-all method relies on detecting the rotational symmetry from the data and subsequently promoting the symmetry in the reconstruction. In contrast, neither prior knowledge or detection of symmetry nor extra post-processing is needed with our method. Comparison of Figure 2.5(b,g,l,q), (c,h,m,l) to (d,i,n,s), respectively, further shows that our method obtains similar NPC structural parameters (the distance between the nuclear and cytoplasmic rings and their radius) as the all-to-all method. The rose plots Figure 2.5(e,j) obtained from the all-to-all method's reconstructions show eight-fold symmetry for each ring, and the number of localizations in each peak is almost the same. The rose plots Figure 2.5(o,t) of our reconstructions also clearly show eight peaks for each ring, but the number of localizations in each peak is slightly different. This is reasonable considering that our method does not rely on symmetry in the reconstruction. Our K-estimation algorithm estimates $K = 34$ for both cases, which is also reasonable as the number of actual binding sites should be 32 given the structure of the EM model [24, 25]. The default value of σ_k does not work here and we used $\sigma_k^0 = \sqrt{1000}$ nm instead. The final center points of the GMMs are nearly all distributed inside the 16 spheres of the 3D NUP reconstructions.

2.4.5. SIMULATION DATA

We explore the limitations of the proposed method in terms of *DOL*, localization precision and the number of particles by applying our method on simulated TUD-logo datasets. Further details and results can be found in Appendix 2.7.4. Even with poor quality datasets (*DOL* as low as 40%, or localization precision as low as 12 nm, or a number of particles as

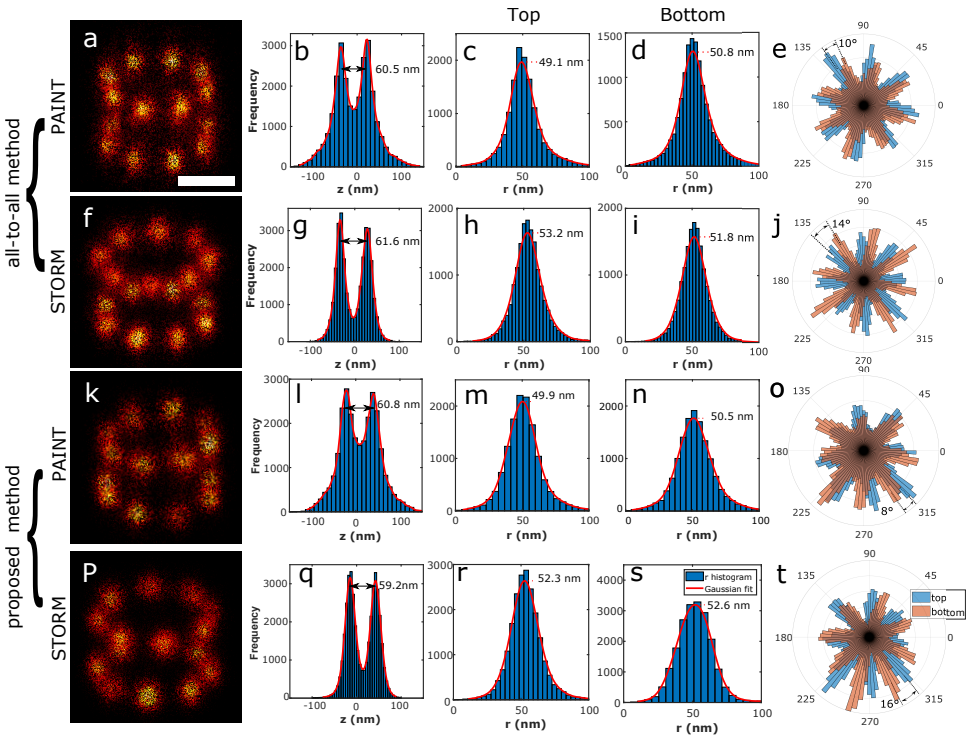


Figure 2.5: Comparison of particle fusion performance between our method and all-to-all method on experimental 3D Nup107 particles acquired by different localization microscopy techniques. (a) Fusion of 306 Nup107 particles obtained from 3D astigmatic PAINT reconstructed by the 3D all-to-all method. (b,g,l,q) Histogram of the z coordinate of localizations in the reconstruction (a). (c,h,m,r) Histogram of the radius of upper ring's localizations, (d,i,n,s) lower ring. (e,j,o,t) Rose plot of the localization distribution over azimuthal angles for the upper and lower rings of the reconstructions. (f) Fusion of 356 Nup107 particles obtained from 3D astigmatic STORM reconstructed by the 3D all-to-all method. (k) Fusion of 306 Nup107 particles obtained from 3D astigmatic PAINT reconstructed by our method. (p) Fusion of 356 Nup107 particles obtained from 3D astigmatic STORM reconstructed by or method. Scale bar indicates 50 nm and applies to a,f,k and p. Rose plots in (e,j) show 8 fold symmetry with nearly equal number of localizations, but symmetry was used here in the reconstruction. Rose plots (o,t) Without any prior knowledge reconstruction with our method also shows 8 clear peaks however with a stronger variation in the number of localizations.

low as 10), our method can obtain reconstructions with registration error in the range 5 to 10 nm.

2.5. DISCUSSION

Several of the results we obtained can be qualitatively understood: In comparison to the all-to-all-method our approach produces better results for poor, underlabeled data. The reason is that in the pairwise registration of the all-to-all method pairs of poor quality particles must be aligned, which is more error prone than our approach where each of the poor quality particles is aligned to the average of all particles. The same line of reasoning applies to the case of symmetric structures. The pairwise registration of the all-to-all method aligns random peaks that occur through the stochastic variations of labeling within the particles, while for our approach each particle is aligned to the average of all particles which smoothens out the stochastic variations in labeling.

We attribute the JRMPC local optima that consist of several distinct clusters with different poses to a difference in convergence rate between the widths of the Gaussian components and the particle rotations. It seems that the Gaussian widths shrink relatively fast, while the particle rotations only change slowly, as the iteration progresses. This results in posterior probabilities α_{kij} for the Gaussian component k that is nearest to localization i of particle j that quickly converge to nearly one and to virtually zero for the other Gaussian components. On the other hand, for the case of 3D NPC particles with a limited range of poses in the dataset, the widths of the Gaussian components appear too large, leading to sets of particle rotations that are distributed too broadly. Summarizing, the reconstruction quality appears to be sensitive to the initial setting and convergence rate of the Gaussian widths.

Next to the limitations of our method on *DOL* and localization uncertainty assessed by the simulation study, there are also some assumptions that go into the proposed method that we wish to emphasize now. Firstly, the underlying biological structure is assumed to be sufficiently rigid for the overall averaging to make sense. Secondly, we assume a single underlying structure. Recently, however, we have studied the detection of structural heterogeneity using particle fusion methods ([14]). Our method can also vastly accelerate the workflow of [14], as there the initial step is to find a global alignment of all particles in the data set, followed by quantification of pairwise (dis)similarity between the particles. Thirdly, the idea of fitting a GMM to localization microscopy data sets matches well structures with a discrete number of binding sites for fluorophores and corresponding data sets with multiple localizations per binding site. In that case the GMM centers will tend to gravitate towards the different binding sites. In case there are only a few localizations per binding site, however, it will be difficult to match the GMM centers to the binding sites, and the quality of the registration process may be compromised.

A number of algorithmic improvements can be envisioned. First of all we could incorporate the localization uncertainties in the JRMPC method, such that the probability of localization i of particle j to fit Gaussian component k is a normal distribution with a variance that is the sum of the variance due to the localization uncertainty and the variance of the Gaussian component. Especially in cases where the localization uncertainty is on the order of the distance between binding sites, or where there is a broad distribution of localization uncertainties, or when the localization uncertainty is anisotropic (for

3D datasets), this may improve the sensitivity to the initial setting of the widths of the Gaussian components, as well as promote convergence to a global optimum. Another improvement may be found in a better description of the quality of the clusters. Now we opt for the simple criterion of number of particles in the cluster. Using the FRC resolution may be a better practice for assessing cluster quality.

2.6. CONCLUSION

We have proposed a fast particle fusion method with computational complexity that scales linearly with the number of input particles. In our method we apply the JRMPC method for multiple initializations and then use classification and connection steps to generate a correct reconstruction with as many particles as possible. The reconstruction quality of our method is measured by the FRC resolution and compared with the all-to-all method, revealing that our results are of comparable or better quality. Our method is fast, even without GPU acceleration, avoids symmetry artifacts, applies to 2D and 3D datasets, and reconstructs poor data with a limited number of particles, a low density of labelling and a large localization uncertainty.

2.7. APPENDIX

2.7.1. SUMMARY OF JRMPC

The JRMPC method [13] is cast as an Expectation Maximization (EM) algorithm. In this framework the observed data are the set of particles $j = 1,2,\ldots,N$ with localizations $i = 1,2,\ldots,M_j$ represented by coordinates $\mathbf{v}_{ji}$. The localization uncertainties are not taken into account in the JRMPC method. The estimated parameters are the parameters of the K Gaussians of the GMM, and the rotations and translations of the particles that best match the GMM, defined as Θ in Equation 2.1. The latent or unobserved data $\mathcal{Z}$ concern the assignment of localizations i in particle j to Gaussian k of the GMM. We have modified the original approach of ref. [13] by ignoring the outlier probabilities, i.e. there is no outlier class where the localizations can be assigned to. The expectation value of the log-likelihood over the distribution of latent data can be expressed as the sum over Gaussians k, particles j and localizations i as:

$$\mathscr{E}(\Theta|\mathbf{V}) = \sum_{k=1}^{K}\sum_{j=1}^{N}\sum_{i=1}^{M_j} \alpha_{jik}\left[\log p_k + \log P(\mathbf{v}_{ji}|G_k)\right]. \tag{2.11}$$

Here, the marginal probability that localization i of particle j fits Gaussian k is given by the normal distribution:

$$\begin{aligned} P(\mathbf{v}_{ji}|G_k) &= \mathcal{N}\left(\phi(\mathbf{v}_{ji})|\boldsymbol{\mu}_k, \sigma_k^2\mathbf{I}_d\right) \\ &= \frac{1}{(2\pi)^{d/2}\sigma_k^d}\exp\left(-\frac{1}{2\sigma_k^2}\|\phi(\mathbf{v}_{ji}) - \boldsymbol{\mu}_k\|_F^2\right), \end{aligned} \tag{2.12}$$

where $\|\cdot\|_F$ denotes the Frobenius norm, the Gaussian weight p_k is the probability $p_k = P(G_k|\Theta)$, and the coefficients α_{jik} represent the posterior probability of the latent variable, i.e. the probability that localization i of particle j is assigned to Gaussian k. Starting

point of each iteration round of the JRMPC is to update the posterior probabilities according to:

$$\alpha_{jik} = \frac{p_k P(\mathbf{v}_{ji}|G_k)}{\sum_{s=1}^{K} p_s P(\mathbf{v}_{ji}|G_s)}. \tag{2.13}$$

The next step is the update of the rotation and translation matrices. It appears that finding the optimum log-likelihood expectation value can be cast as:

$$\begin{cases} \min_{\mathbf{R}_j,\mathbf{t}_j} \|(\mathbf{R}_j\mathbf{W}_j + \boldsymbol{t}_j\boldsymbol{e}^\top - \mathbf{M})\Lambda_j\|_F^2 \\ \text{s.t.} \quad \mathbf{R}_j^\top\mathbf{R}_j = \mathbf{I}_d \text{ and } |\mathbf{R}_j| = 1, \end{cases} \tag{2.14}$$

where $\mathbf{\Lambda}_j \in \mathbb{R}^{K\times K}$ is a diagonal matrix with elements $\lambda_{jkk} = (\sum_{i=1}^{M_j} \alpha_{jik}/\sigma_k^2)^{1/2}$, $\boldsymbol{e} \in \mathbb{R}^K$ is a vector of ones, $\mathbf{M} = [\boldsymbol{\mu}_1,\ldots,\boldsymbol{\mu}_K] \in \mathbb{R}^{d\times K}$ represents a matrix of the means of the Gaussian components, and $\mathbf{W}_j = [\boldsymbol{w}_{j1},\ldots,\boldsymbol{w}_{jK}] \in \mathbb{R}^{d\times K}$ is the weighted-average localizations of the j^{th} particle, where $\boldsymbol{w}_{jk}$ represents the single weighted-average localization of the j^{th} particle assigned to the k^{th} Gaussian component

$$\boldsymbol{w}_{jk} = \frac{\sum_{i=1}^{M_j} \alpha_{jik}\mathbf{v}_{ji}}{\sum_{i=1}^{M_j} \alpha_{jik}}. \tag{2.15}$$

The optimal transformations $\Phi_j = \{\mathbf{R}_j,\mathbf{t}_j\}_{k=1}^K$ are subsequently found using the method of Umeyama [26]. Then, the optimal means and covariances of the Gaussian components are estimated. It turns out that the closed-form expressions:

$$\boldsymbol{\mu}_k = \frac{\sum_{j=1}^{M}\sum_{i=1}^{M_j} \alpha_{jik}\phi(\mathbf{v}_{ji})}{\sum_{j=1}^{M}\sum_{i=1}^{M_j} \alpha_{jik}} \tag{2.16}$$

and

$$\sigma_k^2 = \frac{\sum_{j=1}^{N}\sum_{i=1}^{M_j} \alpha_{jik}\|\phi(\mathbf{v}_{ji}) - \boldsymbol{\mu}_k\|_2^2}{d\sum_{j=1}^{N}\sum_{i=1}^{M_j} \alpha_{jik}} + \epsilon^2, \tag{2.17}$$

provide the sought-for optimum. Here, a small scalar ϵ is added to avoid singularities of λ_{jkk}, as could arise if a Gaussian component has near zero localizations with a substantial posterior probability α_{jik}.

Finally, the weights of the Gaussian components are updated according to:

$$p_k = \frac{1}{\eta}\sum_{j=1}^{N}\sum_{i=1}^{M_j} \alpha_{jik}. \tag{2.18}$$

The different steps of the iterative JRMPC procedure are summarized as below:

Algorithm 1 JRMPC Algorithm

Input: a union of particles $\mathbf{V} = \{\mathbf{V}_j\}_{j=1}^{N}$

Require: number of iterations Q, initial parameter set $\Theta^0 = \left\{\{p_k^0, \boldsymbol{\mu}_k^0, \sigma_k^0\}_{k=1}^{K}, \{\mathbf{R}_j^0, \mathbf{t}_j^0\}_{j=1}^{N}\right\}$

for $q = 1$ to Q **do**
  Update α_{jik}^{q} from Θ^{q-1}
  Update $\mathbf{R}^q$ from α_{jik}^{q}, p_k^{q-1},$\boldsymbol{\mu}_k^{q-1}$ and σ_k^{q-1}
  Update $\mathbf{t}^q$ from $\mathbf{R}^q$, α_{jik}^{q}, p_k^{q-1},$\boldsymbol{\mu}_k^{q-1}$ and σ_k^{q-1}
  Update $\boldsymbol{\mu}_k^{q}$ from $\mathbf{t}^q$, $\mathbf{R}^q$ and α_{jik}^{q}
  Update σ_k^{q} from $\boldsymbol{\mu}_k^{q}$, $\mathbf{t}^q$, $\mathbf{R}^q$ and α_{jik}^{q}
  Update p_k^{q} from α_{jik}^{q}
  $q = q + 1$
end for
return $\Theta^Q = \left\{\{p_k^Q, \boldsymbol{\mu}_k^Q, \sigma_k^Q\}_{k=1}^{K}, \{\mathbf{R}_j^Q, \mathbf{t}_j^Q\}_{j=1}^{N}\right\}$

2.7.2. CONNECTION PIPELINE

The connection pipeline is summarized here:

Algorithm 2 Connection Algorithm

Input: a union of particles $\mathbf{V} = \{\mathbf{V}_j\}_{j=1}^{N}$, clusters $\mathbf{C} = \{\mathbf{C}^l\}_{l=1}^{L}$ and translations $\Phi = \{\Phi^l\}_{l=1}^{L}$

Require: the main cluster $\mathbf{C}_m^{l_m}$, the reconstruction estimate $\mathbf{C}_a$ ($\mathbf{C}_a$ is initialized as $\mathbf{C}_m^{l_m}$)

if $l \neq l_m$ **then**
  for $l < L$ **do**
    for $n < n_l$ **do**
      Find all common particles between $\mathbf{C}_n^l$ and $\mathbf{C}_m^{l_m}$.
      if common particle exists **then**
        Determine the common particle $\mathbf{V}_c$ to use for connection.
        Find unique particles $\mathbf{V}_u$ of $\mathbf{C}_n^l$ from $\mathbf{C}_a$.
        if $\mathbf{V}_u$ exists **then**
          Transfer all $\mathbf{V}_u$ to $\mathbf{C}_a$ with inverse Φ_c^l and $\Phi_c^{l_m}$.
        end if
      end if
      $n = n + 1$
    end for
    $l = l + 1$
  end for
end if
return the final super-particle reconstruction $\mathbf{C}_a$

2.7.3. PARAMETER SETTING TABLE

The default values given in the Table 2.1 are suitable for most of the cases.

definition	notation	default value
# particles used to estimate K	ζ	$\min(20, N)$
# GMM centers used to estimate K	K_0	$\min(\frac{\Sigma_{j=1}^{N} M_j}{N}, 100)$
# GMM centers	K	depends on input
prior probability of G_k	p_k^0	$1/K$
initial mean of G_k	μ_k^0	randomly generated
initial rotation matrix	$\mathbf{R}_j^0$	$\mathbf{I}_d$
initial translation matrix	$\mathbf{t}_j^0$	$\bar{\boldsymbol{\mu}}^0 - \bar{\mathbf{v}}_j$
initial standard deviation of G_k	σ_k^0	depends on input
# repetitions	L	2
# clusters	n_l	2
threshold for good cluster	ϑ	$N/(n_l+1)$

Table 2.1: Parameter Settings

2.7.4. SIMULATION RESULTS

We applied our method to simulated TUD-logo datasets to explore its limitations. These simulated data have the default settings of 200 particles, 2000 detected photons per localization event (corresponding to an average localization uncertainty of 4.85 nm) and 60% *DOL*. When we change one of these three parameters, we keep the other two at the default values. We simulate more challenging conditions than most often encountered in real experiments to probe the performance limitations. We perform ten independent simulations for each setting of the simulated data. We evaluate the reconstruction quality by calculating the average distance of localizations to the corresponding binding sites (*AD* in short) following ref. [12] where this measure was introduced for simulated data. For a simulated structure with around 5 nm distance between binding sites corresponding to an DNA Origami design, then an error of $AD < 10$ nm is needed for the reconstruction to appear reasonably correct; for $AD < 5$ nm, binding sites details can be observed in the reconstruction. Figure 2.6(a) shows that the error (*AD*) decreases with increasing *DOL*. For *DOL* larger than about 40% our method can stably obtain a clear reconstruction. Our method is not sensitive to the number of input particles. For particle numbers varying from 3 to 200, *AD* values are always less than 5 nm and fluctuate in a small range (Figure2.6(b)). Even though the registration for input particles less than 10 appears correct, the underlying structure is still hardly visible in the reconstruction because of the small total number of localizations.

Figure 2.6(c) indicates that the error *AD* decreases with increasing number of photons per localization. With 200 input particles with 60% *DOL*, our method is able to correctly reconstruct the underlying structure as long as the number of photons is greater than 400, corresponding to a localization uncertainty of 12 nm which is 2.4 times larger than the

minimum binding site distance 5 nm.

2.8. ACKNOWLEDGEMENTS AND FUNDING

We thank Sabri Bolkar for initial attempts to apply the JRMPC method to single molecule localization microscopy. This work has been supported by the Dutch Research Council (NWO), VICI grant no. 17046 for B.R. and W.W.

REFERENCES

[1] S. W. Hell, *Microscopy and its focal switch,* Nature Methods **6** (2009), pp. 24–32, doi: 10.1038/nmeth.1291.

[2] G. Vicidomini, P. Bianchini, and A. Diaspro, *STED super-resolved microscopy,* Nature Methods **15** (2018), pp. 173–182, doi: 10.1038/nmeth.4593.

[3] T. Klein, S. Proppert, and M. Sauer, *Eight years of single-molecule localization microscopy,* Histochemistry and Cell Biology **141** (2014), pp. 561–575, doi: 10.1007/s00418-014-1184-3.

[4] R. P. J. Nieuwenhuizen, K. A. Lidke, M. Bates, D. Leyton Puig, D. Grünwald, S. Stallinga, and B. Rieger, *Measuring image resolution in optical nanoscopy,* Nature Methods **10** (2013), pp. 557–562, doi: 10.1038/nmeth.2448.

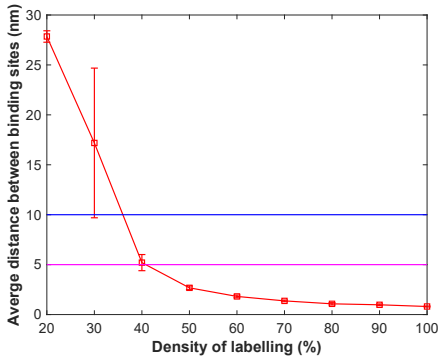
[5] M. Metzger, A. Konrad, S. Skandary, I. Ashraf, A. J. Meixner, and M. Brecht, *Resolution enhancement for low-temperature scanning microscopy by cryo-immersion,* Opt. Express **24** (2016), pp. 13023–13032, doi: 10.1364/oe.24.013023.

[6] P. Fox-Roberts, R. Marsh, K. Pfisterer, A. Jayo, M. Parsons, and S. Cox, *Local dimensionality determines imaging speed in localization microscopy,* Nat. Commun. **8** (2017), pp. 1–10, doi: 10.1038/ncomms13558.

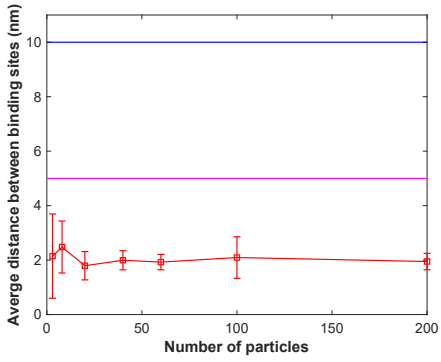
[7] A. Löschberger, S. van de Linde, M.-C. Dabauvalle, B. Rieger, M. Heilemann, G. Krohne, and M. Sauer, *Super-resolution imaging visualizes the eightfold symmetry of gp210 proteins around the nuclear pore complex and resolves the central channel with nanometer resolution,* Journal of Cell Science **125** (2012), pp. 570–575, doi: 10.1242/jcs.098822.

[8] A. Szymborska, A. De Marco, N. Daigle, V. C. Cordes, J. A. Briggs, and J. Ellenberg, *Nuclear pore scaffold structure analyzed by super-resolution microscopy and particle averaging,* Science **341** (2013), pp. 655–658, doi: 10.1126/science.1240672.

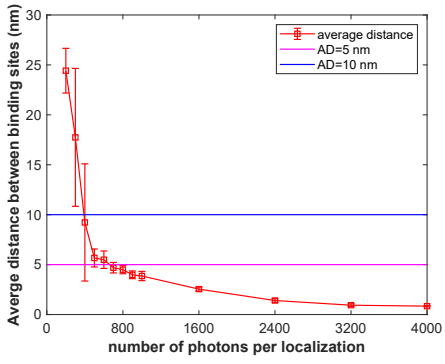
[9] J. Broeken, H. Johnson, D. S. Lidke, S. Liu, R. P. Nieuwenhuizen, S. Stallinga, K. A. Lidke, and B. Rieger, *Resolution improvement by 3D particle averaging in localization microscopy,* Methods and Applications in Fluorescence **3** (2015), p. 014003, doi: 10.1088/2050-6120/3/1/014003.



(a)



(b)



(c)

Figure 2.6: Simulation study of limitation of the proposed method. Each point in the graph indicates ten independent experiments. Reconstructions with AD < 10 *nm (magenta line) are assessed as 'correct' and with AD* < 6 *nm (blue line) as 'clear'. (a) Reconstruction quality as a function of DOL for 200 particles with 2000 photons per localization. (b) Reconstruction quality as a function of number of input particles with 60% DOL and 2000 photons per localization. (c) Reconstruction quality as a function of number of photons per localization for 200 particles with 60% DOL.*

[10] R. D. Gray, C. Beerli, P. M. Pereira, K. M. Scherer, J. Samolej, C. K. E. Bleck, J. Mercer, and R. Henriques, *Virusmapper: open-source nanoscale mapping of viral architecture through super-resolution microscopy,* Sci. Rep. **6** (2016), pp. 1–8, doi: 10.1038/srep2913.

[11] H. Heydarian, F. Schueder, M. T. Strauss, B. Van Werkhoven, M. Fazel, K. A. Lidke, R. Jungmann, S. Stallinga, and B. Rieger, *Template-free 2D particle fusion in localization microscopy,* Nature Methods **15** (2018), pp. 781–784, doi: 10.1038/s41592-018-0136-6.

[12] H. Heydarian, M. Joosten, A. Przybylski, F. Schueder, R. Jungmann, B. v. Werkhoven, J. Keller-Findeisen, J. Ries, S. Stallinga, M. Bates, *et al.*, *3D particle averaging and detection of macromolecular symmetry in localization microscopy,* Nat. Commun. **12** (2021), pp. 1–9, doi: 10.1038/s41467-021-22006-5.

[13] G. D. Evangelidis and R. Horaud, *Joint alignment of multiple point sets with batch and incremental expectation-maximization,* IEEE Trans. Pattern Anal. Mach. Intell. **40** (2017), pp. 1397–1410, doi: 10.1109/TPAMI.2017.2717829 .

[14] T. A. P. M. Huijben, H. Heydarian, A. Auer, F. Schueder, R. Jungmann, S. Stallinga, and B. Rieger, *Detecting structural heterogeneity in single-molecule localization microscopy data,* Nat. Commun. **12** (2021), pp. 1–8, doi: 10.1038/s41467-021-24106-8.

[15] A. Mead, *Review of the Development of Multidimensional Scaling Methods,* J. Roy. Stat. Soc. D-sta **41** (1992), pp. 27–39, doi: 10.2307/2348634.

[16] Y. Cheng, *Mean shift, mode seeking, and clustering,* IEEE T. Pattern Anal. **17** (1995), pp. 790–799, doi: 10.1109/34.400568.

[17] A. K. Jain, M. N. Murty, and P. J. Flynn, *Data clustering: a review,* ACM Comput. Surv. (CSUR) **31** (1999), pp. 264–323, doi: 10.1145/331499.331504.

[18] B. Rieger and S. Stallinga, *The lateral and axial localization uncertainty in super-resolution light microscopy,* Chemphyschem **15** (2014), pp. 664–670, doi: 10.1002/cphc.201300711.

[19] H. Heydarian, F. F. Schueder, M. M. Strauss, B. B. van Werkhoven, M. M. Fazel, K. K. Lidke, R. R. Jungmann, S. S. Stallinga, and B. B. Rieger, *Single-Molecule Localization Microscopy (SMLM) 2D TU Delft logos,* 4TU.ResearchData. Dataset (2018), doi: 10.4121/uuid:0d42a28f-f625-41a3-ba77-25e397685466.

[20] H. Heydarian, B. Rieger, R. Jungmann, F. Schueder, S. Stallinga, and J. Ries, *Single-Molecule Localization Microscopy (SMLM) 3D datasets,* 4TU.ResearchData. Dataset (2021), doi: 10.4121/13797686.v1.

[21] A. Auer, M. T. Strauss, S. Strauss, and R. Jungmann, *nanoTRON: a Picasso module for MLP-based classification of super-resolution data,* Bioinformatics **36** (2020), pp. 3620–3622, doi: 10.1093/bioinformatics/btaa154.

[22] H. Heydarian, T. Huijben, B. Rieger, S. Stallinga, R. Jungmann, F. Schueder, and A. Auer, *Single-molecule localization microscopy (SMLM) 2D digits 123 and TOL letters datasets,* 4TU.ResearchData. Dataset (2021), doi: 10.4121/14074091.v1.

[23] K. Fukunaga and L. Hostetler, *The estimation of the gradient of a density function, with applications in pattern recognition,* IEEE T. Inform. Theory **21** (1975), pp. 32–40, doi: 10.1109/TIT.1975.1055330.

[24] J. Kosinski, S. Mosalaganti, A. Von Appen, R. Teimer, A. L. DiGuilio, W. Wan, K. H. Bui, W. J. Hagen, J. A. Briggs, J. S. Glavy, *et al.*, *Molecular architecture of the inner ring scaffold of the human nuclear pore complex,* Science **352** (2016), pp. 363–365, doi: 10.1126/science.aaf0643 .

[25] J. Thevasthasan, M. Kahnwald, K. Cieslinski, P. Hoess, S. Kumar Penetie, M. Reitberger, D. Heid, K. Chitanaya Kasuba, S. Janice Hoerner, Y. Li, Y.-L. Wu, M. Mund, U. Matti, P. Matos Pereira, R. Henriques, B. Nijmeijer, M. Kuehlbeck, V. Jimenez Sabinina, J. Ellenberg, and J. Ries, *Nuclear pores as versatile reference standards for quantitative superresolution microscopy,* Nature Methods **16** (2019), pp. 1045–1053, doi: 10.1038/s41592-019-0574-9.

[26] S. Umeyama, *Least-squares estimation of transformation parameters between two point patterns,* IEEE T. Pattern Anal. **13** (1991), pp. 376–380, doi: 10.1109/34.88573.

2.9. CORRECTION NOTE

[2] In the results section of the original publication, an oversight occurred in the computational time descriptions for Fig. 2.2(b-f) and Fig. 2.3. The computational times reported for our method were, in fact, based on GPU-accelerated computation. This oversight may have caused confusion, as the implication was that the times stated were for computations using standard CPU acceleration. We would like to clarify this distinction and provide the correct comparative data. A revised comparison of computational times, differentiating between GPU and CPU accelerations, is provided in the table below.

Table 2.2: Comparison of computational times for GPU and CPU acceleration

	particle number	localizations	time in paper (GPU acceleration)	time with CPU acceleration (40 cores as mentioned in the paper)
Fig. 2.2 b	383	788,875	9.5 min	28.1 min
Fig. 2.3 a	4,155	692,139	1.1 h	0.6 h
Fig. 2.3 b	4,943	909,805	1.3 h	0.8 h
Fig. 2.3 c	2,541	4,714,134	48 min	7.5 h
Fig. 2.3 d	5,961	2,359,937	4.8 h	4.7 h

The conclusion drawn from the table is that for datasets with more localizations, GPU acceleration performs better. However, for datasets with a higher particle number, CPU acceleration yields superior results. The bulk computation time is spent on the computation of the Bhattacharya distances in the (dis)similarity matrix. The reason for this scaling behavior can be explained by our choice of parallelization. If we have N particles, then the computation complexity of the dissimilarity matrix scales as N^2 and if each particles contains M localizations, then the computation of one Bhattacharya distance scales as M^2. We have implemented CPU parallelism over N where each particle is assigned a different core and GPU parallelism over M. Given this table a future advance would be a combined approach or automatic decision which parallelism to choose.

This correction does not alter the main conclusions of our work, as the efficiency of our method under both computational conditions remains consistent with the original findings.

[2] This section has been published as: **Wenxiu Wang**, Hamidreza Heydarian, Teun A.P.M. Huijben, Sjoerd Stallinga, and Bernd Rieger, *Correction to: Joint registration of multiple point clouds for fast particle fusion in localization microscopy*, Bioinformatics, Volume 39, Issue 8, August 2023, doi: 10.1093/bioinformatics/btad500

3

PARTICLE FUSION OF SUPER-RESOLUTION DATA REVEALS THE UNIT STRUCTURE OF NUP96 IN NUCLEAR PORE COMPLEX

[1]

Single molecule localization microscopy offers resolution nearly down to the molecular level with specific molecular labelling, and is thereby a promising tool for structural biology. In practice, however, the actual value to this field is limited primarily by incomplete fluorescent labelling of the structure. This missing information can be completed by merging information from many structurally identical particles in a particle fusion approach similar to cryo-EM single-particle analysis. In this paper, we present a data analysis of particle fusion results of fluorescently labelled Nup96 nucleoporins in the Nuclear Pore Complex to show that Nup96 occurs in a spatial arrangement of two rings of 8 units with two Nup96 copies per unit giving a total of 32 Nup96 copies per pore. We use Artificial Intelligence assisted modeling in Alphafold to extend the existing cryo-EM model of Nup96 to accurately pinpoint the positions of the fluorescent labels and show the accuracy of the match between fluorescent and cryo-EM data to be better than 3 nm in-plane and 5 nm out-of-plane.

[1] This chapter has been published as: **Wenxiu Wang**, Arjen Jakobi, Yu-Le Wu, Jonas Ries, Sjoerd Stallinga, and Bernd Rieger, *Particle fusion of super-resolution data reveals the unit structure of Nup96 in Nuclear Pore Complex*, Scientific Reports 13, no. 1 (2023): 13327, doi: 10.1038/s41598-023-39829-5

3.1. INTRODUCTION

The Nuclear Pore Complex (NPC) is an essential molecular machine embedded in the nuclear envelope connecting the nucleus to the cytoplasm [1–3]. The NPC acts as a diffusion barrier that separates the nuclear compartment from the cytoplasm and works as a gateway for gene regulation [4, 5]. It is indispensable in eukaryotic cellular processes such as regulating the transport of proteins and ribonucleoprotein [4, 6, 7]. The structure and molecular composition of the NPC, in particular the scaffold, has been extensively studied. Previous cryogenic electron microscopy (cryo-EM) and cryo-electron tomography (cryo-ET) studies have resolved the structure of the NPC scaffold to high-resolution[1–3, 8–11]. The scaffold is composed of multiple copies of about 34 different nucleoporins (Nups) and these Nups are organized in threerings[12–14]. Previous studies have shown that the cytoplasmic ring (CR) and the nuclear ring (NR) are composed of multiple so-called Y-complexes and that Nup96 proteins are contained in these Y-complexes[10, 15]. Each ring contains 16 Nup96 molecules, organised into units with eight-fold rotational symmetry[15–17].

Super-resolution microscopy is emerging as a complementary technique for the study of biological structure, as it provides 'diffraction-unlimited' resolution[18–20]. Single molecule localization microscopy (SMLM) is one of these super-resolution techniques and obtains super-resolved images with a resolution of 10-20 nm by localizing single fluorescent emitters [18, 21]. If many chemically identical structures, called particles, can be imaged, they can be registered and combined into one 'super-particle'. With this strategy, the often poor degree of labelling of each individual particle can be mitigated and reconstructions with even better resolution than the typical 20 nm can be obtained[22–25]. Different 2D template-based particle fusion methods have been applied in SMLM to demonstrate the eight-fold rotational symmetry of the NPC [22, 26, 27].These methods, however, carry the risk to generate reconstructions with a bias toward the template. Later, a template-free 2D registration approach could reveal the eight-fold symmetry of the NPC in an unbiased manner [28]. This template-free method was extended to 3D and used to reconstruct the 3D structure of Nup107 and Nup96 revealing two phase shifted rings with eight blobs per ring [29]. Yet, this 3D approach suffered from the 'hot spot' artefact, which could only be mitigated by applying prior knowledge about the eight-fold symmetry in a post-processing step. In another 3D particle fusion approach a model is fitted to the individual NPCs [30] and the particles are combined in order of their similarity into the super-particle. This approach also shows two rings with each eight blobs or clusters in the NPC reconstruction, as expected, but interestingly some of the blobs are elongated and tilted in the plane of the rings. Limited by their high computational cost, neither the approach by [29] nor by [30] can reconstruct thousands of particles in a reasonable time. We recently introduced a template free and fast particle fusion approach [31] that overcomes the limitation to computation speed, so that datasets of several thousands of particles (or more) are now accessible for structural analysis. In our iterative fast particle fusion approach [31], we rotate and translate all the particles to fit a Gaussian Mixtures Model (GMM) using the Joint Registration of Multiple Point Clouds (JRMPC) method [32], and subsequently update the positions and widths of the Gaussian centers. In each iteration round both the particles and the GMM are updated. Due to the inherent limitations of the joint registration method [32], we can only obtain a locally optimal alignment of the

particles which consist of several distinct clusters with different poses. We then classify [33] these aligned particles with different poses and recombine them to obtain a globally optimal solution consisting of a single well-aligned structure.

Up to now SMLM of the NPC was able to reveal the eighth-fold rotational symmetry, resolve eight spots individually per ring and could show the separation of the nuclear and cytoplasmic rings in 3D. Although from cryo-EM work [15] it is known that e.g. Nup96s occur as eight distinct units per ring, where a single unit contains two Nup96 copies, this could so far not be resolved by conventional SMLM. Here we show that each of the eight blobs indeed contains two fluorophores attached via a SNAP tag to the Nup96 unit by combining our fast template free particle averaging method [31] with careful data analysis on 5 datasets (originating from five nuclei of five cells) of in total 4,538 NPCs [30]. We made a detailed comparison of the outcome of our analysis to the cryo-EM data. To this end, we extended the incomplete Nup96 model derived from von Appen et al. [15] by Alphafold [34] to find the positions where the fluorescent SNAP-tags are expected to attach to the Nup96. Next, we registered our estimated positions of the fluorescent units to these expected positions of the SNAP-tag from the cryo-EM model and found the average distance between the SNAP positions derived from the cryo-EM model and from our SMLM data analysis to be < 3 nm laterally and 5 nm axially.

3.2. RESULTS

We have applied the methodological steps outlined in the Methods section to five Nup96 datasets. Dataset 1 was described earlier [30]; datasets 2-5 result from the same cell line and staining and imaging protocol. The number of analyzed NPCs per dataset are 368, 568, 706, 1178 and 1,718 for a total of 4,538 NPCs. Since the fast particle fusion method [31] uses randomly generated GMM and leads to different reconstructions for the same dataset, we performed the fusion process 10 times for each dataset (including the combined dataset), resulting in a total of 60 reconstructions. Then we conduct data analysis carefully on these reconstructions to obtain our results.

PARTICLE FUSION INDICATES INCLINED ELLIPTICAL BLOBS

In Fig. 3.1a-c) we show one of the super-particle reconstructions of the combined dataset consisting of 4,538 NPCs. In the top-view of the nuclear (NR) and cytoplasmic (CR) ring (a,b) the elongation of the blobs is clearly visible. Fig. 3.1d) shows a histogram of the z-coordinates of the reconstruction in Fig. 3.1a-c). The localizations of the NR and CR are visible in the two peaks with a distance of 48.5 nm. Fig. 3.1e) shows the average distance between the NR and CR z-histogram peaks for the 5 individual datasets and the combined dataset. The box plots give an indication of the statistical distribution of the fitted parameters for the 10 independent reconstructions that were made for each of the 6 datasets. The lower whisker of the plot extends to the minimum value of 46.6 nm, while the upper whisker extends to the maximum value of 49.1 nm. The median of the data is approximately 48.1 nm.

Next, we analyzed the shape of the 16 separate blobs in the reconstructions. To that end, we first applied a routine for removing outlier localizations (see Methods), reducing the number of localizations by 32%. In Fig. 3.1f), the median value for the ellipticity is con-

sistent, ranging from around 0.90 to 0.93 for all datasets. There are a few outliers ranging from 0.75 to 0.85, which are lower than the median values. Furthermore, it appears that the long axes of the ellipses make a median angle of around $\pm 5.5^{o}$ with the tangent to the rings, with an opposite sign for the CR and NR (Fig. 3.1g). If the ellipticity was due to a registration error or artefact of the particle fusion procedure, then the long axis of the ellipses would be aligned to the tangent of the rings. We can therefore conclude that the non-zero inclination of the elliptical blobs is evidence of a structural feature. The most simple explanation is that each elliptical blob is the composite of localization events of emitters that bind to two distinct binding sites, i.e. the particle fusion directly suggests that Nup96 occurs in an arrangement with two Nup96 copies per unit. We note that the evidence for (at least) two binding sites comes from interpreting the structural feature of ellipticity. The binding sites cannot be directly observed as two distinct spots in the reconstructions. It is noticeable that the blobs in the NR have a better signal-to-noise ratio (SNR) than the blobs in the CR (Fig. 3.1c) and the NR has a significantly higher number of localizations compared to the CR, with around 22% more localizations (Fig. 3.1d). This difference in quality may be caused by the way the JRMPC method fuses the data. In a global optimization, a better registration may be achieved for one part of the structure (here NR) at the expense of the registration quality of another part of the structure (here CR). Additionally, structural heterogeneity could also contribute to the observed differences in registration quality.

UNCONSTRAINED AND SYMMETRY CONSTRAINED FITTING OF BINDING SITES

We investigated two approaches to estimate the positions of the two binding sites in the Nup96 units. Firstly, we used unconstrained fitting of 16 anisotropic Gaussian centers for the NR and CR separately, and secondly we used constrained fitting in which we impose the eight-fold rotational symmetry (according to Eqs. (3.1) and (3.2)). We found the distance between the 16 binding sites of the two fitting approaches for the combined data to be 2.8 ± 1.7 nm for the NR and 3.6 ± 1.4 nm for the CR (see Fig. 3.2). This small difference indicates that our data matches well with the eight-fold symmetry and that the use of the symmetry constraint in the fit is a valid procedure.

STRUCTURAL PARAMETERS OF THE NUP96 UNIT

Fig. 3.3 shows the structural parameters (radius R of CR and NR, distance d between the two binding sites in the unit, out-of-plane tilt angle θ of the unit, in-plane inclination angle ϕ of the unit) for all datasets for the case of symmetry constrained fitting. For comparison we also show the reference values from the cryo-EM data ($R = 54$ nm, $d = 11.8$ nm, $\phi = -32.6^{\circ}$ (NR), $\phi = 32.1^{\circ}$ (CR), $\theta = 76.8^{\circ}$) [15, 34]. The structural parameters obtained from the symmetry constrained fitting are relatively consistent between the different reconstructions and close to the cryo-EM reference values. In particular the radius and distance between the binding sites in the unit match well, where the angles are a bit more off compared to the cryo-EM model. There are some variations between the datasets, where datasets 3 and 4 appear to have more scattered localizations than the others resulting in visually poorer reconstructions for the NR and CR. The estimated uncertainties of the parameters for the NR are smaller than for the CR, consistent with the lower reconstruction quality of the CR compared to the NR. The radius of the NR matches very well with

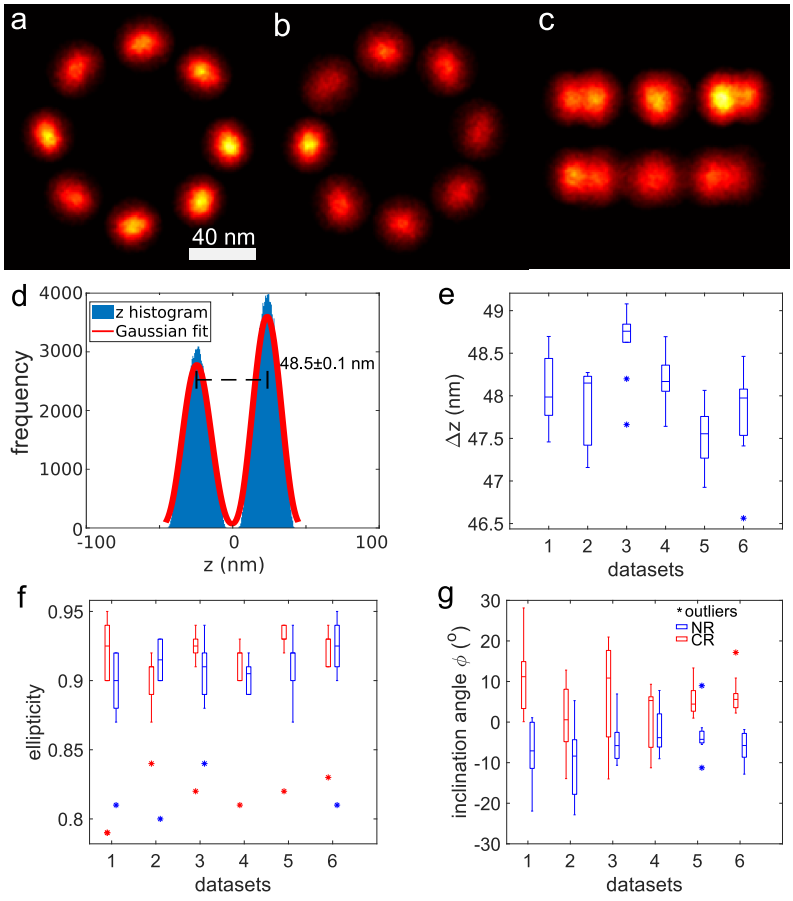


Figure 3.1: (a-c) One of the 3D reconstructions of Nup96 from 4,538 NPC particles of the combined dataset. The output reconstruction includes all input particles. The reconstruction resolves two rings and 8 blobs per ring. (a) top view of the nuclear ring (NR), (b) top view of the cytoplasmic ring (CR), (c) Side view, (d) histogram of z coordinates of localizations in the reconstruction obtained from the combined dataset, and bimodal Gaussian fit to the data. The distance between the two peaks of the histogram is 48.5±0.1 nm (e) Distance between the NR and CR for all the datasets. (f) Average ellipticity e of the elliptical blobs for the NR (blue) and CR (red), (g) Average angle ϕ of the long axis of the elliptical blobs with the tangent to the ring in the xy-plane for the NR (blue) and CR (red) (negative angles indicate clockwise rotation). Scale bar in (a) is 40 nm and applies to (b,c) as well.

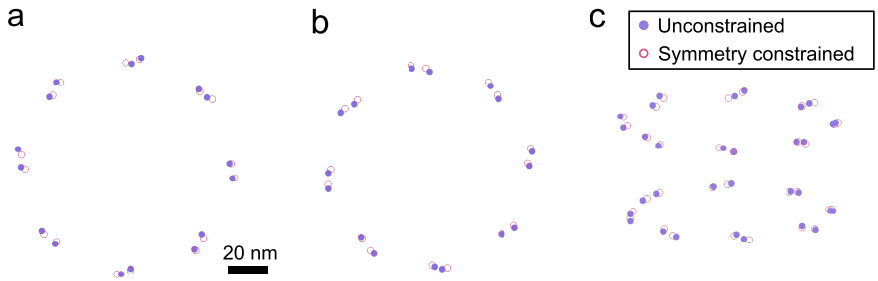


Figure 3.2: Locations of binding sites obtained from an unconstrained fit (purple) and from an eight-fold symmetry constrained fit (pink). These fits are obtained from the reconstruction displayed in Fig. 3.1a-c) of the combined dataset. Top view of binding sites for the NR (a) and CR (b). (c) Oblique view of CR and NR for projection at an angle 15^o *with the* xy*-plane. Scale bar in (a) applies to (b,c) as well.*

the cryo-EM reference value, whereas the radius of the CR is consistently about 2.5 nm smaller. The estimate of the distance between the binding sites in the unit is somewhat smaller than the cryo-EM reference distance. We find that the in-plane inclination angle has the opposite sign for the NR and the CR, which is consistent with the cryo-EM model. The estimated magnitude of the in-plane inclination angle is somewhat smaller than the cryo-EM reference values. A larger quantitative mismatch is found for the out-of-plane tilt angle, which we attribute to the localization uncertainty in the axial direction, which is two to three times larger than the uncertainty in the xy-plane [35].

COMPARISON OF CRYO-EM AND SMLM

Figure 3.4 shows a direct comparison between the positions of the SNAP tags predicted from the cryo-EM data, and the positions of the emitters according to the SMLM particle fusion data. The emitters of SMLM are generated from the overall symmetry constrained fit parameters obtained from averaging all 60 fits. We measured the distance between the corresponding emitters of our symmetry constrained fit and SNAP tags predicted from the cryo-EM. For the NR we find in-plane distances of 2.1 nm and 2.3 nm, and out-of-plane distances of 5.1 nm and 5.0 nm for the two emitter positions in the unit, for the CR we find in-plane distances of 4.2 nm and 0.4 nm, and out-of-plane distances of -3.5 nm and -5.2 nm for the two emitter positions in the unit. Please note that we cannot assign the fluorophore position directly in our structural model, which is limited to modeling the SNAP tag position relative to NUP96, whereas we measure the fluorophore position directly in SMLM. The expected distance between the SNAP-tag and the fluorophore position is 1-2 nm. Overall the agreement in the plane of the NPC rings is rather well, with an overall error of just 2.2 ± 1.4 nm.

The lateral error also seems free of a systematic bias between the cryo-EM based and the SMLM particle fusion based position estimates. This stands in contrast to the axial position estimates where we find a distance between the top emitters in the CR and the NR

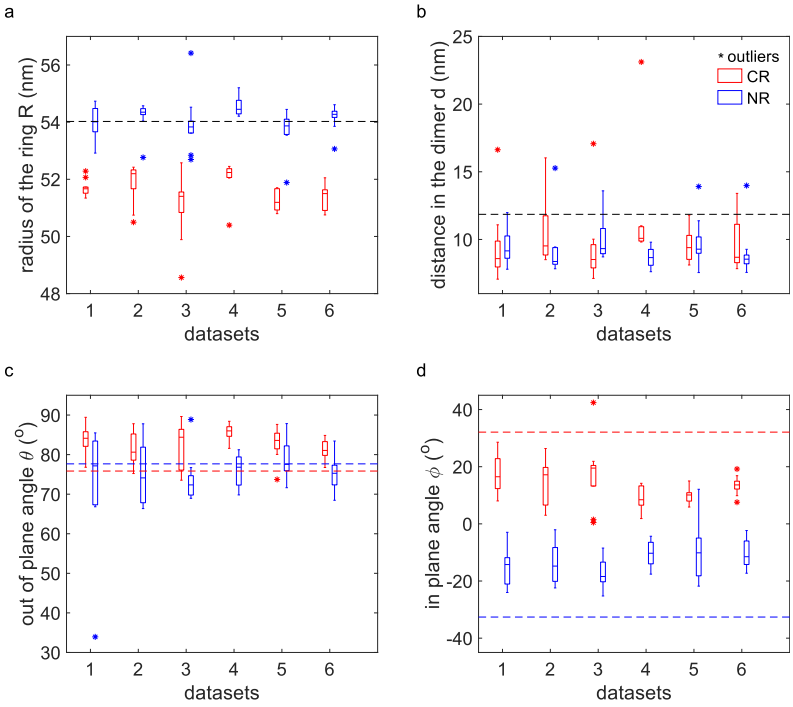


Figure 3.3: Structural parameters for eight-fold symmetry constrained fitting for the NR (blue) and CR (red) for all datasets. Parameters of fluorophore locations derived from the cryo-EM model are shown as dashed lines. (a) Radius for the NR and CR, the NR radius is around 54.1 nm and the CR radius is around 51.6 nm. The reference radius is 54.2 nm. (b) Distance between the binding sites in the units. The distance in the units is 9.7 ± 2.2 *nm and the reference value is* 11.8 *nm. (c) Out-of-plane angle* θ *of the line connecting the two emitters in a unit to the* z*-axis,* $\theta = 79.1^\circ \pm 7.3^\circ$ *and the reference value is* $\theta = 76.8^\circ$*. (d) In-plane angle* ϕ *of the line connecting the two emitters in the unit to the tangent of the ring in the* xy*-plane,* $\phi \sim -12.8 \pm 7.0^\circ$ *for the NR and* $\phi \sim +13.7 \pm 7.5^\circ$ *for the CR. The reference angle is* -32.6° *for the NR (blue dashed line) and* 32.1° *for the CR (red dashed line). For panels (c-d), positive angles indicate counterclockwise rotation.*

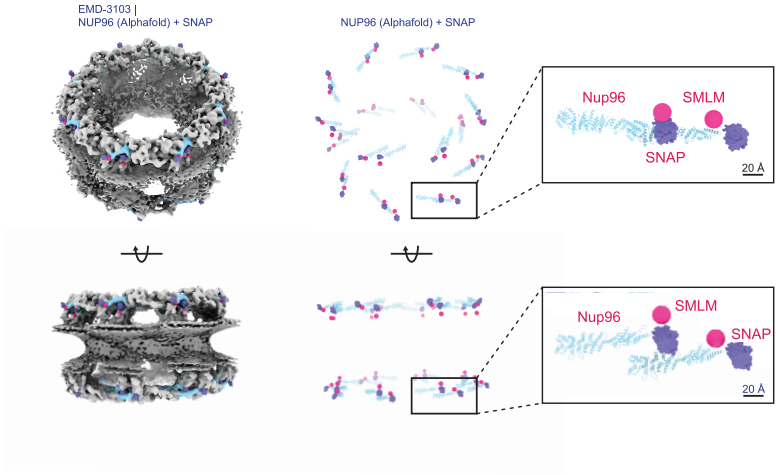


Figure 3.4: Overlay of the fluorophore positions from the SMLM particle fusion data (pink) and the SNAP-tag derived from the cryo-EM data (purple). For our overall SMLM emitters (pink), the lateral distance between a unit are 9.1 nm for NR and 10.0 nm for CR. The axial distances between a unit are 2.4 nm for NR and 1.2 nm for CR. The SNAP tags (purple) have lateral distances between a unit of 11.6 nm for NR and 11.5 nm for CR as well as axial distances of 2.5 nm for NR and 2.9 nm for CR.

of 48.4 nm, and between the bottom emitters in the CR and NR of 47.3 nm. It appears that the SMLM data give a distance between the NR and CR that is systematically smaller than the distance obtained from the cryo-EM model, with an average bias in estimated NPC thickness obtained from the cryo-EM/SMLM comparison of 9.4 nm. Fig. 3.1e) shows the estimated distance between the NR and CR for all datasets as well as Fig. 3.1d) displays the histogram of the z coordinates of the 244,946 localizations in the reconstruction obtained from the combined dataset. We find the distance between the NR and CR in SMLM to be 48.0 ± 0.4 nm, while the distance between the rings in the EM model is 57.2 ± 0.2 nm, a bias of about 9.2 nm. This analysis of the underlying axial localization data suggests that the found bias is not due to an artefact of the particle fusion method.

3.3. DISCUSSION

Overall, our analysis of the SMLM particle fusion points to a Nup96 unit structure that matches well with the cryo-EM data. Compared to the cryo-EM model, however, one major inconsistency remains. Namely, the height of the NPC is estimated from SMLM particle fusion to be about 9 nm less than that from cryo-EM. Similar ring separation values for Nup96 have been reported in other SMLM studies [36, 37], and have a larger deviation from the cryo-EM model than the levels of statistical error. There are a number of factors that could contribute to this discrepancy. Firstly, the differences in cell lines and the functional state of cells may play a role, as discrepancies in the height of NPCs have previously been observed in cryo-ET structures from different cell lines, with a ~5 nm height

difference observed between in situ NPC observed in HeLa and HEK cells. These differences could correspond to distinct functional states involving e.g. dilation of the inner ring [3, 9]. Secondly, the axial localization data is calibrated based on reference images of NPCs that are oriented sideways. In a side view the two rings are clearly separated in the image plane. There the localization precision is much better ($2-3\times$) than in the axial direction and the separation can be estimated easily and used to calibrate the axial coordinate [37]. The side views are obtained from NPCs at the sides of the nucleus, while the other NPCs are located in the nuclear envelope that lies flat against the cover slip. Thirdly, if the fluorsophore motion is (partly) restricted during imaging, dipole orientation effects on single molecule imaging can have an impact on the lateral position estimation, resulting in biases on the order of ~10 nm [38]. A confounding factor is the analysis of 3D SMLM data of Nup107, which is adjacent to Nup96 in the NPC scaffold, and which was found to have a ring separation of 60 nm in SMLM particle fusion [29] - in agreement with the cryo-EM data. It is worth mentioning that the Nup107 particles were obtained from another experiment in which the cell line and the functional state of the cell were different from the datasets analysed in this paper.

The estimated structural parameters for the NR appear to be more consistent between datasets than for the CR (see Fig. 3.3). We attribute this to the fact that the localizations of the CR are more scattered, and in turn the reconstructed super-particle is of lower quality there. The underlying reason could be that the 3D particle fusion favors the alignment of one ring, which leaves the other ring more blurred - in particular in the axial direction. A possible point of improvement could be to re-register the localizations in the CR by our particle fusion method separately, which in the end may give rise to a reconstruction of the CR with better quality.

Other technical improvements to the data analysis can also be envisioned. Firstly, the removal of outlier localizations is now done via a cascade of adding an extra Gaussian center to the GMM and subsequently filtering localizations according to the density of localizations. A single integrated approach may improve robustness of the data analysis procedure. Secondly, initial parameter settings could result in a better convergence to global optima of GMM fitting. A more principal refinement of the current analysis relates to model selection. The current analysis relies on the simplest model that can explain the observations, namely that the Nup96 appears in a unit structure of two copies. A possible improvement may therefore be found in a statistical criterion that supports that there are 32 emitters in the NPC, as opposed to another multiple of 16.

Recently, Helmerich et al.[39] speculated that fluorophores at distances below about 10 nm cannot be reliably resolved with SMLM although the precision given by the microscope, data analysis pipeline, and observed single molecule photon count should allow such a distinction. Energy transfer between the close-by fluorophores is assumed to cause re-excitation of the emitters at the beginning of the experiment followed by bleaching of both, preventing observation of them individually as required for SMLM. In particular with the introduction of MINFLUX and related techniques [36, 40–42] that offer nanometer localization precision at low photon count these observations could frustrate direct imaging of the unit separation in the NPC except for rare cases and leave careful data analysis as presented here as the only remaining tool [43]. Gwosch et al.[44] fused 162 NPC particles obtained from the MINFLUX data of[36] by the all-to-all particle fusion

method [29]. Their result does not depict the unit structure we demonstrate here. This may be due to a number of factors such as low density of labelling or few localizations per particle.

Very recently, a novel DNA-barcoding method, "Resolution Enhancement by Sequential Imaging" (RESI), has significantly improved the localization precision [45]. By applying the data-driven average method[30] to give a fusion of 1,217 particles, individual Nup96 proteins within a unit could be distinguished. The lateral and axial distance were 11.9±1.2 nm and 5.4±0.4 nm, respectively. These values are somewhat larger than the values obtained from our symmetry constrained fit in Fig. 3.4. We measured each blob's long axis inclination angle and ellipticity from Fig. 2g of the RESI paper. We found for the CR that the inclination angle is -9.3±3.1^{o}, while for the NR the inclination angle is 5.9±1.6^{o}. In comparison, Fig. 3.1g gives median inclination angles of ±5.5 ^{o} for the CR and NR blobs with an opposite sign. The seeming difference in handedness of the structure, i.e. the sign of the in-plane inclination angle, of the RESI results compared to our results and those of ref. [30] is due to a difference in projection (viewed from the CR side vs. viewed from the NR side). The average ellipticity for the CR and NR blobs in the RESI result is 0.86±0.02 and 0.88±0.03, respectively. We found from Fig. 3.1f, ellipticity ranges from 0.90 to 0.93 for the CR and NR blobs, respectively, in good agreement. Overall, the conclusions on the unit structure are similar, differences may be attributed to biological variations and differences in imaging conditions.

In conclusion, we have demonstrated that SMLM can reveal 3D structures in the small nanometer range. Our high-resolution particle fusion reconstructions and subsequent data analysis enabled the precise estimation of the positions of 32 emitter binding sites for Nup96 in the NPC. The comparison with cryo-EM data shows consistency better than 2.5 nm laterally, but also a bias in the NPC height estimation of about 9.2 nm. The latter inconsistency is poorly understood, and requires further study by the research community.

3.4. METHODS

3.4.1. PARTICLE AVERAGING

We applied our previously published particle averaging method [31] to five NPC SMLM datasets individually and all 5 combined (total of 4,538 NPCs). Dataset 1 with 368 NPCs was previously described [30] and the other datasets were obtained under the same experimental conditions. Each dataset corresponds to a single cell. Our fast particle fusion method first jointly aligns the input particles with a randomly generated Gaussian Mixture Model (GMM). The number of Gaussian components and the initial Gaussian standard deviation of the GMM are pre-set for better convergence in the iterative optimisation step [32]. Since the joint registration can be trapped at a local optimum, which contains several groups of particles (clusters) with different overlapping poses in the reconstruction, we apply an unsupervised classification method [33] to separate these clusters and reconnect the cluster to the same pose. We use the default values given in [31] for the data fusion algorithm, except for the number of Gaussian components K in the Gaussian Mixture Model (GMM) and the initial Gaussian standard deviation of the Gaussian components. We set $K = 34$, that is 32 Gaussian components for 32 binding sites and 2 to accommodate

false positive localizations of the nuclear ring and the cytoplasmic ring. The initial poses of the NPC particles (long axis of the cylinder) are roughly aligned with the optical axis, as the nuclear membrane runs along the cover slip. We can use this to set an initial Gaussian standard deviation to 33 nm, smaller than the overall size of the NPC, to make the algorithm converge faster. We performed particle averaging 10 times on each dataset and thus obtained a total of 60 super-particles to incorporate the statistical variability induced by the randomly generated GMM into the analysis. All super-particles, shown in the Appendix, consistently exhibit two rings, each containing 8 elliptically shaped blobs across all 5 datasets as well as for the combined dataset.

3.4.2. SPLITTING IN NR AND CR

We obtain super-particles from the particle averaging of the different datasets that have the same shape but different absolute poses in 3D space. For better comparison and easier analysis we align their global pose such that the rings are perpendicular to the z-axis of our global coordinate system. We do so by registering the super-particles to a fixed template which has two rings and each ring contains eight points with eight-fold rotational symmetry. The center of the template is located at the origin of the coordinate system. The distance between the two rings in the template is 50 nm and the radius of each ring is 55 nm. Finally, we rotate the super-particles around the x and y-axis from -3^o to 3^o in steps of 0.2^o to find the minimum full width at half maximum (FWMH) of the histogram of z-coordinates. Then, we split the super-particle into two 8-blob rings according to their z-coordinates for further analysis. These rings represent the nuclear (NR) and cytoplasmic (CR) ring of the NPC.

3.4.3. OUTLIER LOCALIZATION FILTERING

Next, we remove outlier localizations in each ring. To this end, we use all the localizations in a ring and fit a randomly initialized GMM with 9 Gaussian components using again the JRMPC method. In the xy plane all the GMM centers are randomly generated inside a 2D box with a diameter of 100 nm. The z-coordinates of these initial Gaussian centers are set to 0. The standard deviations of the initial Gaussian components are set to 33 nm. Thus, an initial GMM with all Gaussian components locating nearby the localizations and having comparable standard deviations to the size of the ring is suitable for JRMPC to converge. In all cases we find that there are 8 components with similar (small) standard deviations and one component with a large standard deviation, indicating the outlier component. We remove the localizations in the latter component from the data (this is about 5% of the number of localizations). As the automated data analysis is sensitive to outlier localizations, we apply further outlier filtering based on the density of localizations to the blobs. We remove all localizations with a local density smaller than a threshold such that every blob is separated clearly from its adjacent blobs in the reconstruction. We experimented with various density thresholds and determined that a value of 95% of the mean density of a single blob is suitable for all reconstructions. This latter operation removes about 1/3 of the localizations.

3.4.4. ELLIPTICITY MEASUREMENT IN PER BLOB

The long axes of the elliptically shaped blobs in a ring are tilted with respect to the circumference of the ring. To investigate this further, we measure the ellipticity of each projected blob onto the xy-plane. The ellipticity is defined as $e = b/a$, where a and b are the lengths of the long and short axis, respectively. We also measure the inclination angle between the long axis of each blob and the tangent to the overall 2D ring. We then calculate the distribution of the ellipticity and inclination angle over the set of values for the 8 blobs of the 10 reconstructions in each ring.

3.4.5. ANISOTROPIC GAUSSIAN MIXTURE FITTING TO 8 BLOBS PER RING

The inclination of the elliptical shape of the blobs with respect to the rings indicates that more than one emitter is present per blob. As moreover the inclination angle is also opposite in sign for the NR and CR, the inclination is not a reconstruction artifact. The unit of two Nup96, however, is not resolvable due to the limited localization precision, residual drift or registration error. We assume that there are two SNAP emitters per blob based on the prior knowledge from the cryo-EM model that there are two Nup96 copies per blob. We again use a GMM with 16 components to fit the 8 blobs per ring and interpret the Gaussian components' centers as the positions of the emitters. For the fitting procedure we use the following heuristics: The initial standard deviation for the Gaussian fitting is equal in the x and y direction (in-plane), but two times larger in the z-direction (which is aligned with the optical axis), because the axial localization uncertainty is typically two to three times larger than in the xy-plane [35]. Furthermore, we assume that all Nup96 units in a ring are identical, i.e. the Gaussian components in the GMM have identical diagonal covariance matrices.

We use iterative Expectation-Maximization (EM) to find the optimal GMM by maximizing the likelihood for the GMM to fit the localizations [46]. The shared diagonal covariance matrices, centers of Gaussian components and posterior probabilities of component memberships are the fitting parameters and updated iteratively. The obtained Gaussian centers from fitting 16 anisotropic Gaussian mixture model are represented by $G = \{\mathbf{g}_i\}_{i=1}^{16}$ with $\mathbf{g}_i = (g_{ix}, g_{iy}, g_{iz})$.

3.4.6. INCORPORATION OF THE EIGHT-FOLD SYMMETRY INTO THE FITTING

We have employed a second fitting method to make use of the eight-fold symmetry, as unconstrained anisotropic GMM fitting is sensitive to the setting of the initial Gaussian centers. To this end we generate 16 points with eight-fold symmetry as the initial centers of the anisotropic GMM fitting for every ring (see Fig. 3.5). The 16 binding sites defined by the point set $S = \{\{\mathbf{s}_{kj}\}_{j=1}^{2}\}_{k=1}^{8}$ are characterized by 6 parameters: $(c_{1x}, c_{1y}, c_{1z}, d, \theta, \phi)$, where $\mathbf{c}_1 = (c_{1x}, c_{1y}, c_{1z})$ is the center of the first ($k = 1$) unit, d is the distance between the two binding sites in the unit, θ is the angle between the line connecting the two binding sites in a unit and the positive z-axis ($0 \le \theta \le \pi$), and ϕ is the angle between the projection of this line on the xy-plane to the tangent of the ring in xy-plane ($0 \le \phi \le 2\pi$). The in-plane coordinates of the unit center can be parameterized as $\mathbf{c}_1 = (c_{1x}, c_{1y}) = R(\cos\psi, \sin\psi)$, with R the radius of the ring and ψ the in-plane angle. The coordinates of the two binding

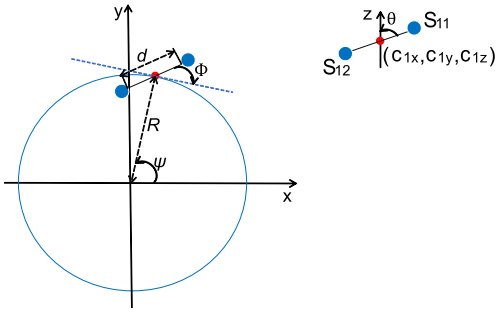


Figure 3.5: Schematic for structural parameters of the 8 pairs of two binding sites with overall eight-fold symmetry. The red point is the center of the first pair and the blue points are the binding sites in the units.

sites in the first unit are thus given by:

$$\mathbf{s}_{11} = (R\cos\psi, R\sin\psi, c_{1z}) + \frac{1}{2}d\left(-\sin\theta\sin\left(\phi+\psi\right), \sin\theta\cos\left(\phi+\psi\right), \cos\theta\right), \tag{3.1}$$

$$\mathbf{s}_{12} = (R\cos\psi, R\sin\psi, c_{1z}) - \frac{1}{2}d\left(-\sin\theta\sin\left(\phi+\psi\right), \sin\theta\cos\left(\phi+\psi\right), \cos\theta\right). \tag{3.2}$$

We choose emitter 1 to be the emitter that lies above the center plane of the ring, i.e. we restrict the polar angle to $0 \leq \theta \leq \pi/2$. The coordinates of the binding sites of the $k = 2,\ldots,8$ other units can be derived by rotating these points by $(k-1)\pi/4$ in the xy-plane. The parameters for the initial GMM centers are (0, 53.5 nm, ±24 nm, 13 nm, $\pi/2, \pi/4$), where + corresponds to NR and − to CR. The in-plane angle ϕ is randomly generated in the interval 0 to π. We find the parameters of the symmetry constrained point sets S from the coordinates of the unconstrained point set G by minimizing the mean square error between the point sets S and G with the quasi-Newton method [47] implemented in MATLAB.

3.4.7. COMPARISON WITH THE CRYO-EM MODEL

In order to relate the found emitter positions to available structural data on the human NPC [15], we modeled the full-length Nup96 structure using Alphafold2 [34] and added the SNAP-tag at the carboxyl-terminal position on the Nup96s. We rigid-body fitted the full-length Nup96 into the cryo-EM density using residues 881-1817 from PDB ID 5a9q as anchor residues [15]. As expected, the SNAP tag protrudes from the NPC assembly, adding support for its correct placement. We then symmetry-expanded the fitted model to generate the full NPC assembly and denote the center-of-mass of the O^6-benzylguanine-AF647 (BG-AF647) to represent the emitter position.

To align the cryo-EM and SMLM models, we first measured the parameters of the SNAP tags from the cryo-EM model. Then, we regenerated the SNAP tags in the EM model and the overall eight-fold symmetric emitter positions in the SMLM reconstructions from their fit parameters individually, ensuring that they share the same center as the first unit. The predicted SNAP tag positions based on the cryo-EM model are represented by position vectors $\mathbf{m}_{kj} = (m_{kj,x}, m_{kj,y}, m_{kj,z})$ for $k = 1,\ldots 8$ and $j = 1,2$. Using this set of position

vectors reference values for the radius R of the CR and NR, the distance d between the two binding sites in the unit, the in-plane inclination angle ϕ of the unit, and the out-of-plane tilt angle θ of the unit can be obtained from equations similar to Eqs. (3.1) and (3.2).

The two models can be directly compared by computing the in-plane and axial distance of the cryo-EM SNAP tag positions to the estimated emitter positions from SMLM particle fusion:

$$b_{xy,jk} = \sqrt{\left(m_{kj,x} - s_{kj,x}\right)^2 + \left(m_{kj,y} - s_{kj,y}\right)^2} \tag{3.3}$$

$$b_{z,jk} = m_{kj,z} - s_{kj,z}. \tag{3.4}$$

As the cryo-EM data satisfies the eight-fold rotational symmetry by construction we only compare the cryo-EM SNAP tag positions to the estimated emitter positions from SMLM particle fusion that are obtained from the symmetry constrained unit fits. That means that there are only four distinct values for $b_{xy,jk}$ and four distinct values for $b_{z,jk}$ (CR and NR, two emitter positions per unit).

3.5. ACKNOWLEDGEMENTS

This work has been supported by the Dutch Research Council (NWO), VICI grant no. 17046 for B.R. and W.W.

3.6. AUTHOR CONTRIBUTIONS STATEMENT

W.W. made numerical analysis and model fits. A.J. provided the analysis of the cryo-EM data. J.R. and Y.W. acquired the Nup96-SNAP SMLM data. S.S. and B.R. supervised the study. All authors contributed to writing the paper.

3.7. DATA AVAILABILITY STATEMENT

The dataset and source codes for this work are publicly available at
https://github.com/wexw/Particle-fusion-Nup96.git

3.8. ADDITIONAL INFORMATION

Competing interests: The authors declare no competing interests.

REFERENCES

[1] B. Hampoelz, A. Andres-Pons, P. Kastritis, and M. Beck, *Structure and assembly of the nuclear pore complex,* Annu. Rev. Biophys. **48** (2019), pp. 515–536, doi: 10.1146/annurev-biophys-052118-115308.

[2] D. H. Lin and A. Hoelz, *The structure of the nuclear pore complex (an update),* Annu. Rev. Biochem. **88** (2019), pp. 725–783, doi: 10.1146/annurev-biochem-062917-011901.

[3] S. Petrovic, G. W. Mobbs, C. J. Bley, S. Nie, A. Patke, and A. Hoelz, *Structure and function of the nuclear pore complex,* Cold Spring Harb. Perspect. Biol. **14** (2022), p. a041264, doi: 10.1101/cshperspect.a041264.

[4] V. Zila, E. Margiotta, B. Turoňová, T. G. Müller, C. E. Zimmerli, S. Mattei, M. Allegretti, K. Börner, J. Rada, B. Müller, *et al.*, *Cone-shaped hiv-1 capsids are transported through intact nuclear pores,* Cell **184** (2021), pp. 1032–1046, doi: 10.1016/j.cell.2021.01.025 .

[5] M. A. D'Angelo, *Nuclear pore complexes as hubs for gene regulation,* Nucleus **9** (2018), pp. 142–148, doi: 10.1080/19491034.2017.1395542.

[6] M. Stewart, *Polyadenylation and nuclear export of mRNAs,* Journal of Biological Chemistry **294** (2019), pp. 2977–2987, doi: 10.1074/jbc.REV118.005594.

[7] C. E. Wing, H. Y. J. Fung, and Y. M. Chook, *Karyopherin-mediated nucleocytoplasmic transport,* Nature Reviews Molecular Cell Biology **23** (2022), pp. 307–328, doi: 10.1038/s41580-021-00446-7.

[8] C. J. Bley, S. Nie, G. W. Mobbs, S. Petrovic, A. T. Gres, X. Liu, S. Mukherjee, S. Harvey, F. M. Huber, D. H. Lin, *et al.*, *Architecture of the cytoplasmic face of the nuclear pore,* Science **376** (2022), p. eabm9129, doi: 10.1126/science.abm9129.

[9] S. Mosalaganti, A. Obarska-Kosinska, M. Siggel, R. Taniguchi, B. Turoňová, C. E. Zimmerli, K. Buczak, F. H. Schmidt, E. Margiotta, M.-T. Mackmull, *et al.*, *AI-based structure prediction empowers integrative structural analysis of human nuclear pores,* Science **376** (2022), p. eabm9506, doi: 0.1126/science.abm9506.

[10] A. P. Schuller, M. Wojtynek, D. Mankus, M. Tatli, R. Kronenberg-Tenga, S. G. Regmi, P. V. Dip, A. K. Lytton-Jean, E. J. Brignole, M. Dasso, *et al.*, *The cellular environment shapes the nuclear pore complex architecture,* Nature **598** (2021), pp. 667–671, doi: 10.1038/s41586-021-03985-3.

[11] S. Petrovic, D. Samanta, T. Perriches, C. J. Bley, K. Thierbach, B. Brown, S. Nie, G. W. Mobbs, T. A. Stevens, X. Liu, *et al.*, *Architecture of the linker-scaffold in the nuclear pore,* Science **376** (2022), p. eabm9798, doi: 10.1126/science.abm9798.

[12] J. Kosinski, S. Mosalaganti, A. Von Appen, R. Teimer, A. L. DiGuilio, W. Wan, K. H. Bui, W. J. Hagen, J. A. Briggs, J. S. Glavy, *et al.*, *Molecular architecture of the inner ring scaffold of the human nuclear pore complex,* Science **352** (2016), pp. 363–365, doi: 10.1126/science.aaf0643 .

[13] D. H. Lin, T. Stuwe, S. Schilbach, E. J. Rundlet, T. Perriches, G. Mobbs, Y. Fan, K. Thierbach, F. M. Huber, L. N. Collins, A. Davenport, Y. Jeon, and A. Hoelz, *Architecture of the symmetric core of the nuclear pore,* Science **352** (2016), p. aaf1015, doi: 10.1126/science.aaf1015.

[14] A. Ori, N. Banterle, M. Iskar, A. Andrés-Pons, C. Escher, H. Khanh Bui, L. Sparks, V. Solis-Mezarino, O. Rinner, P. Bork, *et al.*, *Cell type-specific nuclear pores: a case in point for context-dependent stoichiometry of molecular machines,* Mol. Syst. Biol. **9** (2013), p. 648, doi: 10.1038/msb.2013.4.

[15] A. v. Appen, J. Kosinski, L. Sparks, A. Ori, A. DiGuilio, B. Vollmer, M.-T. Macmull, N. Banterle, . Parca, P. Kastritis, K. Buczak, S. Mosalagnati, W. Hagen, A. Andres-Pons,

E. Lemke, P. Bork, W. Antonin, J. Glavy, K. Bui, and M. Beck, *In situ structural analysis of the human nuclear pore complex,* Nature **526** (2015), pp. 140–143, doi: 10.1038/nature15381.

[16] K.-C. Hsia, P. Stavropoulos, G. Blobel, and A. Hoelz, *Architecture of a coat for the nuclear pore membrane,* Cell **131** (2007), pp. 1313–1326, doi: 10.1016/j.cell.2007.11.038.

[17] T. Stuwe, A. R. Correia, D. H. Lin, M. Paduch, V. T. Lu, A. A. Kossiakoff, and A. Hoelz, *Architecture of the nuclear pore complex coat,* Science **347** (2015), pp. 1148–1152, doi: 10.1126/science.aaa4136 .

[18] S. W. Hell, *Microscopy and its focal switch,* Nature Methods **6** (2009), pp. 24–32, doi: 10.1038/nmeth.1291.

[19] T. Klein, S. Proppert, and M. Sauer, *Eight years of single-molecule localization microscopy,* Histochemistry and Cell Biology **141** (2014), pp. 561–575, doi: 10.1007/s00418-014-1184-3.

[20] G. Vicidomini, P. Bianchini, and A. Diaspro, *STED super-resolved microscopy,* Nature Methods **15** (2018), pp. 173–182, doi: 10.1038/nmeth.4593.

[21] M. Lelek, M. Gyparaki, G. Beliu, F. Schueder, J. Griffie, S. Manley, R. Jungmann, M. Sauer, M. Lakakamyali, and C. Zimmer, *Single-molecule localization microscopy,* Nature Reviews Methods Primers **1** (2021), p. 39, doi: 10.1038/s43586-021-00038-x.

[22] A. Löschberger, S. van de Linde, M.-C. Dabauvalle, B. Rieger, M. Heilemann, G. Krohne, and M. Sauer, *Super-resolution imaging visualizes the eightfold symmetry of gp210 proteins around the nuclear pore complex and resolves the central channel with nanometer resolution,* Journal of Cell Science **125** (2012), pp. 570–575, doi: 10.1242/jcs.098822.

[23] A. Szymborska, A. De Marco, N. Daigle, V. C. Cordes, J. A. Briggs, and J. Ellenberg, *Nuclear pore scaffold structure analyzed by super-resolution microscopy and particle averaging,* Science **341** (2013), pp. 655–658, doi: 10.1126/science.1240672.

[24] J. Broeken, H. Johnson, D. S. Lidke, S. Liu, R. P. Nieuwenhuizen, S. Stallinga, K. A. Lidke, and B. Rieger, *Resolution improvement by 3D particle averaging in localization microscopy,* Methods and Applications in Fluorescence **3** (2015), p. 014003, doi: 10.1088/2050-6120/3/1/014003.

[25] R. D. Gray, C. Beerli, P. M. Pereira, K. M. Scherer, J. Samolej, C. K. E. Bleck, J. Mercer, and R. Henriques, *Virusmapper: open-source nanoscale mapping of viral architecture through super-resolution microscopy,* Sci. Rep. **6** (2016), pp. 1–8, doi: 10.1038/srep2913.

[26] A. Löschberger, C. Franke, G. Krohne, S. van de Linde, and M. Sauer, *Correlative super-resolution fluorescence and electron microscopy of the nuclear pore complex with molecular resolution,* Journal of Cell Science **127** (2014), pp. 4351–4355, doi: 10.1242/jcs.156620.

[27] A. P. Curd, J. Leng, R. E. Hughes, A. J. Cleasby, B. Rogers, C. H. Trinh, M. A. Baird, Y. Takagi, C. Tiede, C. Sieben, *et al.*, *Nanoscale pattern extraction from relative positions of sparse 3d localizations,* Nano Lett. **21** (2020), pp. 1213–1220, doi: 10.1021/acs.nanolett.0c03332.

[28] H. Heydarian, F. Schueder, M. T. Strauss, B. Van Werkhoven, M. Fazel, K. A. Lidke, R. Jungmann, S. Stallinga, and B. Rieger, *Template-free 2D particle fusion in localization microscopy,* Nature Methods **15** (2018), pp. 781–784, doi: 10.1038/s41592-018-0136-6.

[29] H. Heydarian, M. Joosten, A. Przybylski, F. Schueder, R. Jungmann, B. v. Werkhoven, J. Keller-Findeisen, J. Ries, S. Stallinga, M. Bates, and B. Rieger, *3D particle averaging and detection of macromolecular symmetry in localization microscopy,* Nature Communications **12** (2021), pp. 1–9, doi: 10.1038/s41467-021-22006-5.

[30] Y.-L. Wu, P. Hoess, A. Tschanz, U. Matti, M. Mund, and J. Ries, *Maximum-likelihood model fitting for quantitative analysis of SMLM data,* Nature Methods (2022), pp. 1–10, doi:10.1038/s41592-022-01676-z.

[31] W. Wang, H. Heydarian, T. A. P. M. Huijben, S. Stallinga, and B. Rieger, *Joint registration of multiple point clouds for fast particle fusion in localization microscopy,* Bioinformatics **38** (2022), pp. 3281–3287, doi: 10.1093/bioinformatics/btac320.

[32] G. D. Evangelidis and R. Horaud, *Joint alignment of multiple point sets with batch and incremental expectation-maximization,* IEEE Trans. Pattern Anal. Mach. Intell. **40** (2017), pp. 1397–1410, doi: 10.1109/TPAMI.2017.2717829 .

[33] T. A. P. M. Huijben, H. Heydarian, A. Auer, F. Schueder, R. Jungmann, S. Stallinga, and B. Rieger, *Detecting structural heterogeneity in single-molecule localization microscopy data,* Nat. Commun. **12** (2021), pp. 1–8, doi: 10.1038/s41467-021-24106-8.

[34] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. Kohl, A. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. Senior, K. Kavukcuoglu, P. Kohli, and D. Hassabis, *Highly accurate protein structure prediction with AlphaFold,* Nature **596** (2021), pp. 583–589, doi: 10.1038/s41586-021-03819-2.

[35] B. Rieger and S. Stallinga, *The lateral and axial localization uncertainty in super-resolution light microscopy,* Chemphyschem **15** (2014), pp. 664–670, doi: 10.1002/cphc.201300711.

[36] K. C. Gwosch, J. K. Pape, F. Balzarotti, P. Hoess, J. Ellenberg, J. Ries, and S. W. Hell, *MINFLUX nanoscopy delivers 3D multicolor nanometer resolution in cells,* Nature Methods **17** (2020), pp. 217–224, doi: 10.1038/s41592-019-0688-0.

[37] J. Thevasthasan, M. Kahnwald, K. Cieslinski, P. Hoess, S. Kumar Penetie, M. Reitberger, D. Heid, K. Chitanaya Kasuba, S. Janice Hoerner, Y. Li, Y.-L. Wu, M. Mund,

U. Matti, P. Matos Pereira, R. Henriques, B. Nijmeijer, M. Kuehlbeck, V. Jimenez Sabinina, J. Ellenberg, and J. Ries, *Nuclear pores as versatile reference standards for quantitative superresolution microscopy,* Nature Methods **16** (2019), pp. 1045–1053, doi: 10.1038/s41592-019-0574-9.

[38] S. Stallinga and B. Rieger, *Accuracy of the Gaussian point spread function model in 2D localization microscopy,* Opt. Express **18** (2010), pp. 24461–24476, doi: 10.1364/OE.18.024461.

[39] D. Helmerich, G. Beliu, D. Taban, M. Meub, M. Streit, A. Kuhlemana, S. Doose, and M. Sauer, *Photoswitching fingerprint analysis bypasses the 10-nm resolution barrier,* Nature Methods **19** (2022), pp. 986–994, doi: 10.1038/s41592-022-01548-6.

[40] F. Balzarotti, Y. Eilers, K. Gwosch, A. Gynnå, V. Westphal, F. Stefani, J. Elf, and S. Hell, *Nanometer resolution imaging and tracking of fluorescent molecules with minimal photon flux,* Science **355** (2016), pp. 606–612, doi: 10.1126/science.aak9913.

[41] J. Cnossen, T. Hinsdale, R. Ø. Thorsen, M. Siemons, F. Schueder, R. Jungmann, C. S. Smith, B. Rieger, and S. Stallinga, *Localization microscopy at doubled precision with patterned illumination,* Nature Methods **17** (2020), pp. 59–63, doi: 10.1038/s41592-019-0657-7.

[42] P. Jouchet, C. Cabriel, N. Bourg, M. Bardou, C. Poüs, E. Fort, and S. Lévêque-Fort, *Nanometric axial localization of single fluorescent molecules with modulated excitation,* Nature Photonics **15** (2021), pp. 297–304, doi: 10.1038/s41566-020-00749-9.

[43] K. Prakash and A. P. Curd, *Assessment of 3D MINFLUX data for quantitative structural biology in cells,* Nature Methods **20** (2023), pp. 48–51, doi: 10.1038/s41592-022-01694-x.

[44] K. C. Gwosch, F. Balzarotti, J. K. Pape, P. Hoess, J. Ellenberg, J. Ries, U. Matti, R. Schmidt, S. J. Sahl, and S. W. Hell, *Reply to: Assessment of 3D MINFLUX data for quantitative structural biology in cells,* Nature Methods **20** (2023), pp. 52–54, doi: 10.1038/s41592-022-01695-w.

[45] S. C. Reinhardt, L. A. Masullo, I. Baudrexel, P. R. Steen, R. Kowalewski, A. S. Eklund, S. Strauss, E. M. Unterauer, T. Schlichthaerle, M. T. Strauss, *et al.*, *Ångström-resolution fluorescence microscopy,* Nature **617** (2023), pp. 711–716, doi: 10.1038/s41586-023-05925-9.

[46] G. McLachlan and D. Peel, *Finite Mixture Models* (John Wiley & Sons, New York, 2000) doi: 10.1002/0471721182.

[47] D. F. Shanno, *Conditioning of quasi-Newton methods for function minimization,* Mathematics of Computation **24** (1970), pp. 647–656, doi: 10.1090/S0025-5718-1970-0274029-X.

3.9. APPENDIX

3.9.1. DATA FUSION RESULTS WITHOUT FILTERING OUTLIER LOCALIZATIONS

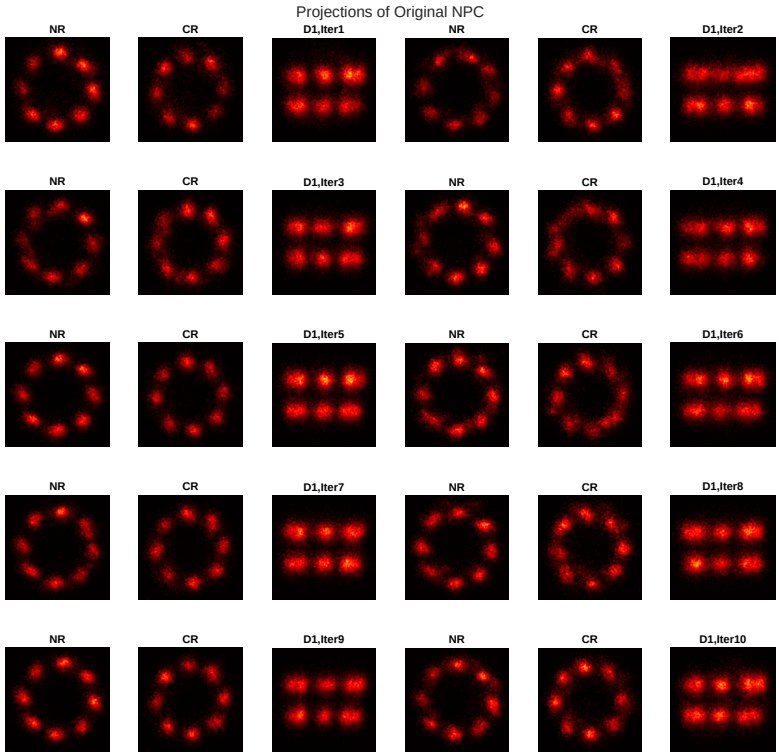


Figure 3.6: Data fusion results for dataset 1 for 10 randomly different GMM initializations without filtering outlier localizations.

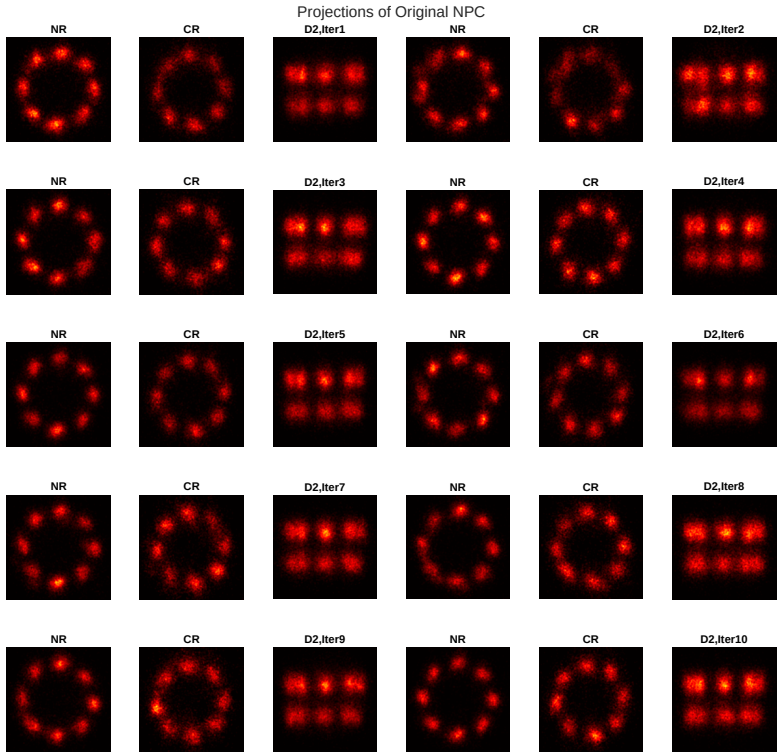


Figure 3.7: Data fusion results for dataset 2 for 10 randomly different GMM initializations without filtering outlier localizations.

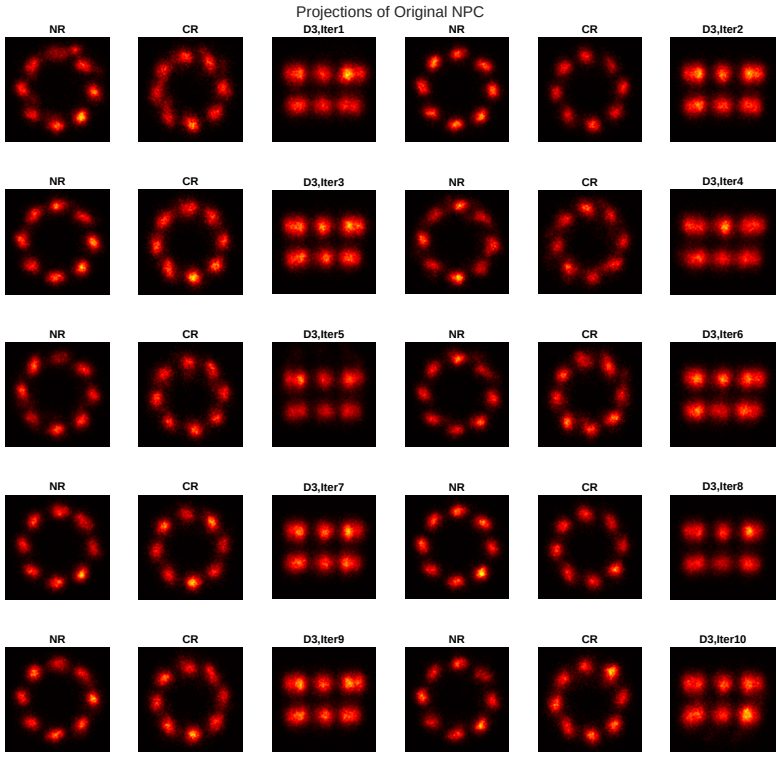


Figure 3.8: Data Fusion results for dataset 3 for 10 randomly different GMM initializations without filtering outlier localizations.

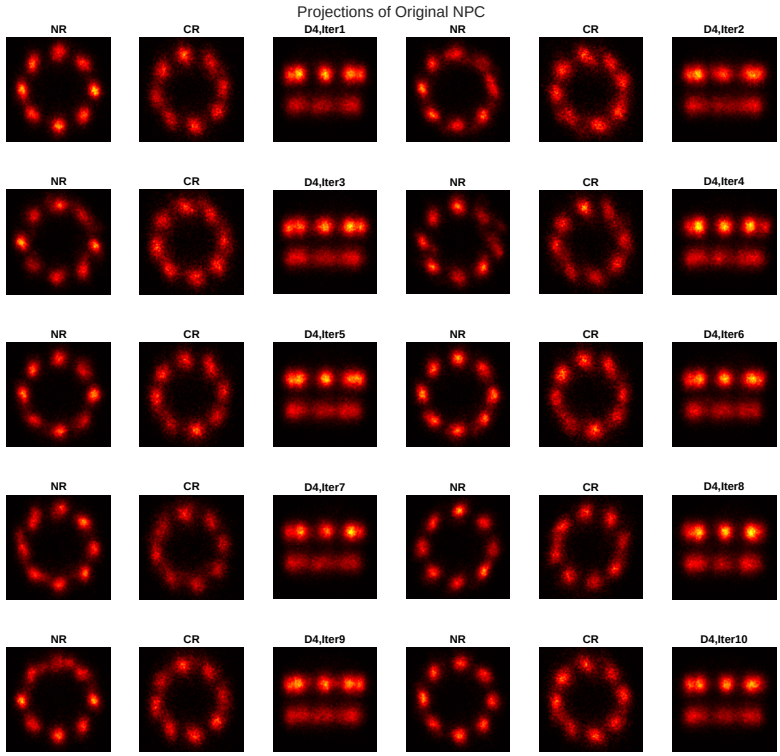


Figure 3.9: Data Fusion results for dataset 4 for 10 randomly different GMM initializations without filtering outlier localizations.

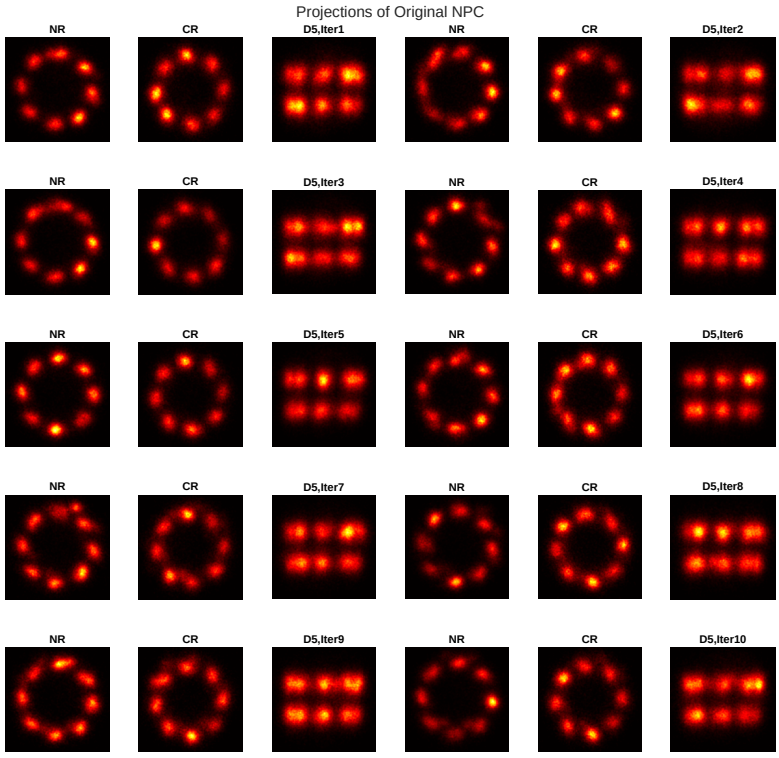


Figure 3.10: Data Fusion results for dataset 5 for 10 randomly different GMM initializations without filtering outlier localizations.

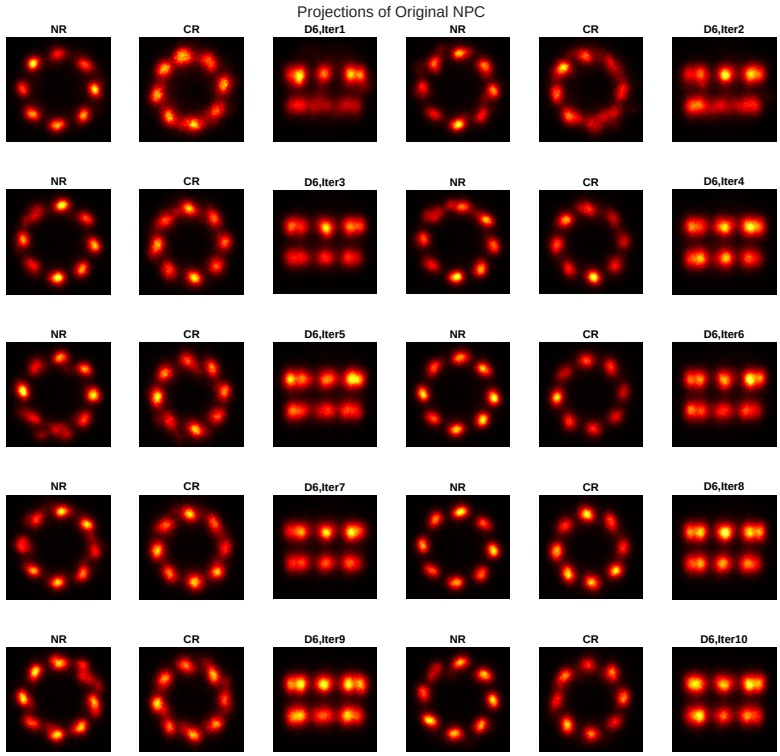


Figure 3.11: Data Fusion results for dataset 6 for 10 randomly different GMM initializations without filtering outlier localizations.

3.9.2. DATA FUSION RESULTS WITH GMM FILTERING OF OUTLIER LOCALIZATIONS

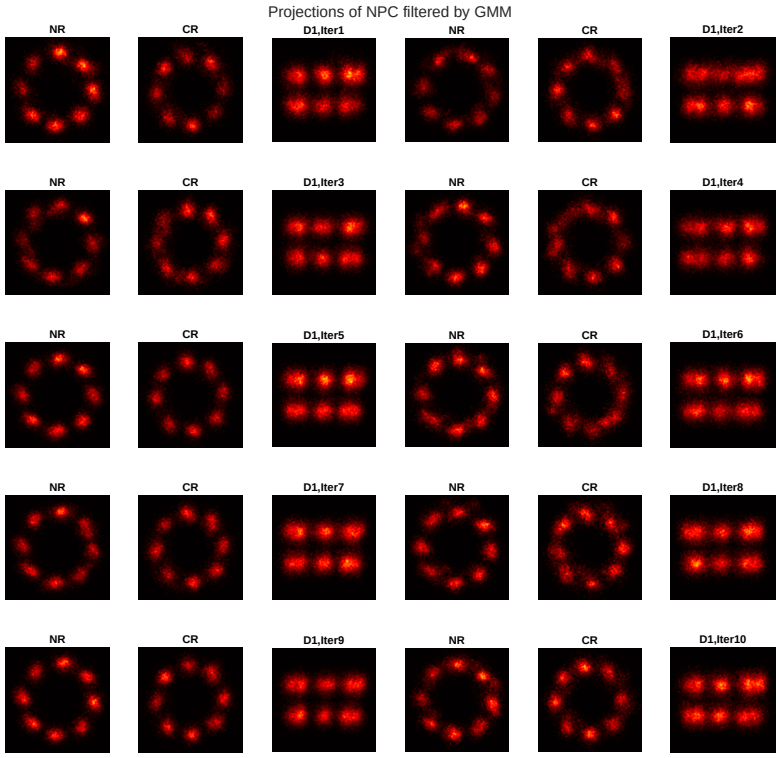


Figure 3.12: Data fusion results for dataset 1 for 10 randomly different GMM initializations with GMM filtering of outlier localizations.

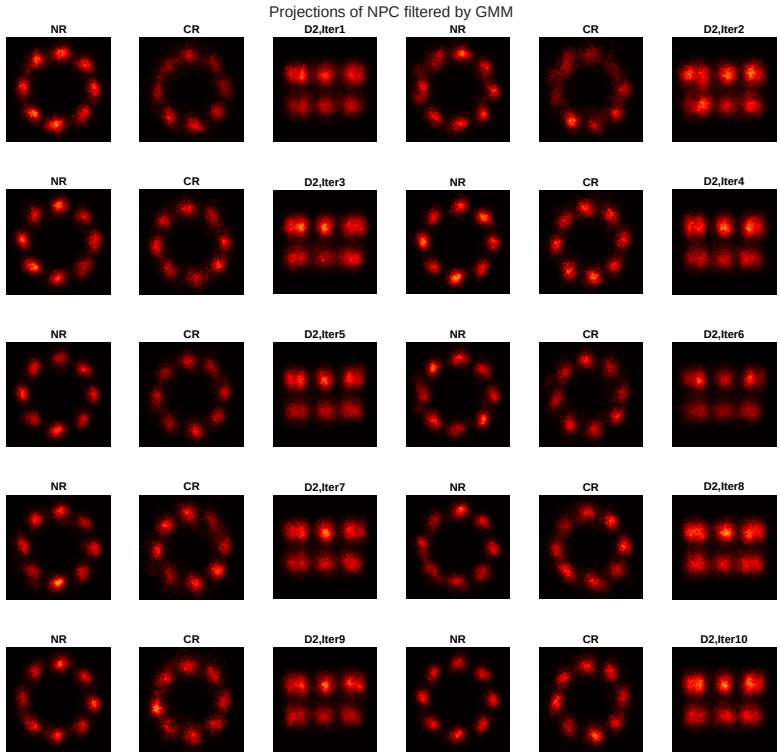


Figure 3.13: Data fusion results for dataset 2 for 10 randomly different GMM initializations with GMM filtering of outlier localizations.

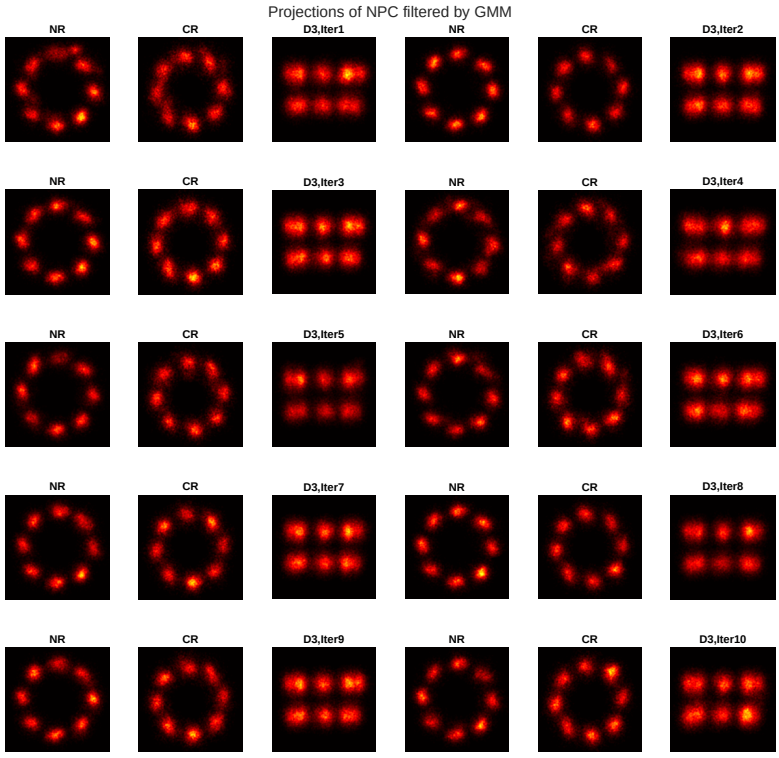


Figure 3.14: Data Fusion results for dataset 3 for 10 randomly different GMM initializations with GMM filtering of outlier localizations.

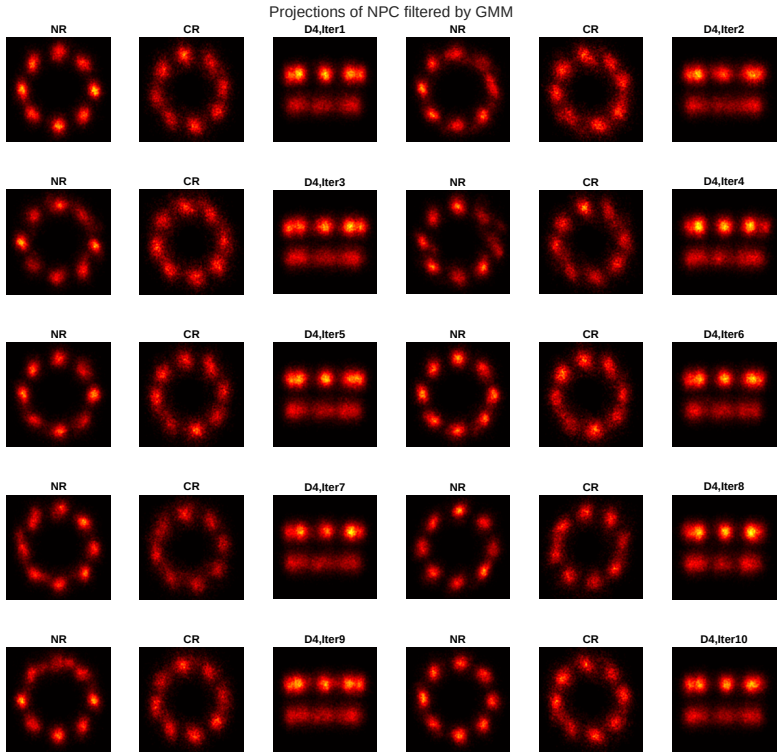


Figure 3.15: Data Fusion results for dataset 4 for 10 randomly different GMM initializations with GMM filtering of outlier localizations.

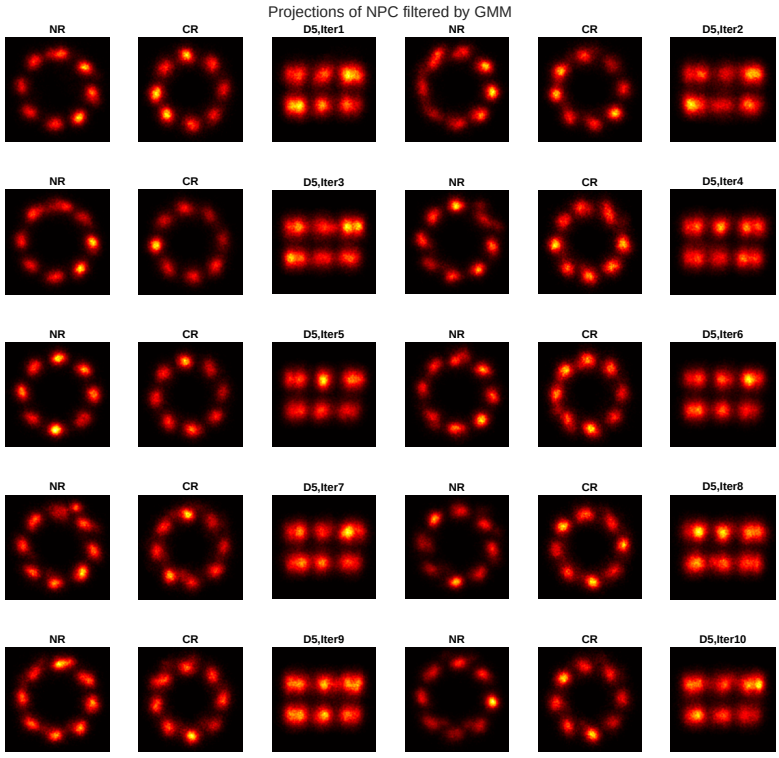


Figure 3.16: Data Fusion results for dataset 5 for 10 randomly different GMM initializations with GMM filtering of outlier localizations.

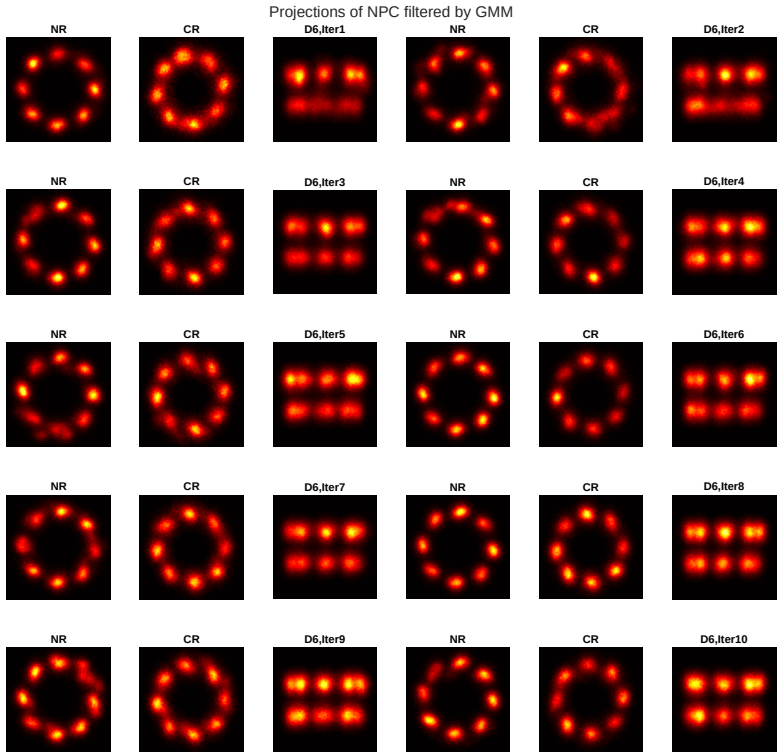


Figure 3.17: Data Fusion results for dataset 6 for 10 randomly different GMM initializations with GMM filtering of outlier localizations.

3.9.3. DATA FUSION RESULTS WITH GMM AND DENSITY FILTERING OF OUTLIER LOCALIZATIONS

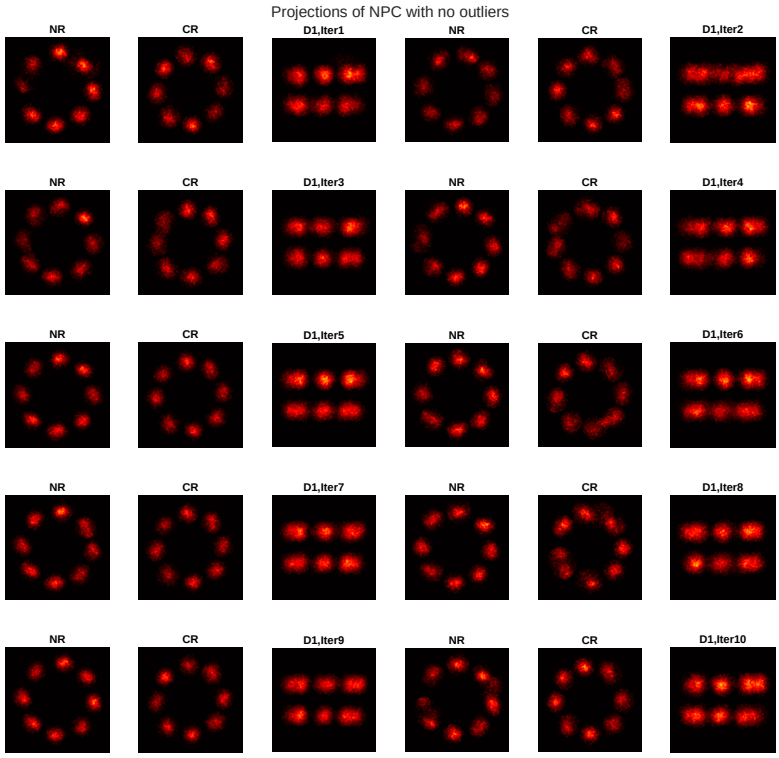


Figure 3.18: Data fusion results for dataset 1 for 10 randomly different GMM initializations with GMM and density filtering of outlier localizations.

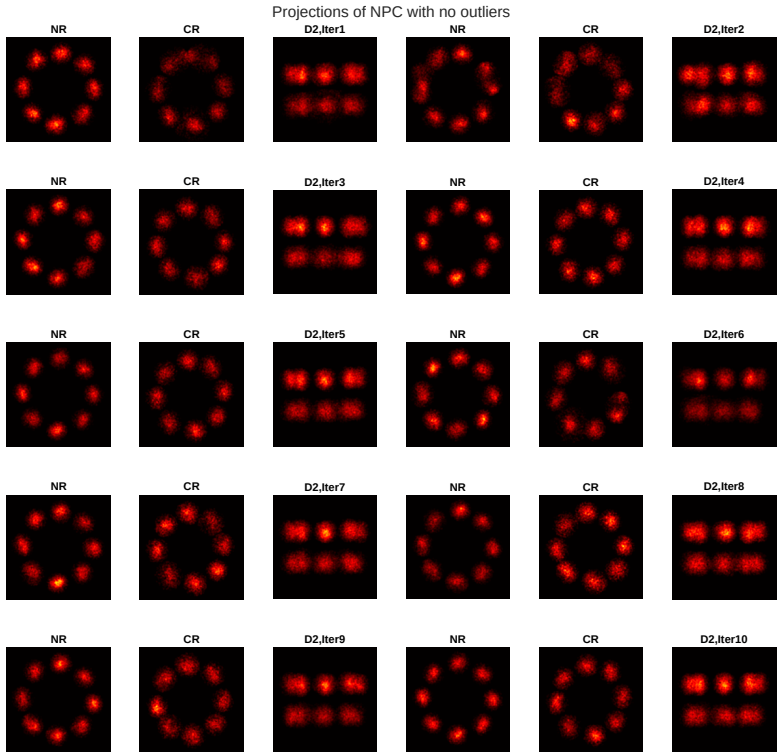


Figure 3.19: Data fusion results for dataset 2 for 10 randomly different GMM initializations with GMM and density filtering of outlier localizations.

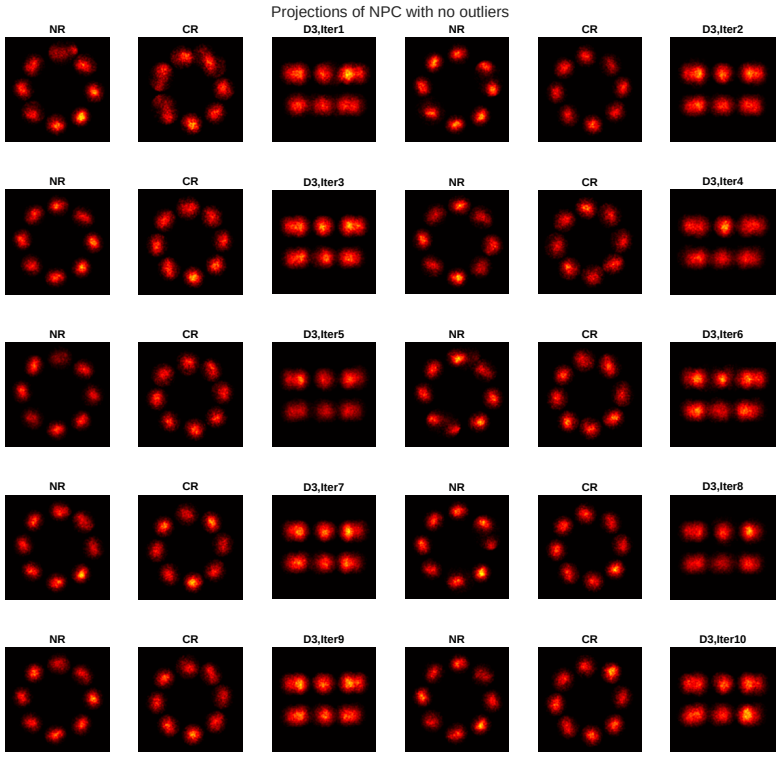


Figure 3.20: Data Fusion results for dataset 3 for 10 randomly different GMM initializations with GMM and density filtering of outlier localizations.

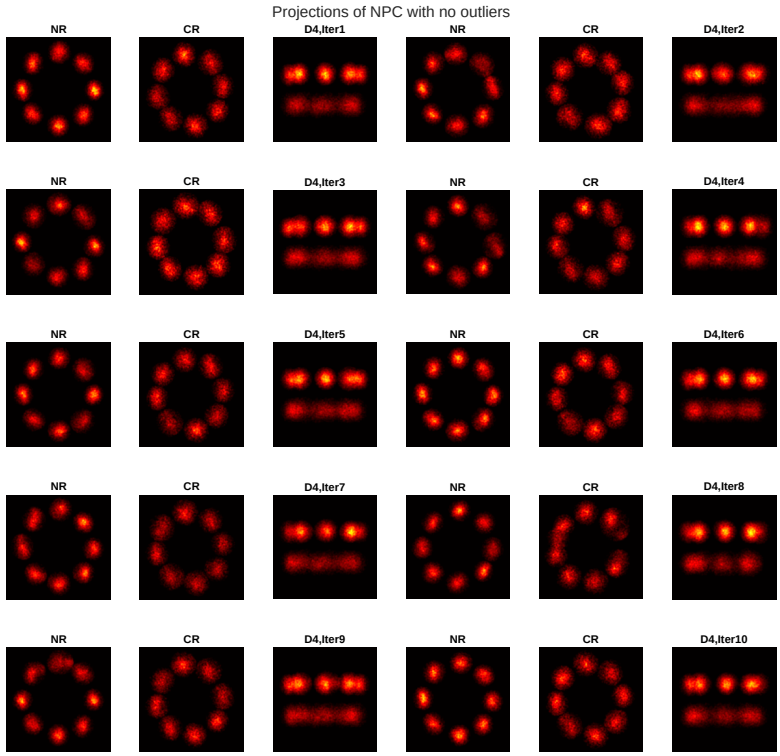


Figure 3.21: Data Fusion results for dataset 4 for 10 randomly different GMM initializations with GMM and density filtering of outlier localizations.

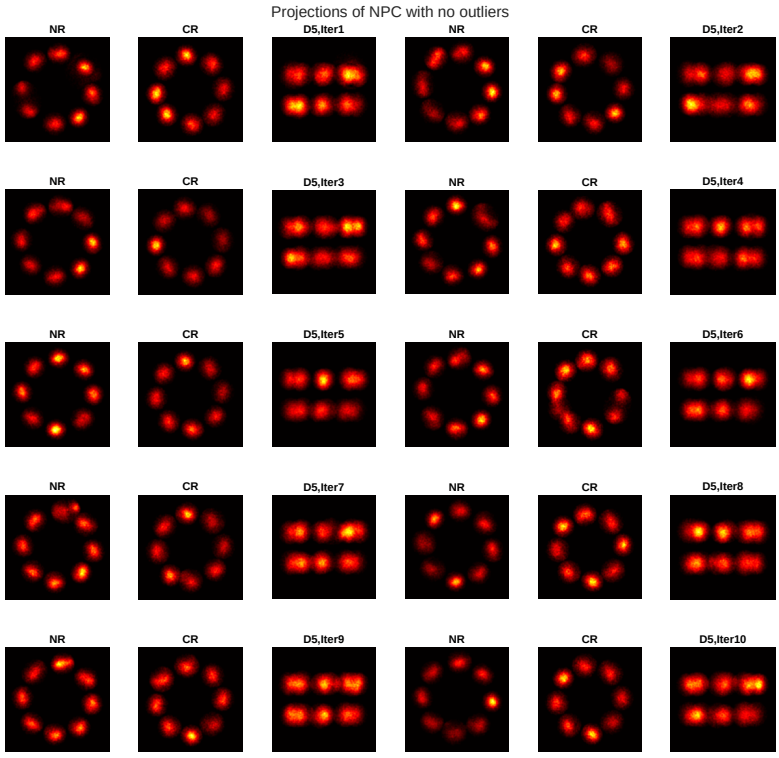


Figure 3.22: Data Fusion results for dataset 5 for 10 randomly different GMM initializations with GMM and density filtering of outlier localizations.

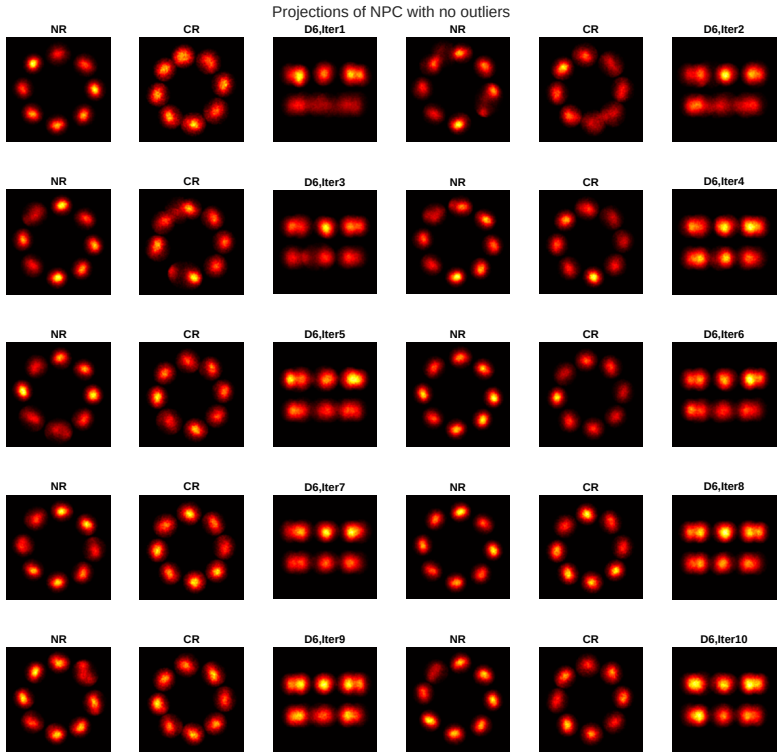


Figure 3.23: Data Fusion results for dataset 6 for 10 randomly different GMM initializations with GMM and density filtering of outlier localizations.

4

IMPACT OF ABSORPTION DIPOLE ORIENTATION ON MINFLUX LOCALIZATION

[1]

MINimal photon FLUXes (MINFLUX) is an emerging super-resolution microscopy technique claiming to achieve 1-3 nm resolution. Dipole orientation effects in the localization method have so far been ignored, which raises concerns about the accuracy and precision of MINFLUX. In our study, we investigate how dipole orientation affects the localization accuracy and precision of MINFLUX. We simulate the image formation process using a physically realistic vector diffraction point spread function (PSF) model for the doughnut excitation, and then localize the emitter based on the simplified Gaussian doughnut PSF model used in MINFLUX so far. Our investigation includes key simulation parameters such as signal photon count, background photons, and distance between the excitation beam centers for both free and fixed dipoles. We find a low to moderate bias, well below the Cramér Rao Lower Bound (CRLB) in case of freely rotating dipoles, except for case of low signal and/or high background. For fixed dipoles, however, we find biases that can be larger than the CRLB, specifically for tilted dipoles. Our findings highlight the need for more accurate fit models to overcome this bias problem, thereby extending the applicability of MINFLUX.

[1] Wenxiu Wang, Bernd Rieger, and Sjoerd Stallinga, in preparation

4.1. INTRODUCTION

The concept of MINimal photon FLUXes (MINFLUX) [1], a novel super-resolution microscopy technique, is built on strategically employing the local minimum of an excitation beam to accurately pinpoint an emitter's position. This approach distinguishes itself from traditional super-resolution methods such as stochastic optical reconstruction microscopy (STORM) [2], photo-activated localization microscopy (PALM) [3], and stimulated emission depletion (STED) [4] microscopy, which typically achieve resolutions between 10-50 nm. MINFLUX, however, boasts an impressive resolution in the small nanometer range.

The principle of MINFLUX is based on a doughnut-shaped excitation beam generated by a vortex phase plate [5]. This beam profile has zero intensity at its center, surrounded by bright ring [6]. If a fluorescent molecule is precisely aligned with the centre of the beam, no photons are detected as the molecule is not excited. However, since the position of the fluorescent molecule is the unknown variable in real experiments, MINFLUX takes a strategic approach. It sequentially illuminates the emitter with the doughnut-shaped beam centered at different positions, in a region of size L. The number of emitted photons generated from the fluorescent molecule excited by these illuminations carries information about the relative distance between the beam centers and the emitter. If the beam centers and the intensity profile of the beam are known, the exact position of the fluorescent molecule can be estimated using Maximum Likelihood Estimation (MLE) [7]. It appears that the localization precision typically scales as $L/\sqrt{N}$, with N the number of detected photons. This scaling is the reason why the localization precision of MINFLUX is favorable compared to conventional single molecule localization approaches, as these scale as $\lambda/\mathrm{NA}/\sqrt{N}$, with λ the wavelength and NA the objective lens' numerical aperture. Namely, the size parameter L can in principle be chosen arbitrarily small, at least much smaller than the diffraction length λ/NA.

In the years since its introduction, MINFLUX has been shown to be capable of three-dimensional imaging [8, 9], of supporting simultaneous imaging in two-color channels [10, 11] and of facilitating observation of live cells [12, 13]. In recent years, several innovative methods derived from MINFLUX have emerged, including SIMFLUX [14], SIMPLE [15], ROSE [16], MODLOC[17], ISM-FLUX [18], DNA-PAINT MINFLUX nanoscopy [19], p-MINFLUX [20] and RASTMIN [21]. MINFLUX has been applied in several sub-cellular imaging studies. It has been used to resolve the active zone of rod photoreceptors [9], to observe motor protein stepping in living cells [12] and to dissect the motions of individual (macro) molecules with unprecedented combinations of spatial and temporal resolution [13].

The foundational paper on MINFLUX [1] claims that "Because C is a prefactor, the molecular orientation has no influence on the solution." Here, "C is a prefactor that is proportional to the molecular brightness and the detection sensitivity, as well as to a parameter describing the molecular orientation in space." This is indeed correct for the emission. However, this perspective overlooks the substantial impact molecular orientation can have on energy *absorption*, which, in turn, affects the number of emitted fluorescence photons. In other words, the emission dipole will not effect MINFLUX but the absorption dipole will. MINFLUX employs a simplified approximation that treats the vectorial electric field at the focal plane as a scalar field and calculates the energy absorbed by fluorescent

molecules without considering their orientation. The disparity in energy absorption due to varying molecular orientations within an identical vectorial electric field could result in model errors, challenging the accuracy and precision of MINFLUX localization.

In this study, we examine the influence of the absorption PSF for both free and fixed dipole orientations on the localization accuracy and precision within the MINFLUX framework. Specifically, we look at the impact of using a simplified doughnut PSF model, as described in the MINFLUX paper [1], compared to a physically realistic vector diffraction PSF model. Our approach is similar in spirit to our previous study on the accuracy and precision of the Gaussian emission PSF model in conventional single molecule localization microscopy [22]. We use the vector diffraction based PSF model to simulate a physically realistic doughnut-shaped PSF. This is subsequently used for generating the photon counts for randomly generated MINFLUX localization instances. These are then fitted using the simplified PSF model and compared to the ground truth parameter values to assess accuracy and precision. We also varied key parameters in this simulation setup, such as the total expected photon count, number of background photons per observation, and the size L of the beam center scan range, to assess their impacts. For absorption dipoles with a fixed orientation we have also investigated the effect of the dipole orientation on the localization outcome.

4.2. THEORY

This section describes our methodology for simulating ground truth MINFLUX data using doughnut-shaped excitation beams computed with vectorial diffraction[23–27] and absorption by fluorescent molecules with fixed or free dipole orientation. We then apply Maximum Likelihood Estimation using a simplified model as described in the original MINFLUX paper [1] to analyze the achievable precision and accuracy in MINFLUX localization.

4.2.1. PHYSICAL OPTICS BASED IMAGE FORMATION MODEL

The incident beam at the back focal plane of the objective lens is characterized by the wavelength (λ) and the amplitude distributions $\left(A_x^{in}, A_y^{in}, A_z^{in}\right)$. To create a doughnut shaped beam in the sample we start with the electric field $\mathbf{E}^{in}$ at any point (x, y) in the back focal plane for a circular polarization which is described by:

$$\mathbf{E}^{in}(x,y) = \begin{bmatrix} E_x^{in}(x,y) \\ E_y^{in}(x,y) \\ E_z^{in}(x,y) \end{bmatrix} = \begin{bmatrix} A_x^{in} \\ A_y^{in} \\ A_z^{in} \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} \\ i/\sqrt{2} \\ 0 \end{bmatrix}. \tag{4.1}$$

The incident beam passes through a vortex phase plate, which has a phase:

$$\gamma(x,y) = \arctan(\frac{y}{x}) + \pi, \tag{4.2}$$

imparted to the beam at any point (x, y) upon passing trough the vortex phase plate. The electric field at the back focal plane is then modified according to:

$$E_m^{back}(x,y) = A_m^{in} \exp\left(i\gamma(x,y)\right), \; m = x, y, z. \tag{4.3}$$

The electric field at the back focal plane is subsequently focused by the objective lens. Given the high numerical aperture (NA) of the objective lens utilized, we must apply the lens matrix to describe the polarization transformation to the front focal plane [23, 24]. The lens matrix is defined as:

$$\mathbf{M}(f_x, f_y) = \begin{bmatrix} \cos\theta\cos^2\phi + \sin^2\phi & (\cos\theta - 1)\sin\phi\cos\phi & -\sin\theta\cos\phi \\ (\cos\theta - 1)\sin\phi\cos\phi & \cos^2\phi + \cos\theta\sin^2\phi & -\sin\theta\sin\phi \\ \sin\theta\cos\phi & \sin\theta\sin\phi & \cos\theta \end{bmatrix}. \tag{4.4}$$

Here $\theta = \arcsin\left(\lambda\sqrt{f_x^2 + f_y^2}/n_r\right)$ and $\phi = \arctan(f_y/f_x)$ are the polar and azimuthal angles in spatial frequency space, respectively, and n_r is the refractive index of the sample. The coordinates f_x and f_y in spatial frequency space are derived from the physical coordinates x, y in the back focal plane as follows:

$$f_x = \frac{x}{\lambda F}, \tag{4.5}$$

$$f_y = \frac{y}{\lambda F}, \tag{4.6}$$

where F is the focal length of the objective lens. The aperture function in the spatial frequency domain is defined as:

$$C\left(f_x, f_y\right) = \mathrm{circ}\left(\frac{\sqrt{f_x^2 + f_y^2}}{NA/\lambda}\right). \tag{4.7}$$

The amplitude of the aperture function is scaled by the aplanatic amplitude factor [24]:

$$C_{norm}\left(f_x, f_y\right) = \frac{E_0}{\sqrt{\cos\theta}} \tag{4.8}$$

where E_0 is a constant factor that is used to normalize the total power passing through the aperture.

The electric field in the spatial frequency domain at the front focal plane is represented as:

$$\begin{bmatrix} E_x\left(f_x, f_y\right) \\ E_y\left(f_x, f_y\right) \\ E_z\left(f_x, f_y\right) \end{bmatrix} = C_{norm}\left(f_x, f_y\right)\mathbf{M}\left(f_x, f_y\right)\begin{bmatrix} E_x^{back}(x, y) \\ E_y^{back}(x, y) \\ E_z^{back}(x, y) \end{bmatrix}. \tag{4.9}$$

The transformation of electric field components from the spatial frequency domain to the spatial domain at the front focal plane can be obtained via inverse Fourier Transform (FT):

$$E_m\left(x, y\right) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} E_m\left(f_x, f_y\right) e^{2\pi i\left(f_x x + f_y y\right)} df_x df_y, \;\; m = x, y, z. \tag{4.10}$$

To obtain the field's overall intensity, we calculate the sum of the squared magnitudes of the vector components of the electric field:

$$I = |E_x|^2 + \left|E_y\right|^2 + |E_z|^2. \tag{4.11}$$

4.2.2. OBSERVED PHOTON COUNTS

The MINFLUX technique uses four exposures to localize a single fluorescent molecule, each utilizing the same doughnut-shaped excitation beam but with differently positioned beam centers. One beam center is placed at the defined origin:

$$[x_0, y_0] = [0,0]. \tag{4.12}$$

The three other beam centers are placed at a distance $L/2$ from the origin, distributed evenly in a circle around it. The positions are defined by:

$$[x_k, y_k] = \frac{L}{2}\left[\cos(\beta_k), \sin(\beta_k)\right], \quad \text{for } \beta_k = k\frac{2\pi}{3} \text{ and } k = 1,2,3. \tag{4.13}$$

The region of interest (ROI) where the emitter is expected to be is inside the circle $x^2 + y^2 \leq (L/2)^2$. We describe the fluorescent molecule as an electric dipole absorber. This dipole is denoted as $\mathbf{d}$ and is located at position (x, y) in the front focal plane. It depends on parameters (ρ, ψ), where the parameter ρ is the polar angle ($0 \leq \rho \leq \pi$), and the parameter ψ is the azimuthal angle ($0 \leq \psi < 2\pi$):

$$\mathbf{d} = \begin{bmatrix} \sin\rho\cos\psi \\ \sin\rho\sin\psi \\ \cos\rho \end{bmatrix}. \tag{4.14}$$

Dipoles with fixed orientation primarily absorb energy from the electromagnetic waves that have a polarization that is aligned with their orientation. This process of energy absorption can be quantitatively described by the following equation:

$$I_k(x, y) = |\mathbf{d} \cdot \mathbf{E}(x - x_k, y - y_k)|^2, \quad k = 0,1,2,3, \tag{4.15}$$

where $I_k(x, y)$ represents the absorbed energy by the dipole at the position (x, y) from the k_{th} exposure, and $\mathbf{E}$ is the vectorial electric field of the doughnut-shaped excitation beam.

For dipoles that can freely rotate during the detector acquisition time, the interaction between the dipole and the electric field is effectively averaged out, resulting in an equal contribution from each direction. Then, the absorbed energy can be expressed as:

$$I_k(x, y) = \frac{1}{3} I(x - x_k, y - y_k), \quad k = 0,1,2,3, \tag{4.16}$$

where $I(x, y)$ represents the overall intensity of the vectorial electric field of the doughnut-shaped excitation beam. The expected observed photon counts μ_k are theoretically proportional to the absorbed energy and include a background offset. This relationship is described by the equation:

$$\mu_k = \eta I_k(x, y) + b, \quad k = 0,1,2,3, \tag{4.17}$$

where η is a proportionality constant and b denotes the expected number of background photons per observation, a non-negligible factor that impacts the detected photon counts. For our simulation we assume that the total number of photon counts (N_p) detected from

the four exposures is constant for better comparability, irrespective of their spatial positions. Thus we have:

$$\eta = \frac{N_p}{\sum_{k=0}^{3} I_k(x,y)}. \tag{4.18}$$

Furthermore, we define the normalized absorbed energy as:

$$I_k^n(x,y) = \frac{I_k(x,y)}{\sum_{k=0}^{3} I_k(x,y)}. \tag{4.19}$$

Consequently, the expected observed photon counts are calculated using:

$$\mu_k = N_p I_k^n(x,y) + b, \quad k = 0,1,2,3. \tag{4.20}$$

The observed photon counts n_k are not equal to the expected photon counts of the forward model μ_k, as the emission is not a deterministic process but rather follows Poisson statistics, which can be mathematically expressed as:

$$n_k = \text{Poisson}(\mu_k). \tag{4.21}$$

This expression accounts for the stochastic nature of photon counting, where n_k is a random variable following a Poisson distribution with the expected value μ_k.

4.2.3. SIMPLIFIED GAUSSIAN DOUGHNUT PSF MODEL

For the estimation of the emitter position we use the simplified model for the doughnut shaped absorption PSF as described in ref. [1]. This simplification characterizes the doughnut shape as a modulated Gaussian function:

$$I_{dk}(x,y) = A_0 \frac{(x-x_k)^2 + (y-y_k)^2}{4\pi\sigma^2} e^{-\frac{(x-x_k)^2+(y-y_k)^2}{2\sigma^2}}, \tag{4.22}$$

where (x_k, y_k) are the coordinates of the k_{th} beam center as described in Equations (4.12) and (4.13), and where A_0 is a pre-factor for characterizing the integrated intensity over the 2D-plane. The width σ is related to the Full Width at Half Maximum (FWHM) of the underlying Gaussian profile by:

$$\sigma = \frac{\text{FWHM}}{\sqrt{2\ln 2}}. \tag{4.23}$$

In our simulation we fit the modulated Gaussian function I_{dk} to the overall intensity distribution I at the focal plane as computed with the physical forward model. This fitting process aims to optimize the parameters A_0, σ , ensuring the best possible approximation of I using I_{dk}.

4.2.4. MAXIMUM LIKELIHOOD ESTIMATOR

Upon retrieving the observed photon counts n_k from the physical optics based forward model and deriving the expression for expected photon counts μ_k from the Gaussian doughnut-shaped model, we are able to infer the location of emitters utilizing these parameters and other prior knowledge by a Maximum Likelihood Estimator.

As the observed photon counts n_k follows Poisson statistics, with probability mass function given by:

$$P(n_k;\mu_k) = \frac{\mu_k^{n_k} e^{-\mu_k}}{n_k!}. \tag{4.24}$$

Given the set of observations n_0, n_1, n_2, n_3, the log-likelihood function of these observations under the Poisson distribution is expressed as:

$$\log L = \sum_{k=0}^{3} [n_k \log \mu_k - \mu_k - \log(n_k!)]. \tag{4.25}$$

Maximizing this log-likelihood function with respect to the fit parameters gives the most likely value for the position of the molecule.

The expected photon count according to the simplified Gaussian doughnut PSF model is, similar to the Poisson rate 4.17 of the forward simulation, given by:

$$\mu_k(\boldsymbol{\tau}) = N_\alpha I_{dk}(x,y) + b. \tag{4.26}$$

In this equation, $\boldsymbol{\tau} = [x, y, N_\alpha, b]$ represents the vector of unknown parameters that we seek to estimate. The parameter N_α is used to control the total number of expected photons. The inclusion of a background parameter b is important to describe realistic experimental data. Background can have a huge impact on the measurements, thereby affecting the accuracy and precision of the MLE.

In the original publication on MINFLUX [1], the actual optimization of the parameters for MLE was tackled using a grid search method. We adopt a more direct Levenberg–Marquardt (LM) non-linear optimization algorithm for this purpose. This choice is driven by the algorithm's efficiency, robustness and effectiveness. The LM algorithm offers a potentially faster convergence compared to the Grid Search (GS) method.

The GS method's duration is relevant to both the search range and the set precision. A large range coupled with a small step size can lead to a substantial time consumption. For the purpose of our comparison, we conducted a uniform sampling over a square range of -50 nm to 50 nm, creating a grid of 51x51 points. The correct emitter positions were then searched within a larger range of -100 nm to 100 nm. We repeated the calculation of emitter positions ten times using both the LM and GS methods. The LM method completed all emitter position calculations within an average time of 6.1±0.8 seconds. For the GS method, rather than recalculating $\log(p_k)$ for each emitter every time, we optimized our computations by performing this calculation just once for the entire ROI at each search precision level, and then reused these results in the GS search. This approach significantly enhanced efficiency. When the precision was set to 1 nm, the computational time for solving 51x51 emitter positions across the entire ROI was approximately 16.5±0.1 seconds. This increased to around 66.6±0.4 seconds when the precision was refined to 0.5 nm. Given the necessity to accurately determine the positions of all points within the entire ROI, the LM algorithm emerges as a better choice over GS. This preference is anchored in the need for high precision and the ability to efficiently process extensive data sets, making LM a strategically sound and efficient approach for our analysis.

To determine the initial estimate for the parameters (x_i, y_i), a straightforward triangulation approach can be employed. This process begins by defining quantities m_k for each

set of observed photon counts:

$$m_k = \left(\sum_{j=0}^{3} n_j\right) - n_k. \tag{4.27}$$

Subsequently, the initial guesses of the coordinates are calculated as:

$$x_i = \frac{\sum_{k=0}^{3} m_k x_k}{\sum_{k=0}^{3} m_k}, \tag{4.28}$$

$$y_i = \frac{\sum_{k=0}^{3} m_k y_k}{\sum_{k=0}^{3} m_k}. \tag{4.29}$$

The rationale behind this choice is that the higher the m_k will be, the closer the emitter is expected to be to the k_{th} beam center.

The initial parameter N_i and b_i are determined by solving the linear least-squares problem formulated with the initial set of coordinates. The optimization problem is expressed as:

$$\min_{\mathbf{x}} \frac{1}{2} \left\| \begin{bmatrix} I_{d_0}(x_i, y_i) & 1 \\ I_{d_1}(x_i, y_i) & 1 \\ I_{d_2}(x_i, y_i) & 1 \\ I_{d_3}(x_i, y_i) & 1 \end{bmatrix} \begin{bmatrix} N_i \\ b_i \end{bmatrix} - \begin{bmatrix} n_0 \\ n_1 \\ n_2 \\ n_3 \end{bmatrix} \right\|_2^2. \tag{4.30}$$

The parameter boundaries are defined to align with the experimental constraints and data characteristics. The boundaries for coordinates x and y are uniformly set at $[-L, L]$, wider than the circle with radius $L/2$ that is the region of ground truth emitter positions that is considered relevant. This specific setting reflects the effective region for localization. For the parameter N, we use a scaling factor of 1.5 expressed as:

$$N \in \left[\frac{N_p}{1.5 \times \max\left(\sum_{k=0}^{3} I_{dk}(x,y)\right)}, \frac{1.5 \times N_p}{\min\left(\sum_{k=0}^{3} I_{dk}(x,y)\right)} \right]. \tag{4.31}$$

This adjustment is crucial for ensuring algorithmic convergence, especially in light of the Poisson distribution's potential to generate excessively large or small values of $\sum n_k$, which could otherwise cause the algorithm to become stuck in a local optimum. The parameter $b \in [0, 10]$ is set to the expected range of small background photon counts.

4.2.5. CRAMÉR RAO LOWER BOUND

The Cramér Rao Lower Bound (CRLB) provides a theoretical benchmark for the minimum variance of an unbiased estimator. As such, the CRLB gives a good evaluation for the performance of the localization performance of MINFLUX. Originally, the CRLB was explicitly formulated for the 2D Gaussian doughnut PSF model, focusing on the two variables x and y, and excluding the impact of estimating the signal strength and background [1]:

$$\sigma_{CRLB} = \sqrt{\frac{1}{2N_p} \frac{\sum_{k=0}^{3} \left(\frac{1}{p_k}\right) \left[\frac{\partial p_k}{\partial x}^2 + \frac{\partial p_k}{\partial y}^2\right]}{\left[\sum_{k=0}^{3} \left(\frac{1}{p_k}\right) (\frac{\partial p_k}{\partial x})^2\right] \left[\sum_{k=0}^{3} \left(\frac{1}{p_k}\right) (\frac{\partial p_k}{\partial y})^2\right] - \left[\sum_{k=0}^{3} \left(\frac{1}{p_k}\right) (\frac{\partial p_k}{\partial x} \frac{\partial p_k}{\partial y})\right]^2}}, \tag{4.32}$$

where p_k is the scaled intensity of the I_{dk},

$$p_k = \frac{I_{dk}(x,y)}{\sum_{k=0}^{3} I_{dk}(x,y)}. \tag{4.33}$$

Given the expansion of our expected photon count model to encompass four variables $\boldsymbol{\tau} = (x, y, N_\alpha, b)$, it becomes imperative to develop a revised CRLB suitable for evaluating the efficacy of our MLE strategy. The Fisher Information is a measure of the amount of information that the observed variables carry about the unknown parameters. For the case of multiple parameters, the CRLB is given by the inverse of the Fisher Information Matrix F, defined by:

$$F_{ij} = \sum_{k=0}^{3} \frac{1}{\mu_k(\boldsymbol{\tau})} \left(\frac{\partial \mu_k(\boldsymbol{\tau})}{\partial \tau_i}\right)\left(\frac{\partial \mu_k(\boldsymbol{\tau})}{\partial \tau_j}\right). \tag{4.34}$$

The appearing partial derivatives are given by:

$$\frac{\partial \mu_k(\boldsymbol{\tau})}{\partial x} = \frac{N_p}{\sum_{k=0}^{3} I_{dk}(x,y)} \frac{\partial I_{dk}(x,y)}{\partial x}, \tag{4.35}$$

$$\frac{\partial \mu_k(\boldsymbol{\tau})}{\partial y} = \frac{N_p}{\sum_{k=0}^{3} I_{dk}(x,y)} \frac{\partial I_{dk}(x,y)}{\partial y}, \tag{4.36}$$

$$\frac{\partial \mu_k(\boldsymbol{\tau})}{\partial N_\alpha} = I_{dk}(x,y), \tag{4.37}$$

$$\frac{\partial \mu_k(\boldsymbol{\tau})}{\partial b} = 1. \tag{4.38}$$

The partial derivatives of $I_{dk}(x,y)$ are given by:

$$\frac{\partial I_{dk}(x,y)}{\partial x} = \frac{A_0}{2\pi\sigma^2}(x - x_k)\left(1 - \frac{r_k^2(x,y)}{2\sigma^2}\right)\exp\left(-\frac{r_k^2(x,y)}{2\sigma^2}\right), \tag{4.39}$$

$$\frac{\partial I_{dk}(x,y)}{\partial y} = \frac{A_0}{2\pi\sigma^2}(y - y_k)\left(1 - \frac{r_k^2(x,y)}{2\sigma^2}\right)\exp\left(-\frac{r_k^2(x,y)}{2\sigma^2}\right), \tag{4.40}$$

where, $r_k(x,y) = \sqrt{(x-x_k)^2 + (y-y_k)^2}$ is the distance between the molecule's position (x,y) and the k_{th} beam center. The CRLB for the estimation of the position of the emitter is finally found from the diagonal elements of the inverse of the Fisher-matrix:

$$\sigma_{CRLBxy} = \sqrt{\frac{1}{2}\left(F_{11}^{-1} + F_{22}^{-1}\right)}. \tag{4.41}$$

4.3. NUMERICAL RESULTS

We set the parameters according to ref. [1] as follows: For the excitation beam, we take a wavelength of $\lambda = 552\ nm$, a numerical aperture of the optical system equal to 1.49 and the sample is on a coverslip with a refractive index of $n_r = 1.51$. This combination gives an I_{dk} with FWHM $\approx$ 248 nm and $A_0 \approx 1.32$.

The vectorial absorption PSF derived from the physical forward model and the optimally fitted Gaussian model are shown in Fig. 4.1. The difference between the vectorial PSF and the Gaussian doughnut PSF model is also calculated and presented in order to quantify the error. Overall, the Gaussian doughnut PSF model works quite well, with errors on the order of a few percent.

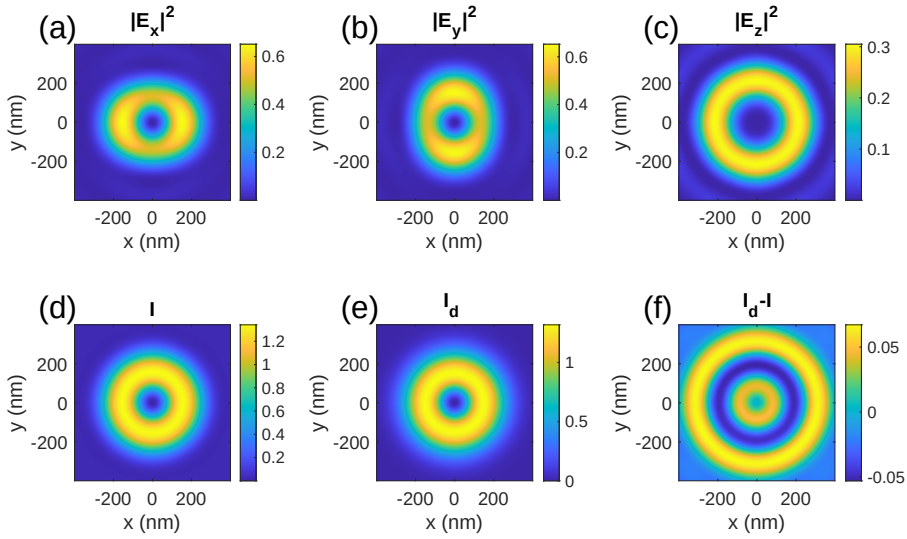


Figure 4.1: Electric field components derived from the physical optics based forward model, the overall intensity, and the intensity of the optimally fitted Gaussian doughnut PSF model. (a) $|E_x|^2$, (b) $|E_y|^2$, (c) $|E_z|^2$, (d) total intensity $I = |E_x|^2 + |E_y|^2 + |E_z|^2$, (e) intensity distribution of the optimal fitted Gaussian model I_d, (f) difference between I_d and total intensity I.

We performed a range of simulations, where we varied the signal photon counts N_p, the distance between beam centers $L/2$, and the constant background photon per observation b, for both freely rotating dipoles and for dipoles with a fixed orientation, with the goal to assess the accuracy and precision of MINFLUX estimation with the simplified PSF model. We change one parameter at a time to control the number of variables in the simulation. This setup enables a focused study of the behaviour of the dipoles in a controlled environment, with particular emphasis on the effect of one variation on their properties and interactions. For each parameter set, we generate the observed photon counts n and apply MLE to localize the dipole position in a single simulation run. Given the random nature of the photon counts in a Poisson distribution, we use 100 simulation instances for each data point. This makes calculation of the the standard deviation and the average bias possible for each point, providing a more robust statistical understanding.

We go through an area within a box, spanning from $-L/2$ to $L/2$, which includes the ROI defined by the circle $x^2 + y^2 \leq (L/2)^2$. Proceeding in steps of $L/50$, we construct a comprehensive performance map for the entire area. In addition, a CRLB map is generated for

each parameter set to facilitate comparison with the statistical performance map in terms of accuracy and precision.

In the series of figures presented in this section from 4.3.1 to 4.3.3 and section 4.3.4, a consistent layout and annotation scheme is employed for clarity and comparability. Each figure is organized into three columns.

The first column shows the bias of the calculated localization relative to the true emitter position:

$$\Delta_{xy} = \sqrt{(x_m - x_r)^2 + (y_m - y_r)^2}, \tag{4.42}$$

where (x_m, y_m) represent the average of the estimated positions computed from 100 simulation instances and (x_r, y_r) stand for the true coordinates of the emitter. The directional bias is indicated by red arrows. These arrows point from the true emitter position towards the estimated emitter location. The size of these arrows is multiplied by a scaling factor to enhance visibility and directional clarity. This scale factor is noted above the figures for reference.

The second column displays the xy standard deviation of the localizations, expressed in the formula:

$$\sigma_{xy} = \sqrt{\frac{\sigma_x^2 + \sigma_y^2}{2}}, \tag{4.43}$$

where σ_x and σ_y are the standard deviations of the x and the y estimations, respectively, computed from 100 simulation instances. The standard deviation σ_{xy} quantifies the precision of each measurement.

The third column shows the CRLB maps, offering a theoretical lower bound on the variance for unbiased estimators. Within the third column, four red asterisks are used to indicate the centers of the excitation beams, while the blue circle denote the ROI. Each row in these figures corresponds to a distinct set of specific parameters, the three columns in each row share a common colorbar, facilitating ease of comparison. Above each subfigure, the average values for the respective map are provided. Additionally, to the left of the images in the most left column, the names and numerical values of the variables pertinent to each figure are displayed.

In Section 4.3.5, the figures displayed have identical first and second columns as those in sections 4.3.1 to 4.3.3. Regarding the third column, since the computation of the CRLB map does not consider variations in dipole orientations, the CRLB maps adds no new information. Instead, we calculate the sum of the absolute differences between two key components: the scaled intensity p_k as outlined in Eq. 4.33 from the modulated Gaussian model, and the normalized absorbed energy I_k^n from Eq. 4.19. The formula used is:

$$I_{diff}(x, y) = \sum_{k=0}^{3} |p_k(x, y) - I_k^n(x, y)|. \tag{4.44}$$

This column can be used as a quantitative indicator for the model mismatch. The expectation is that magnitude of the biases found from the simulations correlate with this quantitative model mismatch.

4.3.1. PHOTON COUNT

In this subsection, we present the results for free dipoles under different conditions of expected total photon count (N_p = 10, 50 and 300). We assume there is no background ($b = 0$) and that the size of search area is $L = 100\ nm$.

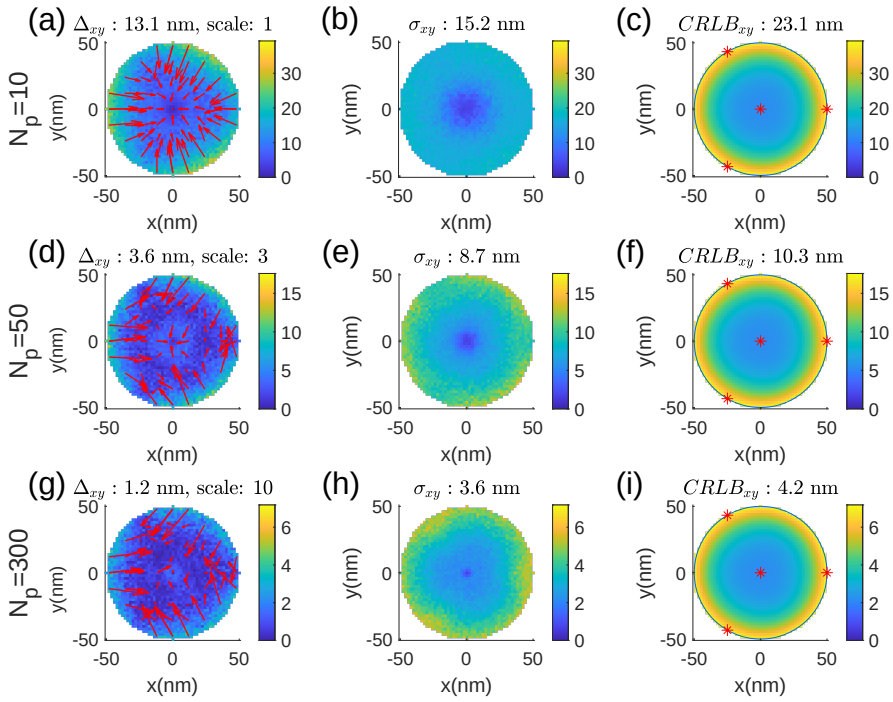


Figure 4.2: Comparative analysis of statistical accuracy and precision in localization of freely rotating dipoles across total photon counts (N_p=10, 50, 300) for $L = 50\ nm$ and $b = 0$.

As the total number of photons increases, a significant reduction in the values of bias, standard deviation and CRLB is observed, meaning a better performance with more photon counts, as expected. All maps show lower values in the center, with values increasing towards the ROI boundary. While the CRLB map shows rotational symmetry, the bias and standard deviation maps show a three-fold rotational symmetry correlated to the orientation of the three outer beam center positions. The bias is smallest for emitter positions close to the beam center positions. In particular, the bias map shows a tendency for the estimated localizations towards the center of the ROI. The precision follows the CRLB for emitter positions close to the center of the ROI. Surprisingly, it is slightly below the CRLB for emitter positions close to the rim of the ROI. This is no contradiction to the lower bound statement, as there is a non-zero bias (the precision is strictly equal to or larger than the CRLB for an unbiased estimator). The bias is approximately 2 (for low N_p) to 4 (for large N_p) times smaller than the CRLB, indicating that bias is a major concern only for

very low photon counts.

4.3.2. DISTANCE BETWEEN BEAM CENTERS

In this subsection, we present the results for freely rotating dipoles under sizes of the search area (L=10, 50, 100 nm). We assume absence of background photons ($b = 0$) and the expected total photon number $N_p = 50$.

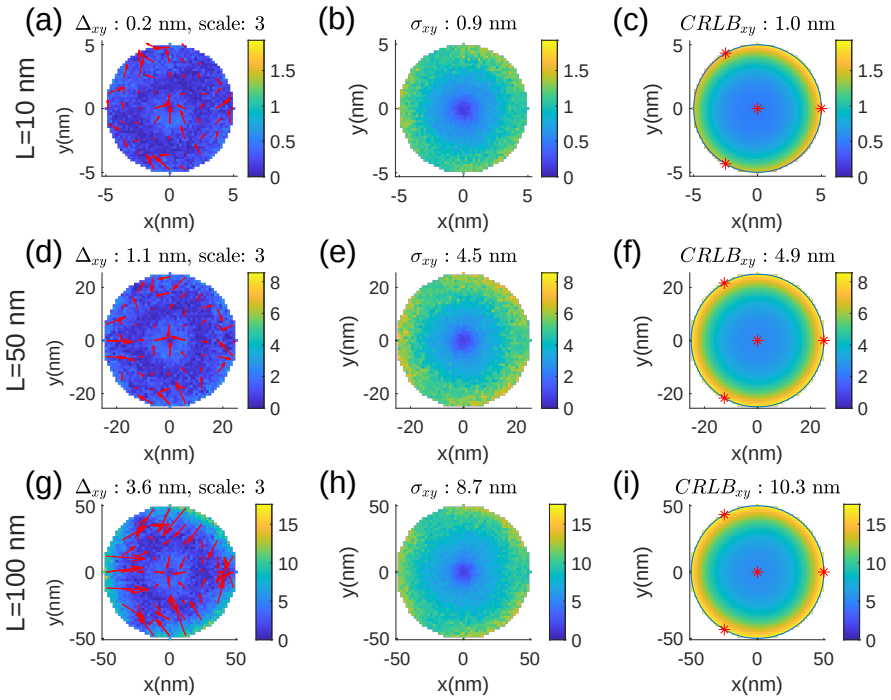


Figure 4.3: Comparative analysis of statistical accuracy and precision in localization of freely rotating dipoles across different sizes of the search area (L=10, 50, 100 nm), for $N_p = 50$ and $b = 0$.

As the distance between beam centers increases, the values of bias, standard deviation, and CRLB increase. This trend indicates better performance with a smaller distance between the beam centers, as expected. The magnitude of the bias compared to the standard deviation and CRLB is smaller for small L, because small model errors in the doughnut are then less important (only the parabolic valley close to the beam center matters for very small L).

4.3.3. CONSTANT BACKGROUND PHOTON PER OBSERVATION

In this subsection, we present the results for freely rotating dipoles for different values of the background photon count per observation ($b = 0, 2, 4, 6$). We assume a size of the

search area $L = 100$ nm and an expected total signal photon number $N_p = 50$.

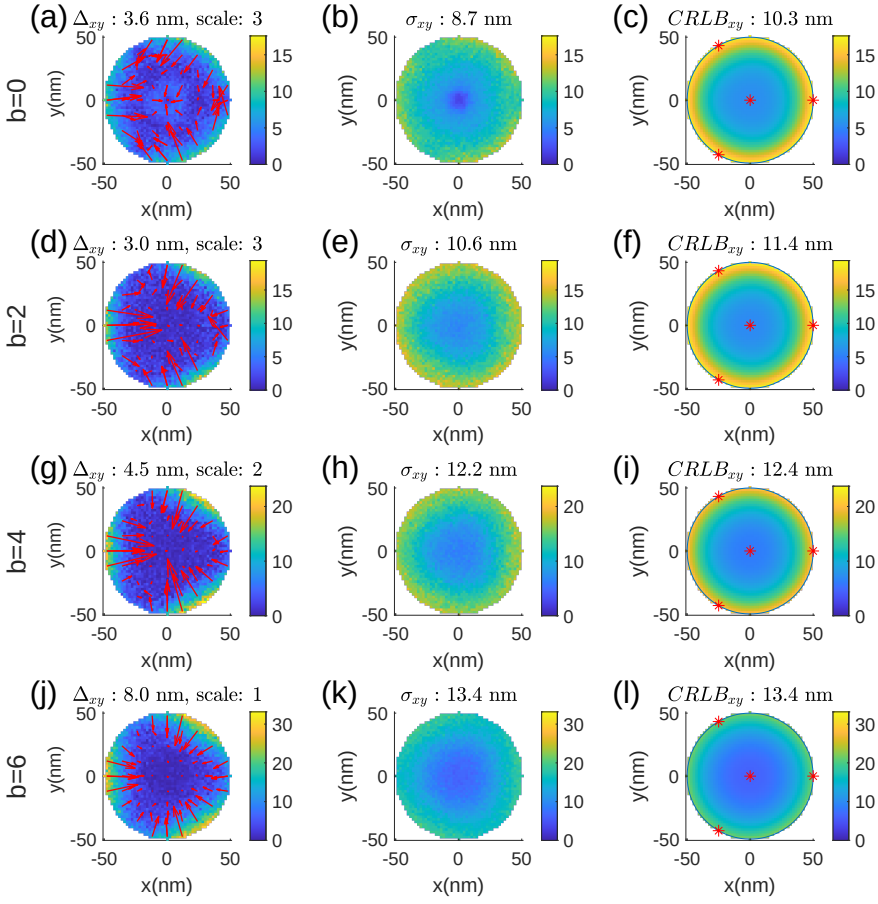


Figure 4.4: Comparative analysis of statistical accuracy and precision in localization of freely rotating dipoles for different values of the background (b=0, 2, 4, 6), for $N_p = 50$ and $L = 100\ nm$.

Overall, an increase in background photons leads to a rise in bias, standard deviation, and CRLB. When the background photon count (b) is lower the standard deviation (σ_{xy}) is slightly less than the CRLB. However, as the background photon count increases($b = 4,6$) the standard deviation becomes more comparable to the CRLB. With an increase in b, the average bias shifts from around 26% of the CRLB to approximately 60%. This indicates that the bias is more significantly influenced by the background noise compared to both the standard deviation or CRLB. When the background photon count increases from 0 to

2, there is a reduction in bias. This might be because at $b = 0$, the fourth parameter $\tau(4) = b$ acts as a redundant term in the MLE (it is fitted, but could have better be ignored), resulting in less accurate estimations of $\tau(1) = x$ and $\tau(2) = y$. In practise this is of course not desirable as we do not know the background count. Additionally, the three-fold symmetry in the bias becomes more marked with higher background levels, suggesting that background has a greater impact on the regions near the edges. The performance for emitter locations close to the beam centers are comparatively less affected by background photons. In the bias map, the estimated positions are still closer to the center than the true positions.

Generally, the average bias values are typically quite a bit smaller than the average CRLB values across most scenarios. This indicates that bias is not a major concern for free rotating dipoles, with the exception of cases of large background and/or low signal photon count. Furthermore, all estimated positions tend to be biased towards the center relative to the true positions. Special attention is needed for regions close to the edges of the ROI, as these areas often exhibit the highest bias and standard deviation. Most of the average values of standard deviation maps in the case of freely rotating dipoles are slightly smaller than the average value CRLB maps, suggesting that the estimation method is efficient. It's important to note that a fair comparison between the standard deviation and the CRLB assumes a near-zero bias. However, in our case, the bias is usually not zero.

4.3.4. ANISOTROPY ANALYSIS OF THE LOCALIZATION

In this subsection we describe the results of simulations on freely rotating dipoles with respect to difference in x and y localization. These experiments were performed under these conditions: the size of the search area is set to $L = 100$ nm, no background photons ($b = 0$), and the expected total number of photons $N_p = 50$. We present the bias, standard deviation and CRLB for both x and y directions, as well as for the (x, y) localization. The results are organized in different rows to highlight the anisotropic characteristics of these results. The bias x and y are calculated by:

$$\Delta_x = x_m - x_r, \tag{4.45}$$

$$\Delta_y = y_m - y_r, \tag{4.46}$$

where (x_r, y_r) are the ground truth emitter coordinates and (x_m, y_m) are the average estimated emitter coordinates.

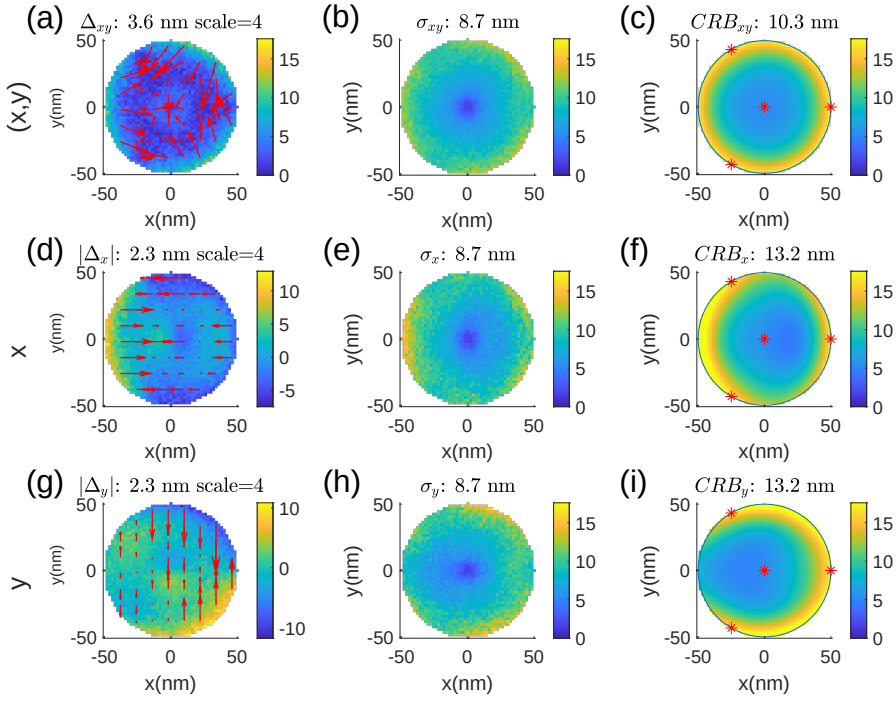


Figure 4.5: Anisotropy analysis of statistical accuracy and precision in localization of free dipoles for $N_p = 50$, $L = 100\ nm$ and $b = 0$.

The central area of the ROI exhibits isotropy, while the anisotropy is pronounced at the edge of the search area. Measurement variability systematically increases away from the center. The bias and standard deviations in the first (for (x, y)) row show three-fold symmetry, while the deviations in the second (for x direction) and third (for y direction) rows show roughly rotational symmetry, and the distribution of the standard deviations is similar to that of the CRLB, with lower values for bias and standard deviation around the beam center.

4.3.5. FIXED DIPOLE ORIENTATION

In this subsection, we present the results for fixed dipoles under different conditions of polar angles ρ and azimuthal angles ψ. We assume the absence of background photons ($b = 0$), the size of search area is $L = 100\ nm$, and the expected total photon number $N_p = 50$. This set of the three parameter (b, L and N_p) share an identical CRLB map as shown in Fig. 4.4(c), as the CRLB map's computation does not account for variations in dipole orientations.

We present a set of plots illustrating the fixed dipole absorption PSF, as formulated in Eq. 4.15. Fig. 4.6 provides characteristics of the fixed dipole PSF and the arrangement within the figure comprises three rows, each describes different polar angles ρ(0, 45, and

90 degrees), with accompanying columns representing the azimuthal angles ψ (0, 45, and 90 degrees). This series of figures share the same FOV.

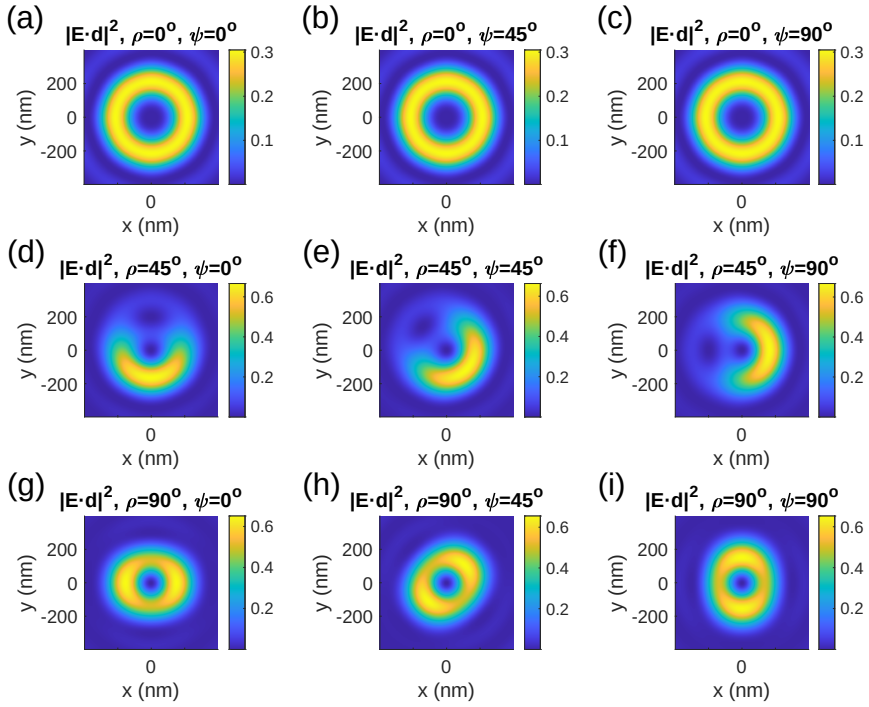


Figure 4.6: Fixed absorption dipole PSF characteristics at different emitter orientations ($\rho = 0^\circ, 45^\circ, 90^\circ$; $\psi = 0^\circ, 45^\circ, 90^\circ$) for $L = 100$ nm and $b = 0$.

The absorption PSF at $\rho = 0^\circ$ is independent of the azimuthal angle, essentially mirroring the $|E_z|^2$ profile as depicted in Fig. 4.1c. In contrast, when ρ is either 45° or 90°, the absorption PSFs within the same row exhibit identical shapes but differ in orientation, rotating in accordance with the changes in the azimuthal angle. At $\rho = 45^\circ$, the pronounced asymmetry in the shape of the absorption PSF is significant, highlighting the substantial impact of the dipole orientation on the absorption PSF. Furthermore, when considering the case at $\rho = 90^\circ$, the relationship between the absorption PSF and the electric field components becomes more evident. Specifically, at $\psi = 0^\circ$, the absorption PSF aligns with $|E_x|^2$ (Fig. 4.1a), and at $\psi = 90^\circ$, it corresponds to $|E_y|^2$ (Fig. 4.1b). This correlation reveals the directional dependency of the absorption PSF and how the dipole orientation and electric field components dictate its spatial distribution.

Through this detailed visualization, it becomes apparent that not all fixed dipole absorption PSFs show the rotationally symmetric doughnut shape associated with freely rotating dipoles. This difference is crucial for our analysis, as it offers a more intricate view of the fixed dipole' absorption PSF behavior.

In Fig. 4.7, we keep the azimuthal angle $\psi = 0^\circ$ and change the polar angle ($\rho = 0^\circ, 45^\circ, 90^\circ$).

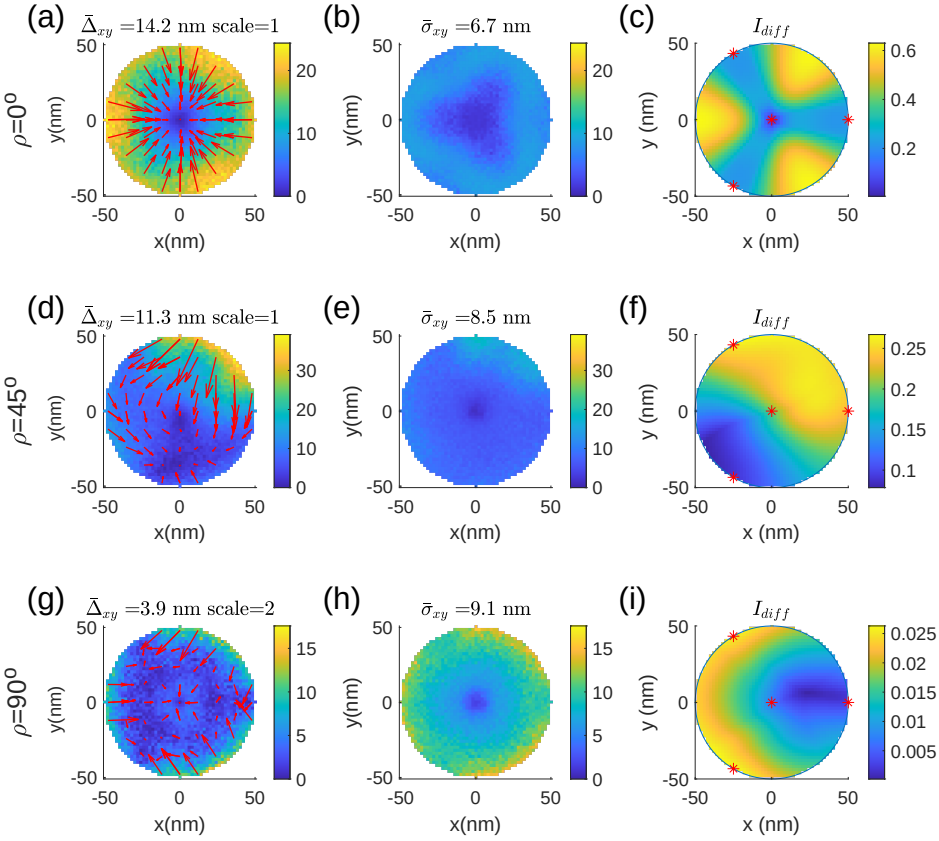


Figure 4.7: Comparative analysis of statistical accuracy and precision in localization of fixed dipoles for different polar angles ($\rho = 0^\circ, 45^\circ, 90^\circ$ and azimuthal angle $\psi = 0^\circ$), for $N_p = 50$, $L = 100$ nm and $b = 0$

Generally, for all polar angles, regions close to the center of the ROI exhibit lower bias and standard deviation. For emitter positions close to the circular edge of the ROI the bias grows considerably, and is larger than for the freely rotating dipoles. When the polar angle $\rho = 0^\circ$, indicating the dipole's perpendicular orientation to the focal plane, both the bias and standard maps exhibit a three-fold symmetry. This symmetry arises because the dipole's energy absorption is influenced solely by the E_z component, which possesses rotational symmetry. Therefore, the placement of excitation beams is the primary factor

affecting this three-fold symmetry. The bias is oriented towards the center of the ROI, which may be related to a mismatch in the actual doughnut size compared to the assumed model doughnut spot. The magnitude of the bias is quite large, about twice as large as the CRLB. For $\rho = 45^\circ$ the bias is smaller in magnitude, and is comparable to the CRLB. The anisotropy in the bias map is large, because the asymmetry induced by the dipole tilt interferes with the three-fold rotational symmetry of the beam center positions. At $\rho = 90^\circ$, where the dipole aligns parallel to the focal plane, the bias is again smaller, and is just a bit larger than for the free dipole case. In all cases, the behaviour found for the bias is strongly correlated with the quantitative model mismatch in the third column.

In Fig. 4.8, we keep the polar angle $\rho = 90^\circ$ (in-plane dipole) and change the azimuthal angle ($\psi = 0^\circ, 45^\circ, 90^\circ$). The impact of the in-plane orientation of the fixed dipole is small. Only a minor perturbation of the three-fold rotational symmetry in the bias map can be observed.

In Fig. 4.9, we keep the polar angle $\rho = 45^\circ$ (tilted dipole) and change the azimuthal angle ($\psi = 0^\circ, 45^\circ, 90^\circ$). In this case we observe a strong asymmetry in the bias map, with the strongest bias appearing in a direction ϕ between 2ψ and $\psi + \pi$. Specifically, the maximum bias value exceeds 30 nm at $\psi = 0^\circ$, surpasses 20 nm at $\psi = 45^\circ$ and exceeds 40 nm at $\psi = 90^\circ$. It is also noteworthy that there is one position in the ROI where the bias is zero.

The distribution patterns of bias and standard deviation in the first and second columns for the different dipole orientations show a notable similarity to the intensity-absorption discrepancy maps in the third column. The emitter locations close to the beam centers mostly result in a small bias and a small standard deviation. Notably, even when the intensity-absorption discrepancy is substantial, the values for bias and standard deviation tend to be lower around the beam center. This is because the measurement with a beam center very close to the emitter position results in a very small photon count in all cases, providing a strong indication to the MLE algorithm that the emitter must be located nearby independent of dipole orientation. The CRLB map, as depicted in Fig. 4.4(c), serves as an effective reference for all standard deviation maps. When $\rho = 90^\circ$, the average bias is only half the average CRLB value, indicating minor concerns regarding bias in this scenario. However, for other dipole orientations, especially for the edge areas, and most particularly those distant from the three outer beam centers, the bias can be much larger than the CRLB. In such cases, MINFLUX localization with a simplified doughnut PSF model cannot be considered reliable.

4.3.6. ABSORPTION PSF PLOTS FOR FIXED-ORIENTATED DIPOLES

To enhance the comprehension of our performance maps, we incorporate a set of absorption PSF plots for a variety of fixed dipole scenarios. We establish the parameters with a fixed length of $L = 100\ nm$ and an azimuthal angle $\psi = 0^\circ$. Given the circular polarization of the input, the influence of the azimuthal angle on the absorption PSF is expected to be insignificant. The absorbed PSF, denoted as I_k, is calculated by Eq. 4.15. This is executed for fixed dipoles under assorted conditions of polar angles ρ, specifically targeting three distinct cases: the axial dipole case ($\rho = 0^\circ$, shown in the second row), the tilted case ($\rho = 45^\circ$, shown in the third row), and the in-plane dipole case ($\rho = 90^\circ$, shown in the fourth row). Furthermore, for a comprehensive analysis, we also present plots of the

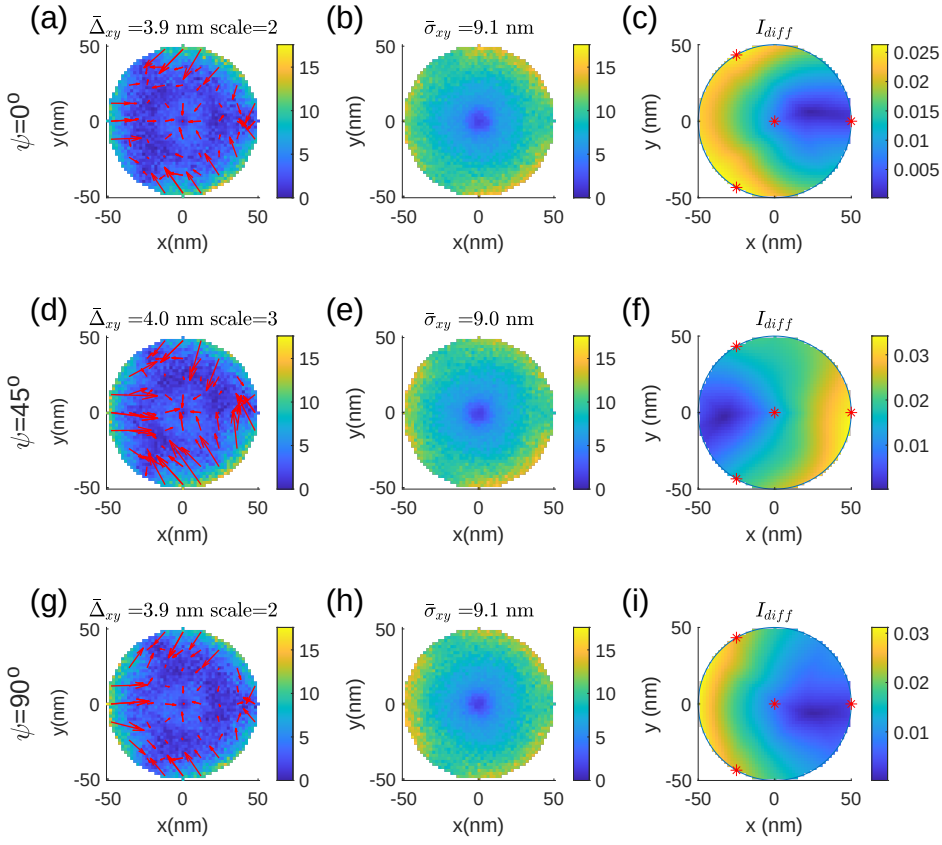


Figure 4.8: Comparative analysis of statistical accuracy and precision in localization of fixed dipoles across different azimuthal angle ($\psi = 0°, 45°, 90°$ and $\rho = 90°$), for $N_p = 50$, $L = 100$ nm and $b = 0$

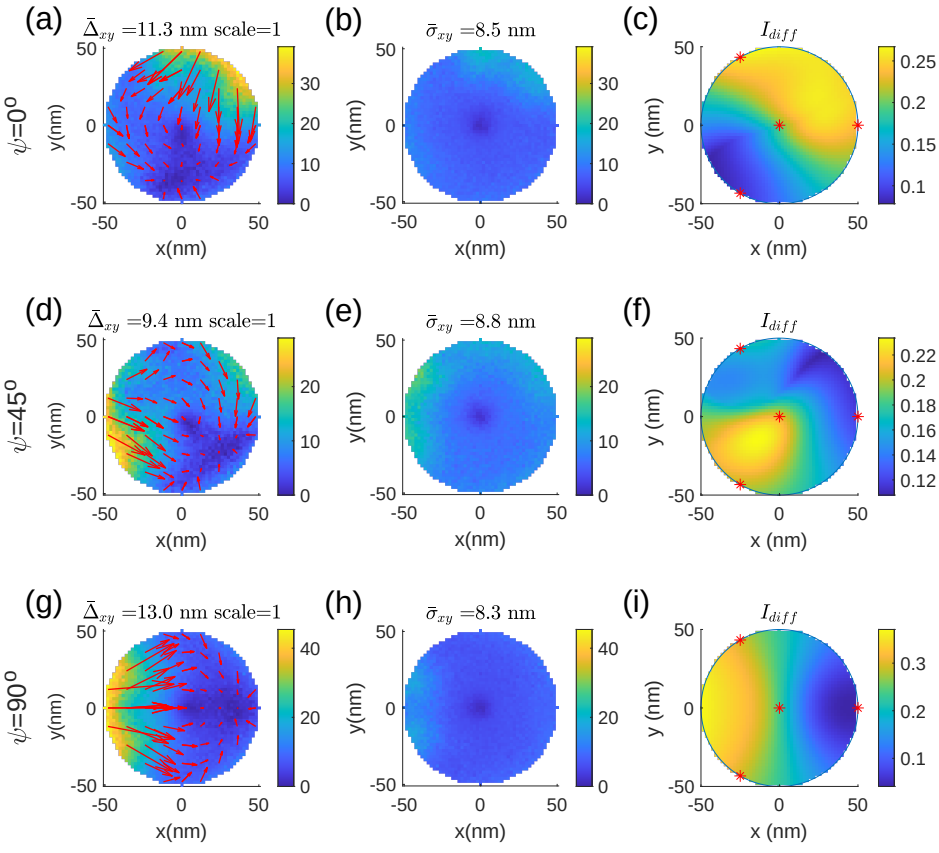


Figure 4.9: Comparative analysis of statistical accuracy and precision in localization of fixed dipoles across different azimuthal angle ($\psi = 0^\circ, 45^\circ, 90^\circ$ and $\rho = 45^\circ$), for $N_p = 50$, $L = 100$ nm and $b = 0$

absorbed PSF I_{dk} (shown in the first row), derived from the Gaussian doughnut-shaped model as described in Eq. 4.22. In each row, different columns represent the energy absorbed by emitters under various excitation beams. A red asterisk marks the center of the excitation beam in each subfigure.

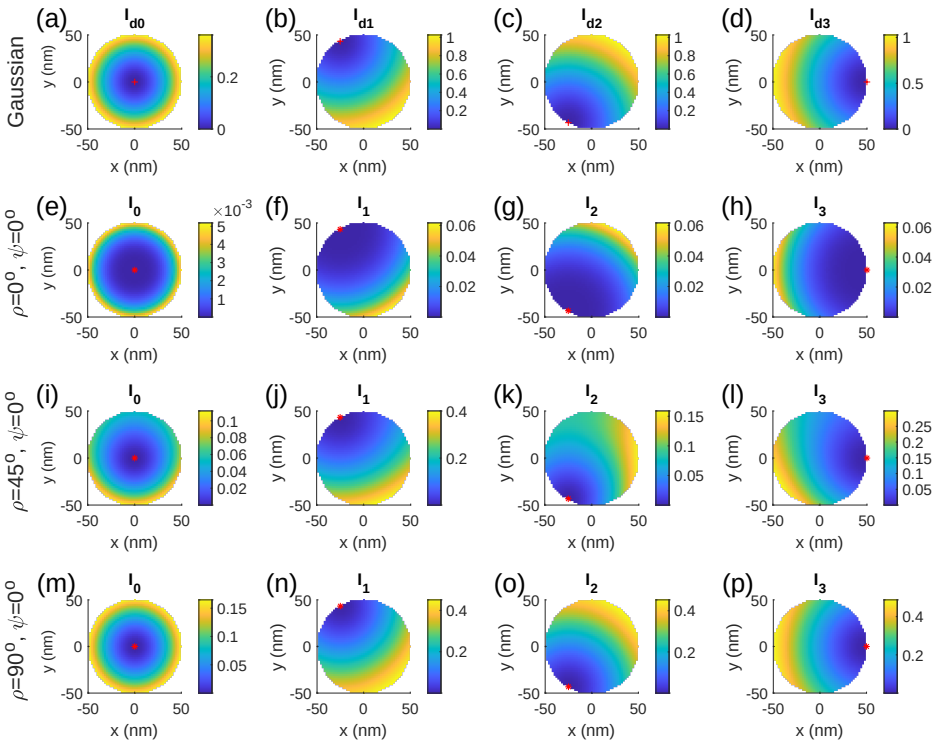


Figure 4.10: Absorption PSF Analysis for fixed dipoles at various polar angles ($\rho = 0^\circ$, 45°, 90°) and constant azimuthal Angle ($\psi = 0^\circ$) with a length of $L = 100\ nm$.

In general, within each plot, for all dipole orientations, the energy absorbed at the location of the beam center approaches zero. The absorption energy distributions generated by the last three illuminations are similar in layout, yet exhibit a change in angle. This angular shift is attributed to the angular change of the beam centers. When the polar angle is $\rho = 45^\circ$, the absorption energy plot from the first illumination is asymmetrical. However, for $\rho = 90^\circ$ or $\rho = 0^\circ$, the absorption energy plots from the first illumination are rotational symmetric. In the case of $\rho = 90^\circ$, the absolute energy absorption plot for the in-plane dipole bears the most resemblance to the reference plot in the first row. Conversely, at $\rho = 0^\circ$, the axial dipole's absolute energy absorption is significantly lower compared to other rows, yet its energy distribution pattern is similar to the others.

Under the condition where the total expected number of photons remains constant,

the number of photons emitted by an emitter at a specific point during a single illumination is determined by the relative amount of energy absorbed during that illumination compared to the total energy absorbed over four illuminations. Therefore, we have also included a normalized absorption energy plot. The basic layout and information of this plot align with the arrangement presented in the Fig. 4.10.

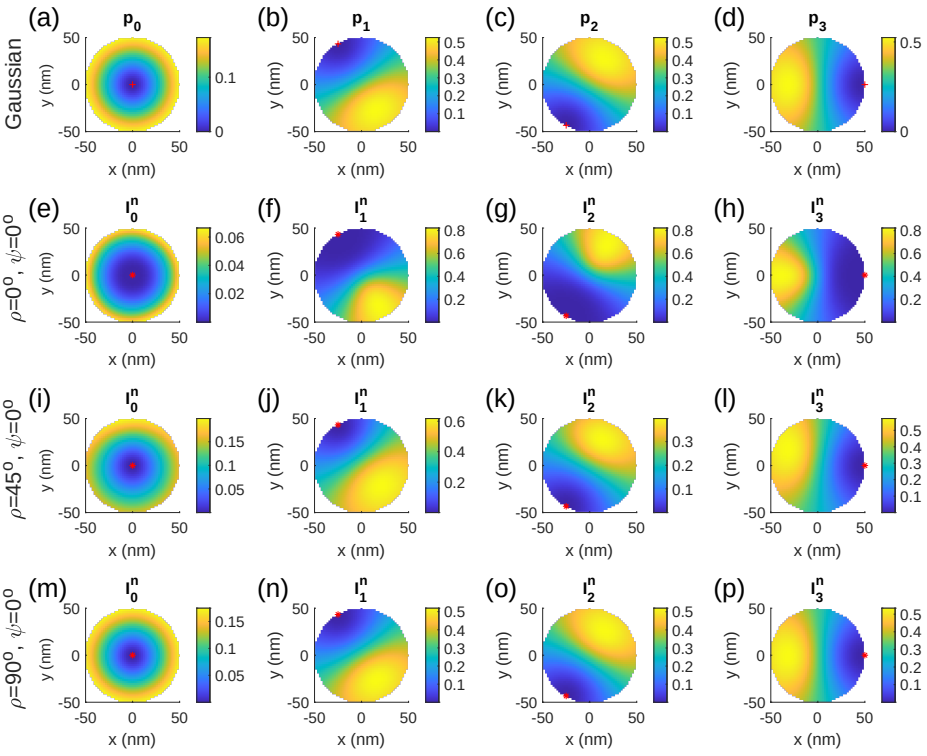


Figure 4.11: Normalized PSF Analysis for fixed dipoles at various polar angles ($\rho = 0^\circ, 45^\circ, 90^\circ$) and constant azimuthal Angle ($\psi = 0^\circ$) with a length of $L = 100\ nm$.

The normalized absorption energy plot reflects patterns consistent with the absolute value plots but focuses on relative energy absorption. Across all dipole orientations in these plots, the trend of relative energy absorption at the beam center also approaches zero. Similar to the absolute energy plots, the relative energy distributions for the last three illuminations display a notable similarity with an angular variation, mirroring the angular change in the beam centers. Particularly, at $\rho = 45^\circ$, the first illumination's relative absorption plot remains asymmetric, reinforcing the observations from the absolute value plot. In the relative value distribution plots, our focus shifts more towards the normalized distribution patterns, allowing us to overlook the differences in absolute values. This approach enables a deeper analysis of the energy absorption characteristics, inde-

pendent of the absolute magnitude of the absorbed energy.

4.4. CONCLUSION AND OUTLOOK

4.4.1. CONCLUSION

Through our simulations, in which we varied key parameters for both freely rotating and fixed-orientation dipoles, we are able to draw several conclusions.

Overall, higher photon numbers N_p, smaller distances between beam centers $L/2$ and fewer background photons per observation b lead to better performance of MINFLUX localization. This is evidenced by lower values of bias, standard deviation and CRLB maps, and confirms expected behavior.

MLE shows robust and efficient localization performance in the central region of the ROI. In the central region, bias, standard deviation and CRLB remain consistently low, regardless of the variations in parameter settings. Conversely, at the periphery of the ROI, the estimated positions of the emitters show a pronounced anisotropy. Furthermore, areas at the edge of the ROI, especially those far from any beam center, are characterised by significant biases and standard deviations. In case the emitter is in the central area of the ROI the number of detected photons for the central beam position is low in all cases, relatively robust to parameter settings and dipole orientations. This provides a strong clue to the MLE, generating robustness to the effects of differences in absorbed energy between the physical model and the Gaussian fit model.

In most cases, the standard deviation of the localization is slightly smaller than or comparable to the CRLB, indicating that the estimation method is efficient. However, the bias is usually not negligible, especially in scenarios with high background noise or low photon counts. In those case the precision can even become lower than the CRLB. The bias in the estimated dipole positions is strongly correlated with the quantitative model mismatch. For freely rotating dipoles the bias is low to moderate, and only becomes problematic in case of a relatively high background and/or low signal photon count. For dipoles with a fixed orientation, however, the bias can become larger than the CRLB, indicating that the simplified doughnut PSF model is not sufficiently accurate for MINFLUX on fixed absorption dipoles.

4.4.2. OUTLOOK

While we have achieved substantial progress in understanding the performance of the MLE and the simplified Gaussian doughnut model in relation to key parameters and across various ROI areas, several aspects could be explored and developed further.

Extension to 3D: Our current research is focused on two-dimensional doughnut illuminations. Expanding this to three-dimensional localization would greatly enhance the applicability and relevance of the findings.

Incorporation of optical aberrations: Currently, our model operates under ideal conditions and does not account for any optical aberrations (astigmatism, coma, spherical aberration). Implementing these aberrations into the model could provide a more realistic representation of experimental conditions. It could also point towards an aberration tolerance criterion, similar to Maréchal's criterion for diffraction limited imaging, specifically for MINFLUX.

Exploration of more accurate models: The bias introduced by the model mismatch suggests the need for more accurate and comprehensive models, especially to account for different dipole orientations. On the one hand, we can pursue scalar models that more closely align with the physical model. On the other hand, we have the option to directly utilize models generated from the physical model for solving MLE.

Optimization of beam center displacement: Enhancing the arrangement of beam centers could improve the precision and accuracy of emitter localization. We propose to adjust parameters such as the number of illuminations and beam center positions in our simulations, aiming to identify the most effective combination of these factors. This approach should explore the trade-off between optimal emitter positioning and time efficiency.

Iterative Localization Strategies: Building on our observation that maximum bias correlates with the distance between beam centers $L/2$, we propose exploring iterative localization, as described in refs. [10, 28]. A sequence of MINFLUX localizations with shrinking scan region size L considered in combination may solve or suppress the bias.

Effective area analysis within ROI: Our observations indicate consistently high precision and accuracy in the central region of the ROI. Investigating the relationship between the parameters and the areas exhibiting low bias and standard deviation is essential. This exploration aims to pinpoint the truly effective areas for specific parameter combinations. Understanding this relationship could enhance the strategic application of these parameters, leading to optimized localization performance in defined ROI regions.

Systematic error reduction: Notably, our studies have revealed systematic errors in bias and standard deviation maps, particularly pronounced in scenarios involving fixed-orientation dipoles. A potential avenue for improvement lies in pre-identifying dipole orientations. With this prior knowledge we may be able to effectively eliminate these systematic errors related to the fixed dipole scenarios.

Quantitative analysis of I_{diff}, beam centers, and localization accuracy: Our current understanding of how I_{diff} (the discrepancy in absorbed energy between fit model and physically realistic simulation) and beam center positions affect localization errors like standard deviation and bias is largely qualitative. There is an opportunity for a more quantitative investigation into these relationships. By developing methodologies for a detailed analysis, we can gain a clearer picture of the direct impact of I_{diff} and beam center placement on the accuracy and precision of localization techniques. This deeper quantitative exploration could help to fine-tune our models and improve the reliability and effectiveness of localization strategies.

REFERENCES

[1] F. Balzarotti, Y. Eilers, K. C. Gwosch, A. H. Gynnå, V. Westphal, F. D. Stefani, J. Elf, and S. W. Hell, *Nanometer resolution imaging and tracking of fluorescent molecules with minimal photon fluxes,* Science **355** (2017), pp. 606–612, doi: 10.1126/science.aak9913.

[2] M. J. Rust, M. Bates, and X. Zhuang, *Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM),* Nature Methods **3** (2006), pp. 793–796, doi: 10.1038/nmeth929.

[3] E. Betzig, G. H. Patterson, R. Sougrat, O. W. Lindwasser, S. Olenych, J. S. Bonifacino, M. W. Davidson, J. Lippincott-Schwartz, and H. F. Hess, *Imaging intracellular fluorescent proteins at nanometer resolution,* science **313** (2006), pp. 1642–1645, doi: 10.1126/science.1127344.

[4] S. W. Hell and J. Wichmann, *Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy,* Optics Letters **19** (1994), pp. 780–782, doi: 10.1364/ol.19.000780.

[5] N. Heckenberg, R. McDuff, C. Smith, and A. White, *Generation of optical phase singularities by computer-generated holograms,* Optics letters **17** (1992), pp. 221–223, doi: 10.1364/OL.17.000221.

[6] S. W. Hell, *Far-field optical nanoscopy,* science **316** (2007), pp. 1153–1158, doi: 10.1126/science.1137395.

[7] I. J. Myung, *Tutorial on maximum likelihood estimation,* Journal of mathematical Psychology **47** (2003), pp. 90–100, doi: 10.1016/S0022-2496(02)00028-7.

[8] R. Schmidt, T. Weihs, C. A. Wurm, I. Jansen, J. Rehman, S. J. Sahl, and S. W. Hell, *MINFLUX nanometer-scale 3D imaging and microsecond-range tracking on a common fluorescence microscope,* Nature Communications **12** (2021), p. 1478, doi: 10.1038/s41467-021-21652-z.

[9] C. P. Grabner, I. Jansen, J. Neef, T. Weihs, R. Schmidt, D. Riedel, C. A. Wurm, and T. Moser, *Resolving the molecular architecture of the photoreceptor active zone with 3D-MINFLUX,* Science Advances **8** (2022), p. eabl7560, doi: 10.1126/sciadv.abl7560.

[10] K. C. Gwosch, J. K. Pape, F. Balzarotti, P. Hoess, J. Ellenberg, J. Ries, and S. W. Hell, *MINFLUX nanoscopy delivers 3D multicolor nanometer resolution in cells,* Nature Methods **17** (2020), pp. 217–224, doi: 10.1038/s41592-019-0688-0.

[11] J. K. Pape, T. Stephan, F. Balzarotti, R. Büchner, F. Lange, D. Riedel, S. Jakobs, and S. W. Hell, *Multicolor 3D MINFLUX nanoscopy of mitochondrial MICOS proteins,* Proceedings of the National Academy of Sciences **117** (2020), pp. 20607–20614, doi: 10.1073/pnas.2009364117.

[12] T. Deguchi, M. K. Iwanski, E.-M. Schentarra, C. Heidebrecht, L. Schmidt, J. Heck, T. Weihs, S. Schnorrenberg, P. Hoess, S. Liu, *et al.*, *Direct observation of motor protein stepping in living cells using MINFLUX,* Science **379** (2023), pp. 1010–1015, doi: 10.1126/science.ade2676.

[13] Y. Eilers, H. Ta, K. C. Gwosch, F. Balzarotti, and S. W. Hell, *MINFLUX monitors rapid molecular jumps with superior spatiotemporal resolution,* Proceedings of the National Academy of Sciences **115** (2018), pp. 6117–6122, doi: 10.1073/pnas.1801672115.

[14] J. Cnossen, T. Hinsdale, R. Ø. Thorsen, M. Siemons, F. Schueder, R. Jungmann, C. S. Smith, B. Rieger, and S. Stallinga, *Localization microscopy at doubled precision with*

patterned illumination, Nature Methods **17** (2020), pp. 59–63, doi: 10.1038/s41592-019-0657-7.

[15] L. Reymond, J. Ziegler, C. Knapp, F.-C. Wang, T. Huser, V. Ruprecht, and S. Wieser, *SIMPLE: Structured illumination based point localization estimator with enhanced precision,* Optics express **27** (2019), pp. 24578–24590, doi: 10.1364/oe.27.024578.

[16] L. Gu, Y. Li, S. Zhang, Y. Xue, W. Li, D. Li, T. Xu, and W. Ji, *Molecular resolution imaging by repetitive optical selective exposure,* Nature methods **16** (2019), pp. 1114–1118, doi: 10.1038/s41592-019-0544-22.

[17] P. Jouchet, C. Cabriel, N. Bourg, M. Bardou, C. Poüs, E. Fort, and S. Lévêque-Fort, *Nanometric axial localization of single fluorescent molecules with modulated excitation,* Nature Photonics **15** (2021), pp. 297–304, doi: 10.1038/s41566-020-00749-9.

[18] E. Slenders and G. Vicidomini, *ISM-FLUX: MINFLUX with an array detector,* Physical Review Research **5** (2023), p. 023033, doi: 10.1103/PhysRevResearch.5.023033.

[19] L. M. Ostersehlt, D. C. Jans, A. Wittek, J. Keller-Findeisen, K. Inamdar, S. J. Sahl, S. W. Hell, and S. Jakobs, *DNA-PAINT MINFLUX nanoscopy,* Nature Methods **19** (2022), pp. 1072–1075, doi: 10.1038/s41592-022-01577-1.

[20] L. A. Masullo, F. Steiner, J. Zähringer, L. F. Lopez, J. Bohlen, L. Richter, F. Cole, P. Tinnefeld, and F. D. Stefani, *Pulsed interleaved MINFLUX,* Nano letters **21** (2020), pp. 840–846, doi: 10.1021/acs.nanolett.0c04600.

[21] L. A. Masullo, A. M. Szalai, L. F. Lopez, M. Pilo-Pais, G. P. Acuna, and F. D. Stefani, *An alternative to MINFLUX that enables nanometer resolution in a confocal microscope,* Light: Science & Applications **11** (2022), p. 199, doi: 10.1038/s41377-022-00896-4.

[22] S. Stallinga and B. Rieger, *Accuracy of the Gaussian point spread function model in 2D localization microscopy,* Opt. Express **18** (2010), pp. 24461–24476, doi: 10.1364/OE.18.024461.

[23] E. Wolf, *Electromagnetic diffraction in optical systems-I. an integral representation of the image field,* Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences **253** (1959), pp. 349–357, doi: 10.1098/rspa.1959.0199.

[24] B. Richards and E. Wolf, *Electromagnetic diffraction in optical systems, II. structure of the image field in an aplanatic system,* Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences **253** (1959), pp. 358–379, doi: 10.1098/rspa.1959.0200.

[25] T. Wilson, R. Juškaitis, and P. Higdon, *The imaging of dielectric point scatterers in conventional and confocal polarisation microscopes,* Optics Communications **141** (1997), pp. 298–313, doi: 10.1016/S0030-4018(97)00226-5.

[26] P. Török, P. D. Higdon, and T. Wilson, *Theory for confocal and conventional microscopes imaging small dielectric scatterers,* Journal of Modern Optics **45** (1998), pp. 1681–1698, doi: 10.1080/09500349808230662.

[27] P. Dedecker, B. Muls, J. Hofkens, J. Enderlein, and J.-I. Hotta, *Orientational effects in the excitation and de-excitation of single molecules interacting with donut-mode laser beams,* Optics Express **15** (2007), pp. 3372–3383, doi: 10.1364/OE.15.003372.

[28] D. Kalisvaart, J. Cnossen, S.-T. Hung, S. Stallinga, M. Verhaegen, and C. S. Smith, *Precision in iterative modulation enhanced single-molecule localization microscopy,* Biophysical Journal **121** (2022), pp. 2279–2289, doi: 10.1016/j.bpj.2022.05.027.

5

CONCLUSION AND OUTLOOK

5.1. CONCLUSION

5.1.1. FAST PARTICLE FUSION METHOD

In **chapter 2**, we introduced a fast particle fusion algorithm for particles acquired by single molecule localization microscopy (SMLM). There exists an established template-free particle fusion approach, known as the all-to-all registration method [1, 2], which has the ability to transform particles into a super-particle and achieve an image resolution down to a few nanometers. However, a significant limitation of the all-to-all method is its computational complexity, which scales quadratically with $O(N^2)$, where N is the number of input particles. The all-to-all method, even with GPU acceleration, requires approximately 2 hours to process 383 particles, highlighting its limitation in efficiently registering larger numbers of particles. Our focus extends beyond mere particle count, emphasizing the significance of capturing the complete biological landscape and natural variations present in samples. Given the increasing number of detectable particles and the potential insights that can be gained from analyzing a larger dataset, developing a faster particle fusion method becomes imperative.

Our proposed fast particle fusion method is orchestrated by a sequential process of registration, classification, and connection. We apply the Joint Registration of Multiple Point Clouds (JRMPC) [3] to particles and obtain mostly locally optimal solutions that contain several groups of overlapping particles. These groups contain particles that are well aligned but different groups have different poses. To address the different poses, classification is performed on these locally optimal reconstructions, to extract clusters of identically posed particles. A connection strategy is used to transform the particles in the different clusters and obtain a single well-aligned structure containing the majority of the particles. The computational complexity of the fast particle fusion method scales linearly with the number of input particles, making it possible to handle thousands of particles in a reasonable time.

The performance of the fast particle fusion method is assessed using multiple datasets obtained from different localization microscopy experiments. To quantify the resolution of our super-particle reconstructions, we employed Fourier Ring Correlation (FRC) [4]. We demonstrated that our particle fusion method is able to reconstruct low-quality particles, process large amounts of particles quickly, and avoid hot-spot artefacts for symmetric structures. We achieved a resolution of 3.3 nm on experimental 2D TUD-logo DNA origami particles for 30% Degree Of Labeling (*DOL*). We resolved the DNA origami structure 'Digit 1', both on a GPU and a CPU platform, utilizing 4,155 particles containing 692,139 localizations. The computation time on the GPU was 1.1 hours, while the computation on the CPU took 0.6 hours.

In summary, our fast, template-free particle fusion method paves the way for efficient fusion of thousands of particles, thereby enhancing the resolution capabilities of SMLM.

5.1.2. DATA ANALYSIS ON NUCLEAR PORE COMPLEX

In **chapter 3**, we analyzed the particle fusion results of 5 cell nuclei and the combined dataset comprising thousands of labelled Nup96 in Nuclear Pore Complexes (NPCs). We pinpoint the positions of the fluorescent labels accurately. The structure of the NPC scaffold has already been resolved at high resolution by extensive studies using cryogenic

electron microscopy (cryo-EM) and cryo-electron tomography (cryo-ET). Conventional SMLM, serving as a complementary technique to study biological structures, however, struggles to resolve the inner structure of the NPC due to factors such as limited localization precision and residual drift. By applying the fast template-free particle fusion methods to thousands of NPCs, we improved the resolution of particles obtained from SMLM, enabling investigation of the Nup96 structure with nanometer scale precision and accuracy.

We applied the particle fusion method described in **chapter 2** to these datasets to obtain super-particles with superior resolution. We conducted particle fusion on each dataset ten times, resulting in 60 super-particles. This approach allowed us to incorporate the statistical variability introduced by the randomly generated Gaussian center of the Gaussian Mixtures Model (GMM) into our analysis. All super-particles show two phase shifted rings with eight inclined elliptical blobs per ring. We then applied the anisotropic GMM with 16 Gaussian centers to the 8 blobs of each ring, integrating the 8-fold symmetry into the final GMM centers. Thus, the positions of the symmetric GMM centers directly represent the positions of the Nup96's fluorescent labels. We also compared these positions of our overall model to the SNAP-tags' positions, which were predicted by Alphafold [5] from an electron microscopy model [6] and found an average discrepancy of less than 3 nm laterally and 5 nm axially. The larger value in the axial direction is due to a small bias in the NPC height as imaged by cryo-EM and by SMLM.

Our research underscores the amplified potential of SMLM in offering profound insights into sub-cellular structures.

5.1.3. IMPACT OF DIPOLE ORIENTATION ON MINFLUX

In **chapter 4**, we investigated the influence of dipole orientation on the localization uncertainty of MINFLUX, which is reported to be in the range 1-3 nm, without considering effects of absorption dipole orientation [7]. For conventional SMLM it is known that a fixed emission dipole orientation can give rise to a bias in the position of several 10s of nanometers [8]. For MINFLUX we need to consider the absorption dipole, not the emission dipole. On the emission side, as long as the emitted light is captured by the objective, all light is then collected by point detection anyway. This merits an investigation if fixed dipole orientation effects can give rise to biases exceeding the typical localization precision of MINFLUX.

To this end we developed a physically realistic forward model for the doughnut excitation Point Spread Function (PSF) to analyze the impact of dipole orientation on MINFLUX localization by simulation. We simulated the vectorial electric field of the excitation beam at the focal plane and absorption dipoles with arbitrary orientations. Background photons are considered in our model as well. By employing the physically realistic forward model, we generated expected photon counts for the different center positions of the doughnut spots in a MINFLUX localization data acquisition, and simulated noisy measurements by adding Poissonian noise. A Maximum Likelihood Estimator (MLE) was then utilized to infer the most likely emitter position, given both the simulated noisy photon count measurements and the effective excitation PSF model used in MINFLUX data analysis [7].

Through our simulations on varying parameters for both rotating and fixed dipoles, we found that MINFLUX localization works better for higher expected photon numbers,

doughnut beam centers that are closer together and for fewer background photons, as evidenced by reduced values of bias, standard deviation and Cramér Rao Lower bound (CRLB). In the central region of the ROI that is scanned by the doughnut beam, the MLE maintains low values for these metrics, in contrast to the peripheral areas, which show significant bias and standard deviation. Notably, even in the peripheral areas, the values for bias and standard deviation are typically lower for emitters that are close to one of the three peripheral doughnut beam centers. This is because a doughnut beam center near an emitter will consistently produce low photon counts, providing strong information to the MLE algorithm that the emitter is likely to be close to the doughnut beam center. The standard deviation of the localization is typically at or below the CRLB, indicating that the localization is generally efficient. The bias, however, remains a significant concern, particularly in scenarios with high background, low signal photon count, and for fixed dipole orientations, in particular for tilted dipoles. This bias originates from the inaccuracy of the simplified absorption PSF model, especially for fixed dipoles.

This research advances our understanding of the MINFLUX technique by highlighting the critical impact of dipole orientation on localization accuracy and precision. Exploration of various key parameters is helpful in establishing criteria for sufficiently good conditions, such as for a minimum photon number, maximum background or maximum doughnut spot scan range. In addition, by revealing the limitations of the simplified doughnut PSF model in accurately predicting fixed dipole positions, the study highlights the need for model improvements.

5.2. OUTLOOK

5.2.1. ENHANCEMENTS IN FAST PARTICLE FUSION

INCORPORATION OF LOCALIZATION UNCERTAINTIES IN JRMPC

In our proposed fast particle fusion method, the joint registration step uses only the coordinates of the localizations, treating each localization as an exact point. However, each particle's localization comes with uncertainty, implying we cannot pinpoint the exact position. It would be beneficial to treat localizations as distributions with uncertainties.

To address this, we can include the uncertainties of each localization in the probability calculation of the joint registration. When calculating the probability of localization i of particle j to fit the Gaussian component k, we can describe the probability as a normal distribution, where the variance is a sum of both the localization uncertainty and the variance of the Gaussian component.

Incorporating this revised probability calculation into each iteration of the JRMPC can potentially enhance convergence towards a global optimum and improve the sensitivity to the initial setting of the widths of the Gaussian components. Especially when the localization uncertainty is comparable to the distance between binding sites, has a large range of variations, or exhibits anisotropic properties in 3D datasets.

INCORPORATION OF PRIOR KNOWLEDGE IN THE PARTICLE FUSION

One of the advantages of our particle fusion method is its independence from the need for prior knowledge. However, in some scenarios, the implementation of prior knowledge could be beneficial for the particle fusion. For instance, configuring the number and width

of Gaussian distributions within the GMM can facilitate the convergence of the JRMPC. Additionally, imposing symmetry for the fusion of symmetric particles could be helpful to solving the 'hot-spot' problem more robustly.

QUALITY ASSESSMENT OF CLUSTERS

In our fast particle fusion pipeline, there is a selection following the classification to advance high quality clusters to the connection stage. At present, the simple criterion for 'cluster quality' is based solely on the number of particles in the cluster, which is useful but not optimal. This is particularly evident when the input particles are of very low quality, resulting in many misaligned particles, and well-aligned ones often end up in sparsely populated clusters. Using the FRC resolution to assess the quality of the clusters could be a better choice for selection.

AUTOMATED PARAMETER SETTING

The fast particle fusion method exhibits robust performance on both simulated datasets and experimental datasets. However, it is sensitive to its initial parameter settings, and with dozens of parameters to configure, it can be complex. While we have implemented algorithms to preset certain parameters and these default values are sufficient for most of our tested datasets, there are still some parameters that need to be adjusted for a specific input. To achieve optimal fusion results often needs an experienced hand to adjust these parameters. Given the fast computational capabilities of the method, these parameters can be tuned through a feedback loop, seeking for the optimal parameters based on the FRC resolution or visualization for the images at each step.

ESTIMATION THE OPTIMAL NUMBER OF PARTICLES

Currently, we input all available particles into the particle fusion algorithm to obtain results. However, it turns out that the quality of the fusion results does not improve indefinitely with an increasing number of input particles. Estimating an optimal or sufficient number of particles required for a certain resolution could improve the efficiency of the particle fusion process. However, the estimation is a complicated task which depends on several factors such as the structure of particles, labelling density, localization uncertainty, repeated localizations per binding site, and typical or minimum distance between binding sites. Developing a comprehensive framework for estimating the optimal number of particles needed to achieve a particular resolution, while considering all the pertinent parameters, may provide us with deeper insights into particle fusion.

MULTI-CHANNEL PARTICLE FUSION ALGORITHMS

Multi-channel SMLM, in comparison to single-channel SMLM, provides a wealth of information which is beneficial for biological application. However, the data processing for multi-channel SMLM data is more complex. Currently, our particle fusion method is specifically designed for registering single-channel particles data. Expanding this algorithm to accommodate multi-channel data holds practical significance for studying cellular interactions and complex biological processes. A possible extension involves initially fusing particles from individual channels and subsequently using the reconstruction results from each channel as prior knowledge to enhance the joint reconstruction of all channels.

5.2.2. IMPROVEMENTS IN SMLM DATA ANALYSIS OF NPC STRUCTURE

OUTLIER LOCALIZATION REMOVAL

The process of outlier localization removal is now done in two steps. Firstly, an extra Gaussian distribution is incorporated into the GMM to capture potential outliers. Secondly, outlier localizations are further filtered out based on the density distribution of the remaining localizations. Considering the intricate nature of data we deal with, this cascading mechanism allows us to effectively separate genuine localizations from potential outliers. However, using a single integrated outlier removal approach may improve the robustness of the data analysis procedure.

MODEL SELECTION OF INITIAL GMM CENTERS

The initial parameter settings play an important role in the convergence to a global optimum of the anisotropic GMM fitting. A refinement in our analysis pertains to the model selection. The current approach relies on the most simple model that can explain our observations. We make an assumption about the number of emitters per ring. This assumption comes from the ground truth obtained from electron microscopy results that Nup96 appears in a unit structure comprising two copies. It might be beneficial to explore other methodologies to further corroborate these findings. One avenue of potential improvement is to explore a statistical criterion. This criterion would provide a strong argument for the presence of 32 Nup96s in the NPC, rather than other multiples of 16. This statistical approach could provide a more comprehensive understanding and validation of our dataset. Additionally, this criterion has the potential to analyze the positions of emitters within various sub-cellular structures different from the NPC.

AUTOMATED POST-PROCESSING AFTER PARTICLE FUSION

The data analysis of Nup96 particle fusion highlights the capability of SMLM in resolving biological structures. In **Chapter 3**, we have designed a dedicated workflow for estimating the position of individual Nup96 copies. The resolution assessment, outlier removal, estimation of the number of emitters, and emitter position determination within the post-fusion phase can be integrated into an automated procedure. Statistical criteria in each step could be used to to objectively quantify the quality of this automated process.

5.2.3. IMPACT OF ABSORPTION DIPOLE ORIENTATION ON MINFLUX

MODEL IMPROVEMENTS

In our outlook for model improvements in MINFLUX localization, we focus on three key areas: extending our current two-dimensional localization to three-dimensional localization, which would significantly broaden the applicability; incorporating optical aberrations into our model to more closely reflect real experimental conditions and improve prediction accuracy; and exploring more accurate and comprehensive models to address biases caused by model mismatches. This could be achieved by developing scalar models that more closely match the physical realities or by directly using models derived from the physically realistic vector diffraction based PSF model in MLE, in particular to account for different dipole orientations.

EXPERIMENTAL PARAMETER OPTIMIZATION

Our investigation has provided insight into how critical parameters affect the accuracy and precision of localization. In the future, this parameter space exploration can be refined, for example by the placement of the doughnut beam centers, the number of doughnut spot illuminations. The goal would be to optimize these elements to achieve an optimal balance between achieving more accurate and precise localizations on the one hand, and maintaining operational efficiency on the other hand.

POST-PROCESSING ADVANCEMENTS

Our exploration of the bias and standard deviation across various parameters and ROI areas provides valuable prior knowledge that can be leveraged to enhance the post-processing of localizations. By conducting a thorough study of the ROI, particularly focusing on areas exhibiting low bias and standard deviation, we can effectively identify the most efficient regions for localization. Additionally, if key parameters are known in advance, we can apply a targeted approach to minimize systematic errors, as guided by our findings. This strategy can be especially relevant in scenarios involving fixed-orientation dipoles, where our results can be directly utilized to refine accuracy and precision in localization post-processing.

REFERENCES

[1] H. Heydarian, M. Joosten, A. Przybylski, F. Schueder, R. Jungmann, B. v. Werkhoven, J. Keller-Findeisen, J. Ries, S. Stallinga, M. Bates, and B. Rieger, *3D particle averaging and detection of macromolecular symmetry in localization microscopy,* Nature Communications **12** (2021), pp. 1–9, doi: 10.1038/s41467-021-22006-5.

[2] H. Heydarian, F. Schueder, M. T. Strauss, B. Van Werkhoven, M. Fazel, K. A. Lidke, R. Jungmann, S. Stallinga, and B. Rieger, *Template-free 2D particle fusion in localization microscopy,* Nature Methods **15** (2018), pp. 781–784, doi: 10.1038/s41592-018-0136-6.

[3] G. D. Evangelidis and R. Horaud, *Joint alignment of multiple point sets with batch and incremental expectation-maximization,* IEEE Trans. Pattern Anal. Mach. Intell. **40** (2017), pp. 1397–1410, doi: 10.1109/TPAMI.2017.2717829 .

[4] R. P. J. Nieuwenhuizen, K. A. Lidke, M. Bates, D. Leyton Puig, D. Grünwald, S. Stallinga, and B. Rieger, *Measuring image resolution in optical nanoscopy,* Nature Methods **10** (2013), pp. 557–562, doi: 10.1038/nmeth.2448.

[5] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. Kohl, A. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. Senior, K. Kavukcuoglu, P. Kohli, and D. Hassabis, *Highly accurate protein structure prediction with AlphaFold,* Nature **596** (2021), pp. 583–589, doi: 10.1038/s41586-021-03819-2.

[6] A. v. Appen, J. Kosinski, L. Sparks, A. Ori, A. DiGuilio, B. Vollmer, M.-T. Macmull, N. Banterle, . Parca, P. Kastritis, K. Buczak, S. Mosalagnati, W. Hagen, A. Andres-Pons, E. Lemke, P. Bork, W. Antonin, J. Glavy, K. Bui, and M. Beck, *In situ structural analysis of the human nuclear pore complex,* Nature **526** (2015), pp. 140–143, doi: 10.1038/nature15381.

[7] F. Balzarotti, Y. Eilers, K. C. Gwosch, A. H. Gynnå, V. Westphal, F. D. Stefani, J. Elf, and S. W. Hell, *Nanometer resolution imaging and tracking of fluorescent molecules with minimal photon fluxes,* Science **355** (2017), pp. 606–612, doi: 10.1126/science.aak9913.

[8] S. Stallinga and B. Rieger, *Accuracy of the Gaussian point spread function model in 2D localization microscopy,* Opt. Express **18** (2010), pp. 24461–24476, doi: 10.1364/OE.18.024461.

ACKNOWLEDGEMENT

First and foremost, I would like to express my deepest gratitude to my two supervisors, **Bernd Rieger** and **Sjoerd Stallinga**. I have always considered myself incredibly fortunate to have had the privilege of working under your guidance. Your understanding, assistance, and unwavering support throughout my doctoral journey have been invaluable. I am particularly grateful for the insightful discussions we have had on topics ranging from academic pursuits to broader aspects of life, all of which have greatly enriched my learning and personal growth. Your mentorship not only provided me with overarching direction in my research but also instilled in me the importance of rigor in formulating hypotheses and meticulousness in code validation. Additionally, I want to express my gratitude for your repeated efforts in helping me refine my writing and academic work.

Furthermore, I wish to extend my heartfelt thanks to my family, especially to my parents **Kecai Wang** and **Limei Tang**. I am deeply grateful for your selfless love and unwavering support over the past two decades. Your encouragement has enabled me to lead a joyful life and pursue personal growth, granting me the freedom and resources to explore the world. Whenever I needed reassurance, you successfully reinstilled confidence in me and provided me with a profound sense of security.

Moreover, I would like to express my gratitude to the friends I have met in the Department of Imaging Physics. I thoroughly enjoyed our casual conversations and discussions during breaks, as well as the exchange and interaction during various activities.

I'm thankful to **Ronald** and **Henry** for your assistance with software and hardware. **Ronald**, your laughter brightens our corridor; **Henry**, I firmly believe in your capacity to traverse the globe on your bicycle. My thanks also go to **Annelies** and **Angela** for your administrative support, making my journey here smoother.

Special thanks to my office mates, **Hamidreza Heydarian** and **Isabel Droste**. **Hamid**, your guidance during my first two years and your advice on PhD studies and job hunting have been incredibly helpful. **Isabel**, your kindness and approach to work and life are deeply appreciated. Our conversations have broadened my perspectives significantly. In my eyes, you are even more delightful than an alpaca.

I am grateful to **Sobhan Haghparast** for our weekly meetings that have been a source of mutual encouragement and support. Thank you, **Peter**, for the endless lunchtime discussions and your enjoyable humor. Thank you, **Kote**, for the cake and ping-pong invitations! Thank you, **Alim** and **Awoke**, for the enlightening conversations and guidance on job hunting. Thank you, **Teun**, for your support in the classification section of my first paper. And to **Rasmus**, your assistance during the initial phase of my PhD and with the Vortex PSF code post-graduation has been invaluable.

Thanks to my PhD and postdoctoral fellows for our lunch breaks and C++ time: **Ali**, **Leo**, **Agis**, **Chih-Hsien**, **Kathleen**, **Christiaan Hulleman**, **Enya Berrevoets**, **Taylor Hinsdale**, **Valerii Brudanin**, **Menglin Wu** and **Spozmai Panezai**.

My thanks extend to the Principal Investigators at the Imaging Physics Department: **Qian**, **Frans**, **Eric**, **Jeroen**, **Miriam**, and **Sebastian**. Even though we do not have direct collaboration, it's my pleasure to work in the same department with you.

Thanks to my buddies in the "Netherlands Roll", I always enjoy our gatherings during "touching fish" time. **Changchun Yang**, you are a person like your name, making people feel as warm as spring. **Qingru Li** and **Chengming Feng**, It's nice to listen to your stories about your cat and fish. Thanks to my dream team members: **Yidong Zhao**, **Xi Chen**, **Xinrui Zu**, and **Huangcheng Shangguan**. For our awesome trips in Austria and France. Look forward to more trips together with more buddies. **Yiming Liu** and **Feiyu Liu**, it's nice to know you.

Yi Zhang, your understanding, help, and support since 2022 have been pivotal to my journey. Whether it was getting my bicycle ready, guiding me through decisions, or offering technical assistance for my thesis cover, you've shown unwavering kindness. For all of these, you have my sincerest and deepest gratitude.

I am thankful to **Jos** and **Yangyang** for the dinners, ceremonies, barbecues, board games, and meaningful conversations in Delft. I look forward to visiting you in Amsterdam, hopefully soon.

Thank you, **Xiufeng Li** and **Xiaocui Wang**, for the delightful Hubei cuisine and the job search updates we've shared.

Thank you, **Xinmin You** and **Qian Feng**, for being like family to me. Meeting fellow Nanjing natives in the Netherlands has been incredibly fortuitous. Your support during both joyful and challenging times has been invaluable.

To my dear neighbors and friends, **Shan He** and **Zejia Zheng**, sharing meals and conversations in your cozy living room has been a joy. Your generosity, from sharing plants to emergency drives for backup keys, has been truly heartwarming.

Yan Guo, thank you for your guidance and advice as both a senior colleague and a friend. Your exemplary work ethic has always served as an inspiration to me.

Shan Qu, I feel incredibly fortunate to have worked alongside you during the first two years of my doctoral journey. Our lunchtime strolls during the initial phase of the pandemic, amidst the spring breeze and cherry blossoms of that year, remain etched in my memory as moments of tranquility and delight.

Haiwei Xie, you are like a heroine in real life, always ready to stand up for what's right and speak out for justice. From Zumba classes with X to concerts in Amsterdam, strolls in Leiden, and gatherings in Utrecht, you've always supported and understood me. **Ting Hu** and **Chenguang Wang**, It's my pleasure to know you and share joyful times with you. Mahjong and hot pot are my warm memories in the winter of 2023.

Acknowledgment is also for my friends **Ulas**, **Yingfu Xu**, **Yifeng Shao**, **Jingyi Liu**, **Kai Wu**, **Zhenzhen Wu**, **Jin Chang**, **Yun Wan**, **Yueer Li**, **Yifan Fu**, **Dongbin Cai**, and **Hong Zhang**, for the shared meals, games, and sporting events. To **Tiexing Wang**, **Junhai Cao**, **Boling Ouyang**, and **Kefei Hei**, I fondly reminisce about the laughter-filled joke sessions and ping-pong matches we shared during my first month here.

I want to express my deepest gratitude to the friends I made in Jena: **Yiming Tu**, **Yichen Liu**, **Xiaoyan Yu**, **Site Zhang**, **Sici Wang**, **Jiaming Liu**, **Jinhao Li**, **Xiao Chen**, and **Hao Meng**. To those who have made the journey to visit me in Delft, your presence has been a source of immense joy. And to all of you, the support we've kept alive throughout online connec-

tions is just as cherished.

Yu Zhang, it's been a pleasure spending two years in Jena and two years in Delft with you. I cherish our deep conversations that touch the soul and confront our true selves. I hope we have the opportunity to meet more often in China.

I owe a debt of gratitude to my mentor during my master's, **Huiying Zhong**. Sister Zhong, your readiness to assist, whether for career advice or life decisions, has been a beacon of support.

To my peers in both master and PhD, **Shih-Te** and **Yutong**, it's great to share 2 periods of study with you. Thank you, **Shih-Te**, for the discussions on optics and the entrepreneurial dreams. Your intelligence and diligence have always inspired me. Thank you, **Yutong**, for your astounding breadth of knowledge in physics and optics. I appreciate your knowledge both on programming and building microscopes.

A special thanks to **Ran Huo** for the unique connection we've formed online, leading to memorable times in cities like Malaga, Lisbon, Porto, Delft, Rotterdam, and Nanjing. Your advice on plants has also added a lovely touch to my balcony.

I'm grateful for the enduring friendship with my bachelor study mates: **Danyang Zhang**, **Jiahui Li**, **Xinda Yi**, and **Yue Yang**. Reuniting in Shanghai or Delft after all these years has been a joy.

Thanks to my undergraduate roommates **Zhouping Lyu**, **Mingjun Xiang**, **Yan Xu**, and **Yifan Wu**. **Pingping**, I do not feel lonely because we share our similar study routines in China, Germany, and the Netherlands. You are always there when I want to see you. **Xiangxiang**, our travels in Germany, the Netherlands, and the Czech Republic are wonderful memories of my time studying abroad. Looking forward to seeing you often in the future. **Yan**, even though we have not seen each other for about 7 years, you are always one of my best friends in the world, and we will see each other someday. **Yifan**, thanks a lot for your emotional support for me during my bachelor, master, and PhD graduations.

My heartfelt appreciation goes out to my dearest friends **Chu Li**, **Jie Yu**, **Qinqin Bao**, **Fan Zhou**, **Huimin Hu**, **Danjing Wu**, and **Weitong Xue**. Watching you evolve from youthful girls into formidable, empowered women is a source of immense pride and joy for me. Your company is where I find solace and comfort. Reuniting with you has become a cherished ritual with each return to Nanjing, a tradition I eagerly anticipate continuing for years to come. I look forward to our future culinary adventures and the new experiences that await us. **Chu Li**, in my heart, you've always been someone who succeeds in whatever you set your mind to, and I believe you always will. I treasure every holiday we've spent together, every meal we've shared, and every trip we've taken. **Jie Yu**, I am in awe of your strong planning and organizational skills.

Thanks to my friends **Zhenyang Shao**, **Shijie Tu**, **Haoyang Zhou**, and **Jing Yu** from my senior high school. Our occasional contact and encouragement helped me get through some difficult times.

I am also grateful to my sisters **Haiyan Tang** and **Lina Xue**, who I grew up with. I am honored to witness all stages of your life, and I will continue to share my growth with you. Thanks to my hometown sister **Simin Sun** and brother-in-law **Peng Wang**. It's so lucky that we meet again in the Netherlands. It's a pleasure to spend weekends with you.

Thank my current Aramco colleagues: **Paul**, **Jewoo**, **Andreas**, **Rob**, **Rolf**, **Abu Yousef**, **Xinru**, **Sven**, **Han Yao**, **Marcin**, **Diego**, **Nauman**, **Maria**, **Luigi**, **Miaoyang**. Special thanks to **Yimin**, for your careful instruction on coding and many things. I look forward to being someone who can research and develop as excellently as you.

CURRICULUM VITÆ

Wenxiu Wang

07-Mar-1995	Born in Jiangning, Nanjing, China.

EDUCATION

2019–2023	Ph.D candidate in Applied Physics Delft University of Technology, Delft, The Netherlands
2017–2019	Master in Photonics Friedrich-Schiller-University Jena, Jena, Germany
2013–2017	Bachelor in Opto-Electronic Information Science and Engineering Harbin Institute of Technology, Harbin, China
2010–2013	Middle High School Student Nanjing Jiangning Senior High School, Nanjing, China
2007–2010	Middle School Student Nanjing Jiangning Tuqiao Middle School, Nanjing, China
2001–2007	Elementary School Student Nanjing Jiangning Tuqiao Central Elementary School, Nanjing, China

RESEARCH EXPERIENCE

2024–present	Lab Scientist Aramco, Delft Global Research Center, Delft, The Netherlands Research on edge computing
2019–2023	Ph.D candidate Delft University of Technology, Delft, The Netherlands Research on super-resolution microscopy
2019–2019	Master thesis Robert Bosch GmbH, Renningen, Stuttgart, Germany Research on wave optical analysis in terms of eye safety
2018–2019	Research labwork Institute of Applied Physics, Friedrich-Schiller-University Jena, Germany Research on modeling of rotational symmetric step-index fiber
2018–2018	Internship Institute of Applied Physics, Friedrich-Schiller-University Jena, Germany Research on mode solver of rotational symmetric step-index fiber
2016–2017	Bachelor Thesis Harbin Institute of Technology, China Research on omnidirectional infrared multitargets Algorithm

List of Publications

2. **Wenxiu Wang**, Arjen Jakobi, Yu-Le Wu, Jonas Ries, Sjoerd Stallinga, and Bernd Rieger, *Particle fusion of super-resolution data reveals the unit structure of Nup96 in Nuclear Pore Complex*, Scientific Reports 13, no. 1 (2023): 13327, doi: 10.1038/s41598-023-39829-5

1. **Wenxiu Wang**, Hamidreza Heydarian, Teun A.P.M. Huijben, Sjoerd Stallinga, and Bernd Rieger, *Joint registration of multiple point clouds for fast particle fusion in localization microscopy*, Bioinformatics 38, no. 12 (2022): 3281-3287, doi: 10.1093/bioinformatics/btac320