Optimum Single Molecule Localization Microscopy
From Single Molecule Detections to Nanoscopic Observations

Smith, Carlas

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

CARLAS S. SMITH

OPTIMUM SINGLE MOLECULE LOCALIZATION MICROSCOPY

FROM SINGLE MOLECULE DETECTIONS TO NANOSCOPIC OBSERVATIONS
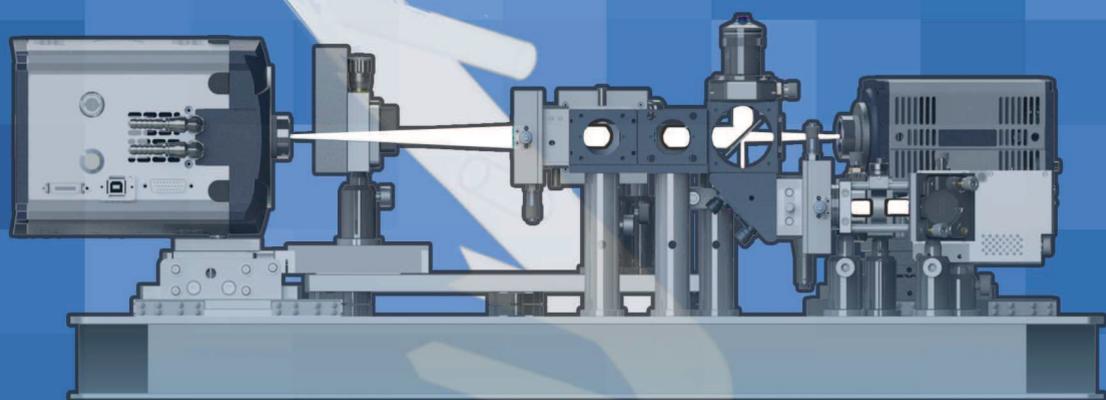
Optimum Single Molecule Localization Microscopy

From Single Molecule Detections to Nanoscopic Observations

Optimum Single Molecule Localization Microscopy

From Single Molecule Detections to Nanoscopic Observations

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof. ir. K.C.A.M. Luyben,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op
maandag 12 september 2016 om 15.00 uur

door

Carlas Sierd Smith

natuurkundig ingenieur & lucht en ruimtevaart ingenieur,
Technische Universiteit Delft,
geboren te 's-Gravenhage.

Contents

1

Introduction

*In the year of 1657 I discovered
very small living creatures in rain water.*

Antonie van Leeuwenhoek

1.1. *The History of Single Molecule Localization Microscopy*

The first single molecule localization microscope became available more than 300 years after the development of the first microscope. In this section, a brief overview is given of the pivotal discoveries that ultimately lead to the single molecule localization microscopy that we know today. Discussed are three key areas: the development of the first microscopes, fluorescence and fluorescence microscopy, and finally single molecule localization microscopy.

1.1.1. The first microscope

The first microscope design can be traced back to 1595 when a Dutch spectacle maker, Zacharias Janssen, developed the first (compound) microscope that could be used at a magnification of up to 30 times. The magnification of his microscope was restricted due to lens manufacturing limitations. It took another fifty years before Antonie van Leeuwenhoek, a Dutch tradesman from Delft, handcrafted a single lens microscope that could be used at a magnification of up to 300 times. His excellent skill in lens grinding and the simple design made the development of this device possible. It allowed him to be the first to discover numerous single cell organisms and revealed a world that was previously not observable by the human eye. The discovery of organisms such as bacteria and fungi were only the start of what would be observed using microscopy.

1.1.2. Fluorescence and fluorescence microscopy

From the 17th century on microscopes were extensively used for research, leading to the development of different types of microscopes specifically designed for a particular application. One major challenge in imaging biological samples is that these samples are more or less optically transparent, which results in poor contrast. The contrast can be significantly increased by staining the biological material as is done in histology, however, the specificity is limited to sub-cellular structures.

Fluorescence microscopy is currently the most valuable tool for visualizing biological processes and structures, as the fluorescent labeling or staining enables both excellent imaging specificity and contrast, which is measured at high sensitivity. The development of the fluorescence microscope goes back to a discovery in 1833 by the preacher David Brewster. In his experiment, he illuminated chlorophyll and observed that the emission light was of a different wavelength than the illumination light. It was George Stokes in 1852 who explained David Brewster's discovery and named it *fluorescence*, after the fluoride that he used for his experiments. In 1934, John Marrack was the first to attach a fluorescent molecule to an

antibody. Subsequently, in 1942, Albert Coons applied this technique for the first time in fluorescence microscopy and by doing so increased the imaging specificity from the level of sub-cellular structures to proteins.

It took another 100 years after George Stokes' experiments for the typical layout of a fluorescence microscope to be drawn and realized in practice (Fig. 1.1). The crucial invention was epi-fluorescence illumination by Evgenii Brumberg and Johan Ploem, where the excitation and emission paths are both on the same side of the sample (Fig. 1.1) [1, 2]. Epi-fluorescence was first introduced by Brumberg in 1948 making use of a dichroic beamsplitter, which reflected the excitation and transmitted the emission light from the sample, therefore hindering the excitation light from going to the detector. In 1962, Ploem worked together with the glass company Schott to increase the efficiency of the dichroic beamsplitter to nearly 100%[1]. This illumination setup increased the fluorescence contrast significantly and made it possible to measure the relatively weak signals from fluorescent labels. The insertion of a beamsplitter into a converging light path can introduce image distortion and aberrations. Therefore, the epi-fluorescence microscope uses an infinity corrected objective to reduce image distortion and aberrations caused by the placement of a dichroic in an otherwise converging light path. Infinity corrected objectives also increased the flexibility in tube length.

A major discovery that boosted the popularity of fluorescence microscopy was the green fluorescent protein (GFP), a fluorescent protein that is endogenously expressed by a jellyfish (*Aequorea victoria*), discovered by Osamu Shimomura in 1962. Essential for the applicability of this protein was the work of Douglas Prasher and Martin Chalfie, who showed in 1994 that one can incorporate the gene that translates for GFP into the genome of other organisms in contrast to the delivery of antibodies to a permeabilized cell. In 1995, Roger Tsien increased the quantum efficiency and photo-stability of GFP, which made GFP a usable probe for fluorescence microscopy. The ability to incorporate the expression of a fluorescent protein into the genome of a live organism allows studying biological processes in a minimally perturbative way, which led to Chalfie, Tsien and Shimomura winning the Nobel Prize in Chemistry in 2008.

1.1.3. Single molecule localization microscopy

In 1873, Ernst Abbe showed that the optical resolution of a microscope is limited by the diffraction of light as it passes through the aperture of the microscope. Consequently, all microscopes will image an infinitesimally small point object

---

[1] A log scale is used to indicate the suppression of the mirror and six orders of magnitude are typical.

a

detector

tube lens

emission
filter

excitation
filter

dichroic
mirror

light source

b

marginal
ray

chief ray

α

high NA
objective

$NA = n\, sin(\alpha)$

fluorescent
sample
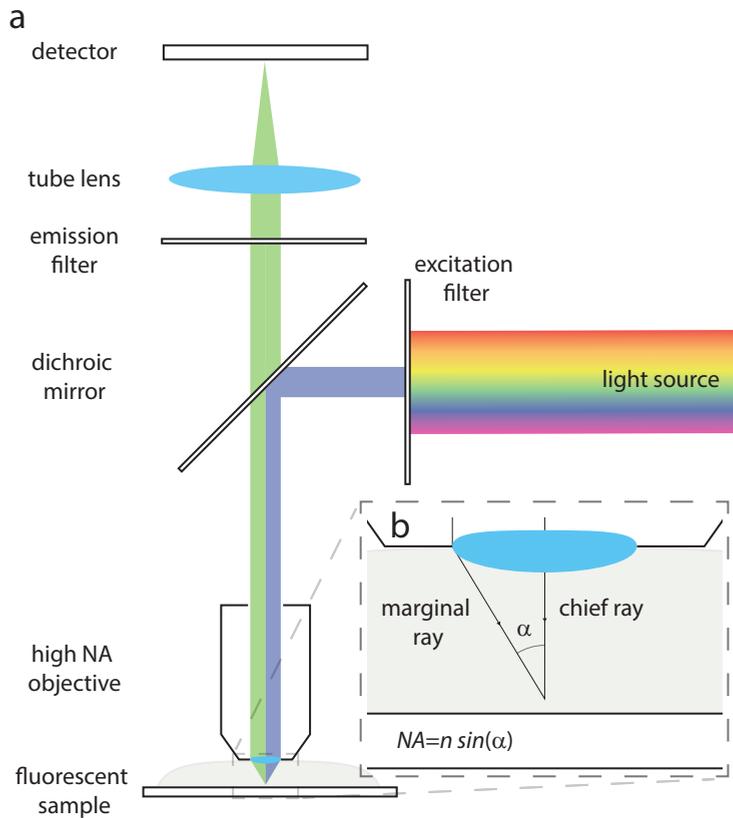
Figure 1.1. a) The typical layout of an epi-fluorescence microscope. The excitation light is filtered by the excitation filter and reflected from the dichroic mirror onto the sample. The objective captures the emission light from the sample and is transmitted through the dichroic mirror, and filtered by the emission filter. Finally, the emitted light is focused by the tube lens onto the detector. b) The definition of the numerical aperture (NA), where n is the refractive index of the immersion medium of the objective, and α the angle between the marginal ray and the chief ray.

as a finite sized spot. This characteristic impulse response function is called the point spread function of the microscope (PSF). The size of features that can still be discriminated is given by the Abbe limit[2] $\lambda/(2NA)$, where λ is the wavelength of the emitted light and $NA = n\sin(\alpha)$ is the numerical aperture (NA), with n the refractive index of the immersion medium, and α the angle between the marginal ray and the chief ray (Fig. 1.1). Considering visible light of $400 - 700$ nm and an NA of 1.4, the Abbe limit is roughly between $150 - 250$ nm, which is too large to resolve viruses (e.g. viral envelope of HIV, 100 nm) or small biological compounds (e.g. amino acids, 0.8 nm; fluorescent proteins, 5 nm; microfilaments, 6 nm). According to Abbe's limit, a better resolution can be obtained by techniques that use a shorter wavelength, such as electron microscopy, which can reach sub-nanometer resolution. This technique, however, does not allow the same imaging specificity and contrast as fluorescence microscopy nor does it allow for live-cell imaging.

Several nanoscopy techniques have been proposed to circumvent Abbe's resolution limit while keeping the benefits of fluorescence microscopy [3–5]. One of these techniques is single molecule localization microscopy, which makes use of the long known concept that the location of a single fluorescent particle can be determined with nanometer precision [6–8]. In the 1990s detectors became sensitive enough to measure single molecule fluorescence from a fluorescence microscope, and Michel Orrit in 1990 was the first person who detected fluorescence from a single molecule [9–11].

The first biological application of fluorescent localization was through Single Particle Tracking (SPT), and one of its pioneers was Thomas Schmidt. He measured space-time trajectories of isolated sub-resolution fluorescent particles in a phospholipid membrane [12–16]. These trajectories were used to extract quantitative information at the nanometer level regarding the molecular mechanisms and interactions (pathways) within a cell [13, 14].

In 1995, Eric Betzig published a theoretical concept, which explained how to circumvent Abbe's limit. The idea was to somehow image all molecules present in the sample separately. At that time, it was unclear how this could be achieved. If this could be accomplished, the single molecule positions are estimated in line with the SPT methodology and used to create an artificial image where the width (w) of the spots is not limited by the diffraction of light ($w \propto \lambda/(2NA)$), but by the localization precision ($w \propto \lambda/(NA\sqrt{N})$, where N is the number of photons collected from the single molecule). It was W.E. Moerner in 1997 who discovered

---

[2]Abbe's criterion should not be confused with the Rayleigh criterion. Abbe's criterion is derived in frequency space, and Rayleigh's criterion is derived in image space.

photo-activatable GFP and showed that it was possible to stochastically switch single molecules on and off. It was not until 2006 that three groups independently proposed single molecule localization microscopy (SMLM), namely Photo-Activated Localization Microscopy (PALM) [17], Fluorescence Photo-Activation Localization Microscopy (FPALM) [18], and Stochastic Optical Reconstruction Microscopy (STORM) [19]. In these methods, single molecules switch stochastically on and off, and over time all single molecules present in the sample are imaged sequentially. The final result would typically approach 20 nm lateral localization precision, thereby circumventing Abbe's diffraction limit [17–20]. Other methods followed, mainly differing in the necessary switching methodology, i.e. sparsity, which include ground state depletion microscopy followed by individual molecule return [21] and direct STORM [22]. For their contribution to super-resolution fluorescence microscopy, which is based on the stochastic nature of the label, Eric Betzig, W.E. Moerner, and Stefan Hell[3] received the Nobel Prize in Chemistry in 2014.

The main advantage of SMLM in comparison to other super-resolution methods, such as stimulated emission depletion and structured illumination microscopy (SIM) [4], is that relatively simple hardware can be used: an epi-fluorescence microscope; a light source; and a camera with low read-out noise and high quantum efficiency [23]. However, unlike a conventional fluorescence microscope, SMLM does not directly return an interpretable image. Image analysis is needed to convert the raw data into a super-resolution image. The pipeline to convert the raw data into a super-resolved image is discussed in the next section.

1.2. *Localization Pipeline*

The typical localization pipeline needed to convert a dataset into a super-resolution image is shown in Fig. 1.2A and presented in this section. Commonly, this pipeline consists of six steps: data acquisition, preprocessing, detection, localization, post-processing and visualization [24–26]. Preprocessing converts the images from analog-to-digital units into photo-electron units; detection identifies pixels with signal potentially arising from the emission of single molecules; localization

---

[3]Stefan Hell was not mentioned earlier in this text, but received this Nobel Prize for the development of a super-resolved scanning microscope developed in 1994, and first experimentally demonstrated in 1999; this scanning technique used a combination of two modulated diffraction limited lasers beams that can effectively activate a region of fluorescent molecules at a size smaller than the diffraction limit, also called stimulated emission depletion (STED).

[4]Structured Illumination Microscopy is a technique, where the sample is illuminated by a grid pattern generated from the diffraction of light [3]. Subsequently, the measurement needs to be processed to obtain an improvement in lateral resolution by a factor of up to two.
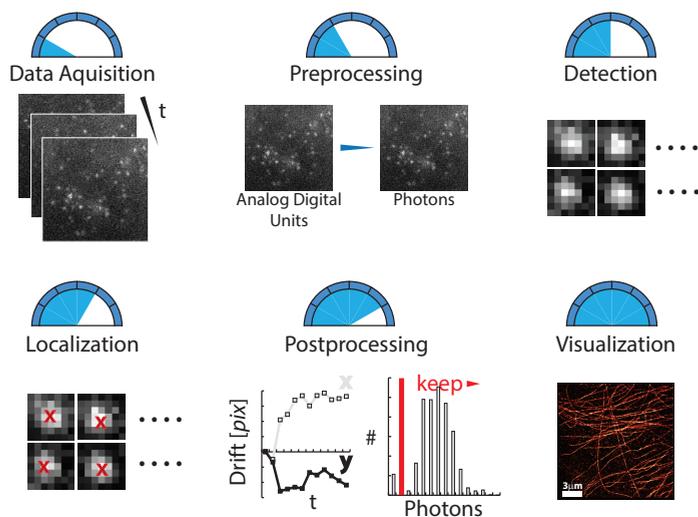
(A) Imaging pipeline to convert raw two-dimensional single molecule localization microscopy data into a super-resolution image.
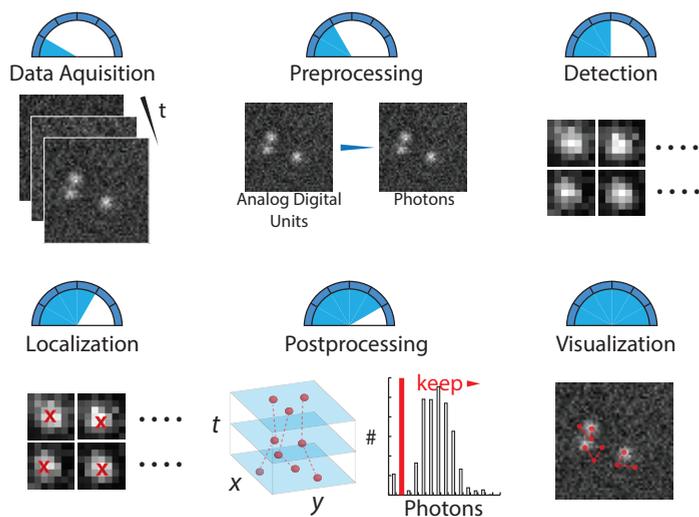
(B) Imaging pipeline to convert raw two-dimensional single molecule localization microscopy data into single particle tracks.

Figure 1.2. Single molecule localization pipelines: (**A**) Single molecule microscopy pipeline; (**B**) Single particle tracking pipeline.The pipelines consist of six steps: data acquisition, preprossessing, detection, localization, postprocessing (e.g. track construction and filtering), and visualization.

7

calculates an estimate of the single molecule position with sub-pixel precision; postprocessing performs tasks like filtering to reject unreliable fit results, drift correction, registration, and clustering of multiple localizations originating from a single on-event of an emitter; and finally, visualization renders reconstruction for display on the screen.

Data acquisition and preprocessing

As mentioned earlier, multiple single-molecule switching techniques exist that can be used to create the sparsity needed for localization microscopy. However, for all these methods one needs to image millions of molecules, using between 10,000 to 100,000 camera frames, to be able to create a super-resolution image. The faster the single molecule switching, the quicker an image can be generated. The switching kinetics of the fluorescent proteins (e.g. Cy5 and Alexa Flour 647) can be increased with higher laser powers [22, 27], and camera speeds of 3,200 frames per second are known to be achievable [28]. The laser power and single molecule switching speed significantly impact the background and the number of photons detected per molecule, and these two parameters determine the localization precision and therefore the final image resolution. A comprehensive study of the key factors that influence the final image quality is presented in Lin et al. [29].

Once the data is obtained, one needs to correct for offset and gain to perform optimal detection and localization [28, 30]. This process will be discussed in more detail, in the subsequent localization paragraph.

Detection

The third analysis step to be conducted, after the experiments and preprocessing have been performed, is the detection of regions that contain signals originating from single molecules, as shown in Fig. 1.2. To select these Regions of Interest (ROIs), usually, a threshold is set on the intensity relative to the background. The detection of these regions becomes particularly difficult when there is only a slight difference between the fluorescence signal and background, or when the background is non-uniform. An example of these conditions can be found in *in vivo* RNA imaging [31–34]. Detection methods that reduce the influence of shot noise and background non-uniformity on the detection quality have been proposed in [35–38].

In general, the two quantities that characterize the quality of detection are the detection efficiency (sometimes called sensitivity or recall), which is defined as the ratio of the true number of events over all detected events, and the false-positive rate, which is defined as the ratio of all false detections over the total number

8

of true and false detections. The detection efficiency and false-positive rate can be determined in simulations, but cannot be determined in an experimental dataset with an unknown ground truth. The currently available methods apply thresholds and filter settings that are determined empirically. Therefore, the detection efficiency and false-positive rate are not directly controllable. A way to limit the number of false positives is to manually tune the false-positive rate based on a visual judgment by using features that are calculated from the pixels above the user-set threshold. These features include the center of mass, circularity, cluster size and nearest cluster distance.

Localization

The location of a single molecule has to be estimated with subpixel precision. The most elementary method to do this estimation is by computing the center of mass (CM) [39]. The drawback of this method is that the underlying imaging model does not take into account background fluorescence, and it therefore leads to biased results in the presence of said background fluorescence. The most appropriate algorithm in localization microscopy for fitting an imaging model to the data is the Maximum Likelihood Estimator (MLE), where the MLE gives optimal results when an accurate noise and imaging model are applied [24, 40].

The type of likelihood function needed for the MLE is determined by the noise model associated with the choice of camera. Two types of cameras are commonly used in SMLM: the Electron Multiplying Charge-Coupled Device (EMCCD), and the scientific Complementary Metal–Oxide–Semiconductor (sCMOS). The advantage of EMCCDs is the high photon sensitivity and zero readout noise, making a Poisson distribution a working description for the noise of an EMCCD camera. sCMOS cameras support a higher frame rate, larger field of view, and smaller physical pixel size enabling Nyquist sampling at lower magnification. Over the past years the quantum efficiency of sCMOS has significantly increased (from $\sim$ 60% to $\sim$ 85%) and the readout noise decreased (from $\sim$ 2 e^- rms to $\sim$ 1 e^- rms), making them, at high enough photon counts (>10 photons/pixel), comparable to EMCCDs in terms of sensitivity [41]. These benefits have increased the use of sCMOS cameras significantly. A drawback of sCMOS cameras is that they suffer from pixel heterogeneity in the readout noise, in pixel offset, and in gain, which cannot be neglected in the localization algorithm[5]. These two types of cameras

---

[5]Not all sCMOS cameras support an emulated global shutter mode and instead use a rolling shutter. An sCMOS with a global shutter can take a snapshot, while the rolling shutter will read line by line, thereby losing time correlation between pixels. The consequence is that objects that are imaged with rolling shutter mode can only be captured undistorted when moving slower than $v_{max} << \Delta x / T$, where Δx the size of a pixel and T the exposure time.

have different calibration methods, which are vital for optimal performance in SMLM [28, 30]. The calibration methods take into account the different noise models and for sCMOS the pixel non-uniformity.

As previously noted, a MLE algorithm also requires a PSF model. Commonly, the PSF is approximated by a Gaussian distribution with a width of σ_0, which is known to be a valid approach in the context of two-dimensional single emitter localization [42]:

$$PSF(x, y) = \frac{1}{2\pi\sigma_0^2} e^{-\frac{1}{2\sigma_0^2}\left[(x-x_0)^2+(y-y_0)^2\right]}. \tag{1.1}$$

This PSF must be integrated over the pixel area to arrive at the expected photon count at each pixel k:

$$\mu_k = \theta_I \Delta E(x_k - \theta_x)\Delta E(y_j - \theta_y) + \theta_{bg}, \tag{1.2}$$

with

$$\Delta E(u) = \frac{1}{2}\left[\text{erf}\left(\frac{u+1/2}{\sqrt{2}\sigma_0}\right) - \text{erf}\left(\frac{u-1/2}{\sqrt{2}\sigma_0}\right)\right], \tag{1.3}$$

where (x_k, y_k) are the pixel coordinates in unit $[pixel]$ of pixel k, (θ_x, θ_y) is the location of the center of the PSF in unit $[pixel]$, θ_I and θ_{bg} are the single molecule intensity and background, respectively, and σ_0 is the PSF width.

Typically, MLE fitting routines assume the PSF width σ_0 to be known [37, 43, 44], but σ_0 can also be treated as an additional fit parameter [37, 45]. For overlapping emitters one can fit multiple emitters simultaneously [28, 46–49]. An alternative to the Gaussian PSF is the use of a PSF obtained from calibration experiments [50], and an extensive overview and benchmark of different localization algorithms can be found in Sage et al. [25].

Any single molecule localization technique is ultimately limited by the localization precision and labeling density of the single molecule, instead of Abbe's limit [51]. Harald Cramér and Calyampudi Rao proved in 1940 that the precision of any unbiased estimator is bounded, and the bound is called the Cramér–Rao Lower Bound (CRLB) or the information inequality [52]. The CRLB can be computed numerically for many different imaging conditions [40], but analytical approximations are also available [53, 54]. A reasonable and concise approximation of the CRLB for single molecule localization in the absence of readout noise from the

camera is [54]:

$$\sigma_x^2 = \frac{\sigma_a^2}{N}\left(1 + 4\tau\sqrt{\frac{2\tau}{1+4\tau}}\right) \tag{1.4}$$

with $\tau = 2\pi\sigma_a^2 b/(Na^2)$ and $\sigma_a^2 = \sigma^2 + a^2/12$, where a is the camera's pixel size, N the total signal photon count, b the background photon count per pixel and σ the width of the Gaussian that is used to fit the PSF. In practice, this results in a localization precision of 5-100 nm. A key element that is often overlooked is that the equation concerns only the position estimate of the fluorescent label. The size of the fluorescent label and of the linker between the label and the binding site on the protein of interest ultimately limit the precision at which the structure can be imaged (e.g. immunoglobulin (IgG) antibody size is 70 nm; coated quantum dot size is 40 nm; fluorescent protein size is 5 nm; organic dye size is 1 nm). The better the localization precision becomes, the more important the size of the fluorescent label and linker will be.

Postprocessing

The first postprocessing step is to conduct quality control on all the segmented and fitted candidate single molecule emission events. Estimated parameters obtained from the detection and localization step are used to maximize detection efficiency and minimize false-positive rate, ensuring a high quality reconstructed super-resolution image.

Another typical postprocessing step is drift correction. Even if the setup contains an active control system, the measurement may contain observable drift, even when the cells are stationary. The correction of stage drift is crucial in most super-resolution experiments [55, 56]. The two most common ways to achieve correction are: i) to attach fiducial markers to the sample, such as fluorescent beads, quantum dots, DNA or nuclear periphery when stained or labeled; or ii) to estimate the drift using the cross-correlation between frames [57]. This second option does not require a change of the experimental setup or the sample preparation.

The third postprocessing step is to combine multiple localizations. In a super-resolution experiment, a single molecule could be 'on' for a couple of frames causing the single molecule to be localized multiple times. Combining multiple localizations into a single localization improves the localization precision. A criterion for this merging is that the distance between the multiple localizations of this single molecule in space and time must be less than a user defined threshold.

Visualization

Once all the subpixel positions of the single molecules are estimated and filtered, a super-resolution reconstruction can be generated. Most often this reconstruction is done by placing a two-dimensional Gaussian profile on the estimated position of the single emitter, where the width of the Gaussian represents the estimation precision of the position [17, 18, 20, 58, 59]. The reported values of the localization precision are generally in the range of 5-100 nm. In the application of super-resolution imaging, it may be required to find the positions of more than 10^6 fluorophores in order to generate one final image for a typical field-of-view of $50 \times 50 \ \mu m^2$. Based on the reconstruction, further quantifications of the observed structure can be carried out such as co-localization or co-orientation [60–62].

Until 2013, a long-standing issue in single molecule localization microscopy was the quantification of the resolution of published reconstructions, since the combined effect of localization precision and labeling density on the resolution could not be quantified.

1.2.1. Single particle tracking

Single molecule localization microscopy has a lot in common with single particle tracking, and as a result, the typical image analysis workflow needed to convert raw data into single particle tracks is similar, see Fig. 1.2. For super-resolved imaging the single molecules have to be detected and localized in a likewise manner. Therefore, the four image processing steps are equivalent to what was described earlier. The workflow differs with regards to the postprocessing and visualization steps. In these two steps, single molecule trajectories are constructed and visualized.

A multitude of algorithms exist for constructing single molecule trajectories from single molecule localizations and multiple comparison studies have been performed of these algorithms [63–66]. The most elementary methods are based on nearest neighbor linking e.g. [67], however, probabilistic methods are superior according to Godinez et al. [66]. Commonly, probabilistic methods take into account parameters such as the bleach rate, birth rate, localization precision, and the distance between a localization and neighboring localizations in the same and adjacent frames [68–70]. The most typical way to display tracking data is to plot the data points of the actual observations and link them linearly to show a perceived trajectory [71].

Important information regarding the pathway and location of single molecules can be estimated from the single molecule trajectories. An example is found in the tracking of RNA molecules. These molecules are known to serve a multitude

of tasks such as being templates for protein translation or acting as enzymes for regulating reactions in the nucleus. The information obtained from studying RNA in living cells can provide new information on RNA function or even human diseases [72–75].

Tracking measurements are often quantified by analyzing the mean squared displacement or the step-size distribution [76–78], where the speed at which a single molecule moves is estimated. Different modes of motion can be discriminated from this analysis, such as random motion and active transport. Additionally, dwell time analysis can provide information on the speed at which a single molecule is moving [34] and more detailed biochemical or kinetic information can be obtained using (hidden) Markov model estimation [79, 80]. Additionally, it has been found that the same single molecule can move at different speeds depending on the micro-environment [81, 82], which can also be modeled using hidden Markov models [83].

1.3. *Imaging in Three Dimensions*

In the previous section, the pipeline of two-dimensional localization microscopy has been discussed. Several methods exist to extend single molecule localization microscopy to three dimensions. Some of these possible extensions will be discussed in this section.

A popular method is to encode three-dimensional information into the shape of the PSF, for example by using astigmatism, double helix, or similar engineered PSF approaches [84–87]. Another option is to estimate the three-dimensional position using multi-focus microscopy (MFM) [88, 89]. Multi-focus microscopy has the advantage of being able to handle a higher density of single molecules since the footprint of the PSF is smaller when the focus planes are imaged separately. The drawback of multi-focus microscopy is that multiple planes have to be registered and that the photons generated by a single molecule are divided over different focal planes. PSF engineering methods and MFM methods spread out the light over a larger area than the area of an in-focus diffraction limited spot, so a higher signal-to-background ratio is needed than in conventional two-dimensional SMLM. An illustrative overview of the different methodologies is given in Deschout et al.[40].

It is advantageous to use PSF models that account for the aberrations as well as polarization and high-NA effects (so-called vectorial effects) to successfully move from two-dimensional to three-dimensional imaging, specifically for thicker samples where aberrations are more dominant [90–92]. These (spherical) aberrations

increase linearly with depth [93–96], and are especially relevant when high-NA lenses are used. Aberrations can already be problematic when imaging single cells such as animal cells that can be as thick as 30 μm and plant cells that can even be as thick as 100 μm [91]. The localization precision not only depends on the number of observed photons, but also on the width of the PSF, which is a function of NA, λ, and optical aberrations. The optical aberrations can be minimized by applying adaptive optics, through the use of a deformable mirror or a spatial light modulator to compensate for them. The reduction of aberrations will decrease the PSF width (σ_0, Equation 1.1) and therefore improve the CRLB (Equation 1.4).

1.4. *Thesis Motivation*

Data collected through single molecule localization microscopy enables us to gain a quantitative understanding of the mechanisms within the living cell at relatively low light intensities. Single molecule localization microscopy permits a relatively simple microscope design, followed by sophisticated image analysis.

The tremendous increase in data that needs to be analyzed is one of the greatest differences between conventional fluorescence microscopy and single molecule localization microscopy. As previously outlined, the diffraction patterns of single molecules are detected, and properties such as position and intensity are estimated. A challenge is to develop image analysis methods that perform the detection and the estimation as efficiently and precisely as possible, and simultaneously estimate the performance of the methods, making the process truly quantitative. A second challenge is that image analysis methods contradict the user's expectation that an image is instantaneously created, as is the case for conventional fluorescence microscopes. This challenge can be addressed by performing the analysis steps as fast as possible to enable near real-time rendering of the images from the list of analyzed single molecules with minimal user intervention.

In contrast to conventional fluorescence microscopy, the outcome of a single molecule localization experiment is a list of estimated parameters, which requires image analysis tools specific to this form of data. These tools need to be able to take into account apriori information from the experiment. If a wrong model is imposed on the data, the chances are high that one creates artifacts leading to misinterpretation of the data [97]. A major challenge is therefore to apply single molecule localization microscopy to biological problems such that insight into biological structures and processes is truly gained. A good example is the imaging of RNA molecules at low signal to background ratios. Often it is known

where and with what molecule the RNA molecule interacts, but the kinetics or the order of interaction are unknown. This creates an evident opportunity for the development of imaging analysis tools that can extract RNA dynamics at a single molecule level. RNA imaging is important as RNA regulates gene expression in cells and thus has the potential to be a target for drugs to treat human disease [98].

It is essential to extend single molecule localization microscopy to three dimensions, to gain a complete understanding of RNA regulation. RNA dynamics happen in a three-dimensional space, which as a whole undergoes major reformation during the cell cycle [99]. As previously mentioned, it is possible to encode three-dimensional information into the measurements. However, there are many options to choose from and the challenge is to find the solution that results in an optimum localization precision. Additionally, one could encode more information into the measurements, such as the emission color of the single molecule, which allows the simultaneous classification and tracking of multiple species of mRNA in a single channel setup.

1.5. *Outline*

Chapter 2 describes how to estimate the position of single molecules with a precision approaching the CRLB. For this purpose, an iterative algorithm is introduced whose result converges to the maximum likelihood estimate for the position and intensity of a single fluorophore, as well as the background fluorescence. The technique assumes a Gaussian PSF model and uniform background. Furthermore, an implementation of the algorithm on Graphics Processing Unit (GPU) hardware achieves more than one million combined fits and CRLB calculations per second enabling nearly real-time data analysis.

Detection of single molecule data heavily relies on user input, such as filter settings and thresholds. Furthermore, at low signal to background levels, there is often a high variability in results obtained by different users with no error indication. The aim of Chapter 3 is to reduce user variability and limit the possible detection errors. For this purpose, the previously developed MLE is applied to perform a hypothesis test in each pixel, where we test if the signal in that pixel is originating from the background, or from signal plus background. A framework is presented where expert knowledge and parameter tweaking is replaced with a probability-based hypothesis test, delivering robust and threshold-free signal detection with an improved detection efficiency of weaker signals at a bounded

false alarm rate. Simulations, experimental data, and tracking of low signal are presented for mRNAs in yeast cells.

In Chapter 4 the spatial and temporal activities of mRNAs during export are studied. Yeast cells were used containing a labeled nuclear pore complex and mRNA, which allows the tracking of a single-particle mRNP, an association that combines RNA and proteins such as splicing factors, export factors, and zip-codes together. The position of the mRNP is determined relative to the nuclear pore complex with high spatial precision and temporal resolution.

In Chapter 5 a method is proposed for registration and (spherical) aberration correction. Multi-focus microscopy (MFM) is applied to instantaneously capture three-dimensional single-molecule real-time (3D-SMRT) images in live cells, visualizing cell nuclei at ten volumes per second [88].

In Chapter 6 an adaptive optics method is presented to correct for aberrations that are present in three-dimensional applications. The aberrations induced by a thick sample need to be initially estimated from the measurements with the microscope, to be able to correct for them. One of the approaches to do so is to estimate the aberrations from two measurements, where a defocus is applied to one of the measurements. This technique is called phase diversity and results in a nonlinear optimization problem. In this chapter, a theoretical study is presented for efficient approximation of this nonlinear optimization problem. The aim of the study is to enable the application of phase diversity in real-time adaptive optics. The new iterative linear phase diversity method presented assumes that the residual phase aberration is small and makes use of a first order Taylor expansion of the Zernike coefficients at the applied defocus, which allows for arbitrary (large) pupil shapes to optimize the phase retrieval.

In Chapter 7 a new method is proposed for the simultaneous measurement of the three-dimensional position and the emission wavelength of single emitters. Here we investigate a diffractive optics based vectorial PSF design in which the spot is split into closely spaced diffraction orders. Experiments were performed using a liquid crystal based SLM for which the nominal phase and amplitude aberrations were calibrated to obtain a match between simulation and experiments.

The final Chapter of this thesis, Chapter 8, presents concluding remarks, recommendations and an outlook on future opportunities in single molecule imaging.

References

[1] F. H. Kasten, *The origins of modern fluorescence microscopy and fluorescent probes,* Cell Structure and Function by Microspectrofluorometry , 3 (1989).

[2] I. B. Buchwalow and W. Böcker, *Immunohistochemistry: Basics and methods* (Springer Science & Business Media, 2010).

[3] A. D. S. J. Gustafsson, M.G.L., US patent, 5671085 **cols. 23–25** (1997).

[4] S. W. Hell and J. Wichmann, *Breaking the diffraction resolution limit by stimulated emission: Stimulated-emission-depletion fluorescence microscopy,* Optics Letters **19**, 780 (1994).

[5] E. Betzig, *Proposed method for molecular optical imaging,* Optics Letters **20**, 237 (1995).

[6] N. Bobroff, *Position measurement with a resolution and noise-limited instrument,* Review of Scientific Instruments **57**, 1152 (1986).

[7] C. W. McCutchen, *Superresolution in microscopy and the abbe resolution limit,* Journal of the Optical Society of America **57**, 1190 (1967).

[8] G. T. Di Francia, *Resolving power and information,* Journal of the Optical Society of America **45**, 497 (1955).

[9] W. Ambrose and W. Moerner, *Fluorescence spectroscopy and spectral diffusion of single impurity molecules in a crystal,* Nature **346**, 225 (1991).

[10] M. Orrit and J. Bernard, *Single pentacene molecules detected by fluorescence excitation in a p-terphenyl crystal,* Physical Review Letters **65**, 2716 (1990).

[11] T. Funatsu, Y. Harada, M. Tokunaga, K. Saito, T. Yanagida, *et al.*, *Imaging of single fluorescent molecules and individual atp turnovers by single myosin molecules in aqueous solution,* Nature **374**, 555 (1995).

[12] G. Schütz, H. Schindler, and T. Schmidt, *Single-molecule microscopy on model membranes reveals anomalous diffusion.* Biophysical Journal **73**, 1073 (1997).

[13] J. Gelles, B. J. Schnapp, and M. P. Sheetz, *Tracking kinesin-driven movements with nanometre-scale precision,* Nature **331**, 450 (1988).

[14] H. Qian, M. P. Sheetz, and E. L. Elson, *Single particle tracking. analysis of diffusion and flow in two-dimensional systems.* Biophysical Journal **60**, 910 (1991).

[15] T. Schmidt, G. Schütz, W. Baumgartner, H. Gruber, and H. Schindler, *Imaging of single molecule diffusion,* Proceedings of the National Academy of Sciences **93**, 2926 (1996).

[16] U. Kubitscheck, O. Kückmann, T. Kues, and R. Peters, *Imaging and tracking of single GFP molecules in solution,* Biophysical Journal **78**, 2170 (2000).

[17] 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**, 1642 (2006).

[18] S. T. Hess, T. P. K. Girirajan, and M. D. Mason, *Ultra-high resolution imaging by fluorescence photoactivation localization microscopy,* Biophysical Journal **91**, 4258 (2006).

[19] M. J. Rust, M. Bates, and X. W. Zhuang, *Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM),* Nature Methods **3**, 793 (2006).

[20] K. A. Lidke, B. Rieger, T. M. Jovin, and R. Heintzmann, *Superresolution by localiza-*

tion of quantum dots using blinking statistics, Optics Express **13**, 7052 (2005).

[21] J. Fölling, M. Bossi, H. Bock, R. Medda, C. A. Wurm, B. Hein, S. Jakobs, C. Eggeling, and S. W. Hell, *Fluorescence nanoscopy by ground-state depletion and single-molecule return,* Nature Methods **5**, 943 (2008).

[22] M. Heilemann, S. Van De Linde, M. Schüttpelz, R. Kasper, B. Seefeldt, A. Mukherjee, P. Tinnefeld, and M. Sauer, *Subdiffraction-resolution fluorescence imaging with conventional fluorescent probes,* Angewandte Chemie International Edition **47**, 6172 (2008).

[23] S. Herbert, H. Soares, C. Zimmer, and R. Henriques, *Single-molecule localization super-resolution microscopy: Deeper and faster,* Microscopy and Microanalysis **18**, 1419 (2012).

[24] A. Small and S. Stahlheber, *Fluorophore localization algorithms for super-resolution microscopy,* Nature Methods **11**, 267 (2014).

[25] D. Sage, H. Kirshner, T. Pengo, N. Stuurman, J. Min, S. Manley, and M. Unser, *Quantitative evaluation of software packages for single-molecule localization microscopy,* Nature Methods **12**, 717 (2015).

[26] B. Rieger, R. Nieuwenhuizen, and S. Stallinga, *Image processing and analysis for single-molecule localization microscopy: Computation for nanoscale imaging,* Signal Processing Magazine, Institute of Electrical and Electronics Engineers **32**, 49 (2015).

[27] M. Bates, T. R. Blosser, and X. Zhuang, *Short-range spectroscopic ruler based on a single-molecule optical switch,* Physical Review Letters **94**, 108101 (2005).

[28] F. Huang, T. M. Hartwich, F. E. Rivera-Molina, Y. Lin, W. C. Duim, J. J. Long, P. D. Uchil, J. R. Myers, M. A. Baird, W. Mothes, *et al.,* *Video-rate nanoscopy using sCMOS camera-specific single-molecule localization algorithms,* Nature Methods **10**, 653 (2013).

[29] Y. Lin, J. J. Long, F. Huang, W. C. Duim, S. Kirschbaum, Y. Zhang, L. K. Schroeder, A. A. Rebane, M. G. M. Velasco, A. Virrueta, *et al.,* *Quantifying and optimizing single-molecule switching nanoscopy at high speeds,* PloS One **10**, e0128135 (2015).

[30] L. J. Van Vliet, D. Sudar, and I. T. Young, *Digital fluorescence imaging using cooled charge-coupled device array cameras,* Cell Biology: A Laboratory Handbook , 109 (1998).

[31] M. F. Juette, T. J. Gould, M. D. Lessard, M. J. Mlodzianoski, B. S. Nagpure, B. T. Bennett, S. T. Hess, and J. Bewersdorf, *Three-dimensional sub–100 nm resolution fluorescence microscopy of thick samples,* Nature Methods **5**, 527 (2008).

[32] S. Manley, J. M. Gillette, G. H. Patterson, H. Shroff, H. F. Hess, E. Betzig, and J. Lippincott-Schwartz, *High-density mapping of single-molecule trajectories with photoactivated localization microscopy,* Nature Methods **5**, 155 (2008).

[33] S. Tyagi, *Imaging intracellular RNA distribution and dynamics in living cells,* Nature Methods **6**, 331 (2009).

[34] D. Grünwald and R. H. Singer, *In vivo imaging of labelled endogenous β-actin mRNA during nucleocytoplasmic transport,* Nature **467**, 604 (2010).

[35] J.-C. Olivo-Marin, *Extraction of spots in biological images using multiscale products,* Pattern Recognition **35**, 1989 (2002).

[36] A. Sergé, N. Bertaux, H. Rigneault, and D. Marguet, *Dynamic multiple-target tracing to probe spatiotemporal cartography of cell membranes,* Nature Methods **5**, 687 (2008).

[37] C. S. Smith, N. Joseph, B. Rieger, and K. A. Lidke, *Fast, single-molecule localization that achieves theoretically minimum uncertainty,* Nature Methods **7**, 373 (2010).

[38] 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**, 2081 (2012).

[39] A. J. Berglund, M. D. McMahon, J. J. Mc-Clelland, and J. A. Liddle, *Fast, bias-free algorithm for tracking single particles with variable size and shape,* Optics Express **16**, 14064 (2008).

[40] H. Deschout, F. C. Zanacchi, M. Mlodzianoski, A. Diaspro, J. Bewersdorf, S. T. Hess, and K. Braeckmans, *Precisely and accurately localizing single emitters in fluorescence microscopy,* Nature Methods **11**, 253 (2014).

[41] J. Chao, E. S. Ward, and R. J. Ober, *Localization accuracy in single molecule microscopy using electron-multiplying charge-coupled device cameras,* in *SPIE Biomedical Optics Conference* (International Society for Optics and Photonics, 2012) pp. 82271P–82271P.

[42] S. Stallinga and B. Rieger, *Accuracy of the gaussian point spread function model in 2D localization microscopy,* Optics Express **18**, 24461 (2010).

[43] R. J. Ober, S. Ram, and E. S. Ward, *Localization accuracy in single-molecule microscopy,* Biophysical Journal **86**, 1185 (2004).

[44] R. Ober and E. Ward, *Estimation tool,* (2009), university of Texas, Southwestern http://www4.utsouthwestern.edu/wardlab.

[45] R. Starr, S. Stahlheber, and A. Small, *Fast maximum likelihood algorithm for localization of fluorescent molecules,* Optics Letters **37**, 413 (2012).

[46] S. J. Holden, S. Uphoff, and A. N. Kapanidis, *DAOSTORM: An algorithm for high-density super-resolution microscopy,* Nature Methods **8**, 279 (2011).

[47] F. Huang, S. L. Schwartz, J. M. Byars, and K. A. Lidke, *Simultaneous multiple-emitter fitting for single molecule super-resolution imaging,* Biomedical Optics Express **2**, 1377 (2011).

[48] T. Quan, H. Zhu, X. Liu, Y. Liu, J. Ding, S. Zeng, and Z.-L. Huang, *High-density localization of active molecules using structured sparse model and bayesian information criterion,* Optics Express **19**, 16963 (2011).

[49] Y. Wang, T. Quan, S. Zeng, and Z.-L. Huang, *PALMER: A method capable of parallel localization of multiple emitters for high-density localization microscopy,* Optics Express **20**, 16039 (2012).

[50] M. J. Mlodzianoski, M. F. Juette, G. L. Beane, and J. Bewersdorf, *Experimental characterization of 3D localization techniques for particle-tracking and super-resolution microscopy,* Optics Express **17**, 8264 (2009).

[51] R. P. Nieuwenhuizen, K. A. Lidke, M. Bates, D. L. Puig, D. Grünwald, S. Stallinga, and B. Rieger, *Measuring image resolution in optical nanoscopy,* Nature Methods **10**, 557 (2013).

[52] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory* (Prentice Hall, 1993).

[53] R. E. Thompson, D. R. Larson, and W. W. Webb, *Precise nanometer localization analysis for individual fluorescent probes,* Biophysical Journal **82**, 2775 (2002).

[54] B. Rieger and S. Stallinga, *The lateral and axial localization uncertainty in super-resolution light microscopy,* ChemPhysChem **15**, 664 (2014).

[55] 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**, 15009 (2011).

[56] M. Bates, B. Huang, G. T. Dempsey, and X. Zhuang, *Multicolor super-resolution imaging with photo-switchable fluorescent probes,* Science **317**, 1749 (2007).

[57] T. Q. Pham, M. Bezuijen, L. J. Van Vliet, K. Schutte, and C. L. L. Hendriks, *Performance of optimal registration estimators,* in

Defense and Security (International Society for Optics and Photonics, 2005) pp. 133–144.

[58] M. Bates, B. Huang, G. T. Dempsey, and X. Zhuang, *Multicolor super-resolution imaging with photo-switchable fluorescent probes,* Science **317**, 1749 (2007).

[59] A. Egner, C. Geisler, C. Von Middendorff, H. Bock, D. Wenzel, R. Medda, M. Andresen, A. C. Stiel, S. Jakobs, C. Eggeling, A. Schonle, and S. W. Hell, *Fluorescence nanoscopy in whole cells by asynchronous localization of photoswitching emitters,* Biophysical Journal **93**, 3285 (2007).

[60] M. Heilemann, D. P. Herten, R. Heintzmann, C. Cremer, C. Müller, P. Tinnefeld, K. D. Weston, J. Wolfrum, and M. Sauer, *High-resolution colocalization of single dye molecules by fluorescence lifetime imaging microscopy,* Analytical Chemistry **74**, 3511 (2002).

[61] W. Trabesinger, B. Hecht, U. Wild, G. Schütz, H. Schindler, and T. Schmidt, *Statistical analysis of single-molecule colocalization assays,* Analytical Chemistry **73**, 1100 (2001).

[62] R. P. Nieuwenhuizen, L. Nahidiazar, E. M. Manders, K. Jalink, S. Stallinga, and B. Rieger, *Co-orientation: Quantifying simultaneous co-localization and orientational alignment of filaments in light microscopy,* PloS One **10**, e0131756 (2015).

[63] I. Smal, M. Loog, W. Niessen, and E. Meijering, *Quantitative comparison of spot detection methods in fluorescence microscopy,* Institute of Electrical and Electronics Engineers Transactions on Medical Imaging **29**, 282 (2010).

[64] P. Ruusuvuori, T. Äijö, S. Chowdhury, C. Garmendia-Torres, J. Selinummi, M. Birbaumer, A. M. Dudley, L. Pelkmans, and O. Yli-Harja, *Evaluation of methods for detection of fluorescence labeled subcellular objects in microscope images,* BioMed Central Bioinformatics **11**, 1 (2010).

[65] N. Chenouard, I. Smal, F. De Chaumont, M. Maška, I. F. Sbalzarini, Y. Gong, J. Cardinale, C. Carthel, S. Coraluppi, M. Winter, *et al.*, *Objective comparison of particle tracking methods,* Nature Methods **11**, 281 (2014).

[66] W. J. Godinez, M. Lampe, S. Wörz, B. Müller, R. Eils, and K. Rohr, *Deterministic and probabilistic approaches for tracking virus particles in time-lapse fluorescence microscopy image sequences,* Medical Image Analysis **13**, 325 (2009).

[67] J. C. Crocker and D. G. Grier, *Methods of digital video microscopy for colloidal studies,* Journal of colloid and interface science **179**, 298 (1996).

[68] P. J. Cutler, M. D. Malik, S. Liu, J. M. Byars, D. S. Lidke, and K. A. Lidke, *Multi-color quantum dot tracking using a high-speed hyperspectral line-scanning microscope,* PloS One **8**, e64320 (2013).

[69] K. Jaqaman, D. Loerke, M. Mettlen, H. Kuwata, S. Grinstein, S. L. Schmid, and G. Danuser, *Robust single-particle tracking in live-cell time-lapse sequences,* Nature Methods **5**, 695 (2008).

[70] M. J. Saxton, *Single-particle tracking: Connecting the dots,* Nature Methods **5**, 671 (2008).

[71] C. Manzo and M. F. Garcia-Parajo, *A review of progress in single particle tracking: From methods to biophysical insights,* Reports on Progress in Physics **78**, 124601 (2015).

[72] G. Wang, Z. Huang, X. Liu, W. Huang, S. Chen, Y. Zhou, D. Li, R. H. Singer, and W. Gu, *Imp1 suppresses breast tumor growth and metastasis through the regulation of its target mRNAs.* Oncotarget (2016).

[73] Z. B. Katz, B. p English, T. Lionnet, Y. J. Yoon, N. Monnier, B. Ovryn, M. Bathe, and R. H. Singer, *Mapping translation "hot-spots" in live cells by tracking single molecules of mRNA and ribosomes,* eLife , e10415 (2016).

[74] B. Wu, A. R. Buxbaum, Z. B. Katz, Y. J. Yoon, and R. H. Singer, *Quantifying protein-mRNA*

interactions in single live cells, Cell **162**, 211 (2015).

[75] S. M. Janicki, T. Tsukamoto, S. E. Salghetti, W. P. Tansey, R. Sachidanandam, K. V. Prasanth, T. Ried, Y. Shav-Tal, E. Bertrand, R. H. Singer, *et al.*, *From silencing to gene expression: Real-time analysis in single cells,* Cell **116**, 683 (2004).

[76] U. Kubitscheck, D. Grünwald, A. Hoekstra, D. Rohleder, T. Kues, J. P. Siebrasse, and R. Peters, *Nuclear transport of single molecules dwell times at the nuclear pore complex,* The Journal of Cell Biology **168**, 233 (2005).

[77] P. J. Bosch, J. S. Kanger, and V. Subramaniam, *Classification of dynamical diffusion states in single molecule tracking microscopy,* Biophysical Journal **107**, 588 (2014).

[78] P. K. Relich, K. A. Lidke, and M. J. Olah, *Estimating the diffusion constant from noisy trajectories,* Biophysical Journal **108**, 317a (2015).

[79] S. T. Low-Nam, K. A. Lidke, P. J. Cutler, R. C. Roovers, P. M. v. B. en Henegouwen, B. S. Wilson, and D. S. Lidke, *ErbB1 dimerization is promoted by domain co-confinement and stabilized by ligand binding,* Nature Structural and Molecular Biology **18**, 1244 (2011).

[80] N. Monnier, Z. Barry, H. Y. Park, K.-C. Su, Z. Katz, B. P. English, A. Dey, K. Pan, I. M. Cheeseman, R. H. Singer, *et al.*, *Inferring transient particle transport dynamics in live cells,* Nature Methods **12**, 838 (2015).

[81] D. Grünwald, R. M. Martin, V. Buschmann, D. P. Bazett-Jones, H. Leonhardt, U. Kubitscheck, and M. C. Cardoso, *Probing intranuclear environments at the single-molecule level,* Biophysical Journal **94**, 2847 (2008).

[82] J. P. Siebrasse, T. Kaminski, and U. Kubitscheck, *Nuclear export of single native mRNA molecules observed by light sheet fluorescence microscopy,* Proceedings of the

National Academy of Sciences **109**, 9426 (2012).

[83] F. Persson, M. Lindén, C. Unoson, and J. Elf, *Extracting intracellular diffusive states and transition rates from single-molecule tracking data,* Nature Methods **10**, 265 (2013).

[84] A. S. Backer and W. Moerner, *Extending single-molecule microscopy using optical Fourier processing,* The Journal of Physical Chemistry B **118**, 8313 (2014).

[85] Y. Shechtman, S. J. Sahl, A. S. Backer, and W. Moerner, *Optimal point spread function design for 3D imaging,* Physical Review Letters **113**, 133902 (2014).

[86] Y. Shechtman, L. E. Weiss, A. S. Backer, S. J. Sahl, and W. Moerner, *Precise three-dimensional scan-free multiple-particle tracking over large axial ranges with tetrapod point spread functions,* Nano Letters **15**, 4194 (2015).

[87] A. S. Backer, M. P. Backlund, A. R. von Diezmann, S. J. Sahl, and W. Moerner, *A bisected pupil for studying single-molecule orientational dynamics and its application to three-dimensional super-resolution microscopy,* Applied Physics Letters **104**, 193701 (2014).

[88] S. Abrahamsson, J. Chen, B. Hajj, S. Stallinga, A. Y. Katsov, J. Wisniewski, G. Mizuguchi, P. Soule, F. Mueller, C. D. Darzacq, *et al.*, *Fast multicolor 3D imaging using aberration-corrected multifocus microscopy,* Nature Methods **10**, 60 (2013).

[89] S. Ram, P. Prabhat, J. Chao, E. S. Ward, and R. J. Ober, *High accuracy 3D quantum dot tracking with multifocal plane microscopy for the study of fast intracellular dynamics in live cells,* Biophysical Journal **95**, 6025 (2008).

[90] S. Quirin, S. R. P. Pavani, and R. Piestun, *Optimal 3D single-molecule localization for superresolution microscopy with aberrations and engineered point spread functions,* Proceedings of the National Academy of Sciences **109**, 675 (2012).

[91] S. Liu, E. B. Kromann, W. D. Krueger, J. Bewersdorf, and K. A. Lidke, *Three dimensional single molecule localization using a phase retrieved pupil function,* Optics Express **21**, 29462 (2013).

[92] J. Enderlein, E. Toprak, and P. Selvin, *Polariztion effect on position accuracy of fluorophore localization,* Optics Express **14**, 8112 (2006).

[93] S. Hell, G. Reiner, C. Cremer, and E. H. Stelzer, *Aberrations in confocal fluorescence microscopy induced by mismatches in refractive index,* Journal of Microscopy **169**, 391 (1993).

[94] P. Török, P. Varga, Z. Laczik, and G. Booker, *Electromagnetic diffraction of light focused through a planar interface between materials of mismatched refractive indices: An integral representation,* Journal of the Optical Society of America A **12**, 325 (1995).

[95] M. J. Booth, M. Neil, and T. Wilson, *Aberration correction for confocal imaging in refractive-index-mismatched media,* Journal of Microscopy **192**, 90 (1998).

[96] S. Stallinga, *Compact description of substrate-related aberrations in high numerical-aperture optical disk readout,* Applied Optics **44**, 849 (2005).

[97] U. Endesfelder and M. Heilemann, *Art and artifacts in single-molecule localization microscopy: Beyond attractive images.* Nature Methods **11**, 235 (2014).

[98] K. V. Morris and J. S. Mattick, *The rise of regulatory RNA,* Nature reviews. Genetics **15**, 423 (2014).

[99] J. E. Sleeman and L. Trinkle-Mulcahy, *Nuclear bodies: New insights into assembly/dynamics and disease relevance,* Current Opinion in Cell Biology **28**, 76 (2014).

2

Fast, Single-Molecule Localization That Achieves Theoretically Minimum Uncertainty

If we betake ourselves to the statistical method, we do so confessing that we are unable to follow the details of each individual case, and expecting that the effects of widespread causes, though very different in each individual, will produce an average result on the whole nation, from a study of which we may estimate the character and propensities of an imaginary being called the Mean Man.

JAMES CLERK MAXWELL

Carlas S. Smith
Nikolai Joseph
Bernd Rieger
Keith A. Lidke

Abstract

We describe an iterative algorithm that converges to the maximum likelihood estimate of the position and intensity of a single fluorophore. Our technique efficiently computes and achieves the Cramér-Rao lower bound, an essential tool for parameter estimation. An implementation of the algorithm on graphics processing unit hardware achieved more than 10^5 combined fits and Cramér-Rao lower bound calculations per second, enabling real-time data analysis for super-resolution imaging and other applications.

2.1. *Introduction*

In many single molecule fluorescence applications, it is often desired to find the position and intensity of a single fluorophore as well as to estimate the accuracy and precision[1] of these parameters. In recent work that uses single-molecule localization to generate super-resolution image [2–6], single emitters are located and on the mosaic of their found positions a two-dimensional Gaussian profile is placed to generate the final super-resolution images. The width of the placed Gaussian blob, σ, is given by the precision of the fluorophore position localization $\sigma = (\sigma_x^2 + \sigma_y^2)^{1/2}$ and in these super-resolution techniques it is therefore necessary to both find the parameters and estimate their precision. Reported values are in the range of 20-70 nm. In the application of super-resolution imaging, it may be required to find the positions of more than 10^6 fluorophores in order to generate one final image of a typical field-of-view of 50×50 μm. In many cases, background rates may vary across the field of view and the fluorophore emission rate of chemically identical fluorophores can vary due to effects such as uneven illumination profile, dipole orientation or different optical path lengths.

In this work, we describe an iterative routine, implemented on a graphics processing unit (GPU) that calculates the Maximum Likelihood Estimate (MLE) for the $x - y - (z)$ position, the photon count of the fluorophore and the background fluorescence rate. We show that our approach achieves the Cramér-Rao Lower Bound (CRLB) over a wide range of parameters. The uncertainties of the fitted parameters are found by calculating their CRLBs [1], and in this sense the estimated σ for building up the super-resolution image is optimal. We provide a software tool (www.diplib.org/add-ons) that only requires an inexpensive graphic card ($100) in order for single molecule fitting speed to be sufficient for real-time data analysis (see discussion in Appendix 2.A.)

Since the speed and precision of single particle localization has long played an important role in single particle tracking as well as in other single molecule biophysical techniques that rely on fluorescent reporters, others have also considered these issues. In 2001, Cheezum et al. [7] compared several algorithms from the literature for finding particle positions, but, without the context of a statistical framework. In 2004, Ober et al. [8] calculated the theoretically, best-possible estimation precision of a fluorophore position by using the well established statistical method of finding the CRLB in an unbiased parameter estimation problem. They considered many of the effects in a real system including background fluorescence, finite camera pixel size, and camera readout noise and they recently

---

[1]Accuracy is a measure of the systematic error or bias and precision is a measure of the statistical error of an estimator [1]. Unfortunately, in the literature these terms are not used consistently.

made a software tool available for estimation [9]. Here we demonstrate a robust, *iterative* routine that finds the particle position, the intensity and the background count rate. Of the above effects, we do not consider camera readout noise since for electron multiplying (EM)CCD cameras, which are generally used for the fast frame rates desired in super-resolution imaging, the readout noise is much less than 1 rms e^- when using large EM gain.

2.2. *Results*

As described further in the Appendix 2.A, the method presented is not restricted to 2D imaging with a symmetric point spread function (PSF), but can be extended to handle super-resolution techniques that encompass astigmatic imaging for z resolution as in Huang et al. [10]. In this case, the z position is also calculated directly (not from intermediate σ_x, σ_y fits) and returned with CRLB based uncertainties. The results of the iterative algorithm compared to CRLB-based theoretical values are shown for a range of background rates and total collected photon counts of the PSF in Fig. 2.1. We show results for $\sigma_{PSF} = 1$ with the size defined in unitless back-projected pixels. The diffraction limit for high NA visible light imaging is $\gtrsim$ 200 nm and $\sigma_{PSF} \gtrsim$ 90 nm [11]. The algorithm both achieves and correctly reports the CRLB uncertainties over a wide range of background and fluorophore intensities. Calculated precision remains within a few percent of the theoretically achievable value even for less than 100 collected photons. We find that under all conditions, when the reported CRLB is less than $\sigma_{PSF}/2$ (here 0.5), the reported CRLB matches the theoretical position, and the routine achieves the CRLB. In typical super-resolution applications this corresponds to $\lesssim$ 50 nm. Addition of significant camera readout noise has effectively the same bad influence on the parameter estimation as a high background. Fortunately, this can be excluded for an EMCCD for the reasons mentioned above. Example images of single fluorophores with intensities and background rates near the $\sigma_{PSF}/2$ value are shown in the Appendix 2.A. The classical approach of solving the position fitting problem via a non-linear least-square optimization is shown in Fig. 2.1 middle. Here we chose a Levenberg-Marquardt (LM) optimization scheme with analytic and computed first derivatives with respect to the optimization parameters. Note that it is common practice to use computed derivatives only. It clearly performs worse in terms of precision than our iterative MLE approach. The reason is mainly the incorrect, i.e. Gaussian, noise model implicitly present in any least-squares based optimization scheme. In the right panel of Fig. 2.1 we compare the predicted uncertainty of the fit by Thompson et al. [12](Eq. 14) with
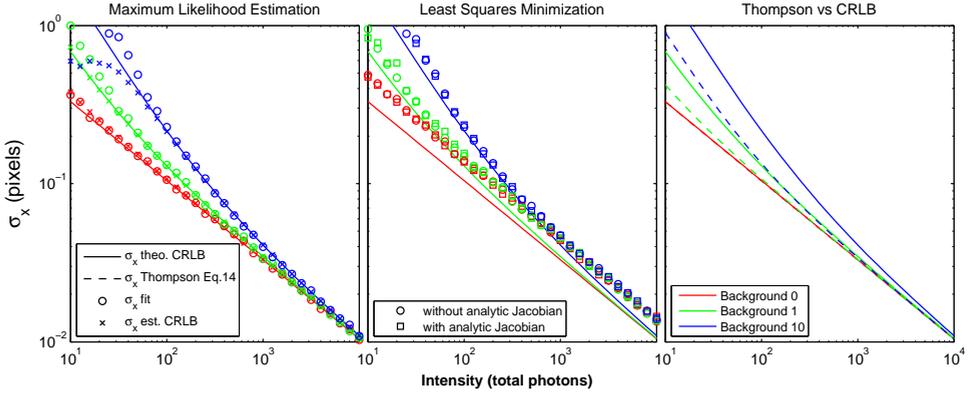
Figure 2.1. Performance comparison on simulated data. Left: The localization precision of the iterative method (circles) is compared to that given by the CRLB (solid lines). Also shown (crosses) are the mean uncertainties reported from CRLB calculations for every image using the found intensity and background rates (constant offset). Calculations are made using a square fitting region of size $2 \times 3\sigma_{PSF} + 1$ and 10 iterations. Center: Fits are performed using non-linear least squares Levenberg-Marquardt with (squares) and without (circles) an analytical Jacobian. Right: The theoretical uncertainty calculated from the four-parameter fit CRLB is compared to the commonly used formula by Thompson [12](Eq. 14) for estimating localization precision. It underestimates the true uncertainty by nearly a factor of two for low light conditions and any background rate.

the theoretical CRLB. Strikingly, they are identical for no background fluorescence, but for any non-zero background and low light conditions the deviation is almost a factor of two. This means that in these cases the suggested uncertainty σ used to constitute the super-resolution image is estimated at nearly half the CLRB based value (overly optimistic). The formula presented by Thompson has the advantage that it can be readily calculated by hand from measurable quantities. However, for a precise estimate, the use of the reported CRLB from our iterative algorithm is to be preferred.

2.3. *Discussion*

Our iterative update scheme is similar to that described by Aguet et al. [13]. We show, however, that a Gaussian approximation for the two-dimensional fluorophore PSF [11] and following localization leads to a compact analytical expression that allows for computationally fast localization without compromising on the localization precision. Our approach achieves the CRLB after a few ($\sim$ 10) iterations. It should be noted that the CRLB predicts the correct precision only when the model function is correct, and the isotropic Gaussian model may not

	iterative MLE method				LM (numeric Jacobian)
Box size [pixel]	AMD phenomII (10^2 fits/second)	Nvidia 8600GTS (10^2 fits/second)	Nvidia 8800GTS (10^2 fits/second)	Nvidia GTX285 (10^2 fits/second)	AMD phenomII (10^2 fits/second)
7×7	31	430	880	2600	15
13×13	9.4	45	100	950	5.2
19×19	4.3	10	22	330	2.3

Table 2.1. Computational performance. CPU and GPU implementations of the iterative MLE and a LM non-linear least-squares fit.

be appropriate when imaging fluorophores with a fixed dipole orientation [14], leading to anisotropic emission. As described in Appendix 2.A, a 'rule-of-thumb' fitting region size of $2 \times 3\sigma_{PSF} + 1$ is used. For z resolution imaging, the Gaussian PSF model is less reliable due to optical aberrations [11].

GPU-based computation has the potential to increase floating point calculation speed by a factor of 10-100 as compared to a modern CPU if the problem is amenable to a parallel processing approach [15]. A generic C-like language interface is available for simplified GPU programming (Nvidia CUDA), and a MATLAB interface has been developed. Operating with a fixed number of iterations complements the GPU's single instruction multiple data strategy (SIMD). A GPU implementation of our iterative method can perform 10^5 combined MLE and CRLB calculations per second of the four parameter model needed to describe the emittance of a fluorophore (Fig. 2.2).

The CPU and GPU performance, characterized by the number of combined position fits and CRLB calculations performed per second, are shown in Tab. 2.1. The slowest GPU tested (Nvidia 8600GTS) outperforms the CPU by more than one order of magnitude, with the Nvidia GTX285 achieving $2.6 \cdot 10^5$ fits per second on a 7×7 fitting box size. We attribute this level of performance gain over the CPU to the fact that this estimation problem is almost ideal for the GPU SIMD architecture. Many iterations are performed on the same data, which are stored in local shared memory, and each thread is independent, eliminating the need for synchronization delays. We also note that all CPU computation only ran on a single thread. In the right column of Tab. 2.1 the performance of a non-linear, least-squares fit (in C-code) on a CPU is shown for reference. It is twice as slow as our iterative algorithm on a CPU. Commonly used MATLAB (The Mathworks, USA) LM optimization only computes about 5 fits/second. Given the readout rate of current high-end (EM)CCD cameras (~ 10 MHz) our GPU implementation performs combined fits and uncertainty estimates at a speed sufficient for real-time analysis (see Appendix 2.A).
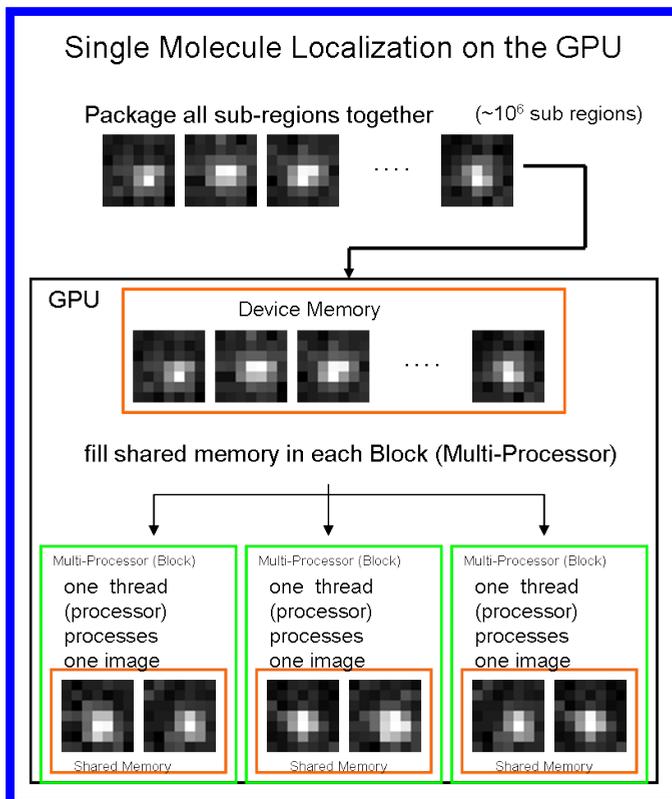
Figure 2.2. Basic concept of single-molecule localization via GPU implementation. Multiple (up to millions) preselected image regions of interest are arranged into a three-dimensional dataset on the CPU and passed into GPU memory. Smaller datasets are arranged and processed in identically sized chunks that fill the multiprocessor shared memory (here shown as just two images, but this can be up to 64 images for 7 x 7 fit regions). Each image is analyzed with the same iterative algorithm. One thread (subprocessor) processes one image. The hundreds of subprocessors available on a GPU result in a speed increase because of massive parallel processing. See Appendix 2.A for more details.

To summarize our findings, we have derived an iterative approach for making a maximum likelihood estimate of the position and intensity of a single fluorophore as well as the background count rate using a two-dimensional Gaussian PSF model and a Poisson noise model. The iterative method achieves the minimal possible estimation uncertainty, as given by the Cramér-Rao Lower Bound, over a wide range of emission and background rates that could be found in single molecule experiments. Implementation of the iterative method on a modern graphics processing unit yields more than 10^5 combined fits and Cramér-Rao Lower Bound calculations per second, greatly facilitating the analysis of large data sets found in single molecule based super-resolution techniques and is suitable for real-time data analysis.

References

[1] A. van den Bos, *Parameter Estimation for Scientists and Engineers* (Wiley & Sons, New Jersey, 2007).

[2] 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**, 1642 (2006).

[3] M. Bates, B. Huang, G. T. Dempsey, and X. Zhuang, *Multicolor super-resolution imaging with photo-switchable fluorescent probes,* Science **317**, 1749 (2007).

[4] A. Egner, C. Geisler, C. Von Middendorff, H. Bock, D. Wenzel, R. Medda, M. Andresen, A. C. Stiel, S. Jakobs, C. Eggeling, A. Schonle, and S. W. Hell, *Fluorescence nanoscopy in whole cells by asynchronous localization of photoswitching emitters,* Biophysical Journal **93**, 3285 (2007).

[5] S. T. Hess, T. P. K. Girirajan, and M. D. Mason, *Ultra-high resolution imaging by fluorescence photoactivation localization microscopy,* Biophysical Journal **91**, 4258 (2006).

[6] K. A. Lidke, B. Rieger, T. M. Jovin, and R. Heintzmann, *Superresolution by localization of quantum dots using blinking statis-tics,* Optics Express **13**, 7052 (2005).

[7] M. K. Cheezum, W. F. Walker, and W. H. Guilford, *Quantitative comparison of algorithms for tracking single fluorescent particles,* Biophysical Journal **81**, 2378 (2001).

[8] R. J. Ober, S. Ram, and E. S. Ward, *Localization accuracy in single-molecule microscopy,* Biophysical Journal **86**, 1185 (2004).

[9] R. Ober and E. Ward, *Estimation tool,* (2009), university of Texas, Southwestern http://www4.utsouthwestern.edu/wardlab.

[10] B. Huang, W. Wang, M. Bates, and X. Zhuang, *Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy,* Science **319**, 810 (2008).

[11] B. Zhang, J. Zerubia, and J. C. Olivo-Marin, *Gaussian approximations of fluorescence microscope point-spread function models,* Applied Optics **46**, 1819 (2007).

[12] R. E. Thompson, D. R. Larson, and W. W. Webb, *Precise nanometer localization analysis for individual fluorescent probes,* Biophysical Journal **82**, 2775 (2002).

[13] F. Aguet, D. Van De Ville, and M. Unser, *A maximum-likelihood formalism for sub-resolution axial localization of fluorescent nanoparticles,* Optics Express **13**, 10503 (2005).

[14] J. Enderlein, E. Toprak, and P. Selvin, *Polariztion effect on position accuracy of fluorophore localization,* Optics Express **14**, 8112 (2006).

[15] NVIDIA, *Compute unified device architecture (CUDA),* (2007), http://www.nvidia.com.

2

Appendix

2.A. *Theory of Image Formation and Parameter Estimation*

The image formation model

In the paraxial limit, the microscope point spread function (PSF) can be taken as the Airy pattern $\text{PSF}(r) = \frac{J_1(\alpha r)^2}{(\alpha r)^2}$, where $\alpha = \frac{2\pi\text{NA}}{\lambda}$ and J_1 is the Bessel function of first kind. However, in fluorescence microscopy, objective lenses with numerical apertures (NAs) of 1.2 or greater are most often used when imaging single fluorescent molecules in order to collect as many emitted photons as possible; the collection efficiency is $\propto \text{NA}^2$. A proper calculation of PSFs for these high NA objectives must include vectorial effects, aberrations in the optical system and apodization in the objective lens [1, 2]. In lieu of such a detailed measurement and description of the PSF, often a two-dimensional Gaussian shape is used as a compact and good expression of the PSF [3]. We use the two-dimensional Gaussian approximation of the PSF to greatly simplify the proceeding analysis. This also holds for extension to 3D confocal laser scanning PSFs [3]. The simplified form of the PSF is then

$$\text{PSF}(x, y) = \frac{1}{2\pi\sigma^2} e^{\frac{-(x-\theta_x)^2 - (y-\theta_y)^2}{2\sigma^2}}, \tag{2.1}$$

where $\theta_{x,y}$ is the position of the emitter. PSF approximations also often neglect the finite size of the detector pixels. In this work, the imaging model, denoted by $\mu_k(x, y)$, will always include the integration over finite pixels which is given by:

$$\mu_k(x, y) = \theta_{I_0} \int_{A_k} \text{PSF}(u, v)\, \mathrm{d}u\mathrm{d}v + \theta_{bg}, \tag{2.2}$$

where $\mu_k(x, y)$ denotes the expected value in the kth pixel, θ_{bg} the expected background count, θ_{I_0} the expected photon count and the integral is over the finite area A_k of the kth pixel, which is centered at (x, y). The Gaussian approximation allows the results to be expressed in terms of error functions (which simplify the implementation as they are supported in CUDA [4]). If we assume uniform pixels

with unit size, the imaging model is given as:

$$\mu_k(x, y) = \theta_{I_0} \Delta E_x(x, y) \Delta E_y(x, y) + \theta_{bg}, \tag{2.3}$$

where the expressions

$$\Delta E_x(x, y) \equiv \frac{1}{2} \text{erf}\left(\frac{x - \theta_x + \frac{1}{2}}{\sqrt{2}\sigma} \right) - \frac{1}{2} \text{erf}\left(\frac{x - \theta_x - \frac{1}{2}}{\sqrt{2}\sigma} \right), \tag{2.4a}$$

$$\Delta E_y(x, y) \equiv \frac{1}{2} \text{erf}\left(\frac{y - \theta_y + \frac{1}{2}}{\sqrt{2}\sigma} \right) - \frac{1}{2} \text{erf}\left(\frac{y - \theta_y - \frac{1}{2}}{\sqrt{2}\sigma} \right), \tag{2.4b}$$

are used to shorten notation. The Gaussian standard deviation, σ, is determined in practice by fitting to images of small beads or single molecules under the operating conditions of the microscope. It can also be computed directly from a theoretical PSF and its FHWM [3]. Eq. 2.3 will be used as the imaging model for the parameter estimation in the following.

Cramér Rao Lower Bound

The Cramér Rao Lower Bound (CRLB) is the limiting lower bound of the variance for any unbiased parameter estimator [5]. The general expression for the CRLB is given by the inverse of the Fisher information matrix

$$var(\hat{\theta}) \geq I(\theta)^{-1}, \tag{2.5}$$

where $var(\hat{\theta})$ is the variance of an estimator and $I(\theta)$ is the Fisher information matrix. The equal sign is the minimum of that estimation and is referred to as the CRLB.

The Fisher information matrix for the imaging model μ

The elements of the Fisher information matrix are given by the expectation of the log-likelihood:

$$I_{ij}(\theta) = E\left[\frac{\partial \ln(L(\vec{x}|\theta))}{\partial \theta_i} \frac{\partial \ln(L(\vec{x}|\theta))}{\partial \theta_j} \right], \tag{2.6}$$

where θ is the set of parameters being estimated $\theta = [\theta_1 ... \theta_N]$ and $L(\vec{x}|\theta)$ is the likelihood of the data set $\vec{x}$ given the model generated by θ. Here the parameters

are $\theta = [\theta_x, \theta_y, \theta_{I_0}, \theta_{bg}]$. For a Poisson process the likelihood is given by

$$L(\vec{x}|\theta) = \prod_k \frac{\mu_k(x,y)^{x_k} e^{-\mu_k(x,y)}}{x_k!}. \tag{2.7}$$

Using Eq. 2.6, Eq. 2.7 and the Stirling approximation ($\ln n! \approx n \ln n - n$ for large n) it is straightforward to show that

$$I_{ij}(\theta) = E\left[\sum_k (x_k - \mu_k(x,y))^2 \frac{1}{\mu_k(x,y)^2} \frac{\partial \mu_k(x,y)}{\partial \theta_i} \frac{\partial \mu_k(x,y)}{\partial \theta_j}\right]. \tag{2.8}$$

Using the fact that $(x_k - \mu_k(x,y))^2$ is the variance and equal to the expected value for a Poisson process, we obtain

$$I_{ij}(\theta) = \sum_k \frac{1}{\mu_k(x,y)} \frac{\partial \mu_k(x,y)}{\partial \theta_i} \frac{\partial \mu_k(x,y)}{\partial \theta_j}. \tag{2.9}$$

The partial derivatives with respect to the parameters are given by

$$\frac{\partial \mu_k(x,y)}{\partial \theta_x} = -\frac{\theta_{I_0}}{\sigma^2} \int_{A_k} (\theta_x - u) \, \text{PSF}(u,v) \, \mathrm{d}u \mathrm{d}v \tag{2.10a}$$

$$\frac{\partial \mu_k(x,y)}{\partial \theta_y} = -\frac{\theta_{I_0}}{\sigma^2} \int_{A_k} (\theta_y - v) \, \text{PSF}(u,v) \, \mathrm{d}u \mathrm{d}v \tag{2.10b}$$

$$\frac{\partial \mu_k(x,y)}{\partial \theta_{I_0}} = \int_{A_k} \text{PSF}(u,v) \, \mathrm{d}u \mathrm{d}v \tag{2.10c}$$

$$\frac{\partial \mu_k(x,y)}{\partial \theta_{bg}} = 1, \tag{2.10d}$$

which upon performing the integration become

$$\frac{\partial \mu_k(x,y)}{\partial \theta_x} = \frac{\theta_{I_0}}{\sqrt{2\pi}\sigma} \left(e^{\frac{-\left(x_k - \theta_x - \frac{1}{2}\right)^2}{2\sigma^2}} - e^{\frac{-\left(x_k - \theta_x + \frac{1}{2}\right)^2}{2\sigma^2}}\right) \Delta \mathrm{E}_y(x,y) \tag{2.11a}$$

$$\frac{\partial \mu_k(x,y)}{\partial \theta_y} = \frac{\theta_{I_0}}{\sqrt{2\pi}\sigma} \left(e^{\frac{-\left(y_k - \theta_y - \frac{1}{2}\right)^2}{2\sigma^2}} - e^{\frac{-\left(y_k - \theta_y + \frac{1}{2}\right)^2}{2\sigma^2}}\right) \Delta \mathrm{E}_x(x,y) \tag{2.11b}$$

$$\frac{\partial \mu_k(x,y)}{\partial \theta_{I_0}} = \Delta \mathrm{E}_x(x,y) \Delta \mathrm{E}_y(x,y) \tag{2.11c}$$

$$\frac{\partial \mu_k(x,y)}{\partial \theta_{bg}} = 1. \tag{2.11d}$$

Note that Eq. 2.9 is general, whereas the Gaussian PSF model allows the simple expressions in Eqs. 2.10 and 2.11. Simultaneous fitting of the position, intensity and background leads to off-diagonal elements in $I(\theta)$. The CRLBs for the estimation of θ are equal to the diagonal elements of the matrix inversion of $I(\theta)$, i.e. $var(\theta_i) = \left(I^{-1}\right)_{ii}$.

Iterative method

We employ essentially a Newton-Raphson method [6] to find the parameters θ that maximize $\ln(L(\vec{x}|\theta))$. This is equal to the maximum likelihood estimate of the parameters $\theta_{ML} = \underset{\theta}{\arg\max} \, L(\vec{x}|\theta)$. The derivatives required for the iterative updates are calculated in a straightforward manner from Eq. 2.7

$$\frac{\partial \ln(L(\vec{x}|\theta))}{\partial \theta_i} = \sum_k \frac{\partial \mu_k(x,y)}{\partial \theta_i} \left(\frac{x_k}{\mu_k(x,y)} - 1 \right). \tag{2.12}$$

Given an adequate guess of starting parameters, the parameters are updated according to:

$$\theta_i \rightarrow \theta_i + \left[\sum_k \frac{\partial \mu_k(x,y)}{\partial \theta_i} \left(\frac{x_k}{\mu_k(x,y)} - 1 \right) \right] \times$$
$$\left[\sum_k \frac{\partial^2 \mu_k(x,y)}{\partial \theta_i^2} \left(\frac{x_k}{\mu_k(x,y)} - 1 \right) - \frac{\partial \mu_k(x,y)}{\partial \theta_i}^2 \frac{x_k}{\mu_k(x,y)^2} \right]^{-1}. \tag{2.13}$$

The first derivatives are given in the previous section, (see Eqs 2.11) and the second derivatives are given by

$$\frac{\partial^2 \mu_k(x,y)}{\partial \theta_x^2} = \tag{2.14a}$$
$$\frac{\theta_{I_0}}{\sqrt{2\pi}\sigma^3} \left(\left(x_k - \theta_x - \frac{1}{2} \right) e^{\frac{-\left(x_k - \theta_x - \frac{1}{2}\right)^2}{2\sigma^2}} - \left(x_k - \theta_x + \frac{1}{2} \right) e^{\frac{-\left(x_k - \theta_x + \frac{1}{2}\right)^2}{2\sigma^2}} \right) \Delta E_y(x,y),$$

$$\frac{\partial^2 \mu_k(x,y)}{\partial \theta_y^2} = \tag{2.14b}$$
$$\frac{\theta_{I_0}}{\sqrt{2\pi}\sigma^3} \left(\left(y_k - \theta_y - \frac{1}{2} \right) e^{\frac{-\left(y_k - \theta_y - \frac{1}{2}\right)^2}{2\sigma^2}} - \left(y_k - \theta_y + \frac{1}{2} \right) e^{\frac{-\left(y_k - \theta_y + \frac{1}{2}\right)^2}{2\sigma^2}} \right) \Delta E_x(x,y),$$

$$\frac{\partial^2 \mu_k(x,y)}{\partial \theta_{\theta_{I_0}}^2} = \frac{\partial^2 \mu_k(x,y)}{\partial \theta_{bg}^2} = 0. \tag{2.14c}$$

36

Extension to 3D astigmatic imaging

Following Holtzer [7] and using σ_0 as usual the PSF near focus can be described by a 2D Gaussian with a z-dependent standard deviation

$$\sigma(z) = \sigma_0 \sqrt{1 + \frac{z^2}{d^2} + A\frac{z^3}{d^2} + B\frac{z^4}{d^2}}, \tag{2.15}$$

where σ_0 is the in focus standard deviation, d is the depth of focus (a constant for a particular objective), A and B are empirical constants ($B \approx 0$ for our experiments, see Results). Introducing an elliptical lens in the beam path splits the focal plane into two perpendicular focal planes at different depth giving an asymmetric PSF. The form of the PSF on the detector is then approximated by

$$\text{PSF}(x, y) = \frac{1}{2\pi\sigma_x(\theta_z)\sigma_y(\theta_z)} e^{-\frac{(x-\theta_x)^2}{2\sigma_x(\theta_z)^2} - \frac{(y-\theta_y)^2}{2\sigma_y(\theta_z)^2}}, \tag{2.16}$$

where $\theta_{x,y,z}$ is the position of the emitter. Half way between the two focal planes we have (assuming Gaussian optics at the disk of least confusion) $\sigma_x = \sigma_y$ and $\sigma_x, \sigma_y \geq \sigma_0$; here we define $z = 0$. Focal planes for each direction are above and below $z = 0$. Assuming the x-direction is focused above and using Eq. 2.15, $z \to z - \gamma$ for the x-direction, $z \to z + \gamma$ for the y-direction

$$\sigma_x(z) = \sigma_{0x}\sqrt{1 + \frac{(z-\gamma)^2}{d^2} + A_x\frac{(z-\gamma)^3}{d^2} + B_x\frac{(z-\gamma)^4}{d^2}}, \tag{2.17a}$$

$$\sigma_y(z) = \sigma_{0y}\sqrt{1 + \frac{(z+\gamma)^2}{d^2} + A_y\frac{(z+\gamma)^3}{d^2} + B_y\frac{(z+\gamma)^4}{d^2}}. \tag{2.17b}$$

The x, y-directions have the same form so we continue only with y. Eq. 2.13 requires the first and second derivatives $\frac{\partial\mu(x_k,y_k)}{\partial\theta_z}$ and $\frac{\partial^2\mu(x_k,y_k)}{\partial\theta_z^2}$ to fit the z position. These are calculated from

$$\frac{\partial\mu(x_k, y_k)}{\partial\theta_z} = \frac{\partial\mu}{\partial\sigma_x}\frac{\partial\sigma_x}{\partial\theta_z} + \frac{\partial\mu}{\partial\sigma_y}\frac{\partial\sigma_y}{\partial\theta_z}, \tag{2.18a}$$

$$\frac{\partial^2\mu(x_k, y_k)}{\partial\theta_z^2} = \frac{\partial^2\mu}{\partial\sigma_x^2}\left(\frac{\partial\sigma_x}{\partial\theta_z}\right)^2 + \frac{\partial\mu}{\partial\sigma_x}\frac{\partial^2\sigma_x}{\partial\theta_z^2}$$
$$+ \frac{\partial^2\mu}{\partial\sigma_y^2}\left(\frac{\partial\sigma_y}{\partial\theta_z}\right)^2 + \frac{\partial\mu}{\partial\sigma_y}\frac{\partial^2\sigma_y}{\partial\theta_z^2}. \tag{2.18b}$$

The first and second derivatives σ_y to θ_z are given by

$$\frac{\partial \sigma_y}{\partial \theta_z} = \frac{\sigma_0 \left(\frac{2z}{d^2} + A\frac{3z^2}{d^2} + B\frac{4z^3}{d^2} \right)}{2\sqrt{1 + \frac{z^2}{d^2} + A\frac{z^3}{d^2} + B\frac{z^4}{d^2}}},$$ (2.19a)

$$\frac{\partial^2 \sigma_y}{\partial \theta_z^2} = \frac{\sigma_0 \left(\frac{2}{d^2} + A\frac{6z}{d^2} + B\frac{12z^2}{d^2} \right)}{2\sqrt{1 + \frac{z^2}{d^2} + A\frac{z^3}{d^2} + B\frac{z^4}{d^2}}} - \frac{\sigma_0 (\frac{2z}{d^2} + A\frac{3z^2}{d^2} + B\frac{4z^3}{d^2})^2}{4(1 + \frac{z^2}{d^2} + A\frac{z^3}{d^2} + B\frac{z^4}{d^2})^{\frac{3}{2}}}.$$ (2.19b)

It does not matter for the result of the differentiation if the substitution $z \to z \pm \gamma$ is done before or after, so $\frac{\partial^2 \sigma_{x,y}}{\partial \theta_z^2}$ is obtained by substituting $z \to z \pm \gamma$ in the derivatives Eqs. 2.19b. Using the expressions

$$G_x^{nm}(x_k) \equiv$$ (2.20a)

$$\frac{1}{\sqrt{2\pi}\sigma_x(\theta_z)^n} \left(\left(x_k - \theta_x - \frac{1}{2}\right)^m e^{\frac{-\left(x_k - \theta_x - \frac{1}{2}\right)^2}{2\sigma_x(\theta_z)^2}} - \left(x_k - \theta_y + \frac{1}{2}\right)^m e^{\frac{-\left(x_k - \theta_y + \frac{1}{2}\right)^2}{2\sigma_x(\theta_z)^2}} \right)$$

$$G_y^{nm}(y_k) \equiv$$ (2.20b)

$$\frac{1}{\sqrt{2\pi}\sigma_y(\theta_z)^n} \left(\left(y_k - \theta_y - \frac{1}{2}\right)^m e^{\frac{-\left(y_k - \theta_y - \frac{1}{2}\right)^2}{2\sigma_y(\theta_z)^2}} - \left(y_k - \theta_y + \frac{1}{2}\right)^m e^{\frac{-\left(y_k - \theta_y + \frac{1}{2}\right)^2}{2\sigma_y(\theta_z)^2}} \right)$$

to shorten notation, we obtain for the derivatives of μ to σ_y

$$\frac{\partial \mu(x_k, y_k)}{\partial \sigma_y} = \theta_{I_0} \Delta E_x(x_k) G_y^{21}(y_k),$$ (2.21a)

$$\frac{\partial^2 \mu(x_k, y_k)}{\partial \sigma_y^2} = \theta_{I_0} \Delta E_x(x_k) \left(G_y^{53}(y_k) - 2G_y^{31}(y_k) \right).$$ (2.21b)

2.B. *Methods*

GPU implementation

The iterative method to solve the MLE problem described above is implemented on a GPU using NVIDIA's compute unified device architecture (CUDA [4]), a C based programming language that makes it possible to readily program parallelized algorithms that are executed on a GPU. High-end gaming and computing GPU's have a (card dependent) large chunk of global device memory, usually several hundred megabytes. Execution is performed on a number (card dependent) of multi-processors. Multi-processors contain eight sub-processors and have 16 kilobytes local memory that is shared between sub-processors. Access to local
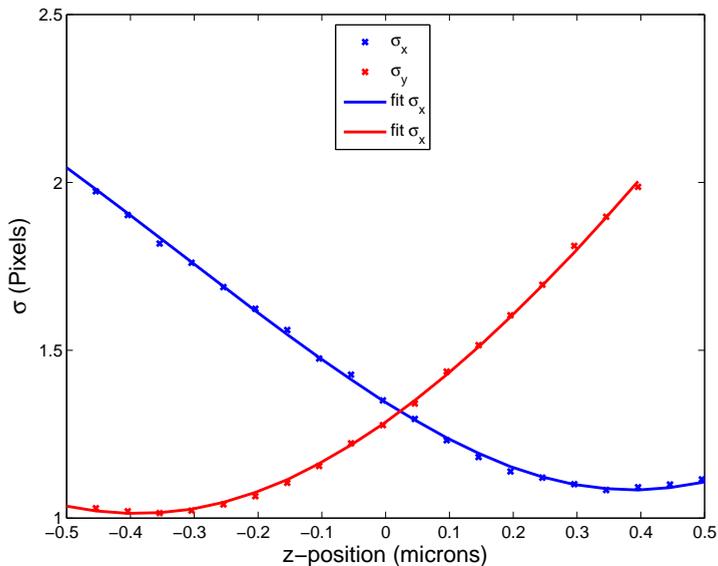
Figure 2.B.1. Fitting $\sigma_x(z)$ and $\sigma_y(z)$ to the astigmatic model of the PSF. Beads were used as calibration markers which were moved in steps of 50 nm by a piezo stage.
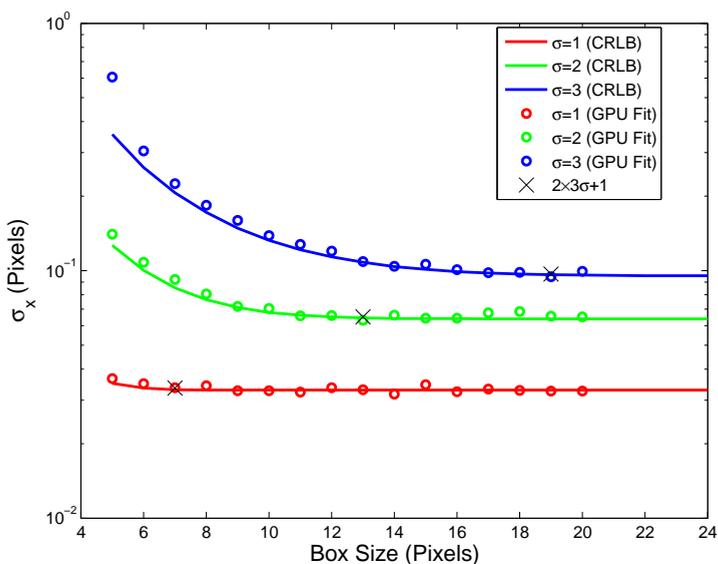
Figure 2.B.2. Localization Accuracy vs Fit Region Size. The best localization accuracy is calculated using the CRLB and compared to the standard deviation of x-position error in simulated data stacks for $I_0 = 1000$ and $bg = 0$. The improvement of localization accuracy diminishes beyond a linear box size of $(2 \times 3\sigma + 1)$.
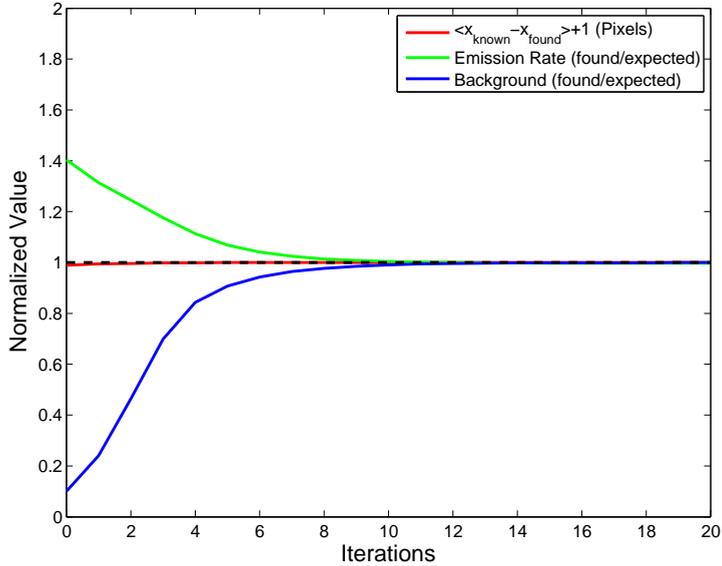
Figure 2.B.3. Convergence of the Iterative Method. The convergence of position (red) intensity (green) and background rate (blue) to known, correct values are shown. Calculation were made using $\sigma_{PSF} = 3$ and a square fitting region of linear size and $2 \times 3\sigma_{PSF} + 1 = 19$, which represents the slowest convergence of any combination of σ_{PSF} and box size used.
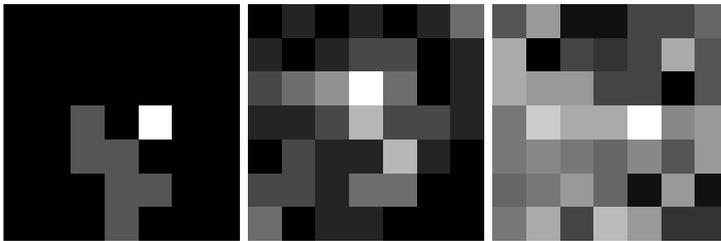
Figure 2.B.4. Example images of simulated single molecules with emission and background rates (in photons) that corresponds with the $\sigma_{PSF}/2$ border. Here the iterative method is still able to both estimate the CRLB and find the position with CRLB accuracy. We show a 7×7 area with $\sigma_{PSF} = 1$. From left to right: background rates 0, 1 and 10. The shown photon counts are 10, 30 and 50.
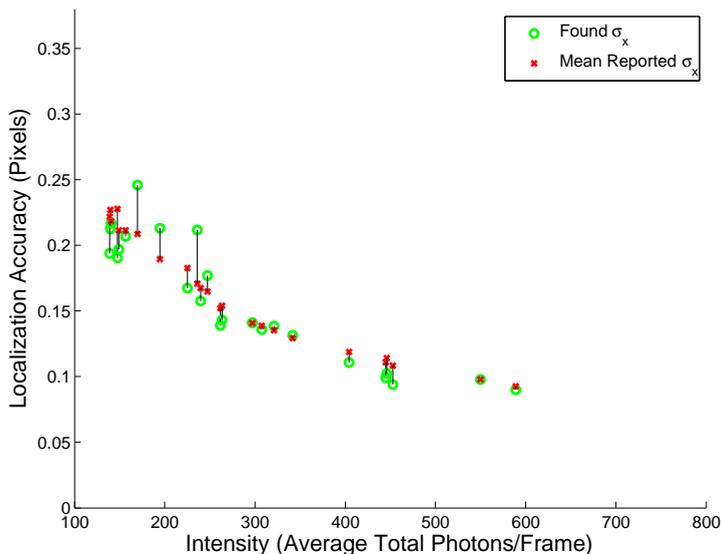
Figure 2.B.5. Analysis of Single Molecule Experiments. The standard deviation of the x-positions found in repeated measurements of single, immobilized Cy5 molecules (green) are compared with the average accuracy calculated frame by frame using the CRLB (red). Found background rates (not shown) vary between single molecules giving a range of possible localization accuracy values for each intensity.
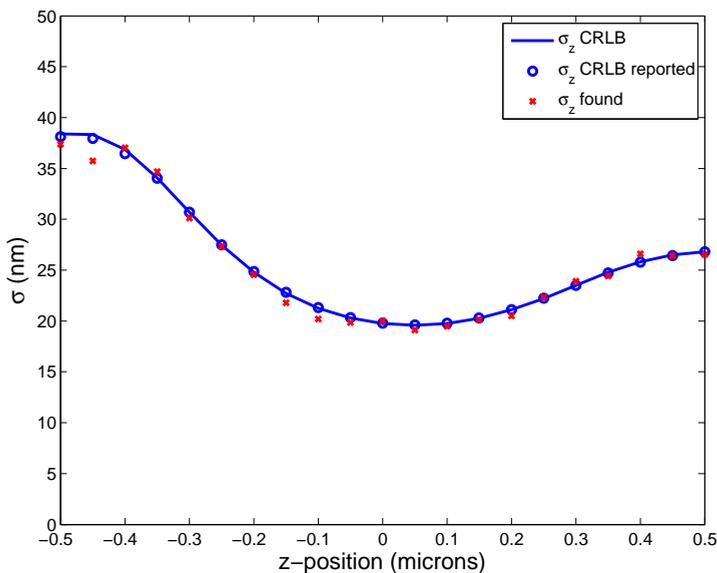
Figure 2.B.6. Simultaneous fits to x, y, z, the emission and background rates synthetic data. Shown is only the precision in z. The reported values of the CRLB of the algorithm are shown as circles whereas the found standard deviation is indicated with crosses.
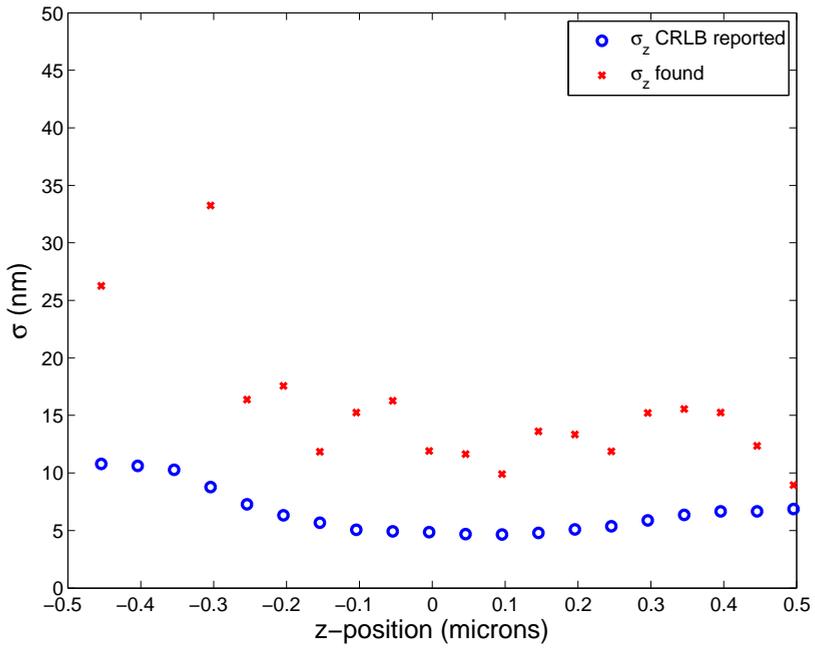
Figure 2.B.7. The standard deviation of found z-positions. Found standard deviations do not match that reported from CRLB calculations because of aberrations present in experimental images that are not accounted for in the imaging model.
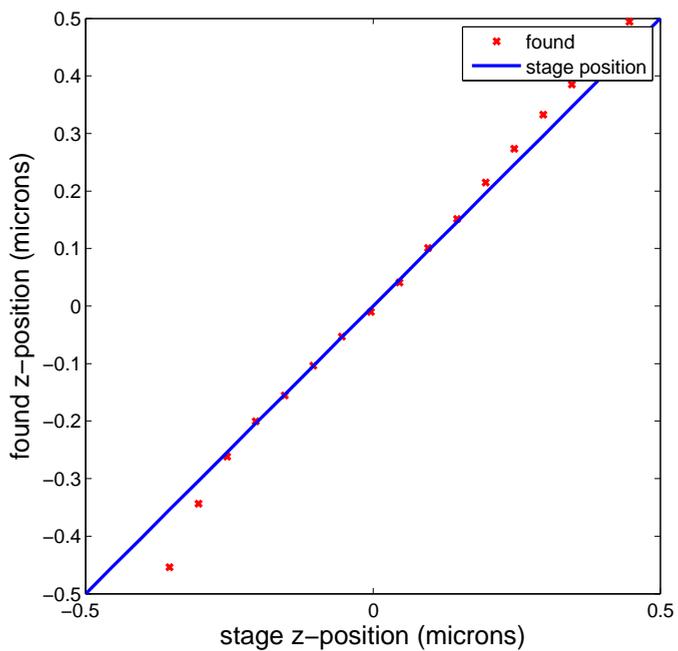
Figure 2.B.8. Experimental bead data. Estimated z-positions versus piezo stage z position. The line has slope one.

shared memory is fast, whereas access to the device global memory incurs a large performance penalty, and should be avoided when possible. The programming model follows the GPU architecture in that parallelized execution is performed by breaking down computations into 'blocks' and 'threads'. Each block executes a set of threads using one multi-processor. Typically, performance is optimal when multiples of 32 threads (called a 'warp') are scheduled and executed using the 8 sub-processors [4]. However, we found for all fit box sizes, e.g. 7×7 pixels, the maximum fits/second occurred when the maximum fits/block were used, which is limited by the available 16KB shared memory per block. This is likely due to compiler optimization and the fact that each fit is independent of all others (no thread synchronization is required).

We map our iterative algorithm on this programming model in the following way. A data stack consisting of identically sized subregions, which are centered single emitter images, are input to the function. This data is copied from host (CPU) memory to device (GPU) main memory. This data set is divided in to blocks, which consist of the largest number of images that can fit into shared memory. The execution of a block begins by copying the data sub-set into local shared memory. Each thread then calculates a complete fit and CRLB calculation for one image. The Fisher information matrix is calculated using Eqs. 2.9 and 2.11 and the CRLB is calculated using the analytical expression for the inverse of a 4×4 matrix. The CPU performance was measured by replacing the block/thread architecture with nested loops which call the same sub-function, and was compiled using Microsoft Visual Studio Express 2008. Images were loaded in MATLAB as arrays (The Mathworks, USA) and the C-Code and the CUDA GPU code where called via mex-files. The CPU was a AMD Phenom II X3 720 @2.8 GHz and only one core was used for the C-code.

Synthetic data generation and analysis

The CUDA routine described above operates on a data set consisting of identically sized images that contain images of (potential) single molecules. For the analysis of synthetic data, a stack is generated using the same finite pixel approximation as described above including background. The center coordinate of each simulated emitter is randomly shifted within the central pixel to prevent a biased result. After the generation of the data stack, the images are corrupted with Poisson noise. This data stack is analyzed by the routine, which returns the estimated position, intensity, background and CRLB calculation for each 2D image in the stack. No camera readout noise is added. Camera readout noise for electron multiplying

(EM) CCD cameras, which are generally used for the fast frame rates desired in super-resolution imaging, is much less than 1 rms e^- with large a EM gain.

Levenberg-Marquardt non-linear least-squares fitting

The standard way of fitting a Gaussian to data is via least-squares fitting. We used 1) an existing implementation of MATLAB via the optimization toolbox `lsqcurefit` where we enabled the Levenberg-Marquardt option as minimalization scheme and 2) an implementation from Numerical Recipes in C [6]. We ran the test on fits with computed Jacobian only and on fits where we supplied analytic first derivatives. We used the following limits in the stopping criterion: tolerance on the parameters 10^{-4}, tolerance on the function 10^{-15} and maximal 10^5 function evaluations. The stopping criterion was in all cases determined by the accuracy put on the parameters. The MATLAB routine is a lot slower (about two orders) than the C implementation altough it makes automatically use of all available cores on the CPU if multi-threading is enabled.

Single molecule imaging

Single molecule imaging experiments were performed in an epi-fluorescence microscope setup consisting of an inverted microscope (IX71, Olympus America Inc.), 1.45 NA objective (U-APO 150x NA 1.45, Olympus America Inc.), 635 nm diode laser (Radius 635, Coherent Inc.), and an electron multiplying CCD camera (Luca DL6581-TIL, Andor Technologies PLC.). The pixel size is 10 μm. The epi-fluorescence filter setup consisted of a dichroic mirror (650 nm, Semrock) and an emission filter (692/40, Semrock). Individual Cy5 molecules were immobilized on an amino-silane ((3-Aminopropyl)triethoxysilane, Sigma-Aldrich) treated 8-well Labtek chambered cover slips (Nalge Nunc International) via an NHS-ester linkage attached to the Cy5 (Cy5 Mono-reactive dye pack, GE Healthcare). An oxygen scavenging system [8] was used to extend fluorophore lifetimes and quench fluorophore triplet states. This was necessary to perform repeated measurements of the same single emitter for several frames while acquiring sufficient photons in order to address localization accuracy. This is not necessary in a dedicated experiment.

Data was recorded by the CCD camera at either 10 or 20 frames per second. All data was post-processed by 1) subtracting a pixel-dependent camera-offset, which was created by averaging 300 dark frames, and 2) multiplying the resulting image by a gain factor to restore correct Poisson statistics, as done in Lidke et al. [9]. Single molecules candidates were identified in each time frame as regions where the 2-D Gaussian filtered image ($\sigma = \sigma_{PSF}$) was greater than one standard

deviation of this image. Note that due to the speed of the GPU implementation, a simple but fast method for identifying candidates is preferred as well as one that errs on the side of including regions that do not contain single molecules. Square regions of a specified number of pixels that included all identified regions in the time series were collected into one stack and input to the GPU routine. The resulting found coordinates were used in building trajectories only if they passed the following criterion: 1) Reported localization accuracy was less than one fifth σ_{PSF} in each dimension, and 2) $1/N\sum_k \ln(L(x_k|\theta)) > -1$, in which N is the number of pixels and $L(x_k|\theta)$ is the likelihood of the data x_k given the parameterized model θ. This criterion essentially performs a shape test and can rule out obvious cases of two proximate emitters. The remaining coordinates were connected into 'trajectories' using an existing single particle tracking routine [10]. Only 'trajectories' that showed little triplet state blinking were used in the final analysis, with a cut-off criterion that $var(I(t)) < 2 \times mean(I(t))$ where $I(t)$ is the sum over all pixels in the analyzed region in frame t. 'Trajectories' were adjusted to compensate for microscope stage drift by subtracting a linear regression line from each single particle trajectory.

The width of the PSF used in the fitting routine was found by minimizing the mean square error between the finite pixel model and the summed projection over a 100 frame time series. The σ for the routine is the average found from analyzing the summed projection of 5 different single emitters.

Astigmatic imaging

The 3D astigmatic imaging was calibrated by imaging 100 nm red ($\lambda = 690$ nm) beads (FluoSphere, Invitrogen) bound to a the bottom of an 8 well coverslip chamber (Lab-Tek II, Nunc). The filter setup used was the same as that used for single molecule Cy5 imaging. We imaged using a 60x 1.2 NA water objective. A 500 mm focal length cylindrical lens was inserted in the emission beam path just after the first lens of a two color beam splitter (OptoSplit II, Cairn Research, UK). A piezo-electric z-stage (Nano-LPS, Mad City Labs) translated the focal plane in steps of 50 nm from -0.5 μm to 0.5 μm. At each focal plane, 20 images of a bead were captured. The fit box size used is again calculated by $2 \times 3\sigma_{PSF} + 1$, but here σ_{PSF} is taken as the maximum value of either σ_{PSFx} or σ_{PSFy}; in this case giving a fit box size of 13×13 pixels.

After gain and background correction, the sum of all images from each focal plane were used to find $\sigma_x(z)$ and $\sigma_y(z)$, which were then fit to model of eq. 2.15. The fit is shown in Fig. 2.B.1. From these calibration the following values for the parameters of eq. 2.15 are found $\sigma_{PSFx} = 1.08, \sigma_{PSFx} = 1.01, Ax = -.0708, Ay =$

0.164, $Bx = -.073$, $By = .0417$, $d = 0.531$, $\gamma = 0.389$. The depth of field for a high NA imaging system is given by

$DOF = \lambda / \left(4n \left(1 - \sqrt{1 - NA^2/n^2} \right) \right) \approx 230$ nm [11] but here is included as a fit parameter.

Real-time data analysis

Initially, the bottleneck for the build-up of a super resolution image was the switching cycle speed for activating only a very small number of particles per image, which resulted in imaging times of many hours [12]. Subsequent speedups were achieved by optimizing the fluorophores for activation based super resolution, protocol improvements or reduction in the number of time frames [13–15].

Small [16] investigated the fundamental relationships between error and acquisition rate (number of activation cycles) in a theoretical study. His findings are relevant to specific chosen or given activation probability. However, to asses the required speed for real-time data analysis, we address the problem somewhat differently. We use the field-of-view V, the size of the footprint of the PSF P, the frame rate F and the fill factor f of the single emitters distribution on the field-of-view. The required fits/second for real-time data analysis are then

$$\text{fits/second} = \frac{VFf}{P}. \tag{2.22}$$

We consider two common cases of emCCD cameras for the maximal fill factor of $f = 1$ and a PSF of $P = 7 \times 7$ pixels: i) $V = 128 \times 128$ pixels, $F = 500$ frames/second and ii) $V = 512 \times 512$ pixels, $F = 30$ frames/second. For the first case $\sim 1.7 \cdot 10^5$ fits/second are required, and for the second $\sim 1.6 \cdot 10^5$ fits/second respectively; these values are about equal as the total readout rate (~ 10 MHz) is the limiting factor and is about equal. The PSF footprint can vary for different physical CCD camera pixel sizes and magnifications. In any case, the fastest GPU ($2.6 \cdot 10^5$ fits/second) tested already full fills this requirement for current fluorophores and CCD cameras! Of course, a fill factor of $f = 1$ is optimistic; more realistic values are $0.1 - 0.5$, but dependent on the experimental conditions and can be chosen according to Small [16]. This means that also the slower (and cheaper) cards already are sufficient in current practice for real-time fitting of positions. The significance of the GPU fitting in the context of the entire process of segmentation (identifying regions of interest for single molecule fits), organizing ROIs, single molecule fitting, and reconstruction, is shown in Tab. S2.B.1. Segmentation and reconstruction are performed as described in [17] with the segmentation performed

Processor	Total Time	Segmentation	ROI collection	Fitting	Reconstruction
	10^4 frames, 10^5 localizations				
GPU	8.8 s	90 %	1 %	8.5 %	0.5 %
CPU	41 s	19.4 %	0.2 %	80.3 %	0.1 %
	10^4 frames, 10^6 localizations				
GPU	14 s	57.8 %	2.7 %	38.9 %	0.6 %
CPU	300 s	2.8 %	0.1 %	97 %	0.1 %

Table 2.B.1. Comparison of total processing time for 10,000 frames of 128x128 pixels. Fitting corresponds to σ_{PSF}=1 and 7×7 pixel fitting areas.

on the GPU. The results show that with even with 10^6 total fits, corresponding to 100 fits per frame, the overall processing could exceed the maximum possible frame rate of 500 Hz of available EMCCDs.

2.C. *Results*

Performance on synthetic data sets

Fig. 2.B.2 shows the CRLB determined best accuracy as a function of a linear box size as well as the results from GPU fits. The results suggest a 'rule of thumb' fitting region size of $2 \times 3\sigma + 1$, which gives near optimal results while keeping computational time down and reducing the probability of including nearby emitters in the fitted region. All further results are shown with fits using this box size. This limit is a direct consequence of the fact that 99% of the volume of a Gaussian is enclosed within 3σ. Increasing the box size effectively only includes more background.

Fig. 2.B.3 shows the convergence of the iterative algorithm for x-position, background and fluorophore intensity. The position estimate converges quickly, whereas the background and fluorophore rates, which are required for calculating the localization accuracy require nearly 10 iterations. The shown example is for $\sigma_{PSF} = 3$ and a box size of 19, which was found to require the most iterations for convergence of all parameter combinations. All subsequent analyses shown were the result of 10 iterations, independent of σ_{PSF} or fit region size. In Fig. 2.B.4 we show example images of simulated single molecules with emission and background rates (in photons) that corresponds with the $\sigma_{PSF}/2$ as discussed in the main text. We find that in all conditions, when the reported CRLB is less than $\sigma_{PSF}/2$ (here 0.5), the reported CRLB matches the theoretical position, and the routine achieves the CRLB.

Performance on experimental single molecule data

2D imaging
Imaging of single Cy5 fluorescent molecules was used to demonstrate the performance of the iterative method under typical single molecule imaging conditions. Since the position, intensity and background rate were not known *a-priori*, we analyzed a set of single molecules that had a steady emission rate over at least one hundred continuous frames. The found standard deviation of the x-position was compared to the mean value of the reported x-dimension localization accuracy, which is calculated for each image that made up a single particle 'trajectory'. As can be seen in Fig. 2.B.5, the reported accuracy gives a good estimate of the actual localization accuracy over a wide range of fluorophore intensities. The results also demonstrate that the Gaussian PSF and pure Poisson noise model with neglection of readout noise are appropriate approximations for single molecule localization in two dimensions.

3D astigmatic imaging
With the found fit parameters on the calibration set (section 2.B) a synthetic data series was generated using 1000 expected photons per frame for the fluorophore, and one expected background count per pixel per frame. 1000 frames for each z-position were generated and analyzed. Fig. 2.B.6 shows the fitting result of the iterative algorithm, which performs a simultaneous fit to x, y, z, and the emission and background rates. The z-position fits both achieve the CRLB value and the CRLB is correctly reported. For fits that include z-position estimates, 20 iterations of the routine were required for convergence.

An analysis of the experimental bead data is shown in Fig. 2.B.7 and 2.B.8. In contrast to fitting simulated 3D data, and all 2D data, the fitting of experimental data does not reach the CRLB. This is attributed to the fact for the CRLB to be justified, the model must be correct.

In 3D imaging, the out-of-focus images are more prone to shape changes of the PSF due to aberrations such as coma and spherical aberration. Therefore, these images do not exactly match the simple astigmatic model. We note, however, that our algorithm does perform a fast MLE given this model, and, as shown with analysis of simulated data, if aberrations can by minimized or eliminated, our routine will also correctly report z-position accuracies. Small amounts of spherical aberration already give rise to a small asymmetry between the two focal lines. Misalignment of the cylinder lens by Δx with respect to optical axis of the tube lens has the effect of a shift of the center of mass of the spot and will result in two different σ_0 fits for x and y as observed in Fig. 2.B.1. The focal line along the

cylinder axis is then displaced by $\sim \left(f_{tub}/f_{cyl}\right)\Delta x = (180/500)\,\Delta x = 0.36\Delta x$ which translates to a shift of $\sim 1/3$ pixel at a $\Delta x = 10\mu$m and a CCD pixel size of 10 μm.

If the optical aberrations cannot be sufficiently reduced, optimal position estimation may be accomplished by fitting the data to a measured PSF [18]. Fitting a measured PSF will be at the expense of the algorithm's speed.

References

[1] B. M. Hanser, M. G. L. Gustafsson, D. A. Agard, and J. W. Sedat, *Phase-retrieved pupil functions in wide-field fluorescence microscopy,* Journal of Microscopy-Oxford **216**, 32 (2004).

[2] A. Schönle and S. Hell, *Calculation of vectorial three-dimensional transfer functions in large angle focusing systems,* Journal of the Optical Society of America A **19**, 2121 (2002).

[3] B. Zhang, J. Zerubia, and J. C. Olivo-Marin, *Gaussian approximations of fluorescence microscope point-spread function models,* Applied Optics **46**, 1819 (2007).

[4] NVIDIA, *Compute unified device architecture (CUDA),* (2007), http://www.nvidia.com.

[5] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory* (Prentice Hall, 1993).

[6] W. Press, S. Teukolsky, W. Vetterling, and B. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed. (Cambridge Univeristy Press, Cambridge, 1992).

[7] L. Holtzer, T. Meckel, and T. Schmidt, *Nanometric three-dimensional tracking of individual quantum dots in cells,* Applied Physics Letters **90**, 1 (2007).

[8] I. Rasnik, S. A. McKinney, and T. Ha, *Non-blinking and longlasting single-molecule fluorescence imaging,* Nature Methods **3**, 891 (2006).

[9] K. A. Lidke, B. Rieger, D. S. Lidke, and T. M. Jovin, *The role of photon statistics in fluorescence anisotropy imaging,* Institute of Electrical and Electronics Engineers Transactions on Image Processing **14**, 1237 (2005).

[10] N. L. Andrews, K. A. Lidke, J. R. Pfeiffer, A. R. Burns, B. S. Wilson, J. M. Oliver, and D. S. Lidke, *Actin restricts Fc epsilon RI diffusion and facilitates antigen-induced receptor immobilization,* Nature Cell Biology **10**, 955 (2008).

[11] I. Young, R. Zagers, L. v. Vliet, J. Mullikin, F. Boddeke, and H. Netten, *Depth-of-focus in microscopy,* in *SCIA'93, Proceedings of the 8th Scandinavian Conference on Image Analysis*, edited by K. A. Høgda, B. Braathen, and K. Heia (Norwegian Society for Image Processing and Pattern Recognition, Tromsø Norway, 1993) pp. 493–498.

[12] 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**, 1642 (2006).

[13] M. J. Rust, M. Bates, and X. W. Zhuang, *Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM),* Nature Methods **3**, 793 (2006).

[14] M. Bates, B. Huang, G. T. Dempsey, and X. Zhuang, *Multicolor super-resolution imaging with photo-switchable fluorescent probes,* Science **317**, 1749 (2007).

[15] A. Egner, C. Geisler, C. Von Middendorff, H. Bock, D. Wenzel, R. Medda, M. Andresen, A. C. Stiel, S. Jakobs, C. Eggeling, A. Schonle, and S. W. Hell, *Fluorescence nanoscopy in whole cells by asynchronous localization of*

photoswitching emitters, Biophysical Journal **93**, 3285 (2007).

[16] A. Small, *Theoretical limits on errors and acquisition rates in localizing switchable fluorophores,* Biophysical Journal **92**, L16 (2009).

[17] S. Wolter, M. Schuttpelz, M. Tscherepanow, S. van de Linde, M. Heilemann, and M. Sauer, *Real-time computation of subdiffraction-resolution fluorescence images,* Journal of Microscopy **237**, 12 (2010).

[18] M. J. Mlodzianoski, M. F. Juette, G. L. Beane, and J. Bewersdorf, *Experimental characterization of 3D localization techniques for particle-tracking and super-resolution microscopy,* Optics Express **17**, 8264 (2009).

Single Molecule Localization

2

3

Probability-Based Particle Detection That Enables Threshold-Free and Robust *In Vivo* Single Molecule Tracking

I believe that we do not know anything for certain, but everything probably.

Christiaan Huygens

Carlas S. Smith
Sjoerd Stallinga
Keith A. Lidke
Bernd Rieger
David Grunwald

Abstract

Single-molecule-detection in fluorescent nanoscopy has become a powerful tool in cell-biology, but can present vexing issues in image analysis like limited signal, unspecific background, empirically set thresholds, image filtering and false-positive-detection limiting overall detection efficiency. Here we present a framework where expert knowledge and parameter tweaking is replaced with a probability based hypothesis test. The result of our method is twofold, delivering robust and threshold-free signal detection with a defined error estimate and improved detection of weaker signals. The probability value has consequences for down stream data analysis, for instance, weighing a series of detections corresponding probabilities, Bayesian propagation of probability or defining metrics in tracking applications. Here we show that the method outperforms all current state-of-the-art approaches, yielding a detection efficiency >70% and a false-positive-detection-rate <5% under conditions of down to 17 photons/pixel-background and 180 photons/molecule-signal, which is beneficial for any kind of photon-limited application. Examples include limited brightness and photostability, phototoxicity in live-cell single-molecule-imaging or use of new labels for nanoscopy. We present simulations, experimental data and tracking of low signal mRNAs in yeast cells.

3.1. *Introduction*

The ability to image single molecules has revolutionized the way molecular interactions can be probed, the environments wherein this is possible and the resolution that can be achieved by use of light microscopy. While the technology is readily available, the analysis of the images often is perceived as challenging as a fair degree of judgment is needed to choose appropriate image filter and intensity thresholds to identify potential signals. In many single-molecule fluorescence applications, such as super-resolution localization microscopy or single-molecule tracking, the position and intensity of a single fluorophore needs to be measured. The first analysis step is the detection of regions that could contain signal originating from single molecules. Detection is especially difficult and important for applications where the fluorescent signal is weak, photo-bleaching is limiting for the observation time or a high background noise is present, such as in vivo RNA imaging or 3D imaging [1–7]. Numerous methods exist to tackle detection [8–11]. However, for all these methods the rate of false positive detection events is unknown, resulting in no quantitative assessment of true and/or false detections at this very first step underlying all further analysis. Practically, even for relatively good data, such as can be expected from smFISH, automated detection and visual inspection of images can diverge such that the user identifies numerous signals that were missed by the automated detection. Here we present a method that employs a probability-based hypothesis test that enables a minimum number of false-negative detections maintaining a fixed number of false positive detections.

The two measures that generally characterize the quality of detection are: The detection efficiency Q (sometimes called sensitivity or recall), which is defined as the ratio of the true number of events over all detected events, and the false-positive rate FP, which is defined as the ratio of all false detections over the total number of true and false detections. Q and FP can only be known in simulations or well-designed test experiments, but are unknown for a real data set. Since image filters, filter settings and intensity thresholds are determined empirically, the Q and FP of existing methods depend intricately on user-set parameters and are not observable or controllable using existing methods. This lack of direct control over Q and FP results in unreliable detection behavior, especially in photon-starved circumstances with a low Signal-to-Background Ratio (SBR).

To overcome the user dependency of current methods we present an alternative approach using pixel based hypothesis testing that delivers a minimum number of false-negative detections at a controlled/fixed number of false positives. This is possible by estimating the probability that a pixel contains signal originating from a single molecule by comparing the likelihood of a foreground

model (emitter present) over that of a background model (no emitter) using a Generalized Likelihood Ratio Test (GLRT)[12, 13]. GLRT uses estimators for which we explicitly use prior knowledge of noise characteristics in light emission; the microscope point spread function (PSF), and camera performance. The maximum likelihood of both models is computed for each pixel x, y of the image using a small region of interest around each pixel, approximately the size of the PSF (see Appendix 3.B), and results in the following test statistic:

$$T_G(x, y) = 2\log\left(\frac{\max_{S_1, b_1}[L(S_1, b_1; d)]}{\max_{b_0}[L(0, b_0; d)]}\right), \tag{3.1}$$

where $L(S, b; d)$ is the likelihood for a single-molecule signal intensity S and background b, given the pixel data d. The test statistic $T_G(x, y)$ is used to calculate the false positive probability $P_{FA}(x, y)$, i.e. the fraction of all tests (pixels) that results in a false positive detection (Appendix 3.B). The false positive probability is corrected for the number of (dependent) hypothesis tests executed to generate a false positive rate $FP(x, y)$ that gives the rate of positive tests that are incorrect. This correction is performed using a statistical tool termed False Discovery Rate (FDR) control (Appendix 3.B)[14]. This enables the algorithm to target a user defined FP in experimental data without knowing the true outcome a-priori, as would be the case in a simulated test data set. Regions-of-interest (ROI) for localization of a potential fluorophore are identified from connected regions of pixels where the $FP(x, y)$ is below a user-set target value. Multiple Target Tracking (MTT)[9] also uses GRLT for detection, but only as a quality control on ROIs that are selected with an initial detection method. Our approach removes the dependency of the arbitrary initial detection by integrating the hypothesis test into the candidate ROI selection and enables control of the number of false positives by our FDR control method.

3.2. *Results*

To develop a quantitative understanding of the performance of our approach, we used simulations of switching single emitters on a randomly labeled Siemens star to represent single-molecule localization based super-resolution data (Fig. 3.1). Switching of the emitters results in variation of signal strength as the emitters can be on for the whole integration time of one frame or for parts of it (Fig. 3.1) We compare Q and the Fourier Ring Correlation resolution (FRC) [15] at fixed FP against the current best methods, MTT and the Scale Space Approach (SSA) (Fig. 3.1 [10, 15–17]. The reason that we are able to fix FP for MTT and SSA, although
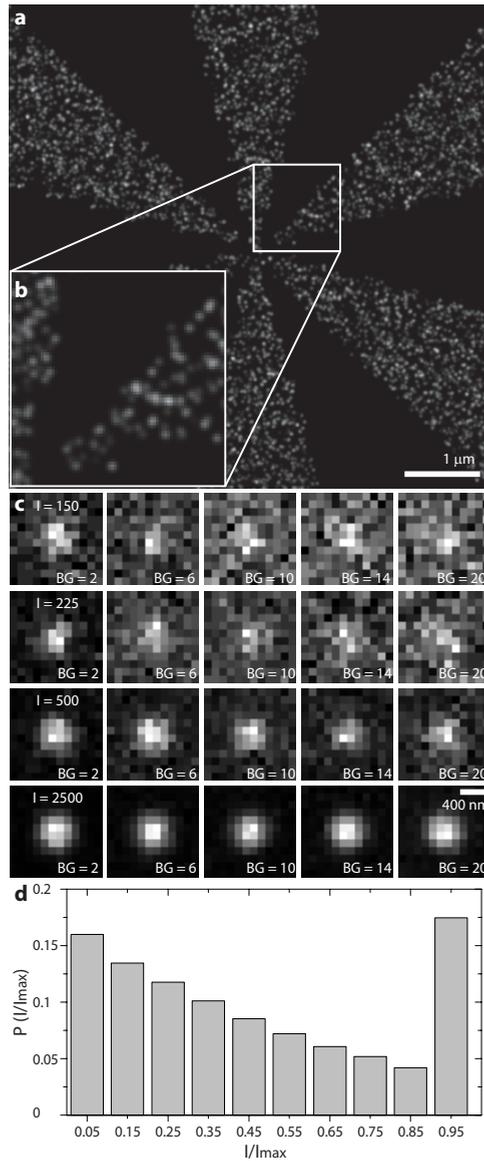
Figure 3.1. Simulation object and PSF examples. a) Super resolution reconstruction of a Siemens star as used in the simulations. Shown is a time projections in which each singe molecule is replaced by its localization precision. b) Zoom in on (a) visualizing the achieved resolution at the highest frequency region of the object. c) Examples of simulated single molecule images as used in simulations with varying signal (I_{max}) and background rates (in photons) as indicated in the panels. An area of 12x12 [pixel] with $\sigma_{PSF} = 1.39$ [pixel] is shown. d) Single molecule intensity distribution resulting from a Markov process simulation for the of the STORM experiment, where the maximum on time is longer than 1 frame.

neither method can do so, is our knowledge of ground truth in the simulations, which we use for optimizing the thresholds instead of relying on empirical settings. In contrast to the GLRT, the FP rate for the SSA and MTT in Fig 2. are bona fide fixed, which is only possible in simulations.

The intensities of the switching single molecules were set to 150, 225, 500, and 2500 photons (I_{max}) per PSF for the duration of a full frame (Fig 1). The PSF had a width of σ_{PSF} = 1.39 pixel (Appendix 3.B). The background intensities varied from 2 to 20 photons per pixel (Fig. 3.1). For MTT and SSA the FP was fixed to <5%. At 150, 225, and 500 photons, the detection efficiency Q for GLRT was at least 10% higher than that for MTT; the detection efficiencies of GLRT and MTT were similar at 2500 photons (GLRT 97%, MTT 94%; Fig. 3.1a,b). Compared to SSA, GLRT detects 10-15% more true spots across all intensity levels (Fig. 3.1a,c). At low photon counts, MTT has higher detection efficiency than SSA. The detection efficiency of GLRT at 225 photons was similar to that of MTT and SSA at 500 photons at low (2 photons) to medium (12 photons) background levels (Fig. 3.1a-c). The FP for GLRT remains well below the specified target of 5% (Fig. 3.1d). We reconstructed images from true positive detections of all methods and computed their FRC resolutions [15]. At high intensities (I = 2500), the FRC resolution of GLRT, MTT, and SSA were all within the uncertainty of each other. However as the intensity decreases the resolution for the GLRT is better with an improvement of 30 nm at 150 photons (Fig. 3.1g-i).

The use of simulated data to test performance of multiple algorithms has the advantage that the true outcome is known and no assumptions, e.g. for variables like noise, background and the shape of point sources, have to be made as they can be set as desired. On the downside the parameters chosen might not reflect the conditions of a real experiment well. Therefore, to experimentally test the performance of GLRT, MTT, and SSA at adjustable SBR levels in a realistic environment, we immobilized 100 nm diameter fluorescent beads on a cover glass and excited them using a white-light laser (Fig. 3.2). The first channel of the white-light laser was tuned to excite the beads, and one or more channels were tuned to overlay with the emission spectrum of the beads acting as artificial background. Ground truth positions of the beads were found at high enough SBR to guarantee 100% detection and beads were subsequently imaged at a constant signal level but at various background settings (Fig. 3.2a-l). Bead data contain non-uniform image background caused by the laser in the emission band to simulate more realistic experimental conditions, e.g. as they could be found during live cell imaging. As background model we assume a homogenous level of background while the signal model is an integrated Gaussian (see Appendix 3.B). We estimate the local
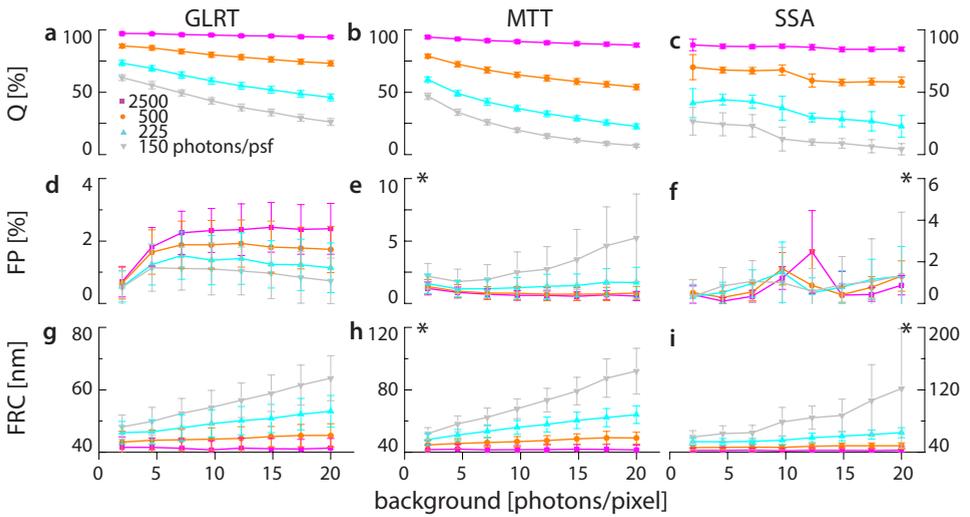
Figure 3.1. Comparison of detection methods in simulation. Localization microscopy images of a Siemens star object have been simulated for different levels of single molecule and background signal. (a-c) Detection efficiency Q, (d-f) false positive rate FP, and (g-i) image resolution according to Fourier-Ring-Correlation (FRC), for the three tested methods, demonstrating superior performance of the Generalized-Likelihood-Ratio-Test (GLRT) over Multiple-Target-Tracking (MTT) and Scale-Space-Approach (SSA). * Indicates a change in scale for visibility.
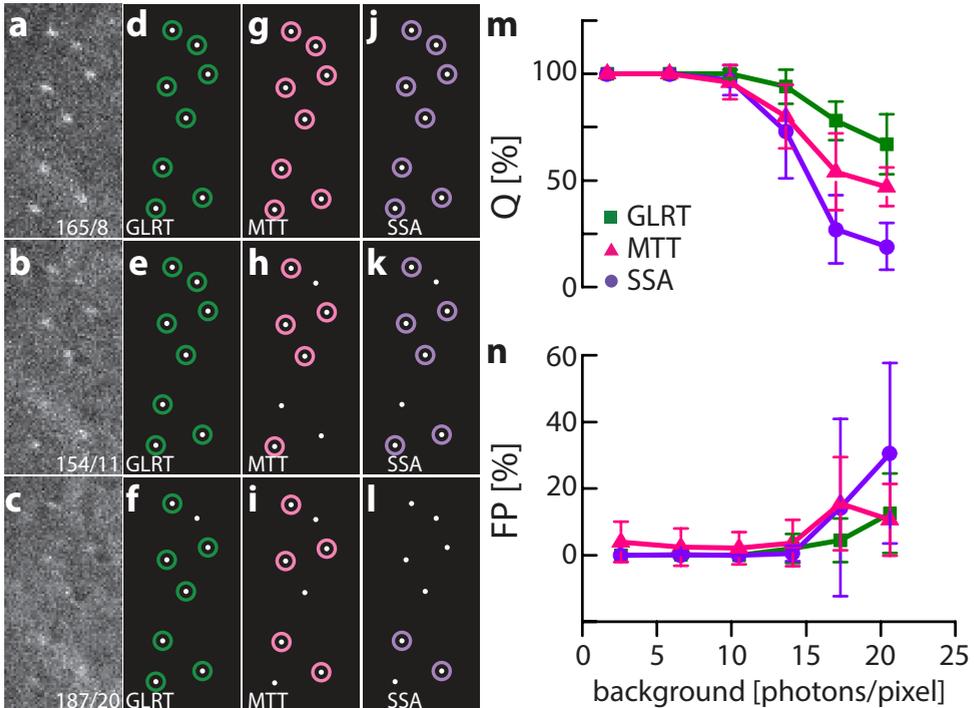
Figure 3.2. Experimental validation of single molecule detection algorithms. The true bead positions are identified at high photon count (not shown). (a-c) 3 examples of bead images. Intensity (photons) and background (photons/pixel) are indicated in raw images. Example of detection results for GLRT (d-f), MTT (g-i), and SSA (j-l). Quantitative comparison of (m) the detection efficiency Q and (n) the false positive rate FP of single molecule detection algorithms, indicating superior performance of GLRT, followed by MTT and SSA.

background within a small box size, which is a fixed ratio to the width of the PSF (see Appendix 3.A) resulting here in a 12x12 pixel box. GLRT allows the use of other signal and background models. At background levels above 7 photons/pixel, GLRT had the highest Q, followed by MTT, agreeing with our simulations (Fig. 3.2m). For high SBRs all methods obtained $Q = 1$, somewhat outperforming simulations- likely because beads in the experiment emit constantly, while in simulation the duty cycle of single fluorophore is less than the duration of a frame as is typically the case in STORM-like experiments. Up to background levels of 17 photons/pixel, GLRT maintains an average FP rate below 5%. At higher background levels, in contrast to our simulations, the FP rate of GLRT did increase but remained well below the FP rate of MTT and SSA, both in its average and error interval (Fig. 3.2n). We attribute the difference between simulation and experiment to the noise and background characteristics of the experimental situation compared to the uniformly distributed noise and background in our simulations. This reflects the limitations of simulations to predict the real-life performance of a method, although GLRT maintains its detection advantage under experimental conditions where noise is not uniform.

Finally, we applied GLRT, SSA, and MTT to detect single fluorescently labeled mRNAs in a living cell, a notoriously low signal (50-150 photons) situation with high background (2-10 photons/pixel) (Fig 4). At a FP target of 5%, GLRT detected 1100 spots over 200 frames. To compare between algorithms, we set thresholds for MTT and SSA such that the same number of spots was detected. Detected spots were then linked using the same tracking algorithm [18, 19]. The GLRT produced longer tracks than the other methods, as the superior detection efficiency resulted in less track interrupts. The mean track length was 39 ± 46 frames (longest track 185 frames) compared to 29 ± 20 frames (longest track 69 frames) for MTT and 27 ± 23 frames (longest track 85 frames) for SSA. These differences are significant at p<0.05 (Mann-Whitney U test).

3.3. *Discussion*

Any single molecule study, e.g. in nanoscopy, smFISH, CoSMoS, single molecule tracking or similar, starts with finding those single molecule signals in the recorded images. Single molecule detection fundamentally suffers from false positive and false negative detections. With GLRT, we present a new method for the detection of single molecules that delivers a minimum number of false negatives at a fixed number of false positive detections for the used background, noise and PSF models. Within the GLRT framework it is possible to treat different
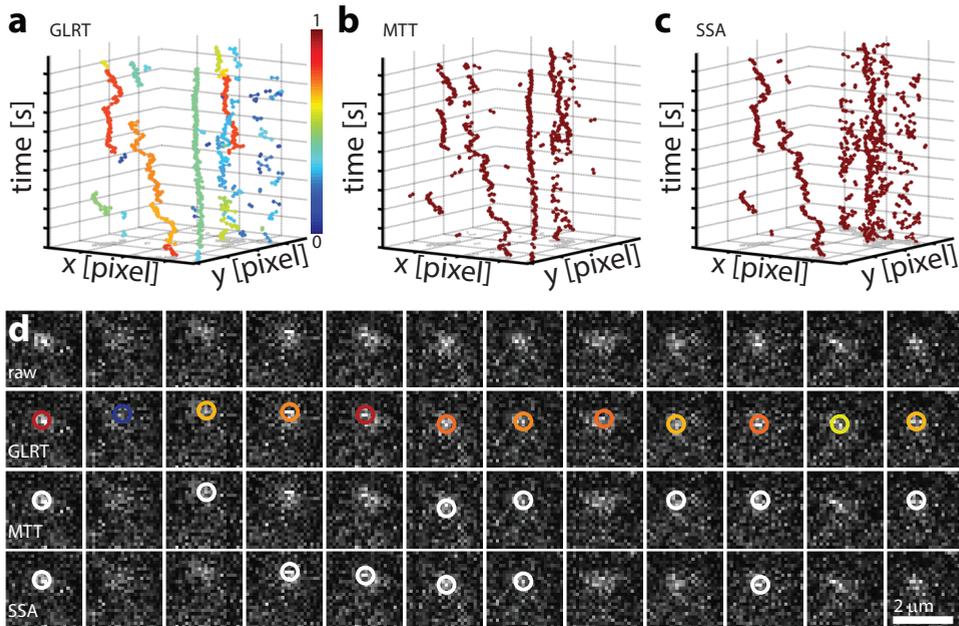
Figure 3.3. Impact of GLRT detection on tracking of single mRNAs in living yeast. (a-c) Constructed traces based on 1072 ± 26 detections and identical linking parameters for (a) GLRT, (b) MTT, and (c) SSA. The color scheme in (a) indicates the average detection probability of the detected position underlying the tracking. d) Sample raw data for detection of fluorescently labeled single mRNAs in a living cell using GLRT, MTT and SSA as indicated. Color scheme of GLRT as in (a).

camera noise models (e.g. sCMOS), PSFs and experiment specific background conditions, which results in a change of the likelihood function, see Appendix 3.B and [20]. In other words GLRT allows setting a target for the FP for a wide range of applications. We showed that this target is stably achieved over a large range of SBR conditions. GLRT is based on statistical testing and significance levels, thereby replacing user-defined thresholds that have an intricate impact on FP and Q. We tested GLRT against multiple state-of-the-art methods under simulated (homogenous background) and experimental (stochastical background and noise distribution) conditions, directly testing the predictive limits of our simulations (Fig. 3.1 & Fig. 3.2). GLRT has the highest detection rate under most conditions, and shows improved resolution in localization-based super-resolution and tracking of single molecules (Fig. 3.3). The robust performance of GLRT at low signal and high background allows lower excitation levels, which will improve the viability of samples during live-cell imaging [21, 22]. It can also open up STORM imaging to a broader palette of fluorophores than is currently in use. In particular the sample handling used for multi antibody labeling with Alexa Fluor647 may be avoided [23, 24]. Most importantly, GLRT provides built-in quality control during signal detection - the first and most fundamental step of data analysis. The framework we present is taking into account the whole imaging process, and is easily adapted to different signal shapes, e.g. a double helix, and adapted to different camera types, see Appendix 3.B. The result of this framework is that we can robustly detect weaker signals, which opens up single molecule studies to applications that were previously not conceivable or can be used to extend the observation time by reducing photobleaching.

References

[1] A. Yildiz, J. N. Forkey, S. A. McKinney, T. Ha, Y. E. Goldman, and P. R. Selvin, *Myosin v walks hand-over-hand: Single fluorophore imaging with 1.5-nm localization,* Science **300**, 2061 (2003).

[2] M. F. Juette, T. J. Gould, M. D. Lessard, M. J. Mlodzianoski, B. S. Nagpure, B. T. Bennett, S. T. Hess, and J. Bewersdorf, *Three-dimensional sub–100 nm resolution fluorescence microscopy of thick samples,* Nature Methods **5**, 527 (2008).

[3] S. Manley, J. M. Gillette, G. H. Patterson, H. Shroff, H. F. Hess, E. Betzig, and J. Lippincott-Schwartz, *High-density mapping of single-molecule trajectories with photoactivated localization microscopy,* Nature Methods **5**, 155 (2008).

[4] S. Tyagi, *Imaging intracellular RNA distribution and dynamics in living cells,* Nature Methods **6**, 331 (2009).

[5] D. Grünwald and R. H. Singer, *In vivo imaging of labelled endogenous β-actin mRNA during nucleocytoplasmic transport,* Nature **467**, 604 (2010).

[6] A. A. Hoskins, J. Gelles, and M. J. Moore, *New insights into the spliceosome by single molecule fluorescence microscopy,* Current Opinion in Chemical Biology **15**, 864 (2011).

[7] E. Cai, P. Ge, S. H. Lee, O. Jeyifous, Y. Wang, Y. Liu, K. M. Wilson, S. J. Lim, M. A. Baird, J. E. Stone, *et al.*, *Stable small quantum dots for synaptic receptor tracking on live neurons,* Angewandte Chemie **126**, 12692 (2014).

[8] J.-C. Olivo-Marin, *Extraction of spots in biological images using multiscale products,* Pattern Recognition **35**, 1989 (2002).

[9] A. Sergé, N. Bertaux, H. Rigneault, and D. Marguet, *Dynamic multiple-target tracing to probe spatiotemporal cartography of cell membranes,* Nature Methods **5**, 687 (2008).

[10] C. S. Smith, N. Joseph, B. Rieger, and K. A. Lidke, *Fast, single-molecule localization that achieves theoretically minimum uncertainty,* Nature Methods **7**, 373 (2010).

[11] 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**, 2081 (2012).

[12] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory* (Prentice Hall, 1998).

[13] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory* (Prentice Hall, 1993).

[14] Y. Benjamini and D. Yekutieli, *The control of the false discovery rate in multiple testing under dependency,* Annals of Statistics , 1165 (2001).

[15] R. P. Nieuwenhuizen, K. A. Lidke, M. Bates, D. L. Puig, D. Grünwald, S. Stallinga, and B. Rieger, *Measuring image resolution in optical nanoscopy,* Nature Methods **10**, 557 (2013).

[16] F. Huang, S. L. Schwartz, J. M. Byars, and K. A. Lidke, *Simultaneous multiple-emitter fitting for single molecule super-resolution imaging,* Biomedical Optics Express **2**, 1377 (2011).

[17] S. T. Low-Nam, K. A. Lidke, P. J. Cutler, R. C. Roovers, P. M. v. B. en Henegouwen, B. S. Wilson, and D. S. Lidke, *ErbB1 dimerization is promoted by domain co-confinement and stabilized by ligand binding,* Nature Structural and Molecular Biology **18**, 1244 (2011).

[18] K. Jaqaman, D. Loerke, M. Mettlen, H. Kuwata, S. Grinstein, S. L. Schmid, and

G. Danuser, *Robust single-particle tracking in live-cell time-lapse sequences,* Nature Methods **5**, 695 (2008).

[19] P. J. Cutler, M. D. Malik, S. Liu, J. M. Byars, D. S. Lidke, and K. A. Lidke, *Multi-color quantum dot tracking using a high-speed hyperspectral line-scanning microscope,* PloS One **8**, e64320 (2013).

[20] B. Rieger and S. Stallinga, *The lateral and axial localization uncertainty in super-resolution light microscopy,* ChemPhysChem **15**, 664 (2014).

[21] P. M. Carlton, J. Boulanger, C. Kervrann, J.-B. Sibarita, J. Salamero, S. Gordon-Messer, D. Bressan, J. E. Haber, S. Haase, L. Shao, *et al.*, *Fast live simultaneous multiwavelength four-dimensional optical microscopy,*

Proceedings of the National Academy of Sciences **107**, 16016 (2010).

[22] F. Huang, T. M. Hartwich, F. E. Rivera-Molina, Y. Lin, W. C. Duim, J. J. Long, P. D. Uchil, J. R. Myers, M. A. Baird, W. Mothes, *et al.*, *Video-rate nanoscopy using sCMOS camera-specific single-molecule localization algorithms,* Nature Methods **10**, 653 (2013).

[23] J. Tam, G. A. Cordier, J. S. Borbely, Á. S. Álvarez, and M. Lakadamyali, *Cross-talk-free multi-color STORM imaging using a single fluorophore,* (2014).

[24] C. C. Valley, S. Liu, D. S. Lidke, and K. A. Lidke, *Sequential superresolution imaging of multiple targets using a single fluorophore,* (2015).

Appendix

3.A. *Materials and Methods*

Single molecule imaging

All experimental data, live cells and beads, were acquired using an inverted, fluorescence microscope equipped with a 150x, 1.45 N.A. objective (Olympus, Tokyo, JP) combined with 200 mm focal length tube lenses (CVI Melles Griot, IDEX, Alburquerque, NM) resulting in an effective magnification of 167x. Images were recorded with an EMCCD (Andor Technology plc., Belfast UK; Model iXon3, 897) featuring 512x512 pixels of 16x16 μm^2 in size, resulting in an effective pixel size in the image of approximately 95 nm. For fluorescence excitation, a 515 nm (Coherent, Carlsbad, CA) and a 561 nm (Cobolt SE, Solna, SE) laser were used for live cell imaging (Saccharomyces Cerevisiae), or a white light laser (NKT Photonics) for imaging of fluorescent beads (TetraSpeck^{TM} Microspheres, 0.1 μm , fluorescent orange, Life Technologies). Wavelength selection and power regulation for the imaging laser was done using an AOTF (AA Optoelectronics, Orsay FR). An emission filter with central wavelength 593 nm and bandwidth 40 nm was placed between the objective and tube lens. To introduce additional background in the bead images, one or more additional wavelengths within the range of the emission filter were selected from the white light laser. Reflection of laser light within the beam path resulted in a non-uniform intensity distribution of background signal and image noise on the camera. The ground truth positions of the beads were found at 220 photons signal and 2.6 photons/pixel background, yielding 100% detection. Beads were then detected at a signal of 180 photons/bead and various background settings (see Fig. 3).

Synthetic data generation and simulation parameters

Synthetic data were generated as a Siemens star with 8 arms on a field of view of 64x64 pixels (6.4x6.4 μm^2). To determine the detection efficiency Q, time series of $T = 100$ frames were produced with an emitter density $\rho = 750$ μm^{-2}, containing on average more than 500 detection events, which was repeated $n = 256$ times. Single molecule switching behavior is modeled as a Markov process with a bleaching rate ($k_b = 0.1$ frame^{-1}), an off rate ($k_{off} = 1$ frame^{-1}). On and off rates are

coupled by the emitter density which is approximated as $\rho = k_{off}/(5k_{on})\ \mu m^{-2}$. The number of simulated emitters N is equal to the density times the area of the object ($N = A_{object}\rho$). This Markov process offers a representation of a STORM experiment, however also biases the resolution calculation, since single emitters are localized more than once [1]. Therefore, for the resolution comparison we performed a much longer ($T = 10^4$), PALM like simulation, where all single emitters could only be localized once, which was repeated $n = 64$ times. The PSF is modeled as a Gaussian with width $\sigma_{PSF} = 0.3\lambda/(NA\Delta x) = 0.3\lambda/(1.45 \cdot 100) = 1.39$ pixels [2, 3]. For our benchmark we have executed four different simulations at single molecule intensities of 150, 225, 500 and 2500 photons of a randomly labeled Siemens star with a density $\rho = 750\ \mu m^{-2}$ (Fig. 1).

Tracking parameters

Using the bacteriophage PP7 RNA-labeling system in live Saccharomyces cerevisiae cells, we tracked single molecule mRNA transport events. We used the same linking parameters for all detection algorithms and tuned the threshold of both the MTT and SSA to detect the same number of spots. The linking parameters that gave the best frame-to-frame connections are: an *on rate* equal to $k_{on} = 100$ frame^{-1}, an *off rate* equal to $k_{off} = 0.1$ frame^{-1}, a *particle density* equal to $\rho = \#detections/\#frames$, and a *diffusion constant* equal to $D = 4$ pixels^2 frame^{-1}. The maximum search distance for a link was chosen as $4D$ and gaps larger than a spatial jump of 3 pixels or temporal 3 frames, was not allowed. These values for jump distance and time gaps were chosen to equal a distance equal to the width of one PSF at our spatial and temporal sampling rates and an RNA mobility in the range of $1\mu m^2/s$ for the diffusive fraction [4]. Furthermore, we weighted the candidates in the cost matrix with their detection probability for GLRT. To obtain similar performance for the MTT and SAA we weighted candidates using a similar quantity stretch$\left(\sqrt{\left(\hat{I}^2 - \hat{bg}^2\right)}\right)$, where $\hat{I}$ is the estimated intensity, $\hat{bg}$ is the estimated background and stretch$(x) = (x - x_{min})/(x_{max} - x_{min})$ with x_{max} and x_{min} as the minimum and maximum value of all linking candidates. For comparing the three methods we identified tracks longer than 8 frames.

Benchmark metrics

The metric with which we judged the performance of the methods is of key importance. For the GLRT we set the target FP at 5% (near 2σ-level), since this is most often used in hypothesis testing. The actual obtained FP rate can be calculated in simulation since the true events are known. We compared the methods based

on their efficiency, Q, and therefore ensure that the FP rate is comparable. Using bisectional optimization for SSA, the detection threshold in each Video was set such that the maximal number of true signals was found while the FP value stayed as close as possible to the target value. To automatically find this threshold we used 8 iterations of bisectional optimization. The intensity threshold for the SSA was set such that the FP rate was matched as closely as possible to that of the GLRT, 5% FP signals. The MTT was set so that the overall average was comparable to that of the GLRT. The FP for the MTT is comparable to that of GLRT (Fig. 2d-f), but is higher for low (150) light-level signals and high (>15) background.

Photon count calibration

The signal is detected by the camera in analog digital units and needs to be converted into photon counts for localizations using Poisson noise characteristics [5]. Conversion was done as described earlier [6].

GPU implementation

We performed two maximum likelihood fits for every pixel, for typical EMCCD we have 512x512 pixels, which requires fitting $5 \cdot 10^5$ sub-regions with side lengths of $3 \cdot (2\sigma_{PSF} + 1)$ pixels to obtain optimal localization precision [5]. Per pixel two MLEs are done: one for the foreground model H_1, and background model H_0. The MLEs are calculated as described earlier [5]. Quantification of the speed-up of MLE fitting on GPUs can be found in earlier work [5]. With recent advances in GPU hardware a 500 times speedup is feasible (Intel-Core-i7-5960X vs. GeForce GTX 780 Ti) [5]. On a GTX 780 graphic card (Nvidia) the processing time per 512x512 pixel image is under one second for $\sigma_{PSF} = 1.39$ pixels. The GPU processing was written in CUDA (Nvidia) and compiled using Visual Studio 2013. The CUDA routine was compiled as a mex-file for Matlab (The Mathworks, MA) and called from Matlab. We provide a software tool and example that implements and demonstrates the detection algorithm as supplemental software, including multi-graphic card support.

Algorithmic details for detection methods

SSA: The two different filter kernels for the noise reduction by smoothing and for background detection are uniform filters with a size of $1.5(2\sigma_{PSF} + 1)$ and $3(2\sigma_{PSF} + 1)$ pixels, respectively. The local maxima are detected using a maximum filter with a size of $3(2\sigma_{PSF} + 1)$ pixels and a user defined intensity threshold is used to create the candidate map.

3.B. *Derivation*

Single molecule detection

A hypothesis test for the presence of a single molecule is performed for each camera pixel. In this test H_1 is the presence of a single molecule, i.e. the intensity of the single molecule being non-zero ($\theta_I \neq 0$), and H_0 is the absence of a single molecule, i.e. the intensity being equal to zero ($\theta_I = 0$):

$$H_0 : \theta_I = 0, \; \theta_{bg},$$
$$H_1 : \theta_I \neq 0, \; \theta_{bg}. \tag{3.2}$$

The intensity θ_I and background θ_{bg} are generally unknown and have to be estimated. The background needs to be estimated, but does not need to be tested (nuisance parameter), in contrast to the intensity. Values, intensity and background, are found by Maximum Likelihood Estimation (MLE), one estimate for each of the two hypotheses. The Maximum Likelihood values found by this procedure are used as input for the likelihood ratio test. The MLE procedure is described in earlier work [5]. The position of the hypothetical emitter is fixed at the center of the pixel. Averaging over multiple positions inside the pixel area did not lead to an appreciable improvement. Including the position in the MLE procedure as an additional fit (nuisance) parameter led to a less robust behavior.

The image formation model

The Point Spread Function (PSF) is approximated by a Gaussian distribution, which is known to be a valid approach in the context of 2D single emitter localization [3]:

$$PSF(x, y) = \frac{1}{2\pi\sigma_0^2} e^{-\frac{1}{2\sigma_0^2}\left[(x-x_0)^2 + (y-y_0)^2\right]}. \tag{3.3}$$

This PSF must be integrated over the pixel area to arrive at the expected photon count at each pixel k:

$$\mu_k = \theta_I \Delta E(x_k - x_0)\, \Delta E\left(y_j - y_0\right) + \theta_{bg}, \tag{3.4}$$

with

$$\Delta E(u) = \frac{1}{2}\left[\mathrm{erf}\left(\frac{u+\frac{1}{2}}{\sqrt{2}\,\sigma_0}\right) - \mathrm{erf}\left(\frac{u-\frac{1}{2}}{\sqrt{2}\,\sigma_0}\right)\right], \tag{3.5}$$

68

where (x_k, y_k) are the pixel coordinates in unit pixel of pixel k, (x_0, y_0) is the location of the center of the PSF in unit $pixel$ and, σ_0 is the PSF width, depending on the numerical aperture (NA), magnification (M), pixel size (Δp)and the wavelength of the light (λ).

Generalized likelihood ratio test statistic

The hypothesis test that best approximates the optimal Neyman-Pearson test [7] is the Generalized Likelihood Ratio Test (GLRT), where the Maximum Likelihood estimates of the parameter vectors ($\hat{\theta}$) for the two hypotheses are used instead of their true values (θ). The GLRT Statistic (T_G) is given by:

$$
\begin{aligned}
T_G &= 2\log\left(\frac{\max_{\theta_I^1, \theta_{bg}^1}\left[P\left(\theta_I^1, \theta_{bg}^1; x\right)\right]}{\max_{\theta_{bg}^0}\left[P\left(0, \theta_{bg}^0; x\right)\right]}\right) \\
&= 2\log\left(\frac{P\left(\hat{\theta}_I^1, \hat{\theta}_{bg}^1; x\right)}{P\left(0, \hat{\theta}_{bg}^0; x\right)}\right),
\end{aligned}
\tag{3.6}
$$

where $P\left(\hat{\theta}_I^1, \hat{\theta}_{bg}^1; x\right)$ is the maximum likelihood of the measured data (x) under hypothesis, H_1 and $P\left(0, \hat{\theta}_{bg}^0; x\right)$ is the maximum likelihood of the measured data (x) under hypothesis, H_0. Here $P(.)$ is the likelihood function, which describes the noise model of the camera used. We assume Poisson noise, which is a good assumption for EMCCD cameras [5], and can easily be modified for sCMOS [8]. The two MLE fits that are performed for each pixel must include photon counts from surrounding pixels where the size of this subregion depends on the width of the PSF (σ_0). We choose the size of this subregion as small as possible without jeopardizing the localization precision [5], $s = 3(2\sigma_0 + 1)$. For pixels near the border of the image that do not have enough neighbors to fill the subregion no MLE and therefore no GRLT is performed.

Up to now we have calculated the GRLT statistic (T_G) per pixel, but this value by itself does not have a useful meaning. As for any other test statistic its value can be converted into a false positive probability ($P_{FA} = P(H_1; H_0)$). This is done using the probability distribution of the test statistic. We have found (proof in the next section) that for this specific problem the false positive probability is given by: $P_{FA} = 2CDF(-\sqrt{T_G})$, where CDF is the cumulative distribution function of the

standard normal distribution defined by:

$$CDF(x) = \frac{1}{2}\left[1 + \text{erf}\left(\frac{x}{\sqrt{2}}\right)\right]. \tag{3.7}$$

The distribution of the generalized likelihood ratio test statistic

Any hypothesis test returns a so-called p-value that measures the likelihood of the current value of the test statistic to be a false positive, i.e. that the value is wrongly classified as signal where it should have been background (probability of H_1 (signal) given H_0 (background)). Using this p-value we decide if the measured difference is significant or not. Apart from the false positive probability, which is the distribution of T_G under the null hypothesis H_0 ($P_{FA} = P(H_1; H_0)$), we can also calculate the detection probability, which is the distribution of T_G under H_1 ($P_D = P(H_1; H_1)$). Below we show that the false positive probability is given by: $P_{FA} = 2CDF(-\sqrt{T_G})$.

The GLRT uses estimated parameters instead of the true parameters and therefore we have to derive the distribution of the test statistic based on the properties of the estimator. The GRLT is formulated using MLE which is unbiased (if it exists and is unrestricted) and asymptotically attains, as the number of data points go to infinity ($n \to \infty$), the minimal possible variance in the parameter estimate, the so called Cramer Rao Lower Bound (CRLB) [9]. Using this property we can derive the distribution of our test statistic. The CRLB is given by the inverse of the Fisher information matrix [9]:

$$I(\theta) = -E\left[\frac{\partial^2 \log P(\theta; x)}{\partial \theta^2}\right], \tag{3.8}$$

where the $E[\cdot]$ is the expectation operation and $P(\theta; x)$ is the likelihood of the parameters (θ) given the data (x); and for an MLE that attains the CRLB ($n \to \infty$) we can make the following Taylor expansion around the true value (θ): $\log P(\hat{\theta}; x) = -\frac{1}{2}(\hat{\theta} - \theta)^T I(\theta)(\hat{\theta} - \theta) + c(\theta)$, which is consistent with the fact that in expectation the second derivative results in the Fisher information. This is an important expression, because equivalently:

$$P(\hat{\theta}; x) = P(\theta; x)\exp\left[-\frac{1}{2}(\hat{\theta} - \theta)^T I(\theta)(\hat{\theta} - \theta)\right], \tag{3.9}$$

which shows us that for large data records ($n \to \infty$) the MLE is normally distributed with a covariance equal to that of the CRLB.

70

In our problem there are two unknown parameters that have to be estimated (θ_I, θ_{bg}), but only one parameter to test $(\theta_I; H_0 : \theta_I = 0, \ H_1 : \theta_I \neq 0)$. This means that we can make a separation of θ into the parameters we need to estimate, but do not need to test (nuisance parameters, θ_n) and parameters we need to estimate and test (parameters of interest, θ_t) as the concatenated vector $\theta = [\theta_t \ ; \theta_n]$. Here we test for $H_0 : \theta_t = \theta_{t_0}, \ H_1 : \theta_t \neq \theta_{t_0}$, where the parameters to be tested are θ_t against θ_{t_0}. We need two MLEs ($\hat{\theta}^0 = [\theta_{t_0}; \hat{\theta}_n^0], \ \hat{\theta}^1 = [\hat{\theta}_t^1; \hat{\theta}_n^1]$), because there are nuisance parameters that have to be estimate under both hypotheses (H_0, H_1). The GLRT statistic follows as

$$T_G = 2\log\left(\frac{P\left(\hat{\theta}_t^1, \hat{\theta}_n^1; x\right)}{P\left(\theta_{t_0}, \hat{\theta}_n^0; x\right)} \right). \tag{3.10}$$

The MLE under H_1 is unrestricted ($\hat{\theta}^1$) and therefore independent of which hypothesis is true, we will find that $E\left[\hat{\theta}_t^1\right] = \theta_t$ and $E\left[\hat{\theta}_n^1\right] = \theta_n$, however the MLE under H_0 is restricted (constrained to $\theta_t = \theta_{t_0}$) and therefore if H_1 is true we obtain a biased estimate, $E\left[\hat{\theta}_n^0\right] \neq \theta_n$, and only if H_0 is true we have that $E\left[\hat{\theta}_n^0\right] = \theta_n$. Asymptotically (as $n \to \infty$), we can make the following second order Taylor expansion around the MLE $\hat{\theta}^1$ for $P\left(\theta_{t_0}, \theta_n; x\right)$. This expansion is valid because the MLE under H_1 ($\hat{\theta}^1$) will maximize the likelihood independent of which hypothesis is true:

$$P\left(\theta_{t_0}, \theta_n; x\right) = \tag{3.11}$$

$$P\left(\hat{\theta}_t^1, \hat{\theta}_n^1; x\right) \exp\left[-\frac{1}{2}\left(\begin{bmatrix} \hat{\theta}_t^1 \\ \hat{\theta}_n^1 \end{bmatrix} - \begin{bmatrix} \theta_{t_0} \\ \theta_n \end{bmatrix} \right)^T I\left(\hat{\theta}_t^1, \hat{\theta}_n^1\right) \left(\begin{bmatrix} \hat{\theta}_t^1 \\ \hat{\theta}_n^1 \end{bmatrix} - \begin{bmatrix} \theta_{t_0} \\ \theta_n \end{bmatrix} \right) \right].$$

However we need, $P\left(\theta_{t_0}, \hat{\theta}_n^0; x\right)$, instead of $P\left(\theta_{t_0}, \theta_n; x\right)$, which is the maximum of $P\left(\theta_{t_0}, \theta_n; x\right)$ for θ_n. To be able to perform the maximization of $P\left(\theta_{t_0}, \theta_n; x\right)$ to θ_n we introduce a short hand notation, where we factorize the Fisher information matrix according to test and nuisance parameters as we have done for the parameter vector:

$$I\left(\hat{\theta}^1\right) = \begin{bmatrix} I_{\theta_t \theta_t}\left(\hat{\theta}^1\right) & I_{\theta_t \theta_n}\left(\hat{\theta}^1\right) \\ I_{\theta_n \theta_t}\left(\hat{\theta}^1\right) & I_{\theta_n \theta_n}\left(\hat{\theta}^1\right) \end{bmatrix}, \tag{3.12}$$

the Fisher information matrix is symmetric, and therefore the partitioned Fisher information matrix following symmetry properties: $I_{\theta_t \theta_t}\left(\hat{\theta}^1\right) = I_{\theta_t \theta_t}\left(\hat{\theta}^1\right)^T, I_{\theta_n \theta_n}\left(\hat{\theta}^1\right) =$

$\mathrm{I}_{\theta_n\theta_n}(\hat{\theta}^1)^T$, $\mathrm{I}_{\theta_n\theta_t}(\hat{\theta}^1) = \mathrm{I}_{\theta_t\theta_n}(\hat{\theta}^1)^T$. Using this short hand notation and the symmetry properties we obtain the maximum by setting the gradient equal to zero

$$\frac{\partial\log\left(P\left(\theta_{t_0},\theta_n;x\right)\right)}{\partial\theta_n} = \mathrm{I}_{\theta_n\theta_t}(\hat{\theta}^1)\left(\hat{\theta}_t^1 - \theta_{t_0}\right) + \mathrm{I}_{\theta_n\theta_n}(\hat{\theta}^1)\left(\hat{\theta}_n^1 - \theta_n\right)$$
$$= 0, \tag{3.13}$$

and solving for $\theta_n = \hat{\theta}_n^0$. We find that the maximum is obtained at

$$\hat{\theta}_n^0 = \hat{\theta}_n^1 - \mathrm{I}_{\theta_n\theta_n}(\hat{\theta}^1)^{-1}\mathrm{I}_{\theta_n\theta_t}(\hat{\theta}^1)\left(\hat{\theta}_t^1 - \theta_{t_0}\right). \tag{3.14}$$

After back substitution into $P\left(\theta_{t_0},\theta_n;x\right)$ we find that

$$P\left(\theta_{t_0},\theta_n;x\right) = P\left(x,\hat{\theta}_t^1,\hat{\theta}_n^1\right) \tag{3.15}$$
$$\exp\left[-\frac{1}{2}(\hat{\theta}_t^1 - \theta_{t_0})^T\left[\mathrm{I}_{\theta_t\theta_t}(\hat{\theta}^1) - \mathrm{I}_{\theta_t\theta_n}(\hat{\theta}^1)\mathrm{I}_{\theta_n\theta_n}(\hat{\theta}^1)^{-1}\mathrm{I}_{\theta_n\theta_t}(\hat{\theta}^1)\right](\hat{\theta}_t^1 - \theta_{t_0})\right],$$

which can be simplified using the block inversion lemma because

$$\left[I(\hat{\theta}^1)^{-1}\right]_{\theta_t\theta_t} = \left[\mathrm{I}_{\theta_t\theta_t}(\hat{\theta}^1) - \mathrm{I}_{\theta_t\theta_n}(\hat{\theta}^1)\mathrm{I}_{\theta_n\theta_n}(\hat{\theta}^1)^{-1}\mathrm{I}_{\theta_n\theta_t}(\hat{\theta}^1)\right]^{-1}. \tag{3.16}$$

We observe that after quite some linear algebra we have obtained exactly what we have expected: the covariance of the parameters θ_t is given by the upper right block $[\cdot]_{\theta_t\theta_t}$ of the inverse of the Fisher information matrix which is the CRLB for the parameters (θ_t). Note that the covariance is dependent on both the nuisance and the test parameters.

After substitution of this expression into the GLRT statistic we find that the limit form follows as

$$T_G = 2\log\left(\frac{P\left(\hat{\theta}_t^1,\hat{\theta}_n^1;x\right)}{P\left(\theta_{t_0},\hat{\theta}_n^0;x\right)}\right) \tag{3.17}$$
$$= 2\log\left(\frac{P\left(\hat{\theta}_t^1,\hat{\theta}_n^1;x\right)}{P\left(\hat{\theta}_t^1,\hat{\theta}_n^1;x\right)\exp\left[-\frac{1}{2}(\hat{\theta}_t^1 - \theta_{t_0})^T\left(\left[I(\hat{\theta}^1)^{-1}\right]_{\theta_t\theta_t}\right)^{-1}(\hat{\theta}_t^1 - \theta_{t_0})\right]}\right)$$

72

$$= \left(\hat{\theta}_t^1 - \theta_{t_0} \right)^T \left(\left[I(\hat{\theta}^1)^{-1} \right]_{\theta_t \theta_t} \right)^{-1} \left(\hat{\theta}_t^1 - \theta_{t_0} \right).$$

Asymptotically, $E\left[[\hat{\theta}_t^1; \hat{\theta}_n^1] \right] = [\theta_t; \theta_n]$, and therefore we may replace the values in the Fisher matrix, if H_1 is true:

$$T_G = \left(\hat{\theta}_t^1 - \theta_{t_0} \right)^T \left(\left[I([\theta_t; \theta_n])^{-1} \right]_{\theta_t \theta_t} \right)^{-1} \left(\hat{\theta}_t^1 - \theta_{t_0} \right), \tag{3.18}$$

and if H_0 is true ($E\left[[\theta_{t_0}; \hat{\theta}_n^1] \right] = [\theta_{t_0}; \theta_n]$):

$$T_G = \left(\hat{\theta}_t^1 - \theta_{t_0} \right)^T \left(\left[I([\theta_{t_0}; \theta_n])^{-1} \right]_{\theta_t \theta_t} \right)^{-1} \left(\hat{\theta}_t^1 - \theta_{t_0} \right). \tag{3.19}$$

We observe that under all circumstances, independent of the true hypothesis (H_i), we have that $T_G \geq 0$, which is in agreement with what we expect, since the MLE under hypothesis H_1, $\hat{\theta}^1$, always results a higher likelihood ($P\left(\hat{\theta}_t^1, \hat{\theta}_n^1; x \right) \geq P\left(\theta_{t_0}, \hat{\theta}_n^0; x \right)$), as there are additional fit parameters present.

Now that the distribution of the GLRT is obtained using the limit form and based on the fact that the MLE of θ is asymptotically ($n \to \infty$) normally distributed:

$$\hat{\theta}_{t_i}^1 \sim \begin{cases} N\left(\theta_{t_0}, \left[I([\theta_{t_0}; \theta_n])^{-1} \right]_{\theta_t \theta_t} \right) & under \quad H_0 \\ N\left(\theta_t, \left[I([\theta_t; \theta_n])^{-1} \right]_{\theta_t \theta_t} \right) & under \quad H_1 \end{cases}, \tag{3.20}$$

Here i denotes the true hypothesis H_i, $N(\mu, \Sigma)$ denotes the normal distribution with mean μ, the true parameter vector, and covariance Σ, which is the estimation uncertainty.

We continue by observing that the limit form of T_G under both hypotheses is of the form:

$$T_G = x^T \Sigma^{-1} x, \tag{3.21}$$

where $x \sim N(\mu, \Sigma)$. The covariance matrix Σ is symmetric and positive semi-definite and can hence be factorized as $\Sigma = LL^T$. Defining the variable $z = L^{-T} x$ we find that:

$$T_G = x^T L^{-1} L^{-T} x = z^T z. \tag{3.22}$$

The variable z is normally distributed with identity covariance. If a random variable z_1 follows a normal distribution with a non-zero mean and identity

covariance ($z_1 \sim N(\mu, I)$) then its square ($z_1^T z_1$) follows a non-central chi-square distribution ($\chi_t^2(\delta)$) with t degrees of freedom ($rank(I) = t$) and a non-centrality parameter [10] ($\delta = \mu^T \mu$). The non-central chi-square distribution simplifies to a central chi-square distribution when the mean of the normal distribution is zero ($z_2 \sim N(0, I) \rightarrow z_2^T z_2 \sim \chi_t^2$). Putting everything together we find that the test statistic follows a non-central chi-square distribution under H_1 and a central chi-square distribution under H_0 (as then the mean of the test parameter is zero):

$$
T_G \sim \left\{ \begin{array}{ll} \chi_t^2 & under\ H_0 \\ \chi_t^2(\delta) & under\ H_1 \end{array} \right. ,
\tag{3.23}
$$

where the non-centrality parameter (δ) is given by:

$$
\delta = \left(\theta_t - \theta_{t_0}\right)^T \left(\left[I([\theta_t; \theta_n])^{-1}\right]_{\theta_t \theta_t}\right)^{-1} \left(\theta_t - \theta_{t_0}\right).
\tag{3.24}
$$

In our particular problem the only test parameter is the intensity of the single molecule ($\theta_t = \theta_I$). This allows for a couple of simplifications and using these we can derive a simple equation for T_G in terms of the normal distribution. This is possible because we can use the fact that the square root of a random variable having a chi-square distribution with one degree of freedom is normally distributed [10]. However, we have to be careful: since T_G is always positive there are two possible values that could be the source of the obtained value of $z = \sqrt{T_G}$. Therefore, the probability of a false positive detection if a significance boundary γ is placed on T_G follows as:

$$
\begin{aligned}
P_{FA} &= P\left(T_G > \gamma ; H_0\right) \\
&= P(z > \sqrt{\gamma} ; H_0) + P\left(z \leq -\sqrt{\gamma}; H_0\right) \\
&= \frac{1}{\sqrt{2p}} \left[\int_{-\infty}^{-\sqrt{\gamma}} dz + \int_{\sqrt{\gamma}}^{\infty} dz\right] exp\left(-\frac{z^2}{2}\right) \\
&= 1 + erf\left(-\sqrt{\gamma/2}\right) \\
&= 2CDF\left(-\sqrt{\gamma}\right),
\end{aligned}
\tag{3.25}
$$

where CDF is the so called cumulative distribution function of the standard normal distribution function. Therefore, reversely, the significance boundary can be calculated from the false positive rate by:

$$
\gamma = CDF^{-1}(P_{FA}/2)^2
\tag{3.26}
$$

74

Similarly, the detection probability $T_G \sim \chi_1^2(\delta)$ can be transformed into $z = \sqrt{T_G} \sim N(\sqrt{\delta}, 1)$, where the detection probability can be calculated as:

$$
\begin{aligned}
P_D &= P\left(T_G > \gamma; H_1\right) \\
&= P(z > \sqrt{\gamma} - \sqrt{\delta}; H_1) + P\left(z \le -\sqrt{\gamma} - \sqrt{\delta}; H_1\right) \\
&= \frac{1}{\sqrt{2p}} \left[\int_{-\infty}^{-\sqrt{\gamma}} dz + \int_{\sqrt{\gamma}}^{\infty} dz\right] exp\left(-\frac{\left(z - \sqrt{\delta}\right)^2}{2}\right) \\
&= 1 - \frac{1}{2}\left[erf\left(\frac{-\sqrt{\delta} + \sqrt{\gamma/2}}{\sqrt{2}}\right) - erf\left(\frac{-\sqrt{\delta} - \sqrt{\gamma/2}}{\sqrt{2}}\right)\right].
\end{aligned}
\tag{3.27}
$$

Multiple comparison problem

Now we are able to calculate the false positive probability as a function of the outcome of T_G and based on this value we can make the decision if a pixel is significant or not, H_1 versus H_0, respectively. Recall that for instance a $P_{FA} = 0.05$ means that there is a 5% probability that the current value is a false positive (decide H_1, while actually the hypothesis H_0 is true).

For a single test a probability of 5% for a false positive might seem acceptable, however, if we perform many tests simultaneously the actual number of false positives can be extremely large. In our case we have to compute millions of tests ($256x256$ pixels for 1000 frames already results in more than 65 million hypothesis tests) which results in millions of false positives, which is not acceptable in practice. In the field of statistics this problem is known as the multiple comparison problem [11]. In the case of performing multiple hypotheses simultaneously, we would like to have a control on the probability that a positive declared test is false (#false positives/(#true positives+#false positives)), instead of the probability that a test gives a false positive (#false positives/#hypothesis tests).

False discovery rate control

There exist a number of approaches to overcome this multiple comparison problem and they are all based on adjusting the p-values for the number of hypothesis tests executed simultaneously. The method of choice to overcome the multiple comparison problem is the procedure for so-called False Discovery Rate (FDR) control devised by Benjamini and Hochberg [12]. The FDR is defined as the expected value of the proportion of false positives among total positives (FDP). This FDP is an (unobserved) random variable:

Probability-Based Particle Detection

$$FDP = \frac{V}{V + S},$$ (3.28)

where V and S are two random variables with outcome equal to the number of false positives and true positives, respectively. Therefore,

$$FDR = E[FDP] \leq a,$$ (3.29)

where the $E[\cdot]$ is the expectation operation and a is the value at which the FDR is controlled. The proof for independent tests can be found in reference [12], appendix A.

In our case the pixel tests are not independent because the PSF extends over multiple pixels. Therefore we apply a modified version of the FDR [13]. The procedure to control the FDR consists of the following steps:

1. The P_{FA} values for the pixels $1, 2, \cdots, m$ (P_1, ..., P_m) are ordered from smallest p-value to highest: $0 \leq P_{(1)} \ldots \leq P_{(m)}$.

2. Find the largest k for which $P_{(k)} \leq (k/m\, c(m))\, a$, where m is the total number of tests and $c(m) = \sum_{i=1}^{m} (1/i)$, and a is the FDR.

3. Finally, declare all $H_{(i)}$ significant for $i = 1 \ldots k$ and calculate the adjusted p-values using $P_{(k)}^* = (m\, c(m)/k)\, P_k$.

This procedure returns the adjusted false positive probabilities, $P_{(k)}^*$, and these adjusted probabilities are declared significant using a user defined significance level which equals the value at which the FDR is controlled. The probabilities below the target FP are the regions where single molecules are detected.

References

[1] R. P. Nieuwenhuizen, K. A. Lidke, M. Bates, D. L. Puig, D. Grünwald, S. Stallinga, and B. Rieger, *Measuring image resolution in optical nanoscopy,* Nature Methods **10**, 557 (2013).

[2] B. Zhang, J. Zerubia, and J.-C. Olivo-Marin, *Gaussian approximations of fluorescence microscope point-spread function models,* Applied Optics **46**, 1819 (2007).

[3] S. Stallinga and B. Rieger, *Accuracy of the gaussian point spread function model in 2D localization microscopy,* Optics Express **18**, 24461 (2010).

[4] J. C. R. Politz, R. A. Tuft, K. V. Prasanth, N. Baudendistel, K. E. Fogarty, L. M. Lifshitz, J. Langowski, D. L. Spector, and T. Pederson, *Rapid, diffusional shuttling of poly(A) RNA between nuclear speckles and the nu-*

cleoplasm, Molecular Biology of the Cell **17**, 1239 (2006).

[5] C. S. Smith, N. Joseph, B. Rieger, and K. A. Lidke, *Fast, single-molecule localization that achieves theoretically minimum uncertainty,* Nature Methods **7**, 373 (2010).

[6] L. J. Van Vliet, D. Sudar, and I. T. Young, *Digital fluorescence imaging using cooled charge-coupled device array cameras,* Cell Biology: A Laboratory Handbook , 109 (1998).

[7] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory* (Prentice Hall, 1998).

[8] F. Huang, T. M. Hartwich, F. E. Rivera-Molina, Y. Lin, W. C. Duim, J. J. Long, P. D. Uchil, J. R. Myers, M. A. Baird, W. Mothes, *et al.*, *Video-rate nanoscopy using sCMOS camera-specific single-molecule localization algorithms,* Na-ture Methods **10**, 653 (2013).

[9] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory* (Prentice Hall, 1993).

[10] N. Ravishanker and D. K. Dey, *A first course in linear model theory* (CRC Press, 2001).

[11] R. G. Miller Jr, *Normal univariate techniques,* in *Simultaneous Statistical Inference* (Springer, 1981) pp. 37–108.

[12] Y. Benjamini and Y. Hochberg, *Controlling the false discovery rate: A practical and powerful approach to multiple testing,* Journal of the Royal Statistical Society. Series B (Methodological) , 289 (1995).

[13] Y. Benjamini and D. Yekutieli, *The control of the false discovery rate in multiple testing under dependency,* Annals of Statistics , 1165 (2001).

3

Probability-Based Particle Detection

4

In Vivo Single Particle Imaging of Nuclear mRNA Export in Budding Yeast Demonstrates an Essential Role for Mex67p

An experiment is a question which science poses to Nature, and a measurement is the recording of Nature's answer.

MAX PLANCK

Carlas S. Smith
Azra Lari
Carina Derrer
Anette Ouwehand
Karsten Weis
David Grunwald
Ben Montpetit

Abstract

Dozens of mRNA export proteins have been identified; yet the spatial and temporal activities of these proteins and how they determine directionality of messenger ribonucleoprotein complex (mRNP) export from the nucleus remain largely undefined. Here the bacteriophage PP7 RNA-labeling system was used in Saccharomyces cerevisiae to follow single-particle mRNP export events with high spatial precision and temporal resolution. These data reveal that mRNP export, consisting of nuclear docking, transport and cytoplasmic release from a nuclear pore complex (NPC), is fast (∼200 ms) and that upon arrival in the cytoplasm mRNPs are frequently confined near the nuclear envelope. Mex67p functions as the principal mRNP export receptor in budding yeast. In a mex67-5 mutant, delayed cytoplasmic release from NPCs and retrograde transport of mRNPs was observed. This proves an essential role for Mex67p in cytoplasmic mRNP release and directionality of transport.

4.1. *Introduction*

In eukaryotes, the physical separation of transcription and translation by the nuclear envelope (NE) allows for additional modes of quality control and regulation to be imposed on the gene expression program, and necessitates transport of mRNA from the nucleus to the cytoplasm. Passage across the NE is predominantly mediated by the nuclear pore complex (NPC), which is composed of ~30 nucleoporin proteins (Nups) that together form a channel connecting the nucleoplasm and cytoplasm [1, 2]. To access this channel, each mRNA must be assembled into a ribonucleoprotein (RNP) complex with export factors (e.g. Mex67p) to allow an mRNP to dock to an NPC, translocate across the NE through the transport channel, and reach the cytoplasm [3]. During export, mRNPs undergo temporally and spatially ordered remodeling where certain proteins are removed prior to export (e.g. Yra1p) while others are removed upon arrival in the cytoplasm (e.g. Mex67p and Nab2p). This provides a mechanism to regulate and impose directionality on the transport process [4–9]. Overall, many proteins have been identified, that as part of an mRNP, are required for nuclear maturation, export, and cytoplasmic release [3, 4, 10, 11]. However, numerous questions remain regarding the assembly and composition of the mRNP and how each protein factor contributes to the export event.

Recent advances in imaging technology (e.g., camera sensitivity and microscope design) and methodology (e.g., RNA-tagging strategies) allow individual mRNAs to be visualized *in vivo*. The ability to image individual mRNPs in living cells has provided important insight into various aspects of mRNP export, including the first measurements of export kinetics [12–14]. Here, we have applied these methods to single-particle imaging of mRNP export in the budding yeast *Saccharomyces cerevisiae*. The availability of mutants that disrupt yeast mRNP export and NPC function allows us to investigate the kinetics and regulation of mRNP export at the level of a single RNA molecule. Using this approach, we tracked hundreds of mRNPs in living cells, analyzed the kinetics of mRNA export events, and characterized the role of the essential mRNA export factor Mex67p. Our results show that mRNP transport across the nuclear envelope inside the living cell is fast (~200 ms), well in agreement with prior findings [12, 14], and prove a critical role for Mex67p in cytoplasmic mRNP release and directional NPC transport.

4.2. *Results*

To observe the mRNP export process in living cells, we employed the bacterio-phage PP7 RNA-labeling system [15, 16]. Twenty-four copies of the PP7 operator sequence were inserted into the 3' UTR of the *GFA1* gene (*GFA1-PP7*), which can be fluorescently labeled when bound by the PP7 coat protein (PP7-CP) fused to YFP (Fig. 4.1a and 4.A.1a). *GFA1* is an essential gene involved in chitin synthesis [17], and was selected because its relatively low expression level was suitable for single-particle tracking [18]. Like most yeast genes, *GFA1* lacks introns [19, 20], and the *GFA1* mRNA is near the average mRNA length in yeast (2154 nt vs. ~1400 nt) [21]. The nucleoporin Ndc1p fused to tdTomato was also co-expressed with *GFA1-PP7* and PP7-CP-3xYFP, resulting in the reference (*REF*) strain, in which the position of a *GFA1* mRNP (PP7-CP-3xYFP multiplexed on the PP7 operator) could be determined relative to NPCs (Ndc1p-tdTomato) to track mRNP export (Fig. 4.1a).

Previous studies using PP7 and the related MS2 RNA-labeling approach in yeast and metazoan systems have demonstrated that the presence of stem-loops in the 3' UTR of a transcript does not alter transcript levels, localization, or RNP assembly [15, 22–25]. Importantly, the additional sequence in the *GFA1* 3' UTR and fusion of tdTomato to Ndc1p, both essential genes, did not impact growth of the *REF* strain (Fig. 4.A.1b), suggesting that the presence of the operator stem-loops, the binding of PP7-CP-3xYFP to the *GFA1-PP7* mRNA and tagging of Ndc1p does not significantly affect *GFA1* mRNP biogenesis or overall cellular fitness. We cannot rule out the possibility that addition of PP7 stem-loops might subtly affect the *GFA1* mRNA, as it was recently shown that the presence of MS2 stem-loops can interfere with the cytoplasmic degradation of an mRNA by Xrn1p (Garcia and Parker, 2015). However, by focusing on nuclear mRNA export events, our analysis selected for functional mRNPs that were transported across an NPC and were not recognized by nuclear surveillance machinery. To further verify that the PP7 stem-loops did not alter steady-state GFA1 transcript levels, the number of transcripts per cell was determined using single-molecule FISH to be 14 ± 7 in the parental strain (no PP7 stem-loops) and 13 ± 6 in the REF strain. This suggests that the presence of the PP7 stem-loops in the 3'UTR of GFA1 does not significantly alter steady-state mRNA expression levels (p = 0.37, two-tailed t-test, n=100 cells, Fig. 4.A.1c).

In the *REF* strain, PP7-CP-YFP positive particles were predominantly observed in the cytoplasm, as would be expected at steady state, and is consistent with our GFA1 single-molecule FISH data. The ability to observe particles was dependent on the presence of PP7 operator loops within the *GFA1* 3' UTR (Fig. 4.A.1a) and

Essential Export Factor for Nuclear mRNA

most particles were relatively uniform in size and brightness, but some cytoplasmic particles appeared brighter and larger. On occasion these large particles within the cytoplasm merged and split suggesting that they may contain multiple mRNAs (Fig. 4.A.1d). This could represent the accumulation of decay intermediates in P-bodies (Garcia and Parker, 2015) or other assemblies containing multiple mRNAs, which have recently been reported in live cultured neurons and *Saccharomyces cerevisiae* [24, 26]. Large assemblies were rarely observed in the nucleus of *REF* cells and were not observed to undergo mRNP export.

Measurements of mRNP export kinetics have shown that mRNA export occurs within a few hundred milliseconds (ms) and involves discrete steps that include NPC docking, translocation, and cytoplasmic release [12, 14]. Imaging the rapid dynamics of cellular processes, including mRNA export, at the single particle level presents a major challenge [27]. For instance, information from multiple channels (i.e. mRNA and NPC signals) must be collected concurrently at high frame rates in a manner that maximizes signal to noise ratios (SNR) and localization precision. To address this challenge, we employed an imaging setup capable of simultaneously capturing two-channel imaging data at high frame rates with precise image registration (see material and methods - *Live cell imaging of mRNP export and image processing)*[12]. Importantly, complications introduced by system drift or cellular movement (e.g. NPC mobility) is negated by our imaging setup that allows us to monitor the position of both the mRNP and Ndc1p in every frame at the same instance in time.

Still, imaging at the rate required to measure export kinetics (67 Hz in this study) limited photon collection, which combined with cellular background and light scatter introduced by the yeast cell wall diminished SNR. To overcome this issue, the yeast cell wall was removed and cells were imaged in medium containing sorbitol for osmotic support(Fig. 4.1b), which substantially increased SNR (see material and methods - *Calculation of signal improvements*). This resulted in reduced widths (s) of single-particle signals (PSF) (p < 0.01, Wilcoxon rank-sum test) and a 23% increase in localization precision (Fig. 4.1c). Cell wall removal also has the effect of inducing *GFA1* expression for the purpose of cell wall synthesis; consequently cells with labeled *GFA1* mRNPs became apparent within 15 minutes after reintroducing growth media and we were able to collect data for ~90 min before the newly forming cell wall increased the background due to light scatter.

Using this approach, we collected two-channel imaging data for 500 frames at 67 Hz from *REF* cells with a measured co-localization precision of 56 ± 20 nm between the two channels (see material and methods - *Registration and co-localization precision*). Due to the small size of yeast, use of a 1.3-numerical
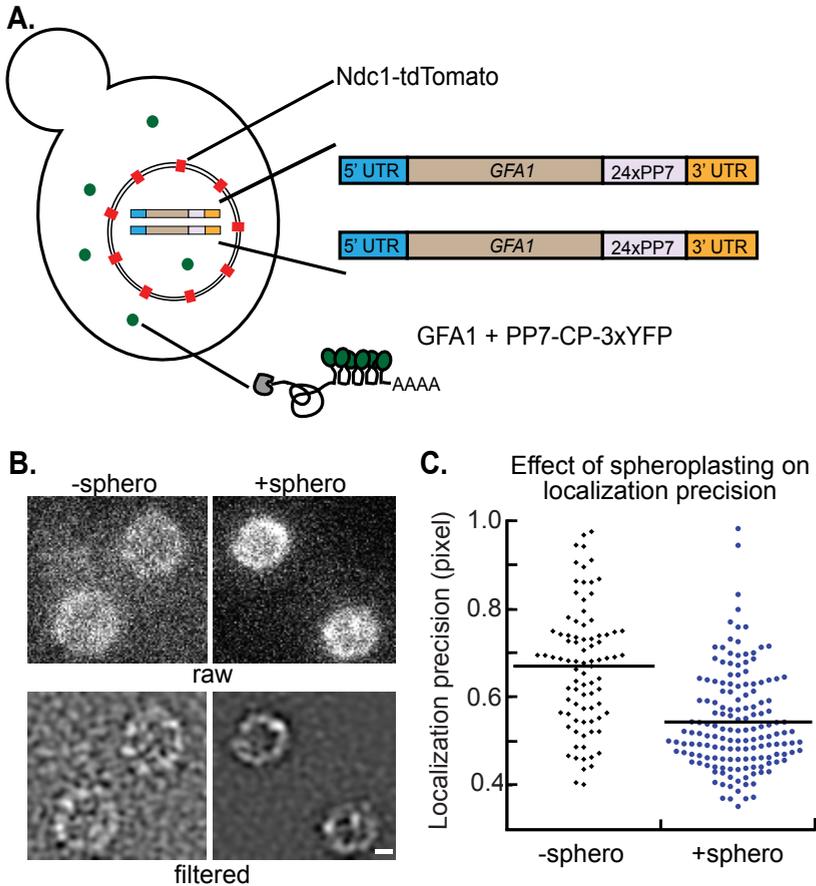
Figure 4.1. *REF* strain design and characterization. (A) Schematic displaying features of the yeast strains used to monitor mRNP export. Upon transcription the *GFA1* mRNA that carries 24 x PP7 loops in the 3'UTR is bound by the PP7-CP-3xYFP appearing as particles that can be tracked in relation to NPCs that are marked by Ndc1-tdTomato. (B) Fluorescent images of the Ndc1-tdTomato signal in *REF* cells using identical image acquisition settings showing the improvement in image quality after removal of the yeast cell wall. Examples of both raw and Laplacian filtered images are shown. Scale bar = 1 μm. (C) Dot plot displaying the localization precision (pixel = 96 nm) obtained when tracking mRNP particles in cells with (-sphero, n=86) and without a cell wall (+sphero, n=156) with the mean denoted by a black line for each.

	REF	mex67-5
Particles tracked	291	203
Nuclear docking events	47	23
Export events	43	9
Retrograde export events	0	7
Mean export time – dwell time analysis (ms)	188**	N.D.
Mean export time – MLE (ms)	215**	N.D.
Mean nuclear docking time during export – dwell time analysis (ms)	32**	362**
Mean nuclear docking time during export – MLE (ms)	39**	202**
Mean transition time during export – dwell time analysis (ms)	87**	406**
Mean transition time during export – MLE (ms)	99**	383**
Mean cytoplasmic docking time during export – dwell time analysis (ms)	62**	1258**
Mean cytoplasmic docking time during export – MLE (ms)	77**	943**
NE associated mRNPs per cell	0.2 ± 0.4*	0.5 ± 0.7*
Nuclear scanning events	4	2
Cytoplasmic scanning events	102	16

Table 4.1. Summary of GFA1-PP7 mRNP imaging data. * Standard deviation. **Export data underlying dwell time analysis may not have a normal distribution, consequently a MLE analysis was performed[28]. The differences between dwell times in REF and mex67-5 strains were tested using the distribution independent Wilcoxon-ranksum test and found to be significant at p<0.05 for the transition and cytoplasmic docking states. For more details and a discussion of error measurements see Material and Methods – Statistical analyses.

aperture objective allowed ~60% of the nuclear volume to be imaged in a single focal plane, thus mRNP particles could be tracked for significantly more frames than in mammalian cells [12].

In the dataset collected from ~450 *REF* cells, we identified 43 successful mRNP export events. Each event contained a tracked particle, which docked to the nuclear side of the NE and moved in successive frames from the interior of the nucleus to the cytoplasm where it was released (Fig. 4.2a). We classified mRNPs within every frame as being in one of the following states: nucleoplasmic, nuclear docked, transition between docked states, cytoplasmic docked, or cytoplasmic. Each state was assigned based on the distance between an mRNP and the NE, plus the dynamic behavior (i.e. the direction and distance the particle moved with

respect to the NE) of the particle in the preceding and subsequent frames (see material and methods - *Definition of transport states and data analysis*). Using these state values, the duration of an export event was calculated from the time of nuclear docking until mRNP release into the cytoplasm, as previously described ;[12, 29, 30]. Using these data, both dwell time analysis [29] and maximum likelihood estimate (MLE)[28], yielded total mean export times of 188 and 215 ms, respectively (Tab. 4.1 and 4.A.2, see material and methods - *Dwell time estimation*). These export times are similar to those reported for transcripts modified with the MS2 RNA-tagging system in murine cells and for unmodified transcripts labeled by fluorescently tagged mRNP proteins injected into insect salivary gland cells [12, 14]. Our findings, therefore, indicate that export in yeast and mammals occur on a similar time scale, consistent with a conservation of the NPC transport mechanism between these species.

During analysis of the *REF* dataset we noted that mobile GFA1 mRNPs would remain in close proximity to the NE and repeatedly enter a docked state, which was observed as confinement of mRNPs near the NE. To quantify this behavior we used the states defined in our tracking data and counted particles that docked at the NE multiple times each separated by a state of nucleoplasmic or cytoplasmic diffusion for less than 7 frames (105 ms). We termed this behavior "scanning". NE scanning was predominantly observed in the cytoplasm (n=102), but also occurred infrequently in the nucleus (n=4). NE scanning has also been observed for *MDN1*, *GLT1*, and *CLB2* mRNAs and was shown to be dependent on the nuclear basket components Mlp1p and Mlp2p (Saroufim et al. 2015, see accompanying paper in this issue). The low frequency of nuclear scanning suggests that docking of the *GFA1* mRNP to the NE often results in a productive NPC interaction and export. Indeed, we found that ~90% of the mRNPs observed to dock at the NE were successfully exported (Tab. 4.1). The functional significance of nuclear scanning remains unclear, but given the low frequency observed here, this may relate to mRNP quality control mechanisms occurring at NPCs, to limited access of an mRNP to a channel engaged with other cargos, or to differences in cellular status (Tutucci and Stutz, 2011; Bonnet and Palancade, 2014). In addition, ~35% of successfully exported mRNPs underwent cytoplasmic scanning directly following export. The observed NE scanning frequency after export is likely an underestimate because many tracked particles moved out of focus following arrival in the cytoplasm. Cytoplasmic scanning may therefore be intimately related to the export event. We could further distinguish two types of cytoplasmic scanning behaviors that confined the mRNP in close proximity to the NE. We observed particles that interacted with the same area of the NE ("local scanning", Fig. 4.2b)
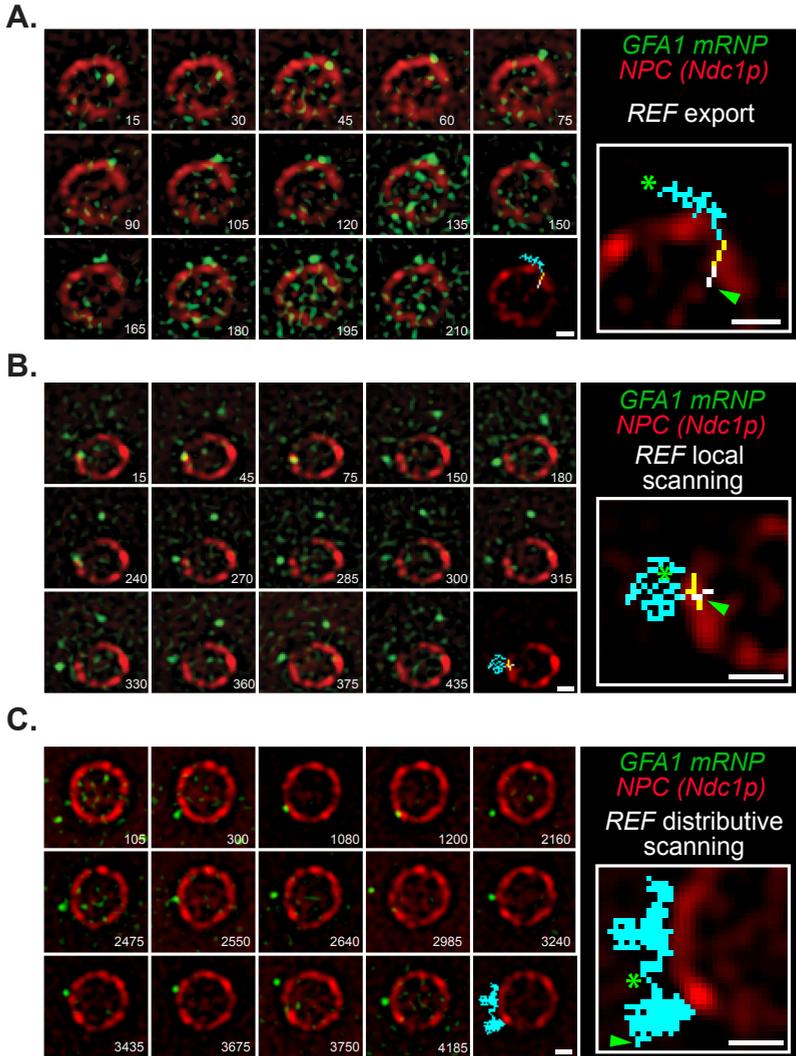
Figure 4.2. *GFA1* mRNP export and NE scanning in *REF* strain. (A) Merged and registered images show consecutive frames of a successful export event based upon tracking of the tagged *GFA1* mRNP across the NE in the *REF* strain. (B,C) Selected non-consecutive frames show local and distributive NE scanning interactions between an mRNP and the NE. For all panels cells were imaged at 26°C and 67 Hz with the time from the start of the event given in the bottom right of each image. The last image and inset show an overlay of the mRNP path that is colour coded based on position (white = nuclear docked, yellow = transition, and blue = cytoplasmic docked / cytoplasm). A green arrowhead and star denote mRNP position at the beginning and end of the track, respectively. Scale bar = 1 μm.

or particles that made contacts over a large area of the NE ("distributive scanning", Fig. 4.2c). Local NE scanning may reflect an mRNP remaining engaged with cytoplasmic Nups within the same NPC that facilitated transport. In this case, the apparent distance that the particle is able to travel into the cytoplasm and along the NE would be determined by the ~50 nm distance that NPC fibrils extend into the cytoplasm and the flexibility of an mRNA molecule [31, 32]. In contrast, distributive NE scanning involves distances far beyond 50 nm suggesting that an mRNP could make repeated contacts with the NE involving multiple NPCs (Fig. 4.2c). Both NE scanning behaviors are consistent with repeated interactions between the mRNP and the NE after translocation. Perhaps some export receptors are not immediately removed from mRNPs upon translocation and multiple NPC-mRNP interactions are required for complete remodeling and cytoplasmic release. It will be interesting to learn if mRNP remodeling can be distributed across multiple NPCs and need not occur solely at the NPC that facilitated export. Yet another possibility is that some instances of NE scanning reflect NPC interactions important for events downstream of mRNA export, as a function in translation was proposed for NPC-associated Dbp5p and the nucleoporin Gle1p [33–36].

Having established an imaging approach in yeast to measure mRNP export kinetics, we examined the impact of a mutant Mex67p protein on nuclear export dynamics of *GFA1* mRNPs. Mex67p (NXF1/TAP in metazoans) is an essential export factor that as part of the mRNP interacts with components of the NPC to facilitate export [7, 37–47]. Together these works support a model of export where: (1) multiple Mex67p molecules bind along the length of mRNA during nuclear maturation; (2) Mex67p promotes mRNP transport through the NPC by binding FG Nups; and (3) Mex67p dissociation from the mRNP on the cytoplasmic face of the NPC preventing further interactions with the NPC. Displacement of Mex67p would therefore impart directionality on nuclear mRNA export [48].

We initially imaged cells carrying the temperature sensitive *mex67-5* allele at the non-permissive temperature of 37°C [43], but mRNPs became static and reduced in number, precluding analysis of export. Consequently, we performed imaging at 26°C, which was also used to collect the *REF* dataset discussed above. At 26°C, the *mex67-5* strain did not have observable growth defects or mRNA export defects, as measured using an oligo-dT FISH assay to determine steady state mRNA localization (Fig. 4.A.2a and 4.A.2b). Using GFP tagged versions of Mex67p and Mex67-5p, we observed an increased nuclear pool of Mex67-5p at 26°C, but the majority remained at the nuclear envelope similar to Mex67p. This is in contrast to Mex67-5p localization at 37°C, where foci within both the nucleus

and cytoplasm formed (Fig. 4.A.2c) [43]. The mean number of *GFA1* mRNAs observed in the *mex67-5* strain (12 ± 5) by single-molecule FISH was significantly different (p < 0.05, two-tailed t-test, n=100 cells) than the *REF* strain (14 ± 6) (4.A.2d). The *mex67-5* mutation, therefore, impacts Mex67p localization and GFA1 mRNA levels, indicating that Mex67-5p has a partial loss-of-function at 26°C, but this does not result in significant changes in mRNA distributions at steady state.

By employing single-particle imaging, we observed that *GFA1* mRNPs in the *mex67-5* mutant at 26°C were 3-fold more frequently associated with the NE (Tab. 4.1). Data from all tracked mRNPs (i.e. independent of being part of a successful export event) were further used to estimate the length of time a particle spent in a single state (e.g., cytoplasmic docked) before transitioning to the next state (e.g., cytoplasmic). We found that mRNPs in the *mex67-5* strain persisted significantly longer in each of the states (p < 0.01, Wilcoxon rank-sum test; Fig. 4.3a). The number of cytoplasmic scanning events was also decreased in the *mex67-5* strain (Tab. 4.1), and we observed mRNPs in the cytoplasmic docked state for the entire duration of a movie (Fig. 4.3b). These data demonstrate that Mex67-5p alters the dynamics of mRNP-NPC binding interactions at 26°C, which can be quantified using our single particle imaging approach.

In the dataset collected from ~250 *mex67-5* cells, only 9 successful mRNP export events were identified (Fig. 4.4a, 4.4b). These ranged in length from 210 to 4080 ms and a comparison of successful export times from *REF* and *mex67-5* cells showed a clear bias (p < 0.01, Wilcoxon rank-sum test) towards long events in *mex67-5* cells (Fig. 4.4b). However, the low number of successful events prevented an accurate calculation of a mean export time. Dwell time analysis of the states occupied by successfully exported mRNPs showed that in *mex67-5* cells the lengthening of export times was due to a ~6-fold increase in nuclear docking times, a ~4-fold increase in transition times, and a ~20-fold increase in cytoplasmic docking times when compared to the duration of these states in *REF* export events (Tab. 4.1 and 4.A.2). The differences between REF and *mex67-5* strains for transition times (p < 0.01, Wilcoxon rank-sum test) and cytoplasmic docking times (p < 0.01, Wilcoxon rank-sum test) were statistically significant, while the six-time increase in nuclear docking times was not (p > 0.05, Wilcoxon rank-sum test). These data fit well with the observed increase in NE-associated mRNPs (Tab. 4.1), the low number of observed successful export events and the extended interactions of tracked mRNPs with the NE in *mex67-5* cells (Fig. 4.3). Of the nuclear mRNPs that docked with the NE in *mex67-5*, only 32% (9 of 23) were ultimately exported, which contrasted with the *REF* strain where 90% of nuclear docking events led to export (Tab. 4.1). Interestingly, from the 14 particles that did
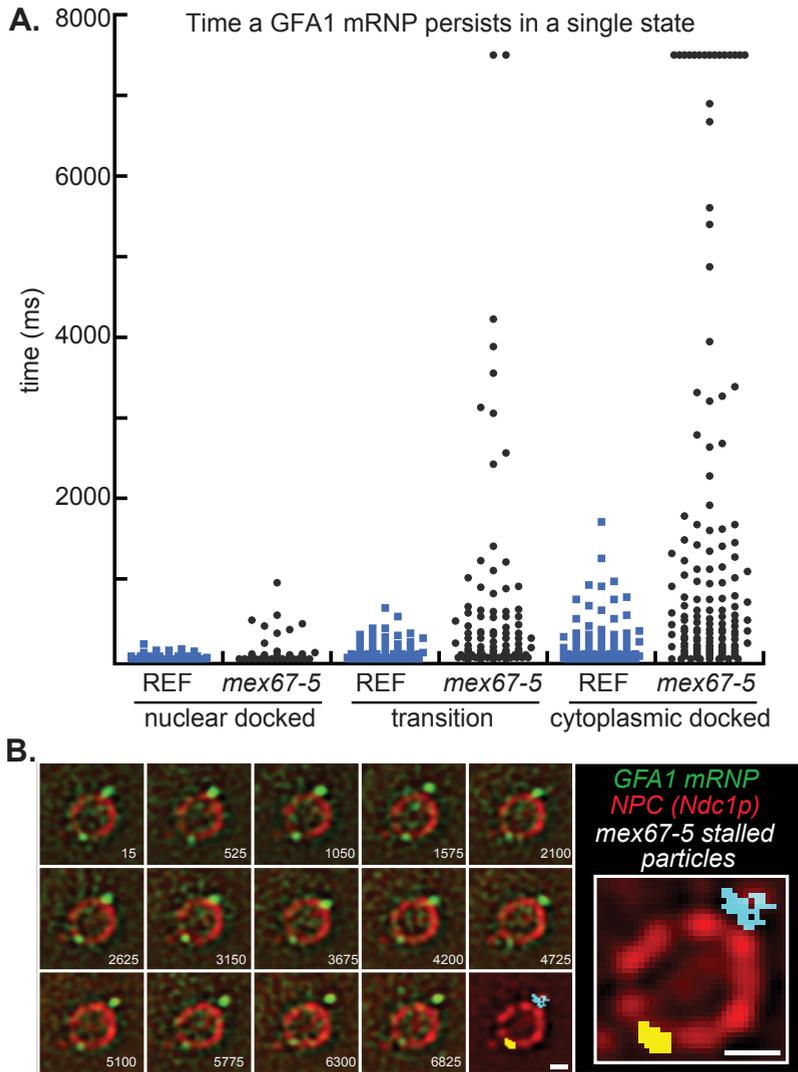
Figure 4.3. Prolonged *GFA1* mRNP interactions with the NE in *mex67-5*. (A) Dot plot displaying the length of time a *GFA1* mRNP persists in a single state (nuclear docked, transition, or cytoplasmic docked) in *REF* (n=49, 82, 108) and *mex67-5* (n=34, 94, 151) strains. Note that the data presented here uses all trace data where a particle interacted with the NE independent of the trace resulting in mRNP export. (B) Selected non-consecutive frames show the persistent interaction of *GFA1* mRNPs with the NE in *mex67-5* cells. Cells were imaged at 26°C and 67 Hz with the time from the start of the event given in the bottom right of each image. The last image and inset show an overlay of the mRNP path that is colour coded based on position (white = nuclear docked, yellow = transition, and blue = cytoplasmic docked / cytoplasm). Scale bar = 1 μm.

not export, 7 retrograde transport events were identified in which mRNPs traverse the NE and achieve a state of cytoplasmic docking before ultimately returning to the nucleoplasm (Fig. 4.4c). Retrograde mRNP transport was never observed in *REF* cells. Taken together, the 7 re-import events and the ~20-fold increase in cytoplasmic dwell times (Tab. 4.1 and 4.A.2) during successful export events strongly argue that the *mex67-5* mutant perturbs directional mRNP export and cytoplasmic release. The *mex67-5* allele contains a histidine to tyrosine amino acid substitution at position 400 and previous studies have shown that there is less mRNA associated with Mex67-5p as compared to Mex67p [7, 43]. Computational models of mRNA export suggest that the efficiency of mRNP export is highly sensitive to both the number and spacing of export factors along an mRNP [49]. As such, a reduced number and/or altered spacing of functional Mex67p molecules on the mRNA could lead to changes in mRNP architecture and altered interactions with the NPC that may affect how efficiently the mRNP is transported and remodeled by the NPC. This could explain the resulting inefficiencies and failures in mRNP export reported here, although further work will be required to characterize the molecular-basis of these transport defects.

4.3. *Discussion*

Overall, our work has measured mRNP export kinetics for the first time in *S. cerevisiae* by taking advantage of two powerful experimental systems: single-particle RNA localization microscopy and yeast genetics. Importantly, this approach provides a platform upon which to address fundamental questions related to nuclear mRNA export, including kinetics, regulation, and mechanism(s) of transport through nuclear pore complexes. For example, we have observed confinement of mRNPs near the NE after arrival in the cytoplasm, suggesting that mRNPs may visit multiple NPCs for the purpose of mRNP remodeling following export. These interactions might function to remodel mRNPs for cytoplasmic release or subsequent events in the mRNA life cycle (e.g., translation). We have further provided direct evidence for the critical role of Mex67p in imparting directionality to mRNP export. Future applications of our approach will include interrogating the many other protein involved in mRNP export, including Nups; studies that will be facilitated by the tractable genetics and the expansive knowledge of mRNP export and NPC function in yeast.

Figure 4.4. mRNP export kinetics and retrograde transport in *mex67-5*. (A) Merged and registered images show select frames of a successful export event in a *mex67-5* cell. (B) Dot plot showing the distribution of *GFA1* mRNP export times in *REF* (n=43) and mex67-5 (n=9) cells with the mean denoted by a black line. (C) Selected non-consecutive frames show a retrograde transport event in which an mRNP on the cytoplasmic side of the NE returns to the nucleus. For A and C, cells were imaged at 26°C and 67 Hz with the time from the start of the event given in the bottom right of each image. The last image and inset show an overlay of the mRNP path that is colour coded based on position (white = nuclear docked, yellow = transition, and blue = cytoplasmic docked / cytoplasm). A green arrowhead and star denote mRNP position at the beginning and end of the track, respectively. Scale bar = 1 μm.

References

[1] S. R. Wente and M. P. Rout, *The nuclear pore complex and nuclear transport,* Cold Spring Harbor Perspectives in Biology **2**, a000562 (2010).

[2] A. G. Floch, B. Palancade, and V. Doye, *Fifty years of nuclear pores and nucleocytoplasmic transport studies: Multiple tools revealing complex rules.* Methods in Cell Biology **122**, 1 (2013).

[3] C. Nino, L. Hérissant, A. Babour, and C. Dargemont, *mRNA nuclear export in yeast,* Chemical Reviews **113**, 8523 (2013).

[4] M. Müller-McNicoll and K. M. Neugebauer, *How cells get the message: Dynamic assembly and function of mRNA-protein complexes,* Nature Reviews Genetics **14**, 275 (2013).

[5] A. Köhler and E. Hurt, *Exporting RNA from the nucleus to the cytoplasm,* Nature reviews Molecular Cell Biology **8**, 761 (2007).

[6] M. Oeffinger and D. Zenklusen, *To the pore and through the pore: A story of mRNA export kinetics,* Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms **1819**, 494 (2012).

[7] M. K. Lund and C. Guthrie, *The DEAD-box protein Dbp5p is required to dissociate Mex67p from exported mRNPs at the nuclear rim,* Molecular Cell **20**, 645 (2005).

[8] E. J. Tran, Y. Zhou, A. H. Corbett, and S. R. Wente, *The DEAD-box protein Dbp5 controls mRNA export by triggering specific RNA: Protein remodeling events,* Molecular Cell **28**, 850 (2007).

[9] N. Iglesias, E. Tutucci, C. Gwizdek, P. Vinciguerra, E. Von Dach, A. H. Corbett, C. Dargemont, and F. Stutz, *Ubiquitin-mediated mRNP dynamics and surveillance prior to budding yeast mRNA export,* Genes & development **24**, 1927 (2010).

[10] A. Bonnet and B. Palancade, *Regulation of mRNA trafficking by nuclear pore complexes,* Genes **5**, 767 (2014).

[11] M. Oeffinger and B. Montpetit, *Emerging properties of nuclear RNP biogenesis and export,* Current Opinion in Cell Biology **34**, 46 (2015).

[12] D. Grünwald and R. H. Singer, *In vivo imaging of labelled endogenous β-actin mRNA during nucleocytoplasmic transport,* Nature **467**, 604 (2010).

[13] A. Mor, S. Suliman, R. Ben-Yishay, S. Yunger, Y. Brody, and Y. Shav-Tal, *Dynamics of single mRNP nucleocytoplasmic transport and export through the nuclear pore in living cells,* Nature Cell Biology **12**, 543 (2010).

[14] J. P. Siebrasse, T. Kaminski, and U. Kubitscheck, *Nuclear export of single native mRNA molecules observed by light sheet fluorescence microscopy,* Proceedings of the National Academy of Sciences **109**, 9426 (2012).

[15] S. Hocine, P. Raymond, D. Zenklusen, J. A. Chao, and R. H. Singer, *Single-molecule analysis of gene expression using two-color RNA labeling in live yeast,* Nature Methods **10**, 119 (2013).

[16] J. A. Chao, Y. Patskovsky, S. C. Almo, and R. H. Singer, *Structural basis for the coevolution of a viral RNA–protein complex,* Nature structural and Molecular Biology **15**, 103 (2008).

[17] G. Watzele and W. Tanner, *Cloning of the glutamine: Fructose-6-phosphate amidotransferase gene from yeast. pheromonal regulation of its transcription.* Journal of Biological Chemistry **264**, 8753 (1989).

[18] D. Lipson, T. Raz, A. Kieu, D. R. Jones, E. Giladi, E. Thayer, J. F. Thompson, S. Letovsky, P. Milos, and M. Causey, *Quantification of*

the yeast transcriptome by single-molecule sequencing, Nature Biotechnology **27**, 652 (2009).

[19] M. Spingola, L. Grate, D. Haussler, and M. Ares, *Genome-wide bioinformatic and molecular analysis of introns in Saccharomyces cerevisiae.* RNA **5**, 221 (1999).

[20] J. R. Rodriguez-Medina and B. C. Rymond, *Prevalence and distribution of introns in non-ribosomal protein genes of yeast,* Molecular and General Genetics MGG **243**, 532 (1994).

[21] E. H. Hurowitz and P. O. Brown, *Genome-wide analysis of mRNA lengths in saccharomyces cerevisiae,* Genome Biology **5**, R2 (2004).

[22] E. Bertrand, P. Chartrand, M. Schaefer, S. M. Shenoy, R. H. Singer, and R. M. Long, *Localization of ash1 mRNA particles in living yeast,* Molecular Cell **2**, 437 (1998).

[23] L. Haim, G. Zipor, S. Aronov, and J. E. Gerst, *A genomic integration method to visualize localization of endogenous mRNAs in living yeast,* Nature Methods **4**, 409 (2007).

[24] H. Y. Park, H. Lim, Y. J. Yoon, A. Follenzi, C. Nwokafor, M. Lopez-Jones, X. Meng, and R. H. Singer, *Visualization of dynamics of single endogenous mRNA labeled in live mouse,* Science **343**, 422 (2014).

[25] T. Song, Y. Zheng, Y. Wang, Z. Katz, X. Liu, S. Chen, R. H. Singer, and W. Gu, *Specific interaction of KIF11 with ZBP1 regulates the transport of β-actin mRNA and cell motility,* Journal of Cell Science **128**, 1001 (2015).

[26] C. E. Simpson, J. Lui, C. J. Kershaw, P. F. Sims, and M. P. Ashe, *mRNA localization to Pbodies in yeast is biphasic with many mRNAs captured in a late Bfr1p dependent wave,* Journal of Cell Science **127**, 1254 (2014).

[27] Z. Liu, L. D. Lavis, and E. Betzig, *Imaging live-cell dynamics and structure at the single-molecule level,* Molecular Cell **58**, 644 (2015).

[28] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory* (Prentice Hall, 1993).

[29] U. Kubitscheck, D. Grünwald, A. Hoekstra, D. Rohleder, T. Kues, J. P. Siebrasse, and R. Peters, *Nuclear transport of single molecules dwell times at the nuclear pore complex,* The Journal of Cell Biology **168**, 233 (2005).

[30] T. Dange, D. Grünwald, A. Grünwald, R. Peters, and U. Kubitscheck, *Autonomy and robustness of translocation through the nuclear pore complex: A single-molecule study,* The Journal of Cell Biology **183**, 77 (2008).

[31] B. Fahrenkrog, E. C. Hurt, U. Aebi, and N. Panté, *Molecular architecture of the yeast nuclear pore complex: Localization of Nsp1p subcomplexes,* The Journal of Cell Biology **143**, 577 (1998).

[32] F. Alber, S. Dokudovskaya, L. M. Veenhoff, W. Zhang, J. Kipper, D. Devos, A. Suprapto, O. Karni-Schmidt, R. Williams, B. T. Chait, *et al.*, *The molecular architecture of the nuclear pore complex,* Nature **450**, 695 (2007).

[33] T. A. Bolger, A. W. Folkmann, E. J. Tran, and S. R. Wente, *The mRNA export factor gle1 and inositol hexakisphosphate regulate distinct stages of translation,* Cell **134**, 624 (2008).

[34] T. A. Bolger and S. R. Wente, *Gle1 is a multifunctional DEAD-box protein regulator that modulates Ded1 in translation initiation,* Journal of Biological Chemistry **286**, 39750 (2011).

[35] A. R. Alcázar-Román, T. A. Bolger, and S. R. Wente, *Control of mRNA export and translation termination by inositol hexakisphosphate requires specific interaction with Gle1,* Journal of Biological Chemistry **285**, 16683 (2010).

[36] T. Gross, A. Siepmann, D. Sturm, M. Windgassen, J. J. Scarcelli, M. Seedorf, C. N. Cole, and H. Krebber, *The DEAD-box RNA helicase Dbp5 functions in translation termination,* Science **315**, 646

(2007).

[37] F. Stutz, A. Bachi, T. Doerks, I. C. Braun, B. Séraphin, M. Wilm, P. Bork, and E. Izaurralde, *REF, an evolutionary conserved family of hnRNP-like proteins, interacts with TAP/Mex67p and participates in mRNA nuclear export.* RNA **6**, 638 (2000).

[38] J. Katahira, K. Sträßer, A. Podtelejnikov, M. Mann, J. U. Jung, and E. Hurt, *The Mex67p-mediated nuclear mRNA export pathway is conserved from yeast to human,* The EMBO Journal **18**, 2593 (1999).

[39] H. Santos-Rosa, H. Moreno, G. Simos, A. Segref, B. Fahrenkrog, N. Panté, and E. Hurt, *Nuclear mRNA export requires complex formation between Mex67p and Mtr2p at the nuclear pores,* Molecular and Cellular Biology **18**, 6826 (1998).

[40] P. Grüter, C. Tabernero, C. von Kobbe, C. Schmitt, C. Saavedra, A. Bachi, M. Wilm, B. K. Felber, and E. Izaurralde, *TAP, the human homolog of Mex67p, mediates CTE-dependent RNA export from the nucleus,* Molecular Cell **1**, 649 (1998).

[41] E. Hurt, K. Sträßer, A. Segref, S. Bailer, N. Schlaich, C. Presutti, D. Tollervey, and R. Jansen, *Mex67p mediates nuclear export of a variety of RNA polymerase II transcripts,* Journal of Biological Chemistry **275**, 8361 (2000).

[42] K. Sträßer, J. Baßler, and E. Hurt, *Binding of the Mex67p]/Mtr2p heterodimer to FXFG, GLFG, and FG repeat nucleoporins is essential for nuclear mRNA export,* The Journal of Cell Biology **150**, 695 (2000).

[43] A. Segref, K. Sharma, V. Doye, A. Hellwig, J. Huber, R. Lührmann, and E. Hurt, *Mex67p, a novel factor for nuclear mRNA export, binds to both poly(A)+ RNA and nuclear pores,* The EMBO Journal **16**, 3256 (1997).

[44] J. P. Rodrigues, M. Rode, D. Gatfield, B. J. Blencowe, M. Carmo-Fonseca, and E. Izaurralde, *REF proteins mediate the export of spliced and unspliced mRNAs from the nucleus,* Proceedings of the National Academy of Sciences **98**, 1030 (2001).

[45] A. Bachi, I. C. Braun, J. P. Rodrigues, N. Panté, K. Ribbeck, C. von Kobbe, U. Kutay, M. Wilm, D. Gorlich, M. Carmo-Fonseca, and E. Izaurralde, *The C-terminal domain of TAP interacts with the nuclear pore complex and promotes export of specific CTE-bearing RNA substrates,* RNA **6**, 136 (2000).

[46] A. C. Tuck and D. Tollervey, *A transcriptome-wide atlas of RNP composition reveals diverse classes of mRNAs and lncRNAs,* Cell **154**, 996 (2013).

[47] C. Baejen, P. Torkler, S. Gressel, K. Essig, J. Söding, and P. Cramer, *Transcriptome maps of mRNP biogenesis factors define pre-mRNA recognition,* Molecular Cell **55**, 745 (2014).

[48] M. Stewart, *Ratcheting mRNA out of the nucleus,* Molecular Cell **25**, 327 (2007).

[49] M. Azimi, E. Bulat, K. Weis, and M. R. Mofrad, *An agent-based model for mRNA export through the nuclear pore complex,* Molecular Biology of the Cell **25**, 3643 (2014).

4

Appendix

4.A. *Materials and Methods*

Yeast strain construction

Strains and plasmids used in this study are listed in Tab. 4.A.2 and 4.A.1. To generate imaging strains, a set of *24xPP7* stem-loops with a KanMX selectable marker flanked by loxP sites was integrated into the 3' UTR of the *GFA1* gene using pDZ417 [1] in the diploid yeast strain BY4743 (BMY008). Cre recombinase was expressed from pSH47 [2] to remove the selectable marker and restore the 3'UTR with the exception of the PP7 loops and a single loxP site. NDC1 was c-terminal tagged in the *GFA1-PP7* heterozygous diploid with tdTomato [3] followed by sporulation and tetrad dissection to isolate a haploid of each mating type that carried *GFA1-PP7* and the Ndc1-tdTomato fusion. The *mex67-5* allele was subsequently integrated into the genome of each haploid using a PCR-based homologous recombination approach, which was confirmed by PCR, temperature sensitivity, and an mRNA export phenotype. Haploids were then mated to form diploids homozygous for the PP7 loops and NDC1 fusion with and without the *mex67-5* allele. Finally, to allow for the visualization of the PP7 containing *GFA1* transcripts, a pRS313-P_{Met25}PP7-CP-3xYFP plasmid (pBM242) was introduced into the diploid strain to generate the *REF* (BMY083) and *mex67-5* (BMY135) strains. To assess growth after PP7 stem-loop addition, growth rates were measured for a control strain with no PP7 loops (BMY642) and the *REF* strain (BMY83). Strains were grown overnight in a 24-well plate format with shaking at 26°C and O.D$_{600}$ measurements were carried out using the CLARIOstar plate reader (BMG Labtech). The *mex67-5* strain (BMY135) was also assessed for growth defects in comparison to a control strain (BMY129).

GFA1 transcript counting and mRNP NE association

Yeast were grown O/N at 26°C, diluted to an O.D$_{600}$ of ~0.1 the next morning, and grown at 26°C to allow at least three doublings. For transcript counting using smFISH, cells were fixed and GFA1 mRNAs were detected using 48 probes (20mers) directed against GFA1 (BioSearch Technologies) in strains with (BMY83) and without PP7 stem-loops (BMY642) as previously described with the noted

changes [1]. Briefly, cells were fixed by the addition of 37% formaldehyde to the cultures (3.7% final concentration) for 30minutes at 26°C. Yeast cell walls were digested with Zymolyase (CedarLane) and spheroplasted cells were applied to 8-well slides coated with poly-L lysine. Cells were permeabilized using ice-cold methanol for 6 minutes followed by ice-cold acetone for 30 seconds. Following rehydration and incubation with hybridization buffer for 1 hour at 37°C, 30 ul of hybridization buffer containing 20 ng of the GFA1 probes was added to each well and incubated overnight at 37°C. Wells were then washed and mounting media containing DAPI was added followed by imaging on a DeltaVision Elite microscope equipped with a Front Illuminated sCMOS camera driven by Softworx 6 (GE Healthcare) at 23°C using an Olympus 60x 1.4 N.A. oil objective.

To determine GFA1 mRNP subcellular localization with respect to the NE, REF and *mex67-5* strains were fixed in 2% paraformaldehyde for 15min, washed with media, and spheroplasted (as described below). GFA1 transcript number was determined by manually counting GFA1 foci and the frequency of NE-associated mRNPs was determined by scoring co-localization between the mRNP (YFP) and NPC (tdTomato) signals. Imaging of both datasets was performed on a DeltaVision Elite microscope equipped with a Front Illuminated sCMOS camera driven by Softworx 6 (GE Healthcare) at 23°C using an Olympus 60x 1.4 N.A. oil objective. Prior to analysis, images were deconvolved in Softworx 6 and processed in ImageJ. Specifically, images were adjusted for brightness/contrast, background subtraction was performed, and a gaussian blur 3D filter was applied.

Live cell imaging of mRNP export and image processing
The overall system design and methodology for imaging was as previously described with the exceptions noted below [4]. Briefly, imaging was performed on a custom dual channel set-up using an Olympus 60× 1.3 N.A. silicone oil immersion objective (R.I. 1.405) combined with 500 mm focal length tube lenses resulting in an effective 167x magnification and 95.8 nm sized pixel with the emission split in the primary beam path onto two EMCCDs (Andor iXon, Model DU897 BI). For excitation of fluorescent proteins solid-state 514 and 561-nm laser lines (Cobolt SE) were used, intensity and on/off were controlled by an AOTF (AA Opto-electronics). Simultaneous imaging of NPCs and mRNPs was performed using sub-frames (~2/5 of each chip, 200×200 pixel) on both cameras at a frame rate of 67 Hz equaling a time resolution of 15 ms.

Prior to imaging, cells were grown O/N at 26°C in synthetic complete media lacking histidine (SC - HIS) with methionine at 150 mg/L, diluted to an $O.D_{600}$ of 0.1 the next morning, and grown at 26°C to allow at least three doublings. To

remove the cell wall, ~2 ODs of cells were collected by centrifugation, washed with water, and re-suspended in 50 mM Tris pH 9.5 and 10 mM DTT solution at room temperature for 15min. Cells were then collected and re-suspended in 0.3 ml spheroplast buffer (150 mM KPO_4 pH=7.5, 1 mM $MgCl_2$, and 250 μg/ml Zymolyase) and incubated at 26°C for 45 minutes. Cells were then placed in 35-mm glass-bottom dishes (MatTek), coated with Concanavalin A (Sigma), and centrifuged at 500xg for 3min to adhere cells. Un-adhered cells were removed by washing with media containing 1.2 M sorbitol and left to recover in 2 ml of fresh media for 30 minutes at 26°C prior to imaging. For each cell four data sets, two in each color, were acquired. First, a registration image was recorded for 375 ms, automatically saved, and 500 ms later the tracking data set was recorded for 7.5 sec (500 frames) in parallel for both channels.

All image processing for visual analysis was done using FIJI [5]. First, each data-set was tested for drift during acquisition by creating 10 average projections of 50 frames of the NPC channel movie that were normalized and fused into a color-coded hyper-stack. Color separation in the resulting stack indicated drift and these data sets were discarded. Second, in the registration images the td-Tomato signal of Ndc1 was made visible in both channels using 'cross-talk on demand' by using 10x more excitation power from the 561 nm laser than for the tracking movies. Using the sensitivity of our EMCCD cameras and the surface reflection of the dichroic, the NPC signal was visible in both the mRNA (one image taken for 375 ms) and the NPC channel (25 images in 375 ms, average time projected for analysis). The mRNA and NPC signals were fine-registered post-experimentally by shifting the NPC channel registration image onto the mRNA channel registration image to calculate the parameters to be used for registration of the tracking movies [6]. To make this alignment more robust, the mRNA channel registration image was filtered with a Gaussian kernel (1.5 pixel width) before registration. RGB images of the two registration images before and after registration were saved and visually compared if the correlation factor of the linear shift was better than 0.95. Registration failed at a frequency of ~50%, which we attribute to aberrations caused by heterogeneity in spheroplasting, and at later times, due to rebuilding of the cell wall. The resulting registration precision was determined to be 0.14 pixel, corresponding to 14 ± 17 nm. After these initial quality checks we created two copies of each tracking movie (NPC and mRNA channel), one being the raw data for quantitative image analysis and the other being enhanced for visual inspection. Raw data images were always displayed next to the enhanced images during subsequent visual analysis and all traces of interest were double checked in the raw data set to prevent 'false-positive' event

identification due to image processing. Tracking was done as described [4] using a supported fit routine were signals were identified visually in either the filtered or raw image and the routine would execute a CoM within 5 pixel around the click position to identify the coordinate for a 2D Gaussian fit. All fits were done in raw data and all fit parameter and initiation parameter reported to the user. For enhancement of images for visual inspection we used running average and a subtraction of Laplacian filter for the NPC channel and a Laplacian filter for the RNA channel. The kernel size was set relative to the theoretical width of the emission PSF and contrast was adjusted in the final RGB movies after processing. After filtering, the transition matrix was applied to the NPC channel movie to overlay it onto the mRNA channel movie.

Calculation of signal improvements

We quantify an increase or decrease in signal improvement as a change in the mean estimated localization precision. The smallest localization precision possible can be calculated using the Cramer Rao lower Bound (CRLB) and is attained using MLE estimation [7]. The CRLB depends on the width of the point spread function, the intensity of the single molecule and the background fluorescence. To quantify the signal improvement we estimated these parameters, in addition to assessing the location of each mRNP, and calculated the corresponding CRLB [7]. These calculations were performed based on mRNP signals from cells with (n=86) and without (n=156) a cell wall and the localization precision was determined to be 64 ± 13 vs. 52 ± 11 nm before and after cell wall removal. Note that total improvement in localization precision was 23% and the shape of the localization distributions significantly changed (Fig. 4.1c).

Registration and co-localization precision

The registration precision between channels of 14 nm was determined by calculating the remaining offset between the registration data after linear translation. We chose the rather conservative linear registration model as the quality of the registration data does not reach the level of individual pores [4]. The standard deviation using this method is in the order of the mean. As a result we do not report spatial binding site distributions, arguing that our total measurement precision is similar to the width of the expected binding site distributions. For kinetic analysis, we employ five classification states, but it is important to note that a four-state model also strongly supports our findings. Colocalization precision is given by the square root of the sum of the squared localization (54 nm) and registration precision (14 nm) values and is 56 nm. From binding profiles of β-actin mRNA at

the NPC we know the peak binding sites on the cytoplasmic and nucleoplasmic surface for mRNA transporting within ~ 200 ms to be 275 nm apart [4].

Definition of transport states and data analysis

Data analysis was peformed using a manual-tracking interface in which the filtered and raw data were presented simultaneously and a particle of interest tracked by consecutive clicking through image frames. The maximal displacement from frame to frame was displayed in the tracking channel to identify situations were two particles could be interchanged. In such cases tracking was ended and the track dismissed. During manual tracking a descriptive state was assigned to the particle in each frame based on the distance from the NE using the following guidelines: nuclear/cytoplasmic diffusion if the distance was > 250 nm, nuclear/cytoplasmic docked if the distance was between 250 and 100 nm, and transition if the distance was < 100 nm. The dynamic behavior of the particle (i.e. the direction and distance the particle moved with respect to the NE) in prior and subsequent frames was also used to inform state decisions. Using these descriptors, analysis was performed in Matlab using routines to search for specific events (e.g. export or scanning) based on the 5 states defined above. We can make this classification because the localization precision of single molecules follows a Gaussian distribution described by $\theta - \hat{\theta} \sim N(0, C(\theta))$ where $\theta = (x, y, I, bg)$, $\hat{\theta}$ the corresponding MLE, and $C(\theta)$ is CRLB [8]. Using our current techniques we have a total colocalization precision of 56 nm. C_{do} and N_{do} are 275 nm apart and therefore we can calculate the false classification probability of a C_{do} being a N_{do} event (and the other way around) as $P(C_{do} | N_{do}) = 1 - normcdf(x = 135, \mu = 0, \sigma = 56)$ that defines the false classification rate as 0.01, or an 0.5% error to each side. When using $P(T|N_{do}, C_{do})$ this increases to 0.23 or an 11.5% error to each side $(2(1 - normcdf(x = 135/2, \mu = 0, \sigma = 56)))$. For a distance of 200 nm between the peak potions $P(T|N_{do}, C_{do})$ this becomes 37% compared to 7% for $P(C_{do}|N_{do})$. In other words, we are able to describe a two-state (C_{do}-N_{do}) model at the NPC (four states in total with the diffusive nuclear and cytoplasmic states) with very high confidence (less than 0.5% error), while for a three state model (C_{do}-T-N_{do}) at the NPC (5 states in total) the classification of the transition state has a error probability between 10 to 17% on each side based on our obtained localization precision. As the transition state T is a shift in between docking states its identification is partially based on our knowledge about the past and future of the particle within the trace. Therefore, we have included the transition state description, but note the related error, which does not impact the major findings of this work related to cytoplasmic docking differences in *mex67-5*.

Dwell time estimation

Due to the limited number of observations we estimated dwell times using two methods, the dwell time fit based on the histogram (exponential distribution) and a maximum likelihood estimate based on the assumption that the data follows an exponential distribution [9, 10]. In the first method: (i) a histogram is constructed from all the observed dwell times, (ii) the histogram is smoothed using a uniform filter having a width of 10 frames, (iii) inverted cumulative distribution is constructed [10] and (iv) a least squares fit is performed on the histogram. The second approach is a maximum-likelihood-estimation (MLE) [8]. The transport times per condition are assumed to be an independently identically set of random variables having an exponential distribution. The probability density function of observing a dwell time x_j is given by $f(x_j; \lambda) = \begin{cases} = \lambda\, e^{-\lambda\, x_j} & x \geq 0 \\ 0 & x < 0 \end{cases}$, where the average dwell time is equal to λ^{-1}. The likelihood of a sequence of observed dwell times is given by $L(\lambda) = \prod_{j=1}^{n} \lambda e^{-\lambda x_j}$, and the value for $\hat{\lambda}$ that maximizes the likelihood is given by $\hat{\lambda} = \frac{n}{\sum_{j=1}^{n} x_j}$. In both cases, based on the data having an exponential distribution, standard deviation is equal to the mean.

Mex67-5 strain characterization

To assay for mRNA export defects, fluorescent in situ hybridization against poly(A)-RNA was performed as previously described [11]. Briefly, REF (BMY83) and *mex67-5* (BMY135) strains were grown to mid-log phase at permissive temperature (26°C) and then shifted to non-permissive (37°C) temperature for 30 minutes with pre-warmed media. Following fixation, poly(A)-RNA was detected using a fluorescein-labeled dT_{50} probe and DNA was visualized using DAPi. Imaging was performed on a DeltaVision Elite microscope equipped with a Front Illuminated sCMOS camera driven by Softworx 6 (GE Healthcare) at 23°C using an Olympus 60x 1.4 N.A. oil objective. To localize Mex67p, haploid strains were generated (KWY5566 and KWY5567) expressing Ndc1p-tdTomato, *GFA1-PP7*, and GFP tagged Mex67. To avoid cross talk from the PP7-CP tagged with YFP, we used strains that did not express the coat protein. Cells were grown in synthetic complete medium at 26°C and then imaged in a 384 well plate coated with Concanavalin A at 26°C using an inverted epi-fluorescence microscope (Nikon Ti) equipped with a Spectra X LED light source and a Hamamatsu Flash 4.0 scMOS camera using a 100x Plan-Apo objective NA 1.4 and the NIS Elements software. All image processing was done using FIJI [5].

Statistical tests applied

Reported p-values were calculated using either t-tests or a Wilcoxon rank-sum test. The latter can be used in place of t-tests when it cannot be assumed that the population is distributed normally [12]. For all results the test used is stated in the text.

In analyzing transition times across the NE, the shape of the distribution determines the interpretation of the SD or SEM reported. While often associated with the symmetric interval of errors around normal distributed data the SD or SEM can also be used to report on non-normal distributed data. In this case the interpretation is based on Chebyshev's theorem specifying that no more than $1/k^2$ fraction of values can be more than k standard deviations away from the mean. In case of an exponential distribution this translates into the standard deviation being equal to the mean. In our case, the cumulative distribution of translocation times is equal to the cumulative distribution function of an exponential distribution, as expected [9, 10], or arrival time distributions in general. Due to the limited number of observations we estimate the dwell time using two methods, the dwell time fit based on the histogram (exponential distribution) and a maximum likelihood estimate based on the assumption of the data following an exponential distribution. For exponential distributions the standard deviation is expected to be equal to the mean. We like to point out that a very limited amount of data (as indicated) underlying some of the values in Tab. 4.1 and 4.A.2 and that the error reported in Tab. 4.A.2 is the error of the measurement.

A.

No loops   GFA1::24xPP7

B.

O.D.$_{600}$

- No loops
- *REF*

Comparison of growth rates at 26°C

Time (min)

C.

GFA1 transcripts per cell

Number of GFA1 foci per cell

No loops   *GFA1::24xPP7*

D.

15   165   225   315   390

570   615   705   795   915

1020   1275   1410   1605   1950
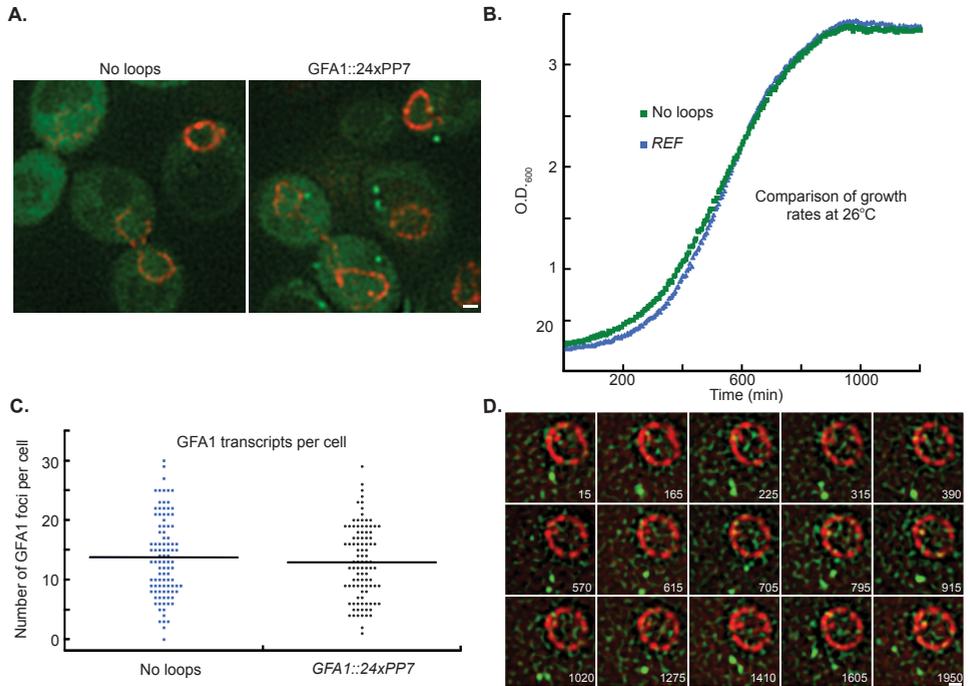
Figure 4.A.1. *REF* strain characterization. (A) Comparison of PP7-CP-3xYFP localization in a strain with no PP7 stem-loops (BMY642) and the *REF* strain with *GFA1*-24xPP7. (B) Growth curves of a control strain with no PP7 stem-loops (BMY642) and the *REF* strain with *GFA1*-24xPP7 at 26°C. The data shown are from a single representative experiment out of 3 repeats. (C) Dot plot shows the number of *GFA1* mRNAs per cell observed using smFISH probes against GFA1 in a strain with no PP7 stem-loops (BMY642) and the *REF* strain at 26°C with the mean denoted by a black line (n=100 cells). (D) Selected non-consecutive frames showing the splitting and merging of cytoplasmic particles in a *REF* cell with the time from the start of the event given in the bottom right of each image. Scale bar = 1 μm.

102
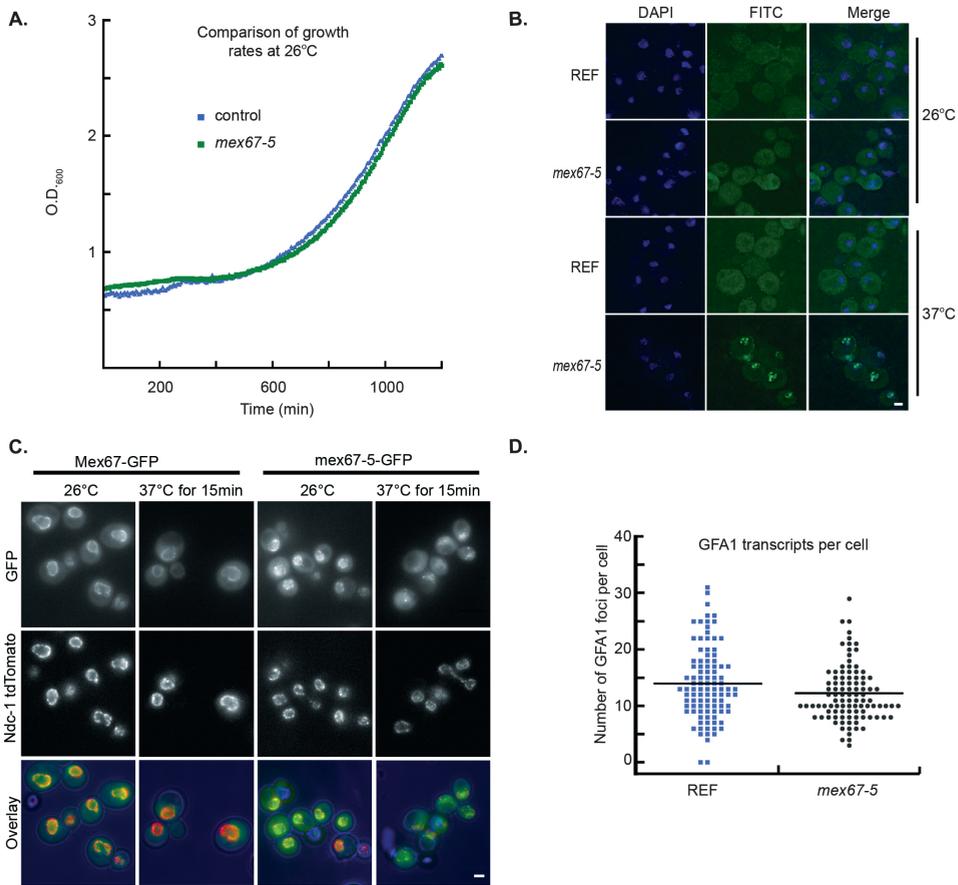
Figure 4.A.2. *mex67-5* strain characterization. (A) Growth curves of control (BMY129) and *mex67-5* (BMY135) strains at 26°C. The data shown are from a single representative experiment out of 5 repeats. (B) Representative images showing poly(A)-RNA localization in the *REF* and *mex67-5* strains at 26°C and 37°C. FISH was performed using a FITC-labeled oligo-(dT) probe and DNA was stained with DAPI. (C) GFP-tagged Mex67 localization in control (KWY5566) and *mex67-5* (KWY5567) strains at 26°C and 37°C as compared to Ndc1-tdTomato. Overlay displays the green and red channels as well as the bright field image. (D) Dot plot shows the number of *GFA1* mRNAs per cell in logarithmically growing *REF* (14±6) and *mex67-5* (12±5) strains at 26°C determined by smFISH with the mean denoted by a black line (n=100 cells). Scale bars = 1 μm.

Name	Description	Reference
pKT178	*pFA6–link–tdimer2–KanMX* (integrative plasmid, SP6 promoter for c-terminal tdimer2 protein fusion with KanMX-based selection)	(Sheff and Thorn, 2004)
pBM242	*pRS313-P$_{MET}$PP7-CP-3xYFP* (HIS3 CEN plasmid, MET17 promoter for PP7-CP-3xYFP expression)	This study
pDZ417	*pDZ417-24xPP7-loxP-KanMX-loxP* (integrative plasmid, T7 promoter for 24xPP7-loxP-KanMX-loxP cassette integration)	(Hocine et al., 2013)
pYM28	*pFA6-yEGFP-HIS3MX* (integrative plasmid, SP6 promoter for c-terminal yEFG protein fusion with HIS3MX-based selection)	(Janke et al., 2004)
pSH47	pRS416-GAL1-Cre (URA3 CEN plasmid, GAL1 promoter for Cre recombinase expression)	(Güldener et al., 1996)

Table 4.A.1. Plasmids used in this study.

	REF (n=43)	mex67-5(n=9)
Mean export time – dwell time analysis (ms)	188 ± 27	N.D.
Mean export time – MLE (ms)	215 ± 33	N.D.
Mean nuclear docking time during export – dwell time analysis (ms)	32 ± 5	362 ± 121
Mean nuclear docking time during export – MLE (ms)	39 ± 6	202 ± 67
Mean transition time during export – dwell time analysis (ms)	87 ± 13	406 ± 135
Mean transition time during export – MLE (ms)	99 ± 15	383 ± 128
Mean cytoplasmic docking time during export – dwell time analysis (ms)	62 ± 10	1258 ± 419
Mean cytoplasmic docking time during export – MLE (ms)	77 ± 12	943 ± 314

Table 4.A.2. Summary of dwell time analysis for successful mRNP export events. Reported errors are the SEM.

Name	Genotype	Reference
BMY008	BY4743 (MATaMatα his3Δ1his3Δ1 leu2Δ0leu2Δ0 LYS2lys2Δ0 met15Δ0MET15 ura3Δ0ura3Δ0)	(Brachmann et al., 1998)
BMY083	BY4743 (MATaMatα his3Δ1his3Δ1 leu2Δ0leu2Δ0 LYS2lys2Δ0 met15Δ0MET15 ura3Δ0ura3Δ0) GFA1-24PP7GFA1-24PP7 NDC1-tdTomato::KanMX NDC1-tdTomato::KanMX+[pBM242]	This study
BMY129	BY4743 (MATaMatα his3Δ1his3Δ1 leu2Δ0leu2Δ0 LYS2lys2Δ0 met15Δ0MET15 ura3Δ0ura3Δ0) NATMX::DBP5NATMX::DBP5 GFA1-24PP7GFA1-24PP7 NDC1-tdTomato::KanMX NDC1-tdTomato::KanMX+[pBM242])	This study
BMY135	BY4743 (MATaMatα his3Δ1his3Δ1 leu2Δ0leu2Δ0 LYS2lys2Δ0 met15Δ0MET15 ura3Δ0ura3Δ0) mex67-5::NATMXmex67-5::NATMX GFA1-24PP7GFA1-24PP7 NDC1-tdTomato::KanMX NDC1-tdTomato::KanMX+[pBM242])	This study
BMY642	BY4743 (MATaMatα his3Δ1his3Δ1 leu2Δ0leu2Δ0 LYS2lys2Δ0 met15Δ0MET15 ura3Δ0ura3Δ0) NDC1-tdTomato::KanMX NDC1-tdTomato::KanMX + [pBM242]	This study
KWY5566	BY4741 (MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0) MEX67-EGFP::HIS3MX GFA1-24PP7 NDC1-tdTomato::KanMX	This study
KWY5567	BY4741 (MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0) mex67-5EGFP::HIS3MX GFA1-24PP7 NDC1-tdTomato::KanMX	This study

Table 4.A.3. Yeast strains used in this study.

4

References

[1] S. Hocine, P. Raymond, D. Zenklusen, J. A. Chao, and R. H. Singer, *Single-molecule analysis of gene expression using two-color RNA labeling in live yeast,* Nature Methods **10**, 119 (2013).

[2] U. Güldener, S. Heck, T. Fiedler, J. Beinhauer, and J. H. Hegemann, *A new efficient gene disruption cassette for repeated use in budding yeast,* Nucleic Acids Research **24**, 2519 (1996).

[3] M. A. Sheff and K. S. Thorn, *Optimized cassettes for fluorescent protein tagging in saccharomyces cerevisiae,* Yeast **21**, 661 (2004).

[4] D. Grünwald and R. H. Singer, *In vivo imaging of labelled endogenous β-actin mRNA during nucleocytoplasmic transport,* Nature **467**, 604 (2010).

[5] J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, *et al.*, *Fiji: An open-source platform for biological-image analysis,* Nature Methods **9**, 676 (2012).

[6] S. Preibisch, S. Saalfeld, and P. Tomancak, *Globally optimal stitching of tiled 3D microscopic image acquisitions,* Bioinformatics **25**, 1463 (2009).

[7] C. S. Smith, N. Joseph, B. Rieger, and K. A. Lidke, *Fast, single-molecule localization that achieves theoretically minimum uncertainty,* Nature Methods **7**, 373 (2010).

[8] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory* (Prentice Hall, 1993).

[9] D. Colquhoun and A. G. Hawkes, *On the stochastic properties of bursts of single ion channel openings and of clusters of bursts,* Philosophical Transactions of the Royal Society B: Biological Sciences **300**, 1 (1982).

[10] U. Kubitscheck, D. Grünwald, A. Hoekstra, D. Rohleder, T. Kues, J. P. Siebrasse, and R. Peters, *Nuclear transport of single molecules dwell times at the nuclear pore complex,* The Journal of Cell Biology **168**, 233 (2005).

[11] C. N. Cole, C. V. Heath, C. A. Hodge, C. M. Hammell, and D. C. Amberg, *Analysis of RNA export,* Methods in Enzymology **351**, 568 (2002).

[12] J. D. Gibbons and S. Chakraborti, *Nonparametric statistical inference* (Springer, 2011).

5

Nuclear Accessibility of β-Actin mRNA is Measured by 3D Single-Molecule Real Time (3D-SMRT) Tracking

To tell you the truth,
I've never met anybody who can envision more than three dimensions. There are
some who claim they can, and maybe they can; it's hard to say.

EDWIN HUBBLE

Carlas S. Smith
Stephan Preibisch
Aviva Joseph
Sara Abrahamsson
Bernd Rieger
Eugene Myers
Robert H. Singer
David Grunwald

Abstract

Imaging single proteins or RNAs allows direct visualization of the inner workings of the cell. Typically, three-dimensional (3D) images are acquired by sequentially capturing a series of 2D sections. The time required to step through the sample often impedes imaging of large numbers of rapidly moving molecules. Here we applied multi-focus microscopy (MFM) to instantaneously capture 3D single-molecule real-time (3D-SMRT) images in live cells, visualizing cell nuclei at 10 volumes per second. We developed image analysis techniques to analyze mRNA diffusion in the entire volume of the nucleus. Combining multi-focus microscopy with precise registration between fluorescently labeled mRNA, nuclear pore complexes and chromatin, we obtained globally optimal image alignment within 80 nm precision using transformation models. We show that β-actin mRNAs freely access the entire nucleus, fewer than 60% of mRNAs are more than 0.5 μm away from a nuclear pore and we do so for the first time accounting for spatial inhomogeneity of nuclear organization.

5.1. *Introduction*

The nucleus of a cell is a crowded, compartmentalized volume wherein dynamic and complex biochemical and molecular events occur. For example, mRNAs are transcribed, spliced, released from the transcription site and subsequently move to the nuclear periphery where they are exported to the cytoplasm to be translated into proteins.

The nuclear landscape in which these processes take place is spatially complex. The genome is organized into topological domains, which in turn organize into non-random chromosome territories (reviewed in [1]. Adding to this complexity are functional distinct compartments or 'nuclear bodies', such as the nucleolus, histone locus bodies, splicing speckles and others (reviewed in [2]). The nuclear landscape is also temporally complex [3, 4]; nuclear bodies show high turnover rates of their components (reviewed in [5]), and the nucleus as a whole undergoes major reformation during the cell cycle [5–8]. Directly studying the dynamics of nuclear components, such as mRNAs in the nucleus of a living cell, will help to define the rules that govern the kinetics, locations, and interactions of proteins and nucleic acids relative to nuclear structure.

Advanced microscopy techniques have improved image resolution or enabled fast tracking of individual molecules in living cells, allowing the nuclear mobility of different proteins, RNAs and other molecules to be probed [9–12]. Currently available single-molecule imaging methods share the limitation that they can only image fast enough to accurately track single molecules in one optical plane (2D), or their 3D capability only allows visualization of small numbers of molecules within a limited field of view [13–15]. Conventional 3D imaging with wide-field light microscopes requires a series of images to be taken along the optical (z) axis. The time required to move the objective and sample relative to each other introduces a significant time delay that can be significant enough to prevent 3D tracking of fast-moving molecules. Furthermore, measuring the kinetics of single molecules relative to nuclear structure requires the accurate registration of image information from two or more different channels [16]. Perhaps more challenging is the need to image at physiologically tolerable excitation powers and the ability to detect weak signals [17]. Thus, sensitive microscopy methods that can quickly acquire high-resolution images and track single molecules in 3D volumes are needed [18].

Extracting information from a 3D volume into a single image plane, for example using astigmatism, double helix spiral phase microscopy, or techniques that simultaneously image multiple focal planes in biplane or multi-focus microscopy [13, 14, 19, 20], is one way to circumvent sequential z-stack imaging and instead si-

multaneously image 3D volumes. We recently developed multi-focus microscopy (MFM) as a method to track single molecules in 3D volumes. Here we combine the technique with precise image registration between fluorescently labeled mRNA, nuclear pore complexes, and chromatin for 3D single molecule real time tracking (3D SMRT). We present an image processing solution to convert the recorded images into well-aligned z-stacks. This solution consists of image registration between each plane, calculation of the correct z-position of each plane in each color channel, and registration between color channels. The same multi-focus optics are used for all color channels, causing a color dependent difference in z-spacing between the focal planes - as well as a slight magnification difference. We developed a transformation model to compensate for sample-induced aberrations and chromatic differences to enable global alignment of images within half-pixel precision. Finally z-stack images are deconvolved and further analyzed. As a result we can resolve the relationship between DNA-dense regions, the nuclear periphery, and the spaces occupied by mRNA moving through the nucleus. We show statistically that β-actin mRNAs freely access the entire nuclear space and that most (60%) are within 0.5 μm of nuclear pores.

5.2. *Results*

5.2.1. Reconstructing multicolor movies of single-molecules in 3D

We have applied multi-focus microscopy to image weak fluorescent signals of labeled mRNAs in the nucleus of a live cell and developed an image-processing framework [21] that relies on inherent registration markers [16]. To understand if nuclear structure, particular chromatin organization, impacts β-actin mRNA movement between its site of transcription and nuclear pores, we built a microscope to simultaneously image multiple focal planes with two different fluorescent channels. After the tube lens of the microscope, we attached the multi-focus optics that split the detection light into nine planes along the optical axis of the objective [19] (Fig. 5.1A). Mouse fibroblast cells were imaged in which all β-actin mRNAs are labeled with eYFP fused to the MS2 coat protein, nuclear pores fluorescently labeled with tdTomato fused to POM121, and heterochromatin stained with a violet vital dye (Vybrant DyeCycle Violet) that emits fluorescent signal over the entire spectral range. Zooming in on the nuclear pore stain (Fig. 5.1B) illustrates the power of simultaneous 3D imaging, even before quantitative image analysis. The section of the nucleus shown has an invagination of the nuclear envelope resulting in patches of nuclear pores being in focus (white arrowheads Fig. 5.1B) and out of focus (circle Fig. 5.1B).

We acquired time series of both the nuclear pore and β-actin mRNA simultaneously on two cameras with a frame rate of 10 Hz for up to 30 seconds using a 514 nm and a 561 nm laser for excitation. The power at the back-focal plane was measured to be ~600 μW. Using eYFP to label mRNA via MS2 coat protein, photobleaching was the limiting factor for acquisition duration. Within one second after acquiring the nuclear pore and mRNA data, we image the DNA stain on both cameras for 1 second followed by the acquisition of a piezo-driven z-stack with 20 nm steps. DNA images were used as a reference for image processing, to align images in each channel and to register the β-actin mRNA and nuclear pore channels.

5.2.2. Lateral image registration

During imaging, the emission signal is split into nine planes using a diffraction grating and the nine planes are recorded in an array on the camera. Due to imperfections in the MFM optical train, the nine sub-images (see Fig. 5.1A) are slightly rotated relative to each other (Fig. 5.2A) and need to be aligned laterally. The two fluorescent channels are recorded on separate cameras resulting in spatial offsets between them that need to be corrected for. The reference point in this
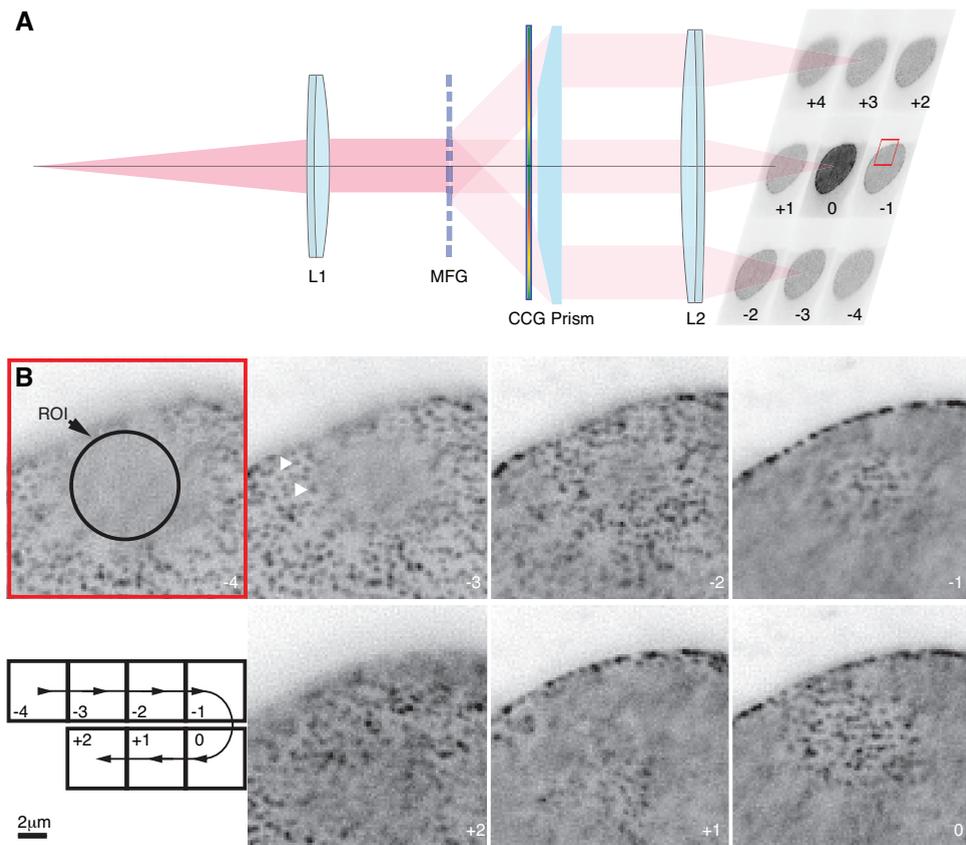
Figure 5.1. Simultaneously acquired 3D image of the nuclear envelope. (A) Schematic drawing of the multi-focus relay optic enabling simultaneous 3D image acquisition. The lens L1 forms a secondary Fourier plane (an image of the objective pupil) in which the MFG is placed to split the beam into zero- and first-order diffraction beams. Lens L2 forms the secondary image plane on the camera. The first-order diffraction forms 8 beams around the zero'th order with only the "above (+3)" and "below (-3)" beams depicted here. The zero-order beam travels straight through an empty panel in the chromatic-correction-grating (CCG), the central, flat facet of the prism, and is focused on the camera by L2. The first-order beams are separated and refocused by the MFG, corrected for chromatic dispersion by the CCG, redirected to proper positions on the camera by the prism, and focused onto the camera by L2. An example of the nine image planes (NPC image recorded in the red channel) projected in a 3×3 array and simultaneously recorded by a CCD camera is shown. Numbers on the subfields of the image correspond to the relative position of the plane in the stack. (B) A magnified series of a ROI from panel A (red box) shown over seven planes (planes +3 and +4 not shown as out of focus). Arrows (plane -3) point to nuclear pores, circle marks "unexpected" invagination. The region of the nuclear envelope that is out of focus near the bottom of the cell comes into focus near the middle of the cell. Image is shown with black and white values inverted.
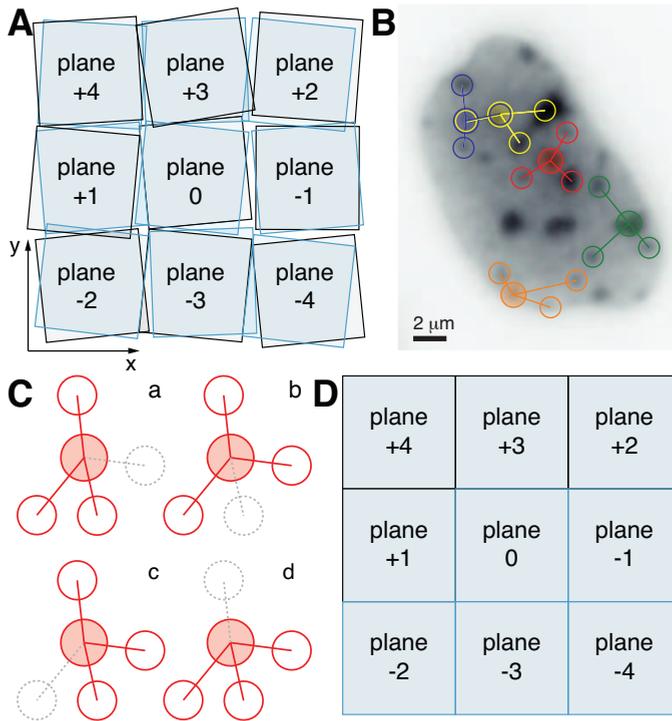
Figure 5.2. Lateral image registration. (A) Two-channel alignment of lateral image data as acquired by the camera. Black and blue panels represent the two color channels. Numbers indicate the relative position of each panel in the stack. (B) Image of the DNA stain used for image alignment overlaid the detection of blob-shaped interest points used as redundant geometric local descriptors to compute the lateral rigid transformations in (A). (C) Illustration of the redundancy used for matching of geometric local descriptors. (D) Illustration of camera image after successful lateral alignment.

alignment is the inherent DNA stain that can be excited specifically at 405 nm and emits over the whole visible range such that we can record identical images in both fluorescence color channels. The DNA stain fluorescence is used to align each channel over the 9 planes and both color channels relative to each other. The lateral image data provides sufficient resolution (~250x250 pixel/plane) to compute 2D rigid transformation models and apply them to each imaging plane in each channel using image interpolation (Fig. 5.2B).

To achieve a globally optimal lateral image alignment, we first performed geometric local descriptor matching [21]. Gaussian-shaped signals in the nine registration images of the DNA stack were extracted for each image plane using the Difference-of-Gaussian (DoG) detector combined with a 2D-quadratic fit for

subpixel localization [22]. The goal is to identify corresponding points and to use those corresponding points to determine the transformation for all planes. To achieve this, each candidate point is expressed as a geometric local descriptor consisting of the candidate point itself and its *n* nearest neighboring points (Fig. 5.2C), where translation-invariant geometric local descriptor matching is used to compute the similarity between two descriptors[21]. As DoG detections inside the sample are likely to contain random, non-corresponding detections, we add redundancy to the matching process (Fig. 5.2D). For each candidate point *n+1* nearest neighbor points are extracted and all combinations of *n* neighboring points are matched, where the best combination of any subset of both descriptors defines their similarity. Corresponding candidate pairs between two descriptors are assigned if the best match for a descriptor shows significantly better similarity (3x) than the second best matching descriptor. Comparing all descriptors of one image plane against all descriptors of another image plane yields a collection of corresponding candidates. Spurious matches are filtered using RANSAC [23] on a rigid transformation model, yielding a list of true corresponding descriptors. Exhaustively matching the set of all image planes provides all corresponding descriptors, linking together all image planes over all channels. Globally minimizing the distance between all corresponding points [24] results in an optimal rigid transformation model for each plane (Fig. 5.2D). This allows us to create a 3D stack for each image recorded and assess how well it is aligned.

5.2.3. Axial image registration

The spectral red-shift between eYFP and tdTomato causes deviations in z-step size between both color-channels (Fig 3A). Sample-induced aberrations along the z-axis cause an additional defocus in the image stack above and below the center plane. This effective defocus must be estimated quantitatively before further image restoration can be performed by image deconvolution. Therefore, additional axial image registration is required. Axial image alignment was performed independent of the lateral registration due to the significant sampling difference, 160 nm lateral with ~250 pixels per plane and 250 nm axial with only nine planes, as well as the different type of transformation models used in each dimension, lateral (x,y) and axial (z).

The axial position of all individual planes is described by 1D translation models, whereby the axial position of each image plane is computed by combining pairwise matching of auto-focus like functions with model-based outlier removal [23] and global optimization [24]. To do so we make use of the finely resolved second registration data set that is acquired using the stepping of the MFM volume
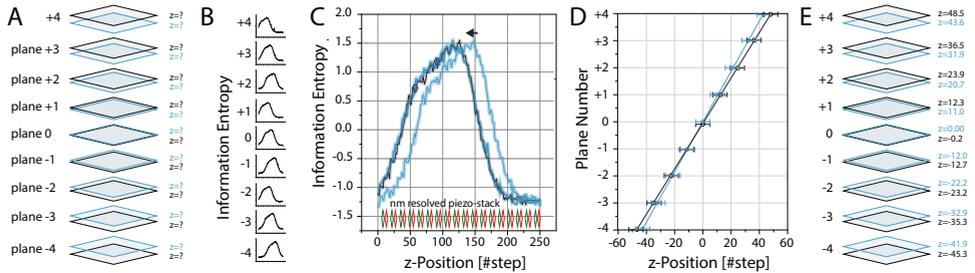
Figure 5.3. Axial image registration. (A) Final lateral (x,y) and missing axial (z) position of each plane in the two fluorescent channels (grey, black frames). (B) To determine the correct axial (z) position for each imaging plane the information entropy is computed over the entire axial range of the imaged cell using a 'stack of stack' acquired for each cell using a piezo-stage with nm step size. Maxima in the entropy function of each plane mark the point where the center of the image is in focus in each plane. (C) As the DNA signal is recorded in both channels, pairwise matching of the nine entropy functions from both channel stacks yields a set of 153 relative distances. For robustness of the method outliers are removed. The axial offset is the distance the grey entropy curve needs to be shifted (arrow) to optimally overlap with the black entropy curve. Black and grey represent the two fluorescent channels. (D) Data in each channel were fit to individual linear functions yielding the final relative position of each image plane in its stack. Due to the resolution of nine z-planes transformations are not applied to avoid image artifacts but instead applied to all extracted coordinates during subsequent analysis steps. (E) Fully aligned stack of all nine planes recorded by the camera. Each cell is sampled with 20 nm step size once, resulting in n=153 distance measures for registration. The error is dominated by the step width of the piezo to ± 5 nm indicated either as line width in C) or as error bar in D).

along the optical axis of the microscope using a piezo stage. Firstly, the entropy function for the piezo-stacks of all image planes and channels is computed (Fig. 5.3B) and the information content for each individual z-slice of every piezo-stack is measured, with the information content being highest in the center of the sample. As the finely resolved stack of 3D-SMRT volumes covers the entire sample for every image plane, the shape of each function is similar (Fig. 5.3B) but shifted in axial position (Fig. 5.3C). All entropy functions are pairwise registered with each other using a simple gradient descent (Fig. 5.3C) yielding all pairwise axial distances between all image planes. For each channel all axial positions can be fitted using a linear function (Fig. 5.3D). This allows us to remove outliers in the pairwise entropy alignment (Fig. 5.3C) using RANSAC on a linear function fit. Global minimization of all pairwise axial distances between all remaining inliers from all channels yields the final positions of all image planes of all channels (Fig. 5.3D). The slope of the linear fit along the z-axis is a function of the emission wavelength, as to be expected by the design of the MFG [19]. The registration image data directly yields the chromatic shift along the z-axis. The correction is not applied to the image data to avoid interpolation artifacts that would result from the correction value not exactly matching the 250 nm axial sampling between the nine planes. Instead, the axial position of each plane is directly applied to all coordinates during subsequent analysis steps (Fig. 5.3E). Following deconvolution it is now possible to quantify distinct features in the image volume.

5.2.4. Image registration precision

We found it both important and necessary to have a measure for success or failure of the image alignment, as many parameters impact the image registration performance. Such parameters include: the number of descriptor pairs; cut-off values for outlier removal; the transformation model used (translation vs. rigid vs. affine vs. homography), or the homogeneity of feature distribution over the sample; stain and shape of features; and the general geometry of the application. In general, the fewer the corresponding features (signals used to register the images) to be aligned the more precise the registration will be. To achieve uniform alignment over the whole object of interest, in our case the nucleus, the more sampling points, the more homogenous their distribution and the lower the remaining error, the better the global alignment will be.

We used a rigid transformation model and a global alignment target of 0.5-pixel cutoff value for outlier removal (RANSAC). The image-registration precision therefore is determined as the higher bound of an image alignment that we deem acceptable. To assess if the registration value targeted is achieved or realistic for
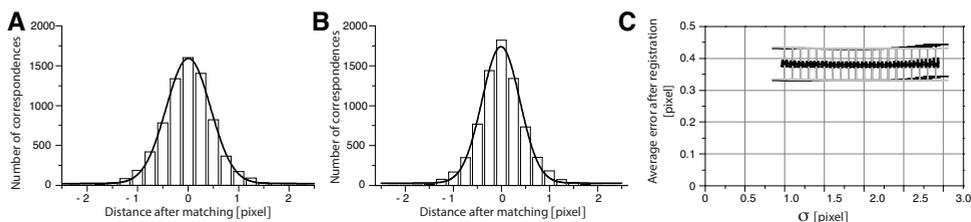
Figure 5.4. Global registration precision. A, B) Distribution of remaining distances between n = 7387 registration features after registration with a rigid two-dimensional transformation model for a single cell. No bias (mean value is zero and distribution symmetric) was introduced and the width σ for x (A) is 0.44 pixel or 70 nm and for y (B) is 0.39 pixel or 62 nm. A & B represent the error of the lateral alignment in x and y as laid out in Fig. 2 for a single cell with a target of σ = 0.5 pixel global registration precision. C) Two kinds of beads were embedded in an agarose gel. First the diffraction limited beads were aligned using a Gaussian fit for the center of the bead location which was then used for alignment. Alternatively, the beads (1 μm diameter) were aligned using a quadratic fit for corresponding features. The remaining average error after registration relative to the σ used for fitting is shown. The differences between both methods are within the noise level at less than 1 nm. The total error of the alignment is dominated by the localization precision (here 10 nm, error bar in C).

the data set, the remaining distance of all corresponding features is measured and plotted as a histogram (Fig. 5.4A and B). We found that both x and y components are distributed normally and are well described by a Gaussian curve. The 'width' of these resulting distributions is a measure of precision, σ, for the achieved registration. We found that the remaining error after alignment of the cell nucleus data using the inherent DNA marker for 7387 corresponding features had a width of 0.44 pixel for the x-ordinate (Fig. 5.4A) and 0.39 pixel for the y-ordinate (Fig. 5.4B), well below our target of 0.5 pixel. An even higher precision can be achieved after focus shift correction, but only for the registration signal. This is because the registration signal is not a single MFM volume, but a nm resolved z-scan of many MFM volumes over the z-range of the MFM volume used for tracking. We used these alignment piezo stacks to calculate the z-offset that we correct for after localization in the x,y plane (Fig. 5.2 and 5.2). If one overlays these two perfectly z-matched image planes, the x,y registration is higher than for the experimental data. The next point of interest is if the alignment has a bias. Or what is the accuracy of the alignment? As can be seen in Fig. 4 the mean of the distribution is zero and symmetric (measure of accuracy) which agrees with the nature of stochastic errors, while the width σ is less than 0.5 pixel (measure of precision).

118

Next we used an artificial sample of known structure to test image alignment under defined conditions. Beads of 1 μm diameter were labeled with plasmid DNA, which was stained with Vybrant DyeCycle Violet and embedded together with 200 nm TetraSpeck beads (Life Technologies) in 1% agarose. The green and red image stacks were aligned in one of two ways: using feature-based alignment of the larger beads with a 0.5-pixel precision target, or based on the point signals from the TetraSpeck beads. The agarose introduced minor aberrations compared to the cells, providing a best-case test scenario. As expected, the resulting alignment was better than for the cell data in both cases (Fig. 5.4C). Gaussian fitting of the fluorescent beads with widths close to that of the point spread function (PSF) in a low-scattering, homogenous medium or sample, such as the gel used in our control experiment (Fig. 5.4), was not superior to the feature-based alignment providing a sub-nanometer smaller final registration error. A sub-nanometer difference is within the noise limit of the method and therefore insignificant. The feature-based alignment (Fig. 5.4C) proved more robust for irregular or blurred signals/image features. In practice, the Gaussian fitting of point-like bead signals neglects aberrations introduced by the sample itself, whereas feature-based registration (Fig. 5.4A and B) allows sample-induced aberrations described by the transformation model to be corrected for, and it is also more robust (Fig. 5.4C).

5.2.5. 3D-SMRT imaging of live cells

Our aim was to test if specific areas exist in the nucleus of living cells that are correlated with how mRNAs reach the nuclear periphery and if heterochromatin contains a lower level of mRNAs than found, on average, within the nucleoplasm. To validate our nuclear pore tag as a marker for the nuclear envelope we counted the number of nuclear pore complexes (NPCs) per μm^2, at the focal planes closest to the objective (Fig 1B). We found that there are ~1 to 4 NPCs per μm^2, as previously reported using confocal microscopy [25]. 3D-SMRT microscopy directly visualizes β-actin mRNA movement with relation to dense and sparse DNA volumes within the nucleus (Fig 5A), which correlates to heterochromatin rich and poor regions. The ability to detect the relative positions of DNA, mRNA and NPCs from different angles, instead of in 2D planes that often contain out-of-focus signal, allows rigorous interrogation of possible colocalization events (Fig. 5.5B and C). The ability to record volumetric movies makes it possible to track molecules and nuclear structure over time (Fig. 5.5D-G). After image deconvolution the image volumes still contain noise, but the signal is defined by its localized intensity due to the multiple binding sites on the mRNA (Fig. 5.5H). To understand nu-

clear occupation by β-actin mRNA intensity, thresholds were manually selected to best fit the dense DNA volume, which we refer to as 'heterochromatin'. An additional lower intensity threshold was selected that corresponded with volumes surrounding the high intensity DNA signal, which we refer to as 'periphery'. Any remaining space within the nuclear envelope, defined by the nuclear pore stain is referred to as 'nucleoplasm'. Thresholds were confirmed to be comparable between cells using an ANOVA test. Visual comparison of staining patterns of Vybrant DyeCycle Violet live cell DNA stain to fixed cell DNA stains (e.g. Hoechst or DAPI) indicates that they correspond well with each other. This result is in agreement with previous reports [7]. By measuring the volume within the nucleus that is occupied by heterochromatin, we find that heterochromatin constitutes ~4% (+/- 2%) of the nuclear volume.

Distribution of β-actin mRNA in the nucleus

We were particularly interested in determining if β-actin mRNA is able to access heterochromatin regions, and if so to what degree. We found that β-actin mRNA distributed almost uniformly across 'nucleoplasm', 'periphery' and 'heterochromatin' (Fig. 5.5I). We do not find a significant difference in β-actin mRNAs between the 'nucleoplasm' and 'periphery', or between the 'periphery' and 'heterochromatin' volumes (Fig. 5.5I). The definition of 'periphery' attributes to the limited resolution of light microscopes and both 'nucleoplasm' and 'periphery' are defined without a positive marker. If we ignored the 'periphery' and hence bias the two-sided nature of the appropriate statistical test, we found a significant 17% reduction of β-actin mRNA in 'heterochromatin' compared to the 'nucleoplasm'.

The only area within the nucleus with evident exclusion of β-actin mRNA is the nuclear envelope other than at nuclear pore complexes during export or scanning along the edge of the nuclear envelope. The number of export events recorded is small, likely caused by the longer integration time compared to the original study on β-actin mRNA export in these cells. We do observe that those pores that transported β-actin mRNA before are repeatedly active over time (Fig. 5.5D). In summary, the distribution of β-actin mRNA in the nucleus is not uniform, but no areas are devoid of β-actin mRNA, including DNA dense areas. Rather we find that over a 5 sec time interval β-actin mRNAs inhabit on average ~5% +/- 1% of the total nuclear volume. With the advantage of being able to simultaneously follow mRNA distribution and movement within the living cell along x, y, and z axes, in real time, this number accounts for 3D diffusion of β-actin mRNAs. The total number of voxels were counted that were occupied by RNA signals at a given time point, corresponding to a 100 msec time window that minimizes mobility
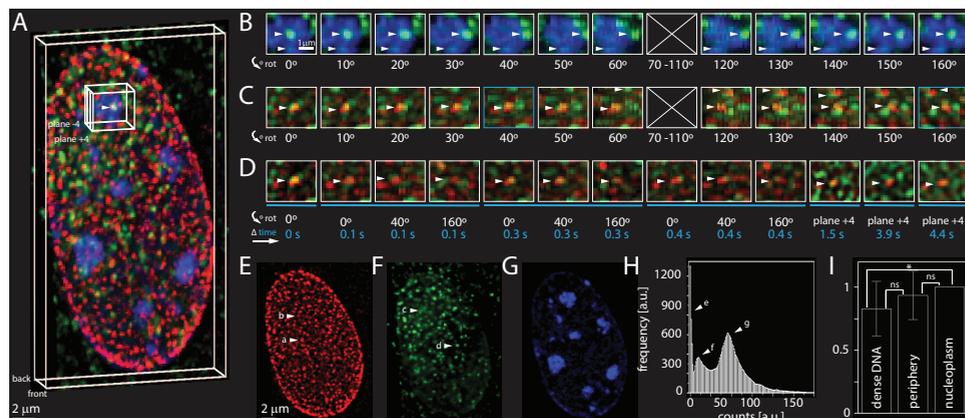
Figure 5.5. 3D distribution of β-actin mRNA within the nucleus of living cells. A) 3D projection of a nucleus showing β-actin mRNA carrying a 24× MS2 stem-loop cassette labeled with eYFP-MS2 coat protein (green), NPCs labeled with POM121-tdTomato (red), and heterochromatin DNA labeled with the vital dye Vybrant DyeCycle Violet (blue). The view is a time projection (β-actin mRNA - 5 sec; NPC - 0.3 sec; DNA - 1 sec) rotated 10° around the y-axis. The less intense clouds of signal represent mobile mRNA over this time span, spots represent immobile mRNAs. The white box indicates the ROI of the nucleus shown in panels B-D. B) Zoomed in rotational view of β-actin mRNA (green) and heterochromatin (blue), numbers below the images indicate the rotation angle. Arrows point at two mRNA clusters that appear to colocalize with heterochromatin. While the upper cluster sits on top of the chromatin (40°-120°) the lower cluster is located within the heterochromatin. Projection was done without interpolation resulting in 'line-patterns' close to the 90° view (e.g. see 60°) and are therefore not displayed. (C) Rotational view of β-actin mRNA and NPCs analog to panel B. At all angles the β-actin mRNA partially overlaps with an NPC (lower arrow). Panel B & C are static projections of the first frame of the movie. D) Time course of β-actin mRNA and NPCs from three rotational perspectives. The 40° and 160° views for the 0 sec time point are shown in panel C) with blue frames. The mRNA - NPC interaction is best visible at 0° angle and invisible from the opposite side of the cell (160°) and ultimately non-productive (mRNA arrives and leaves NPC on the same side). The last three views are from the first slices of the stack at 0° angle at later time points of the same movie illustrating repeated activity of a specific nuclear pore over time. Cyan bars under images indicate identical time groups. E) Nuclear pore 3D projection from panel A. Arrows point to nuclear pores at the top/plane +4 (a) and bottom/plane -4 (b) of the nucleus. F) 3D projection of the RNA channel from panel A. Arrows point to examples of a mobile β-actin mRNA resulting in a blurred projection (c) and a stationary β-actin mRNA resulting in a sharp projection (d). G) 3D projection of the DNA channel from panel A. H) Intensity frequency histogram of the cell (n=1) in F) showing three distributions: to the very left background resulting from rotation of the deconvolved image (arrow e); next a lower intensity peak resulting from mobile mRNAs in the time projection (arrow f); and to the right a higher intensity peak resulting from stationary mRNAs in the time projection (arrow g). (I) Bar graph showing the level of β-actin mRNA localized to heterochromatin and heterochromatin periphery relative to nucleoplasm. On average a slight reduction of mRNA occupation in the heterochromatin was found, but it was not statistically relevant (two-sided test). * = significant within 2σ confidence interval, ns = not significant. Images were processed as described in Materials and Methods and contrast was adjusted for visibility.

121

effects on the distribution of the mRNAs. We found cell-to-cell variations in total voxel numbers ranging from a lower limit of ~2200 voxels to an upper limit of ~7200 voxels with an average of ~5100 +/- 1400 voxels being occupied by β-actin mRNAs. At the achieved spatial sampling of 160 nm in x and y directions and 250 nm in z direction, a diffraction-limited spot will generate 12 voxels of signal. The variation in β-actin mRNA numbers is well explained by variation in expression levels from cell to cell. To exclude the possibility that the observed variation in β-actin mRNA occupancy could be due to image processing, we performed 2-way ANOVA analysis of the variation in total number of β-actin mRNA occupied voxels between cells, which we found to be significant. When testing for variation in β-actin mRNA density rather than total number of β-actin mRNA occupied voxels the ANOVA yields no significant differences, indicating that the observed β-actin mRNA density is comparable between cells.

β-Actin mRNAs are always close to nuclear pores

In the equatorial plane the distance between mRNA and NPCs can be multiple μm and is limited by the size of the nucleus (Fig. 5.6A). However, in cultured cell lines, the nucleus is often disk-shaped, and the top and bottom are visible within our imaging volume. The radius of the nucleus in the z-direction is hence comparable to the height of our imaging volume (2.25 μm). Logically the height of the nucleus should dominate the actual distance between mRNA and the nuclear envelope. To understand this actual distance distribution between mRNA and the nuclear periphery, we identified the shortest lateral and axial distance between every mRNA and the closest nuclear pore using a nearest neighbor algorithm. Query of the lateral (x and y axis) distance between β-actin mRNA in any of our nine image planes to the closest nuclear pore complex revealed that ~95% of β-actin mRNA are observed within 1.5 μm from a nuclear pore complex (Fig. 5.6A and C). The 2D contour size of a nucleus in our MEF cell line is roughly given by an oval with a shorter axis of ~8 μm and a longer axis of ~20 μm. In other words, if the analysis of nuclear space occupancy by mRNA was based on data acquired only in the equatorial plane of the cell, it could be incorrectly concluded that the center of the nucleus contains less β-actin mRNA than the nuclear periphery. If we instead analyze the 3D occupancy of nuclear volume by mRNA, querying the distance distribution for β-actin mRNA to the closest nuclear pore complex in the z-direction we find more than 60% of β-actin mRNAs within 0.5 μm and more than 90% of β-actin mRNA within one μm distance of a nuclear pore (Fig. 5.6B and D). The size correlation between the thickness of the nucleus and the imaging volume, both in the range of 2 μm implies that a distance of 1 μm between mRNA
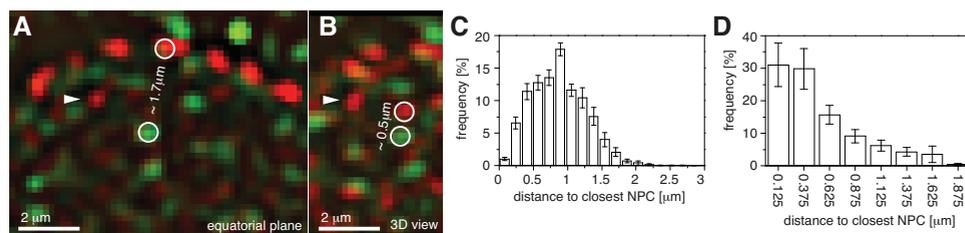
Figure 5.6. Distance between mRNA and nuclear envelope. A) Region of interest view of the equatorial plane of Fig. 5.5A. The closest distance between the circled mRNA and an NPC is close to 2μm. The arrow points at a presumably closer NPC. B) Rotated view of a 3D projection of the same mRNA. A NPC is found in close proximity to the mRNA two planes above. Arrow points to the same NPC as in A) showing that the NPC is located 1 plane above the mRNA in panel A). C) For all identified mRNA signals the closest lateral nuclear pore is identified and the distance measured. Most mRNAs (>80%) are further than 0.5 μm away from the nuclear rim. D) For all identified mRNAs the shortest distance to a nuclear pore is calculated in 3D. At least 60% of mRNAs are within 0.5 μm of a nuclear pore and no more than 15% of mRNAs are further than one μm from a nuclear pore. The data are from 12 cells in which >51000 distances between mRNAs and NPCs were measured.

and NPC is an upper limit along the z-axis. Only ~30% of β-actin mRNA were found within 0.5 μm of a nuclear pore in the lateral direction.

5.3. *Discussion*

Early studies of intranuclear mRNA movement presented the possibility that distinct tracks or pathways exist from the transcription site to the nuclear pore [26–30]. Later studies, however, have shown that this is not the case; that is, mRNA movement within the nucleus is diffusive and distributes throughout the intranuclear space [31–36]. There are, however, findings that demonstrate this is not the entire picture. For example, gene activity was shown to correlate with chromatin territories [37]; a study based on the kinetic analysis of single human dystrophin-mRNAs in U2OS cells suggests that mRNA mobility is facilitated in regions of low DNA density [38]; the TREX complex was shown to couple to the nuclear pore complex (NPC) thereby linking gene expression to nuclear transport [39–42]; and upon depletion of ATP from the nucleus the movement of β-actin mRNA is inhibited, indicating that mRNA mobility in the nuclear landscape requires energy [12].

In the last decade, kinetic studies using modern light microscopy in living cells painted a highly dynamic picture of the nucleus [4, 10, 12, 43]. Recombinant

and endogenously fluorescent tagged nuclear proteins and RNAs or fluorescent beads loaded into the nucleus were used to probe nuclear mobility and interactions with specific nuclear compartments like heterochromatin or the nucleolus [9, 34, 44, 45]. In most of these biophysically oriented studies, different mobile particles were recorded in one fluorescent channel simultaneously with fluorescently tagged nuclear compartments in a second fluorescent channel. From these studies a divergent picture of nuclear organization arises: single molecules and particles of almost any size (up to ~50nm) can access any portion of the nuclear space, even including densely packed heterochromatin regions [9, 10]. Large populations of molecules, such as polyA mRNAs, have been shown to move in the nucleus in the range of 1 μm^2/sec by use of fluorescence correlation spectroscopy (Politz et al., 2006). Interestingly, what was thought to be an exclusion of proteins from the nucleolus instead correlated with a lower retention rate and hence a higher mobility of proteins in the nucleolus compared to the nucleoplasm and heterochromatin regions [10]. What these studies had in common was that the molecules in the nucleus were not constrained.

Our data represent the first full 3D analysis of the dynamic mobility, occupation, and interaction of a typical nuclear particle, β-actin mRNA, within the nuclear environment. We do not find evidence of exclusion from, or enrichment in, the heterochromatin, nor do we find repeated tracks of β-actin mRNAs occupying the same region of the nucleus on their travel to the nuclear envelope. Our data add that most β-actin mRNA (>60%) are less than 500 nm from the central channel of a nuclear pore complex. At the same time β-actin mRNA appears to be traveling freely through the complex landscape of the nucleus.

These findings were possible by applying multi-focus microscopy to the simultaneous 3D imaging of mRNAs in the nucleus of a live cell. The acquisition of 9 planes in a 3D signal allows the tracking of particles with a slower frame rate than in 2D, as particles are not lost due to being out of focus. However, faster frame rates than 100 milliseconds per stack are needed to reduce blurring of the signal and to enable precise tracking of individual mRNAs over time. As a result we based our analysis on the measurement of voxels that mRNA occupied for each time point. A slower imaging rate means that more photons can be collected per time point, but a substantial amount of signal is lost due to lower efficiencies of the 3D components since MFG (multi-focus grating), CCG (chromatic correction grating) and prisms are theoretically ~60% light-efficient. The distribution of the signal in z-direction suffers from image distortion. We were able to measure these distortions in each cell and apply image registration and restoration to compensate. This was possible using a cell-internal registration marker. We developed three

major image analysis components: geometric descriptor matching to orient the image planes correctly, image entropy analysis to find the chromatically corrected z position for each plane, and analytical description of spherical aberration along the z-axis necessary before image deconvolution.

We showed that 3D-SMRT microscopy is applicable to living cells at tolerable excitation power levels. Using a cell internal registration marker, similar to the super-registration concept for 2D alignment of nuclear pores [16], we developed a robust, feature-based image-alignment method that can be adjusted to the amount of registration features, supports different transformation models and in this case was used to achieve global alignment of the MFM data with 0.5-pixel precision. With the alignment signal originating from within the sample, the degree of homogeneity in distribution of registration features across the image is known, and possible optical aberrations introduced by the cell are corrected for. While the current sampling rates, both in time and space, leave room for improvement, they are in the range of some 2D single-molecule imaging data [46] and provide better time resolution than in comparable studies on nuclear organization [38]. Although substantially faster 2D tracking data have been reported [16, 47, 48], the ability to acquire 3D data translates into an extended focal depth (here about three times larger than in the 2D case) and a reduced chance of losing particles in the same time interval used in 2D imaging. Acquisition speed therefore can be slower, depending on the field of view and the total focal depth. The technology for faster imaging is already available with sCMOS cameras and larger, faster chips in EMCCD cameras. This means that signal currently limits the image acquisition speed. Improvements in signal strength are hence desirable and should ideally come from improved technology for labeling bio-molecules and background reduction, as reduced excitation power is desirable in parallel to increased signal. The complex shape of the nuclear envelope and the distance between mRNA and nuclear pore in a living cell are two examples of why studies of molecular interactions in living cells need to account for anisotropy or non-linearity of the cellular environment using simultaneous 3D-imaging approaches.

References

[1] J. H. Gibcus and J. Dekker, *The hierarchy of the 3D genome,* Molecular Cell **49**, 773 (2013).

[2] J. Padeken and P. Heun, *Nucleolus and nuclear periphery: Velcro for heterochromatin,* Current Opinion in Cell Biology **28**, 54 (2014).

[3] T. Misteli, A. Gunjan, R. Hock, M. Bustin, and D. T. Brown, *Dynamic binding of histone H1 to chromatin in living cells,* Nature **408**, 877 (2000).

[4] R. D. Phair and T. Misteli, *High mobility of proteins in the mammalian cell nucleus,* Nature **404**, 604 (2000).

[5] J. E. Sleeman and L. Trinkle-Mulcahy, *Nuclear bodies: New insights into assembly/dynamics and disease relevance,* Current Opinion in Cell Biology **28**, 76 (2014).

[6] P. R. Cook and D. Marenduzzo, *Entropic organization of interphase chromosomes,* The Journal of Cell Biology **186**, 825 (2009).

[7] L. Schermelleh, P. M. Carlton, S. Haase, L. Shao, L. Winoto, P. Kner, B. Burke, M. C. Cardoso, D. A. Agard, M. G. Gustafsson, *et al.*, *Subdiffraction multicolor imaging of the nuclear periphery with 3D structured illumination microscopy,* Science **320**, 1332 (2008).

[8] S. P. Shevtsov and M. Dundr, *Nucleation of nuclear bodies by RNA,* Nature Cell Biology **13**, 167 (2011).

[9] S. M. Görisch, M. Wachsmuth, C. Ittrich, C. P. Bacher, K. Rippe, and P. Lichter, *Nuclear body movement is determined by chromatin accessibility and dynamics,* Proceedings of the National Academy of Sciences of the United States of America **101**, 13221 (2004).

[10] D. Grünwald, R. M. Martin, V. Buschmann, D. P. Bazett-Jones, H. Leonhardt, U. Kubitscheck, and M. C. Cardoso, *Probing intranuclear environments at the single-molecule level,* Biophysical Journal **94**, 2847 (2008).

[11] J. C. R. Politz, R. A. Tuft, K. V. Prasanth, N. Baudendistel, K. E. Fogarty, L. M. Lifshitz, J. Langowski, D. L. Spector, and T. Pederson, *Rapid, diffusional shuttling of poly(A) RNA between nuclear speckles and the nucleoplasm,* Molecular Biology of the Cell **17**, 1239 (2006).

[12] Y. Shav-Tal, X. Darzacq, S. M. Shenoy, D. Fusco, S. M. Janicki, D. L. Spector, and R. H. Singer, *Dynamics of single mRNPs in nuclei of living cells,* Science **304**, 1797 (2004).

[13] M. P. Backlund, M. D. Lew, A. S. Backer, S. J. Sahl, G. Grover, A. Agrawal, R. Piestun, and W. Moerner, *Simultaneous, accurate measurement of the 3D position and orientation of single molecules,* Proceedings of the National Academy of Sciences **109**, 19087 (2012).

[14] B. Huang, W. Wang, M. Bates, and X. Zhuang, *Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy,* Science **319**, 810 (2008).

[15] T. Ragan, H. Huang, P. So, and E. Gratton, *3D particle tracking on a two-photon microscope,* Journal of Fluorescence **16**, 325 (2006).

[16] D. Grünwald and R. H. Singer, *In vivo imaging of labelled endogenous β-actin mRNA during nucleocytoplasmic transport,* Nature **467**, 604 (2010).

[17] P. M. Carlton, J. Boulanger, C. Kervrann, J.-B. Sibarita, J. Salamero, S. Gordon-Messer, D. Bressan, J. E. Haber, S. Haase, L. Shao, *et al.*, *Fast live simultaneous multiwavelength four-dimensional optical microscopy,* Proceedings of the National Academy of

5

Sciences **107**, 16016 (2010).

[18] L. Trinkle-Mulcahy and A. I. Lamond, *Toward a high-resolution view of nuclear dynamics,* Science **318**, 1402 (2007).

[19] S. Abrahamsson, J. Chen, B. Hajj, S. Stallinga, A. Y. Katsov, J. Wisniewski, G. Mizuguchi, P. Soule, F. Mueller, C. D. Darzacq, *et al., Fast multicolor 3D imaging using aberration-corrected multifocus microscopy,* Nature Methods **10**, 60 (2013).

[20] S. Ram, P. Prabhat, J. Chao, E. S. Ward, and R. J. Ober, *High accuracy 3D quantum dot tracking with multifocal plane microscopy for the study of fast intracellular dynamics in live cells,* Biophysical Journal **95**, 6025 (2008).

[21] S. Preibisch, S. Saalfeld, J. Schindelin, and P. Tomancak, *Software for bead-based registration of selective plane illumination microscopy data,* Nature Methods **7**, 418 (2010).

[22] M. Brown and D. G. Lowe, *Invariant features from interest point groups.* in *BMVC,* s 1 (2002).

[23] M. A. Fischler and R. C. Bolles, *Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography,* Communications of the ACM **24**, 381 (1981).

[24] S. Saalfeld, A. Cardona, V. Hartenstein, and P. Tomančák, *As-rigid-as-possible mosaicking and serial section registration of large ssTEM datasets,* Bioinformatics **26**, i57 (2010).

[25] U. Kubitscheck, T. Kues, and R. Peters, *Visualization of nuclear pore complex and its distribution by confocal laser scanning microscopy,* Methods in Enzymology **307**, 207 (1999).

[26] P. S. Agutter, *Models for solid-state transport: Messenger RNA movement from nucleus to cytoplasm.* Cell Biology International **18**, 849 (1994).

[27] G. Blobel, *Gene gating: A hypothesis,* Proceedings of the National Academy of Sciences **82**, 8527 (1985).

[28] I. Davis and D. Ish-Horowicz, *Apical localization of pair-rule transcripts requires 3' sequences and limits protein diffusion in the drosophila blastoderm embryo,* Cell **67**, 927 (1991).

[29] B. A. Edgar, M. P. Weir, G. Schubiger, and T. Kornberg, *Repression and turnover pattern fushi tarazu RNA in the early drosophila embryo,* Cell **47**, 747 (1986).

[30] J. B. Lawrence, R. H. Singer, and L. M. Marselle, *Highly localized tracks of specific transcripts within interphase nuclei visualized by in situ hybridization,* Cell **57**, 493 (1989).

[31] B. Daneholt, *Pre-mRNP particles: From gene to nuclear pore,* Current Biology **9**, R412 (1999).

[32] R. W. Dirks, K. C. Daniël, and A. K. Raap, *RNAs radiate from gene to cytoplasm as revealed by fluorescence in situ hybridization,* Journal of Cell Science **108**, 2565 (1995).

[33] A. M. Femino, F. S. Fay, K. Fogarty, and R. H. Singer, *Visualization of single RNA transcripts in situ,* Science **280**, 585 (1998).

[34] J. C. Politz, R. A. Tuft, T. Pederson, and R. H. Singer, *Movement of nuclear poly(A) RNA throughout the interchromatin space in living cells,* Current Biology **9**, 285 (1999).

[35] O. P. Singh, B. Björkroth, S. Masich, L. Wieslander, and B. Daneholt, *The intranuclear movement of balbiani ring premessenger ribonucleoprotein particles,* Experimental Cell Research **251**, 135 (1999).

[36] Z. Zachar, J. Kramer, I. P. Mims, and P. M. Bingham, *Evidence for channeled diffusion of pre-mRNAs during nuclear RNA transport in metazoans.* The Journal of Cell Biology **121**, 729 (1993).

[37] S. Stadler, V. Schnapp, R. Mayer, S. Stein, C. Cremer, C. Bonifer, T. Cremer, and S. Diet-

zel, *The architecture of chicken chromosome territories changes during differentiation,* BioMed Central Cell Biology **5**, 44 (2004).

[38] A. Mor, S. Suliman, R. Ben-Yishay, S. Yunger, Y. Brody, and Y. Shav-Tal, *Dynamics of single mRNP nucleocytoplasmic transport and export through the nuclear pore in living cells,* Nature Cell Biology **12**, 543 (2010).

[39] T. Cremer, M. Cremer, S. Dietzel, S. Müller, I. Solovei, and S. Fakan, *Chromosome territories–a functional nuclear landscape,* Current Opinion in Cell Biology **18**, 307 (2006).

[40] D. Jani, S. Lutz, E. Hurt, R. A. Laskey, M. Stewart, and V. O. Wickramasinghe, *Functional and structural characterization of the mammalian TREX-2 complex that links transcription with nuclear messenger RNA export,* Nucleic Acids Research **40**, 4562 (2012).

[41] D. Umlauf, J. Bonnet, F. Waharte, M. Fournier, M. Stierle, B. Fischer, L. Brino, D. Devys, and L. Tora, *The human TREX-2 complex is stably associated with the nuclear pore basket,* Journal of Cell Science **126**, 2656 (2013).

[42] V. O. Wickramasinghe, R. Andrews, P. Ellis, C. Langford, J. B. Gurdon, M. Stewart, A. R. Venkitaraman, and R. A. Laskey, *Selective nuclear export of specific classes of mRNA from mammalian nuclei is promoted by GANP,* Nucleic Acids Research **42**, 5059 (2014).

[43] D. Grünwald, A. Hoekstra, T. Dange, V. Buschmann, and U. Kubitscheck, *Direct observation of single protein molecules in aqueous solution,* ChemPhysChem **7**, 812 (2006).

[44] K. E. Handwerger, J. A. Cordero, and J. G. Gall, *Cajal bodies, nucleoli, and speckles in the xenopus oocyte nucleus have a low-density, sponge-like structure,* Molecular Biology of the Cell **16**, 202 (2005).

[45] K. Richter, M. Nessling, and P. Lichter, *Experimental evidence for the influence of molecular crowding on nuclear architecture,* Journal of Cell Science **120**, 1673 (2007).

[46] S. Semrau, L. Holtzer, M. González-Gaitán, and T. Schmidt, *Quantification of biological interactions with particle image cross-correlation spectroscopy (PICCS),* Biophysical Journal **100**, 1810 (2011).

[47] D. Grünwald, B. Spottke, V. Buschmann, and U. Kubitscheck, *Intranuclear binding kinetics and mobility of single native U1 snRNP particles in living cells,* Molecular Biology of the Cell **17**, 5017 (2006).

[48] S. Manley, J. M. Gillette, and J. Lippincott-Schwartz, *Single-particle tracking photoactivated localization microscopy for mapping single-molecule dynamics,* Methods in Enzymology **475**, 109 (2010).

5

Appendix

5.A. *Materials and Methods*

5.A.1. Cell lines

In the mouse fibroblast cell line used all β-actin mRNA carry a 24X MS2 stem loop cassette in the 3' UTR that is labeled by enhanced yellow fluorescent protein (eYFP) fused to a MS2 protein tag. This cell line also harbors POM121 protein labeled with tandem Tomato (tdTomato) fluorescent protein. POM121 exists in at least eight copies per nuclear pore complex (NPC) but tdTomato intensity only shows a three modal labeling distribution [1, 2]. Cells were maintained in IMDM, supplemented with 10% fetal bovine serum (FBS) and penicillin/streptomycin. For microscopy, cells were cultured overnight on glass bottom dishes (Mattek) in IMDM+FBS. Prior to imaging, the media was changed to L15 (Gibco, Life Technologies) supplemented with 10% FBS and cells were transferred to a CO_2 free incubator until use. To label DNA rich compartments Vybrant DyeCycle Violet (Life Technologies) was used. A total of 1 ul of dye at 5mM concentration was added directly to the L15 media (2 mLs) and cells were incubated at 37 °C for 30 minutes prior to imaging. Cells were imaged at room temperature using Solis for acquisition (Andor).

5.A.2. Imaging system

We used a multi-focus microscope (MFM, Fig. 1) [3]. To achieve separation of the light collected by the objective into multiple planes a secondary pupil plane (Fourier plane) was formed inside a relay optic (L1 and L2) in which a distorted diffractive multi-focus grating (MFG) was placed, similar to [4]. The MFG was followed by a chromatic correction grating (CCG) and a refractive prism (Fig. 1). In the primary image plane behind the microscope's tube lens a square iris was placed to restrict the size of the image or field of view. The MFG splits the fluorescence light emitted from the sample into its 0 and 1st diffractive orders resulting in nine beams of similar intensity [5]. The pattern of the MFG is distorted introducing an angular offset to each image corresponds to a 2D diffractive order (m_x, m_y) of the grating. The CCG corrects for chromatic dispersion and was placed behind the MFG at a position where the diffractive orders were separated.

A multifaceted refractive prism was used to re-direct the individual beams towards the second lens of the relay optic that forms nine images on the camera. Color separation and band-pass filtering was done using a custom-made optic (dichroic mirror 6 μm thickness, 2" diameter with λ/10 flatness, band-pass for eYFP and tdTomato, Chroma).

The design of MFG, CCG and prism assumed a refractive index (R.I.) of 1.4, and a numerical aperture (NA) of 1.3. These values are closely matched using a 60x silicone immersion objective with an NA of 1.3 and R.I. of 1.41 for the immersion media. The effective maximal detection NA is given by multiplying the NA of the objective with the R.I. of the sample/buffer divided by the R.I. of the objective immersion. The expected effective or detection NA of the microscope is hence 1.3x1.41/1.33=1.23. Using beads for calibration we estimated the effective detection NA at 1.2, well in agreement with our expectation. The deviation from the design value for the NA, the effective NA being 1.2 rather than the expected NA of 1.3, results from imaging live cells in aqueous solution with a R.I. different from that of the objective immersion.

The need for a large field of view, high sensitivity and high time resolution (100 milliseconds per frame) was met using 780x761 pixel sub-regions on two back-illuminated EMCCDs (Andor 888, 2011 built). The pixels of this camera have a side length of 13 μm. The relay optic with L1 being a 150 mm lens and L2 being a 200 mm lens, introduces a slight magnification of 1.3 x. Together, the total magnification of objective (60x) and MFM (1.3x) was 80x, resulting in a pixel size in image space of 162.5 nm in x and y. The z-spacing is given by the MFG design and is 250 nm for the used grating. The size of the PSF can be estimated and $\sigma_{PSF-fullwidth}$ at 534 nm equals ~270 nm in x and y and ~520 nm in z. The full width at half maximum (FWHM) of the detected signal is defined as 2.355 σ, and is accordingly ~320 nm in x and y, and ~610 nm in z. For the mRNA, labeled with eYFP, these values result in lateral under-sampling of the PSF and axial Nyquist sampling. Color separation on the two cameras was done with an ultra flat dichroic mirror positioned after L2.

For fluorescence excitation, three lasers, 405 nm (Coherent, Cube), 514 nm (Coherent Sapphire) and 561 nm (Melles Griot), were combined into a single optical fiber and delivered to the microscope. Image acquisition was done with two identical cameras where mRNA (eYFP, 514 nm laser) and nuclear pores (tdTomato, 561 nm laser) were acquired simultaneously, while registration data were recorded immediately following (Vybrant DyeCycle Violet, 405 nm laser). The maximal cumulative excitation power was ~600 mW. For registration, two data sets were taken. First a 10 frame (= 1 sec) long 3D time series was recorded of the nucleus

with the same parameters used for mRNA and nuclear pore imaging. Next a fine registration data set was recorded where the 3D acquisition beam was z-stacked through the nucleus in 20 nanometer steps using a piezo objective stage.

5.A.3. Image registration marker

Use of cell external registration markers was impossible due to the need to fit 9 planes on one camera chip while maximizing the acquisition speed. The resulting field of view allows capturing the whole nucleus but not much of the cytoplasm or even the extracellular space. To achieve the highest degree of physical alignment and confidence, a cell internal marker is favorable. Such a marker would need to emit over the whole spectrum used in the experiment and would have to be switchable between on and off states so as to not overlay the fluorescent signal of the sample. A DNA marker, Vybrant DyeCycle Violet, fits this bill. Binding to double stranded DNA the dye emits a strong signal spanning from 400-700 nm when excited at ~400 nm. The strong fluorescence emission of the dye allows for excitation with minimal amounts of light.

Photodamage through low-level laser light excitation has been described in the literature [6]. In a straightforward experiment the doubling rate of yeast cells on the microscope was measured and a reduction was found with less than 100 mW input power at the back-focal plane. At 0.8 mW input power cell growth was practically arrested. The experiment was done using 488 nm and 561 nm laser light demonstrating higher sensitivity to blue than red light. Shifting the wavelength down to 405 nm hence is likely to impact cells even at very low power levels.

To avoid the possibility that our mRNA data are impacted by the registration procedure, the registration images were recorded immediately after acquiring the mRNA and NPC data. Similarly the possibility of damage to neighboring cells by general imaging conditions cannot be excluded or monitored in our experiments [7]. We hence imaged only a few well-separated cells per dish. Under these imaging conditions the Vybrant DyeCycle Violet provided an acceptable independent alignment marker based on the distribution of DNA in the nucleus, yielding additional information about nuclear occupation by DNA. The fact that the registration is taken independently from data acquisition is unavoidable, and adds a possible source of error to the assumed static distribution of DNA during the duration of imaging.

5.A.4. Image registration in silico

The software as described in the paper is implemented in ImgLib2 [8] and available on GitHub (https://github.com/StephanPreibisch/MFM-Align), published under the GPLv3 license. The redundant geometric local descriptor matching is also available as a Fiji plugin (http://fiji.sc) distributed through the Fiji Updater [9]. It can be found under Plugins > Registration > Descriptor based Registration (2d/3d) and Plugins > Registration > Descriptor based Series Registration (2d/3d+t)

5.A.5. Image registration precision

Image plane registration is achieved through detection of corresponding image features and fitting of a transformation model that minimizes the distances between all corresponding features. The number and distribution of corresponding features within the image, localization precision, and the transformation model influence the final quality of the registration.

To achieve optimal registration quality, the transformation model should require a minimal amount of parameters to solve, while still being able to describe all significant transformations that occur during acquisition. As the remaining distance between corresponding features after applying the transformation is normally distributed (Fig. 4A and B), maximizing the number of corresponding features will minimize the final registration error and at the same time provide a high confidence that the registration result is correct. In other words, being able to identify thousands of corresponding detections that all support the same transformation model (requiring two corresponding features to be solved) shows that the conservative approach of using a rigid transformation model is sufficient to achieve registration precision of 0.5 pixel (Fig. 4). Feature based image registration currently supports regularized translation, rigid, similarity, affine and homography transformation models and is available both as source code or FIJI function.

To test and compare feature based registration against alignment with diffraction limited beads we labeled 1 μm beads (non-fluorescent NeutrAvidin labeled FluoSpheres, Life Technologies) with biotinylated plasmid DNA (*Label* IT, Mirus) and stained with Vybrant DyeCycle Violet. The beads were embedded in an agarose gel and imaged in the green and red MFM channel. Registration was done for the beads as for the cells. As a control we embedded 0.2 μm diameter TetraSpeck beads (Life Technologies) in the same gel. This allows to directly compare a Gaussian fit based alignment of diffraction limited spots to our quadratic fit using feature based alignment (Fig. 4C).

5.A.6. Image deconvolution

As originally reported, multi-focal microscopy using a single MFG for a wide range of emission light bandwidth (here eYFP and tdTomato) will produce a wavelength dependent defocus. In real live applications the experimental conditions might not always match the optical design to the number, adding an additional source of image distortion. As discussed above (see "Imaging System"), the use of living cells in an aqueous buffer resulted in a mismatch of the refractive index between objective immersion and sample. This situation is common to the use of high NA objectives, but due to the inherent 3D nature of our image acquisition the spherical aberrations introduced on top of the expected defocus need to be taken into account in image processing.

A defocus is necessary to focus at a distance Δz in the sample depends on the optical path difference between the ray over the optical axis and under an angle θ [3]. This is described as follows:

$$
\begin{aligned}
D_i &= \Delta z \left(k_i - k_{z_i} \right) \\
&= n \, \Delta z \, k_0 \left(1 - \sqrt{1 - \frac{k_x^2 + k_y^2}{k_i^2}} \right) \\
&= n_i \, \Delta z \, k_0 \left(1 - \cos\theta_i \right)
\end{aligned}
\tag{5.1}
$$

Here n_i is the refractive index of medium i, k_0 is the length of the wave vector in vacuum, k_{z_i} is the wave vector length of the z component in medium i, Δz is the thickness of the medium. The differences in sign and constant phase offset to the original publication can be ignored. In addition to the original approach, we had to compensate for the spherical aberration introduced by refractive index mismatch. The spherical aberration (SA) due to index of refraction mismatch [10, 11] is given by:

$$
\begin{aligned}
SA_{1,2} &= \Delta z \left(k_1 - k_{z_1} \right) - \Delta z \left(k_2 - k_{z_2} \right) \\
&= n_1 \, \Delta z \, k_0 \left(1 - \cos\theta_1 \right) - n_2 \, \Delta z \, k_0 \left(1 - \cos\theta_2 \right)
\end{aligned}
\tag{5.2}
$$

Here Δz is the thickness of medium 2, θ_1 is the angle in medium 1 and θ_2 the angle in medium 2. Note that this expression does not only contain spherical aberration (of all orders) but also defocus. After estimation and correction of the z-dependence of spherical aberrations in each data set (Fig. 3), a momentum

preserving deconvolution with the distorted PSF is done to completely correct for limits of the used MFG, using Huygens professional software (SVI, The Netherlands). Momentum preserving deconvolution was done on the aligned image volumes without further image processing. For display, the image contrast was adjusted and the Dots Per Inch (DPI) were adjusted without interpolation. All analysis was preformed on the deconvolved, registered images. A 3D smoothing (σ = 1 pixel) was applied to the DNA channel.

5.A.7. mRNA detection

Signal variability causes difficulty in separating noise from real fluorescence. This is due to unbound fluorescently labeled MS2 coat proteins, out of focus mRNA particles located in other planes along the optical axis, auto-fluorescence of the cell as well as the limited spectral separation of eYFP and tdTomato [1]. A simple threshold for particle identification is highly error prone which lead us to using the methodology of scale space where we can search for Gaussian structures that have the width of a diffraction limited spot while simultaneously estimating and subtracting the local background variations. To create comparability between different cells, a uniform parameter is needed to be able to discriminate between spots originating from noise and from signal. This was done by sorting the intensities of all candidate mRNAs for intensity and selecting the stationary point of the obtained graph as the discriminatory threshold of noise versus signal (β-actin mRNA). The DNA channel was smoothed with a Gaussian filter (σ=1) and then binary thresholds were used to segment the nucleus.

References

[1] D. Grünwald and R. H. Singer, *In vivo imaging of labelled endogenous β-actin mRNA during nucleocytoplasmic transport,* Nature **467**, 604 (2010).

[2] T. Lionnet, K. Czaplinski, X. Darzacq, Y. Shav-Tal, A. L. Wells, J. A. Chao, H. Y. Park, V. de Turris, M. Lopez-Jones, and R. H. Singer, *A transgenic mouse for in vivo detection of endogenous labeled mRNA,* Nature Methods **8**, 165 (2011).

[3] S. Abrahamsson, J. Chen, B. Hajj, S. Stallinga, A. Y. Katsov, J. Wisniewski, G. Mizuguchi, P. Soule, F. Mueller, C. D. Darzacq, *et al.*, *Fast multicolor 3D imaging using aberration-corrected multifocus microscopy,* Nature Methods **10**, 60 (2013).

[4] P. M. Blanchard and A. H. Greenaway, *Simultaneous multiplane imaging with a distorted diffraction grating,* Applied Optics **38**, 6692 (1999).

[5] J. N. Mait, *Understanding diffractive optic design in the scalar domain,* Journal of the Optical Society of America A **12**, 2145 (1995).

[6] P. M. Carlton, J. Boulanger, C. Kervrann, J.-B. Sibarita, J. Salamero, S. Gordon-Messer, D. Bressan, J. E. Haber, S. Haase, L. Shao, *et al.*, *Fast live simultaneous multiwavelength four-dimensional optical microscopy,*

Proceedings of the National Academy of Sciences **107**, 16016 (2010).

[7] J. W. Dobrucki, D. Feret, and A. Noatynska, *Scattering of exciting light by live cells in fluorescence confocal imaging: Phototoxic effects and relevance for frap studies,* Biophysical Journal **93**, 1778 (2007).

[8] T. Pietzsch, S. Preibisch, P. Tomančák, and S. Saalfeld, *ImgLib2—generic image processing in java,* Bioinformatics **28**, 3009 (2012).

[9] J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, *et al.*, *Fiji: An open-source platform for biological-image analysis,* Nature Methods **9**, 676 (2012).

[10] J. Braat, *Influence of substrate thickness on optical disk readout,* Applied Optics **36**, 8056 (1997).

[11] S. Stallinga, *Compact description of substrate-related aberrations in high numerical-aperture optical disk readout,* Applied Optics **44**, 849 (2005).

5

Nuclear Accessibility of mRNA

6

Iterative Linear Focal-Plane Wavefront Correction

Equipped with our five senses — along with telescopes and microscopes and mass spectrometers and seismographs and magnetometers and particle accelerators and detectors sensitive to the entire electromagnetic spectrum — we explore the universe around us and call the adventure science.

EDWIN HUBBLE

Carlas S. Smith
Raluca Marinica
Arjan J. den Dekker
Michel Verhaegen
Christoph U. Keller
Niek Doelman

Abstract

We propose an efficient approximation to the nonlinear phase diversity method for wavefront reconstruction and correction from intensity measurements with potential of being used in real-time applications. The new iterative linear phase diversity method assumes that residual phase aberration is small and makes use of a first order Taylor expansion of the point spread function, which allows for arbitrary (large) pupil shapes in order to optimize the phase retrieval. For static aberrations, in each step the residual phase aberrations are estimated based on one defocused image by solving a linear least squares problem, and corrected with a deformable mirror. Due to the fact that the linear approximation does not have to be updated and it becomes more accurate with each correction step, the computational complexity of the method is a fraction of $\mathcal{O}\left(m^2\right)$, where m^2 is the number of pixels. The convergence of the iterative linear phase diversity correction steps has been investigated and numerically verified. The comparative study that we make demonstrates the improved performance in computational time with no decrease in accuracy with respect to existing methods that also linearize the point spread function.

6.1. *Introduction*

All optical measurements are subject to optical aberrations either coming from exterior sources or intrinsic to the instrument. If the aberrations can be estimated, they can be compensated for either through adaptive optics during image acquisition or post-processing. One method that has been used mostly in post processing is phase diversity (PD) [1]. PD estimates wavefront aberrations using nonlinear optimization techniques from multiple images of the same unknown scene acquired simultaneously, which contain additional user introduced aberrations, the latter referred to as diversities. To be able to uniquely estimate wavefront aberrations, more than one in focus image is needed [2], because rotating a wavefront by 180° and flipping its sign produces the same point spread function (PSF) as the original wavefront [3]. The resulting optimization problem is nonlinear and is known to be computationally complex due to the repetitive evaluations of Fourier transforms. In addition, the method is also prone to converge to local minima [3]. As a consequence, nonlinear phase diversity has a limited usage in real-time correction algorithms [2], and different ideas have been presented to decrease the complexity of the calculations. These ideas can be split up into Fourier domain [4–7] and spatial domain [8, 9] techniques. The Gerchberg–Saxton (GS) [4] algorithm is one of the oldest and best known Fourier domain techniques, which is an iterative algorithms for retrieving the phase from intensity measurements. Spatial domain techniques make use of a local model for the PSF, but do not use the Fourier transform. The common idea in decreasing the computational complexity is the approximation of the PSF based on the assumption that the total aberration is small [5, 6, 10]. This small-phase assumption is associated in the literature with the Born approximation [5, 11, 12], which implicitly assumes that the diversity used is small.

Recently, in [6], it was shown that using a second order expansion of the generalized pupil function (GPF), wavefront retrieval algorithms give more accurate results than using the Born approximation, which results from a linear expansion of the GPF. The key assumption of these methods is that the sum of the diversity and the aberration is small. However, as has been shown in [13], the optimal diversity depends on the present aberration and can generally not be considered small. In the present paper, we overcome this shortcoming by the use of an alternative approximation of the PSF. The linearization of the PSF is done around zero aberration and a (possibly large) diversity and it is suited for small values (~ 0.5 radians (rad) root mean square (rms) [11, 12]) of the phase aberration. The iterative manner in which the method is applied compensates for this small phase

assumption. In this context, the use of one image is enough for the uniqueness of the phase estimate [14].

A similar approximation is used in [15] or [8]. Both [15] and [8] use only one image for the phase retrieval. In [15], an analysis is made for the best defocus measurement plane for robust phase retrieval. The method in [8], linearized focal-plane technique (LIFT), performs several iterations using the same recorded image. In the first step, the linearization of the PSF is obtained from a Taylor expansion of the nonlinear PSF taken around zero (aberration) and a least squares (LS) problem is solved yielding an estimate of the wavefront aberration. For the next iterations, the linearization of the PSF is taken around the current estimate of the Zernike coefficients and again a linearization step is performed and a linear least squares problem is solved using the same image as in the previous step. The authors mention that more than three iterations do not yield significant improvements. Using only the PSF approximation in the first iteration in [8], we present a novel iterative linear phase diversity (ILPD) method, which consists in iteratively collecting one defocused image with a fixed known defocus, solving a least squares problem obtained from the linearization of the PSF around zero aberration and correcting for the wavefront aberrations by the least squares estimate. As opposed to LIFT, which can estimate several modes from a full pupil image by approximating the PSF iteratively around the current estimate of the aberrations, we use the approximation around zero aberration and collect one new image (which includes the previous corrections) at each iteration. In this way, we speed up the algorithm due to the fact that the linear coefficients of the PSF do not change from iteration to iteration. The method reduces to a matrix-vector multiplication and has computational complexity a fraction of $\mathscr{O}\left(\left(2m^2 - 1\right)n\right)$ - where m^2 is the number of pixels, and n is the number of Zernike coefficients used in the wavefront expansion. This is due to the fact that only a part of the pixels can be used for the estimation. The gain in computational time sets the premises for using this method in a dynamic setting for time varying wavefronts.

The paper is organized as follows. In Section 6.2.1 we present the general problem and introduce the PSF of the optical system and the noise model. In Section 6.2.2 we review four linear and quadratic PSF approximations and show the advantages and disadvantages of each of them, which we prove in Appendix 6.A. In Section 6.2.3 we use the previously mentioned approximations and present the iterative linear phase diversity solution. In Section 6.2.4 we discuss results of numerical simulations and compare them to the ones in [8]. We end with conclusions in Section 6.3.

Some mathematical notations used are standard: ^{T} and * denote transposition and transpose conjugation, respectively, $\star$ denotes the convolution operator, $\| \bullet \|$ denotes the vector 2-norm, $\mathscr{O}(\bullet)$ describes the complexity of a function when the argument tends towards a particular value, usually in terms of simpler functions, $\mathscr{O}(\| \bullet \|^a)$ is the a-th order Lagrange residue, $\mathbb{R}$ and $\mathbb{C}$ are the sets of real and complex numbers, respectively, $\mathbb{R}^{m \times n}$ and $\mathbb{C}^{m \times n}$ are the sets of $m \times n$ matrices with elements in the set of real or complex numbers, respectively.

6.2. *Results*

6.2.1. The optical system

In this section, a model is presented for the image formation of a point source in the presence of phase aberrations $\phi \in \mathbb{R}^{m^2 \times 1}$, approximated using a normalized Zernike basis [16]

$$\phi(u_j, v_j) = Z(u_j, v_j)^T \alpha, \tag{6.1}$$

where $\alpha \in \mathbb{R}^{n \times 1}$ contains the Zernike coefficients corresponding to the unknown aberration and $Z \in \mathbb{R}^{n \times m^2}$ is a matrix containing the n Zernike polynomials evaluated in the pupil plane coordinates (u_j, v_j). Besides the "in focus" image, PD uses additional images with known diversities. The phase aberration in the i-th diversity image is

$$\phi_i(u_j, v_j) = Z(u_j, v_j)^T (\alpha + \beta_i), \tag{6.2}$$

where $\beta_i \in \mathbb{R}^{n \times 1}$ is a known diversity. These phase aberrations nonlinearly influence the PSF. The incoherent image formation of a point source is given by [17]

$$y_{i,j} = \mu_i h\left(s_j, t_j, \alpha, \beta_i\right) + n_i(s_j, t_j), \tag{6.3}$$

where $y_{i,j}$ denotes the j-th pixel of the i-th diversity image, μ_i is the number of photons (the expected arrival rate multiplied with the integration time of the camera), h denotes the spatially invariant PSF expressed in the spatial coordinates (s_j, t_j) with aberration α and user introduced diversity β_i, and $n_i(s_j, t_j)$ is Gaussian white noise with standard deviation $\sigma_{i,j}$, which we assume to be equal for all pixels by dropping the index j. If only a defocus aberration is present, the schematic representation of adding a defocus diversity is given in Fig. 6.1.

140
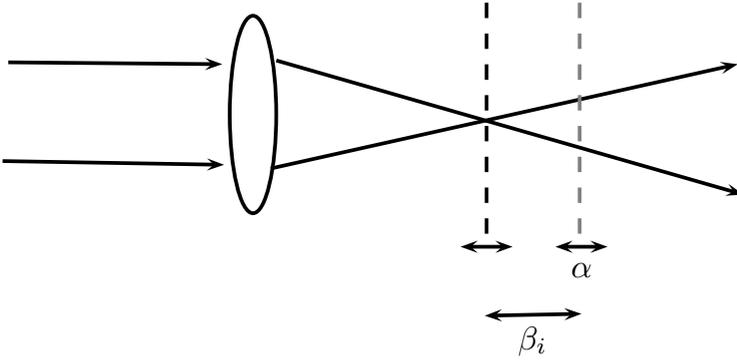
Figure 6.1. Optical system; Focal plane (black); Defocus plane (gray); Unknown small aberration Zernike coefficients (α); known arbitrary diversity Zernike coefficients (β_i)

In iterative linear phase diversity, assuming that the aberrations do not change in the time window considered, at time k in that time window we obtain a least squares estimate $\hat{\alpha}_k$ of α_k as described in Section 6.2.3. Next, we assume that we have a deformable mirror (DM) that acts perfectly in the space of Zernike coefficients, which we use to correct for the wavefront. The residual wavefront $\Delta\alpha_k = \alpha_k - \hat{\alpha}_k$, will be again estimated and corrected for at the next step $\alpha_{k+1} = \Delta\alpha_k$, until a desired tolerance is reached.

In Subsection 6.2.1 the aberrated point spread function presented in Eq. (6.3) is derived. Subsequently, in Subsection 6.2.1 the measurement noise is presented.

Image formation
The spatially invariant PSF of the i-th optical path in Eq. (6.3) is given by [17]

$$
\begin{aligned}
&h\left(s_j, t_j, \alpha, \beta_i\right) \\
&= \mathscr{F}\left(\Pi(u,v)\exp\left(i\phi_i(u,v)\right)\right)\left(s_j, t_j\right) \\
&\quad \times \mathscr{F}\left(\Pi(u,v)\exp\left(i\phi_i(u,v)\right)\right)^*\left(s_j, t_j\right) \\
&= \mathscr{F}\left(\left(\exp\left(i\phi_i\left(u',v'\right)\right)\Pi\left(u',v'\right)\right.\right. \\
&\quad \left.\left. \star \exp\left(-i\phi_i\left(-u',-v'\right)\right)\Pi\left(-u',-v'\right)\right)(u,v)\right)\left(s_j, t_j\right),
\end{aligned}
\tag{6.4}
$$

where $\mathscr{F}(\cdot)$ is the Fourier transform, $\left(u_j, v_j\right) = \left(2\pi s_j/f\lambda, 2\pi t_j/f\lambda\right)$, f is the focal length, λ is the wavelength, ϕ_i is the phase and Π is the pupil function. Next, we define the GPF as

$$
p\left(u_j, v_j, \alpha, \beta_i\right) = \Pi\left(u_j, v_j\right)\exp\left(i\phi_i\left(u_j, v_j\right)\right).
\tag{6.5}
$$

Using Eq. (6.5), the optical transfer function (OTF) is given by

$$
\begin{aligned}
& W(u_j, v_j, \alpha, \beta_i) \\
& = \left(p\left(u, v, \alpha, \beta_i\right) \star p\left(-u, -v, \alpha, \beta_i\right)^* \right)(u_j, v_j).
\end{aligned}
\tag{6.6}
$$

Next, we introduce the short hand notations

$$
\begin{aligned}
p_j(\alpha, \beta_i) &:= p\left(u_j, v_j, \alpha + \beta_i\right), \\
p_{-j}(\alpha, \beta_i) &:= p\left(-u_j, -v_j, \alpha + \beta_i\right),
\end{aligned}
\tag{6.7}
$$

and

$$
\begin{aligned}
W_j(\alpha, \beta_i) &:= W(u_j, v_j, \alpha, \beta_i), \\
h_j(\alpha, \beta_i) &:= h\left(s_j, t_j, \alpha, \beta_i\right).
\end{aligned}
\tag{6.8}
$$

Using Eqs. (6.7), (6.8), and (6.4), the PSF becomes

$$
\begin{aligned}
h_j(\alpha, \beta_i) &= \mathscr{F}\left(p_j(\alpha, \beta_i) \star p_{-j}^*(\alpha, \beta_i) \right)(s_j, t_j) \\
&= \mathscr{F}\left(W_j(\alpha, \beta_i) \right)(s_j, t_j),
\end{aligned}
\tag{6.9}
$$

where, for simplicity, we drop the convolution brackets and its new coordinates and only use the star operator.

Measurement noise

We consider here two main noise processes that are dependent on the exposure time and luminosity of the object, namely the Gaussian read-out noise and the photon counting noise. We approximate the photon noise by an additive zero-mean Gaussian noise, with a variance equal to the flux. The read-out noise is the same for each pixel and follows a Gaussian distribution, which has the property that all the pixels are mutually uncorrelated. The signal to noise (SNR) level is given by

$$
\mathrm{SNR} = \frac{1}{m^2} \frac{\sum_{j=1}^{m^2} \mu_i h_j(\alpha, \beta_i)}{\sqrt{\frac{1}{m^2} \sum_{j=1}^{m^2} \sigma_i^2}} \left(= \frac{\mu_i}{m^2 \sigma_i} \right),
\tag{6.10}
$$

where μ_i and $h_j(\alpha, \beta_i)$ follow from Eq. (6.3) and σ_i is the standard deviation of the read-out noise in the i-th diversity image. The total noise has zero mean and its variance is the sum of the two variances. The approximation of the photon noise is made only in the theoretical part of the paper. For the numerical simulations, photon noise modeled using a Poisson distribution.

142

6.2.2. Approximations of the PSF

In the previous section, we have derived the relation between the aberrations and the observed image. The resulting nonlinear integral form in Eq. (6.4) is computationally expensive to evaluate and needs to be approximated. The approximations we use enable us to present the iterative linear phase diversity solution in the next section. The generally used approach is the Born approximation, which results into a valid mapping for a small phase of the wavefront aberration (up to 0.5 rad rms) [12]. It has been shown in [13] that the lower bound on the variance of any unbiased estimator of the wavefront aberration is much lower for large diversities. In [13], it was further shown that e.g. a defocus with an rms of 2 rad on average results in the lowest bound for Poisson noise. However, the Born approximation relies on the fact that the diversities are also small. Therefore, it is of high importance to investigate other approximations of Eq. (6.4) that allow the use of large diversities. In this section, we motivate our choice for a simplified model. We start by describing 4 possible approximations of the PSF/OTF and the disadvantages/advantages of each of them. The proofs of the properties stated in this section are given in Appendix 6.A.

First order approximations

The approximations presented here are all based on a linear Taylor expansion of the GPF or of the PSF, respectively. In Subsection 6.2.2, the assumption is that both the wavefront phase and the diversity used are small. We approximate the GPF with a linear expression and compute the coefficients of the resulting quadratic PSF. The approximation given in Subsection 6.2.2 can be used for small wavefronts and a general diversity. This is simply the Taylor expansion of the PSF. It approximates the PSF around the diversity with a linear expression.

Small total phase approximation

The Born approximation assumes a small phase, $\phi_i = Z^T (\alpha + \beta_i)$, such that the GPF can be approximated using only a first order Taylor expansion around $\alpha + \beta_i = 0$. The consequence is that the GPF can be written as

$$
\begin{aligned}
p_j(\alpha, \beta_i) &= p_j(\alpha, \beta_i)\big|_{\alpha+\beta_i=0} + \frac{\partial p_j(\alpha,\beta_i)}{\partial(\alpha+\beta_i)}\Big|_{\alpha+\beta_i=0} (\alpha + \beta_i) \\
&\quad + \mathcal{O}\left(\|\alpha + \beta_i\|^2\right), \\
&= \Pi\left(u_j, v_j\right)\left(1 + i Z^T\left(u_j, v_j\right)(\alpha + \beta_i)\right) \\
&\quad + \mathcal{O}\left(\|\alpha + \beta_i\|^2\right).
\end{aligned}
\tag{6.11}
$$

Substituting this first order (Born) approximation of the GPF in Eq. (6.4) yields a quadratic PSF

$$h_j(\alpha, \beta_i) = A_{0,j} + A_{1,j}(\alpha + \beta_i) + (\alpha + \beta_i)^T A_{2,j}(\alpha + \beta_i)$$
$$+ \mathcal{O}\left(\|\alpha + \beta_i\|^2\right), \tag{6.12}$$

where

$$
\begin{aligned}
A_{0,j} &:= \left. \mathcal{F}\left(p_j(\alpha, \beta_i) \star p_{-j}^*(\alpha, \beta_i)\right)(s_j, t_j) \right|_{\alpha + \beta_i = 0} \\
&= \left. h_j(\alpha, \beta_i) \right|_{\alpha + \beta_i = 0},
\end{aligned}
\tag{6.13}
$$

$$
\begin{aligned}
A_{1,j} &:= \mathcal{F}\left(\frac{\partial p_j(\alpha, \beta_i)}{\partial(\alpha + \beta_i)} \star p_{-j}^*(\alpha, \beta_i) \right. \\
&\quad + \left. \left. p_j(\alpha, \beta_i) \star \frac{\partial p_{-j}^*(\alpha, \beta_i)}{\partial(\alpha + \beta_i)^T} \right)(s_j, t_j) \right|_{\alpha + \beta_i = 0} \\
&= \left. \frac{\partial h_j(\alpha, \beta_i)}{\partial(\alpha + \beta_i)} \right|_{\alpha + \beta_i = 0},
\end{aligned}
\tag{6.14}
$$

$$
\begin{aligned}
A_{2,j} &:= \mathcal{F}\left(\frac{\partial p_j(\alpha, \beta_i)}{\partial(\alpha + \beta_i)} \star \frac{\partial p_{-j}^*(\alpha, \beta_i)}{\partial(\alpha + \beta_i)^T} \right. \\
&\quad + \left. \left. \frac{\partial p_j(\alpha, \beta_i)}{\partial(\alpha + \beta_i)} \star \frac{\partial p_{-j}^*(\alpha, \beta_i)}{\partial(\alpha + \beta_i)^T} \right)(s_j, t_j) \right|_{\alpha + \beta_i = 0}.
\end{aligned}
\tag{6.15}
$$

Property 6.2.1. *The linear term of the approximated PSF in Eq. (6.12) is invariant in the even aberrations.*

Property 6.2.1 makes it impossible to neglect the quadratic term of the PSF when the Born approximation is used to formulate an estimation problem. This is also what [15] states - the even modes are not observable in the "in focus" intensity image and you have to go out of focus in order to be sensitive to them. This will turn out to be equivalent to the approximation we present in the next subsection. But we also go further and present a quadratic approximation for an out of focus PSF. We do not solve the quadratic estimation problem here, but this should increase the estimation accuracy. The downside would be an increase in the computational time for solving the quadratic problem if no structure is assumed on the matrices involved.

Small aberration approximation
Another first order model is obtained by directly approximating the PSF in Eq. (6.4) for small aberrations and non-zero diversities. The first order Taylor approxi-

mation of the PSF in $\alpha = 0$ is given by

$$h_j(\alpha, \beta_i) = B_{0,j}(\beta_i) + B_{1,j}(\beta_i)\alpha + \mathcal{O}\left(\|\alpha\|^2\right), \tag{6.16}$$

where

$$
\begin{aligned}
B_{0,j}(\beta_i) &:= \left. h_j(\alpha, \beta_i)\right|_{\alpha=0}, \\
B_{1,j}(\beta_i) &:= \left. \frac{\partial h_j(\alpha, \beta_i)}{\partial \alpha}\right|_{\alpha=0}.
\end{aligned}
\tag{6.17}
$$

Property 6.2.2. *The linear terms of the first order approximation of the PSF and of the PSF resulting from the first order Taylor approximation of the GPF are equal.*

Property 6.2.3. *The approximation in Eq. (6.16) has the property that for $\phi \neq 0$ the even modes do not cancel in the linear term.*

Remark 6.2.1. *Note that this approximation is valid for any diversity. Due to the fact that the linear term is not invariant in the even modes, we can estimate the even and odd modes with just a linear equation as will be shown in a later section.*

Second order approximations

In this section, we present two quadratic approximations of the PSF, namely, starting from the second order Taylor approximation of the GPF, and the second order Taylor approximation of the PSF, respectively. As we mention in the next section, second order approximations could also be used to formulate linear estimation problems. Another motivation for presenting them is that they give more accurate phase estimates and can easily be used in a dynamic setting where a Kalman filter [18] can be included.

Small total phase approximation

It has been shown in [6] that an additional quadratic term leads to a more accurate PSF approximation than using the Born approximation. This term is obtained using a second order Taylor expansion of the GPF in $\phi = 0$ and neglecting the 3rd and the 4th orders of the resulting PSF. The resulting approximation is given by

$$
\begin{aligned}
h_j\left(\alpha, \beta_i\right) = {}& C_{0,j} + C_{1,j}(\alpha + \beta_i) + (\alpha + \beta_i)^T C_{2,j}(\alpha + \beta_i) \\
& + \mathcal{O}\left(\|\alpha + \beta_i\|^3\right),
\end{aligned}
\tag{6.18}
$$

where

$$
\begin{aligned}
C_{0,j} &:= h_j(\alpha, \beta_i)\big|_{\alpha+\beta_i=0} \left(= A_{0,j}\right), \\
C_{1,j} &:= \frac{\partial h_j(\alpha,\beta_i)}{\partial(\alpha+\beta_i)}\bigg|_{\alpha+\beta_i=0} \left(= A_{1,j}\right), \\
C_{2,j} &:= \frac{\partial^2 h_j(\alpha,\beta_i)}{\partial(\alpha+\beta_i)\partial(\alpha+\beta_i)^T}\bigg|_{\alpha+\beta_i=0} \\
&= A_{2,j} + \mathscr{F}\left(\frac{\partial^2 p_j(\alpha,\beta_i)}{\partial(\alpha+\beta_i)\partial(\alpha+\beta_i)^T} \star p^*_{-j}(\alpha,\beta_i)\right. \\
&\quad + \left. p_{-j}(\alpha,\beta_i) \star \frac{\partial^2 p^*_{-j}(\alpha,\beta_i)}{\partial(\alpha+\beta_i)\partial(\alpha+\beta_i)^T}\right)(s_j, t_j)\bigg|_{\alpha+\beta_i=0}.
\end{aligned}
\tag{6.19}
$$

Property 6.2.4. *The expression in Eq. (6.18) is also obtained when the PSF is approximated using a second order Taylor expansion around $\phi = 0$.*

Small aberration approximation
The second order Taylor approximation of the PSF is given by

$$
\begin{aligned}
h_j\left(\alpha, \beta_i\right) = {}& D_{0,j}(\beta_i) + D_{1,j}(\beta_i)\alpha + \alpha^T D_{2,j}(\beta_i)\alpha \\
& + \mathcal{O}\left(\|\alpha\|^3\right),
\end{aligned}
\tag{6.20}
$$

where

$$
\begin{aligned}
D_{0,j}(\beta_i) &:= h_j(\alpha, \beta_i)\big|_{\alpha=0} \left(= B_{0,j}(\beta_i)\right), \\
D_{1,j}(\beta_i) &:= \frac{\partial h_j(\alpha, \beta_i)}{\partial \alpha}\bigg|_{\alpha=0} \left(= B_{1,j}(\beta_i)\right), \\
D_{2,j}(\beta_i) &:= \frac{\partial^2 h_j(\alpha, \beta_i)}{\partial \alpha \partial \alpha^T}\bigg|_{\alpha=0}.
\end{aligned}
$$

Property 6.2.5. *The second order Taylor approximation of the PSF in $\phi = 0$ is more accurate than the PSF obtained from the first order GPF approximation in $\phi = 0$, while the quadratic form remains.*

Property 6.2.6. *The second order Taylor approximation of the PSF in $\phi \neq 0$ has the property that the even modes do not cancel in the linear term of the PSF.*

6.2.3. Iterative linear phase diversity
In the previous section, we have presented different first and second order approximations of the PSF. In this section we study static aberration estimation and correction techniques based on linear least squares.

Apart from Eq. (6.16), all the other approximations of the PSF derived in Section 6.2.2 given by Eqs. (6.12), (6.18), and (6.20) are quadratic in the unknown

aberration, as represented by α. Here, we aim to obtain a linear relationship between the measured intensity and the unknown aberration due to the fact that a linear model has low computational complexity and gives rise to fast algorithms. The approximations in Eqs. (6.12) and (6.18) are based on the Taylor series expansion of the GPF, first order and second order, respectively, and only the terms of the PSF up to the second order are retained. This means that Eq. (6.18) is more accurate than Eq. (6.12) which motivates its preferred use. In order to obtain a linear formulation using the approximation in Eq. (6.18), we could take the difference of two measurements as done in [12]. Note that this artifice can not be performed on Eq. (6.20), because the coefficients $D_{0,j}$, $D_{1,j}$ and $D_{2,j}$ are functions of the diversities β_i and they do not cancel when two measurements are subtracted, such that the only solution when using this type of approximation is to retain only the linear term as in Eq. (6.16). Nevertheless, subtracting two measurements, first of all implies obtaining two measurements, which we want to avoid in this work, and secondly, the numerical estimation problem is not well conditioned when noise is also subtracted. Also, the SNR decreases when taking differences and the following property states this.

Property 6.2.7. *Taking the difference between two images significantly decreases the SNR.*

In what follows, we form a linear system using the approximation in Eq. (6.16). For this, we use one defocused image. The solution of iterative linear phase diversity using the previously mentioned approximation is compared with the solution given in [8]. The approximation in Eq. (6.16), is already linear, and we have as follows

$$Y_1 = b_S + A_S \alpha + \Delta b_S(\alpha) + n_1, \tag{6.21}$$

$$\Delta b_S(\alpha) := \mathcal{O}\left(\|\alpha\|^2\right), \tag{6.22}$$

where $Y_i := \left[y_{i,1}...y_{i,j}...\right]^T$, $b_S := \left[\tilde{b}_{S,1}...\tilde{b}_{S,i}...\right]^T$, with $\tilde{b}_{S,i} := \left[B_{0,i,1}...B_{0,i,j}...\right]$, $A_S := \left[\tilde{A}_{S,1}^T...\tilde{A}_{S,i}^T...\right]^T$, with $\tilde{A}_{S,i} := \left[B_{i,1}(\beta_i)^T...B_{i,j}(\beta_i)^T...\right]^T$, and $B_{0,j}(\beta_i)$, $B_{1,j}(\beta_i)$ are defined in Eq. (6.17). The index i is kept when defining the quantities above to suggest that this method can easily be generalized to more than one image if the optical system can facilitate this, while still keeping the linear system formulation.

The main advantages of the first order Taylor approximation of the PSF in Eq. (6.16) are that it is possible to approximate the PSF at large diversities and that the first order term is not invariant in the even modes, which makes it possible to

estimate them (except for piston). Therefore, we do not have to subtract images, which significantly decreases the SNR.

In iterative linear phase diversity, the residual aberration is repeatedly estimated and compensated for with a DM using the residual aberration estimate. Assuming that the DM can fully compensate for the estimated residual aberration, then, denoting the residual aberration estimate at the k-th correction step by $\hat{\alpha}_{k-1}$, and denoting the residual aberration at the k-th correction step by α_{k-1}, we obtain

$$
\begin{aligned}
\alpha_k &= \Delta\alpha_{k-1} \\
&= \alpha_{k-1} - \hat{\alpha}_{k-1}.
\end{aligned}
\tag{6.23}
$$

At the k-th correction step, one image, $Y_{1,k}$, is recorded with diversity β_1 assuming the wavefront aberration does not change. The additional index k of Y denotes the correction step. From the new image, a new estimate of α_k is obtained via the solution of a least squares problem based on Eq. (6.21). The algorithm continues until the strength of the aberration decreases under a certain threshold or a finite number of correction steps has been performed. Let Eq. (6.21) (where the step index k has been added) be rewritten as

$$
\overline{b}_{S,k} - \Delta b_S(\alpha_k) = A_S\alpha_k + n_k,
\tag{6.24}
$$

where $\overline{b}_{S,k} := Y_{1,k} - b_S$, $n_k = n_{1,k}$, and

$$
\begin{aligned}
\Delta b_S(\alpha_k) &= \mathcal{O}\left(\|\alpha_k\|^2\right) \\
&= C_S\|\alpha_k\|^2,
\end{aligned}
\tag{6.25}
$$

with C_S a constant defined by the Lagrange remainder. Then, the LS problem that needs to be solved is

$$
\min_{\alpha_k} \|\overline{b}_{S,k} - A_S\alpha_k\|^2.
\tag{6.26}
$$

The solution of (6.26) after each correction step k (no correction for at the 0-th correction step) with the deformable mirror will be indicated by the ILPD method for joint wavefront estimation and correction.

Convergence analysis

In this section, we study the convergence behavior of the iterative linear phase diversity method in the absence of measurement noise. Using Eq. (6.23), the

relative residue after correction step k is given by

$$r_{\mathrm{LS}} := \frac{\|\alpha_{k+1}\|}{\|\alpha_k\|}. \tag{6.27}$$

The relative residue has to be smaller than one to ensure that the rms value of the wavefront is reduced. If this is not the case, the rms value increases or remains constant. Therefore, the convergence can be quantified using this quantity. We validate by Monte Carlo simulation that using the PSF approximation proposed in Eq. (6.16) we converge to an unbiased estimate faster than the method in [8].

For the linear system in Eq. (6.24), in the noiseless case, we can compute an approximate upper bound on the relative error in the solution [19] as

$$\frac{\|\hat{\alpha}_k - \alpha_k\|}{\|\alpha_k\|} \lesssim \frac{\|\Delta b_S(\alpha_k)\|}{\|\overline{b}_{S,k}\|} \left\{ \frac{2\kappa(A_S)}{\cos(\theta)} + \tan(\theta)\,\kappa(A_S)^2 \right\}, \tag{6.28}$$

where $\kappa(A_S)$ denotes the condition number of A_S and θ is the acute angle between the vectors $A_S\hat{\alpha}_k$ and $\overline{b}_{S,k}$. For a well conditioned matrix, the bound depends on $\Delta b_S(\alpha_k)$. As $\|\alpha_k\|$ decreases, $\|\Delta b_S(\alpha_k)\|$ decreases and the bound becomes zero in the limit

$$\lim_{\alpha_k \to 0} \frac{C_S\|\alpha_k\|^2}{\|\overline{b}_{S,k}\|} \left\{ \frac{2\kappa(A_S)}{\cos(\theta)} + \tan(\theta)\,\kappa(A_S)^2 \right\} = 0. \tag{6.29}$$

Using the approximation in Eq. (6.16) to formulate our problem, it is clear that the model error in Eq. (6.25) only depends on the unknown aberration. This would be different when differences of two PSFs modeled by the approximation in Eq. (6.20) are taken, when the model error also depends on the chosen diversity. Then, a compromise should be made between a small diversity which leads to a small model error and a large diversity which ensures that the difference between two images does not become zero and the information content is lost.

6.2.4. Simulations

In this section, we present numerical simulations for the iterative aberration correction problem using **ILPD** and **LIFT**. We first describe the simulation setup. Secondly, we give one example of iterative linear phase diversity. Next, we analyze the behavior of both methods using a Monte Carlo simulation by varying the noise level and the rms value of the initial aberration. The computer employed for these simulations is a 2.67 GHz quad-core Intel(R) Core (TM)2 Quad CPU Q8400 with 4.0 GB of RAM.

We perform the comparison over 5 iterations for each method. One important difference between the two methods is that one iteration has a different meaning. **LIFT** collects one image at iteration 1 and based on this image performs 5 steps that lead to a phase estimate, which is then used for the correction. On the other hand, **ILPD** collects a new image at each of the 5 steps (image which includes the previous corrections) and performs a correction of the wavefront by a DM after each step. When estimating the wavefront using **ILPD**, we take 1 image (per correction step) of the same point-like object at defocus 2 radians. For a fixed diversity, like the one used here, the linear coefficients in Eq. (6.17) can be computed in advance. **LIFT** uses only one 1 image with an astigmatism diversity of $\pi/8$ radians as in [8] and computes at consecutive steps the gradient of the PSF in a different point. The pre-computation of the gradient makes our algorithm faster while preserving the accuracy, which we will quantify in this section. Of course, also the gradient for the first step of **LIFT** can be pre-computed.

We compute the corresponding nonlinear PSFs in Eq. (6.4) and the coefficients of the linear Taylor expansion given in Eq. (6.17). We consider a pupil of radius r sampled on a 32×32 grid embedded in a $4r \times 4r$ image to satisfy the Nyquist sampling criterion. We ensure that the wavefront does not contain jumps larger than $\pi/2$ which would be problematic for the sampling process. The signal to noise ratio corresponding to the read-out noise is calculated over the image ($m = 32$). Also, all treatment is monochromatic. We assume that the deformable mirror is able to produce known diversity shapes with an error that is negligible compared to other sources. This assumption motivates our choice to model only the first $n = 14$ modes that can be corrected by the DM. To obtain a PSF of unit surface, the pupil function Π is chosen as

$$
\Pi(u_j, v_j) = \begin{cases} 1/\sqrt{S} & u_j^2 + v_j^2 \le r^2 \\ 0 & u_j^2 + v_j^2 > r^2 \end{cases}, \tag{6.30}
$$

where S is the physical surface of the pupil.

If the frame-rates of the imaging camera and of the DM are sufficiently fast, it is an acceptable approximation that a few sequential wavefronts are assumed to be identical. The static aberration consisting of n modes is generated using the Kolmogorov model [20], which assumes aberrations with zero mean and covariance matrix C_ϕ. The parameters used to generate the Kolmogorov model are: diameter $D = 1$ [m], outer scale $L_0 = 42$ [m], Fried parameter $r_0 = 0.3$ [m].

First, we give an example of iterative linear phase diversity in Subsection 6.2.4. Secondly, in Subsection 6.2.4, for the same aberration, in the noiseless case, we

show the convergence and the corresponding rate of convergence in terms of residual wavefront error and relative residual wavefront error. Subsequently, in Subsection 6.2.4, we study the convergence properties in terms of the residual error for **ILPD** and **LIFT** as a function of increasing read-out noise SNR, photon count and wavefront rms.

One example of iterative phase diversity

We first show the convergence of the algorithms for a particular choice of the wavefront (with rms of 1 rad and intensity of 1000 photons per image) in the noiseless case in Fig. 6.2 and Tab. 6.1, and for a particular choice of the read-out noise realization (with SNR 3.16) in Fig. 6.3 and Tab. 6.2, respectively. Starting from the initial aberration, we perform 5 correction steps. The figures and tables mentioned before show both visually and numerically that the two methods converge to similar small residual errors. One important difference, as we show in the next subsection, is the convergence time, that for **ILPD** is much shorter. Tab. 6.1 and Tab. 6.2 list the residual rms values obtained after each iteration for the two methods. Inspecting the tables, it seems that **ILPD** is more robust to noise than **LIFT**, but the error difference between them is not significant. One advantage of **LIFT** is that it only uses one image, while **ILPD** uses 1 image per iteration, but the later method is faster. In order to quantify how much faster, we need to make a Monte Carlo analysis. This is the subject of the next subsection.
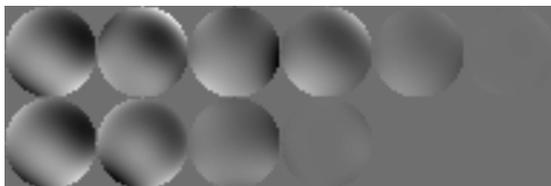
Figure 6.2. Convergence in terms of wavefront error: 1 rad rms, no read-out/photon noise, 1000 photons per image - **LIFT** (top), **ILPD** (bottom)

iterations / method	0	1	2	3	4	5
LIFT	1	0.85	0.82	0.65	0.31	0.039
ILPD	1	0.45	0.067	0.0039	1e-5	3.4e-9

Table 6.1. Rms values of the corrected wavefronts for no read-out/photon noise, 1000 photons per image, and 1 rad initial rms.
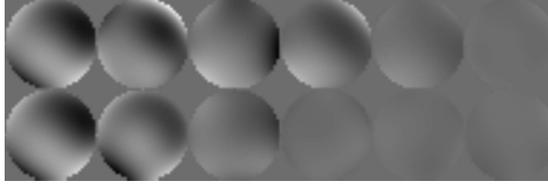
Figure 6.3. Convergence in terms of wavefront error: 1 rad rms, read-out noise with SNR = 3.16, no photon noise, 1000 photons per image - **LIFT** (top), **ILPD** (bottom)

iterations / method	0	1	2	3	4	5
LIFT	1	0.85	0.8	0.66	0.31	0.061
ILPD	1	0.45	0.09	0.05	0.061	0.057

Table 6.2. Rms values of the corrected wavefronts for read-out noise SNR = 3.16, no photon noise, 1000 photons per image, and 1 rad initial rms.

Iterative linear phase diversity without noise

We now repeat the experiment in the previous section 128 times for random aberrations with 1 rad rms, intensity of 1000 photons per image, and no read-out/photon noise. We use boxplots to visualize the results. On each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points not considered outliers, and outliers are plotted individually. The diamond signs represent mean values.

In Fig. 6.4 we plot the residual error in the aberration vector $\|\alpha_k - \hat{\alpha}_k\|$ at each iteration versus the number of iterations. The residual error decreases with each step for almost all the samples of **LIFT** and for all the samples considered for **ILPD** at this particular rms value. In the noiseless case, **ILPD** converges to a residual error ~ 0. **LIFT** converges to a small bias different from zero, but not as fast. There are also cases when **LIFT** diverges.

In Fig. 6.5 we plot, for the same data as in the previous figure, the relative residue $\frac{\|\alpha_{k+1}\|}{\|\alpha_k\|}$. We have mentioned before, in Subsection 6.2.3, that the error bound on this relative error depends on the model error for well conditioned matrices. The condition number of A_S in this example is 4.1810, such that Eq. (6.28) is valid. Therefore, the remark made using Eq. (6.29) in the previous section is sustained by Fig. 6.5: **ILPD** converges to a relative residual error equal to zero and it is independent with respect to different realizations of $\Delta b_S(\alpha_k)$. In both plots it can be seen that the error variance of **ILPD** is smaller.

The computational time necessary for **LIFT** to complete 5 iterations is 10.9978 seconds on average, while **ILPD** performs them in 0.0028 seconds on average.
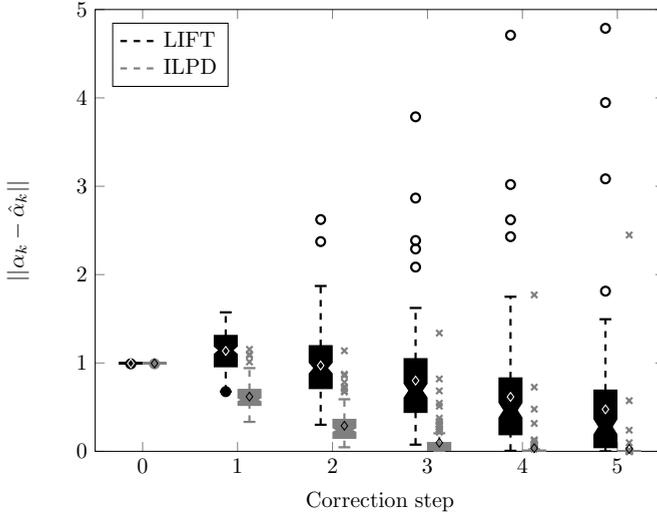
Figure 6.4. Residual error in the aberration vector. On each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points not considered outliers, and outliers are plotted individually. The diamond signs represent mean values.
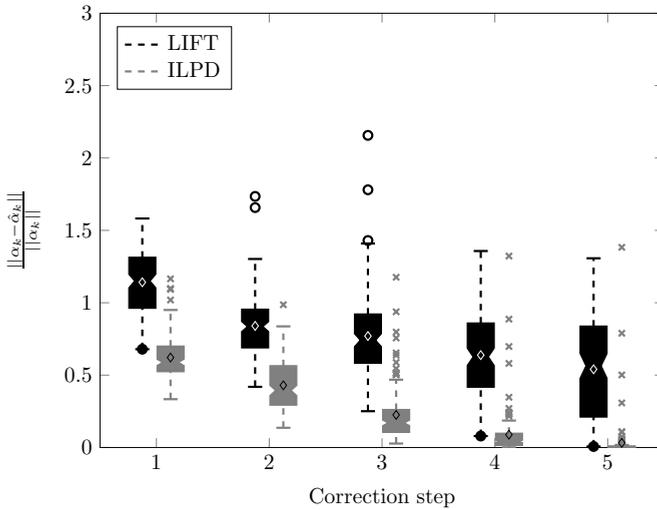
Figure 6.5. Relative residue
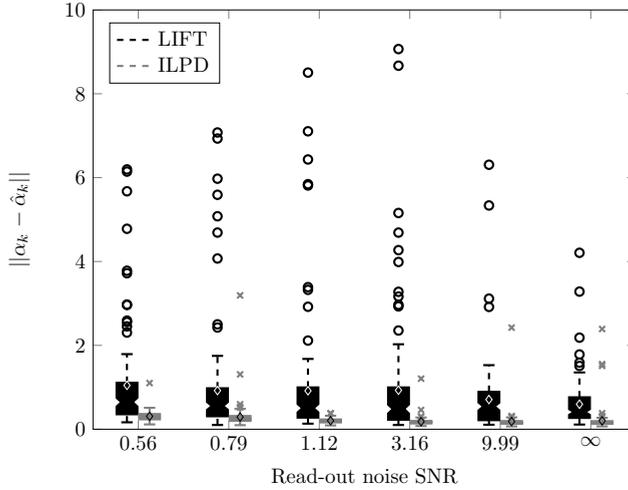
Figure 6.6. Wavefront residual error versus increasing SNR

Note that the integration time of the CCD is not included in computing these times. This makes **ILPD** 3927.8 times faster. When we also count the CCD integration time, **ILPD** is 40 times faster. For a fair comparison with **LIFT**, we have used here all the pixels in order to compute the estimate at each step, but the computational time for **ILPD** further decreases when using just a fraction of the pixels.

Error residue in the presence of noise
In the previous section, we have only analyzed a wavefront with rms equal to 1 rad and noise was not taken into account. However, more information about the properties of the proposed method can be obtained from a Monte Carlo simulation if we vary the read-out noise SNR, the photon noise, and the wavefront rms. We make a detailed analysis of the convergence properties of the presented methods using the residual error which gives us a measure for how much of the aberration we can correct for. The analysis is presented in Figs. 6.6, 6.7, and 6.8. Fig. 6.6 plots the residual error after 5 correction steps versus increasing read-out noise SNR, considering 1000 photons per image and Poisson photon noise. For each SNR, we repeat the experiment 128 times. The considered initial aberration has 1 rad rms. For **ILPD**, the residual error decreases with the increase of the SNR, which is what we expected. For **LIFT** this behavior is not very visible. One reason is the high value of the rms. In our simulations, we have noticed that for smaller rms values, e.g. 0.5 rms, **LIFT** starts to show this decrease in bias for increasing
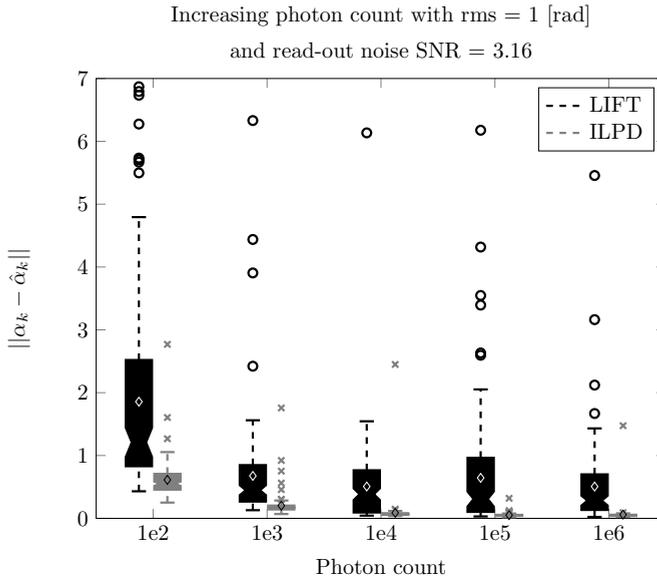
Figure 6.7. Wavefront residual error versus increasing photon count

rms, which shows that **LIFT** is more appropriate for small rms values. **ILPD** has a lower error variance and a lower error mean. Clearly, for **ILPD**, the error variance would converge to zero for an SNR equal to ∞ if no Poisson photon noise were considered.

Fig. 6.7 plots the residual error after 5 correction steps versus increasing photon count. For each photon count, we repeat the experiment 128 times. The considered initial aberration has 1 rad rms. Besides the Poisson photon noise, we also added read-out noise with SNR = 3.16. The residual error decreases with the increase of the photon count, which is what we expected. **ILPD** has a lower error variance and a lower error mean. Furthermore, it is visible in Fig. 6.7 that at low photon counts **LIFT** diverges.

The same type of analysis is made in Fig. 6.8 for increasing rms of the initial wavefront, a constant read-out noise level of 3.16 and Poisson photon noise. Both methods are based on a small-aberration assumption, so the bias of the estimation increases with increasing rms or it takes more iterations to converge. It is visible that **LIFT** starts to diverge for rms values larger than 0.5 rad, while **ILPD** still corrects for the aberration. This is due to the fact that with each iteration the aberration becomes smaller and the linear model in Eq. (6.21) is more and more accurate.
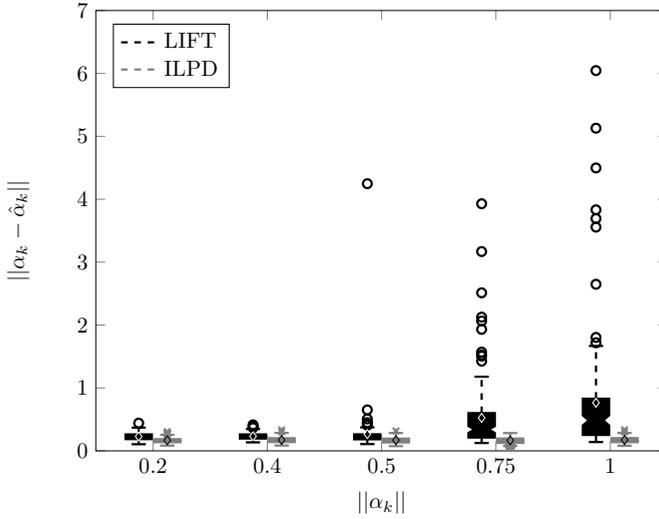
Figure 6.8. Wavefront residual error versus increasing rms

6.3. *Conclusions*

We have presented a novel method for wavefront estimation and correction suitable for several applications in astronomy or microscopy. Under the assumption of small phase aberrations, which is the typical situation in a control loop, the point spread function of an incoherent imaging system has been approximated with a linear model, which can be precomputed if the diversity used is a fixed one. This model allows for arbitrary phase diversities to be introduced in the system. Our iterative approach uses at each step one image of a point-like object, which includes a known phase diversity, and estimates the aberration using a least squares approach. In this way we increase the computational speed of phase retrieval methods that linearize the point spread function at each iteration around the current estimate of the aberration. Also, as the residual aberration decreases, the precomputed model of the point spread function becomes a better fit to the real one. This creates the premises for the method to be applied in a real-time correction system.

References

[1] R. G. Paxman, T. J. Schulz, and J. R. Fienup, *Joint estimation of object and aberrations by using phase diversity,* Journal of the Optical Society of America A **9**, 1072 (1992).

[2] L. M. Mugnier, A. Blanc, and J. Idier, *Phase diversity: A technique for wave-front sensing and for diffraction-limited imaging,* Advances in Imaging and Electron Physics **141**, 3 (2006).

[3] D. R. Gerwe, M. M. Johnson, and B. Calef, *Local minima analysis of phase diverse phase retrieval using maximum likelihood,* in *AMOS Technical Conference, Kihei HI* (2008).

[4] R. Gerchberg and W. Saxton, *Phase retrieval by iterated projections,* Optik **35**, 237 (1972).

[5] R. A. Gonsalves, *Small-phase solution to the phase-retrieval problem,* Optics Letters **26**, 684 (2001).

[6] C. U. Keller, V. Korkiakoski, N. Doelman, R. Fraanje, R. Andrei, and M. Verhaegen, *Extremely fast focal-plane wavefront sensing for extreme adaptive optics,* in *SPIE Astronomical Telescopes and Instrumentation* (International Society for Optics and Photonics, 2012) pp. 844721–844721.

[7] F. Martinache, *Kernel phase in Fizeau interferometry,* The Astrophysical Journal **724**, 464 (2010).

[8] S. Meimon, T. Fusco, and F. Cassaing, *A focal plane sensor for low-order sensing on laser tomographic systems: LIFT,* in *SPIE Astronomical Telescopes and Instrumentation* (International Society for Optics and Photonics, 2010) pp. 773611–773611.

[9] C. Van der Avoort, J. Braat, P. Dirksen, and A. Janssen, *Aberration retrieval from the intensity point-spread function in the focal region using the extended Nijboer–Zernike ap-* proach, Journal of Modern Optics **52**, 1695 (2005).

[10] I. Mocœur, L. M. Mugnier, and F. Cassaing, *Analytical solution to the phase-diversity problem for real-time wavefront sensing,* Optics Letters **34**, 3487 (2009).

[11] W. J. Wild, *Linear phase retrieval for wavefront sensing,* Optics Letters **23**, 573 (1998).

[12] J. Antonello, M. Verhaegen, R. Fraanje, T. van Werkhoven, H. C. Gerritsen, and C. U. Keller, *Semidefinite programming for model-based sensorless adaptive optics,* Journal of the Optical Society of America A **29**, 2428 (2012).

[13] D. J. Lee, M. C. Roggemann, and B. M. Welsh, *Cramer–Rao analysis of phase-diverse wavefront sensing,* Journal of the Optical Society of America A **16**, 1005 (1999).

[14] A. Tokovinin and S. Heathcote, *Donut: Measuring optical aberrations from a single extrafocal image,* Publications of the Astronomical Society of the Pacific **118**, 1165 (2006).

[15] A. Polo, S. F. Pereira, and P. H. Urbach, *Theoretical analysis for best defocus measurement plane for robust phase retrieval,* Optics Letters **38**, 812 (2013).

[16] ANSI, *Methods for reporting optical aberrations of eyes.* American national standard for ophtalmics , ANSI (2004).

[17] J. W. Goodman, *Introduction to Fourier optics*, 3rd ed. (Roberts and Company, 2005).

[18] F. Gustafsson, *Statistical sensor fusion* (Studentlitteratur, 2010).

[19] G. H. Golub and C. F. Van Loan, *Matrix computations*, Vol. 3 (JHU Press, 2012).

[20] R. J. Noll, *Zernike polynomials and atmospheric turbulence,* Journal of the Optical Society of America **66**, 207 (1976).

Appendix

6.A. *Proof of Propositions*

In Sections 6.2.2 and 6.2.3 multiple claims have been made and in this appendix these claims are proved. The claims are invariant of the Fourier transform between the PSF and OTF. Therefore, to shorten the proves, the approximations and their properties will be given in terms of the OTF. For clarity we introduce the following short hand notations

$$p_{i,j} := p_j(\alpha, \beta_i), \ W_{i,j} := W_j(\alpha, \beta_i). \tag{6.31}$$

Proof of Property 6.2.1. We introduce the short hand notation $\gamma_i = \alpha + \beta_i$ and for the small total phase approximation we have $\gamma_0 = 0$. Using Eqs. (6.12) and (6.15) the OTF is equal to

$$
\begin{aligned}
W_{i,j} &\approx \left(p_{i,j} + \frac{\partial p_{i,j}}{\partial \gamma_i^T} (\gamma_i - \gamma_0) \right) \star \\
&\quad \left. \left(p_{i,-j}^* + \frac{\partial p_{i,-j}^*}{\partial \gamma_i^T} (\gamma_i - \gamma_0) \right) \right|_{\gamma_i = \gamma_0} \\
&= \left. p_{i,j} \star p_{i,-j}^* \right|_{\gamma_i = \gamma_0} \\
&\quad + \left. \left[\frac{\partial p_{i,j}}{\partial \gamma_i^T} \star p_{i,-j}^* + p_{-j}(\alpha, \beta_i) \star \frac{\partial p_{i,-j}^*}{\partial \gamma_i^T} \right] \right|_{\gamma_i = \gamma_0} (\gamma_i - \gamma_0) \\
&\quad + (\gamma_i - \gamma_0)^T \left[\frac{\partial p_{i,j}}{\partial \gamma_i^T} \star \frac{\partial p_{i,-j}^*}{\partial \gamma_i} + \right. \\
&\quad \left. \left. \frac{\partial p_{i,j}}{\partial \gamma_i} \star \frac{\partial p_{i,-j}^*}{\partial \gamma_i^T} \right] \right|_{\gamma_i = \gamma_0} (\gamma_i - \gamma_0).
\end{aligned}
\tag{6.32}
$$

The first order term is

$$\mathcal{L} := \left. \left[\frac{\partial p_{i,j}}{\partial \gamma_i^T} \star p_{i,-j}^* + p_{-j}(\alpha, \beta_i) \star \frac{\partial p_{i,-j}^*}{\partial \gamma_i^T} \right] \right|_{\gamma_i = \gamma_0}. \tag{6.33}$$

To show that Eq. (6.33) is invariant in the even modes we reorder γ_i and Z^T by even and odd parts and Eq. (6.33) becomes

$$\begin{aligned}
\mathscr{L} &= iZ_j^T \Pi_j \star \Pi_{-j}^* - \Pi_j \star iZ_{-j}^T \Pi_{-j}^* \\
&= i \begin{bmatrix} Z_{e,j}^T & Z_{o,j}^T \end{bmatrix} \Pi_j \star \Pi_{-j}^* - \Pi_j \star i \begin{bmatrix} Z_{e,-j}^T & Z_{o,-j}^T \end{bmatrix} \Pi_{-j}^*,
\end{aligned} \tag{6.34}$$

where the subindexes j and $-j$ are short hand notations for the coordinates (u_j, v_j) and $(-u_j, -v_j)$. Next, because Π is even and real we have that

$$\begin{aligned}
\mathscr{L} &= i \begin{bmatrix} Z_{e,j}^T & Z_{o,j}^T \end{bmatrix} \Pi_j \star \Pi_j - \Pi_j \star i \begin{bmatrix} Z_{e,j}^T & -Z_{o,j}^T \end{bmatrix} \Pi_j \\
&= i \begin{bmatrix} 0 & 2Z_{o,j}^T \end{bmatrix} \Pi_j \star \Pi_j.
\end{aligned} \tag{6.35}$$

$\square$

Proof of Property 6.2.2. The OTF is given by

$$\begin{aligned}
W_{i,j} &\approx p_{i,j} \star p_{i,-j}^* \Big|_{\gamma_i=\gamma_0} \\
&+ \left[\frac{\partial p_{i,j}}{\partial \gamma_i^T} \star p_{i,-j}^* + p_{-j}(\alpha, \beta_i) \star \frac{\partial p_{i,-j}^*}{\partial \gamma_i^T} \right]\Bigg|_{\gamma_i=\gamma_0} (\gamma_i - \gamma_0)
\end{aligned} \tag{6.36}$$

and the linear term is equal to Eq. (6.33). $\square$

Proof of Property 6.2.3. We show that Eq. (6.33) for $\gamma_0 \neq 0$ is not invariant for even modes. We again reorder γ_i and Z^T by even modes Z_e^T and odd modes Z_o^T to obtain

$$\begin{aligned}
\mathscr{L} &= iZ_j^T \Pi_j \exp\left(iZ_j^T \gamma_0\right) \star \Pi_{-j}^* \exp\left(-iZ_{-j}^T \gamma_0\right) \\
&- \Pi_j \exp\left(iZ_j^T \gamma_0\right) \star iZ_{-j}^T \Pi_{-j} \exp\left(-iZ_{-j}^T \gamma_0\right) \\
&= i \begin{bmatrix} Z_{e,j}^T & Z_{o,j}^T \end{bmatrix} \Pi_j \exp\left(iZ_j^T \gamma_0\right) \star \Pi_{-j}^* \exp\left(-iZ_{-j}^T \gamma_0\right) \\
&- \Pi_j \exp\left(iZ_j^T \gamma_0\right) \star i \begin{bmatrix} Z_{e,-j}^T & Z_{o,-j}^T \end{bmatrix} \Pi_{-j} \exp\left(-iZ_{-j}^T \gamma_0\right) \\
&= i \begin{bmatrix} Z_{e,j}^T & Z_{o,j}^T \end{bmatrix} \tilde{\Pi}_j \star \tilde{\Pi}_{-j}^* - \tilde{\Pi}_j \star i \begin{bmatrix} Z_{e,-j}^T & Z_{o,-j}^T \end{bmatrix} \tilde{\Pi}_{-j}^*,
\end{aligned} \tag{6.37}$$

where $\tilde{\Pi}_j := \Pi_{-j} \exp\left(-iZ_{-j}^T \gamma_0\right)$. Next, because $\tilde{\Pi}_j$ is neither even nor real we have that the two terms are different and the even modes do not cancel. $\square$

Proof of Property 6.2.4. The 2nd order Taylor approximation of the GPF is

$$p_{i,j} \approx p_{i,j}\big|_{\gamma_i=\gamma_0} + \frac{\partial p_{i,j}}{\partial \gamma_i}\bigg|_{\gamma_i=\gamma_0} (\gamma_i - \gamma_0)$$

$$+ \frac{1}{2} (\gamma_i - \gamma_0)^T \frac{\partial^2 p_{i,j}}{\partial \gamma_i \gamma_i^T}\bigg|_{\gamma_i=\gamma_0} (\gamma_i - \gamma_0). \tag{6.38}$$

Dropping terms of order 3 and higher the resulting approximated OTF reduces to

$$W_{i,j} \approx p_{i,j} \star p_{i,-j}^*\big|_{\gamma_i=\gamma_0}$$

$$+ \left[\frac{\partial p_{i,j}}{\partial \gamma_i^T} \star p_{i,-j}^* + p_{-j}(\alpha, \beta_i) \star \frac{\partial p_{i,-j}^*}{\partial \gamma_i^T} \right]\bigg|_{\gamma_i=\gamma_0} (\gamma_i - \gamma_0)$$

$$+ (\gamma_i - \gamma_0)^T \left[\frac{\partial^2 p_{i,j}}{\partial \gamma_i \partial \gamma_i^T} \star p_{i,-j}^* + p_{-j}(\alpha, \beta_i) \star \frac{\partial^2 p_{i,-j}^*}{\partial \gamma_i \partial \gamma_i^T} \right.$$

$$\left. + \frac{\partial p_{i,j}}{\partial \gamma_i^T} \star \frac{\partial p_{i,-j}^*}{\partial \gamma_i} + \frac{\partial p_{i,j}}{\partial \gamma_i} \star \frac{\partial p_{i,-j}^*}{\partial \gamma_i^T} \right]\bigg|_{\gamma_i=\gamma_0} (\gamma_i - \gamma_0), \tag{6.39}$$

which is exactly the second order Taylor approximation. $\square$

Proof of Property 6.2.5. The difference between the approximated PSF following from a first order approximation of the GPF in Eq. (6.32) and the second order Taylor approximation in Eq. (6.39) is given by

$$\frac{\partial^2 p_{i,j}}{\partial \gamma_i \partial \gamma_i^T} \star p_{i,-j}^* + p_{-j}(\alpha, \beta_i) \star \frac{\partial^2 p_{i,-j}^*}{\partial \gamma_i \partial \gamma_i^T}. \tag{6.40}$$

The addition of the missing term from the first order GPF approximation results in a residue of order $\mathcal{O}\left(\|\alpha\|^3\right)$ instead of $\mathcal{O}\left(\|\alpha\|^2\right)$. Therefore, the second order Taylor expansion of the PSF is more accurate than the first order approximation of the GPF, while the quadratic form remains. $\square$

Proof of Property 6.2.6. Inspecting Eq. (6.40) we observe that the linear term is not affected therefore Property 6.2.1 and Property 6.2.3 still hold for the linear terms of Eq. (6.39). $\square$

Proof of Property 6.2.7. The intensity of both signals is positive and subtracting 2 images decreases the mean signal at each pixel. Recall that we assume that all

160

the camera pixels are mutually independent and that the measurement noise is Gaussian distributed.

$$\mu_{\text{diff}} = \mathbb{E}\left[\Delta y_j\right] \tag{6.41}$$

$$= \mu_1 h_{1,j} - \mu_2 h_{2,j}. \tag{6.42}$$

μ_{diff} is smaller than either $\mu_1 h_{1,j}$ and $\mu_2 h_{2,j}$. Next, the variance of the signal increases

$$\sigma_{\text{diff}}^2 = \mathbb{E}\left[\left(\Delta y_j - \mu_{\text{diff}}\right)^2\right] \tag{6.43}$$

$$= \sigma_{1,j}^2 + \sigma_{2,j}^2. \tag{6.44}$$

Therefore, the resulting SNR at pixel j is given by

$$\text{SNR}_{\text{diff}} = \frac{\mu_1 h_{1,j} - \mu_2 h_{2,j}}{\sqrt{\sigma_{1,j}^2 + \sigma_{2,j}^2}}. \tag{6.45}$$

If we assume that the noise is the same for each pixel with $\sigma_{i,j} := \sigma_i$, the total signal to noise ratio is equal to

$$
\begin{aligned}
\text{SNR}_{\text{diff}} &= \frac{1}{m^2} \sum_{j=1}^{m^2} \frac{\mu_1 h_{1,j} - \mu_2 h_{2,j}}{\sqrt{\frac{1}{m^2} \sum_{j=1}^{m^2} \sigma_1^2 + \frac{1}{m^2} \sum_{j}^{m^2} \sigma_{2,j}^2}} \\
&= \frac{1}{m^2} \frac{1}{\sqrt{\sigma_1^2 + \sigma_2^2}} \sum_{j=1}^{m^2} \mu_1 \left(h_{1,j} - \mu_2 h_{2,j}\right) \\
&= \frac{1}{m^2} \frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2 + \sigma_2^2}}.
\end{aligned}
\tag{6.46}
$$

$\square$

7

Simultaneous Measurement of Emission Color and 3D Position of Single Molecules

Measure what is measurable, and make measurable what is not so.

GALILEO GALILEI

Carlas Smith
Max Huisman
Marijn Siemons
David Grünwald
Sjoerd Stallinga

Abstract

We show that the position of single molecules in all three spatial dimensions can be estimated alongside its emission color by diffractive optics based design of the Point Spread Function (PSF). The phase in a plane conjugate to the aperture stop of the objective lens is modified by a diffractive structure that splits the spot on the camera into closely spaced diffraction orders. The distance between and the size of these sub-spots are a measure of the emission color. Estimation of the axial position is enabled by imprinting aberrations such as astigmatism and defocus onto the orders. The overall spot shape is fitted with a fully vectorial PSF model. Proof-of-principle experiments on quantum dots indicate that a spectral precision of 10 to 20 nm, an axial localization precision of 25 to 50 nm, and a lateral localization precision of 10 to 30 nm can be achieved over a 1 μm range of axial positions for on average 800 signal photons and 17 background photons/pixel. The method appears to be rather sensitive to PSF model errors such as aberrations, giving in particular rise to biases in the fitted wavelength of up to 15 nm.

7.1. *Introduction*

The resolution in single molecule localization microscopy based on widefield fluorescence imaging [1–4] is limited by the localization precision and the density of fluorescent labels [5] and can be on the order of several tens of nanometers in practice. This leap in resolution has opened up new venues in the study of macromolecular complexes at the interface of chemistry, biology and physics. The study of molecular interaction and function necessitates the detection of multiple molecular species, which is conventionally done by labeling the target molecules with fluorophores that differ in their emission color. These species can be differentiated in a number of ways. The most straightforward ways are to image the species subsequently by changing the illumination laser and the dichroics [6] or to image the species simultaneously by color splitting and projecting images side by side on one or on multiple cameras [7–9]. Hyperspectral imaging with a line scan camera offers a more direct measurement of the emission color [10]. An entirely different way of color detection relies on activator-reporter labeling in which the illumination wavelength is switched but the detection wavelength is kept the same [11]. This approach requires a calibration procedure for handling the cross-talk between the different activator molecules. A similar approach is the label washing method in which the target molecules are labeled sequentially with the same fluorophore and the labels are washed away in intermediate sample preparation steps [12, 13]. The key advantage of this approach is that it benefits from the high photon count per on-event and the relatively high photostability of the Alexa647 fluorophore.

Recently, a direct estimation of the emission color has been proposed making use of a diffraction grating with large pitch placed in the detection branch of the microscope [14]. The diffraction orders generated by the grating appear as closely spaced spots on the camera, the distance between the orders being a measure for the wavelength. In this way the simultaneous measurement of the in-plane emitter position and the emission color with a single camera setup is enabled, reaching an estimation precision on the order of 10 nm for both the position and the (peak) emission wavelength. A similar technique with similar performance has recently been proposed in [15], the difference being the use of a prism as dispersing optical element and the use of a 4π setup to increase photon collection efficiency. The technique described in [14] belongs to the family of Point Spread Function (PSF) engineering or optical Fourier processing techniques [16]. A relay optical path is added to the detection branch of the microscope and a phase profile is added to the emission beam in a plane conjugate to the pupil plane, often making use of a Spatial Light Modulator (SLM). So far, PSF

engineering has mainly been applied to 3D single emitter localization. A wide range of 3D methods have been described in the literature. The main flavors are the astigmatic method [17–19], the bi-/multifocal method [20–24], the double-helix method [25–29], the phase-ramp method [30], the self-bending method [31] and the interferometric method [32]. Pupil optimization of 3D localization has been investigated in [33, 34], leading to so-called saddle-point and tetrapod PSFs. These are variants of the astigmatic method that use higher order astigmatism for splitting the astigmatic focal lines into halter shaped spot pairs, which improves precision and axial range. Based on these developments we set out to explore if 3D spatial and color information can be encoded simultaneously on a single camera.

The scale of a diffraction limited point source image is on the order of λ/NA with λ the wavelength and NA the numerical aperture of the objective lens. Estimation of the spot width would seemingly be enough then to estimate the wavelength. In standard 2D imaging, however, variations in spot width could also be due to small focus errors, so that the effects of wavelength and axial position cannot be disentangled. This results in a divergent Crámer Rao Lower Bound (CRLB), the best possible precision for an unbiased estimator, for these two parameters in diffraction limited imaging at the focal plane. This divergence is overcome by PSF engineering, in which spot shape parameters are used to encode the axial position. For example, the axial position is encoded by the elongation of the spot for the astigmatic method, and by the orientation of the spot for the double helix method. Now that the *shape* of the PSF is used for estimation of the axial position, it follows that in principle the wavelength can be fitted from the *scale* of the PSF.

Here, we propose a further advance by making use of diffractive optics based design principles. Fig. 7.1 shows the gist of the idea. The phase of the emission beam is modified at a plane conjugate to the exit pupil of the objective lens. The imprinted phase profile is divided into repetitive zones that results in a split of the spot on the camera into different diffraction orders (sub-spots) that are closely positioned on the camera, similar to [14]. The shape of the diffractive zones is no longer straight as in a conventional grating but instead curved so as to generate a desired aberration (astigmatism, defocus) in the different diffraction orders. This enables the detection of the emission color from the distance and relative magnitude of the sub-spots and of the axial position from the shape of and the shape differences between the sub-spots. In this way both the shape and scale are used in a mixed way to extract the information on wavelength and axial position with a better precision.

7.2. *Results*
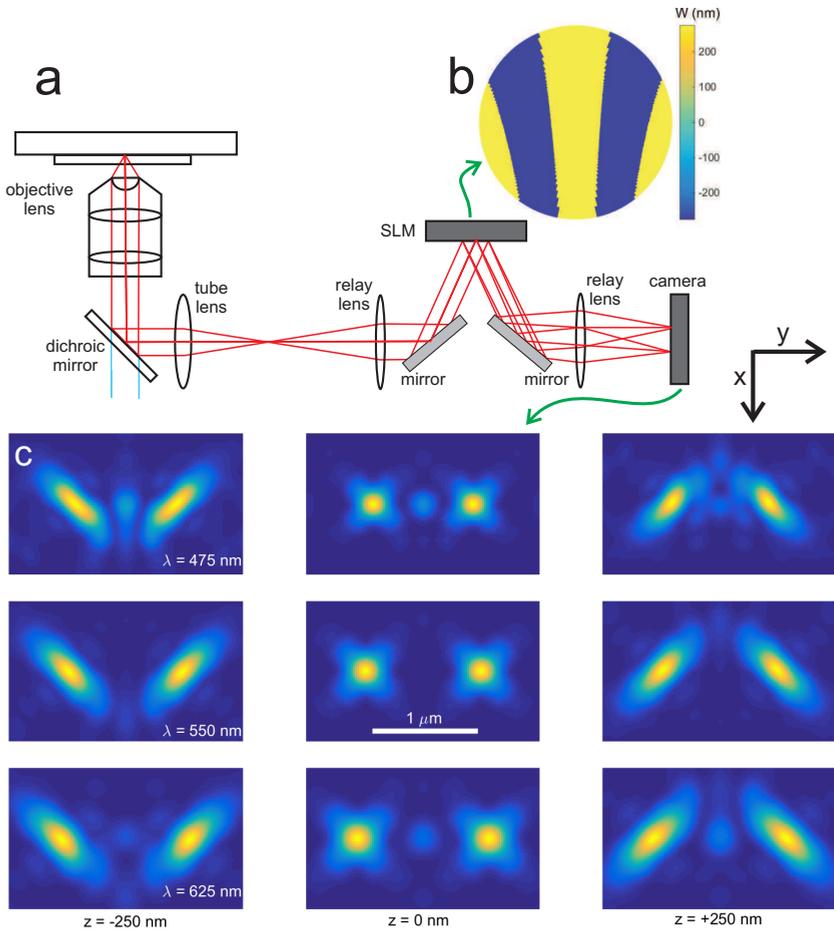
Figure 7.1. Principle of 3D+λ measurement for single emitters. (a) Schematic of the microscope setup. A 4f-relay path is added to the detection branch of a microscope, the SLM is placed conjugate to the back focal plane of the objective lens. (b) The shape of the zones is curved for inducing aberrations to the sub-spots (diffraction orders) captured at the camera. The profile shown is a binary phase profile with phase step close to π in the visible spectrum for splitting the spot into -1st and +1st diffraction order sub-spots, and with zone shapes inducing astigmatism. (c) Modeled spots at the camera plane for different axial positions and emission wavelengths of the emitter. The computation of the spots is over a 2.5μm×1.5μm region discretized with a 125 × 75 grid for a medium refractive index $n = 1.33$ and a numerical aperture NA = 1.25, with a pupil discretized with a 164 × 164 grid.

7.2.1. Theory

Diffractive Optical Elements (DOEs)

The aberration profile of the SLM is that of a DOE. We describe the DOE as a thin surface, locally altering the phase of the passing beam, independent of the local direction of incidence. This fits within the framework of scalar diffraction theory and is justified in the paraxial (low NA) limit, which is appropriate at the SLM plane. We also assume that the SLM is placed at a plane conjugate to the pupil plane. The thin surface of the DOE can be divided in zones, labeled with a discrete index j, the zone indices. Points with normalized pupil coordinates $\vec{\rho} = (\rho_x, \rho_y)$ are in zone j if:

$$j\lambda_0 \leq K(\vec{\rho}) < (j+1)\lambda_0, \tag{7.1}$$

where λ_0 is the design wavelength, chosen to be in the center of the spectral region of interest, and where the function $K(\vec{\rho})$ is called the zone function. For a standard grating structure the zone function is a linear function of $\vec{\rho}$. Within each zone j a variable t that takes values $0 \leq t < 1$ is defined by:

$$t = \frac{K(\vec{\rho})}{\lambda_0} - j = \frac{K(\vec{\rho})}{\lambda_0} - \mathrm{floor}\left(\frac{K(\vec{\rho})}{\lambda_0}\right), \tag{7.2}$$

where floor (x) is the largest integer smaller than x. The aberration profile added to an incoming beam is given by the so-called profile function $f(t)$:

$$W(\vec{\rho}) = f(t) = f\left(\frac{K(\vec{\rho})}{\lambda_0} - \mathrm{floor}\left(\frac{K(\vec{\rho})}{\lambda_0}\right)\right). \tag{7.3}$$

This profile function $f(t)$ describes the shape of the grating within each zone. For example, a blazed structure has $f(t) = h_{\max} t$ with $h_{\max}$ the step in optical path length, and a sinusoidal structure has $f(t) = h_{\max}(1 + \cos(2\pi t))/2$ with $h_{\max}$ the optical path length modulation. The transmission function $T(\vec{\rho}) = \exp(2\pi i W(\vec{\rho})/\lambda)$ can be written as a sum over diffraction orders with index m:

$$
\begin{aligned}
T(\vec{\rho}) &= \exp\left(\frac{2\pi}{\lambda} f\left(\frac{K(\vec{\rho})}{\lambda_0} - \mathrm{floor}\left(\frac{K(\vec{\rho})}{\lambda_0}\right)\right)\right) \\
&= \int_0^1 dt'\, \delta\left(\frac{K(\vec{\rho})}{\lambda_0} - \mathrm{floor}\left(\frac{K(\vec{\rho})}{\lambda_0}\right) - t'\right) \exp\left(\frac{2\pi i f(t')}{\lambda}\right) \\
&= \int_0^1 dt'\left[\sum_{m=-\infty}^{+\infty} \exp\left(2\pi i m\left(\frac{K(\vec{\rho})}{\lambda_0} - \mathrm{floor}\left(\frac{K(\vec{\rho})}{\lambda_0}\right) - t'\right)\right)\right] \exp\left(\frac{2\pi i f(t')}{\lambda}\right)
\end{aligned}
$$

$$= \sum_{m=-\infty}^{+\infty} C_m \exp\left(\frac{2\pi i W_m(\vec{\rho})}{\lambda}\right), \tag{7.4}$$

with amplitude:

$$C_m = \int_0^1 dt \exp(-2\pi i m t) \exp\left(\frac{2\pi i f(t)}{\lambda}\right), \tag{7.5}$$

and aberration function:

$$W_m(\vec{\rho}) = \frac{m K(\vec{\rho}) \lambda}{\lambda_0}. \tag{7.6}$$

The diffraction efficiency is $\eta_m = |C_m|^2$, and obeys conservation of energy, i.e. $\sum_{m=-\infty}^{\infty} \eta_m = 1$. In summary, the profile function $f(t)$ can be designed to give a desired distribution of light over the diffraction orders, and the zone function $K(\vec{\rho})$ can be designed to give the desired aberration to the different contributing orders m. On top of the DOE profile a nominal aberration profile $W_{\mathrm{nom}}(\vec{\rho})$ can be introduced giving each diffraction order a total aberration $W_{\mathrm{nom}}(\vec{\rho}) + W_m(\vec{\rho})$.

Design configurations
Diffractive optics based designs based on adding astigmatism or defocus to the diffraction orders have been described in [35]. In the following we will provide details on these designs and investigate the performance. A first configuration is based on a binary grating, which has $f(t) = h_{\max} \mathrm{sign}(\sin(2\pi t)) t$ with $h_{\max} = \lambda_0/2$. This will split the spot in two dominant diffraction orders, $m = \pm 1$, with efficiencies $\eta_{\pm 1} = 4\sin^2(\pi \lambda_0/(2\lambda))/\pi^2$, the zeroth order being essentially suppressed. The zone function is taken to be the sum of tip and diagonal astigmatism:

$$K(\vec{\rho}) = A_{11}\rho_x + 2A_{2-2}\rho_x\rho_y, \tag{7.7}$$

giving a distance between the $\pm$1st orders $\Delta x = 2A_{11}\lambda/(\lambda_0 \mathrm{NA})$ and astigmatism of equal magnitude but opposite sign to the two orders. The wavelength can be estimated from the spot separation, and the axial position from the (opposite) elongation of the two orders, as indicated by the simulated spot shapes shown in Fig. 7.1. The particular design shown there has a zone function with Zernike coefficients $A_{11} = 1.4\lambda_0$, $A_{2-2} = 0.27\lambda_0$ for $\lambda_0 = 550$ nm. It is also possible to generate variations on this design by mixing in higher order astigmatism, i.e. having non-zero Zernike coefficients $A_{4-2}, A_{6-2}, \ldots$. Note that we use the convention of labeling the Zernike coefficients A_{nl} with a radial index n and an azimuthal index l [36].

168
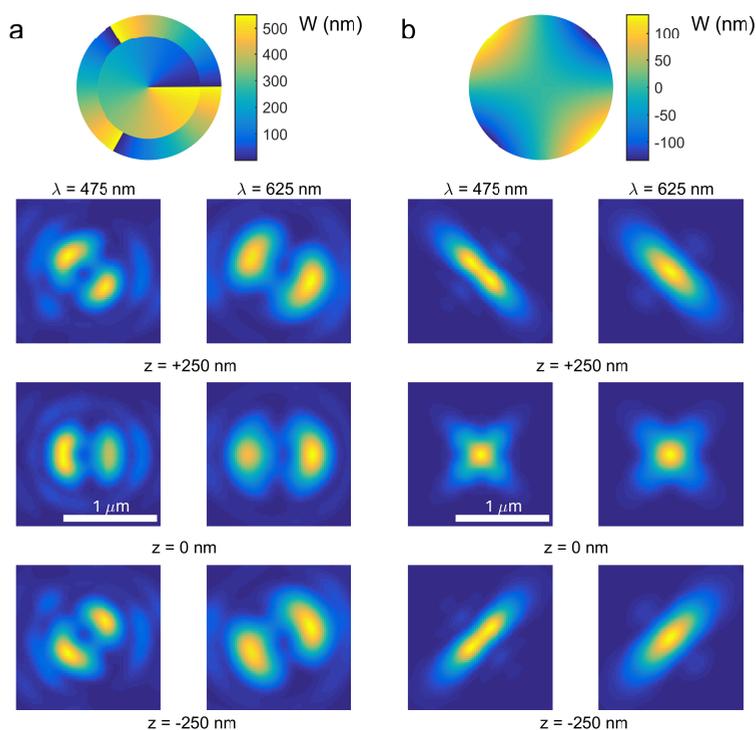
Figure 7.2. Phase profile and spot shapes of existing axial localization principles for fitting the emission wavelength λ from the *scale* of the spot. (a) Double helix PSF, (b) astigmatic PSF. The computation of the spots is over a $1.5\mu\text{m}\times1.5\mu\text{m}$ region discretized with a 75×75 grid for a medium refractive index $n = 1.33$ and a numerical aperture NA = 1.25, with a pupil discretized with a 164×164 grid.

Color and 3D Position Estimation of Single Molecules

A design variation on this first configuration is based on the blazed grating, which has $f(t) = h_{\max} t$ with $h_{\max} = \lambda_0/2$. This will split the spot in two diffraction orders, $m = 0$ with efficiency $\eta_0 = \mathrm{sinc}(\pi\lambda_0/(2\lambda))^2$, and $m = 1$ with efficiency $\eta_1 = \mathrm{sinc}(\pi - \pi\lambda_0/(2\lambda))^2$, with losses to higher orders amounting to less than $1 - 8/\pi^2 = 0.19$. The zone function is again taken to be the sum of tip and diagonal astigmatism giving only one of the two spots (the first order) that is astigmatically aberrated. Symmetry between the two spots can be restored by adding an overall aberration $W_{\mathrm{nom}}(\vec{\rho}) = -K(\vec{\rho})/2$ so that order $m = 0$ has an overall aberration $-K(\vec{\rho})/2$ and order $m = 1$ has an overall aberration $K(\vec{\rho})(2\lambda - \lambda_0)/(2\lambda_0)$, i.e. they are (nearly) equal in magnitude but of opposite sign.

A second configuration uses a zone function with tip and defocus:

$$K(\vec{\rho}) = A_{11}\rho_x + A_{20}\left(2\left(\rho_x^2 + \rho_y^2\right) - 1\right), \tag{7.8}$$

and can be implemented with the binary or blazed flavor in the same way as described above. In this configuration the spot is split in two orders that mutually defocused. The wavelength can again be estimated from the spot separation, and the axial position from the relative spot size of the two orders. Here too design variations can be generated by mixing in spherical aberration, i.e. having non-zero Zernike coefficients of higher order $A_{40}, A_{60}, \ldots$.

Instead of splitting the spot in two sub-spots it is also possible to split in three sub-spots. By varying the modulation of a binary or sinusoidal profile function the ratio between a central $m = 0$ spot and two $m = \pm 1$ side spots can be tuned. The central spot is unaberrated and the side spots have aberrations (astigmatism, defocus) of equal magnitude but opposite sign.

The double helix PSF designed according to [28, 29] can also be framed in the currently proposed DOE formalism. Now the zone function cannot be conveniently expressed as the sum of a limited number of Zernike functions as it contains singular points. The pupil is divided into L annular zones $l = 1, 2, \ldots, L$ with outer radius $\rho_l = (l/L)^\alpha$ with α an exponent, within annular zone l the zone function is:

$$K(\vec{\rho}) = (2l + 1)\phi/(2\pi), \tag{7.9}$$

with $\phi = \arctan(\rho_y/\rho_x)$ the azimuthal pupil coordinate. The profile function is a blaze with $h_{\max} = \lambda_0$ so that predominantly order $m = 1$ is selected. Fig. 7.2 shows a simulation of such a double helix PSF with $L = 2$ and $\alpha = 1/2$ for $\lambda_0 = 550$ nm. The orientation encodes for the axial position, and the distance between the two lobes for the wavelength. The axial range of the resulting double helix PSF

increases with L, but at the expense of an increase in the lateral footprint of the spot on the camera. For the purpose of comparison we show simulated astigmatic spots in Fig. 2(b) (nominal aberration function with Zernike coefficients $A_{2-2} = 0.34\lambda_0$ for $\lambda_0 = 550$ nm). It is possible to add higher order astigmatism (non-zero A_{4-2}) to mimic the saddle-point/tetrapod PSF of [33, 34], which has a lower and more constant precision for different axial positions. For the sake of simplicity this is not further pursued here.

Single emitter localization by Maximum Likelihood Estimation (MLE)

The procedure of MLE for single emitter localization has been documented in e.g. [37–39]. In brief, parameters of interest $\theta_1, \ldots, \theta_L$ are found by fitting the expected photon count μ_k to the observed photon counts n_k for all pixels $k = 1, \ldots, N$ inside a ROI by optimization of the log-likelihood function for the model given the observation. The log-likelihood that incorporates both Gaussian readout noise and shot noise is given by [40–42]:

$$\log L = \sum_k \left[\left(n_k + \sigma^2 \right) \log \left(\mu_k + \sigma^2 \right) - \left(\mu_k + \sigma^2 \right) - \log \Gamma \left(n_k + \sigma^2 + 1 \right) \right], \qquad (7.10)$$

where $\Gamma(x) = \int_0^\infty dt\, t^{x-1} \exp(-x)$ is the Gamma-function, and σ is the root mean square (rms) readout noise. The average and variance of the observed photon count n_k are given by:

$$\langle n_k \rangle \;=\; \mu_k, \qquad (7.11)$$

$$\langle n_k^2 \rangle - \langle n_k \rangle^2 \;=\; \mu_k + \sigma^2. \qquad (7.12)$$

The derivatives of the log-likelihood w.r.t. the fit parameters are needed in optimization routines such as the Levenberg-Marquardt algorithm:

$$\frac{\partial \log L}{\partial \theta_j} \;=\; \sum_k \frac{n_k - \mu_k}{\mu_k + \sigma^2} \frac{\partial \mu_k}{\partial \theta_j}, \qquad (7.13)$$

$$\frac{\partial^2 \log L}{\partial \theta_j \partial \theta_l} \;=\; -\sum_k \frac{n_k + \sigma^2}{\left(\mu_k + \sigma^2 \right)^2} \frac{\partial \mu_k}{\partial \theta_j} \frac{\partial \mu_k}{\partial \theta_l} + \sum_k \frac{n_k - \mu_k}{\mu_k + \sigma^2} \frac{\partial^2 \mu_k}{\partial \theta_j \partial \theta_l}. \qquad (7.14)$$

where the contribution from the second order derivatives of μ_k with respect to the fit parameters θ_j is usually neglected as it is small close to the optimum of $\log L$. The Fisher matrix is found as:

$$F_{jl} = \left\langle \frac{\partial \log L}{\partial \theta_j} \frac{\partial \log L}{\partial \theta_l} \right\rangle = \sum_k \frac{1}{\mu_k + \sigma^2} \frac{\partial \mu_k}{\partial \theta_j} \frac{\partial \mu_k}{\partial \theta_l}. \qquad (7.15)$$

The CRLB is found from the diagonal elements of the inverse of the Fisher matrix:

$$CRLB_j = \sqrt{F^{-1}{}_{jj}}, \tag{7.16}$$

for $j = 1,\ldots,L$.

Vectorial PSF model
The complexity of the spot shapes resulting from diffractive optics inspired PSF designs warrants a more exact PSF model than standard Gaussian based effective PSF models. We will use an exact vectorial PSF model to fit spots, as originally described in [43], but now for fitting the axial position and emission wavelength as well.

The expected photon count at pixel k depends on the emitter position $\vec{r}_0 = (x_0, y_0, z_0)$, the emission wavelength λ, the signal photon count N, and the number of background photons per pixel b, giving a total of 6 fit parameters. The expected photon count is given by the integration of the PSF over the pixel area $\mathcal{D}_k$ of size $a \times a$:

$$\mu_k = \int_{\mathcal{D}_k} dx\,dy\, H(\vec{r} - \vec{r}_0), \tag{7.17}$$

where the integration is carried out at the image plane $z = 0$. The emitted radiation is assumed to be imaged by an aplanatic and telecentric combination of objective lens and tube lens, and the emission dipole is assumed to rotate freely during the image acquisition time of the camera. The PSF of a freely rotating dipole in the presence of background is given by [44–48]:

$$H(\vec{r}) = \frac{N}{3} \sum_{l=x,y} \sum_{j=x,y,z} \left| w_{lj}(\vec{r}) \right|^2 + \frac{b}{a^2}, \tag{7.18}$$

where the functions $w_{lj}(\vec{r})$ represent the electric field component $l = x, y$ in the image plane proportional to the emission dipole component $j = x, y, z$. These functions can be expressed as integrals over the pupil plane:

$$w_{lj}(\vec{r}) = \frac{1}{w_n} \int_{|\vec{\rho}| \leq 1} d^2\rho\, A(\vec{\rho}) \exp\left(\frac{2\pi i W(\vec{\rho})}{\lambda} \right) q_{lj}(\vec{\rho}) \exp\left(-i\vec{k}(\vec{\rho}) \cdot \vec{r} \right), \tag{7.19}$$

where the integration is over normalized pupil coordinates $\vec{\rho}$, w_n is a normalization factor, $A(\vec{\rho})$ is the amplitude, including the well-known aplanatic correction factor [44, 45], the $q_{lj}(\vec{\rho})$ are polarization vector components defined elsewhere [48], and where $W(\vec{\rho})$ is the aberration function induced by the SLM. The wavevec-

tor $\vec{k}(\vec{\rho})$ depends on the normalized pupil coordinates $\vec{\rho}$ by:

$$\vec{k}(\vec{\rho}) = \frac{2\pi}{\lambda}\left(\text{NA}\rho_x, \text{NA}\rho_y, \sqrt{n^2 - \text{NA}^2\vec{\rho}^2}\right), \tag{7.20}$$

with n the medium refractive index and NA the objective NA.

The μ_k are linear functions of the signal photon count N and the background per pixel b, making the first order derivatives of the μ_k with respect to these fit parameters easy to evaluate. The derivatives with respect to the fit parameters $\theta_m = x_0, y_0, z_0, \lambda$ are considerably more elaborate:

$$\frac{\partial\mu_k}{\partial\theta_m} = \frac{2N}{3}\sum_{l=x,y}\sum_{j=x,y,z}\int_{\mathcal{D}_k}dxdy\,\text{Re}\left\{w_{lj}(\vec{r}-\vec{r}_0)^*\frac{\partial w_{lj}(\vec{r}-\vec{r}_0)}{\partial\theta_m}\right\}. \tag{7.21}$$

The derivatives of the functions $w_{lj}(\vec{r}-\vec{r}_0)$ are given by:

$$\frac{\partial w_{lj}(\vec{r}-\vec{r}_0)}{\partial\vec{r}_0} =$$
$$\frac{i}{w_n}\int_{|\vec{\rho}|\leq 1}d^2\rho\,A(\vec{\rho})\exp\left(\frac{2\pi i W(\vec{\rho})}{\lambda}\right)q_{lj}(\vec{\rho})\vec{k}(\vec{\rho})\exp\left(-i\vec{k}(\vec{\rho})\cdot(\vec{r}-\vec{r}_0)\right), \tag{7.22}$$

and:

$$\frac{\partial w_{lj}(\vec{r}-\vec{r}_0)}{\partial\lambda} =$$
$$-\frac{2\pi i}{w_n\lambda^2}\int_{|\vec{\rho}|\leq 1}d^2\rho\,A(\vec{\rho})\exp\left(\frac{2\pi i W(\vec{\rho})}{\lambda}\right)q_{lj}(\vec{\rho})W(\vec{\rho})\exp\left(-i\vec{k}(\vec{\rho})\cdot(\vec{r}-\vec{r}_0)\right)$$
$$+\frac{1}{\lambda}(\vec{r}-\vec{r}_0)\cdot\frac{\partial w_{kj}(\vec{r}-\vec{r}_0)}{\partial\vec{r}_0}, \tag{7.23}$$

All relevant functions can be evaluated efficiently by 2D Fourier transforms, implemented using the chirp z-transform technique [49–51].

7.2.2. Numerical results

We have tested the performance limits of the different configurations by computing the CRLB as a function of the axial position and the emission wavelength. This has been done for an ROI with an $x \times y$-size of 25×15 pixels (diffractive astigmatic PSF, blazed grating design described in section 2.2) or 15×15 pixels (double helix PSF and astigmatic PSF) of pixel size $a = 100$ nm, for $N = 1500$ signal photons and $b = 10$ background photons/pixel, taking only shot noise into account. Fig. 7.3

shows the results of this analysis. It appears that the position (x, y, z) precision limit is comparable for all cases with the double helix PSF giving the most constant precision across all z-values. The axial (z) precision is typically twice as large as the lateral (x, y) precision. The spectral (λ) precision is best for the diffractive astigmatic PSF and worst for the astigmatic PSF. For an axial range on the order of 1 μm for the double helix design it appears that $L = 2$ or $L = 3$ annular zones suffice.
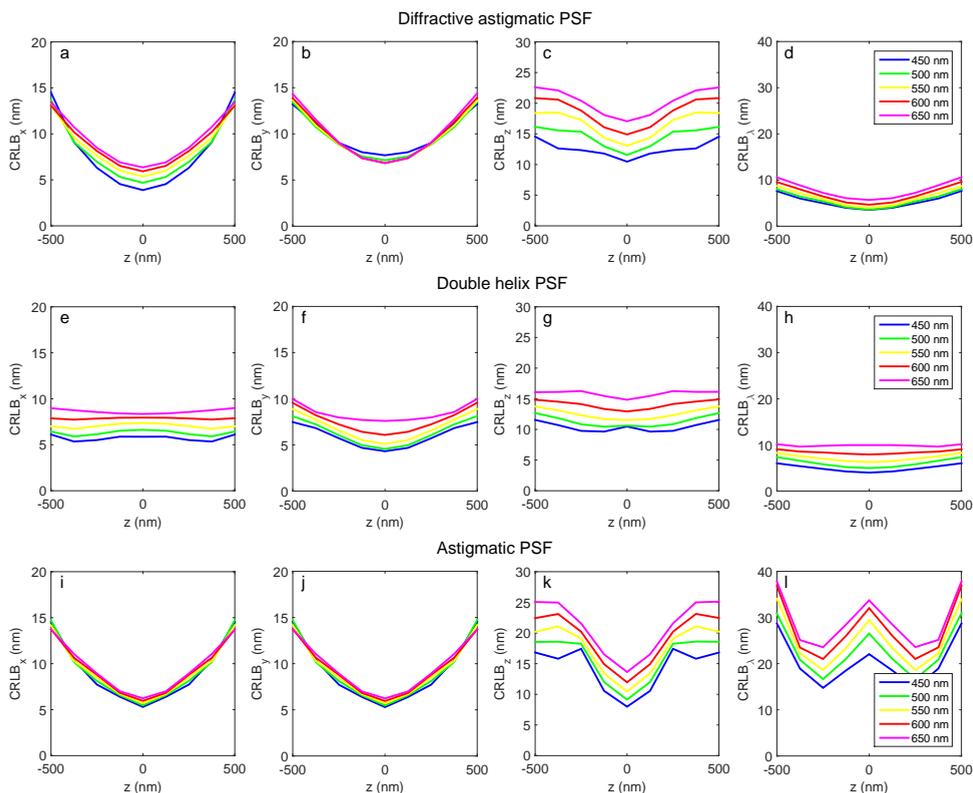
Figure 7.3. CRLB in emitter 3D-position and emission wavelength as a function of the z-position of the emitter and the emission wavelength λ for 1500 signal photons and 10 background photons/pixel. (abcd) Diffractive astigmatic PSF using two diffraction orders. (efgh): Double helix PSF. (qrst): astigmatic PSF.

The vectorial PSF fitting method has been tested with a simulation study. Fig. 4 shows the numerically determined fit error and the CRLB in x, y and z-position of the emitter and in the emission wavelength λ as a function of signal photon count for two background levels. For each signal photon count and background setting $N_{\text{inst}} = 100$ random instances for the 3D-position and wavelength were

generated (axial range ±500 nm, spectral range ±75 nm around central wavelength of 550 nm), the fit error was determined as the standard deviation of the fit parameter values at convergence. The uncertainty in the fit error was taken to be a factor $\sqrt{2/N_{\text{inst}}}$ times the fit error, assuming a Gaussian distribution of errors. This numerical fit error was compared to the mean and standard deviation of the CRLB-values for the fit parameters at the found optimum. Outliers were identified by a merit function at convergence that was more than three times the standard deviation of the distribution of final merit function values less than the mean of that distribution or by an x or y-position more than 1.5 pixels away from the center of the ROI. In the simulations only shot noise is taken into account, the readout noise is set to zero. Initial values for the MLE-fit were found as follows. The background was estimated from the median of all the pixels at the edge of the ROI, the signal photon count was found from the sum signal after background subtraction, and the lateral positions from the centroid position of the spot. The initial estimation for the axial position and wavelength depends on the PSF type. For the diffractive astigmatism design we proceeded with dividing the image in two halves corresponding to the two diffraction orders. The distance between the centroid estimates for the two halves and the ratio of the signal photon count in the two halves were used in an estimate for the wavelength, the second order moments in the two halves were used for an estimate of the diagonal astigmatism, which measures the axial position of the emitter. For the double helix and the astigmatic PSF the center wavelength of the spectral range was chosen as initial value. The second order moments of the light distribution were used to infer an estimate for the orientation (double helix) or elongation (astigmatism) of the spot and in this way to provide an initial estimate for the axial position. This set of initial values was used for a Levenberg-Marquardt optimization of the log-likelihood. Reasonable initial values promote a speedy and robust convergence of the optimizer.

The numerical fitting routine works well for most of the signal and background settings, as may be concluded from the correspondence between the numerically determined fit error and the CRLB, and from the fraction of outlier fits that is below 1 %. The method breaks down for the least favorable signal-to-background levels (200 signal photons at 10 background photons/pixel), where the precision exceeds the CRLB and where the fraction of outlier fits is increased to about 5 % for the diffractive astigmatic PSF and even to 20 to 40 % for the double helix and astigmatic PSF. The numerically found fit uncertainty agrees reasonably well with the CRLB values reported in Fig. 7.3. The quantitative level of deterioration of fit uncertainty by high background is comparable for all PSF cases considered, which
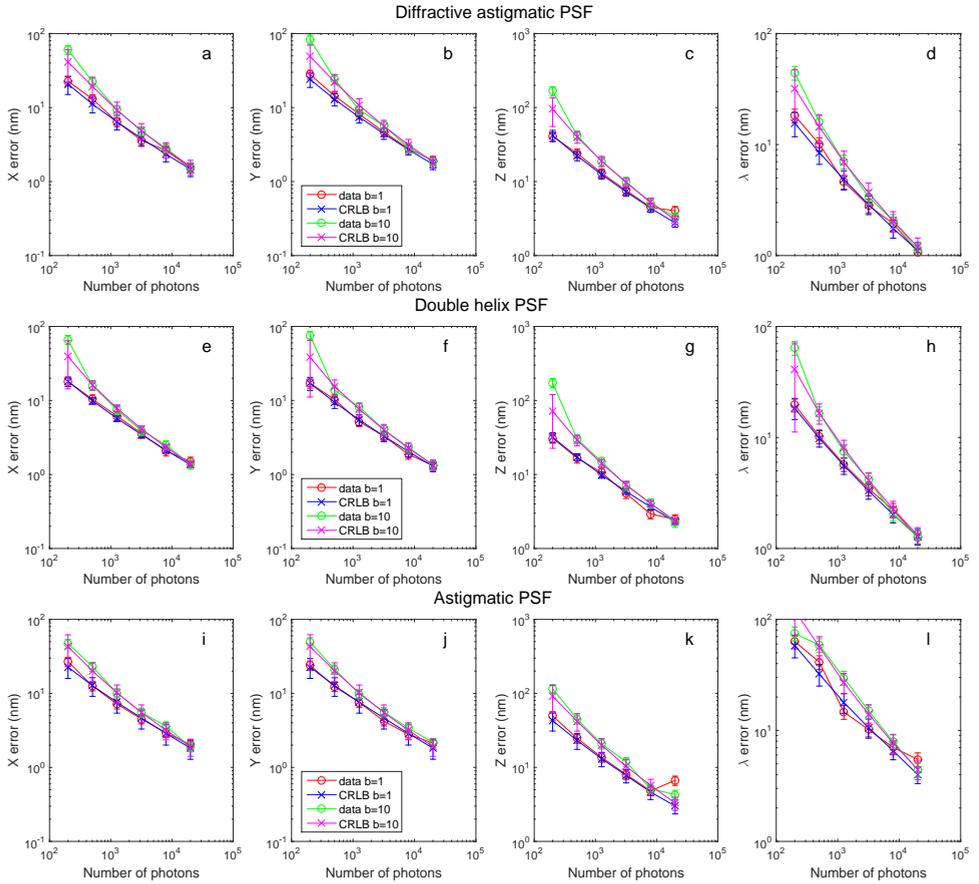
Figure 7.4. Numerically determined fit error and CRLB in the x, y and z-position of the emitter and in the emission wavelength λ as a function of signal photon count for two background levels, (abcd) Diffractive astigmatic PSF case. (efgh) Double helix PSF case, (ijkl) astigmatic PSF case.

can be understood from the comparable size of the spot footprint on the detector. The fit uncertainty in x and y is comparable for all methods, the fit uncertainty in z is typically twice as large, and the fit uncertainty in λ is best for the diffractive astigmatic PSF, slightly worse for the double helix PSF, and significantly worse for the astigmatic PSF.

Additional simulations have been done to test the robustness of the vector PSF fitting method. In the simulations of Fig. 4 monochromatic synthetic data was generated and fitted with a monochromatic vector PSF model. Real fluorescence signals have a non-zero spectral bandwidth, which gives rise to differences with the monochromatic model PSF. In order to test the effect of these differences synthetic data is generated by computing a weighted average of the PSF over a Gaussian emission spectrum with standard deviation σ_{em} (full-width-half maximum $2\sqrt{2\log(2)}\sigma_{\mathrm{em}}$) and fitting this spectrally averaged PSF with a monochromatic PSF for estimating the 3D-position and the peak wavelength of the emission spectrum. Fig. 7.4(abcd) show the fit error and CRLB as a function of the spectral bandwidth σ_{em}. For values below roughly 20 nm no significant loss in precision is found, implying that the use of a monochromatic model PSF can safely be used for most emitters. It is mentioned that an asymmetric emission spectrum, which organic fluorophores typically have, is expected to give rise to a bias in the estimation of the peak wavelength but not to a loss in precision.

The microscope setup may suffer from (unknown) aberrations, which will result in deviations of the image data from the model PSF, which only takes into account the designed aberrations generated by the SLM. In order to assess the gravity of unknown aberrations synthetic data is generated with random aberrations of a given root mean square (rms) value W_{rms} and fitted with the vector model PSF without these random aberrations. Fig. 7.4(efgh) show the fit error and CRLB as as function of the rms value W_{rms}. The data points are averaged over 100 random instances, where we take randomly different aberration coefficients for the 12 lowest order aberrations for each of the 100 instances. It appears that the precision for all parameters increases steeply with aberration level. Control of aberrations at the conventional level of the diffraction limit ($W_{\mathrm{rms}} = 72\,m\lambda$) is not good enough, about half the diffraction limit seems a more sensible tolerance limit. This implies that the vectorial fitting method is relatively sensitive to model PSF errors such as aberrations.

7.2.3. Experiment

We have done proof-of-principle experiments on different quantum dots. The setup uses an Olympus set of lenses consisting of a 150X/NA1.4 UPLSAPO objec-
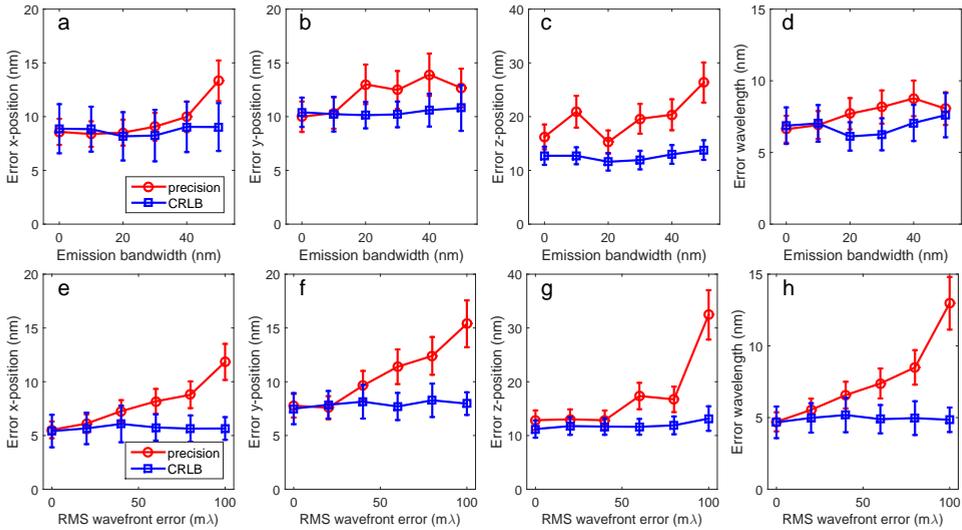
Figure 7.5. Simulated effect of non-zero emission bandwidth and of unknown aberrations on fit precision in the x, y and z-position of the emitter and in the emission wavelength λ. (abcd) Precision and CRLB averaged over 100 random instances as a function of the width of Gaussian emission spectrum. (efgh) Precision and CRLB averaged over 100 random instances as a function of the rms value of unknown aberrations.

tive lens and a standard LU095500 tube lens (180 mm focal length). An intermediate relay path is built with two Olympus LU095500 180 mm focal length lenses and a Meadowlark SLM (model XY-SLM, reflective Liquid Crystal on Silicon design, 256×256 pixels, pixel size 24 μm). A wire grid polarizer (WP25M-VIS, Thorlabs) was used to filter out the polarization component that is not modulated by the SLM. The SLM was configured to produce the diffractive astigmatic blazed grating PSF design with a . The raw images were captured on an Andor iXon Ultra 897 EM-CCD camera (512×512 pixels, pixel size 16 μm). Excitation was done with a 405 nm diode laser, directed towards the objective via a 458 nm long pass dichroic. The emitted light was captured and filtered with a notch filter (405 StopLine, Semrock) and a 450-650 nm bandpass emission filter (FF01-550/200, Semrock). The setup was used to image Quantum Dots (QDs) with specified emission peaks at 525 nm and 585 nm (ThermoFisher Q21341MP and Q21311MP) prepared on separate glass slides and immersed in glycerol (refractive index 1.47). Through-focus image stacks were taken by moving the objective lens with a stage (nanoXYZ, Physik Instrumente).

Visual inspection of the measured spots revealed a large asymmetry between the spot shape above and below the nominal focal plane. Such an asymmetry

178

is usually indicative for spherical aberration. It appears that the observed spot shapes can be reasonably accounted for by assuming that the setup suffers from about 0.06λ rms spherical aberration. The parameters characterizing the SLM pattern were adjusted as well to improve the correspondence between observed spot shapes and the vector PSF model. We found that a tip equal to 2.2λ and astigmatism equal to 0.49λ for a center wavelength of 550 nm, and a blaze step height of about 145 nm agree best with the observed spot shapes. It turned out that the spots appeared rotated over about 14 deg with the x, y axes of the camera frame due to a rotational misalignment of the SLM with respect to the camera. This rotation of the SLM phase pattern was incorporated into the vector PSF model. With these parameter settings we have applied our vectorial PSF fitting model to the data, the results are shown in Fig. 7.5. The fits were done over 25×15 pixel ROIs. The precision was determined from fits to 30 repeated acquisitions of the same QD. The mean and standard deviation of each of the fit parameters (x, y, z, λ) was found from a weighted average over the 30 fit values with weight proportional to the fitted signal photon count. The same procedure was followed for the mean and standard deviation of the CRLB. Fit error and CRLB data was obtained in this way for 5 different QDs, the displayed data points are the weighted averages over the fit error and CRLB for the 5 QDs with weight inversely proportional to the square of the standard deviation. The mean background over the 30 repeated acquisitions was typically 17 photons per pixel, the mean signal photon count per acquisition was typically around 800 photons. The experimental precision (Fig. 6(abcd)) does not quite reach the CRLB. The resulting precision in x and y ranges from 10 to 30 nm, the precision in z ranges from 25 to 50 nm, and the precision in λ ranges from 10 to 20 nm, an axial localization precision of 25 to 50 nm, and a lateral localization precision of 10 to 30 nm. There is an asymmetry in precision between positive and negative z values that we attribute to the apparent spherical aberration. In addition, there is a substantial bias in the fitted wavelength (up to 15 nm) and the axial position that depends on the axial position of the stage (Fig. 6(ef)). The resulting spread in fitted wavelengths amounts to about 20 nm (Fig. 6(g)).

The experimental PSF for all axial stage positions is determined by first up-sampling the raw images (with a factor 5), then displacing the images to the center of the ROI using the fitted lateral (x, y) position values, and then summing all individual images (30 repeated acquisitions of 5 QDs). Fig. 7 shows a slice of the through-focus experimental PSF and the model PSF at the fitted axial position. The asymmetry in spot shape below and above focus due to spherical aberration is clearly visible. There is a reasonable match between the data and the model but
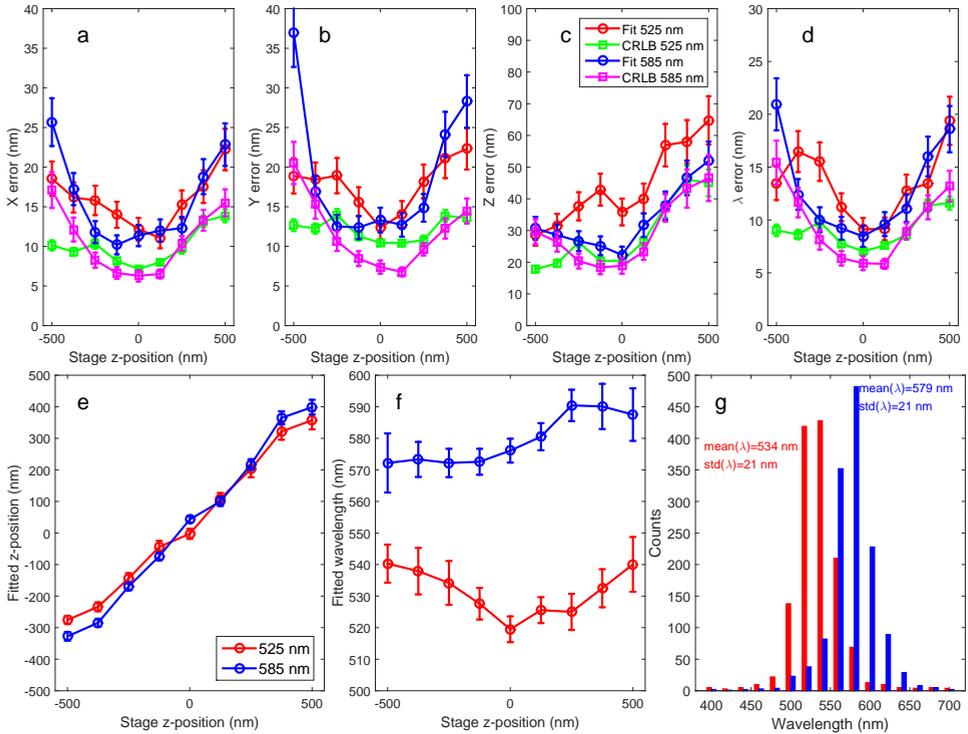
Figure 7.6. Results on experiments on quantum dots emitting at 525 nm and 585 nm. (abcd) Experimental precision in x, y, z, λ determined from repeated fits of the same quantum dot compared to the CRLB for the fitted parameter values. (ef): Fitted z position and fitted wavelength as a function of stage position, showing a z-dependent bias in the fitted wavelength of up to 15 nm. (g) Histogram of fitted wavelengths for all repeated acquisitions at all focus levels of all QDs ($30 \times 9 \times 5 = 1350$ data points).

not a very good quantitative match. The differences are bigger than the blurring arising from the shift of the measured spots to the center of the ROI with the fitted x, y positions, which are subject to errors on the order of 25 nm. The reasons for the mismatch are not clear but could be found in unknown aberrations, optical alignment errors, and errors in the calibration of the SLM. An additional factor may be found in the spectral diffusion of QDs, which can have jumps in emission wavelength of around 10 nm after blinking events [52].
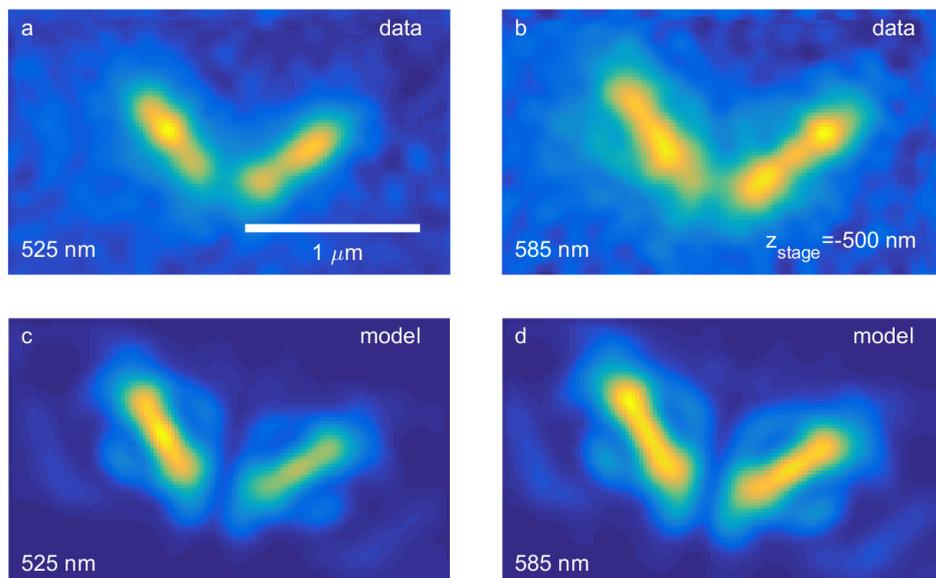
Figure 7.7. Single focal slice of the experimental through-focus PSF and the corresponding model through-focus PSF for the QD525 and the QD585 species. The orientational misalignment of the SLM with respect to the camera (over an angle of 14 deg) has been incorporated to the fitting routine. The asymmetry in spot shape between positive and negative defocus is due to spherical aberration (estimated as 0.06λ rms).

We have also done a proof-of-principle experiment in which we have imaged a mixture of the two types of QDs. Fig. 8 shows a frame of the measured through-focus images and ROIs for 6 QDs that were visible in most focal slices. The ROIs have been assigned a pseudo-color derived from the fit of the measured spots with our vectorial PSF model. First, a spectrally weighted average of the RGB-values corresponding to monochromatic light is computed, where the weight is taken to be a Gaussian function with peak at the fitted wavelength and width equal to the CRLB of the fitted wavelength. Subsequently, this spectrally weighted RGB is used as colormap in the MATLAB plot, thus giving a visual impression of the

measured color of the QDs. It appears that three of the QDs have a red-amber color and may be identified as belonging to the species with emission peak at 585 nm, and three of the QDs have a green-yellow color and may be identified as belonging to the species with emission peak at 525 nm. Close to the nominal focus the spots are clearly visible and the fitting procedure works reasonably well. Away from the nominal focus (the smallest and largest axial coordinate values) the signal-to-noise ratio is so low that the spots can not or hardly be seen in the recorded image and the fitting becomes unreliable.
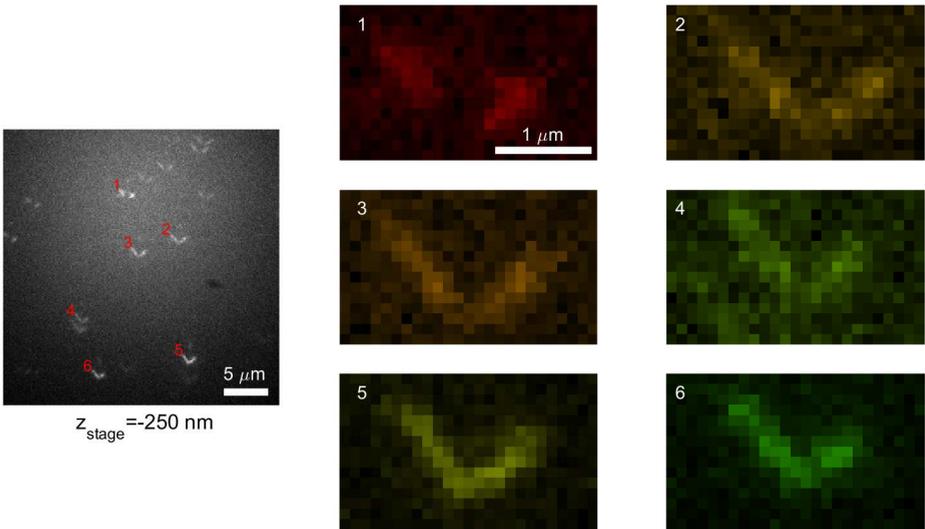
Figure 7.8. Single frame showing the recorded images for different axial positions (left), and the 6 indicated ROIs (right). All images are first contrast stretched and then rendered with an RGB colormap corresponding to the fitted wavelength, so they do not accurately represent the actual signal level. QDs 1, 2, and 4 have the red-amber appearance of the species with emission peak at 585 nm, QDs 4,5, and 6 have the green-yellow appearance of the species with emission peak at 525 nm.

7.3. *Discussion*

In summary, we have proposed a new method for the simultaneous measurement of the 3D-position and the emission wavelength of single emitters. In particular, we have investigated a diffractive optics based PSF design in which the spot is split into closely spaced diffraction orders, and we have evaluated, both numerically and experimentally, a fitting method based on a vector PSF model. A lateral localization precision between 10 and 30 nm, an axial localization precision between

25 and 50 nm, and a spectral precision between 10 and 20 nm was achieved. The main limiting factor for these figures of merit is the accuracy of the PSF model. A full characterization and possibly compensation of aberrations [53] may be expected to bring down the level of precision in x, y, z, λ to the numbers predicted by simulations.

An avenue which we have not explored in full is the study of fully optimized SLM profiles, which could possible lead to new solutions with further improved performance. In particular, the addition of higher order astigmatism (non-zero A_{4-2}, A_{6-2}) may significantly improve precision. An open issue here is how to handle or incorporate aberration profiles that have one or more singularities, such as the pattern needed for double-helix type PSFs. In view of the singularities an expansion in Zernike modes does not seem suitable, and a different way to parametrize a broad class of patterns with a small number of parameters would be needed.

The main drawback of the proposed technique is the increased footprint of the PSF. In standard 2D imaging with diffraction-limited spots the smallest possible footprint is achieved with the highest possible signal-to-background ratio. This results in the best possible (lateral) localization precision, which depends significantly on the signal-to-background ratio [39]. The price of an increased spot footprint, however, is already paid upon making the transition to 3D imaging. The additional increase in the PSF support size of the newly proposed PSFs is on the order of several tens of percents and can possibly be reduced by tweaking the design parameters. An additional drawback is related to possible applications in which the single molecules are imaged and randomly activated simultaneously. In order to prevent having too big a fraction of overlapping spots per recorded frame, the requirements on the ratio of the off time τ_{off} and the on time τ_{on} and/or on the maximum allowable labeling density ρ are more stringent. These should satisfy $\rho A_{\text{PSF}} \tau_{\text{on}} / \tau_{\text{off}} \ll 1$, with A_{PSF} the PSF footprint area, so the upper limit for ρ is smaller by the same percentage as the PSF footprint is larger.

In our experiment we have made use of a liquid crystal based SLM, which is polarization sensitive as the SLM aberration profile is added to only one of the two orthogonal linear polarization states. For the sake of simplicity we have filtered out the unaffected polarization state, but this is not desirable in photon starved applications

An alternative to the fitting of the emission wavelength is to perform a Generalized Likelihood Ratio Test (GLRT), similar to the one recently proposed by us for detecting dim single emitter events at relatively high background [54]. The GLRT turns the original wavelength estimation into a classification between the

multiple fluorescent species. In a GLRT approach each spot is fitted N times assuming the known emission wavelength for the N different species. The fit with the highest likelihood would then lead to the identification of the fluorescent species at hand.

References

[1] 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**, 1642 (2006).

[2] M. J. Rust, M. Bates, and X. W. Zhuang, *Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM),* Nature Methods **3**, 793 (2006).

[3] S. T. Hess, T. P. K. Girirajan, and M. D. Mason, *Ultra-high resolution imaging by fluorescence photoactivation localization microscopy,* Biophysical Journal **91**, 4258 (2006).

[4] S. W. Hell, *Far-field optical nanoscopy,* Science **316**, 1153 (2007).

[5] R. P. Nieuwenhuizen, K. A. Lidke, M. Bates, D. L. Puig, D. Grünwald, S. Stallinga, and B. Rieger, *Measuring image resolution in optical nanoscopy,* Nature Methods **10**, 557 (2013).

[6] H. Shroff, C. G. Galbraith, J. A. Galbraith, H. White, J. Gillette, S. Olenych, M. W. Davidson, and E. Betzig, *Dual-color superresolution imaging of genetically expressed probes within individual adhesion complexes,* Proceedings of the National Academy of Sciences **104**, 20308 (2007).

[7] M. Bossi, J. Foolling, V. N. Belov, V. P. Boyarskiy, R. Medda, A. Egner, C. Eggeling, A. Schonle, and S. W. Hell, *Multicolor far-field fluorescence nanoscopy through isolated detection of distinct molecular species,* Nano Letters **8**, 2463 (2008).

[8] I. Testa, C. A. Wurm, R. Medda, E. Rothermel, C. Von Middendorf, J. Folling, S. Jakobs, A. Schonle, S. W. Hell, and C. Eggeling, *Multicolor fluorescence nanoscopy in fixed and living cells by exciting conventional fluorophores with a single wavelength,* Biophysical Journal **99**, 2686 (2010).

[9] D. Grunwald and R. H. Singer, *In vivo imaging of labelled endogenous -actin mRNA during nucleocytoplasmic transport,* Nature **467**, 604 (2010).

[10] P. J. Cutler, M. D. Malik, S. Liu, J. M. Byars, D. S. Lidke, and K. A. Lidke, *Multi-color quantum dot tracking using a high-speed hyperspectral line-scanning microscope,* PloS One **8**, e64320 (2013).

[11] M. Bates, B. Huang, G. T. Dempsey, and X. Zhuang, *Multicolor super-resolution imaging with photo-switchable fluorescent probes,* Science **317**, 1749 (2007).

[12] J. Tam, G. A. Cordier, J. S. Borbely, Á. S. Álvarez, and M. Lakadamyali, *Cross-talk-free multi-color STORM imaging using a single fluorophore,* PloS One **9**, e101772 (2014).

[13] C. C. Valley, S. Liu, D. S. Lidke, and K. A. Lidke, *Sequential superresolution imaging of multiple targets using a single fluorophore,* PloS One **10**, e0123941 (2015).

[14] J. Broeken, B. Rieger, and S. Stallinga, *Simultaneous measurement of position and color of single fluorescent emitters using diffractive optics,* Optics Letters **39**, 3352 (2014).

[15] Z. Zhang, S. J. Kenny, M. Hauser, W. Li, and K. Xu, *Ultrahigh-throughput single-molecule spectroscopy and spectrally resolved superresolution microscopy,* Nature Methods **12**,

935 (2015).

[16] A. S. Backer and W. Moerner, *Extending single-molecule microscopy using optical Fourier processing,* The Journal of Physical Chemistry B **118**, 8313 (2014).

[17] H. P. Kao and A. Verkman, *Tracking of single fluorescent particles in three dimensions: Use of cylindrical optics to encode particle position.* Biophysical Journal **67**, 1291 (1994).

[18] L. Holtzer, T. Meckel, and T. Schmidt, *Nanometric three-dimensional tracking of individual quantum dots in cells,* Applied Physics Letters **90**, 1 (2007).

[19] B. Huang, W. Wang, M. Bates, and X. Zhuang, *Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy,* Science **319**, 810 (2008).

[20] E. Toprak, H. Balci, B. H. Blehm, and P. R. Selvin, *Three-dimensional particle tracking via bifocal imaging,* Nano Letters **7**, 2043 (2007).

[21] S. Ram, P. Prabhat, J. Chao, E. S. Ward, and R. J. Ober, *High accuracy 3D quantum dot tracking with multifocal plane microscopy for the study of fast intracellular dynamics in live cells,* Biophysical Journal **95**, 6025 (2008).

[22] M. F. Juette, T. J. Gould, M. D. Lessard, M. J. Mlodzianoski, B. S. Nagpure, B. T. Bennett, S. T. Hess, and J. Bewersdorf, *Three-dimensional sub–100 nm resolution fluorescence microscopy of thick samples,* Nature Methods **5**, 527 (2008).

[23] M. J. Mlodzianoski, M. F. Juette, G. L. Beane, and J. Bewersdorf, *Experimental characterization of 3D localization techniques for particle-tracking and super-resolution microscopy,* Optics Express **17**, 8264 (2009).

[24] S. Abrahamsson, J. Chen, B. Hajj, S. Stallinga, A. Y. Katsov, J. Wisniewski, G. Mizuguchi, P. Soule, F. Mueller, C. D. Darzacq, *et al., Fast multicolor 3D imaging using aberration-corrected multifocus microscopy,* Nature Methods **10**, 60 (2013).

[25] S. R. P. Pavani and R. Piestun, *Three dimensional tracking of fluorescent microparticles using a photon-limited double-helix response system,* Optics Express **16**, 22048 (2008).

[26] S. R. P. Pavani, M. A. Thompson, J. S. Biteen, S. J. Lord, N. Liu, R. J. Twieg, R. Piestun, and W. Moerner, *Three-dimensional, single-molecule fluorescence imaging beyond the diffraction limit by using a double-helix point spread function,* Proceedings of the National Academy of Sciences **106**, 2995 (2009).

[27] G. Grover, K. DeLuca, S. Quirin, J. DeLuca, and R. Piestun, *Super-resolution photon-efficient imaging by nanometric double-helix point spread function localization of emitters (SPINDLE),* Optics Express **20**, 26681 (2012).

[28] S. Prasad, *Rotating point spread function via pupil-phase engineering,* Optics Letters **38**, 585 (2013).

[29] C. Roider, A. Jesacher, S. Bernet, and M. Ritsch-Marte, *Axial super-localisation using rotating point spread functions shaped by polarisation-dependent phase modulation,* Optics Express **22**, 4029 (2014).

[30] D. Baddeley, M. B. Cannell, and C. Soeller, *Three-dimensional sub-100 nm super-resolution imaging of biological samples using a phase ramp in the objective pupil,* Nano Research **4**, 589 (2011).

[31] S. Jia, J. C. Vaughan, and X. Zhuang, *Isotropic three-dimensional super-resolution imaging with a self-bending point spread function,* Nature Photonics **8**, 302 (2014).

[32] G. Shtengel, J. A. Galbraith, C. G. Galbraith, J. Lippincott-Schwartz, J. M. Gillette, S. Manley, R. Sougrat, C. M. Waterman, P. Kanchanawong, M. W. Davidson, *et al., Interferometric fluorescent super-resolution microscopy resolves 3D cellular ultrastructure,* Proceedings of the National Academy of Sciences **106**, 3125 (2009).

[33] Y. Shechtman, S. J. Sahl, A. S. Backer, and W. Moerner, *Optimal point spread function design for 3D imaging,* Physical Review Letters **113**, 133902 (2014).

[34] Y. Shechtman, L. E. Weiss, A. S. Backer, S. J. Sahl, and W. Moerner, *Precise three-dimensional scan-free multiple-particle tracking over large axial ranges with tetrapod point spread functions,* Nano Letters **15**, 4194 (2015).

[35] S. Stallinga, *Diffractive optical element for localization microscopy,* NL patent, 2012187 (2014).

[36] M. Born and E. Wolf, *Principles of optics: Electromagnetic theory of propagation, interference and diffraction of light* (Cambridge University, 1997).

[37] R. J. Ober, S. Ram, and E. S. Ward, *Localization accuracy in single-molecule microscopy,* Biophysical Journal **86**, 1185 (2004).

[38] C. S. Smith, N. Joseph, B. Rieger, and K. A. Lidke, *Fast, single-molecule localization that achieves theoretically minimum uncertainty,* Nature Methods **7**, 373 (2010).

[39] B. Rieger and S. Stallinga, *The lateral and axial localization uncertainty in super-resolution light microscopy,* ChemPhysChem **15**, 664 (2014).

[40] F. Huang, T. M. Hartwich, F. E. Rivera-Molina, Y. Lin, W. C. Duim, J. J. Long, P. D. Uchil, J. R. Myers, M. A. Baird, W. Mothes, *et al.*, *Video-rate nanoscopy using sCMOS camera-specific single-molecule localization algorithms,* Nature Methods **10**, 653 (2013).

[41] J. Chao, E. S. Ward, and R. J. Ober, *Fisher information matrix for branching processes with application to electron-multiplying charge-coupled devices,* Multidimensional Systems and Signal Processing **23**, 349 (2012).

[42] S. Stallinga, *Single emitter localization analysis in the presence of background,* in *SPIE Optical Systems Design* (International Society for Optics and Photonics, 2015) pp.

96300V–96300V.

[43] K. I. Mortensen, L. S. Churchman, J. A. Spudich, and H. Flyvbjerg, *Optimized localization analysis for single-molecule tracking and super-resolution microscopy,* Nature Methods **7**, 377 (2010).

[44] T. Wilson, R. Juškaitis, and P. Higdon, *The imaging of dielectric point scatterers in conventional and confocal polarisation microscopes,* Optics Communications **141**, 298 (1997).

[45] P. Török, P. Higdon, and T. Wilson, *Theory for confocal and conventional microscopes imaging small dielectric scatterers,* Journal of Modern Optics **45**, 1681 (1998).

[46] J. Enderlein, E. Toprak, and P. Selvin, *Polariztion effect on position accuracy of fluorophore localization,* Optics Express **14**, 8112 (2006).

[47] S. Stallinga and B. Rieger, *Accuracy of the gaussian point spread function model in 2D localization microscopy,* Optics Express **18**, 24461 (2010).

[48] S. Stallinga, *Effect of rotational diffusion in an orientational potential well on the point spread function of electric dipole emitters,* JOSA A **32**, 213 (2015).

[49] X. Deng, B. Bihari, J. Gan, F. Zhao, and R. T. Chen, *Fast algorithm for chirp transforms with zooming-in ability and its applications,* JOSA A **17**, 762 (2000).

[50] J. L. Bakx, *Efficient computation of optical disk readout by use of the chirp z transform,* Applied optics **41**, 4897 (2002).

[51] M. Leutenegger, R. Rao, R. A. Leitgeb, and T. Lasser, *Fast focus field calculations,* Optics Express **14**, 11277 (2006).

[52] R. Neuhauser, K. Shimizu, W. Woo, S. Empedocles, and M. Bawendi, *Correlation between fluorescence intermittency and spectral diffusion in single semiconductor quantum dots,* Physical Review Letters **85**, 3301 (2000).

7

[53] I. Izeddin, M. El Beheiry, J. Andilla, D. Ciepielewski, X. Darzacq, and M. Dahan, *PSF shaping using adaptive optics for three-dimensional single-molecule super-resolution imaging and tracking,* Optics Express **20**, 4957 (2012).

[54] C. S. Smith, S. Stallinga, K. A. Lidke, B. Rieger, and D. Grunwald, *Probability-based particle detection that enables threshold-free and robust in vivo single-molecule tracking,* Molecular Biology of the Cell **26**, 4057 (2015).

7

Color and 3D Position Estimation of Single Molecules

8

Outlook

*I think that only daring speculation can lead us further
and not accumulation of facts.*

ALBERT EINSTEIN

8.1. *Conclusions*

One of the fundamental differences between conventional fluorescence microscopy and single molecule localization microscopy is the tremendous increase in data that needs to be analyzed. The challenge for single molecule localization microscopy is to develop image analysis methods that perform the detection and the estimation as efficiently and precisely as possible with minimal user input. For this purpose methods are proposed in this thesis that improve the analysis by employing optimum statistical techniques that do not need manual fine-tuning.

For wide-spread use of single molecule localization microscopy, the two-dimensional methods need to be extended to allow for three-dimensional imaging. A pixelated Gaussian PSF is a reasonable approximation to an in-focus PSF. However, the validity of this approximation decreases significantly for three-dimensional imaging purposes. PSF models that account for the aberrations and vectorial effects are presented here to move successfully to imaging in three dimensions. In addition, multiple three-dimensional single molecule localization approaches are studied and are extended to four dimensions by encoding the emission color of the single molecule into the measurements. As noted in the introduction, RNA imaging is selected as both a challenging and biologically interesting application for single molecule localization microscopy. RNA imaging is relevant since RNA regulates gene expression in cells and thus has the potential to be a target for drugs to treat human disease [1]. Often it is known where and with what molecule an RNA molecule interacts, but the kinetics or order of interaction are unknown, and this is where single molecule localization microscopy can add its value. RNA imaging is challenging due to the low signal to background ratios, making the processing of the data a cumbersome task.

8.1.1. Single molecule localization and detection

In Chapter 2 a localization algorithm is proposed that is executed on a GPU-based parallel processing platform for near real-time processing of the tremendous amount of data that single molecule localization microscopy experiments generate. This iterative approach gives a maximum likelihood estimate of the position and intensity of a single fluorophore as well as the background count. For the derivation of the iterative algorithm a two-dimensional pixelated Gaussian PSF model and a Poissonian noise model are assumed. The method achieves the minimal possible estimation uncertainty, as given by the CRLB, over a wide range of emission and background rates that are found in single molecule experiments. Implementation of the iterative method on a modern GPU yields more than one million combined fits and CRLB calculations per second on a single GPU (NVIDIA

GeForce GTX 790 - 4992 Cores). The computing ability of currently available high-end GPU cards enables single molecule based super-resolution techniques to be executed in real-time, which was previously inconceivable.

8.1.2. Single molecule detection

Single molecules need to be detected in the raw data before they can be localized. Historically a scale space method is used for this purpose. However, a significant challenge in the detection of single molecules is the unknown number of false positives and true positives, especially at low signal to background ratios (SBR). In Chapter 3 an approach is presented to control both the number of false positives and true positives with minimal user input. The approach gives a significant improvement over the current detection methods, as it results in a controlled number of false positives and a maximal number of true positives. This approach is based on a detection theory framework, which uses the maximum likelihood estimators introduced in Chapter 2 to perform a pixel-based generalized likelihood ratio based hypothesis test (GLRT). The GLRT best approximates the optimal likelihood ratio test as proven by Neyman and Pearson. In other words, GLRT allows setting a target for the false-positive rate for a broad range of applications, and the results show that this objective is stably achieved over a large range of SBR conditions.

GLRT is based on statistical testing and significance levels, thereby replacing user-defined thresholds that have an intricate impact on the false-positive rate and detection efficiency. The result of the proposed framework is robust detection of weaker signals, which opens up single molecule studies to a wider range of applications, as well as more robust and user independent performance that is suitable for non-expert users.

8.1.3. Single molecule tracking

In Chapter 4 single molecule localization microscopy is applied to a biological problem, where the nuclear pore complex and the kinetic interaction with mRNA are investigated. In this study the export kinetics of mRNP in *S. cerevisiae* is measured for the first time. This is accomplished by taking advantage of two powerful experimental systems: single-particle RNA localization microscopy and yeast genetics. From biochemical experiments, mex67p was identified to be a component for nuclear export in yeast. However, this interaction was never visualized in a live cell. Single molecule localization microscopy is used to provide direct evidence for the critical role of the mex67p protein in nuclear export. It is shown that a point mutation of mex67p caused the export efficiency to drop

significantly and gave rise to phenomena such as re-import and mRNPs scanning of the nuclear periphery in search of a suitable nuclear pore. These findings are made possible using single particle tracking, and future experiments using this technology can result in essential insight into gene regulation and protein transcription in the cell.

8.1.4. Three-dimensional single molecule tracking

In Chapter 5 the first full three-dimensional analysis of mRNA in the nucleus of a living cell is presented, at a time resolution relevant to the diffusion driven populations of mRNA. Multi-focus microscopy is used to instantaneously image the whole three-dimensional volume of mRNAs in the nucleus of a living cell. Instead of using a fiducial marker for registration of the different focal planes, the heterochromatin is stained, which makes it possible to adjust for cell-induced chromatic distortions using this cell internal marker. Subsequently, image registration and restoration are applied to compensate for the distortions. Three major image analysis components are used: geometric descriptor matching, to orient the image planes correctly; image entropy analysis, to find the chromatically corrected z position for each plane; and image deconvolution, accounting for varying spherical aberrations along the z-axis. To study the mobility and occupation within the nucleus, β-actin mRNA, a typical nuclear particle, is investigated within the nucleus. From this study, no evidence is obtained for exclusion or enrichment of the heterochromatin. However, the results of the study show that most β-actin mRNAs are found less than 500 nm away from the central channel of the nuclear pore complex.

8.1.5. Adaptive optics

In Chapter 6 an adaptive optics strategy is presented that can correct for sample induced phase aberrations. The optical aberrations influence the CRLB, therefore, the correction of sample induced aberrations can improve the attainable localization precision. The proposed method can quickly estimate aberrations from camera images and compensate for them by the use of adaptive optics. In the proposed approach the aberrations are determined from the PSF using a least squares approach. This method increases the computational speed and decreases the average CRLB with respect to existing methods that solely calibrate for aberrations in postprocessing. The proposed method is suitable for a multitude of applications besides single molecule localization.

8.1.6. Four-dimensional single molecule localization

In Chapter 7 multiple approaches are presented that extend two-dimensional single molecule localization into three dimensions. These approaches use the insertion of a phase mask into a wide field microscope, where the phase mask distorts the PSF and thereby encodes the three-dimensional position into the shape of the PSF. This allows for the extension of the algorithms derived in Chapters 2 and 3 to three-dimensional single molecule localization for any single plane PSF engineered imaging modality. Additionally, an approach is presented for encoding the emission color of a single molecule into the measurements. This novel technique obtains a measurement of the three-dimensional position while simultaneously measuring the emission wavelength of single emitters. The proposed technique combines a diffractive optics approach and a vectorial PSF design to split the PSF into closely spaced diffraction orders.

8.2. *Recommendations*

In this section recommendations are presented for further improvements following the work put forth in this thesis.

Short term

- The localization and detection methods presented in this thesis assume a locally homogeneous background, but as single molecule localization is attempted in thicker samples, this approximation loses its validity. A study should be performed to assess the impact of using a homogeneous background model on the estimation accuracy and precision. A straightforward extension of the homogeneous background model is to add extra degrees of freedom by assuming an additional gradient. Adding degrees of freedom to the background model benefits the detection algorithm since the probability that the signal is mistaken for background decreases when the background model is more realistic.

- The detection method presented in Chapter 3 is based on a single measurement. For single molecule tracking applications, additional information is present in adjacent frames, and thus the detection method presented in Chapter 3 should be extended from single frame to multi-frame detection. Multi-frame detection can be accomplished by use of apriori information on the movement of the single molecule, such as diffusion or velocity, which enables the use of multiple frames in the generalized likelihood ratio test. If the single molecule movement is only driven by diffusion, a ten percent de-

tection increase may be expected compared to the single frame generalized likelihood ratio test. A similar Bayesian framework has been proposed in the context of aerospace defense for the detection of hostile missiles, and is called *track before detect* [2].

- The four-dimensional single molecule localization technique presented in Chapter 7 assumes a monochromatic model. The model is sufficient for the narrow spectrum expected from quantum dots, however, for more general applicability it is worth considering generalization to a polychromatic model as broadband emission spectra can induce an estimation bias. Schemes for calculating a polychromatic model require convolution with the wavelength and are therefore more computationally demanding to evaluate. An alternative is to include prior knowledge of the emission spectra, to turn the color estimation into a color classification problem. Color classification could be accomplished by use of a generalized likelihood ratio test, where each spot is fitted N times assuming the known emission spectra for the N different labels used. The fit with the highest likelihood would then lead to the identification of the fluorescent species at hand.

Long term
- The optimal and fast two-dimensional detection and localization algorithms in Chapter 2 and 3 are developed for two-dimensional single molecule localization. The work in Chapter 7 enables the extension of these two methods to three-dimensional imaging, which would lead to a detection and localization algorithm that allows for a more general PSF model and therefore a wider application of the algorithms. It can be accomplished by adding a general phase mask as an input parameter and simultaneously estimating local aberrations, which would lead to generalization of the localization and segmentation over multiple imaging modalities such as double-helix, astigmatism, and tetrapod PSFs.

- In three-dimensional single molecule localization, one has to account for aberrations that are expected to vary over depth and field of view. These aberrations could be estimated locally, and the average aberration can be corrected by use of adaptive optics. This correction will improve the PSFs and therefore the average CRLB. This global correction should be applied to maximize the envisioned information, which can be done using the minimization of the CRLB.

8

The aberrations vary over the field of view, therefore, in minimization of the CRLB, the robustness of the PSF model with regard to the influence of local variations of aberrations should be accounted for. This cannot be achieved by using the standard CRLB, but could be incorporated into the CRLB by applying the Bayesian van Tree's inequality [3]. The result would be a robust three-dimensional single molecule localization technique with broad applicability.

8.3. *Outlook*

In this section, an outlook on future developments in single molecule localization microscopy is put forth. A key observation is that biologists and medical researchers are continuously looking to obtain more information out of thicker samples, at a higher resolution, in 3D, and at real time. In an ideal scientific future, it would be possible to walk into a 3D holographic room that shows exactly the machinery of the cell functions at a high resolution in real-time. Here the position (x,y,z) as a function of time (t), the molecular species, and potentially other properties of single molecules would be visualized. From the data visualized in this room, one could obtain quantitative measurements without imposing a model on the data. The technology needed to realize such a room is not yet in existence, if ever possible, but we can envision the steps required to move into this direction.

8.3.1. One to five years
In this section speculations are made on what kind of developments can be expected in the next five years.

From biochemistry on a coverslip to biochemistry in the cell
Recently, the single molecule localization field took the first pioneering steps to move to *in vitro* imaging of cell extracts. Here, dynamical processes are measured *in vitro* by using cell extracts on coverslips that provide scientists a controlled environment for probing biochemical processes within the cell. These single molecule biochemistry experiments can be performed using a variety of techniques: CoS-MoS [4, 5], SiMPull [6, 7], Immunofluorescence [8, 9], or smFRET [10, 11]. A major drawback of traditional *in vitro* imaging is that the micro-environment within a live cell is different from that of the artificial environment created on a coverslip. Cell extracts bring us closer to live cell imaging. However, this method pools many cells in different cell states together into one extract. The consequence is that these type of experiments can only give insight into single molecule dynamics relative to other controlled experiments. In the next five years, more single molecule

biochemistry experiments are expected to be performed within live cells by use of single molecule tracking.

To enable biochemistry experiments in the cell, both the experimental conditions and data processing tools will need to be improved. A major improvement is expected through the reduction of background noise from unbound fluorescent proteins or dyes (e.g. by using HaloTag and Snap-tag) due to imaging techniques that only excite a small region within the cell. An example of such a method is (high NA) light-sheet microscopy. Another improvement could be achieved through fluorescent proteins exhibiting optimized quantum efficiency and photo-stability. The algorithms that quantify the tracking data would need to be improved as well. The trajectories from tracking data are commonly quantified by fitting a fixed number of mobility states [12–15]. Misinterpretation and the influence of the model on the data are decreased when the number of dynamical states of the single molecules are estimated from the data [16, 17]. Still, these methods have limited applicability because the user imposes experimental specific apriori information on the data. Future, developments will be focused on generalizing these models and making them applicable over a wider range of applicators.

From a list of positions to quantitative information

Currently, image processing tools for segmentation operations and analysis assume data that is encoded as an array of pixels. The data in single molecule localization microscopy, however, consists of a list of positions for which not many image processing tools exist. In the next five years, an increase is expected in the development of basic image processing tools that can make sense of this type of data. Such methods will enable automatic extraction of quantitative information directly from the list of single molecule positions.

Standardization of image analysis tools

Single molecule localization microscopy could become a more widely used tool if a standardized image analysis framework becomes available. At the moment, many research groups work on the development of their own pipelines for processing data. However, as more algorithms are likely to be designed over the next decade than over the previous decade, it will become increasingly challenging for non-experts to choose a suitable algorithm for each experimental condition and imaging modality. A standardized framework should be offered for each camera or imaging configuration, allowing an easy switch between various algorithms for segmentation, localization and image analysis. This would minimize the amount of time spent by non-experts to determine suitable data processing methods.

8.3.2. Five to ten years

In this section speculations are made on what kind of developments are expected in five to ten years.

From dead to live cell single molecule localization microscopy

Many super-resolution imaging experiments are currently performed on lifeless samples. These experiments are relevant because they lead to biological understanding of the structure and organization of cells. Single molecule experiments performed on live cells would enable far deeper insight, where besides visualizing biological structures at a high resolution they can show how structures are formed and what their function is.

One of the major drawbacks of single molecule localization experiments is the toxicity of the experimental techniques to live cells. Examples are the toxicity from the buffer and the high illumination intensities typically used. It is possible to acquire an FRC resolution of 50 nm in 10 s at 800 fps on Alexa Fluor 647-immunolabeled microtubules using the current techniques, however, the sample has to be exposed to 62 kW cm^{-2} light intensity [18]. Exposures with these intensities are in all likelihood toxic, but satisfactory cell internal measures of toxicity are not yet developed. A way to reduce the photo-toxicity is to reduce the cell volume that is exposed to these laser powers, which can be accomplished using selective plane illumination microscopy or light sheet microscopy [19–22]. Major obstacles that need to be addressed to increase the imaging speed to a frame rate relevant to most live cells (> 50 Hz) include: i) a significant increase in single molecule density at which raw single molecule images are processed, since currently at least 10,000 frames are needed; ii) imaging at a higher camera speed, as the current limit is 3,200 fps with sCMOS - restricting the imaging to around 0.3 Hz; and iii) dyes that have a significantly higher quantum efficiency and photo-stability.

In the next years, we will witness a big push for processing of data containing a higher single molecule density through the development of new algorithms. One of the possibilities to achieve this is by incorporating apriori structural information, which allows for the estimation of whole structures that are composed of many molecules. Another approach is to combine aspects from single molecule tracking with those of single molecule microscopy, like single particle tracking PALM [23].

Super-resolution sensing

Single molecule localization microscopy has other benefits over techniques such as SIM, apart from the higher attainable resolution. This includes the ability to

encode more information into the PSF of the single molecule. The wavelength of the emitter can be encoded into the PSF so that various proteins with different emission wavelengths are captured at the same time on a single camera, and biological experiments could be multiplexed. Alternatively, a single fluorescent protein could be measured that reports on the surrounding pH, glucose content, and protein maturation of the protein by use of the change in the emission spectrum [24–26]. One could even engineer a protein that reports on its diffusion speed. Other properties that can be extracted from the PSF are fluorescence lifetime, single molecule switching rate, and emission dipole orientation. These properties could also be used to report on the local environment within the cell.

Limits of localization precision

The localization precision is expected to increase in the next decade due to improved fluorescent proteins and dyes and increased sensitivity of the fluorescence microscope. A wider variety of fluorescent proteins and dyes will become available with higher quantum yield and stability. Furthermore, the sensitivity of the fluorescence microscope will increase due to an improved photon yield of the optical system together with further maximization of the quantum efficiency and minimization of the readout noise of the detectors. The consequence is that the tag length of the proteins will need to decrease to prevent that the length becomes a limiting factor. A decreased tag length could limit the ability of the fluorescent label to move freely, thereby causing vectorial effects in the PSF to get visible while imaging, as the field emitted by the fluorescent protein will be polarized. Therefore, for high accuracy, single molecule localization routines should be able to take these effects into account. Furthermore, a higher photon budget would lead to the execution of experiments in thicker samples that were previously only possible in thin samples. Imaging in thicker samples is essential for three-dimensional microscopy to obtain a full understanding of mRNA regulation since these processes are occurring in three dimensions within the cell.

Besides smaller tags, also the variety of tags should increase to allow for various colors to be used in a single experiment with high specificity. A way to increase the efficiency and stability of fluorescent proteins is to image them at cryogenic temperatures [27], which has been demonstrated recently for single molecule localization microscopy with beneficial results [28]. Cryogenic temperatures also narrow the excitation and emission band, which increases the number of fluorescent labels that one can use without spectral overlap. Ideally, improvements made to label qualities would lead to the development of a tag that is non-fluorescent when unbound, and delivery of this fluorescent label to the model system, plus

exposure to the light intensity used for imaging, will be negligible and known. This would ensure reliable experiments and help separate fact from artifact.

References

[1] K. V. Morris and J. S. Mattick, *The rise of regulatory RNA,* Nature reviews. Genetics **15**, 423 (2014).

[2] D. Hall and J. Llinas, *Multisensor data fusion* (CRC press, 2001).

[3] H. L. Van Trees, *Detection, estimation, and modulation theory* (John Wiley & Sons, 2004).

[4] Y. Lu, W. Wang, and M. W. Kirschner, *Specificity of the anaphase-promoting complex: A single-molecule study,* Science **348**, 1248737 (2015).

[5] C. Yao, H. M. Sasaki, T. Ueda, Y. Tomari, and H. Tadakuma, *Single-molecule analysis of the target cleavage reaction by the drosophila RNAi enzyme complex,* Molecular Cell **59**, 125 (2015).

[6] K. M. Herbert, S. K. Sarkar, M. Mills, H. C. D. De la Herran, K. C. Neuman, and J. A. Steitz, *A heterotrimer model of the complete microprocessor complex revealed by single-molecule subunit counting,* RNA **22**, 175 (2016).

[7] M. L. Rodgers, J. Paulson, and A. A. Hoskins, *Rapid isolation and single-molecule analysis of ribonucleoproteins from cell lysate by SNAP-SiMPull,* RNA (2015).

[8] A. B. Loveland, S. Habuchi, J. C. Walter, and A. M. van Oijen, *A general approach to break the concentration barrier in single-molecule imaging,* Nature Methods **9**, 987 (2012).

[9] Y. V. Fu, H. Yardimci, D. T. Long, T. V. Ho, A. Guainazzi, V. P. Bermudez, J. Hurwitz, A. van Oijen, O. D. Schärer, and J. C. Walter, *Selective bypass of a lagging strand roadblock by the eukaryotic replicative dna helicase,* Cell **146**, 931 (2011).

[10] D. R. Semlow, M. R. Blanco, N. G. Walter, and J. P. Staley, *Spliceosomal DEAH-box ATPases remodel pre-mRNA to activate alternative splice sites,* Cell **164**, 985 (2016).

[11] M. R. Blanco, J. S. Martin, M. L. Kahlscheuer, R. Krishnan, J. Abelson, A. Laederach, and N. G. Walter, *Single molecule cluster analysis dissects splicing pathway conformational dynamics,* Nature Methods (2015).

[12] P. J. Bosch, J. S. Kanger, and V. Subramaniam, *Classification of dynamical diffusion states in single molecule tracking microscopy,* Biophysical Journal **107**, 588 (2014).

[13] D. Grünwald and R. H. Singer, *In vivo imaging of labelled endogenous β-actin mRNA during nucleocytoplasmic transport,* Nature **467**, 604 (2010).

[14] S. T. Low-Nam, K. A. Lidke, P. J. Cutler, R. C. Roovers, P. M. v. B. en Henegouwen, B. S. Wilson, and D. S. Lidke, *ErbB1 dimerization is promoted by domain co-confinement and stabilized by ligand binding,* Nature Structural and Molecular Biology **18**, 1244 (2011).

[15] W. E. Salomon, S. M. Jolly, M. J. Moore, P. D. Zamore, and V. Serebrov, *Single-molecule imaging reveals that argonaute reshapes the binding properties of its nucleic acid guides,* Cell **162**, 84 (2015).

[16] N. Monnier, Z. Barry, H. Y. Park, K.-C. Su, Z. Katz, B. P. English, A. Dey, K. Pan, I. M. Cheeseman, R. H. Singer, *et al.*, *Inferring transient particle transport dynamics in live cells,* Nature Methods **12**, 838 (2015).

[17] F. Persson, M. Lindén, C. Unoson, and J. Elf, *Extracting intracellular diffusive states and transition rates from single-molecule tracking data,* Nature Methods **10**, 265 (2013).

[18] Y. Lin, J. J. Long, F. Huang, W. C. Duim, S. Kirschbaum, Y. Zhang, L. K. Schroeder, A. A. Rebane, M. G. M. Velasco, A. Virrueta, *et al.*, *Quantifying and optimizing single-molecule switching nanoscopy at high speeds,* PloS One **10**, e0128135 (2015).

[19] Y. S. Hu, M. Zimmerley, Y. Li, R. Watters, and H. Cang, *Single-molecule super-resolution light-sheet microscopy,* ChemPhysChem **15**, 577 (2014).

[20] F. C. Zanacchi, Z. Lavagnino, M. P. Donnorso, A. Del Bue, L. Furia, M. Faretta, and A. Diaspro, *Live-cell 3D super-resolution imaging in thick biological samples,* Nature Methods **8**, 1047 (2011).

[21] M. Palayret, H. Armes, S. Basu, A. T. Watson, A. Herbert, D. Lando, T. J. Etheridge, U. Endesfelder, M. Heilemann, E. Laue, *et al.*, *Virtual-"light-sheet" single-molecule localisation microscopy enables quantitative optical sectioning for super-resolution imaging,* PloS One **10**, e0125438 (2015).

[22] J. G. Ritter, R. Veith, A. Veenendaal, J. P. Siebrasse, and U. Kubitscheck, *Light sheet microscopy for single molecule tracking in living tissue,* PloS One **5**, e11639 (2010).

[23] S. Manley, J. M. Gillette, G. H. Patterson, H. Shroff, H. F. Hess, E. Betzig, and J. Lippincott-Schwartz, *High-density mapping of single-molecule trajectories with photoactivated localization microscopy,* Nature Methods **5**, 155 (2008).

[24] F. V. Subach, O. M. Subach, I. S. Gundorov, K. S. Morozova, K. D. Piatkevich, A. M. Cuervo, and V. V. Verkhusha, *Monomeric fluorescent timers that change color from blue to red report on cellular trafficking,* Nature Chemical Biology **5**, 118 (2009).

[25] D. M. Chudakov, S. Lukyanov, and K. A. Lukyanov, *Fluorescent proteins as a toolkit for in vivo imaging,* Trends in Biotechnology **23**, 605 (2005).

[26] M. Sauer, *Single-molecule-sensitive fluorescent sensors based on photoinduced intramolecular charge transfer,* Angewandte Chemie International Edition **42**, 1790 (2003).

[27] R. Kaufmann, C. Hagen, and K. Grünewald, *Fluorescence cryo-microscopy: Current challenges and prospects,* Current Opinion in Chemical Biology **20**, 86 (2014).

[28] W. Li, S. C. Stein, I. Gregor, and J. Enderlein, *Ultra-stable and versatile widefield cryo-fluorescence microscope for single-molecule localization with sub-nanometer accuracy,* Optics Express **23**, 3770 (2015).

8

List of Publications

Journal Publications - peer reviewed

8. A. Noma[†], **C. S. Smith**[†], M. More, D. Grunwald *Single molecule RNA FISH in Drosophila larval muscle reveals location dependent mRNA composition of megaR-NPs* , Journal of Cell Biology, In preparation, (2016)

7. **C. S. Smith**, M. Huisman, M. Siemons, D. Grunwald, S. Stallinga *Simultaneous measurement of emission color and 3D position of single molecules* , Optics Express, vol. 24, no. 5, pp. 4996-5013, (2016)

6. **C. S. Smith**, S. Stallinga, K. A. Lidke, B. Rieger, D. Grunwald, *Probability-based particle detection that enables threshold-free and robust in vivo single-molecule tracking,* Molecular Biology of the Cell, vol. 26, no. 22, pp. 4057-4062, (2015)

5. **C. S. Smith**[†], A. Lari[†], C. Derrer[†], A. Ouwehand, A. Rossouw, M. Huisman, T. Dange, M. Hopman, A. Joseph, K. Weis, D. Grunwald, B. Montpetit, *In vivo single particle imaging of nuclear mRNA export in yeast,* Journal of Cell Biology, vol. 211, no. 6, pp. 1121-1130, (2015)

4. **C. S. Smith**[†], S. Preibisch[†], A. Joseph, S. Abrahamsson, B. Rieger, E. Myers, R. H. Singer, D. Grunwald *Nuclear accessibility of β-actin mRNA measured by 3D single-molecule real time (3D-SMRT) microscopy,* Journal of Cell Biology, vol. 209, no. 4, pp. 609-619, (2015)

3. A. Haber, A. Polo, **C. S. Smith**, S. F. Pereira, P. Urbach, and M. Verhaegen *Iterative learning control of a membrane deformable mirror for optimal wavefront correction,* Applied Optics, vol. 52, no. 11, pp. 2363-2373, (2013)

2. **C. S. Smith**, R. Marinica, A. J. den Dekker, M. Verhaegen, V. Korkiakoski, C. U. Keller, and N. Doelman, *Iterative linear focal-plane wavefront correction,* JOSA A, vol. 30, no. 10, pp. 2002-2011, (2013)

1. **C. S. Smith**, N. Joseph, B. Rieger & K. A. Lidke *Fast, single-molecule localization that achieves theoretically minimum uncertainty,* Nature Methods, vol. 7, no. 5, pp. 373 - 375 (2010)

[†]These authors contributed equally to this work

Conference Publications - peer reviewed

2. R. Marinica, **C. S. Smith**, M. Verhaegen, *State feedback control with quadratic output for wavefront correction in adaptive optics,* Annual Conference on Decision and Control (CDC), IEEE, vol. 52, pp. 3475-3480, (2013)

1. **C. S. Smith**, R. Marinica, J. Antonello, A. J. den Dekker & M. Verhaegen (2012, October). *Focal-plane wavefront estimation and control using the extended Kalman filter.* International Symposium on Optomechatronic Technologies (ISOT), IEEE, vol. 1, no. 6, (2012)

Conference Publications

3. **C. S. Smith**, R. Andrei, J. Antonello, A. J. den Dekker, M. Verhaegen , *Focal-plane wavefront estimation and control using the extended Kalman filter,* Adaptive Optics Systems, SPIE, vol. 3, (2014)

2. **C. S. Smith**, A. J. den Dekker, R. Andrei, R. Fraanje, M. Verhaegen , *Fast phase diversity wavefront sensing using object independent metrics,* Adaptive Optics Systems, SPIE, vol. 3, (2012)

1. R. Andrei, **C. S. Smith**, R. Fraanje, M. Verhaegen, V. A. Korkiakoski ; C. U. Keller ; N. Doelman*Linear analytical solution to the phase diversity problem for extended objects based on the Born approximation,* Adaptive Optics Systems, SPIE, vol. 3, (2012)

Editorial

1. **C. S. Smith**, T. Li-Chun, D. Grunwald, *A 4D view on mRNA* Oncotarget. vol. 6, no. 30, pp. 28515–28516, (2015)

8

Summary

Single molecule localization microscopy is a powerful tool that delivers high contrast and imaging specificity at a resolution beyond that of conventional microscopes. To obtain a super-resolved image, one needs to image at least hundred thousand camera frames and estimate the position of millions of molecules with nanometer precision. The tremendous amount of data that needs to be analyzed is one of the challenges scientists face when applying single molecule localization techniques. For this reason, a maximum likelihood estimation method is proposed in this thesis that attains the Cramer-Rao Lower Bound and estimates the position of single molecules in parallel on a GPU to achieve near real-time processing with high precision.

A major drawback of current methods that detect single molecules is that the number of false positives is unknown. Therefore, a generalized likelihood ratio test is proposed here, which can control both the number of false positives and true positives with minimal user input. This target is stably achieved in the experiment over a large range of signal to background conditions.

A key application of single molecule localization microscopy can be found in *in vivo* RNA imaging. In this thesis two RNA studies are presented: i) The nuclear pore complex structures and the kinetic interaction with mRNA are investigated, where a point mutation of mex67p triggered the nuclear export efficiency to drop significantly; and ii) A study on the mobility and occupation of mRNA within the nucleus is performed by instantaneously imaging the whole three-dimensional volume of mRNAs in the nucleus of a living cell. From this study, no evidence was obtained for exclusion or enrichment of the heterochromatin by mRNAs.

For the general applicability of single molecule localization microscopy, the two-dimensional methods will need to be extended to three-dimensions. In three-dimensional imaging small aberrations will become significant when imaging away from focus and therefore need to be compensated or calibrated. This thesis shows that one can extend single molecule localization into three-dimensions, or even a fourth-dimension such as color when the point spread function of a microscope is correctly distorted. Additionally, an adaptive optics strategy

is presented that can correct for sample induced aberrations, and a method is proposed to encode emission color of the single molecule into the measurement.

Samenvatting

Single molecule lokalisatiemicroscopie is een superresolutiemethode die een hoog contrast en specificiteit in beeldvorming bewerkstelligt. Deze methode gebruikt meer dan honderdduizend camerabeelden waarin de positie van miljoenen moleculen met nanometer precisie geschat worden om een beeld met superresolutie te genereren. Een significant obstakel voor wetenschappers om te overbruggen wanneer ze single molecule lokalisatie technieken toepassen is de enorme hoeveelheid data die geanalyseerd moet worden. In dit proefschrift wordt een meest aannemelijke schattingsmethode gepresenteerd die de Cramer-Rao ondergrens bereikt en de positie van de afzonderlijke moleculen in parallel schat op een GPU met hoge precisie. De dataverwerking is met behulp van deze techniek bijna instantaan.

Een probleem van hedendaagse methoden voor single molecule detectie is dat de hoeveelheid valse detecties niet bekend is. Daarom wordt hier een generale meest aannemelijke ratio test gepresenteerd die de hoeveelheid valse detecties en echte detecties kan beheersen met minimale interventie van de gebruiker. Dit doel wordt experimenteel behaald voor verschillende signaal-tot-achtergrond verhoudingen.

Een belangrijke applicatie voor single molecule lokalisatiemicroscopie is de beeldvorming van RNA. In dit proefschrift worden twee RNA studies gepresenteerd: i) Een onderzoek naar de structuren van de nucleaire porie en zijn interactie met mRNA, waar een puntmutatie in mex67p een significante afname in de exportefficiëntie veroorzaakt; en ii) een onderzoek naar de mobiliteit en locatie van mRNA in de celkern door middel van het instantaan vastleggen van het volledig driedimensionaal volume van mRNAs in de celkern van een levende cel. Geen bewijs werd hier gevonden voor uitsluiting of verrijking van het heterochromatine door mRNAs.

De tweedimensionale methoden moeten worden uitgebreid naar drie dimensies om wijdverspreide toepassing van single molecule lokalisatiemicroscopie mogelijk te maken. Kleine aberraties zijn al snel significant in driedimensionale beeldvorming wanneer de afbeelding buiten focus vastgelegd wordt, en daarom is compensatie of kalibratie van de aberraties essentieel. Dit proefschrift toont

aan dat single molecule lokalisatie uitgebreid kan worden naar drie dimensies, en zelfs naar een vierde dimensie zoals kleur, wanneer de puntspreidingsfunctie van een microscoop correct verstoord wordt. Daarnaast wordt hier een strategie in adaptieve optica gepresenteerd die aberraties kan corrigeren die veroorzaakt worden door het preparaat. Vervolgens wordt een methode besproken waarmee de emissiekleur van de afzonderlijke moleculen versleuteld kan worden in de meeting.

Acknowledgments

First of all, I would like to express my sincere gratitude to my mentor Prof. David Grunwald for his support during my Ph.D., and for the endless patience and motivation he showed me. His guidance contributed significantly to my research, my ability to set up fruitful collaborations, and of course the writing of this thesis. I am very grateful for having benefitted from him as my advisor and mentor, and having learned from him how to be more diplomatic when dealing with the politics of science. He is truly a supervisor like no other I have ever worked with, and I mean this in the most positive way imaginable.

Beside my mentor in the United States, I would like to express my sincere gratitude to my mentors Dr. Bernd Rieger and Dr. Sjoerd Stallinga, at Delft University of Technology. It was their understanding of my adventurous nature that led to the opportunity for this Ph.D. thesis. In particular, I would like to thank them for their insightful comments and encouragement along the way, but also, the difficult questions they always dared to ask that either focused my research or broadened the scope of it, enabling me to perceive my research from other perspectives.

Throughout my career at the Delft University of Technology, during numerous courses that were part of my BSc. and MSc. programs, Prof. Lucas van Vliet interested me for the imaging sciences and without him, I would not have had the opportunity to be put forth as a Ph.D. candidate at the Delft University of Technology. I am forever grateful that he made this possible for me, and that he always believed in me.

Furthermore, my sincere thanks go to Prof. Keith Lidke, who provided me an opportunity to join his team at UNM as a BSc. Thesis intern. My time at UNM kindled my love for microscopy but also my love for the USA, which changed my life for the better and, in the end, led me to do my Ph.D. at UMass.

Most importantly and above all, I would like to thank my wife, Kirstine Schiebel, for being my best friend and supporting me spiritually and emotionally throughout my academic career with her constant love and guidance. Thank you for

being my proofreader, editor, sounding board, and world travel buddy. I owe you everything.

Last but not least, I would like to thank my fellow lab mates that were always there for inspiring discussions. I sincerely appreciate all the fun we have had during the last years, the camping trips in New England, the cultural differences that we had heated deliberations about but always knew how to overcome, and the endless 3D printing of handy and not-so-handy knickknacks. My appreciation also goes out to my friends in the Quantitative Imaging group at TU-Delft, whom all made me feel welcome every time I visited.

8

Acknowledgments

Curriculum Vitæ

Carlas Sierd Smith, born in The Hague, The Netherlands on 10 January 1986, graduated in 2005 with a high school diploma from Scholengemeenschap Dalton Voorburg after which he enrolled for the BSc. program in Applied Physics at Delft University of Technology. In 2008 he obtained his BSc. degree after completion of his thesis on the topic of Single Molecule Localization, which he carried out at the University of New Mexico, Albuquerque, USA, in the Keith Lidke lab.

In February 2009 he started a dual degree MSc. program in Applied Physics and Aerospace Engineering at Delft University. For his MSc. research project he worked on Adaptive Optics at Delft Center for Systems and Control under Prof. Michel Verhaegen. After obtaining both MSc. degrees in 2012, Carlas started as a Research Associate in Prof. Michel Verhaegen's lab for a period of one year. Subsequently, he moved to Boston, Massachusetts, USA in 2013 to join Prof. David Grunwald's lab at the RNA Therapeutics Institute (RTI) of the University of Massachusetts. Here he worked in collaboration with Delft University's Department of Imaging Physics on the Single Molecule Localization Microscopy research described in this thesis.