Dissertation
submitted to the
Combined Faculties for the Natural Sciences and for Mathematics
of the Ruperto-Carola University of Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

Put forward by
Diplom-Physiker Claas von Middendorff
Born in Gifhorn
Oral examination: Oct 8, 2008

Experimental Stochastics in High-Resolution Fluorescence Microscopy

- Imaging Theory of PALMIRA Microscopy -

- Improved Models for FCS -

Referees:
Prof. Dr. Stefan W. Hell
Prof. Dr. Dr. Christoph Cremer

Abstract. This thesis presents a statistical imaging theory for photo-activation localization microscopy with independently running acquisition (PALMIRA). In this type of sub-resolution microscopy the switching of the fluorescence capability of macromolecules reduces imaging to the high-precision localization of individual fluorescent molecules. The point-spread function and the imaging equation of a PALMIRA imaging system are calculated and stochastic expressions for the measurement time and the confidence level of the image as a function of the spatial resolution are provided. Different localization schemes like astigmatic imaging, multi-channel defocus imaging, 4pi imaging and a multi-point setup using photo-diodes are analyzed. The theory for multi-color and polarization-resolved measurements is addressed and estimators for data evaluation procedures are provided. The role of background noise in producing artefacts is studied. Finally, it is assessed whether the quality of images can be augmented by a suitable deconvolution procedure.

Furthermore, stochastic methods are applied to solve a couple of persistent problems in fluorescence correlation spectroscopy (FCS). The development of computational methods for the simulation of FCS experiments necessitates the analytical description of the architecture of a multiple-lag-time correlator that is used to estimate autocorrelations from intensity time traces. Recently, FCS has been combined with stimulated emission depletion (STED) focal volumes. The general phenomenology of STED-FCS correlation curves is studied as a function of the STED beam intensity. It is shown that the quality of a measurement is mainly determined by the fraction of signal originating from the focal plane. Then, an improved fit model taking into account the exact spatial dependency of intersystem crossing rates is presented and tested on synthetic data. Finally, the influence of second-order correlations among the points of the FCS curve on the determination of fit parameters is studied.

Analytical results are provided wherever possible. Otherwise, Monte-Carlo computations are performed.

Zusammenfassung. In dieser Arbeit wird eine statistische Theorie zur Bildentstehung in der PALMIRA-Mikroskopie vorgestellt. Die Schaltbarkeit der Fluoreszenz passender Makromoleküle hat den Effekt, mikroskopische Bildentstehung auf das Lokalisieren von Einzelmolekülen zurückführen zu können. Dabei kann die Abbesche Beugungsgrenze überwunden werden. Die Punktabbildungsfunktion sowie die Bildgebungsgleichung werden berechnet. Zudem werden stochastische Ausdrücke für die Aufnahmezeit und die Konfidenz des Bildes als Funktion der räumlichen Auflösung hergeleitet. Verschiedene Methoden zur dreidimensionalen Positionsbestimmung von Fluorophoren - astigmatische, defokussierte und 4pi-Abbildung - und ein zweidimenionaler Mehrpunktaufbau aus Photodioden werden analysiert. Die Theorie für Mehrfarben-Messungen und für polarisationsabhängige Mikroskopie wird entwickelt und die Rolle von Hintergrundrauschen bei der Objekterkennung wird diskutiert. Eine Entfaltungstrategie wird vorgestellt.

Weiterhin werden stochastische Methoden eingesetzt, um Probleme auf dem Gebiet der Fluoreszenz-Korrelations-Spektroskopie zu behandeln. Die Entwicklung von numerischen Verfahren zur Simulation von FCS-Experimenten erfordert es, die Theorie eines multi-lag-time Korrelators analytisch zu formulieren. FCS wurde kürzlich mit den grössenreduzierten Fokalvolumina der STED-Mikroskopie kombiniert. Die Phänomenologie von STED-FCS Korrelationskurven wird in Abhängigkeit von der STED-Intensität studiert. Hierbei zeigt sich, dass die Qualität der Messung entscheidend durch die Grösse des Signals aus der Fokalebene bestimmt ist. Weiterhin wird ein verbessertes Fitmodell für Übergangsraten in die Triplet-Mannigfaltigkeit, das den exakten geometrischen Fokalverlauf berücksichtigt, vorgeschlagen und an synthetischen Daten getestet. Schliesslich wird untersucht, inwieweit die nicht-diagonale Kovarianz von FCS-Korrelationskurven das Fitergebnis von Standardverfahren verfälscht und ob ein detaillierteres Fitverfahren hier Abhilfe schaffen kann.

Wo immer möglich werden analytische Resultate hergeleitet, andernfalls empirische Monte-Carlo-Rechnungen verwendet.

Contents

Introduction

Statistical methods have gained a constantly increasing importance in the quantitative sciences [1, 2]. Both exact analytical results from probability theory and stochastics, as well as numerical computation abilities, have found widespread applications.

High-resolution fluorescence microscopy techniques overcoming Abbe's diffraction barrier [3] are no exception to this rule. The statistical aspects are more decisive the smaller the size of the ensemble of recorded fluorophores becomes because ensemble averaging is gradually eliminated [4]. The extreme case consists of just a single isolated fluorescing molecule. A convenient way to activate single molecules out of dense ensembles of large size is the use of fluorophores with a fluorescence transition that can be switched on and off by light. Originally, this type of molecules has been introduced into fluorescence imaging in the realm of stimulated emission depletion microscopy (STED) [5, 6] while recently it has turned out that the switching of fluorophores permits to reduce imaging to localization of individual molecules. A new type of sub-diffraction microscopy that sequentially reads out stochastic sequences of single molecules has been developed over the last two years [7]. Thereby, the image is built up by a histogram of individual fluorophore positions that are retrieved with sub-diffraction accuracy from non-overlapping images of single molecules that are activated at constant rate at random positions. The method has been denoted as photo-activation localization microscopy with independently running acquisition (PALMIRA). It involves a multitude of inherently stochastic variables: the random positions of activated molecules, delivered photon numbers, detected numbers of photo-electrons, recorded background noise levels and the total number of times a molecule might be switched before it finally bleaches irreversibly. Also, the creation of the image as a histogram of absolute frequencies from a limited number of statistical samples is genuinely stochastic. An imaging theory for PALMIRA microscopy should encompass an imaging equation, a point-spread-function (PSF), a resolution measure, some measure of the degree of confidence of the image obtained after a finite measurement time and a quantitative treatment of generalizations like multi-color imaging. In the first part of this thesis, such a statistical imaging theory for PALMIRA microscopy is presented. This theory has provided foundations for a number of successful experiments [7, 8, 9, 10].

In PALMIRA microscopy the molecule under consideration is assumed to be fixed over at least a couple of successive image frames. This assumption is abandoned in fluorescence fluctuation techniques like fluorescence correlation spectroscopy (FCS). There, the fluorescence time trace originating from molecules diffusing through an excitation spot and possibly undergoing photo-chemical transformations is subjected to an autocorrelation analysis. Fitting the resulting correlation curve yields the time constants of these processes and geomet-

rical parameters like the size of the fluorescing ensemble. FCS is a well-established method [11]. However, some new developments necessitate further theoretical treatments. Firstly, it has almost always been assumed that the focal volume might be modeled by a Gaussian shape. The Gaussian model implies certain interpretations of the correlation curve of FCS like the one that the amplitude is inversely proportional to the size of the volume of excited molecules. Here, it is assessed to which degree this interpretation can be maintained for FCS measurements using the differently shaped STED focal volume [12, 13]. Secondly, in many data interpretations the spatial variation of photo-chemical transformation rates has been assumed to be constant. This is not the case in reality where, for example, the inter-system crossing rate between the fluorescent singlet system and the non-fluorescent triplet state of a molecule is strongly dependent on the shape of the excitation beam [14]. An optimized fit model is needed. Thirdly, fitting the FCS correlation curve is prevalently done by using the Gaussian least-squares procedure. This assumes that different points of the correlation curve are statistically independent. The assumption is questionable in real applications since the correlation curve at different lag time values is estimated from the same data trace. A better noise model taking into account the coupling between different lag times leads to a higher quality of the fit results. The second part of this thesis is devoted to the study of these topics.

In the following the thematic background is presented in more detail and the content of the thesis is outlined briefly.

PART I. The wave-nature of light leads to diffraction. For decades this has been viewed as an obstacle to the ability to discern different emitters. In a classical work Abbe calculated the resolution barrier, the lowest distance at which particles could be discerned in a lens-based instrument [3]. Following Abbe the minimum lateral distance r_{min} at which two objects might still be discerned by an imaging system reads

$$r_{\text{min}} = \frac{\lambda}{2n \sin \alpha}. \tag{0.1}$$

The value λ denotes the wavelength of the light, n the refractive index and α the half-aperture angle of the system. Since Abbe's work, many efforts have been undertaken to lower the limit r_{min}. These methods can be separated into two classes. The first class which will be termed *conservative* in the following tries to achieve the aim by optimizing the values of λ, n and α. Typical representatives are x-ray [15] and electron microscopy [16]. The former uses electromagnetic radiation of a wavelength $\lambda \approx 1..5$ nm, the latter electron matter waves with a de-Broglie wavelength of $\lambda \approx 10^{-3}$ nm. However, these methods are problematic in biological imaging since they require fixation procedures that a living cell could not survive. 4pi-microscopy making use of two opposed lenses augments the aperture angle α and thereby mainly increases axial resolution [17]. Of course, the aperture cannot exceed $\sin \alpha = 1$. Furthermore, the simple method of employing the embedding medium with the highest refractive index n should be mentioned. In contrast to these conservative methods are the methods called *progressive* in the following. These try to create experimental schemes that add further parameters to Equation (0.1). The first far-field method of this class has been STED microscopy [5, 18, 19]. In this technique a fluorescence microscope [20] is equipped

with a second, red-shifted laser beam that features a zero intensity node in the center of the beam. This is achieved by a characteristic phase retardation using a phase mask [21]. As a result, the fluorescence volume is depleted by stimulated emission at the outer regions of the doughnut around the zero value. Such a reduced-size optical 'sensor' is then scanned across the sample. Although the STED beam is diffraction-limited the resulting focal volume may be much smaller in width than the classical r_{min} due to the high nonlinearity in the rate equation kinetics of stimulated emission [22]. The parameter governing the resolution is the ratio $\zeta = I_{\text{STED}}/I_{\text{sat}}$ of the STED beam intensity I_{STED} and the saturation intensity I_{sat} that describes the strength of stimulated emission. The resulting resolution equation is then of the following form [23]

$$r_{\text{min}} = \frac{\lambda}{2n \sin \alpha} \frac{1}{\sqrt{1+\zeta}}. \tag{0.2}$$

It is seen that by raising the STED intensity, arbitrarily low values of r_{min} can be reached. Typical values can be as low as $r_{\text{min}} \approx 20$ nm [24]. The maximum achievable resolution depends crucially on the photo-fatigue properties of the dye with respect to the STED beam [25]. It should be mentioned that STED is just one realization of a wider category of reversible saturable optical fluorescence transitions (RESOLFT) methods [4] that rely on a dual-state system with a detectable state A and an inactive state B, for instance a complex macromolecule featuring a switching transition [6] or a ground-state and a triplet state [26].

The switching of molecules also lies at the heart of a second class of progressive sub-diffraction methods. Thereby, a sample, possibly densely packed with switchable fluorophores, the conventional images of which would overlap if they were activated all at the same time, is stochastically activated at only a few positions that are separated by a distance exceeding the classical r_{min}. Then, the molecules are localized and their center position is stored. This process is repeated until enough position estimates have been calculated to estimate the true molecule distribution from these statistical samples. The main physical insight is that molecular switching can be used for a temporal un-mixing of dense samples. In contrast to STED, where the number of photons is stochastic and the scanning deterministic, PALMIRA features both a stochastic photon number and a stochastic molecular 'sensor' position.

The fact that sub-diffraction accuracy of position estimates can be achieved by determining the center-of-mass of single-molecule images containing many photons was already noticed by Heisenberg 80 years ago when he calculated the momentum transfer in a quantum-mechanical analysis of the optical measurement process [27]. The task has become technically feasible since the advent of high-quality CCD cameras in the 1980s.

PALMIRA has some predecessors. Temporal un-mixing by irreversible bleaching was proposed in 2004 [28]. However, with bleaching the image frames are of different quality since the number of active molecules decreases with time. The first image frame contains much more active fluorophores than the last. Hence, the first frames have the tendency to be useless for sub-diffraction localization of single molecules. Subtracting later image frames from earlier ones is possible but is strongly degraded by background noise. The original idea of stochastic temporal un-mixing by molecular switching is credited to Eric Betzig who, in 2006, termed the method photo-activation localization microscopy (PALM) [29]. Hereby, every frame contains the same number of active molecules on average. Soon after-

wards, several different groups proposed very similar methods such as fluorescence photoactivation localization microscopy (FPALM) [30] and stochastic optical reconstruction microscopy (STORM) [31]. However, the acquisition time of the experiment can amount to up to 10 h which requires extremely stable samples. The problem has been solved by introducing the concept of asynchronous activation where molecules are not only activated at the beginning of an acquisition frame but constantly with a certain rate [7, 8]. This leads to typical acquisition durations of a few minutes. In PALMIRA microscopy the entry and the exit point of newly activated molecules are stochastic while in PALM and STORM only the exit point is stochastic while the entry point is deterministic. Multi-color STORM [32], multi-color PALM [33] and a generalization to three dimensions [34] have been published in 2007 and 2008, respectively. However, these are experimental considerations and no theoretical foundations are provided. For instance, multi-color data acquisition is done sequentially for each color there and not like it is considered in this thesis in parallel for all colors with the distinguishability of molecules relying on the difference in emission characteristics.

In PALMIRA the photon number effectively follows a geometric distribution whereas the position is a random number drawn from the molecule distribution under consideration. The photon number is the quality index giving the credibility level of each position estimate, much as the STED intensity I_{STED} is the fluorescence ensemble size index in STED microscopy. Mathematically, PALMIRA is equivalent to a decomposition of the underlying molecule distribution into wavelets of stochastic widths and positions [35]. The basic wavelet is the single-molecule diffraction pattern, the Airy-function or its generalizations. The detection mode of PALMIRA is called stochastic readout as opposed to targeted readout of scanning techniques like STED [4]. Stochastic readout displays the property of importance sampling, since it addresses different regions of space directly proportional in weight to the number of molecules residing in the volume in question. In PALMIRA only such events with a photon number exceeding a certain threshold T are accepted. The photon numbers above threshold feature a shifted geometrical distribution. In this thesis the resolution formula for PALMIRA microscopy is derived from such considerations and reads

$$r_{\text{min}} = \frac{\lambda}{2n\sin\alpha}\sqrt{\frac{\phi(\frac{\bar{N}}{\bar{N}+1},1,T)}{\bar{N}+1}}. \tag{0.3}$$

Here, $\bar{N}$ denotes the mean photon number of a molecule, T is the total photon number threshold that each event is subjected to and ϕ denotes a special function, Lerch's ϕ-transcendent, which describes the stochastic spread that still remains in the accepted photon values. The function ϕ plays the same role in PALMIRA microscopy as the Airy function in classical microscopy or the depletion doughnut in STED microscopy. Again, it is seen that a new dependency on the photon number $\bar{N}$ and on the lower quality bound T is added to Equation (0.1). By inspecting ϕ it is found that the resolution can be arbitrarily squeezed down by increasing the threshold T or the photon number $\bar{N}$.

The text of the first part is organized as follows. In Chapter 1 the imaging process of PALMIRA microscopy is formulated mathematically and the important parameters are defined. A proof-of-principle Monte-Carlo simulation for dual-color PALMIRA imaging is presented. Algorithms for data generation and evaluation are described. Chapter 2 is devoted to the calculation of the effective point-spread function and the imaging equation of

PALMIRA microscopy. The resolution equation is derived. Like all sub-diffraction methods developed so far, PALMIRA gains spatial resolution at the cost of temporal stability of the sample. This effect is quantified by calculating an expression for the measurement time as a function of the recognition threshold. Then, it is assessed to what degree an image can be termed credible after a finite measurement time. Chapter 3 studies the quality of different localization algorithms for single molecules with a focus on full three-dimensional position estimation. Different experimental setups that include multichannel defocusing, astigmatic, 4pi and multi-point imaging using photodiodes are investigated. The analysis employs the basic statistical tool of Fisher information as well as empirical simulations. Chapter 4 dwells on the recognition of objects. It is assessed in what way pixel thresholding methods influence the frequency of recognition of objects. The full probability for an event to be recognized in the presence of background noise is derived. Multi-color or polarization-dependent measurements add a new degree of freedom to the analysis. It is shown that three- or more color imaging is possible with only two channels and a very simple but efficient estimator is presented. The accuracy with which the polarization of the emitted light may be determined is investigated. The last Chapter 5 applies Wiener filtering to the task of deconvolving PALMIRA images. In a classic PALMIRA image event positions are tabulated disregarding their photon number. This neglects available information. Events with a higher photon number have a higher level of credibility and should be privileged over events with a lower quality in the final object reconstruction. This is accounted for by deconvolving each event with an individual PSF depending on the photon number.

PART II. Fluorescence correlation spectroscopy is a technique that uses a correlation analysis of the fluorescence time trace to unravel molecular dynamics [11]. These dynamics can stem from different processes like translational and rotational diffusion, conformational changes or photochemical transitions between states of different molecular brightness. The method is closely related to dynamic light scattering methods in solid-state physics [36] and was initially applied to investigate the binding of molecules onto strands of DNA [37]. By combining FCS with confocal microscopy a variety of applications and extensions have been achieved [38].

Monte-Carlo simulations were originally developed in the field of statistical mechanics in order to compute partition sums of spin systems [39, 40] but have gained importance in almost all areas of science where the system is too complex to permit an analytical solution but sufficiently elementary to be transformable into algorithmic form. Monte-Carlo simulations in FCS are well established [41, 42]. The main hurdles here are the correct treatment of boundary conditions, the calculation time and the correlation algorithm. Linear correlation is easily done by Fourier transformation but causes severe memory problems if a large range of lag times is to be covered. A way out of this dilemma is multiple lag time correlation [43, 44] that uses a progressive correlator bin size at higher lag times. Although the principle has been stated in the literature and commercial systems are available a comprehensive mathematical treatment that provides clear formulas for all computed variables is still missing and is therefore provided here.

The combination of FCS with STED focal volumes is conceptually new and has been investigated experimentally for the first time with the fluorophore MR121 and the use of a

circular phase mask which leads to a strong axial resolution increase [12]. Even more recent work focuses on the application of STED-FCS to two-dimensional membranes [13]. There, new insights are expected when the size of the reduced focal volume becomes comparable to the particle size and when it is no longer averaged over nanoscale dynamics like trapping of molecules.

Assuming a Gaussian-shaped focal volume leads to fit artefacts if the real focal volume does not display this shape. This is already the case for a simple diffraction limited focal volume and has been investigated experimentally [45]. A similar effect follows from the assumption of spatially constant inter-system crossing rates. In this case too, wrong assumptions about the spatial shape of the excitation volume lead to systematic errors in the fit results for inter-system crossing [14] and photo-bleaching rates [46]. In practice, it has been tried to correct for this by replacing the excitation profile by a constant profile with smaller height. Here, it is shown that this still causes characteristic errors, especially in the case of pulsed excitation.

Once the question of the fit model has been assessed it is necessary to inquire whether the noise model is correct. The fluctuation of the correlation curve as a function of the molecular brightness and the measurement time has been studied thoroughly [47, 48]. Higher-order covariance on correlation curves has been considered in dynamic light scattering in connection with the relaxation time spectrum [49]. A multivariate fit model has been proposed [50, 51]. However, the concrete situation of FCS with a fluorescence signal being degraded by diffusion, blinking and detection noise has either been neglected or the study did not investigate the quality of fit results and merely described the correlation.

The text of the second part is organized as follows. In Chapter 6 the mathematical and the algorithmic framework is outlined. The correlation curve is expressed as a double-integral over brightness functions times a propagator. The two special cases of a single diffusing particle and a fixed particle undergoing inter-system crossing are treated. The algorithm for a Monte-Carlo simulation of the molecular ensemble of FCS is described. There, it is of importance to construct the right boundary conditions in order to suppress typical artefacts. Then, a mathematical description of multiple-lag-time correlation and its algorithmic implementation is provided. In Chapter 7 the developed computation capabilities are applied to FCS of particles diffusing in three dimensions through a STED focal volume. Thereby, emphasis is placed on the dependence of the focal volume on the STED intensity and on possible artefacts due to non-correlated background for the case of non-ideal STED point-spread-functions. It turns out that artefacts could be minimized by working in two dimensions only. In Chapter 8 a new fit function considering the exact spatial dependency of inter-system crossing rates is presented. The model is tested on Monte-Carlo generated FCS data for the fluorophore Rh110 in continuous-wave and pulsed excitation mode and is shown superior to the standard procedure. In Chapter 9 the assumption of non-correlated noise on autocorrelation curves is relaxed and a generalized multivariate fit is formulated. First the covariance is estimated from an ensemble of correlation curves. Then its effect on fit outcomes is examined using artificial correlation data following exactly the assumed statistics and more realistic data from Monte-Carlo computations. Again the more sophisticated model yields results of higher quality.

METHODOLOGY. All problems treated in this thesis have the same underlying structure: they are estimation problems. This applies equally to whether one searches for the probability of a single-molecule image given a molecule position or whether one aims at understanding the dependency of the correlation curve on the STED intensity. The problem is completely formulated once the conditional probability

$$p\left(\{x\}|\{y\}\right) \tag{0.4}$$

to obtain the data $\{x\}$ provided the parameters $\{y\}$ are known is specified.

(i) Often one is interested in the mean data $\{\hat{x}\}$ resulting from certain parameters. If $p(\{x\}|\{y\})$ is complicated, the transition

$$\{y\} \rightarrow \{\hat{x}\} \tag{0.5}$$

will be performed by a computer simulation. Otherwise expectation values are calculated directly by integration or summation.

(ii) Equally often, some measured data $\{\tilde{x}\}$ is given and the parameters are to be estimated. The transition

$$\{\tilde{x}\} \rightarrow \{y\} \tag{0.6}$$

is often performed as a maximum-likelihood argument.

A large variety of different effects considered in this thesis is describable by basic distributions like the binomial, the multinomial, the Poisson and the Gaussian distribution and associated distributions such as waiting time distributions. Further information can be found in canonical texts on probability theory and statistics [52, 53].

Part I

Stochastic Theory of Photo-Activation Localization Microscopy

1 The Principles

In the subsequent the measurement protocol of PALMIRA microscopy is formulated in mathematical terms and the method is exemplified with a simulation example. At first, qualitatively different steps of the experimental procedure are isolated. Notations are defined. Statistical models corresponding to the experimental steps are identified. Then, the PALMIRA method is cast into algorithmic form and the results of a Monte Carlo simulation for dual-color PALMIRA imaging are presented.

1.1 Description of the Imaging Process

Classical fluorescence microscopy relies basically on a single fluorescent electronic transition, ensemble averaging of the signal over the focal volume and deterministic scanning of the sample space. In contrast, PALMIRA microscopy [29, 30, 31, 7] requires at least two sufficiently independent molecular transitions, works properly on single molecules only and entails stochastic scanning.

The molecular design of the fluorophore used for PALMIRA usually follows a scheme like in Figure (1.1): an optically accessible switching transition (A) enables a fluorescent readout transition (B). For later calculations each transition is characterized with the mean

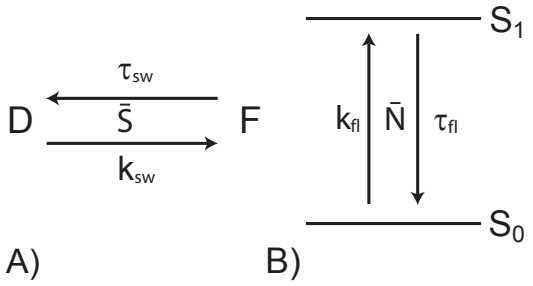

Figure 1.1: The structure of the state-space of PALMIRA fluorophores. The molecule is required to possess a switching transition (A) which is used for scanning and a fluorescence transition (B) which is used for readout. $\bar{S}$ denotes the mean number of switching cycles until irreversible photo-bleaching, $\bar{N}$ the mean number of photons until spontaneous return to the dark state. The former determines the sampling degree of the image, the latter its resolution. The excitation rates k and the lifetimes τ are indicated.

number of associated cycles before its deactivation, the lifetimes and the excitation rates. $\bar{N}$ denotes the mean signal photon number until switching back spontaneously to the non-fluorescent state. $\bar{S}$ is the mean number of switching cycles until the molecule undergoes irreversible photo-bleaching. τ_{sw} is the lifetime of the on-state (F) of the switching transition,

τ_{fl} is the lifetime of the excited fluorescent state S_1. $k_{\text{sw}}(\mathbf{r})$ is the excitation rate of the switching transition and $k_{\text{fl}}(\mathbf{r})$ that of the fluorescence transition. The mean number of photons $\bar{N}$ is given by

$$\bar{N} = \frac{\tau_{\text{sw}}}{\tau_{\text{fl}} + 1/k_{\text{fl}}(\mathbf{r})}, \tag{1.1}$$

i.e. the mean number of photons that can be observed equals the mean number of fluorescence cycles per lifetime of the fluorescent state (F). $\bar{S}$ is given independently.

A typical PALMIRA experiment is composed of a sequence of cycles of which Figure (1.2) sketches the temporal structure. A constantly running activation laser leads to a

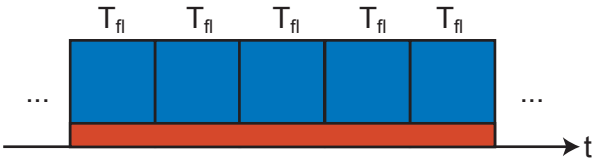

Figure 1.2: Temporal diagram for PALMIRA cycles. New molecules are constantly activated at random positions by an activation laser (red). The fluorescence of the activated molecules is recorded on a number of camera frames (blue). The length of the readout interval is T_{fl}. After having contributed a random number of photons to the image frames the molecules switch back spontaneously or undergo bleaching. The entry and exit points of fluorophores are random.

stochastic activation of the fluorescent state of random molecules. The fluorescence transition is probed over several intervals of length T_{fl} and the resulting image frames are recorded. After having contributed a random number of total photons the molecules switch back to the dark state spontaneously. The entry point and the exit point of new molecules is stochastic in PALMIRA microscopy. It is important to notice that the mean numbers, $\bar{N}$ and $\bar{S}$, do measure the fatigue properties of the transitions, the total lifetimes in units of corresponding cycles. In contrast, the actual occurrence of the switching transition or the fluorescence transition are determined by the excitation rates $k_{\text{sw}}(\mathbf{r})$ and $k_{\text{fl}}(\mathbf{r})$, respectively. These are determined by the laser power, the molecular cross-sections and the pulse length. A simple rate equation analysis yields the following probabilities for activation of either the switching transition, $P_{\text{sw}}(\mathbf{r})$, or the fluorescence transition, $P_{\text{fl}}(\mathbf{r})$, during a laser pulse duration ΔT

$$\begin{aligned} P_{\text{sw}}(\mathbf{r}) &= 1 - \exp\left[-k_{\text{sw}}(\mathbf{r})\Delta T\right] \\ P_{\text{fl}}(\mathbf{r}) &= 1 - \exp\left[-k_{\text{fl}}(\mathbf{r})\Delta T\right]. \end{aligned} \tag{1.2}$$

Uniform excitation facilitates the image reconstruction but is not strictly mandatory. With sufficient knowledge of the functions $k_{\text{sw}}(\mathbf{r})$ and $k_{\text{fl}}(\mathbf{r})$ sampling inhomogenities could be corrected for in an adequate way. It is merely necessary to retain $k_{\text{sw}}(\mathbf{r}) \neq 0$ and $k_{\text{fl}}(\mathbf{r}) \neq 0$ in regions where fluorophores may exist. In order to simplify the analysis these functions are assumed to be made as spatially constant as possible by physical preparation of the experiment, e.g. uniform illumination or scanning schemes. This achieves a homogeneous sampling of the configuration space of molecule positions. Also, the excitation probabilities have to be low enough to make multi-molecule events with overlapping diffraction patterns

sufficiently improbable in order for a single-molecule analysis to be applicable. Image segments with considerable overlap of single-molecule optical patterns will be called higher-order events in the following. Data with non-negligible amount of higher-order events is not worthless as such, but the mixing of sources inevitably leads to a loss of information.

A PALMIRA experiment may be subdivided into the following parts.

(i) Activation of molecules. Whereas the number of activated molecules follows a binomial distribution the distribution of their positions is multinomial. This may be justified as follows. A physical sample is a distribution of M molecules at one, two or three-dimensional positions $\{\mathbf{r}_i\}, i = 1..M$. Since the spatial extent of molecules is far below resolution limits the molecules might be considered as point-like objects leading to the subsequent object fluorophore distribution,

$$o(\mathbf{r}) = \frac{1}{M} \sum_{i=1}^{M} \delta(\mathbf{r} - \mathbf{r}_i). \tag{1.3}$$

The excitation of the switching transition in activation cycle j leads to a stochastic activation of molecules at random positions $\left\{\mathbf{r}_i^j\right\}, i = 1..E_j$ with the number of activated molecules E_j being a random number following the distribution

$$p\left(E_j|M, P_{\text{sw}}\right) = \binom{M_j}{E_j} P_{\text{sw}}^{E_j} P_{\text{sw}}^{M_j - E_j}. \tag{1.4}$$

Here M_j is the number of molecules which could still be switched and have not undergone photo-bleaching, yet. In other words, the stochastic activation sites of cycle j, $\left\{\mathbf{r}_i^j\right\}$, are a random sample without double occurrences of stochastic size E_j out of the set of fixed fluorophore positions $\{\mathbf{r}_i\}$.

(ii) Total number of photons per molecule during the active period. The scheme of Figure (1.1) featuring a binary decision between two alternatives ($D \leftrightarrow F$ or $S_0 \leftrightarrow S_1$) allows the use of the binomial model also for the modeling of the switching and the fluorescence transition. For instance, the photon number N until switching off is a random variable following a geometric distribution

$$p(N) = \frac{\bar{N}^N}{(\bar{N}+1)^{N+1}} = p_{\text{off}}(1 - p_{\text{off}})^N, \tag{1.5}$$

with the mean photon number $\bar{N}$ implying a certain probability

$$p_{\text{off}} = (1 + \bar{N})^{-1} \tag{1.6}$$

of switching off spontaneously. The geometric distribution is simply the waiting time distribution for this process to occur for the first time. From Equation (1.1) the probability to switch back from F to D spontaneously is given by

$$p_{\text{off}} = \frac{\tau_{\text{fl}} + 1/k_{\text{fl}}(\mathbf{r})}{\tau_{\text{fl}} + 1/k_{\text{fl}}(\mathbf{r}) + \tau_{\text{sw}}}. \tag{1.7}$$

Likewise, the dichotomy of the switching process leads to a geometric distribution of the number of switching cycles S a molecule can support until irreversible photo-bleaching.

(iii) Observed photon numbers per frame. Once the total number of photons being emitted by the molecule before it switches off is given it is still not clear which fraction of it is recorded in a single image frame. If the frame time T_{fl} is long all photons will be read-out during one frame whereas if it is short the photons will be distributed among several frames. Moderate frame times are described by an interpolation between these two extreme cases. The photon number N recorded per frame is then distributed as

$$p(N) = \left[\frac{\bar{N}}{\bar{N}+1}\right]^N \left\{ p_{\text{poi}}(N|\bar{N}_{\text{fl}}) + \frac{1}{\bar{N}+1}\left(1 - \sum_{n=0}^{N} p_{\text{poi}}(n|\bar{N}_{\text{fl}})\right)\right\}. \tag{1.8}$$

Here, the Poisson distribution [52] with mean λ is defined by

$$p_{\text{poi}}(n|\lambda) = \frac{1}{n!}\lambda^n e^{-\lambda}. \tag{1.9}$$

A study of Equation (1.8) interpolating between the geometric and the Poisson distribution can be found in the appendix. The main point in the derivation is that after having radiated N photons there can be two processes: either the frame is filled or the frame is not yet filled but the molecule has switched back. The mean photon number per frame $\bar{N}_{\text{fl}}$ for a molecule with an infinitely long-lived state F ($\tau_{\text{sw}} = \infty$) is given by

$$\bar{N}_{\text{fl}} = \frac{T_{\text{fl}}}{\tau_{\text{fl}} + 1/k_{\text{fl}}(\mathbf{r})} \tag{1.10}$$

and equals the mean number of fluorescence cycles that can take place during one image frame.

(iv) Position and molecule type estimation. Localization of activated molecules yields estimates of molecule positions and photon numbers

$$\{\bar{\mathbf{r}}_i^j, N_i^j\}, i = 1..E_j \tag{1.11}$$

These position estimates follow a distribution very similar to a Gaussian centered around the real position and featuring a variance inversely proportional to the photon number. For spectrally resolved multi-channel imaging additionally the molecule type t_i has to be inferred from the distribution of the signal among the channels. Polarization-sensitive measurements add another degree of freedom to the analysis.

Such cycles consisting of activation and readout are repeated either until all molecules have bleached or the required number of position samples has been reached. In the end,the result is a set of stochastic estimates for the position, the photon number and the types of the molecules,

$$\left\{\bar{\mathbf{r}}_i^j, N_i^j, t_i^j\right\}, j = 1..E_i, i = 1..M, \tag{1.12}$$

with i being the deterministic index of the molecule and j the index of the stochastic total number of events E_i sampled from the molecule over all frames. Here, the values of Equation (1.11) have been reordered in such a way that the leading index is the molecule index i. It does not matter in which activation cycle j the molecule showed up. These positions then are subjected to a histogram representation. All steps are studied in more detail in the remaining chapters.

1.2 Algorithms and Forward Monte-Carlo Simulations

In order to perform a numerical simulation of a PALMIRA experiment the different steps of the preceding chapter have to be cast into algorithmic form. Figure (1.3) displays a flow diagram for the image generation algorithm of a single activation cycle for a single molecule. For simplicity it is assumed that activation takes place at the beginning of an imaging frame only. The algorithm comprises the activation step (i), the generation of the total photon number (ii) and the observed photon number per frame (iii) as well as generation of the stochastic image including background noise. Firstly, a random molecule is chosen by drawing a multinomial random number following the molecule distribution under question. This distribution can be artificial, but easy to interpret like just two molecules at a certain distance that may be

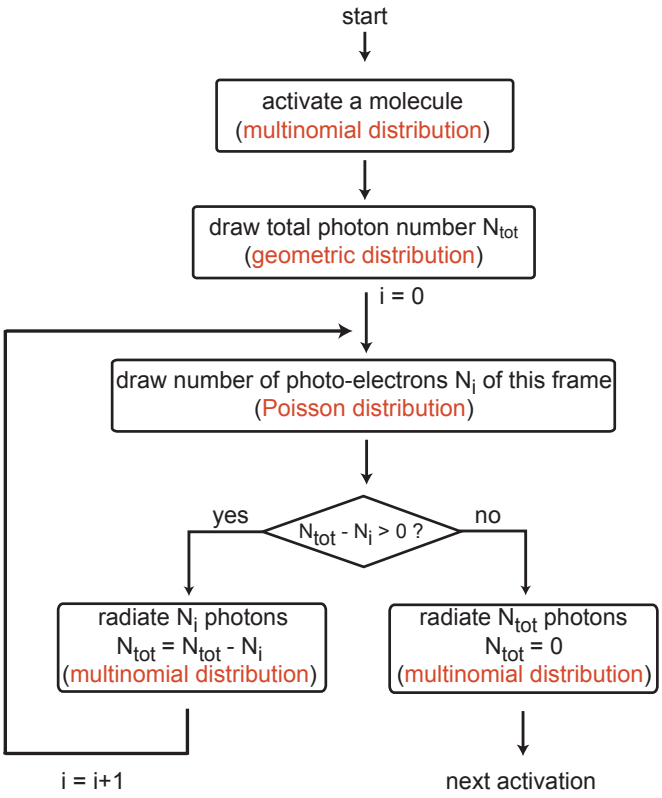

Figure 1.3: Flow diagram for one switching cycle of a single molecule for the Monte-Carlo PALMIRA image generation algorithm. For steps involving the generation of a random number the associated distribution is reported in red. Firstly, a random molecule is activated. Its stochastic position is given by a multinomial random number drawn from the spatial fluorophore distribution. Next, the total photon number N_{tot} that the molecule contributes to the current switching cycle is determined. This is a geometric random number of mean $\bar{N}$. In a third step, these photons are distributed among the camera frames until the next activation takes place. The actual number N_i of detected photo-electrons in a frame is distributed according to a Poisson with mean N_{fl} where N_{fl} is the mean number of photo-electrons detected in a frame for a fluorophore radiating without being deactivated of Equation (1.10). The N_i photon-electrons are distributed multinomially according to the detection PSF until either their sum attains the total photon number N_{tot}. In this case the molecule just radiates the remaining photons. Then the molecule switches off and can be activated again.

well below the optical wavelength. However, the fluorophore distribution can also be more complicated and closer to reality. Next, the total photon number contributed by the molecule during its activation period is determined by sampling a geometric random number. Then, the corresponding image frames are constructed. The candidate photon number radiated by the molecule in the current frame is given by a sample of a Poisson distribution. Now, two different situations can happen. Either the current Poisson sample is smaller than the photon number that can still be delivered by the molecule before switching back spontaneously or it is larger. In the latter case the molecule radiates the rest of its photons and is deactivated. In the former case it radiates a photon number equal to the Poisson candidate sample and the remaining photons are lowered by this value. For multi-color imaging random numbers are drawn to specify the spectral channel into which the photons are radiated. In all cases, the photons are distributed among the image pixels by drawing multinomial random samples from the detection PSF of the molecule. The process is iterated until the radiated photons attain the total photon number. Then the molecule can be activated again. Too high an activation rate may lead to a growing population of molecules in the active state. This is undesirable as far as higher-order events are concerned. In contrast, too low an activation rate leads to many frames staying empty without any molecules activated. Here, the measurement time is raised without need. The optimum condition is in between these two cases.

Once the images have been generated stochastically the molecule positions are estimated from their optical patterns. Figure (1.4) presents the flow diagram of a typical localization algorithm that is applied to each image frame. Firstly, the image is smoothed by a Gaussian filter in order to suppress background noise. The width of the filter is chosen well below the width of the detection PSF. Next, the brightest pixel of the image is searched for. If it does not exceed a threshold the whole image frame is useless and the next image frame is evaluated. If it does exceed the threshold the region around this pixel defines a segment and an algorithm retrieving the pattern center position is started with the brightest pixel as initial value. In two-dimensional PALMIRA the localization algorithm is frequently a centroid algorithm calculating iteratively the center-of-mass of properly weighted data [55]. Alternatively, it is possible to directly use minimization of the least-squares deviation between a theoretical or measured PSF and the segment data. A third possibility is maximum likelihood-estimation relying likewise on a model PSF and additionally on a more sophisticated statistical model for the pixel statistic than the one underlying the least-squares fitting. If the localization algorithm converged, the position and the photon number of the event are saved in a list, otherwise discarded. In both cases, a theoretical PSF properly scaled to contain the segment photon number is subtracted from the image. In such a way the pixel values in the segment become smaller than the threshold which marks this area of space as processed. The whole algorithm is called recursively with the altered image until no pixel is above threshold. Then, the next frame is evaluated.

Subtracting a theoretical signal from a measured image in order to create a segmentation effect is closely related to the CLEAN algorithm frequently used to evaluate astronomical images of far galaxies [54]. To achieve image segmentation it would also be possible to use a simple segmentation routine which always attributes the area around the maximum pixel to a single object. The CLEAN procedure is superior to the latter routine as far as proper

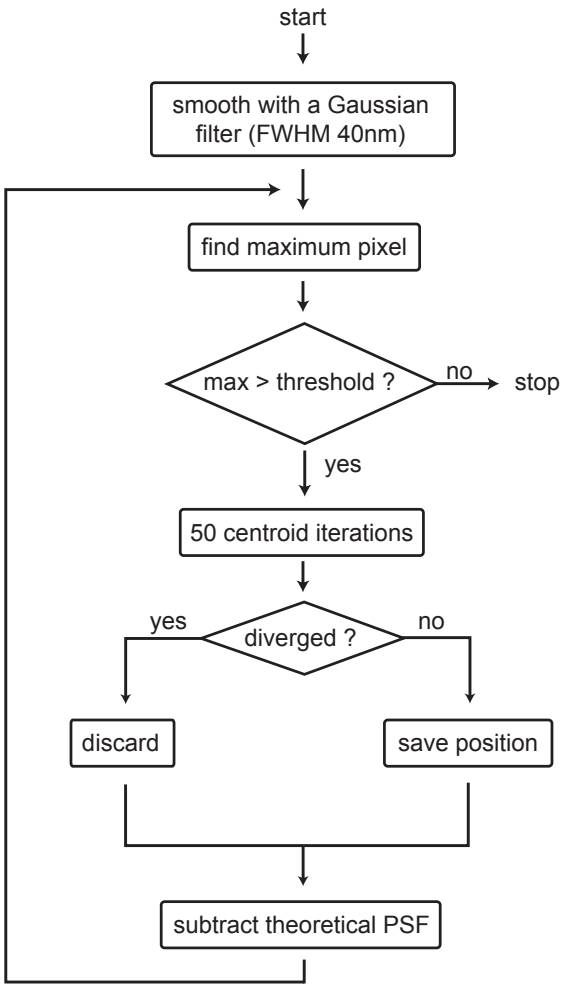

Figure 1.4: Flow diagram for the image evaluation algorithm used for PALMIRA. Objects are identified by subjecting the image to a pixel thresholding procedure. Then, the position of the molecule is determined. In two-dimensional PALMIRA the most frequent method is a centroid algorithm yielding the-center-of-mass of the optical pattern. Three-dimensional localization requires more complicated procedures. If the iteration did not diverge the position and the associated photon number are stored and the theoretical signal is subtracted from the image. If the iteration did diverge the position is discarded but nonetheless the theoretical signal centered at the starting position is subtracted. The algorithm is repeated until all pixel are below the threshold. The procedure is closely related to Hogbom's CLEAN algorithm in astronomy [54].

treatment of partially overlapping events is concerned. As long as the overlap is not too strong the CLEAN algorithm differentiates the objects while the simple segmentation always assigns the segment to one object.

In the subsequent a proof-of-principle for dual-color two-dimensional PALMIRA imaging is presented. Therefore, a Monte-Carlo simulation using the above algorithms has been performed. For further information on the particle type estimation the reader is referred to Chapter (4.4). In order not to complicate the experiment one and only one molecule is activated per frame. The frame length is long enough for practically all molecules radiating their

photons in one frame. The test distribution, Figure (1.5, A), consists of an equally spaced array with lattice constant $a = 50$ nm. At each lattice point an assembly of fluorophores is placed. In mathematical terms, the object distribution is a function as follows

$$o(\mathbf{r}) = \sin^4(\pi x/a)\sin^4(\pi y/a). \tag{1.13}$$

In order to have more than one spatial frequency in the object the sinus-functions are taken to the power of 4. The FWHM F of the peaks turns out to be roughly $F \approx a/3 \approx 17$ nm. Two lattices with fluorophores that feature emission characteristics as in Figure (4.7) are combined at a distance of $d = 35$ nm. The separation wavelength λ_d between the two color channels is taken as $\lambda_d = 592$ nm to achieve a realistic sorting matrix like in Equation (4.23). The wavelengths for the optical PSFs differ according to the type of fluorophore and are $\lambda_1 = 575$ nm and $\lambda_2 = 620$ nm. The mean photon number is $\bar{N} = 250$, the event threshold is $T = 250$. Figure (1.5, B) displays the reconstruction of the molecule distribution by formation of a histogram of relative occurrences of the estimated molecule positions. The histogram bin size is 1 nm $\times$ 1 nm. The number of samples is $M = 1 \cdot 10^5$ per molecule type. It is obvious that features smaller than the optical wavelength by almost a factor of ten can be discerned without effort. The peaks are slightly broadened due to the error in position estimation. Localization errors for the object indicated in red are slightly larger than those for the object in blue since the wavelength of the latter is smaller, $\lambda_1 < \lambda_2$. This manifests itself in the width of the distributions of the PALMIRA histogram. The estimation

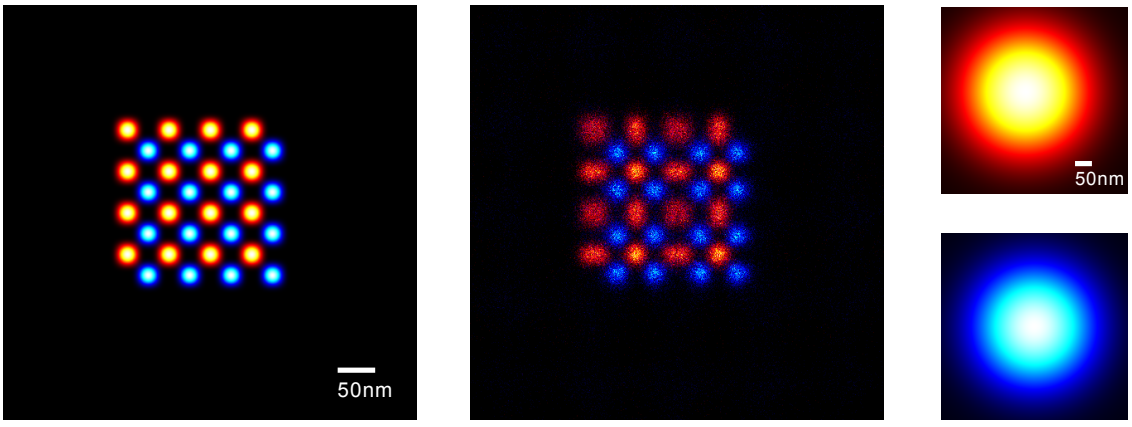

Figure 1.5: Results of a Monte-Carlo simulation for dual-color two-dimensional PALMIRA imaging of a hypothetical lattice object. The object is composed of peaked distributions of two molecules arranged on arrays with a lattice constant of $a = 50$ nm (A). The peaks of different fluorophores have a distance of $d = 35$ nm. The wavelengths for the optical PSFs are $\lambda_1 = 575$ nm (indicated in blue) and $\lambda_2 = 625$ nm (indicated in red), respectively. The reconstruction of this distribution from the PALMIRA position stack is shown in the middle (B). The classical wide-field images for the two channels are shown on the right (C). Obviously, features of several orders of magnitude smaller than the optical wavelengths might be discerned in the PALMIRA image whereas they stay unnoticed in the conventional image. Despite of the considerable crosstalk between the two channels - 35% of the 'blue' photons end up in the 'red' channel - both sublattices can be discerned with certainty.

of the molecule type is sufficiently good for the two sublattices to be distinguished properly. The classical wide-field images, Figure (1.5, C), exemplify that PALMIRA microscopy is capable of revealing details that wide-field microscopy ignores completely. Strong crosstalk between the spectral channels cannot be circumvented in wide-field microscopy because the molecules are not recorded transiently in time and space. This creates 'ghost' images in the channels. For the concrete simple distribution used here, this effect is present but not predominant because the distance of the ghost images is just the sublattice distance $d = 35$ nm which is small compared to the wavelengths involved. The effect may be predominant for objects with stronger spatial separation of the different types of molecules. This would even more degrade the quality of the conventional image. In PALMIRA microscopy crosstalk can even be an advantage because molecules are distinguished according to their crosstalk characteristics. Crosstalk need not to be absent but merely sufficiently distinct. For the simulation presented here the channel signal sorting characteristics are 0.1:0.9 (red emission) and 0.65:0.35 (blue emission). Even more colors are feasible without changing the setup. For instance, three-color imaging with only two channels is possible in PALMIRA microscopy for crosstalk characteristics 1:0, 0:1 and 1:1. The latter is as suitable for PALMIRA as the other two although it would be useless in classical microscopy with non-individual readout.

2 Mean Field Theory of the PALMIRA Technique

In this chapter a mathematical discussion of single-molecule image formation in PALMIRA microscopy is provided. Essentially, PALMIRA marks a shift of interpretation as to how an image is formed: although photons continue to play an important role, the basic entities are the single fluorophore positions. PALMIRA allows to draw direct statistical samples of the spatial molecule distribution to be investigated. These samples are of different qualities depending on its associated photon numbers. The most suggestive way of forming an image from spatial samples is a histogram of absolute occurrences. This treats every position sample irrespective of its quality in the same way and thereby disregards the degree of freedom linked to the event photon number. Imaging theory of PALMIRA incorporating an averaging over this degree of freedom will be termed mean field theory in the subsequent in order to distinguish it from deconvolution methods that take into account the photon degree of freedom.

In the subsequent the mean PSF of PALMIRA microscopy is calculated and its properties are studied. It is investigated how spatial sub-diffraction leads to a higher extension of the experiment in the temporal domain. The question of the level of confidence of the PALMIRA image is addressed.

2.1 The Mean Point-Spread-Function of PALMIRA

If the frequency of occurrence of overlapping optical patterns of different molecules is sufficiently low in a single image frame so that molecules can be imaged independently from each other. In that case, it is possible to restrict the discussion to the imaging of a single molecule. At first the localization properties at fixed photon number will be treated, then the discussion progresses on to an incorporation of photon thresholding into the theory.

2.1.1 Point-Spread-Function at Fixed Photon Number

In general, the distribution of photons in an optical image is a two-dimensional spatial multinomial probability distribution with probabilities given by the diffraction integral $h(x, y|z)$ [56, 57, 58]. Here, the optical point-spread-function $h(x - x_0, y - y_0|z - z_0)$ describes the probability to receive a photon at position (x, y) in an detection plane z if the fluorophore resides at (x_0, y_0, z_0). In scalar diffraction theory for a circular aperture, the focal part $h(x, y|0)$ is just the well-known Airy-function while the out-of-focus point-spread-function

is described by Lommel functions [57]. Vectorial diffraction theory incorporating the polarization properties of the light and theories for high-numerical aperture lenses (Sommerfeld approximation) make the analysis more complex, but essentially do not change the overall character of the PSF as a Fourier pattern of the aperture displaying pronounced oscillations as a result of electromagnetic interference [59].

All localization procedures, e.g. fitting or a center-of-mass analysis, have a strong tendency to produce results following a Gaussian distribution. The structural fact responsible for that is the data reduction inherent in retrieving the pattern position: many different photon positions are reduced to one center position. This makes applicable the central limit theorem. For the case of a centroid localization procedure that simply calculates the center-of-mass of Gaussian-weighted data [55] this fact is most obvious. At this place, the discussion is restricted to one radial dimension. For instance, let N random photon positions $\{x_i\}$, originating from a fluorophore at x_0, be given. Then, the mean value (center-of-mass)

$$\bar{x} = \frac{1}{N} \sum_i x_i \tag{2.1}$$

follows a distribution $r(\bar{x}|x_0)$ given by the N-fold self-convolution of the detection PSF

$$r(\bar{x}|x_0) = N \underbrace{[h * h * ... * h]}_{1..N} ((N(\bar{x} - x_0)) \,. \tag{2.2}$$

This is a direct consequence of the fact that the probability density of a sum is the convolution of the densities of the summands. The progressive smoothing with increasing photon number N makes the convolution more and more Gaussian, Figure (2.1, A). The error is propagated according to the standard result [53, 60, 61]

$$\sigma^2(N) = \frac{\sigma_0^2}{N} \tag{2.3}$$

with the variance σ_0^2 given by the well-known Rayleigh result for the radial width of the detection PSF [57],

$$\sigma_0 \approx \frac{0.22\lambda}{n \sin \alpha}, \tag{2.4}$$

where λ denotes the optical wavelength, n the index of refraction and α the semi-aperture angle of the lens system. As long as the photon number N is high enough it makes sense to approximate Equation (2.2) by a Gaussian and therefore assume the probability to obtain the centroid position $\bar{x}$ given an original molecule position x_0 to be

$$p(\bar{x}|x_0, N) = \frac{1}{\sqrt{2\pi}\sigma_0/\sqrt{N}} \exp\left\{ -\frac{(\bar{x} - x_0)^2}{2\sigma_0^2/N} \right\}. \tag{2.5}$$

In the case of axial localization the square-root law for the error and the Gaussianity of its distribution survive, but the dependency of σ_0 on the radial moments of the detection PSF gets more complicated. This just reflects the fact that a simple centroid analysis is no longer possible in the $z-$ dimension.

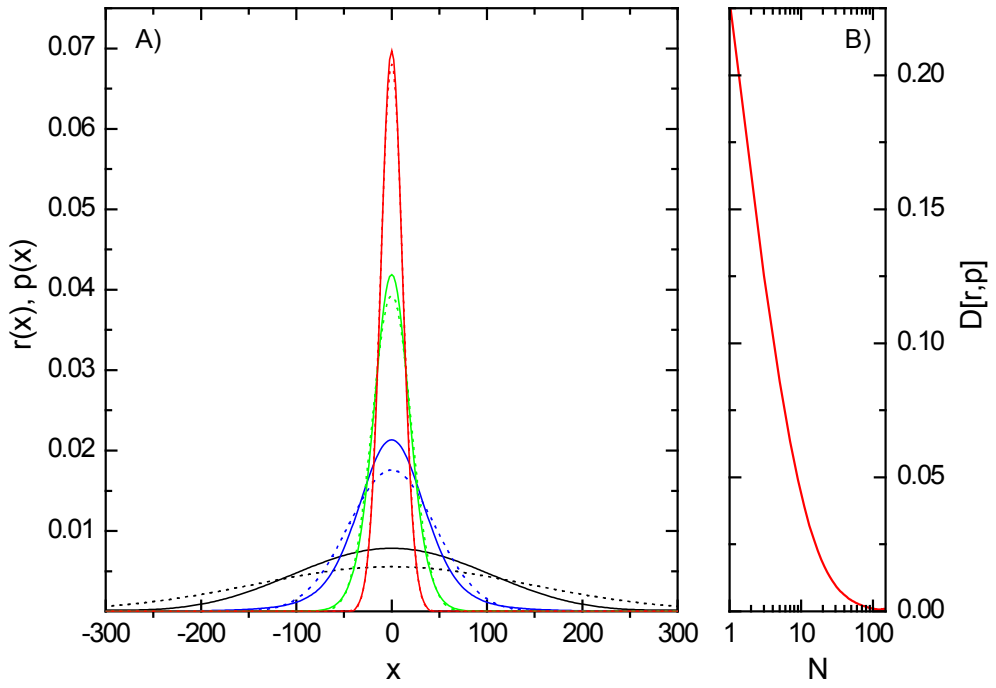

Figure 2.1: Distribution of the center-of-mass of an optical pattern. The exact distribution (solid) of Equation (2.2) is shown here together with its Gaussian approximation (dotted) of Equation (2.5) for different photon numbers $N = 1$ (black), $N = 10$ (blue), $N = 50$ (green) and $N = 150$ (red). The more photons an image contains the more the distribution of the center of the optical pattern approaches a Gaussian (A). The degree of similarity between these two distributions might be quantified using a distance measure (B). The distance approaches zero as N increases.

A quantification of the degree of similarity between the two distributions, the exact distribution of the center-of-mass in Equation (2.2) and the Gaussian in Equation (2.5), may be obtained by making use of the Kullback-Leibler distance of two probability distributions $p(x)$ and $r(x)$ [62]

$$D[r,p] = -\int p(x) \ln \left[\frac{p(x)}{r(x)}\right] dx. \tag{2.6}$$

Figure (2.1, B) shows the results of the Kullback-Leibler distance as a function of the number of photons.

2.1.2 Point-Spread-Function in the Presence of Thresholding

In order for the localization error of every event to be lower than σ_{min} with probability one, the detected events are subjected to a thresholding procedure. Exclusively events with photon numbers exceeding the threshold

$$T = \frac{\sigma_0^2}{\sigma_{\text{min}}^2} \tag{2.7}$$

are taken into account while the others are rejected. This leads to a truncation of the geometric distribution of Equation (1.5). The distribution of accepted photon numbers $N \in [T, \infty]$ now reads

$$p(N|\bar{N}, T) = \frac{\bar{N}^{N-T}}{\left[\bar{N}+1\right]^{N-T+1}}. \tag{2.8}$$

The geometric distribution is scale-invariant and thus retains its shape. To obtain the photon-averaged PSF taking into account the thresholding effect it is necessary to sum Equation (2.5) over the accepted photon numbers $N \geq T$:

$$p(\bar{\mathbf{r}}|\mathbf{r}_0, \bar{N}, T) = \sum_{N=T}^{\infty} p(\bar{\mathbf{r}}|\mathbf{r}_0, N) p(N|\bar{N}, T). \tag{2.9}$$

The localization PSF with the worst accuracy, that with $N = T$, experiences admixtures of narrower, but less probable PSFs. The result is a mean PSF which is reported here for d dimensions to cover the cases of $d = 2$ and $d = 3$ ate the same time,

$$p(\bar{\mathbf{r}}|, \mathbf{r}_0, \bar{N}, T) = \frac{b(\bar{\mathbf{r}}, \mathbf{r}_0)^T}{(\bar{N}+1)\sqrt{2\pi}^d \prod_{i=1}^{d} \sigma_i} \sum_{N=0}^{\infty} (N+T)^{d/2} \left[\frac{\bar{N}}{\bar{N}+1} b(\bar{\mathbf{r}}, \mathbf{r}_0)\right]^N, \tag{2.10}$$

with the abbreviation

$$b(\bar{\mathbf{r}}, \mathbf{r}_0) = \prod_{i=1}^{d} \exp\left(-\frac{(x_i - x_{0,i})^2}{2\sigma_i^2}\right) \tag{2.11}$$

and σ_i being the width of the detection PSF in dimension i. Since the bracket term in Equation (2.10) is strictly < 1 the series converges and might be expressed with the help of a special function, Lerch's ϕ-transcendent [63, 64],

$$\phi(x, s, a) := \sum_{i=0}^{\infty} \frac{x^i}{(a+i)^s}, \tag{2.12}$$

as

$$p(\bar{\mathbf{r}}|\mathbf{r}_0, \bar{N}, T) = \frac{b(\bar{\mathbf{r}}, \mathbf{r}_0)^T}{(\bar{N}+1)\sqrt{2\pi}^d \prod_{i=1}^{d} \sigma_i} \phi\left(\frac{\bar{N}}{\bar{N}+1} b(\bar{\mathbf{r}}, \mathbf{r}_0), -\frac{d}{2}, T\right). \tag{2.13}$$

This is the PSF of PALMIRA microscopy. The ϕ-function plays the same role in PALMIRA microscopy as the depletion doughnut beam in STED microscopy and the Airy function in classical microscopy. The properties of this function will be studied in Chapter (2.2).

2.1.3 Data Representation and the Imaging Equation

In this paragraph it is shown that the assumption of independent recording of fluorophores, a sufficiently high number of activation cycles per molecule and translation symmetry of the localization accuracy result in an imaging equation in the form of a convolution.

Two approaches of representing the stochastic data $\left\{\bar{\mathbf{r}}_i^j, N_i^j\right\}$, Equation (1.12), have been reported in the literature so far: (i) the sorting of $\bar{\mathbf{r}}_i^j$ into a histogram of relative occurrences

and the (ii) formation of a sum of localization distributions as in Equation (2.5) centered at $\bar{\mathbf{r}}_i^j$ with a width determined by the photon number N_i^j. The latter will be referred to as the 'Gaussian view' [29], the former as 'histogram view'.

For a simple histogram view (i) with a finite number of excitations E_i of molecule i the final image $b(\bar{\mathbf{r}})$ is a function constructed as follows

$$b(\bar{\mathbf{r}}) = \frac{1}{F}\sum_{i=1}^{F}\sum_{j=1}^{E_i}\frac{1}{E_i}p(\bar{\mathbf{r}}|\mathbf{r}_i, N_i^j). \tag{2.14}$$

as long as all molecules are imaged independently from each other. Here the summation is over all molecules i and their excitation cycles j taking into account that each retrieved pattern position fluctuates around the real fluorophore position $\mathbf{r}_i$ as described by Equation (2.5). At this stage, the image is still 'unstable' in a stochastic sense since the number of samples from molecule i, E_i, and the photon number from molecule i in sample j, N_i^j are random numbers and hence the image $b(\bar{r})$ is also a random field. Instead of summing stochastically over observed photon numbers, one could also sum over all possible photon numbers N using a counting measure $H_i(N)$ of the photon number N. Then, Equation (2.14) reads

$$b(\bar{\mathbf{r}}) = \frac{1}{F}\sum_{i=1}^{F}\sum_{N=0}^{\infty}\frac{H_i(N)}{E_i}p(\bar{\mathbf{r}}|\mathbf{r}_i, N). \tag{2.15}$$

If the number of events E_i is sufficiently high the relative frequencies of photon numbers approach the distribution in Equation (2.8). Thus, it is obtained

$$\frac{H_i(N)}{E_i} \to p(N|\bar{N}, T) \tag{2.16}$$

for all i and the imaging equation, Equation (2.15), becomes

$$b(\bar{\mathbf{r}}) = \frac{1}{F}\sum_{i=1}^{F}\sum_{N=0}^{\infty}p(\bar{\mathbf{r}}|\mathbf{r}_i, N)p(N|\bar{N}, T) = \int p(\bar{\mathbf{r}}|\mathbf{r}, \bar{N}, T)o(\mathbf{r})d\mathbf{r}. \tag{2.17}$$

Here, Equations (1.3) and (2.13) were used. If the variances of the effective PSF, Equation (2.13), are not position-dependent it is obtained

$$p(\bar{\mathbf{r}}|\mathbf{r}, \bar{N}, T) = p(\bar{\mathbf{r}} - \mathbf{r}, \bar{N}, T), \tag{2.18}$$

and the imaging equation turns out to be a convolution of the object distribution and a point-spread-function as in classical confocal microscopy,

$$b(\bar{\mathbf{r}}) = p(\mathbf{r}, \bar{N}, T) * o(\mathbf{r}). \tag{2.19}$$

For the Gaussian view (ii) the situation is similar: stochastic scanning takes place in the original fluorophore coordinate $\mathbf{r}$, a fluctuating estimate $\bar{\mathbf{r}}$ is returned. In contrary to (i) a localization Gaussian is centered around $\bar{\mathbf{r}}$, which leads to the imaging equation

$$b(\mathbf{r}) = \sum_{N=T}^{\infty}\int\left\{\int p(\mathbf{r}|\mathbf{r}'', N)p(\mathbf{r}''|\mathbf{r}', N)d\mathbf{r}''\right\}p(N|\bar{N}, T)o(\mathbf{r}')d\mathbf{r}'. \tag{2.20}$$

The inner convolution can be performed analytically and yields again a Gaussian as in Equation (2.5), but with higher variance $\sigma_0^2 \rightarrow 2\sigma_0^2$. The summation over the photon number reproduces the mean PDF of Equation (2.13).

In summary, the Gaussian view features the disadvantages of a broader effective PSF and the mixing of position information and accuracy information. It is not obvious whether an image value $b(\mathbf{r}') \neq 0$ originates from a localization event or the outer part of a 'credibility' Gaussian. Especially, the events with a poor quality of position estimation are stretched out in space very strongly. The Gaussian view pretends too high a degree of sampling. Nonetheless, it is a first attempt towards taking into account properly the accuracy information. This point will be discussed further in Chapter (5).

2.2 Measure of Resolution

The width of the PSF, Equation (2.13), is a measure of the attainable resolution since it describes how strong a single point of the object is spread out in the image. Here, the discussion is resumed in one dimension. As above, it is necessary to sum over different contributions above the threshold. The variance $\Sigma^2(\bar{N}, T, \sigma_0)$ of the effective mean PSF then reads

$$\begin{aligned}
\Sigma^2(\bar{N}, T, \sigma_0) &= \sum_{N=T}^{\infty} \frac{\sigma_0^2}{N} \frac{\bar{N}^{N-T}}{(\bar{N}+1)^{N-T+1}} \\
&= \frac{\sigma_0^2}{\bar{N}+1} \sum_{N=0}^{\infty} \frac{\left(\frac{\bar{N}}{\bar{N}+1}\right)^N}{N+T} \\
&= \frac{\sigma_0^2}{\bar{N}+1} \phi\left(\frac{\bar{N}}{\bar{N}+1}, 1, T\right). \qquad (2.21)
\end{aligned}$$

Thus, the standard deviation $\Sigma(\bar{N}, T, \sigma_0)$ of the single molecule PALMIRA-PSF given $\bar{N}$ photons on average, a threshold T and a standard deviation σ_0 of the single molecule photon distribution is given by

$$\Sigma(\bar{N}, T, \sigma_0) = \frac{\sigma_0}{\sqrt{\bar{N}+1}} \sqrt{\phi\left(\frac{\bar{N}}{\bar{N}+1}, 1, T\right)}. \qquad (2.22)$$

Figure (2.2) shows this function for different $\bar{N}$ compared to the square-root law for the fixed photon number case, Equation (2.3). As a result of the admixture of narrower states to the local PSF at $N = T$ the effective standard deviation is always smaller than the value given by the square-root law,

$$\Sigma(\bar{N}, T, \sigma_0) < \frac{\sigma_0}{\sqrt{T}}. \qquad (2.23)$$

It is instructive to compare the resolution formula in Equation (2.22) to the resolution formula of STED microscopy [4, 23] that is also given by a standard deviation,

$$\Sigma_{\text{STED}}(I, I_{\text{sat}}, \sigma_0) = \frac{\sigma_0}{\sqrt{1 + I/I_{\text{sat}}}}. \qquad (2.24)$$

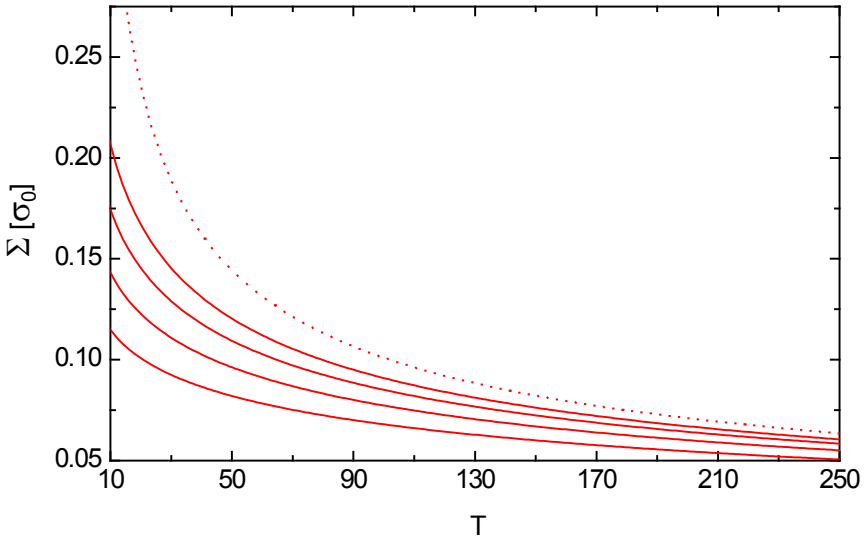

Figure 2.2: The standard deviation $\Sigma(\bar{N}, T, \sigma_0)$ of the PALMIRA-PSF as a function of the threshold T for different mean photon numbers $\bar{N}$ (solid lines) in comparison to the square-root law for fixed photon number $1/\sqrt{T}$ (dotted). From top to bottom the photon numbers are increasing $\bar{N} = 25, 50, 100$ and 200. Due to the admixture of narrower states the effective standard deviation lays always below the square-root result. The resolution might be raised by both increasing $\bar{N}$ or T. Super-resolution might be achieved even with vanishing threshold $T = 0$ if $\bar{N}$ is sufficiently large.

Here, I_{sat} denotes the saturation intensity of the quenched molecular transition and I the actual STED beam intensity. The saturation intensity I_{sat} depends on the orientation of the dye, the molecular cross-section, the wavelength, the temporal structure and the slope of the doughnut-shaped STED light. Just as Equation (2.24) is a correction to the classical Abbe formula due to saturation effects, Equation (2.22) is a correction to single molecule localization accuracy due to photon thresholding.

The localization error at fixed photon number needs not always to follow a simple square root law as Equation (2.3), but can lead to more complicated dependencies if effects of pixelation or background noise are incorporated. For instance, including readout and background noise Equation (2.3) is replaced by

$$\sigma^2(N) = \frac{\sigma_0^2 + a^2/12}{N} + \frac{4\sqrt{\pi}\sigma_0^3\bar{b}^2/a}{N^2}, \tag{2.25}$$

with a denoting the pixel length and $\bar{b}$ the mean background per pixel [55]. In this case Equation (2.22) is changed into

$$\Sigma^2(\bar{N}, T, \sigma_0, a, b) = \frac{\sigma_0^2 + a^2/12}{\bar{N}+1}\phi\left(\frac{\bar{N}}{\bar{N}+1}, 1, T\right) + \frac{4\sqrt{\pi}q\sigma_0^3\bar{b}^2/a}{\bar{N}+1}\phi\left(\frac{\bar{N}}{\bar{N}+1}, 2, T\right).$$

All summations over a truncated geometric distribution times a polynomial terms can be expressed as a ϕ-function.

2.3 Temporal Extent of a Measurement

In classical microscopy the statistical sampling degree of a pixel is determined by the number of photons falling into that pixel. In PALMIRA microscopy this role is adopted by the number of accepted position estimates of the molecule to be imaged. If the measurement time or the maximum number of switching cycles is limited PALMIRA microscopy suffers a conflict of objectives between resolution and statistical sampling. If the threshold is lowered the number of available position estimates increases. That leads to better stochastic sampling but worse spatial resolution. If the threshold is increased localization is made better but the number of events being available to build up the image is lowered. In other words: spatial resolution is gained at the cost of a certain temporal stationarity of the image. This is a demand both on (i) the fatigue properties of the switching transition of the molecule and (ii) on the timescale on which the image might change over time. Both, the number of reversible switching cycles and the stationarity time interval of the image have to be large enough in order to collect sufficiently many stochastic samples.

This effect is quantified by establishing a functional relationship between the mean photon number $\bar{N}$, the recognition threshold T and the mean waiting time $\bar{W}$ for an event producing a photon number above threshold. The number of samples S required per molecule for gaining a good estimate of the single-molecule spatial distribution is assumed to be given.

The cumulative distribution for the geometric distribution of Equation (1.5), i.e. the probability for the photon number to be contained in the interval $[N_0, N_1]$, is given by

$$P([N_0, N_1]) = \sum_{N=N_0}^{N_1} p(N|\bar{N}) = \left(\frac{\bar{N}}{\bar{N}+1}\right)^{N_0} - \left(\frac{\bar{N}}{\bar{N}+1}\right)^{N_1+1}. \tag{2.26}$$

The state space of the problem in question is made up by the two events $N < T$ and $N \geq T$ with the probabilities $P([0, T-1])$ and $P([T, \infty])$, respectively. The dichotomy again leads to the binomial model. For that case, the probability of obtaining k failures given r successes in $n = r + k$ samples is described by the negative binomial distribution [52]

$$p(k|r) = \binom{k+r-1}{k} p^r (1-p)^k \tag{2.27}$$

with the mean

$$\bar{k} = r/p - r. \tag{2.28}$$

For the problem under consideration, we have

$$\begin{aligned} p &= P[T, \infty] \\ r &= S \\ k &= W - S \end{aligned} \tag{2.29}$$

Together with Equation (2.26) the probability distribution of the waiting time W for S events being above the threshold T reads

$$p(W|S, T, \bar{N}) = \binom{W-1}{W-S} \left(\frac{\bar{N}}{\bar{N}+1}\right)^{TS} \left(1 - \frac{\bar{N}}{\bar{N}+1}\right)^{W-S} \tag{2.30}$$

with, of course, $W \geq S$. For the mean waiting time it is obtained from Equation (2.28)

$$\bar{W}(S,T,\bar{N}) = S\left(1+\frac{1}{\bar{N}}\right)^{T}. \tag{2.31}$$

This equation may also be used in backward direction to obtain the threshold attainable $T(\bar{W},\bar{N},S)$ if the mean number of switching cycles is prescribed. In that case, the threshold compatible with the structure of the events is given by

$$T(\bar{W},\bar{N},S) = \frac{\ln\bar{W} - \ln S}{\ln(\bar{N}+1) - \ln\bar{N}}. \tag{2.32}$$

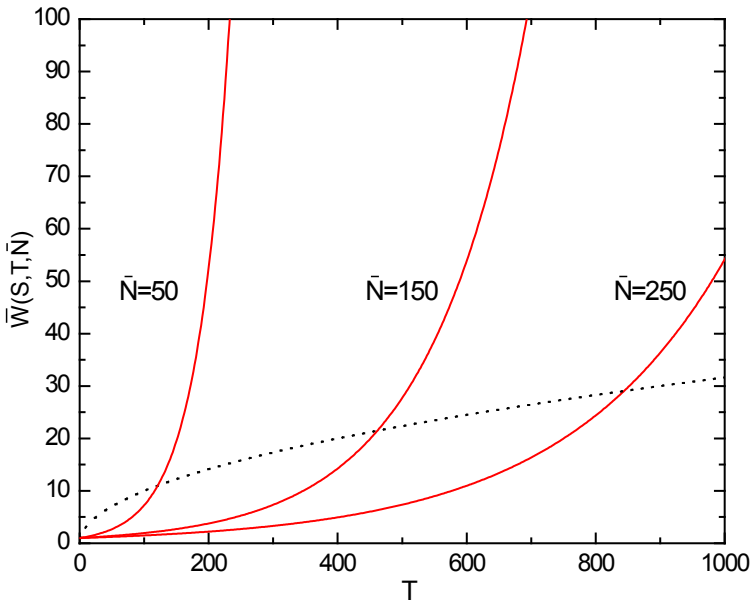

Figure 2.3: The waiting time $\bar{W}(S,T,\bar{N})$ in numbers of switching cycles for one event, $S = 1$, to exceed the threshold T given the fluorophore yields $\bar{N}$ photons on average (red). In addition, the resolution augmentation factor at threshold, $\sqrt{T}$, is shown (dotted). The smaller the mean photon number $\bar{N}$ the longer the waiting time for an event to be above threshold. Spatial resolution is gained at the cost of temporal stationarity. The hyperbolic dependency of the rate in Equation (2.31) manifests itself in a pronounced increase of the slope as $\bar{N}$ decreases.

Figure (2.3) shows a plot of the function $\bar{W}(S,T,\bar{N})$ for some realistic values of $\bar{N}$ together with the resolution reduction factor at T for $S = 1$. The waiting time for other values of S is directly proportional to that for $S = 1$. Essentially, the waiting time $\bar{W}(S,T,\bar{N})$ depends exponentially on the threshold T with a rate proportional to the logarithm of the inverse of the mean photon number $\bar{N}$. The hyperbolic dependency of the base in Equation (2.31) on $\bar{N}$ is striking. The smaller the mean photon number $\bar{N}$ the more severe is the growth of the mean waiting time $\bar{W}$ when going to an infinitesimally higher threshold T.

Equation (2.31) assumes that one and only one molecule gets activated. In reality, the specimen contains M molecules which are switched on with a probability P_{sw}, i.e. in each

step a mean number $\bar{M} = MP_{\text{sw}}$ of molecules is activated. If S samples are to be drawn from M molecules there may also occur events in parallel and the waiting time $\bar{W}$ measured in activation cycles reads

$$\bar{W}(S,T,\bar{N},\bar{M}) = \frac{S}{\bar{M}}\left(1+\frac{1}{\bar{N}}\right)^{T} = \frac{S}{M}P_{\text{sw}}^{-1}\left(1+\frac{1}{\bar{N}}\right)^{T}. \tag{2.33}$$

Equation (2.33) is equivalent to inquiring for the mean waiting time for a molecule to be activated AND to yield a photon number above threshold. The probability P_+ of this statistical event is simply

$$P_+ = P_{\text{sw}}\left(\frac{\bar{N}}{\bar{N}+1}\right)^{T}. \tag{2.34}$$

The fraction γ of the sample number S relative to the number of molecules M,

$$\gamma = \frac{S}{M}, \tag{2.35}$$

defines a sort of observance fraction of the molecules. For instance, for $\gamma = 1$ each molecule is seen for one time on average. The waiting time in Equation (2.33) depends crucially on γ and on the probability P_+.

2.4 Level of Confidence of a Measurement

After having studied how many switching cycles $\bar{W}$ have to be attended on average to achieve a number of samples S from a single molecule it is interesting how many of these samples are required to achieve a certain level of confidence α of the final image. Of course, this will depend on the object molecule distribution $o(\mathbf{r})$: an object consisting of a few molecules might be mapped authentically with much lower a number of samples than a more complicated object. The focus is put here on the degree of convergence of the final image. With a finite number of samples the final image $b(\mathbf{r})$ is approximated by a histogram of relative occurrences. With infinitely many samples $S \to \infty$ this histogram converges against the function $b(\mathbf{r})$ given in the imaging Equation (2.20) just as the classical image in confocal microscopy acquires stability when the photon number goes to infinity. The theoretical PALMIRA-PSF of Equation (2.13) is dependent on the mean photon number $\bar{N}$, the threshold T and the width of the underlying detection PSF and plays a decisive role here. The question is to which extent the relative frequencies of the PALMIRA histogram do approach the final image density $b(\mathbf{r})$. The image $b(\mathbf{r})$ can assumed to be given in discrete form as $\{b_i\}$. The $\{b_i\}$ form a normalized probability. The probability $p(S_i|b_i,S)$ to receive S_i molecule positions in the pixel i when S samples have been collected is given by the binomial distribution

$$p(S_i|b_i,S) = \binom{S}{S_i} b_i^{S_i}(1-b_i)^{S-S_i}. \tag{2.36}$$

In a standard PALMIRA histogram the value b_i is estimated by the relative frequency estimator

$$\hat{b}_i(S_i) = \frac{S_i}{S}. \tag{2.37}$$

The crucial question is how large the number of samples S has to be in order for $\hat{b}_i$ having a relative error ε smaller than ε_{rel} with probability of at least α. Mathematically, it is inquired from which $S \geq S_{\text{min}}$ on the probability of the statistical event

$$\hat{b}_i(S_i) \in [\max\{0, b_i - b_i\varepsilon_{\text{rel}}/2\}, \min\{1, b_i + b_i\varepsilon_{\text{rel}}/2\}] \tag{2.38}$$

is larger than α. This is equivalent to finding S with

$$P(S, \varepsilon_{\text{rel}}, b_i) := \sum_{S_i = S\max\{0, b_i - b_i\varepsilon_{\text{rel}}/2\}}^{S\min\{1, b_i + b_i\varepsilon_{\text{rel}}/2\}} p(S_i|b_i, S) \geq \alpha \tag{2.39}$$

with the limits of summation defining the lower and upper tolerable fluctuation of the absolute number of position estimates falling into the pixel. The limits of the summation are rounded to the nearest integer. $S > 20$ is a convenient choice as long as $b_i > 0.01$. $P(S, \varepsilon_{\text{rel}}, b_i)$ is the probability that S position estimates lead to a relative error smaller than ε_{rel}.

Fortunately, Equation (2.39) can be transformed into a relationship of special mathematical functions that is suitable for numerical computations. A non-approximative result is required here since no restrictions should be made on the the pixel probability $b_i \in [0, 1]$. The function

$$P(m, n; b_i, S) = \sum_{m}^{n} p(S_i|b_i, S) \tag{2.40}$$

occurring in Equation (2.39) describes the cumulation of the binomial distribution of total sample number S and probability b_i over the interval $[m, n]$ and is given by [65]

$$P(m, n; b_i, S) = K(b_i; m, S - a + 1) - K(b_i; n + 1, S - n) \tag{2.41}$$

with the function $K(p; m, n)$ defined by a quotient of beta functions

$$K(p; m, n) = \frac{B(p; m, n)}{B(m, n)}. \tag{2.42}$$

Here, the beta functions are [66]

$$B(p; m, n) := \int_0^p s^{m-1}(1 - s)^{n-1} ds \tag{2.43}$$

and

$$B(m, n) := \frac{(m - 1)!(n - 1)!}{(m + n - 1)!}. \tag{2.44}$$

Figure (2.4) evaluates the probability $P(S, \varepsilon_{\text{rel}}, b_i)$ of Equation (2.39) for different values of S, the error ε_{rel}, the error level α and the pixel probability b_i to be estimated. The probability $P(S, \varepsilon_{\text{rel}}, b_i)$ to have a tolerable error approaches one for growing sample number S irrespectively of the pixel probability b_i. At fixed sample number S this probability is higher for larger b_i. This makes sense since the relative spread of the binomial distribution

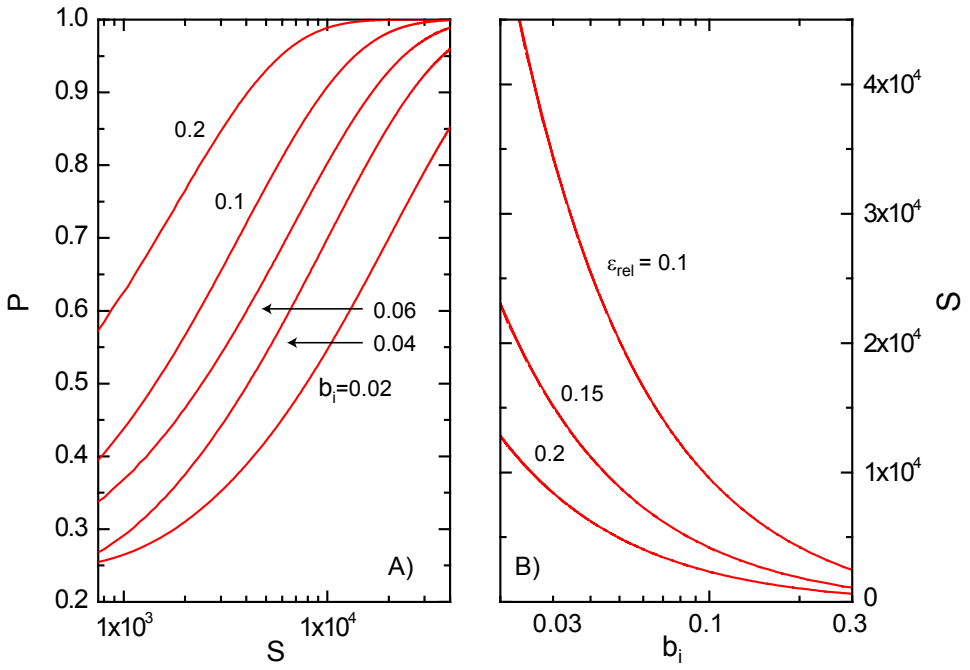

Figure 2.4: The probability $P(S, \varepsilon_{\text{rel}}, b_i)$ to have a relative error $\varepsilon < \varepsilon_{\text{rel}}$ as a function of S and the pixel mean values b_i is displayed as an explicit plot for different b_i and $\varepsilon_{\text{rel}} = 0.1$ (A) and as an implicit plot for the credibility level $\alpha = 0.9$ for different ε_{rel} (B). The left figure reveals that in a natural way the probability to have a bearable error is increasing with increasing number of samples S. The more b_i approaches zero the smaller becomes this probability. The right figure points out how many samples are needed to achieve a prescribed error at 90% probability. Of course, this number is smaller for a higher relative error.

$\sqrt{Sb_i(1-b_i)}/b_i$ in Equation (2.36) diverges as $b_i \to 0$. For this configuration fluctuations are most likely to fall outside the tolerance interval given in Equation (2.38). The implicit plot on the right provides the number of samples needed for a confidence level of 90% for different error strengths ε_{rel}.

A concrete example is the imaging of just two molecules at a distance $d = 20$ nm with mean photon number $\bar{N} = 250$ and threshold $T = 50$. Then, the width of the PSF is about 16 nm and the objects are not overlapping in the PALMIRA image. Using a pixel of 3 nm the two peaks of the image can be shown to be $b_i = 0.02$ by using the PSF in Equation (2.13). From Figure (2.4, B) it is seen that for $S > 2.5 \cdot 10^4$ samples the relative error is $\varepsilon_{\text{rel}} < 15\%$ with more than 90% probability. Since both molecules reside at sub-diffraction distance the switching probability P_{sw} has to be low enough to switch on one molecule only. From Figure (2.3) it is seen that for the values the waiting time is $\bar{W} = 1.22 \cdot S \approx 3 \cdot 10^4$ if in every cycle one and only one object is switched on. The high number of cycles is due to the small pixel-size of 3 nm. Using a pixel-size of 30 nm would result in $b_i = 0.125$ and $S = 2000$.

3 Localization Methods and their Performance

The task of measuring the position or tracking the path of molecules has been investigated since the advent of high-quality CCD cameras in the 1980s. Early investigations focus strongly on optimization of the pixel size or on statistical fluctuations of the light field [67, 68]. The biophysical literature reports systematic examinations of particle tracking [55, 69] and radial localization of entirely focused objects with idealized detection PSFs [70]. Statistical models based on Poisson pixel statistics and Gaussian object functions have been considered thoroughly in the related but different field of electron microscopy [71, 72]. All these authors discuss radial localization only. In the last years, single-particle tracking has brought with it the experimental implementation of three-dimensional localization procedures like defocused imaging [73, 74] and encoding of the axial information in asymmetries of the detection PSF using an astigmatism [75]. A detailed study of a maximum-likelihood estimator for the full three-dimensional position has been given recently [76]. However, the authors restrict their theory to aberrated detection PSFs modeling the diffraction of light traversing a cover-slip before entering the sample.

Here, full vectorial-diffraction theory detection PSFs are employed. In first instance, the imaging model and the way of its statistical analysis is introduced. Next, the theory is applied to localization via multi-channel defocused imaging (i), astigmatic single- and double-channel imaging (ii), use of a 4pi-microscope (iii) and a multi-point setup using photodiodes instead of CCDs to obtain a better temporal resolution (iv). The last two methods have not been described in the literature so far.

3.1 Theoretical Framework

In the subsequent the coordinate system is defined, the stochastic model for description of the imaging process is set up and the general method of its information theoretic analysis is presented.

3.1.1 The Stochastic Image Model

Let the image consist of C channels with $M_x \times M_y$ pixels recording transversal cross-sections of the detection PSF $h(x, y, z)$ at axial positions $\{z_l\}, l = 1..C$. Then, the mean value of photo-electrons in pixel ij of channel l may be denoted by

$$\bar{N}_{ij}^{l}(x, y, z) = \bar{N} h(x_i - x, y_i - y, z_l - z) + \bar{b}, \tag{3.1}$$

where $\bar{N}$ is the mean total photon number, $\bar{b}$ is the mean background, x, y, z are the emitter positions and $\{x_i, x_j\}$ the pixel centers. All z values are measured relative to the focal points in the image and the sample space, respectively. Figure (3.1) shows a plot clarifying the

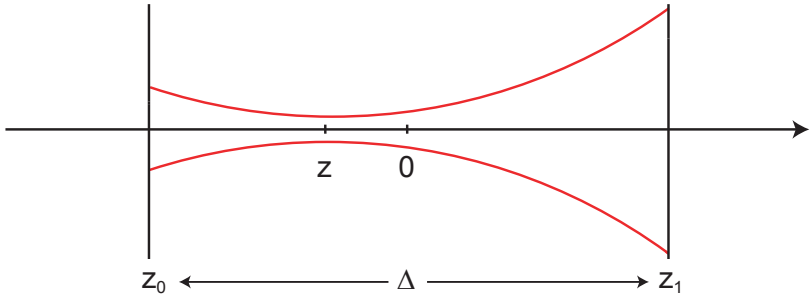

Figure 3.1: Definition of the coordinate system used for the Poisson double-channel imaging model. The typical case of two acquisition planes of distance Δ at z_0 and $z_1 = -z_0$ is shown. The molecule is located at z. The red lines indicate the beam radius. The detection PSF is centered axially at z.

coordinate system used. The shifted PSF defines sorting probabilities $h_{ij}^l(\mathbf{r})$ that depend on the absolute spatial position $\mathbf{r}$ of the fluorophore,

$$h_{ij}^l(\mathbf{r}) = h(x_i - x, y_i - y, z_l - z). \tag{3.2}$$

The detection PSF is assumed to be normalized so as to assure that it is guaranteed that

$$\sum_{ijl}^{M_x, M_y} h_{ij}^l(\mathbf{r}) = 1 \tag{3.3}$$

for all x, y, z. In the simplest case, the photons are distributed equally among the C channels.

Assuming Poisson statistics, see Equation (1.9), for the photo-detection signal of each pixel [71] the overall probability for the image of a single molecule reads [1]

$$p(\{N_{ij}^l\}|\mathbf{r}) = \prod_l \prod_{ij} \frac{1}{N_{ij}^l!} [\bar{N}_{ij}^l(\mathbf{r})]^{N_{ij}^l} e^{-\bar{N}_{ij}^l(\mathbf{r})}, \tag{3.5}$$

with $\{N_{ij}^l\}$ being the photon numbers recorded in pixel ij of channel l.

[1] An alternative image model featuring a fixed photon number N instead of the mean number $\bar{N}$ is given by the multinomial distribution

$$p(\{N_{ij}^l\}|\mathbf{r}) = \frac{N!}{\prod_{ijl} N_{ij}^l!} \prod_{ijl} \left[h_{ij}^l(\mathbf{r})\right]^{N_{ij}^l}. \tag{3.4}$$

Superposition of such a multinomial signal and Poisson background noise would be consequent in terms of rigorousness but unnecessarily complicated. The reality of experiments is fully captured by the Poisson model that allows straightforward incorporation of background noise.

3.1.2 Fisher Information and Cramer-Rao-Inequality

The characterization of estimation procedures requires the systematic investigation of errors, i.e. of the variance of the methods applied. Given a statistical model $p(o|x)$ for the data o conditional on the parameter x its information content can be measured using the Fisher information $I(x)$ [53, 77, 78, 79, 80],

$$I(x) := E_o\left[\frac{\partial \ln p(o|x)}{\partial x}\right]^2 = \int \left[\frac{\partial \ln p(o|x)}{\partial x}\right]^2 p(o|x)do \tag{3.6}$$

Here, $E_o[]$ is the expectation value taken with respect to the stochastic parameter o. Mathematically, Fisher information measures the variance of the derivative of the likelihood function and quantifies its dependency on the parameter to be estimated. The variance of any unbiased estimator $\hat{x}(o)$ of x is bounded from below by the inverse of $I(x)$,

$$\text{var}\left[\hat{x}(o)\right] \geq \frac{1}{I(x)}. \tag{3.7}$$

This is the well-known Cramer-Rao-uncertainty [53, 81, 82].

The expectation value in Equation (3.6) may be calculated without much effort for the case of basic distributions like the Gaussian distribution or the Poisson distribution. For the latter case, let the distribution given by

$$p(o|f(x)) = \frac{1}{o!}[f(x)]^o e^{-f(x)} \tag{3.8}$$

with the mean $f(x)$ depending on a parameter x under consideration. For this case, the Fisher information is essentially a measure of the first derivative of the mean relative to the mean,

$$I_x(x) = \frac{1}{f(x)}\left[\frac{\partial f(x)}{\partial x}\right]^2. \tag{3.9}$$

For the optical model of Equation (3.5) the function $f(x)$ is provided by Equation (3.1) and the Fisher information on the coordinate x with the other coordinates y and z, the optical parameters $\bar{N}, \bar{b}$ and the setup parameters z_l given reads

$$I_x(x, y, z, \bar{N}, \bar{b}) = \sum_l \sum_{ij} \frac{1}{\bar{N}^l_{ij}}\left[\frac{\partial \bar{N}^l_{ij}(x, y, z, \bar{N}, \bar{b})}{\partial x}\right]^2. \tag{3.10}$$

The additivity stems from the combined effect of the logarithm in Equation (3.6) together with the product in Equation (3.5). The Fisher information for the Poisson single-particle multi-channel imaging model of Equation (3.5) becomes [2]

$$I_x(x, y, z, \bar{N}, \bar{b}) = \sum_{ijl} \frac{\left[\bar{N}\partial_x h^l_{ij}(x, y, z)\right]^2}{\bar{N}h^l_{ij}(x, y, z) + \bar{b}}. \tag{3.11}$$

[2] Fisher information is equal for the multinomial model of Equation (3.4) and the Poisson model of Equation (3.5) in the absence of noise. This might be shown by computing the covariance $\langle N^l_{ij} N^k_{mn}\rangle$. Additional Poisson background noise will even more favor the Poisson model.

With a sufficiently large support the summation over i, j leads to a vanishing of the $x, y-$ dependence of $I_x(x, y, z, \bar{N}, \bar{b})$. The remaining free parameters are the mean photon number $\bar{N}$, the background noise $\bar{b}$, the axial position z and the detection mode parameters, i.e. the number of channels C , the focal plane positions $\{z_l\}$ and the concrete PSF. For the other spatial coordinates y and z the expression is analogous. The derivative is always with respect to the coordinate under consideration.

3.2 Position Estimation Methods

In the following the information theoretic framework of the preceding paragraphs is applied to different position estimation methods. It is studied to which precision the full three-dimensional position can be retrieved using a defocus, an astigmatic or a four-channel 4pi-setup. Special emphasis is put onto estimation of the axial coordinate. Additionally, a planar multi-point setting consisting of several photo-diodes is examined.

3.2.1 Defocus

A standard diffraction detection PSF shows reflection symmetry with respect to the focal plane. Consequently, the image originating from a particle dislocated out of the focus does only depend on the relative distance to the focal plane. It is obtained that $h(x, y, z) = h(x, y, -z)$. This symmetry has to be broken if the full three-dimensional position is to be inferred unambiguously. In principle, it would also be possible to work with off-focus excitation and to use only the positive z-branch of the detection PSF to obtain a non-ambiguous axial position estimate. Nonetheless, it may be desirable to work close but symmetrically to the focal plane in order to achieve good radial localization. For that reason, a double-channel defocus setup as indicated in Figure (3.1) may be used. Figure (3.2) shows radial cross-sections of the expected image for different positions of the molecule. The channels

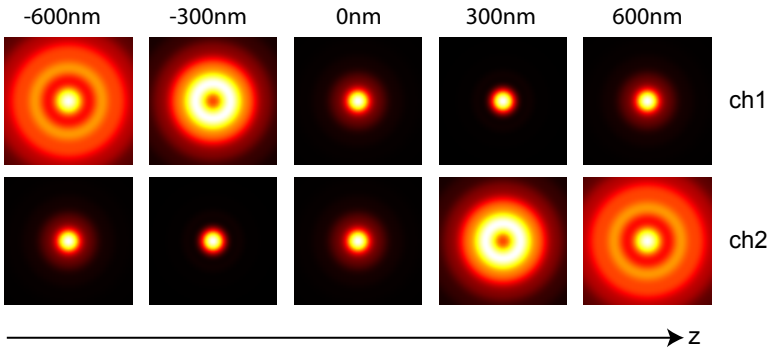

Figure 3.2: The expected mean images at different molecule z positions for a two-channel defocus setup. The channels are focused at 300nm and -300nm, respectively. For $z = 0$, both channels display the same image since the detection PSF is symmetric under reflection at the focal plane. Moving out of the focus leads to images of increasing width in one channel, while the other channel is more strongly focused initially and after all will also get defocused.

are focused at ± 300 nm, respectively. The choice of a focus distance $\Delta = 600$ nm will be justified in the subsequent. At $z = 0$ nm both channels display the same slightly defocused image. Moving away from zero defocuses one channel while the other is temporarily focused more heavily and reaches the same degree of defocusing not until at a higher value of z.

The defocusing behavior is exemplified in more detail in Figure (3.7, I) describing the FWHM of the images as a function of z. This figure is presented later in the text to allow a direct comparison with the widths of the astigmatic PSFs. In general, there is a conflict of objectives between axial and radial localization: while axial localization draws upon changes in the geometry of the image, radial localization requires this geometry to be as narrow as possible. Unfortunately, such a change in the geometry usually may only be achieved as an increase of the width of the pattern. However, this can often be tolerated since axial position estimation is more critical in terms of the measurement error than radial estimation.

The performance of position estimation depends strongly on the focus distance Δ. Figure (3.3) presents the Cramer-Rao bounds resulting from Equation (3.11) for radial $\sigma_r(z)$ and axial estimation $\sigma_z(z)$ as a function of the axial position z for different focus distances $\Delta = 200$ nm, 400 nm, 600 nm and 800 nm. The mean photon number is $\bar{N} = 250$ as has been stated for Rh6B derivatives [83]. The pixel-length of the camera is 100 nm with a mean

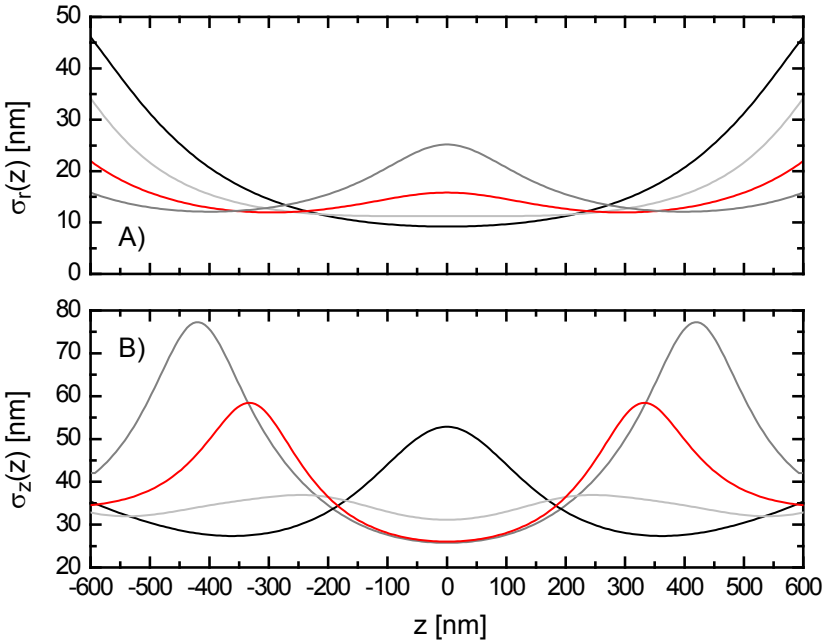

Figure 3.3: Cramer-Rao bounds for radial $\sigma_r(z)$ (A) and axial estimation $\sigma_z(z)$ (B) for different camera distances $\Delta = 200$ nm (black), 400 nm (light grey), 600 nm (red) and 800 nm (grey). The photon number is $\bar{N} = 250$ and the background noise $\bar{b} = 1$. Radial localization has an optimum at $z = \pm\Delta/2$ for $\Delta > 200$ nm while axial localization can feature a single error maximum at $z = 0$ for $\Delta = 200$ nm as well as double error peaks at $z \approx \pm\Delta/2$ for larger Δ. $\Delta = 400$ nm is a convenient choice for comparatively homogeneous axial error properties over a large support $z \in [-600\text{ nm}, 600\text{ nm}]$. $\Delta = 600$ nm is an option when axial position estimation is to be as accurate as possible on an even smaller support $z \in [-200\text{ nm}, 200\text{ nm}]$.

background $\bar{b} = 1$ as is frequently observed in experiments [7]. High-resolution numerical detection PSFs were re-sampled to this size via smoothing with a square function in order to take into account the coarse-graining of the CCD. Radial localization is optimal at $z = 0$ for the comparatively small distance $\Delta = 200$ nm. For increasing Δ two secondary minima appear at $\pm\Delta/2$ while precision worsens directly at $z = 0$ due to the increase of the width of the PSF. Axial position estimation at low distance $\Delta = 200$ nm features two optimum regions at $z \approx \pm 375$ nm but a maximum in the error at $z = 0$. This error decreases with increasing Δ while the outer optimum regions turn into regions of maximum error. It is seen that $\Delta = 400$ nm produces an axial error that is quite homogeneous over the full $z-$range while smaller and larger distances Δ lead to stronger variations. $\Delta = 600$ nm leads to an even smaller error of average ~ 30 nm on the support $z \in [-200\,\text{nm}, 200\,\text{nm}]$ and shall be used in the subsequent. Supports of such a length can generally be tolerated since axial localization is to be of highest quality directly at the center of symmetry $z = 0$ of the setup because the activation beam will possess a maximum at this place and consequently the majority of fluorophores will be detected in the plane $z = 0$.

Therefore, the canonical choice of $\Delta = 600$ nm might further be justified by studying the behavior of $\sigma_z(z = 0)$ as a function of Δ. Figure (3.4) demonstrates this for all parameters being the same as above. Obviously, σ_z approaches a broad minimum beginning at $\Delta \approx 600$ nm that extends up to higher values $\Delta \approx 900$ nm. Indeed, not too high a value of Δ is tolerable in order for radial localization to be affected minimally. Thence, the minimum Δ leading to a sufficiently constant σ_z has been chosen.

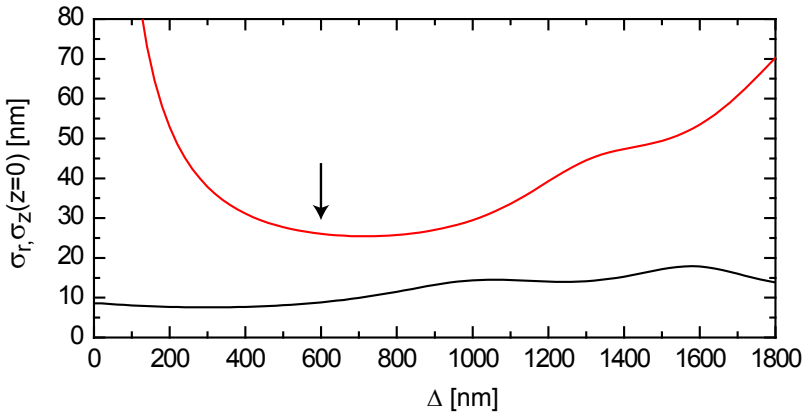

Figure 3.4: Cramer-Rao bounds σ_z (red) and σ_r (black) at z=0 for different camera distances Δ. Clearly, $\Delta = 600$ nm marks the onset of a broad minimum of σ_z that extends up to much higher distances $\Delta \approx 900$ nm. Nevertheless, the smallest Δ on this plateau is desirable in order not to affect the quality of radial localization.

The localization procedure depends crucially also on the number of observation planes. Figure (3.5) imparts the evaluation of the Cramer-Rao-bound of Equation (3.11) for a two and a three-channel setup with observance planes at ± 300 nm and $0, \pm 300$ nm, respectively. Both, radial and axial localization accuracies are in the range of a few tens of nm which is far below the Rayleigh length of typically several hundreds of nanometers. The Fisher information bounds still retain signatures of the differential properties of the underlying detection

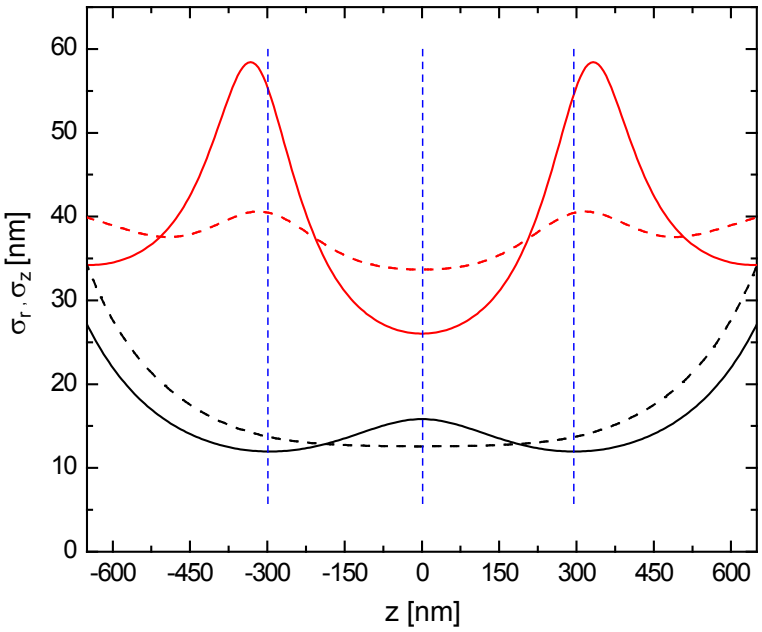

Figure 3.5: Cramer-Rao bounds $\sigma_r(z)$ for radial (black) and $\sigma_z(z)$ for axial localization (red) in the defocus setup for a mean photon number $\bar{N} = 250$ and a mean background $\bar{b} = 1$. The solid lines display the results for two acquisition planes at ± 300 nm while the dashed lines represent those for three imaging planes at $0, \pm 300$ nm (dashed blue vertical lines). Using a higher number of channels worsens the signal-to-noise ratio locally, but leads to a higher degree of homogeneity of the localization properties.

PSF. For instance, radial localization is optimum when one channel is maximally focused because then its image has reached the highest possible degree of radial contraction. In contrast, axial localization is of low quality at that point since the z-derivative of the detection PSF is zero at focus. Introducing a further acquisition plane at $z = 0$ nm decreases the signal-to-noise ratio but leads to a higher degree of homogeneity which is advantageous in terms of the translation symmetry of the estimation properties. The transition to full three-dimensional detection could be studied by introducing more and more image planes.

3.2.2 Astigmatism

The necessity of breaking the reflection symmetry of a standard detection PSF can also be achieved by means of an astigmatic aberration. In this case, just a single acquisition channel is already sufficient to retrieve the full axial information which is desirable as far as the signal-to-noise ratio is concerned. Nonetheless, this is gained at the expense of higher complexity in radial localization since x and y are no longer equivalent. Therefore, it presents itself to study also double-channel astigmatic imaging analogous to the preceding chapter. Such an approach restores not the rotational symmetry but the interchangeability of x and y if the second channel is rotated by $\pi/2$ with respect to the first.

Figure (3.6) delivers the expected mean single-channel images for two astigmatisms of different strengths. The axial position is encoded in the ratio of the horizontal width

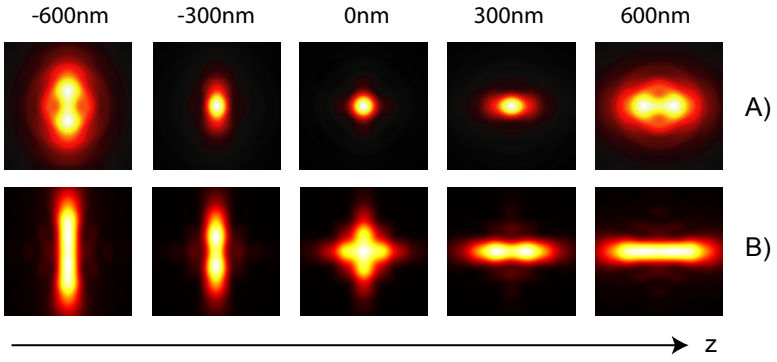

Figure 3.6: The expected mean images at different z for a weak (A) and a strong astigmatism (B) in a single-channel setup. The image is invariant under a reflection at the focal plane in conjunction with a rotation by $\pi/2$.

to the vertical one. These detection PSFs were calculated employing vectorial diffraction theory with a suitable apodization function. The diffraction integrals to be solved are a generalization of vectorial diffraction theory [56] and are defined in the appendix **A**. The aberration function $\Psi(\phi, \theta)$ for the case of astigmatism is [57]

$$\Psi(\phi, \theta) = \exp\left(iS\frac{1-\cos^2\theta}{\sin^2\alpha}\cos(2\phi)\right) \tag{3.12}$$

with α being the semi-aperture angle. The weak astigmatism (A) has strength $S = 1$, the strong one (B) $S = 3$. The S-values were adjusted to yield PSF widths typically for

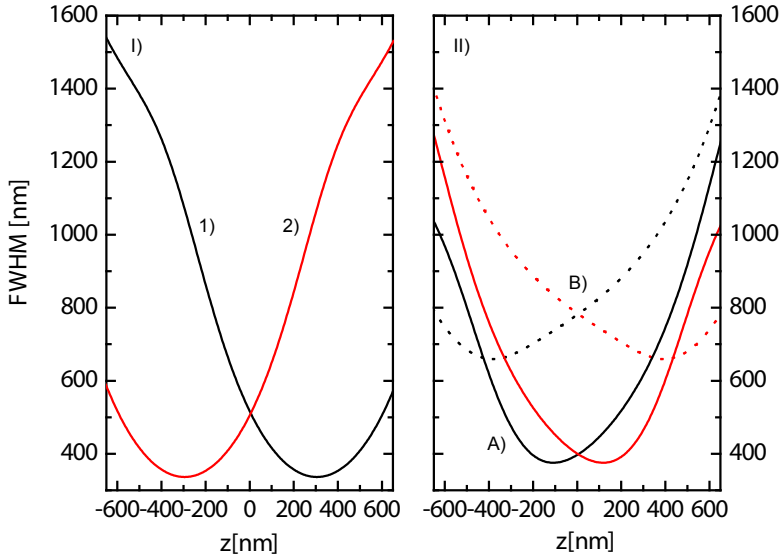

Figure 3.7: The FWHM of the PSF for double-channel defocus imaging (I) is strongly differing for channel 1 (black) and channel 2 (red) and amounts to 400 nm..1500 nm. Single-channel astigmatic imaging (II) uses a weak (A) or a strong astigmatism (B). The widths are in the same range as for the case of the defocus but the axial information is achieved at the cost of an asymmetry in x (black) and y (red).

fluorescence experiments. The behavior of the widths over the full z-support can be found in Figure (3.7, II) and might be compared to those of the defocus setup of the preceding paragraph. The fact that the differences in widths can amount up to several hundreds of nanometers constitute a realistic situation and have frequently been observed [34].

Radial and axial localization are discussed here in two steps. Figure (3.8) presents the Cramer-Rao bounds for axial localization for single-channel and double-channel astigmatic imaging for the weak and the strong astigmatism. Again, the photon number is $\bar{N} = 250$ with the background $\bar{b} = 1$. The defocus result is added to allow a quantitative comparison with the method of the preceding section. Apparently, the resulting precision curve $\sigma(z)$

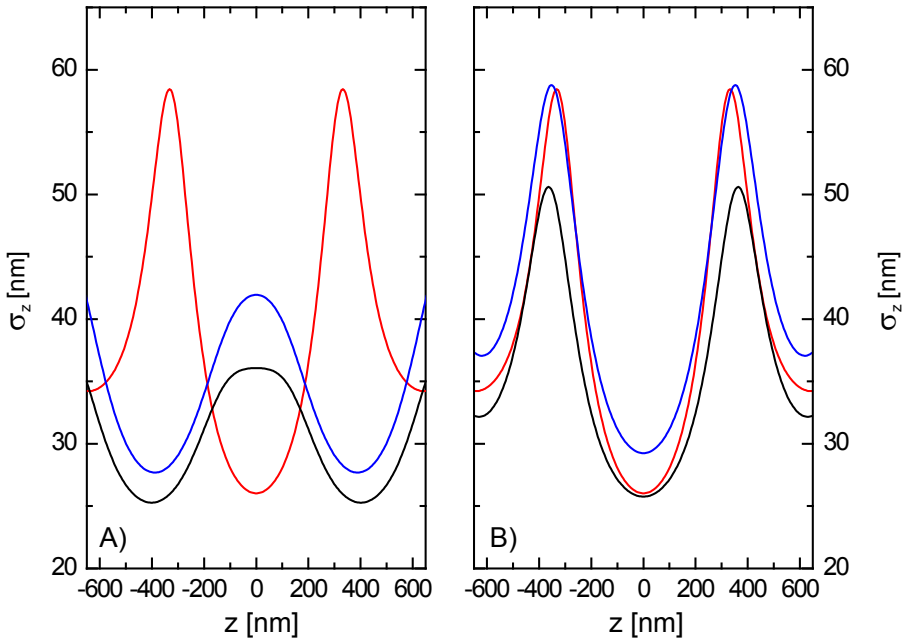

Figure 3.8: Cramer-Rao bound $\sigma_z(z)$ for axial localization using astigmatic aberrations for single-channel imaging (black) and double-channel imaging with a rotated second channel (blue). The results for the two-channel defocus setting are shown as a reference (red). (A) is the weak astigmatism, (B) the strong one. Of course, double-channel astigmatic position estimation is always inferior to single-channel astigmatic estimation because the signal is split among two channels and thereby the signal-to-noise ratio is lowered.

may have large similarity with the one for the defocus like in (B) or be related to it by a simple transformation, for instance a horizontal reflection, for (A). Nevertheless, the weak astigmatism is less optimal than the defocus in the interval $z \in [-200, 200]$ nm by a factor up to $1.4..1.6$ for both configurations, the single- and the two-channel setup. The weak astigmatism features better results at $\sim \pm 350$ nm but usually this is of minor use since radial localization is less precise here. The results for the strong astigmatism are only slightly different from the defocus.

Eventually, Figure (3.9) contributes the Cramer-Rao bounds for x and y. The weak astigmatism outperforms the defocus setting on the central interval $[-200, 200]$ nm in both, the one- and two-channel variants and in both, x and y. Relative amelioration amounts up to a factor of 1.5 at $z = 0$. Outside this interval the astigmatism leads to a strongly deteriorating

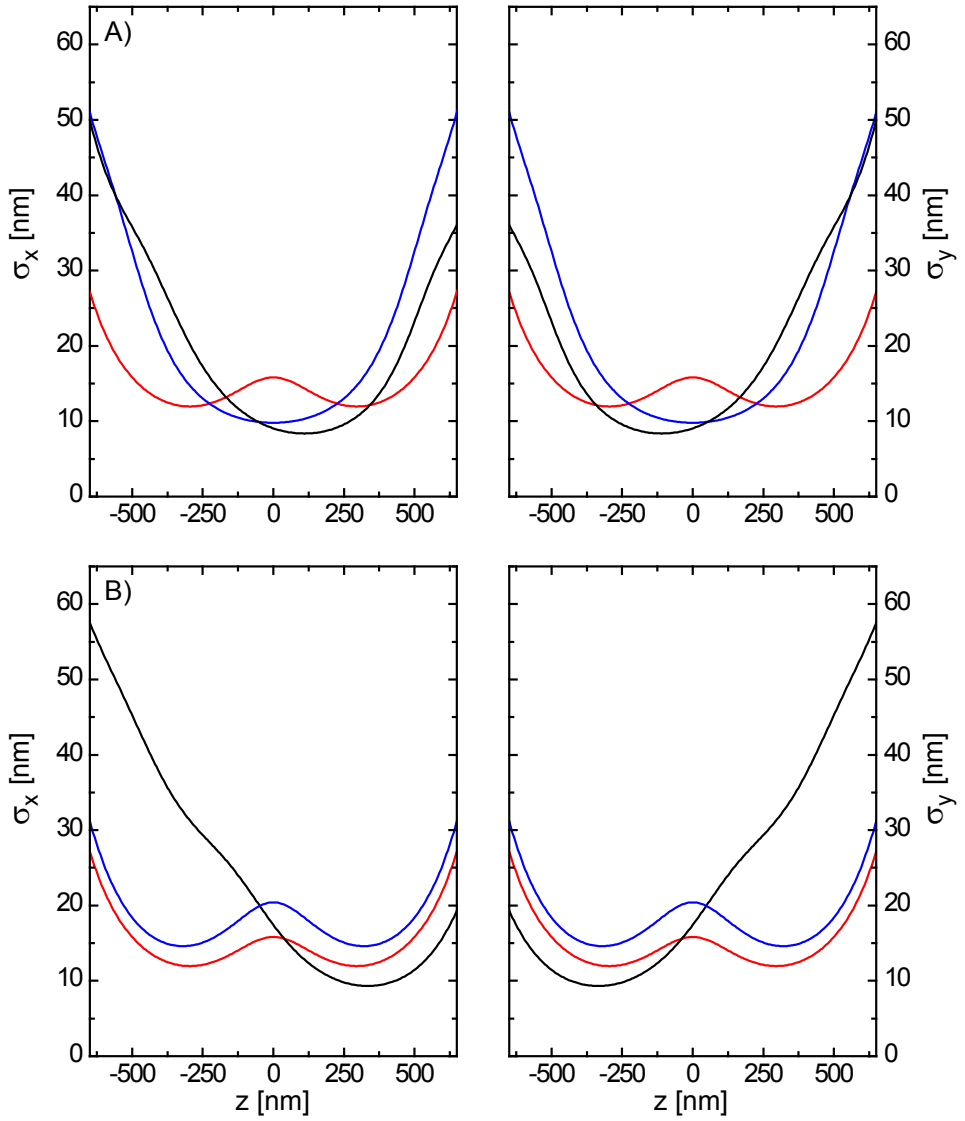

Figure 3.9: Cramer-Rao bounds $\sigma_x(z)$ and $\sigma_y(z)$ for radial localization using astigmatic aberrations for single-channel imaging (black) and double-channel imaging with a rotated second channel (blue). The results of the double-channel defocus setting (red) are added to facilitate a comparison. The first row originates from the weak astigmatism (A), the second from the strong one (B). x and y are displayed in the first and second column, respectively.

accuracy which is due to the strong beam divergence. This difference can grow up to a factor 2 at $z = \pm 400$ nm. In contrast, the strong astigmatism does never outperform the defocus in x and y at the same time. In summary, the astigmatic methods do not outperform the defocus setup in x, y or z at the same time.

3.2.3 4pi-Scheme

Conventional fluorescence microscopy collects merely the signal being radiated into one half of the solid angle. By using a second lens opposed to the other lens the signal photon number can be almost doubled [17]. Interfering the two beams with a 50%-50%-beamsplitter produces a constructive and a destructive detection PSF [84]. The constructive detection PSF is usually used for imaging and leads to images which are in need of deconvolution due to axial side-lobes. Fortunately, the central node has an axial width that is far below the one of confocal microscopy. Just as 4pi - microscopy was combined with STED microscopy to benefit from the advantages of both techniques [85] a combination of the 4pi-concept with PALMIRA can be aimed at. In the defocus and the astigmatism scheme radial position estimation was superior to axial in terms of precision. At this place, an extension of the 4pi-setup is presented that reverses this order and privileges axial localization over radial localization. The physical effects responsible are (i) the increased number of signal photons on the grounds of the enlarged solid angle and (ii) the strong z-modulation of the interference detection PSFs.

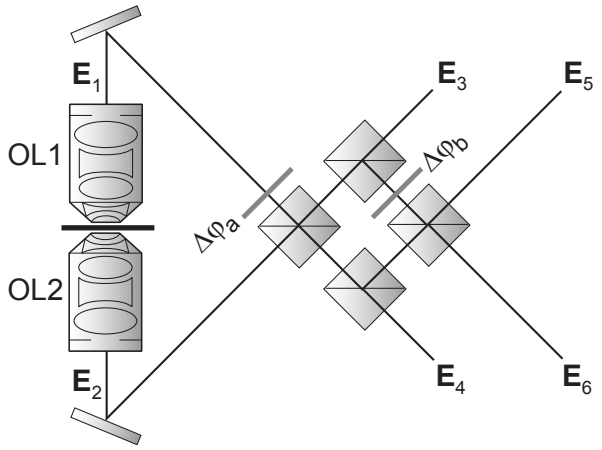

Figure 3.10: An extended 4pi-setup with four identical 50%-50%-beamsplitters and two phase retardation plates. $\Delta\varphi_b = \pi/2$ leads to interference patterns with a relative phase shift of $0, \pi/2, \pi$ and $3\pi/2$, respectively. The effective detection PSFs are so strongly dependent on z that unlike the preceding methods the precision of $z-$estimation becomes superior to $x, y-$estimation.

Figure (3.10)[3] displays the experimental setup. A standard 4pi-setup with two opposing objective lenses OL1, OL2 and one beam-splitter creates the usual constructive and destructive electric fields. These waves are then transmitted by another two beam-splitters leading to the measured fields $\mathbf{E}_3$ and $\mathbf{E}_4$. One of the reflected parts is retarded by $\Delta\varphi_b$ and then recombined with the second reflected wave. The constructive and deconstructive part of this process are measured as $\mathbf{E}_5$ and $\mathbf{E}_6$. The second beam-splitter and the associated channels $\mathbf{E}_5$ and $\mathbf{E}_6$ serve to eliminate undesired ambiguities in the determination of the z-position that still reside in the first two channels $\mathbf{E}_3$ and $\mathbf{E}_4$. A wave tracing analysis has to take into account that each reflection implies a phase shift of $\pi/2$ and that $\mathbf{E}_2$ arises from $\mathbf{E}_1$ by a

[3]The figure is courtesy of Dr. A. Schönle, MPI f. Biophysical Chemistry, Göttingen.

reflection at z, i.e.

$$\begin{aligned} E_{2,x}(x,y,z) &= E_{1,x}(x,y,-z) \\ E_{2,y}(x,y,z) &= E_{1,y}(x,y,-z) \\ E_{2,z}(x,y,z) &= -E_{1,z}(x,y,-z). \end{aligned} \tag{3.13}$$

After careful tracking of the phases the resulting field $\mathbf{E}$ may be compared to the standard form

$$\mathbf{E} = \mathbf{E}_1 + e^{i\Delta\varphi}\mathbf{E}_2 \tag{3.14}$$

to yield the following relative phase shifts

$$\begin{aligned} \Delta\varphi_3 &= \Delta\varphi_a + \pi/2 \\ \Delta\varphi_4 &= \Delta\varphi_a + 3\pi/2 \end{aligned} \tag{3.15}$$

and

$$\begin{aligned} e^{i\Delta\varphi_5} &= e^{i(\Delta\varphi_a - 3\pi/2)}\frac{e^{i\Delta\varphi_b}-1}{e^{i\Delta\varphi_b}+1} \\ e^{i\Delta\varphi_6} &= e^{i(\Delta\varphi_a + 3\pi/2)}\frac{1+e^{i\Delta\varphi_b}}{1-e^{i\Delta\varphi_b}}. \end{aligned} \tag{3.16}$$

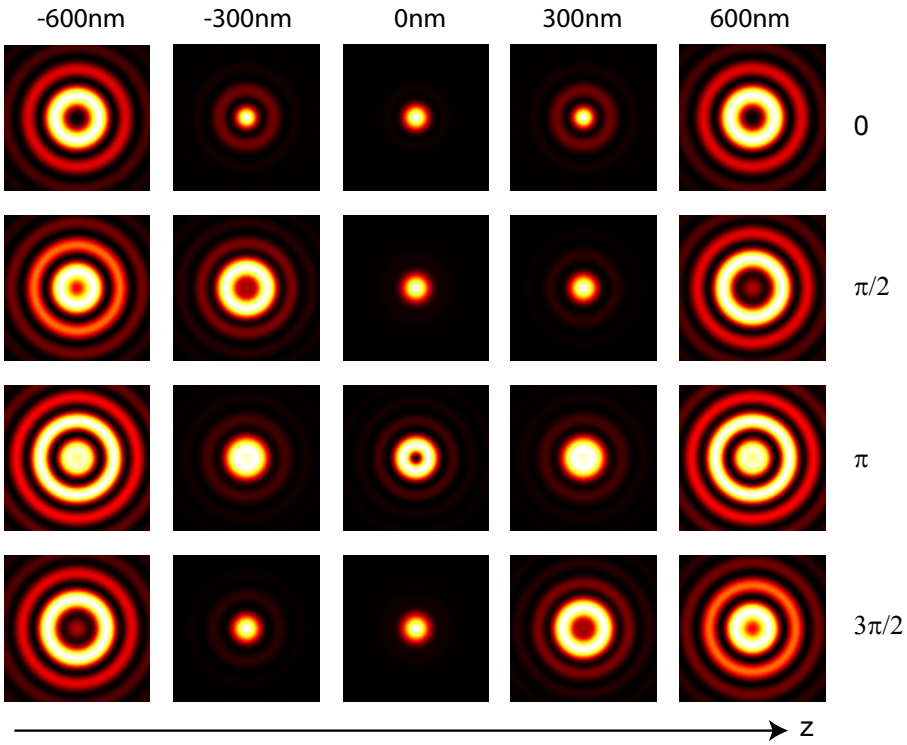

Figure 3.11: Axial cross sections for the 4-channel 4pi-setup with $\Delta\varphi_a = 0$ and $\Delta\varphi_b = \pi/2$. The channels with relative phase shifts 0 ($\mathbf{E}_6$) and π ($\mathbf{E}_5$) are symmetric with respect to $z = 0$. Those with relative phase shifts $\pi/2$ ($\mathbf{E}_3$) and $3\pi/2$ ($\mathbf{E}_4$) are symmetric if interchanged.

Setting $\Delta\varphi_b = \pi/2$ results in

$$\begin{aligned} \Delta\varphi_5 &= \Delta\varphi_a + \pi \\ \Delta\varphi_6 &= \Delta\varphi_a. \end{aligned} \tag{3.17}$$

From now on, $\Delta\varphi_a = 0$ is used. Figure (3.11) displays axial cross sections for the associated detection PSFs at different positions z. The four channels display strong variations and show symmetry properties under reflection at $z = 0$ and reflection combined with interchange if necessary. Meanwhile, Figure (3.12) reveals the behavior of the center of the four detection PSFs $h_i(x = y = 0, z)$ and the width $\omega_i(z) = 2\sqrt{2\ln 2}\sigma(z)$ over the full z-support. The multi-path interference leads to pronounced axial oscillations in a single detection PSF with a period of roughly $\lambda/2$. The maxima of the different channels are alternating during this period. Furthermore, for the double beam-splitter setup the first ambiguity in widths arises at $z_0 \sim \pm 120$ nm while it occurs already at $z = 0, \sim \pm 120$ nm for the one beam-splitter scheme where only channels 3 and 4 are used. The ambiguity in widths can jeopardize axial localization if - while featuring the same width - the shape of the PSF is not different enough. This is likely to be the case for the one beam-splitter setup

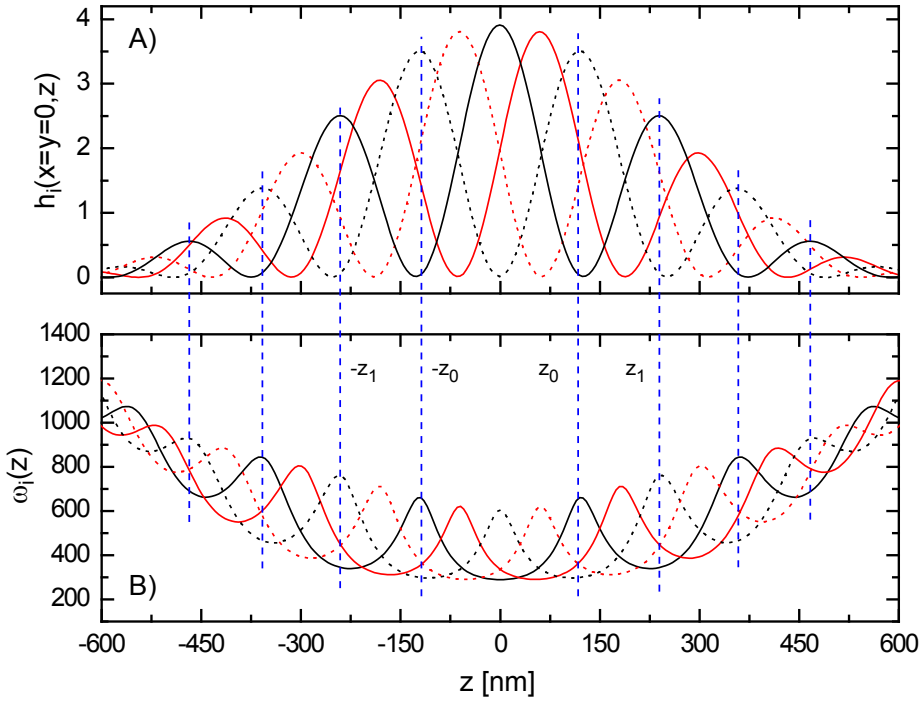

Figure 3.12: Full $z-$trace of the point $x = y = 0$ (A) and the widths $\omega_i(z)$ as a function of z (B) of the 4pi-setup with $\Delta\varphi_a = 0$ and $\Delta\varphi_b = \pi/2$. The different relative phase shifts are 0 (solid black), $\pi/2$ (solid red), π (dotted black) and $3\pi/2$ (dotted red). Characteristic oscillations are occurring. The different channels alternate in providing the maximum peak. The PSFs for the phase shifts 0 and π are symmetric with respect to $z = 0$, those for $\pi/2$ and $3\pi/2$ are invariant under reflection and interchange. Certain axial positions z_i lead to a full symmetry of the widths (dotted blue lines). Although the widths are equal at $\pm z_i$ the details of the PSFs $h_{\pi/2}$ and $h_{3\pi/2}$ may be different. This is seen by prolonging the lines to the upper graph: $h_{\pi/2}(x = y = 0, z = -z_i) = h_{3\pi/2}(x = y = 0, z = z_i)$.

at $z = 0, \sim \pm 120$ nm, see Figure (3.12, A, red lines). The double beam-splitter scheme supports two times larger a localization support. In addition, the two beam-splitter scheme avoids simultaneously vanishing derivatives of PSFs and widths, i.e. diverging local Cramer-Rao bounds.

Finally, Figure (3.13) renders an account for the Cramer-Rao bounds of the four-channel 4pi-setup in comparison to the two-channel defocus setting. It is taken into account that the photon number is doubled in the case of the 4pi-setting, $\bar{N} = 500$, whereas it stays $\bar{N} = 250$ for the case of the defocus. The background noise is assumed as usual $\bar{b} = 1$. The 4pi-setting outperforms the defocus in all categories : (i) radial localization is ameliorated by a factor of 2, (ii) axial localization rises even by a factor of 8 and (iii) radial and axial estimation feature a higher degree of spatial homogeneity. Already a standard 4pi scheme using merely a single beam-splitter would have yielded qualitatively related results but with a much lower degree of spatial homogeneity. In the 4pi scheme axial localization is privileged over radial localization.

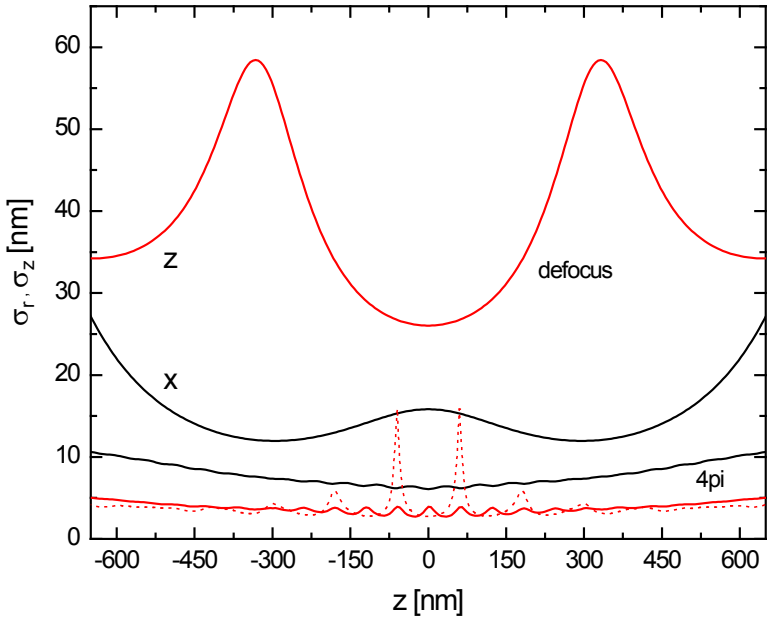

Figure 3.13: Cramer-Rao bounds $\sigma_{\mathrm{r}}(z)$ for radial (black) and $\sigma_{\mathrm{z}}(z)$ for axial localization (red) for the four-channel 4pi-setup with $\Delta\varphi_a = 0$ and $\Delta\varphi_b = \pi/2$. The photon number is $\bar{N} = 500$ for the 4pi-setting to account for the doubled solid angle whereas it stays $\bar{N} = 250$ for the defocus. The 4pi-setup not only achieves higher precision but leads to smaller spatial variations of the localization properties. A standard 4pi scheme featuring only one beam-splitter would yield two images with higher photon number but with much more strongly varying $\sigma_{\mathrm{z}}(z)$ (red, dotted).

3.2.4 Multi-Point Scheme

The methods presented so far rely on the use of a CCD camera. Typical CCD frame rates are of the order of 500 Hz. This implies a temporal resolution of about 2 ms [7, 8]. In contrast,

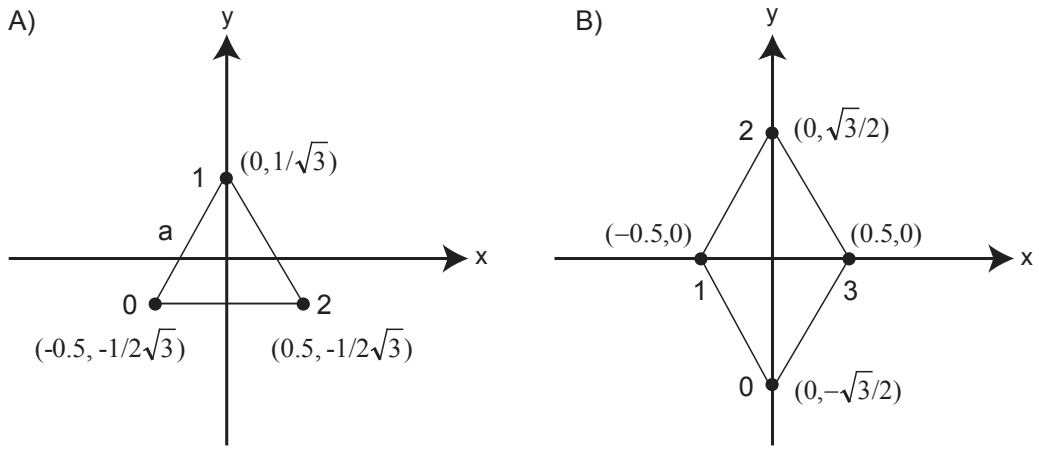

Figure 3.14: The arrangement of the APD foci for a multi-point setup. Both, an equilateral triangle (A) and a rhombus (B) are investigated. The center of mass coincides with the origin of the coordinate system. The only free parameter is the side-length a of the triangles (thin black lines) joining the APD positions. For simplicity, $a = 1$ has been chosen for this figure. A typical value is $a = 200$ nm.

APDs allow integration times in the range of a few tens of μs rendering possible a much higher temporal sampling. Here, the Fisherian method is used to study the precision of a position estimation setup consisting of an array of APDs. Figure (3.14) sketches the two versions of this setup. Arrangement of the APDs on an equilateral triangle (A) as well as on a rhombus (B) are considered. The coordinate system is chosen in such a way that its origin coincides with the center of mass of the detector setup. The single free geometric parameter is the side-length a which is typically $a = 200$ nm.

The statistical model for the photon signal on the different APDs is formulated in close analogy to Equations (3.1) and (3.5). The detection PSF $h_{\text{det}}(\mathbf{r})$ giving the probability to detect a photon emitted by a molecule at position $\mathbf{r}$ if the detection efficiency is 1 is assumed to be normalized in such a way that $h_{\text{det}}(\mathbf{r} = 0) = 1$. When denoting the APD centers by $\{\mathbf{r}_i\}, i = 1..A$, the detection efficiency by η, the mean background by $\bar{b}$ and the mean photon number by $\bar{N}$ the following mean signal number $\bar{N}_i(\mathbf{r})$ is obtained for the diode i

$$\bar{N}_i(\mathbf{r}) = \eta \bar{N} h_{\text{det}}(\mathbf{r} - \mathbf{r}_i) + \bar{b}. \tag{3.18}$$

In other words, the probability to be detected by APD i is $\eta h_{\text{det}}(\mathbf{r} - \mathbf{r}_i)$ while

$$p_{\text{ovs}}(\mathbf{r}, \eta) = 1 - \eta \sum_{i=1}^{A} h_{\text{det}}(\mathbf{r} - \mathbf{r}_i) \tag{3.19}$$

is the probability to be overlooked by all APDs. To avoid unphysical results η has to be chosen in such a way to guarantee $p_{\text{ovs}}(\mathbf{r}, \eta) \geq 0$ for all $\mathbf{r}$. The Fisher information for $x-$localization with all other parameters given reads

$$I(x) = \sum_{i=1}^{A} \frac{\left[\eta \bar{N} \partial_x h_{\text{det}}(\mathbf{r} - \mathbf{r}_i)\right]^2}{\eta \bar{N} h_{\text{det}}(\mathbf{r} - \mathbf{r}_i) + \bar{b}}. \tag{3.20}$$

Here, it has been assumed that the number of photo-electrons fluctuates with the number of photons being radiated onto the detector according to a Poisson distribution[4]. The expression for y is analogous. Equation (3.20) displays a pronounced dependency on the radial coordinates x, y. This is different from Equation (3.11) where a large image support leads to translational invariance of $I(x)$ in the radial direction.

The setup, Figure (3.14), suggests that the mean photon signals $\bar{\mathbf{N}}(\mathbf{r})$ allow for a one-to-one determination of the molecule position $\mathbf{r}$. However, the number A of recording sites is quite low. Usually, small numbers of observables frequently lead to ambiguities due to an inherent reduction in dimensionality. To show that such ambiguities can effectively be excluded here the vector of signals $\bar{\mathbf{N}}(\mathbf{r})$ is investigated in the following.

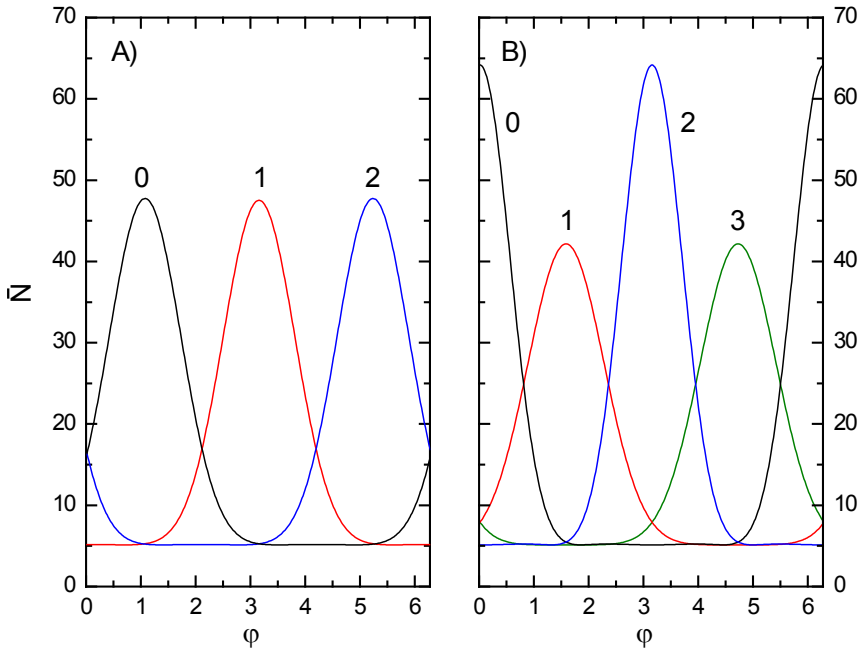

Figure 3.15: Signal traces $\bar{\mathbf{N}}(\varphi)$ as a function of the polar angle φ for a molecule moving along a circle of radius $r = 200$ nm with $\eta = 0.25$, $\bar{N} = 250$ and $\bar{b} = 4$. Both, for the three-point (A) and for the four-point-setup (B) these traces display an unambiguous mapping between the particle position $\mathbf{r}(\varphi)$ and the signals $\bar{\mathbf{N}}(\varphi)$.

Figure (3.15) displays the signal vector $\bar{\mathbf{N}}(\mathbf{r})$ as a function of the polar angle φ for a molecule located on a circle of radius $r = 200$ nm. The trajectory starts at $\mathbf{r} = (0, -r)$ and then proceeds counter-clockwise. The total signal is distributed in a characteristic manner among the detectors. The 3-fold rotational symmetry of (A) and the double mirror-symmetry of (B) is evident. Detectors that can be mapped onto each other by these operations are called equivalent. Equivalent detectors feature signals with similar but shifted functional shape over φ. For instance, the graph belonging to configuration (A) display a uniform relative shift of $2\pi/3$. The assignment $\mathbf{r}(\varphi) \mapsto \bar{\mathbf{N}}(\varphi)$ is always invertible to yield an angle φ from

[4] A multinomial image model can be discussed on the same grounds as in connection with Equation (3.11).

observations $\mathbf{N}$. This fact remains valid as long as the radius is not too large in order to collect enough signal in all detectors.

While Figure (3.15) showed that ambiguities do not occur in the polar angle φ Figure (3.16) clarifies this point as far as the radius r is concerned. The figure presents the implicit plots of the mean signal curve $\bar{\mathbf{N}}(\varphi)$ of the three-point setup for two circles of radii $r = 100$ nm and $r = 200$ nm, respectively. The angle φ now plays the role of the arc-length of the signal curves. The total signal is gradually shifted from one detector to the other. Larger radii decrease the maximum signal experienced by a single detector. In the limit of $r \to \infty$ the curves shrink down to the point $(\bar{b}, \bar{b}, \bar{b})$ and no inference about the position is possible any more. Nonetheless, the curves do not intersect for different radii as long as they are in a range of a few hundreds of nanometers. This behavior may be reproduced for different combinations of two radii.

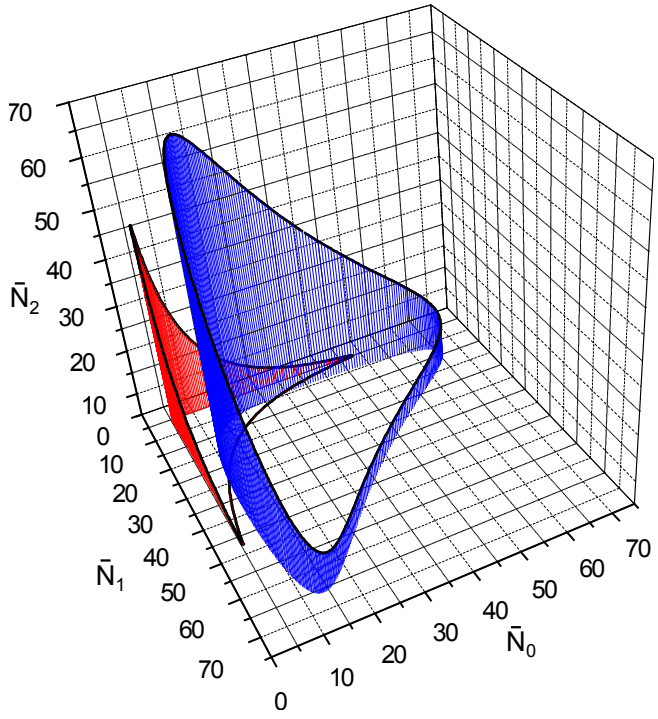

Figure 3.16: Implicit signal traces $\bar{\mathbf{N}}(\varphi)$ for the three-point setup with particle trajectories on circles of radii $r = 100$ nm (blue) and $r = 200$ nm (red), respectively. Optical parameters are $\eta = 0.25$, $\bar{N} = 250$ and $\bar{b} = 4$, as above. Naturally, the small radius features a higher maximum signal since the particle gets closer to the detector centers in this case. For $r \to \infty$ the trajectories shrink to the point $(\bar{b}, \bar{b}, \bar{b})$ and no inference about the position is possible any more. For moderately small radii of the order of some hundreds of nm both curves do not intersect. Estimation of the position from intensity vectors is then free from global ambiguities.

Once it is established that the measured observables do not have global ambiguities the local behavior of the likelihood function may be studied using the Fisher information expression of Equation (3.20). Figure (3.17) displays two-dimensional plots of the Fisher information $I_x(x, y)$ and $I_y(x, y)$ associated with $x-$ and $y-$localization, respectively. The optical parameters $\eta = 0.25$, $\bar{N} = 250$ and $\bar{b} = 4$ have been chosen as before. The highest values of $I_i(x, y)$ correspond to Cramer-Rao bounds for any unbiased position estimator of

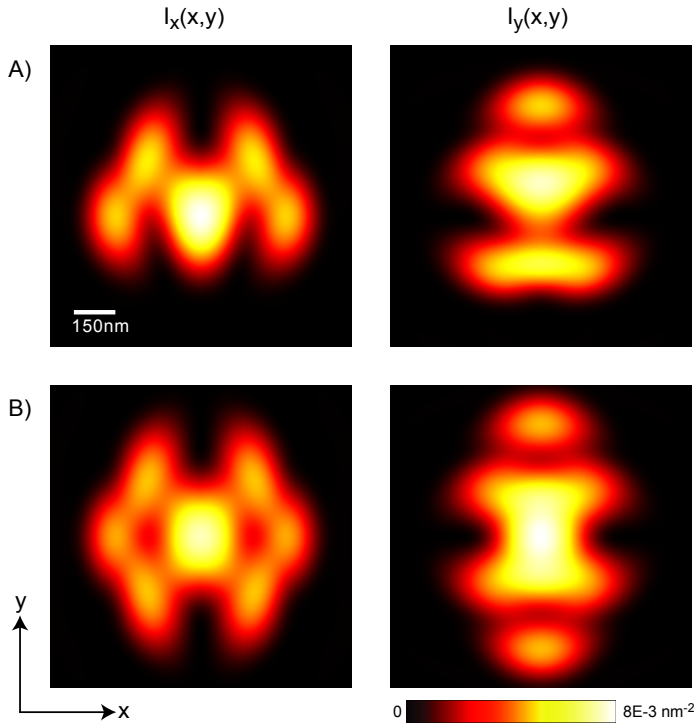

Figure 3.17: Fisher information $I_x(x, y)$ and $I_y(x, y)$ for $x-$ and $y-$localization, respectively, for the three-point (A) and the four-point-setup (B). The photon number is $\bar{N} = 250$, the mean background noise of the individual detector $\bar{b} = 4$. The detection efficiency directly at the center of the detection spot is $\eta = 0.25$. At maximum, the information reaches values corresponding to a Cramer-Rao bound of 12 nm. The functions I_x and I_y show pronounced spatial variations with a distinct area of high estimation quality in the center of the scheme (yellow, white). It is worth noting that the Fisher information can approach extraordinarily low values close to detector sites as for instance for I_x (A). This implies that unbiased estimators may have diverging variance although the fluorophore might be located close to a detector. Nonetheless, biased position estimators may exist in this region that have a variance well below the Cramer-Rao bound of the unbiased estimators.

about 12 nm. Irrespective of x or y and of the setup (A) or (B) the Fisher information fields feature regions of high localization quality before decaying of to zero. The comparatively low number of detectors leads to strong spatial variations with a characteristic geometric shape. This was not the case for the CCD camera setups of the preceding chapters because the number of pixels was high and edge effects could be neglected. It is worth noting that the Fisher information value may be extraordinarily small even close to detector sites as for instance for $I_x(x, y)$ of the three-point setup. This leads to very large error bounds. At a first sight, this is astonishing since high-quality estimation is expected at positions of maximum signal-to-noise ratio. Nevertheless, these two points are compatible when being aware of the fact that biased estimators need not follow the unbiased Cramer-Rao bound of Equation (3.7). Although there need not exist unbiased estimators meeting the expected accuracy, there may exist biased estimators with variances well below the unbiased Cramer-Rao bound. Quantitatively, if an estimator $\hat{x}$ of the parameter x with bias $b(x)$

$$b(x) := \langle \hat{x} \rangle - x \tag{3.21}$$

is given the Cramer-Rao inequality replacing Equation (3.7) becomes [78, 80]

$$\mathrm{var}[\hat{x}] \geq \frac{[1+b'(x)]^2}{I(x)} + b^2(x). \tag{3.22}$$

Thus, a biased estimator mapping a certain interval of the parameter x on a constant value x_0 can even decouple the variance from Fisher information since $b(x) = x_0 - x$ is obtained for this case. The variance is then simply $(x_0 - x)^2$. In other words, if the unbiased Cramer-Rao bound is super-parabolic in an important region it is always good practice to stick to a simple biased interval estimator instead of using a necessarily bad unbiased point estimator. For the Fisher information $I(x)$ and an unbiased estimator given an artificial bias could be designed by optimizing a parameterized bias function $b(x, p)$ over the parameter p to yield a minimum mean biased Cramer-Rao bound in an interesting interval.

Nevertheless, unbiased estimators too may have Cramer-Rao bounds as small as those for the estimation settings of the preceding chapters in large regions of space. Figure (3.18) displays the behavior of the unbiased Cramer-Rao bounds $\sigma_x(\mathbf{r})$ and $\sigma_y(\mathbf{r})$ for both radial directions along the abscissa $(x; y = 0)$ and the ordinate $(x = 0; y)$. Unbiased radial position estimation in the multi-point setup is capable of yielding the same maximum quality as for the defocus and astigmatism settings.

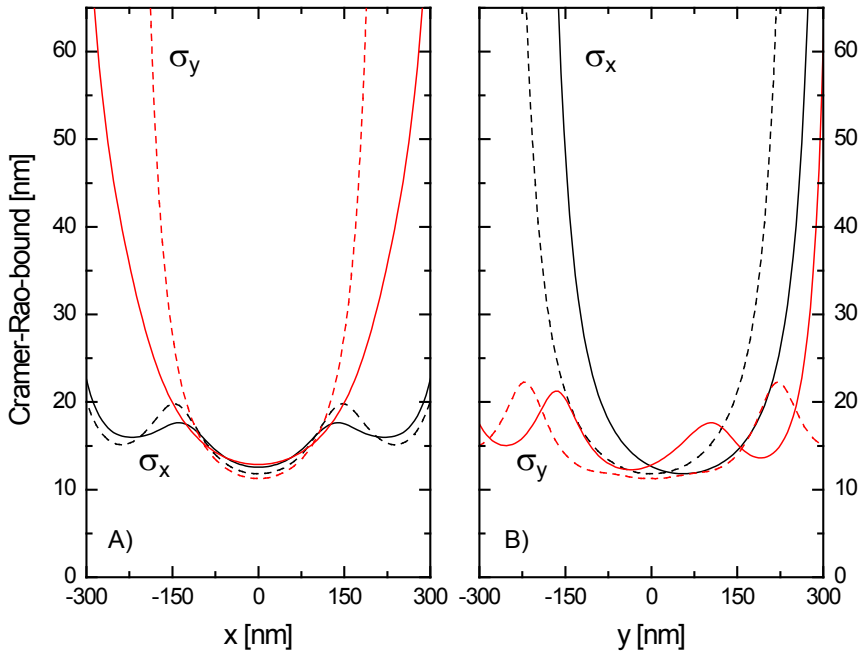

Figure 3.18: One-dimensional Cramer-Rao-bounds $\sigma_x(\mathbf{r})$ (black) and $\sigma_y(\mathbf{r})$ (red) for $x-$ and $y-$localization for the three-point (solid) and the four-point-setup (dashed) along the abscissa $(x; y = 0)$ (A) and the ordinate $(x = 0; y)$ (B). It is seen that the unbiased position estimators of the multi-point setup feature errors in center regions as low as those of the defocus or astigmatism methods of the preceding chapters.

3.3 Performance under Realistic Conditions

Once, the Cramer-Rao bounds have been established it is of interest to assess the question whether estimators employed with stochastic data are efficient, i.e. reach this bound. Furthermore, a systematic bias should be revealed. For an estimator given in explicit form the mean and variance might be calculated directly by evaluating expectation values. For estimators that are known merely implicitly a Monte-Carlo simulation could be employed to generate a sufficiently large amount of samples from which the interesting values are estimated. Here, the maximum-likelihood estimator is studied. Since the radial dimensions have been studied thoroughly in the literature [69] the focus is put on the axial coordinate z.

3.3.1 Defocus-, Astigmatism- and 4pi-Scheme

The channel methods using defocused, astigmatic or 4pi-imaging are characterized by the fact that a huge number of measured values, the pixel photon numbers $\{N_{ij}^l\}$ per channel l, are reduced to a few three-dimensional position vectors. Typical reduction factors are of the order of 20, i.e. about 60 measured values are compressed into a single three-dimensional position. This makes it complicated to state the maximum-likelihood estimator [86] explicitly. Given the position $\mathbf{r}$ Equation (3.5) states the probability to measure $\{N_{ij}^l\}$. The associated maximum-likelihood estimate $\hat{\mathbf{r}}_{\text{ML}}\left(\{N_{ij}^l\}\right)$ is the solution of the equation

$$\nabla_{\mathbf{r}} p\left(\{N_{ij}^l\}|\mathbf{r}\right) = 0. \tag{3.23}$$

Defocus- and Astigmatism-Scheme

Figure (3.19) delivers a plot of the mean axial position estimated from a Monte-Carlo simulation using the maximum-likelihood estimator. The mean photon number is $\bar{N} = 250$ and the background noise amounts to $\bar{b} = 1.\,400$ image samples per $z-$position featuring a camera pixel size of 100 nm are drawn from the image model in Equation (3.5). The axial

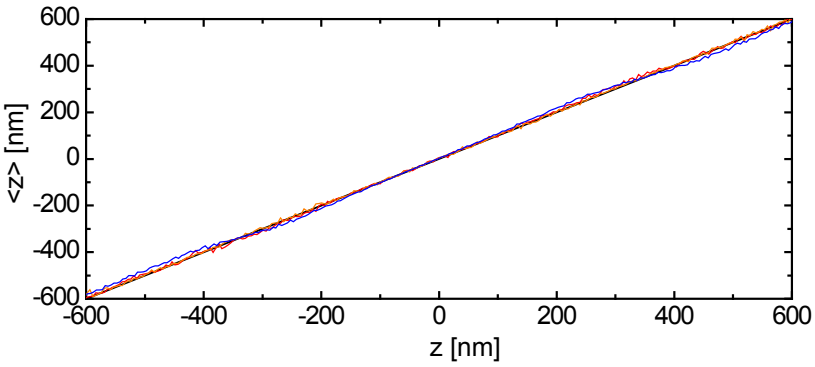

Figure 3.19: Mean axial position $\langle z \rangle$ for the defocus (blue)as well as the weak (orange) and strong astigmatisms (red) estimated from a Monte-Carlo simulation making use of a maximum-likelihood fit. The methods are bias-free and fluctuate merely marginally around the line through origin (black).

sampling step is $\Delta z = 5$ nm. The maximum-likelihood estimate is determined by employing a Nelder-Mead downhill simplex algorithm [87] to search the minimum in Equation (3.5). The fit iteration is initialized radially with the xy-maximum pixel $(xy)_{\text{max}}$ and axially with the maximum position z_{max} of the likelihood with $(xy) = (xy)_{\text{max}}$ fixed. The stopping criterium of the fit is to feature consecutively more than 10 times a spatial step < 5 nm. The defocus and astigmatism methods are bias-free.

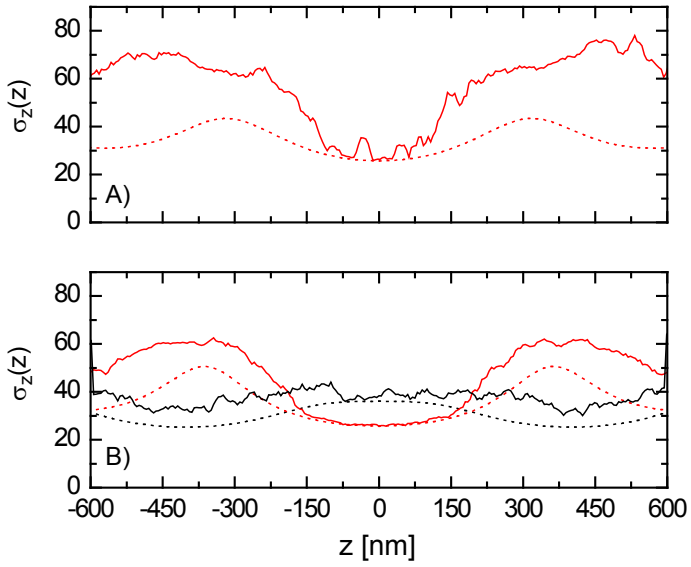

Figure 3.20: Standard deviation $\sigma_z(z)$ for the defocus (A), the weak (black) and strong (red) astigmatisms (B) retrieved from a Monte-Carlo simulation for the maximum-likelihood fit. The unbiased Cramer-Rao bounds calculated in the preceding chapter are shown as dotted lines. For the case of defocus as well as astigmatism the standard deviation approaches the Cramer-Rao bound in the central region.

Figure (3.20) states the standard deviations estimated from the same simulation. The unbiased estimators for the defocus and the astigmatisms are bound from below by the Cramer-Rao bound and even almost approach this bound in the central region around $z = 0$. In the outer regions the experimental error is approximately twice as large as the Cramer-Rao bound.

4pi-Scheme

For the 4pi setting the mean photon number amounts to $\bar{N} = 500$ to account for the larger aperture. All other values remain the same as in the foregoing paragraph. Figure (3.21) shows the distribution $p(\hat{z}|z)$ of estimated axial positions $\hat{z}$ conditional on the original position z. The pixel-size is 5 nm, i.e. a value of p in the histogram means that a fraction p of all estimates falls into a single interval of size 5 nm. The method is bias-free and of small standard deviation, $\sigma_z(z) \approx 7$ nm, for all axial values. This is mirrored by the similarity of the distribution to the line through origin and its width of $1 - 2$ pixels. No anomalies occur at the positions $\pm z_i$ of global ambiguities of the widths in Figure (3.12).

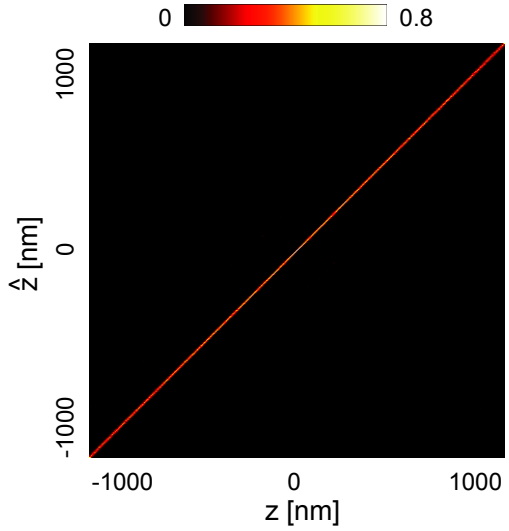

Figure 3.21: The distribution $p(\hat{z}|z)$ for the retrieved values $\hat{z}$ as a function of the original axial coordinate z in the 4pi-setup. The distribution is a bisecting line through origin. Thus, the method is bias-free at all positions. The width of the line describes the variance and leads to values $\sigma_z \approx 7$ nm. This is in accordance with the Cramer-Rao results of Figure (3.13) where the lower bound was $\sigma_z \approx 5$ nm.

3.3.2 Multi-Point-Scheme

The multi-point setup has been studied theoretically in section (3.2.4). Here, a lookup-estimator to retrieve the particle position is evaluated experimentally for the three-point setup of Figure (3.14, A). The estimator assumes the mean photon number $\bar{N}$ and the background $\bar{b}$ to be known. Then, the mean signals $\bar{N}_i(\mathbf{r})$ follow from Equation (3.18) and the position $\mathbf{r}$ is estimated using the estimator

$$\hat{\mathbf{r}}(\{N_i\}) = \min\left\{\sum_{i=1}^{3}\left[N_i - \bar{N}_i(\mathbf{r})\right]^2\right\}. \tag{3.24}$$

Thus, the position $\mathbf{r}$ leading to a minimal euclidian distance between the measured vector $\boldsymbol{N}$ and the expected one $\bar{\boldsymbol{N}}(\mathbf{r})$ is taken. Figure (3.22) presents the results of Monte-Carlo simulation. The bias $b_{\mathrm{x}}(x,y)$ and $b_{\mathrm{y}}(x,y)$ as well as the standard deviations $\sigma_{\mathrm{x}}(x,y)$ and $\sigma_{\mathrm{y}}(x,y)$ are estimated from 400 samples at each $(x,y)-$position on a $150 \times 150-$ grid of length 750 nm $\times$ 750 nm. As before the mean photon number is $\bar{N} = 250$, the detection efficiency $\eta = 0.25$ and the background noise $\bar{b} = 1$. Events with too low a photon number are excluded by accepting only events with $[N_1 N_2 N_3]^{\frac{1}{3}} > 7.5$. The bias is of considerable strength. In regions where a significant signal is distributed among the three channels the bias carries signatures of the geometry of the setup. In regions where the setup records almost exclusively noise the bias is almost linear in x or y. This results from the fact that in this case every position is equally probable. The standard deviation reaches minimum values as low as 20 nm directly at the detector centers and grows cylindrically around these positions. As an effect of the strong systematic bias the similarity with the Fisher results of Figure (3.17) vanishes almost entirely. This is possible and reflects the lower bound in Equation (3.22).

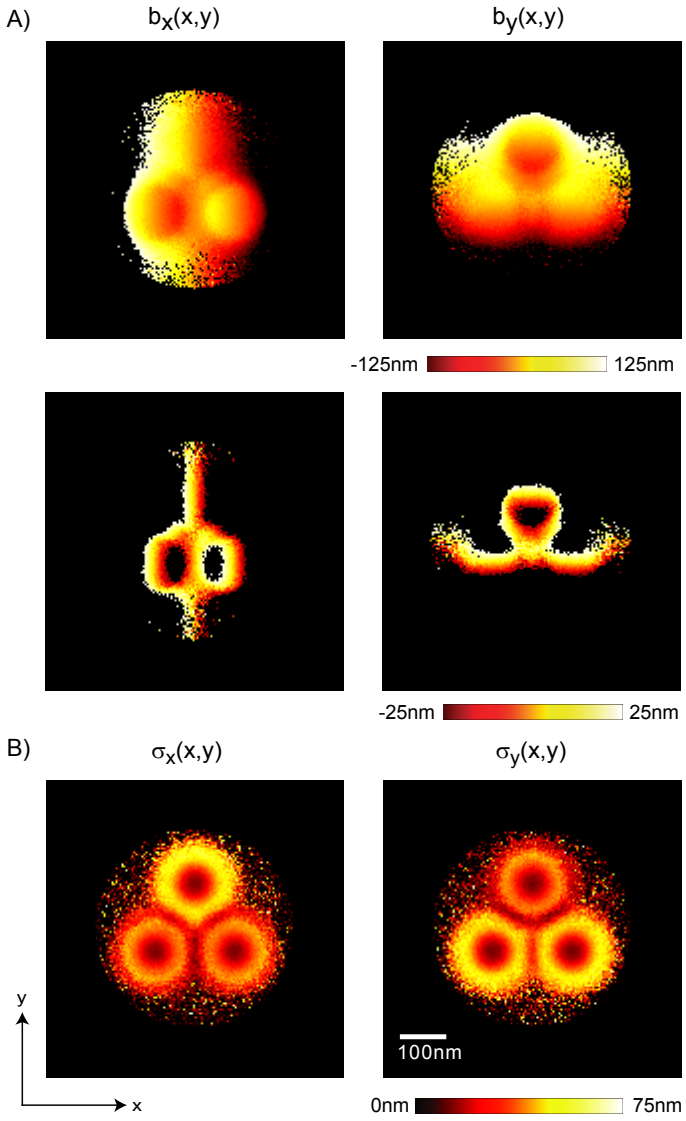

Figure 3.22: Bias $b_x(x,y), b_y(x,y)$ (A) and standard deviations $\sigma_x(x,y), \sigma_y(x,y)$ (B) estimated during a Monte-Carlo simulation of the three-point setup, see Figure (3.14, A). Signal triples yielding geometric mean values smaller than 7.5 have been discarded to exclude events consisting purely of noise. The bias is shown for different color maps: the first row $[-125\text{ nm}, 125\text{ nm}]$ to investigate the overall behavior, the second row $[-25\text{ nm}, 25\text{ nm}]$ to indicate the regions of comparatively small bias. The standard deviations show clear signatures of the geometry of the setup and are on the order of 50 nm with minima of 20 nm directly at the APD foci. This might compared to the Fisher results of Figures (3.17) and (3.18).

4 Recognition of Objects

The analysis in the preceding chapters has assumed that recognition of molecules is ideal in the sense that fluorophores featuring a signal photon number higher than a certain threshold are detected while those with lower signals are rejected with absolute certainty. Real experiments are more complicated. Pixel thresholding methods create a spatial dependency of the effective threshold value. Background noise contributing to measured signals can be over- or underestimated leading to events misinterpreted or overseen, respectively. Spatial overlap of optical patterns originating from more than one fluorophore puts into question a single-molecule analysis. Furthermore, molecules can be detected depending not only on their position but also either on the spectral properties or of the state of polarization of their fluorescence. Thereby, further degrees of freedom are added to the analysis.

4.1 Pixel Thresholding and the Homogeneity of Space

All scanning methods require a systematic sampling of the space in which the molecules can reside. In standard scanning confocal microscopy the total number of excitation cycles is kept constant for each pixel. In principle, similar ensembles of molecules in different pixels should give similar signals. Although scanning is stochastic for PALMIRA it is no exception to this rule. A molecule in different regions of space has to be given the chance to be recognized equally frequently in order for the PALMIRA image to be a credible estimate of the real fluorophore distribution.

Generally, one has to distinguish between the value of the maximum pixel of the optical pattern and its total photon content. In two dimensions these two values are merely multiples of each other. In three dimensions a fixed pixel threshold T_{pix} leads to a $z-$dependent total event threshold $T_{\text{tot}}(z)$. Their relationship is dictated by the detection PSF $h(\mathbf{r})$,

$$\frac{T_{\text{pix}}}{T_{\text{tot}}(z)} = \frac{h_{\text{max}}(z)}{\sum_{ij} h_{ij}(z)} = h_{\text{max}}(z). \tag{4.1}$$

The last equality sign holds since the PSF is normalized radially for all z. If the pixel threshold T_{pix} is given the effective event threshold reads

$$T_{\text{tot}}(z) = \frac{T_{\text{pix}}}{h_{\text{max}}(z)}. \tag{4.2}$$

For the case of the two-channel defocus setup an event is considered to be relevant if in at least one of the two frames the maximum signal exceeds $T_{\text{pix}}/2$. Using half the value

takes into account that the signal is split equally among the two channels. For two channels focused at $z = \pm z_0$ the effective total threshold follows as

$$T_{\text{tot}}(z) = \min\left\{\frac{T_{\text{pix}}}{h_{\text{max}}(z - z_0)}, \frac{T_{\text{pix}}}{h_{\text{max}}(z + z_0)}\right\}. \tag{4.3}$$

Figure (4.1, A) displays a plot of Equations (4.2) and (4.3) for the two-channel defocus and the single-channel weak astigmatism setup of Chapter (2.2). The event threshold is considerably homogeneous in intervals of length 200-300 nm for the single-channel setup but displays stronger variations for the two-channel scheme. Consequently, a space-dependent event threshold leads to a space-dependent probability of recognizing an object. Figure (4.1,

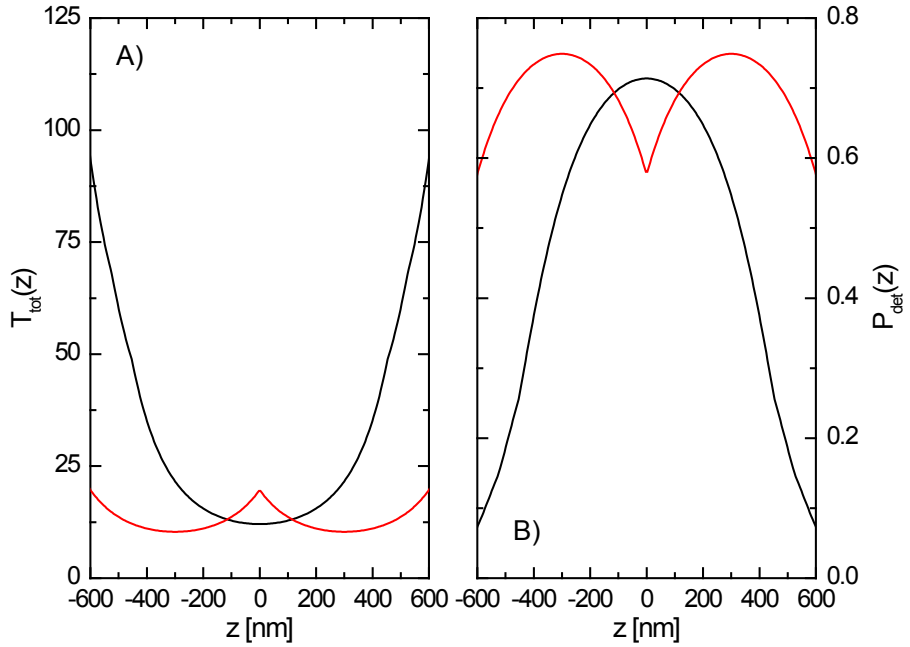

Figure 4.1: (A) Effective event threshold $T_{\text{tot}}(z)$ as a function of z for the two-channel defocus (red) and the one-channel weak astigmatism setup (black). For the two-channel setup an event is considered to be relevant if on at least one of the two images a pixel exceeds $T_{\text{pix}}/2$. The pixel threshold T_{pix} has been set to unity. To resolve all regions statistically equitably it is necessary to introduce $z-$scanning or to correct numerically for the space-dependent detection probability. (B) Effective detection probability $P_{\text{det}}(z)$ as a function of z for the two-channel defocus (red) and the one-channel weak astigmatism setup (black). The photon number is $\bar{N} = 250$ and the pixel threshold $T_{\text{pix}} = 7$. The maximum detection probability for the defocus is slightly larger than the one for the astigmatism since the astigmatism is never focused as tightly as the non-aberrated detection PSF and therefore the detection PSF features a higher center value in the focal plane.

B) gives the associated detection probability $P_{\text{det}}(z)$ for a mean photon value $\bar{N} = 250$ and a pixel threshold $T_{\text{pix}} = 7$. It is given by the probability to exceed the threshold in the current image frame. Here it has been assumed that the detection process is ideal in the sense that all photons are observed and no noise is present. Then simply Equation (2.26), the cumulative sum of the geometric distribution, might be used. As has already been obvious from Figure (4.2) scanning schemes have to be employed if regions where the threshold is position-dependent in a relevant way are to be mapped.

4.2 Recognition in the Presence of Noise

The interplay of the real photon signal and the background noise leads to subtle effects. Background noise may be underestimated which causes objects being recognized although they have a signal photon number below the threshold. Poisson fluctuations of the number of measured photo-electrons can yield values lower values than the actual signal which leads to objects being omitted though their signal exceeds the photon threshold. Here, these effects are studied quantitatively. It is assumed that thresholding is a one-stage process during recognition and that no secondary thresholding rejecting already recognized events according to their total photon number takes place. The discussion is restricted to two dimensions. Hence, the event threshold T_{tot} and the pixel threshold T_{pix} may be used interchangeably. The standard detection PSF used in this thesis and an assumed pixel size of 100 nm x 100 nm induces that $T_{\text{tot}} \approx 10 T_{\text{pix}}$.

In a first instance, the probability $P_{\text{det}}(N, T_{\text{pix}}, \bar{b})$ for an event to be detected as a function of the the photon number N, the threshold T_{pix} and the background value $\bar{b}$ is studied. So far, this has been assumed to be a simple step function,

$$P_{\text{ideal}}(N, T_{\text{tot}}, \bar{b}) = \begin{cases} 0 & N < T_{\text{tot}} \\ 1 & N \geq T_{\text{tot}} \end{cases} . \tag{4.4}$$

Now, this ideal shape is altered due to the effects of the background noise. The problem is equivalent to inquiring for the probability for all pixels being below the threshold T_{pix} because one and only one pixel above threshold suffices for an event to be recognized. Hence, the probability to be detected, $P_{\text{det}}(N, T_{\text{pix}}, \bar{b})$, reads

$$P_{\text{det}}(N, T_{\text{pix}}, \bar{b}) = 1 - \prod_{ij} \left(\sum_{N_{ij}=0}^{T_{\text{pix}}-1} p(N_{ij} | \bar{N}_{ij}) \right) , \tag{4.5}$$

with the mean value of pixel ij

$$\bar{N}_{ij} = N h_{ij} + \bar{b} \tag{4.6}$$

and $p(N_{ij}|\bar{N}_{ij})$ being a Poisson distribution. The cumulation in Equation (4.5) may be simplified further to obtain

$$P_{\text{det}}(N, T_{\text{pix}}, \bar{b}) = 1 - \prod_{ij} Q(T_{\text{pix}}, \bar{N}_{ij}) \tag{4.7}$$

with the cumulative Poisson distribution

$$Q(T, \lambda) = \frac{\Gamma(T+1, \lambda)}{T!}. \tag{4.8}$$

$\Gamma(T+1, \lambda)$ denotes the generalized gamma function defined as [52, 88]

$$\Gamma(T+1, \lambda) = \int_{\lambda}^{\infty} t^T e^{-t} dt. \tag{4.9}$$

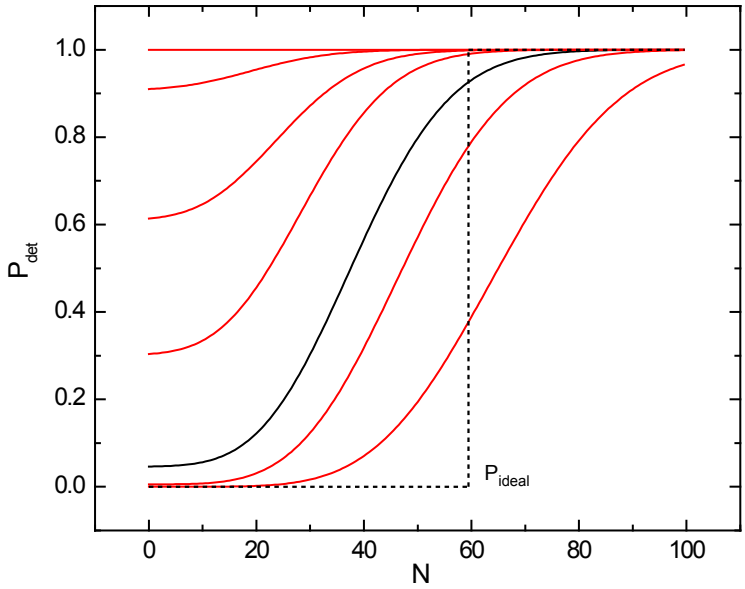

Figure 4.2: Detection probability $P_{\text{det}}(N, T_{\text{tot}}, \bar{b})$ according to Equation (4.7) in the presence of pixel background noise $\bar{b} = 1$ as a function of the photon number N for different values of the total photon threshold $T_{\text{tot}} = 20, 40, 45, 50, 60, 70$ and 90 (from left to right). Noise leads to non-ideal recognition: the probability to be detected is larger than zero for $N < T_{\text{tot}}$ (overestimation) and smaller than one for $N > T$ (underestimation). The former effect is several orders of magnitude stronger than the latter. The graph belonging to the threshold value $T_{\text{tot}} = 60$ is highlighted (black) and the idealized detection probability of Equation (4.4) is added for this case (dotted).

Figure (4.2) delivers $P_{\text{det}}(N, T_{\text{tot}}, \bar{b})$ as a function of the photon number N for a variety of different total thresholds T_{tot}. The background noise has been set to $\bar{b} = 1$ as is typical for high-quality CCDs. It can clearly be seen how objects are overestimated in regions with $N < T_{\text{tot}}$ where the probability is unequal zero and underestimated in regions where $N > T_{\text{tot}}$ where the probability is smaller one.

It is of considerable importance to notice that $P_{\text{det}}(N, T_{\text{tot}}, \bar{b})$ is not a probability distribution for the photon number N. The full distribution to yield N photons and being detected is

$$p(N|\bar{N}, T_{\text{tot}}, \bar{b}) = P_{\text{det}}(N, T_{\text{tot}}, \bar{b}) p(N|\bar{N}) \tag{4.10}$$

with $p(N|\bar{N})$ given by the geometric distribution of Equation (1.5). Then, four alternatives can be discerned of which the first two are desirable and the other two are not:

1. $N \geq T_{\text{tot}}$ and the object is recognized,

$$P_1(\bar{N}, \bar{b}, T_{\text{tot}}) = \sum_{N=T_{\text{tot}}}^{\infty} P_{\text{det}}(N, T_{\text{tot}}, \bar{b}) p(N|\bar{N}), \tag{4.11}$$

2. $N < T_{\text{tot}}$ and the object is not recognized,

$$P_2(\bar{N}, \bar{b}, T_{\text{tot}}) = \sum_{N=0}^{T_{\text{tot}}-1} (1 - P_{\text{det}}(N, T_{\text{tot}}, \bar{b})) p(N|\bar{N}), \tag{4.12}$$

3. $N \geq T_{\text{tot}}$ and the object is not recognized,

$$P_3(\bar{N}, \bar{b}, T_{\text{tot}}) = \sum_{N=T_{\text{tot}}}^{\infty} (1 - P_{\text{det}}(N, T_{\text{tot}}, \bar{b}))p(N|\bar{N}), \tag{4.13}$$

4. $N < T_{\text{tot}}$ and the object is recognized,

$$P_4(\bar{N}, \bar{b}, T_{\text{tot}}) = \sum_{N=0}^{T_{\text{tot}}-1} P_{\text{det}}(N, T_{\text{tot}}, \bar{b})p(N|\bar{N}). \tag{4.14}$$

To classify for which types of fluorophores, characterized by their mean photon number $\bar{N}$, the recognition process is error-prone Figure (4.3) shows a plot of these probabilities as a function of the mean photon number $\bar{N}$ for a realistic value of the event threshold $T_{\text{tot}} = 100$. The pathological case of exceeding the threshold but not being detected (3) is weaker by an order of magnitude $1 \cdot 10^{-4}$ and therefore does not play any practical role. At mean photon numbers $\bar{N}$ close to zero practically no event is recognized, the case of correct rejection (2) is dominating. At high mean photon numbers desired recognition (1) dominates. The probability to be detected without wanting to do so, case (4), approaches zero for large photon numbers. For mean photon numbers close to the threshold this pathological case

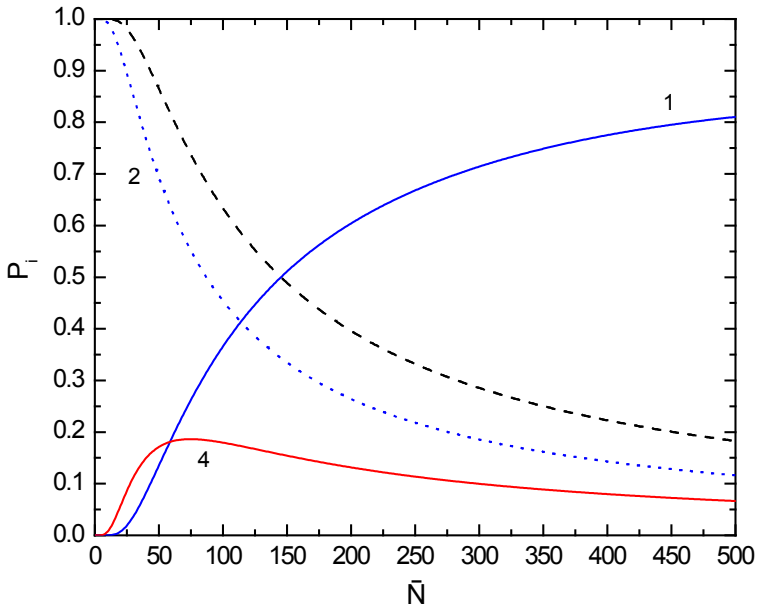

Figure 4.3: Probabilities $P_i(\bar{N}, \bar{b}, T_{\text{tot}})$ for the different detection modes of Equations (4.11,4.12,4.14) as a function of the mean photon number $\bar{N}$. This graph allows to categorize fluorophores with different $\bar{N}$ according to their affinity to produce a recognition error. The desired configurations (1) and (2) are displayed in blue, the pathological one (4) in red. Configuration (3) is weaker by a factor $1 \cdot 10^{-4}$ and therefore practically non-existent. Recognition is represented by solid lines, non-recognition by dotted lines. The event threshold is $T_{\text{tot}} = 100$, the background noise is $\bar{b} = 1$. To facilitate a comparison, the dashed black line shows the probability to be below threshold for the noise-free case. For a further explanation see the text.

displays a maximum. In this case up to 20% of all events are pathological. It is also worth noting that at moderately low photon numbers $\bar{N} \sim 40$ undesired recognition dominates over desired recognition. As a comparison, the probability to be below threshold for the ideal case without any background noise is added. This describes what type of event the molecule originally offers.

The event detection theory developed so far does describe experimental data in sufficient detail. Figure (4.4) shows a comparison of a measured photon histogram[1] recorded under realistic biophysical conditions [7] and the theoretical prediction of Equation (4.10). The fluorophore is the fast reversibly photo-switching protein rsFastLime [89], a variant (V157G) of the reversibly switchable fluorescent protein Dronpa [90]. The molecule was attached to α-tubulin inside intact PtK2 cells. The data was recorded by analyzing 60,000 frames. The overall image acquisition time was 120 s (500 frames/s) and the light intensity (488 nm) was increased stepwise from 3.5 kW/cm^2 to 5.0 kW/cm^2 in order to keep the number of activated molecules sufficiently stable in spite of photo-bleaching. The same experimental data will be used in the following whenever theoretical results are to be compared to the experimental reality. For instance, in the later Chapter (5.2.2) it will be subjected to a pointwise deconvolution scheme suitable for PALMIRA.

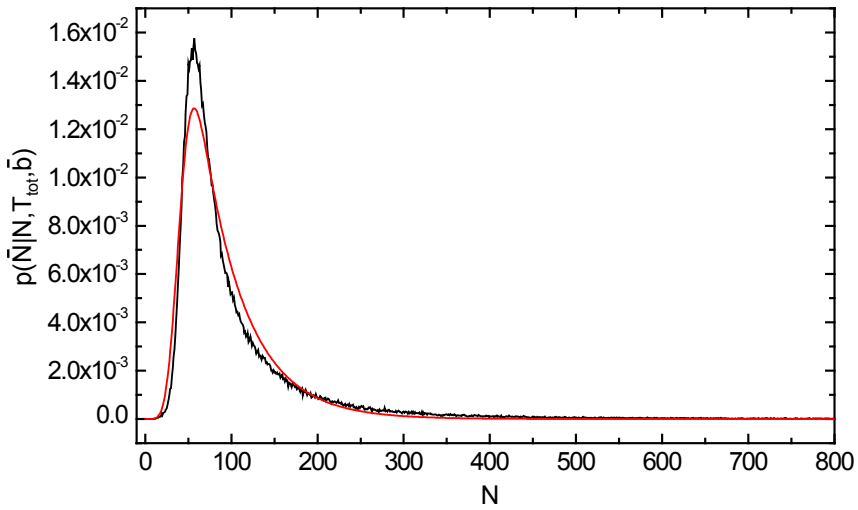

Figure 4.4: Comparison of a measured photon histogram (black) and the prediction of the detection theory according to Equation (4.10) (red). The fluorophore is the fast reversibly photo-switching fluorescent protein rsFastLime embedded in an intact mammalian cell. Model parameters here are $\bar{N} = 50$, $\bar{b} = 1$ and $T_{\text{pix}} = 9$. The background value was extracted from a pure-noise measurement of the setup. The mean photon number coincides with the geometric mean of the measured histogram for values $N > 100$ where the detection probability has effectively reached $P_{\text{det}} = 1$.

[1]The measurement is courtesy of C. Geisler, MPI f. Biophysical Chemistry, Göttingen.

4.3 The Problem of Higher-Order Events

In the last sections the recognition of events has been studied as far as spatial variations or the role of background noise are concerned. Thereby, it has been assumed that all molecules are imaged independently from each other. Higher-order events are events for which the optical patterns of more than one molecule overlap. Here, the focus is laid on the effect of neighboring molecules. Can molecules really be imaged independently from each other ? How can the independence of molecules be quantified? How strong is the effect of information loss for overlapping single-molecule images? At first, attention is turned to the independence of molecules.

In the subsequent the probability for higher-order events as a function of the excitation probability of the switching transition P_{sw} is calculated. It is important to notice that one image frame might contain a huge number of activated molecules at the same time. The only necessity is that the associated patterns do not overlap. The critical area which defines the onset of higher order events is the Airy disc. The number of molecules being activated and sharing an Airy disc at the same time must not be higher than one for independent single-molecule imaging to apply. Therefore, it makes sense to characterize the fluorophore distribution under investigation by the maximum number M of molecules being distributed on any arbitrary area as large as the optical Airy disc. Then the probability $p(n, M)$ for n molecules to be switched on in this area in one cycle is distributed binomially as Equation (1.4),

$$p(n, M) = \binom{M}{n} P_{\text{sw}}^{n} P_{\text{sw}}^{M-n}, \tag{4.15}$$

where P_{sw} denotes the switching probability. The probability $P_{>1}(P_{\text{sw}}, M)$ to have a higher-order event then reads

$$P_{>1}(P_{\text{sw}}, M) = \sum_{n=2}^{\infty} p(n, M). \tag{4.16}$$

The cumulation of the binomial distribution is given by beta functions, see Equation (2.40), and may be expressed in a simple polynomial form here since the summation limits are trivial,

$$P_{>1}(P_{\text{sw}}, M) = 1 - (1 - P_{\text{sw}})^{M} - M P_{\text{sw}} (1 - P_{\text{sw}})^{M-1}. \tag{4.17}$$

Figure (4.5) delivers a plot of $P_{>1}(P_{\text{sw}}, M)$ as a function of the molecule number M for different values of the switching probability $P_{\text{sw}} = 1 \cdot 10^{-3} .. 6 \cdot 10^{-3}$ in steps of $1 \cdot 10^{-3}$. Lower switching probabilities lead to a higher number of molecules tolerable in an Airy disc. High switching probabilities require samples of low fluorophore density in order for higher-order events to occur sufficiently rarely.

Once the probability for higher-order events is known it is of considerable interest to quantify the information loss inherent in these events. The overlap of optical patterns of molecules entails a loss of knowledge about the separation of the molecules. This loss is more severe the closer the objects become. Again, Fisher information lends itself as a measure of this process. For two molecules of equal mean photon number $\bar{N}$ and separation

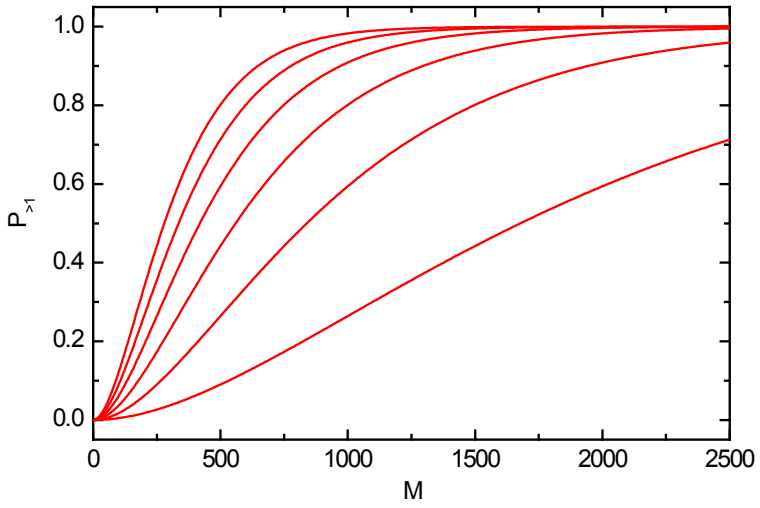

Figure 4.5: The probability for higher order events $P_{>1}(M, P_{\text{sw}})$ as a function of the molecule number M and the switching probability P_{sw}. From the left to the right the switching probability is decreasing and takes the values $P_{\text{sw}} = 6{\cdot}10^{-3}, 5{\cdot}10^{-3}, 4{\cdot}10^{-3}, 3{\cdot}10^{-3}, 2{\cdot}10^{-3}$ and $1{\cdot}10^{-3}$, respectively. The order of magnitude of the molecule number M is appropriate for conventional biological samples. Clearly, lower switching probabilities lead to a higher tolerable number of molecules in one Airy disc.

distance d the mean signal $\bar{N}_{ij}$ in a pixel indexed by i, j is given by

$$\bar{N}_{ij} = \bar{N}h(x_i - d/2, y_j) + \bar{N}h(x_i + d/2, y_j) + \bar{b}, \tag{4.18}$$

where $h(x, y)$ is the normalized detection PSF, x_i, y_j are the CCD pixel center coordinates and $\bar{b}$ the mean background noise. Assuming Poisson pixel statistics the Fisher information measure for the particle distance d follows in close analogy to Equation (3.11),

$$I(d) = \sum_{i,j} \frac{\bar{N}^2}{4} \frac{[h'(x_i + d/2, y_j) - h'(x_i - d/2, y_j)]^2}{\bar{N}h(x_i - d/2, y_j) + \bar{N}h(x_i + d/2, y_j) + \bar{b}}. \tag{4.19}$$

The derivative is with respect to the $x-$coordinate, here. The spatial distribution of the two molecules could be chosen arbitrarily since the detection PSF is rotationally symmetric when averaging over equally-distributed orientations of the emission dipole.

Figure (4.6) represents this Fisher information measure for different mean photon numbers $\bar{N} = 50, 150$ and 250 at a background level of $\bar{b} = 1$. As usual, the CCD pixel size amounts to 100 nm squared. The standard detection PSF with $\lambda = 575$ nm is employed. The molecule distance is varied in an interval $d \in [0, 1000]$ nm. Perfectly overlapping molecules are not discernible. For increasing particle distance d the Fisher information $I(d)$ grows and stabilizes on a plateau for $d > 300$ nm. This plateau corresponds to Cramer-Rao bounds for unbiased distance estimators of $12, 16$ and 32 nm, respectively. Small variations on the plateau are residuals of the oscillating spatial behavior of the detection PSF. Figure (4.6) provides a measure as to what extent algorithms treating higher-order events like the CLEAN algorithm of Figure (1.4) could be successful in discerning double objects. Predictions of a

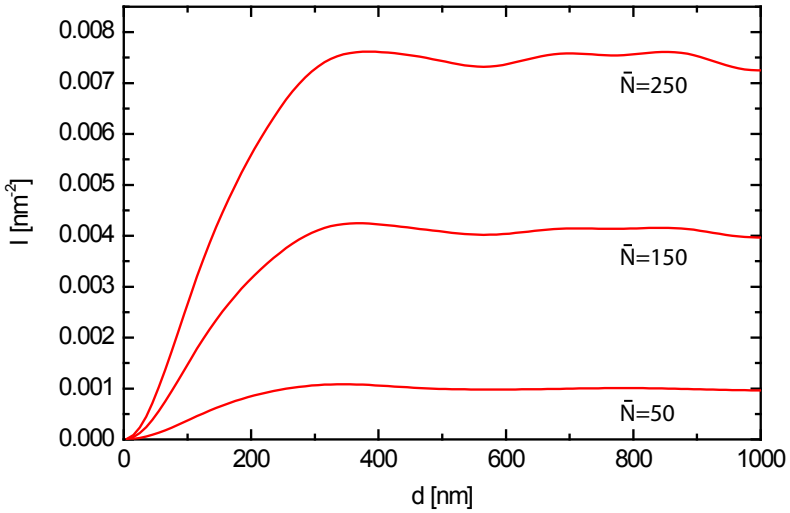

Figure 4.6: Fisher information for the distance of two equally bright molecules as given by Equation (4.19) with $\bar{N} = 50, 150$ and 250. The pixel size of the CCD is 100 nm, the mean background noise per pixel $\bar{b} = 1$. The molecule distance is chosen at 100 equally spaced points in the interval $d \in [0, 1000]$ nm. Entirely overlapping molecules are not discernible. For increasing particle distance Fisher information grows and stabilizes on a plateau for $d > 300$ nm.

Fisherian analysis for the estimation accuracy of the inter-particle distance have already been stated for the simpler case of the PSF given by a Gaussian or an Airy function [91] and been verified experimentally [92]. The above findings using the more realistic PSF are consistent with these results.

4.4 Spectrally Resolved Molecule Recognition

Use of molecules with different emission characteristics that are recorded in different spectral channels enhances the level of complexity of the analysis by adding a particle type degree of freedom. In statistical terms, the problem is again described by a multinomial distribution. Assume that L different types of molecules with spectra $\{S_l(\lambda)\}, l = 1..L$ give the mean photon signals $\{\bar{n}_l\}, l = 1..L$ and are recorded on C different channels with wavelength supports $\{I_c\}, c = 1..C$. Then, the mean signal in the channels $\bar{\boldsymbol{c}}$ is

$$\bar{\boldsymbol{c}} = \boldsymbol{K}\bar{\boldsymbol{n}} \tag{4.20}$$

with the $C \times L$ sorting matrix $\boldsymbol{K}$ given by

$$K_{cl} = \int_{I_c} S_l(\lambda) d\lambda. \tag{4.21}$$

K_{cl} is the probability that photons originating from a molecule of type l will be detected in channel c. The normalization

$$\sum_{c=1}^{C} K_{cl} = 1 \quad \forall l \tag{4.22}$$

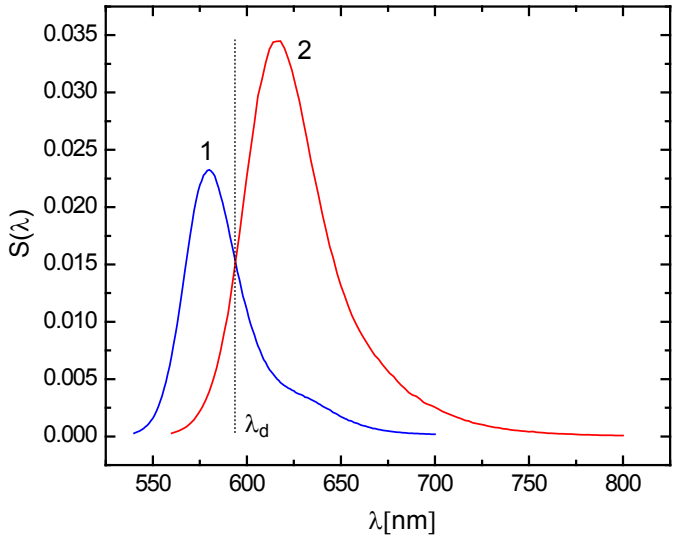

Figure 4.7: Typical spectra of two dyes,a RhB-derivative (1) and a Rh590s variant (2), with sufficiently distinct emission characteristics for use in dual-color PALMIRA microscopy. The more blue-shifted fluorophore has an emission peak at $\lambda_1 = 575$ nm, the more red-shifted dye at $\lambda_2 = 625$ nm. The spectra are normalized such that the integral yields unity. The channel separation wavelength $\lambda_d = 592$ nm leading to a maximum occurrence of correct assignments is shown as a dotted line. There is considerable crosstalk between the two channels. Nonetheless, statistical estimation procedures can heavily profit of the fact that the crosstalk is quantitatively different for the two molecule types.

is assured for all molecules l as long as the channels cover the whole emission wavelength support without overlap.

Figure (4.7) presents an example[2] for the spectra of a RhB-derivative (1) and a Rh590s variant (2) [83]. The channel intervals are chosen as $I_1 = [0, \lambda_d]$ and $I_2 = [\lambda_d, \infty]$ with $\lambda_d = 592$ nm. In this special case, the sorting matrix $\boldsymbol{K}$ takes the values

$$\boldsymbol{K} = \begin{pmatrix} 0.65 & 0.1 \\ 0.35 & 0.9 \end{pmatrix}. \tag{4.23}$$

Given a stochastic signal sample $\boldsymbol{c}$ the probability $P(\boldsymbol{c}|l)$ of the channel signal configuration assuming that it originates from molecule l reads

$$P(\boldsymbol{c}|l) = \frac{N!}{c_1!c_2!...c_C!} K_{1l}^{c_1} ... K_{Cl}^{c_C}, \tag{4.24}$$

where the total measured signal N is given by

$$N = \sum_{j=1}^{C} c_j. \tag{4.25}$$

Equation (4.24) just reflects the process of sorting N photons in C channels with multiple occurrences being allowed. This is the standard model leading to a multinomial probability

[2]The measurement is courtesy of J. Fölling, MPI f. Biophysical Chemistry, Göttingen.

distribution. The associated maximum-likelihood estimator $\hat{L}(\mathbf{n})$ giving the molecule type l for which the observed data $\mathbf{c}$ has the highest probability may be constructed as

$$\hat{L}(\boldsymbol{c}) = \max\left\{P(\boldsymbol{c}|l), l = 1..L\right\}. \tag{4.26}$$

The estimator $\hat{L}(\boldsymbol{c})$ is extraordinarily simple for the case of just two channels $C = 2$. Then, the setup is characterized by the probabilities $\{K_l\}, l = 1..L$ to end up in the channel 1 and the estimator is just

$$\hat{L}(c_1, c_2) = \max\left\{K_l^{c_1}\left(1 - K_l\right)^{c_2}, l = 1..L\right\}. \tag{4.27}$$

This is equivalent to finding the maximum of the log-likelihood $L(c_1, c_2|l)$

$$L(c_1, c_2|l) = c_1 \ln(K_l) + c_2 \ln(1 - K_l). \tag{4.28}$$

Interestingly, maximizing Equation (4.28) is equivalent to minimizing the entropy H

$$H(K_l) = -K_l \ln(K_l) - (1 - K_l)\ln(1 - K_l) \tag{4.29}$$

subject to the constraint that the channel mean values

$$\begin{aligned} \bar{c}_1 &= NK_l \\ \bar{c}_2 &= N(1 - K_l) \end{aligned} \tag{4.30}$$

are fixed to c_1, c_2. In other words, the molecule type is chosen in such a way that it leads to the most pronounced, i.e. minimum entropy, distribution effect while maintaining the correct mean values. Neglecting a scaling factor the likelihood in Equation (4.28) solely depends on the the signal fraction γ

$$\gamma = \frac{c_1}{c_2}. \tag{4.31}$$

Hence, maximizing $L(c_1, c_2|l)$ is equivalent to maximizing the function

$$L(K, \gamma) = \gamma \ln(K) + \ln(1 - K). \tag{4.32}$$

The maximum $K_{\max}(\gamma)$ given γ is determined by

$$\frac{K_{\max}}{1 - K_{\max}} = \gamma \tag{4.33}$$

and describes the type of fluorophore, characterized by the sorting probability $K_{\max}$, that maximizes the probability of the observed data.

Figure (4.8) displays a plot of $L(K, \gamma)$ for different values of the signal fraction γ. The function displays a unique maximum at $K_{\max}(\gamma)$. This value describes that molecule spectrum for which the data is maximum probable. If a choice has to be made among a finite discrete set $\{K_i\}$ of sub-optimal sorting probabilities there might exist ambiguities. For instance, if the measurement would yield the same signal in both channels, $\gamma = 1$, and the probabilities would be given as $K_1 = 0.3$ and $K_2 = 0.7$ it could not be inferred whether the

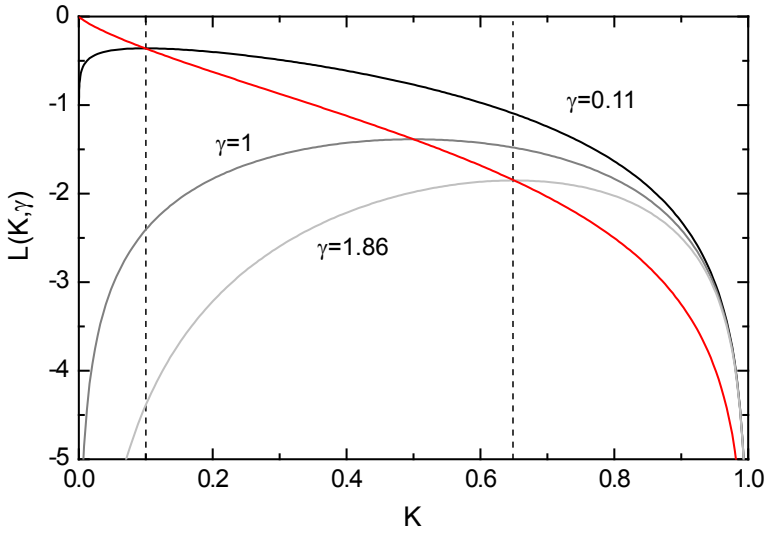

Figure 4.8: Likelihood $L(K,\gamma)$ as a function of the channel sorting probability K for different values of the signal ratio γ. The values $\gamma = 0.11$ (black) and $\gamma = 1.86$ (light grey) are the mean ratios obtained for the sorting probabilities (vertical lines) associated with the spectra in Figure (4.7). The value $\gamma = 1$ (grey) is added to cover the symmetric case. The function $L(K,\gamma)$ always features a single maximum which gives the sorting probability K that leads to a highest probability of the observed data. The curve $[K_{\max}(\gamma), L_{\max}(\gamma)]$ covering the maxima of the likelihood is added (red). This is the trajectory of the most probable sorting with the arc length given by γ. Multi-color imaging with only two channels is rendered feasible because the maximum of the likelihood covers a broad range as K is varied.

signal originated from the first or the second molecule type since $L(K,\gamma=1)$ is symmetric around $K = 0.5$. Fortunately, sub-optimal γ are less frequent since the mean of γ is proportional to $K_l/(1-K_l)$, see Equation (4.30). Additionally, the curve of maximum likelihood $[K_{\max}(\gamma), L_{\max}(\gamma)]$ is displayed. This is the trajectory of sorting optima. The arc length on this curve measured from the origin yields the mean γ for a given K. Multi-color imaging using two channels is possible because the maximum of the likelihood spans a broad range as K is varied.

Provided shot noise is dominating the performance of the estimator of Equation (4.26) is characterized by the transformation matrix $\boldsymbol{T}(\boldsymbol{K}, C, N)$

$$T_{il}(\boldsymbol{K}, C, N) = \sum_{\boldsymbol{c}} P(\boldsymbol{c}|l) \begin{cases} 1 & \hat{L}(\boldsymbol{c}) = i \\ 0 & \hat{L}(\boldsymbol{c}) \neq i \end{cases} \tag{4.34}$$

where $T_{il}(\boldsymbol{K}, C, N)$ describes the probability that a molecule of type l will be recognized as of type i. $\boldsymbol{T}(\boldsymbol{K}, C, N)$ depends on the sorting matrix $\boldsymbol{K}$, the total number of photons N and the channel number C only. In Equation (4.34), the probabilities for all configurations $\boldsymbol{c}$ leading to the assignment $l \rightarrow i$ are accumulated. The trace of $\boldsymbol{T}(\boldsymbol{K}, C, N)$,

$$Q(\boldsymbol{K}, C, N) = \sum_{i=1}^{L} T_{ii}(\boldsymbol{K}, C, N), \tag{4.35}$$

is a figure of merit for the goodness of the estimator since it measures the frequency of correct assignments. A maximization of $Q(\lambda_d)$ for the spectra of Figure (4.7) yields $\lambda_d = 592$ nm. Typical examples for the transformation matrix $\boldsymbol{T}(N)$ at $N = 5$ or $N = 50$ photons are

$$\boldsymbol{T}(N=5) = \begin{pmatrix} 0.92 & 0.06 \\ 0.08 & 0.94 \end{pmatrix} \tag{4.36}$$

and

$$\boldsymbol{T}(N=50) = \begin{pmatrix} 1 & < 1 \cdot 10^{-5} \\ < 1 \cdot 10^{-5} & 1 \end{pmatrix}. \tag{4.37}$$

In real measurements the signal is polluted by background noise $\bar{b}$. Figure (4.9) renders a Monte-Carlo simulation of the transition matrix $T_{ii}(N)$ that has been performed in order to clarify this point. The experimental spectra of Figure (4.7) have been employed and the sorting signal $\boldsymbol{c}$ has been degraded with Poisson background noise of mean value $\bar{b} = 27$. This is motivated by the fact that each optical pattern extends over roughly 3×3 pixels if a pixel length 100 nm is used. Then, the mean noise value has been subtracted and the estimator of Equation (4.28) was used on the corrected data. For signal photon values $N \gg \bar{b}$ where

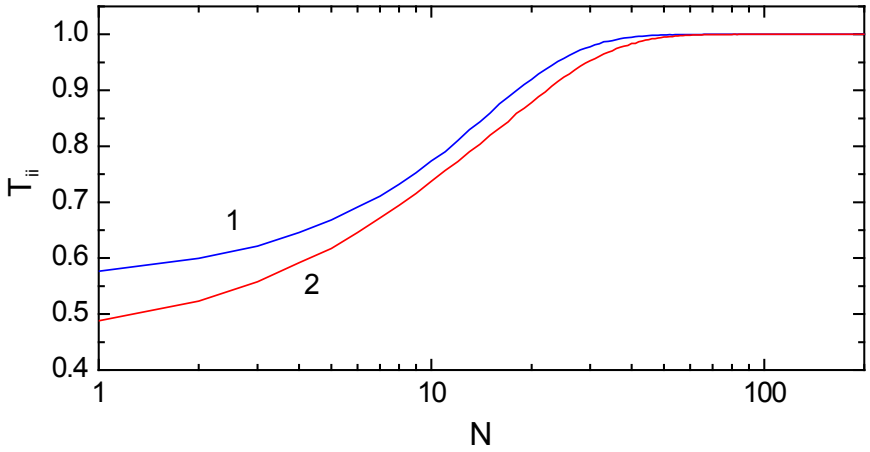

Figure 4.9: The transition matrix $T_{ii}(N)$ for correct assignments as a function of the photon number N calculated by a Monte-Carlo simulation of noise-degraded data. The experimental spectra of Figure (4.7) and a background of $\bar{b} = 27$ per channel have been used. At higher photon numbers $N > 50$ object recognition is practically error-free. At lower signal photon number N mis-assignments are encountered more frequently. In this case, the background noise tends to make the signals equal in both channels. Thus, the fluorophore (1) with the more equilibrated sorting $K_1 = 0.65$ is preferred over the molecule (2) with the more asymmetric sorting $K_2 = 0.1$.

the sorting effect dominates the results are not different from the noise-free case in Equation (4.37). At low signal-to-noise ratio $N < 50$ erroneous assignments take place and the results are worse than those in Equation (4.36). The background noise inclines to equalize the signals in both channels. Consequently, the fluorophore with the more equilibrated sorting $K_1 = 0.65$ is privileged over the more asymmetric molecule with $K_2 = 0.1$.

In summary, with moderate photon numbers $N > 50$ and a simple estimator it is possible to discern the molecule types to almost absolute certainty. Unlike in classical microscopy

crosstalk in PALMIRA microscopy is of no disadvantage as long as the crosstalk characteristics, i.e. the columns of $\boldsymbol{K}$, are sufficiently different [93]. The method has been shown to work in an experiment [10].

4.5 Polarization Resolved Molecule Recognition

Fluorescence may be detected depending on the state of its polarization. Thereby, the electric field vector is projected onto certain spatial directions. At first, the distribution of photon numbers is studied for a random projection of the electric field on a plane formed by two orthogonal detected polarizations. Secondly, the connection of pixel thresholding and sampling rate of different polarization angles is elaborated. In a third step, a Fisherian analysis of the estimation error of the polarization angle is presented and the results are compared to a Monte-Carlo simulation using a common polarization estimator.

Usually, the electric field vector $\mathbf{E}$ of the emitted fluorescence light depends on the orientation of the emission dipole and can point in any direction of space unless special properties of the sample force this vector into a preferred direction. Here, it is assumed that the emission dipole is static during acquisition. The fluorescence signal is recorded in two channels corresponding to the polarization pointing into the orthogonal $x-$ and $y-$direction, respectively. The plane spanned by these two vectors is referred to as projection plane. Figure (4.10, A) provides a definition of the coordinate system for the problem. The azimuthal angle ϑ is formed by $\mathbf{E}$ together with the $z-$direction. The $xy-$projection of $\mathbf{E}$ is given by $\mathbf{E}_\mathrm{p}$ and defines the polar angle φ formed together with the $x-$direction.

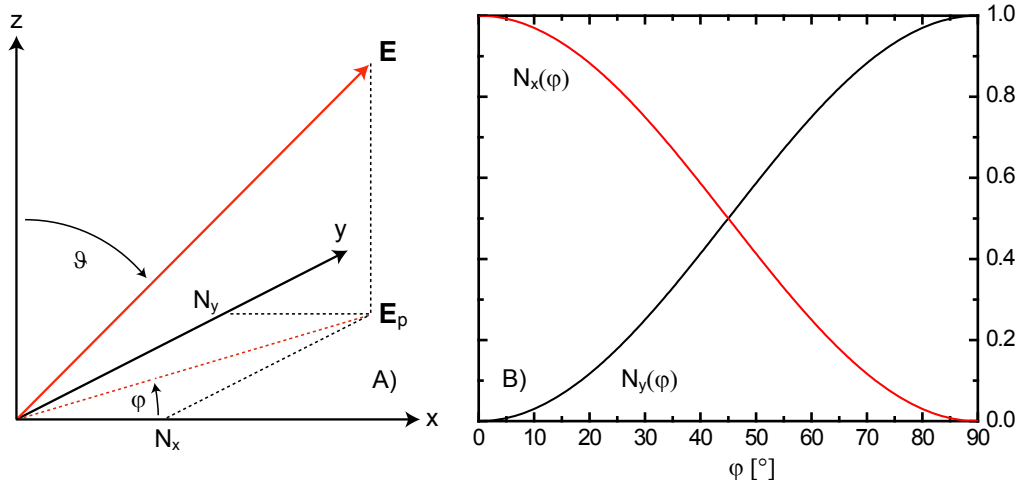

Figure 4.10: The coordinate system (A) is given by a conventional three-dimensional cartesian system. The electric field vector $\mathbf{E}$ of the emitted fluorescence is shown in red, its projection on the detection plane $\mathbf{E}_\mathrm{p}$ in dotted red. The azimuthal angle with the $z-$axis is ϑ, the polar angle formed between the $x-y-$projection and the $x-$axis is φ. The projected signals $N_x(\varphi)$ and $N_y(\varphi)$ for a unit vector $\mathbf{E}_\mathrm{p}$ are shown on the right (B). The signal is split among the channels proportionally to $\cos^2(\varphi)$ and $\sin^2(\varphi)$ for (x) and (y), respectively.

In the subsequent, it is assumed that the absolute square of $\mathbf{E}$, that is the total photon number, is distributed geometrically with mean $\bar{N}$ according to Equation (1.5) and that all

azimuthal angles ϑ are equally probable. Then, the photon number N in the projection plane given a certain azimuth ϑ is conditionally distributed as

$$p(N|\bar{N},\vartheta) = \frac{\left[\bar{N}\sin^2(\vartheta)\right]^N}{\left[\bar{N}\sin^2(\vartheta)+1\right]^{N+1}} \tag{4.38}$$

and the azimuth angle itself is equally distributed as

$$p(\vartheta) = \sin(\vartheta) \tag{4.39}$$

such that the integral of $p(\vartheta)$ is normalized to unity. The $z-$component of $\mathbf{E}$ is not observed experimentally and therefore the distribution $p_{\text{eff}}(N|\bar{N})$ of in-plane projections reads

$$p_{\text{eff}}(N|\bar{N}) = \int_0^{\pi/2} p(N|\bar{N},\vartheta)p(\vartheta)d\vartheta. \tag{4.40}$$

Figure (4.11) shows a plot of the function $p_{\text{eff}}(N|\bar{N})$ for a mean photon value $\bar{N} = 250$. A geometric distribution featuring the same mean value is given as a comparison. The difference between the former and the latter function is positive for values $N < 120$ and negative for values $N > 120$. This reflects the fact that probability is shifted down towards smaller values of N. If all azimuthal emission dipole orientations can be assumed to occur equally probable in the sample the problem is effectively two-dimensional with the photon distribution given in Equation (4.40).

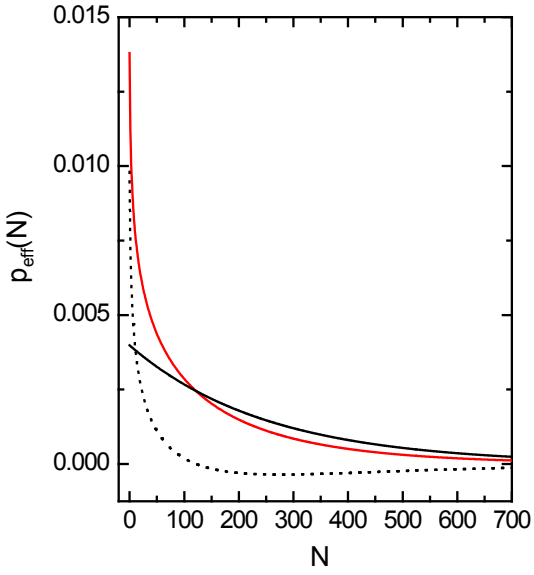

Figure 4.11: The effective distribution $p_{\text{eff}}(N|\bar{N})$ of in-plane projections of the emitted electric field for an equally-distributed azimuthal angle ϑ (red). A geometric distribution featuring the same mean value is added (black). The difference of $p_{\text{eff}}(N|\bar{N})$ and $p_{\text{geo}}(N|\bar{N})$ is positive for small N and negative for larger values (dotted). Smaller values of N are more frequent than larger values since projection always shortens the field vector.

In real experiments the image frames of different channels may be misaligned because their light follows different optical paths. If the geometric transformation between the channels is not known it is advantageous to firstly localize the events of both channels separately and then calculate the alignment transformation by using the most probable regrouping of the event positions. This entails subjecting the images of the $x-$ or $y-$channel to an individual thresholding procedure during object recognition which is equivalent to introducing a polarization-dependent total event threshold. No disadvantage follows as long as the highest effective total threshold is below the minimum total threshold that is required for the subsequent image analysis. To affirm the validity of this assumption it is of interest to study the effective event threshold as a function of the polarization angle φ. The discussion is closely related to the one of Chapter (4.1) where the z-dependency has been treated. Assume that an event is accepted if the photon number in at least one of the two channels is higher than T. This number is called individual threshold in the following. The effective event threshold $T(\varphi)$ is then given by

$$T(\varphi) = \min\left\{\frac{T}{\sin^2(\varphi)}, \frac{T}{\cos^2(\varphi)}\right\}. \tag{4.41}$$

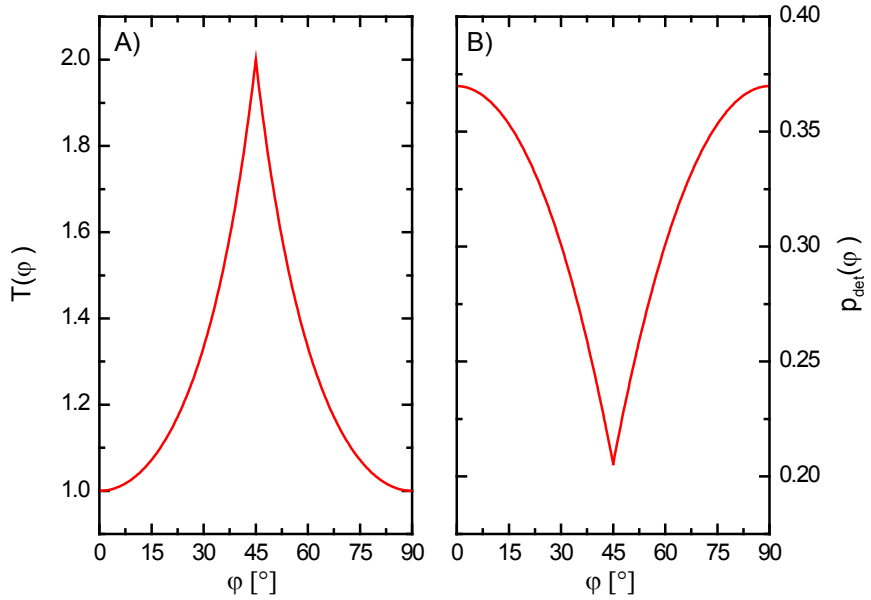

Figure 4.12: The factor between the effective total threshold $T(\varphi)$ and the individual threshold T during object recognition as a function of the polarization angle becomes 2 for $\varphi = 45°$ (A). The associated detection probability for an individual threshold $T = 100$ and a mean photon number $\bar{N} = 250$ drops by a factor of about 1.5 in this case (B). To achieve a homogeneous sampling of the polarization space the retrieved events have to be subjected to a second thresholding procedure with total threshold $T_{\text{tot}} \geq 2T$.

A plot of the threshold is given in Figure (4.12, A) for an individual threshold $T = 1$. The threshold is one for the pure polarizations $\varphi = 0°, 90°$ and rises up to two when approaching $\varphi = 45°$. Clearly, the graph is symmetric with respect to $\varphi = 45°$. Making use of Equation (4.40) the probability $P_{\text{det}}(\varphi, \bar{N}, T)$ for a molecule to be detected during one switching cycle

may be calculated as

$$P_{\text{det}}(\varphi, \bar{N}, T) = \sum_{N=N(\varphi,T)}^{\infty} p_{\text{eff}}(N|\bar{N}). \tag{4.42}$$

Here, capital letters are employed to distinguish probabilities from probability densities. This equation replaces Equation (2.26) which was valid for the case of a pure geometric distribution. Figure (4.12, B) presents $P_{\text{det}}(\varphi, \bar{N}, T)$ for the reasonable values $T = 100$ and $\bar{N} = 250$. It is seen that events with a polarization $\varphi = 45°$ are about 1.5 times less frequently observed than the events with pure polarizations. Therefore, to achieve a homogeneous sampling of the polarization space $\{\varphi = 0°..90°\}$ it is necessary to subject the final retrieved events to a second total signal thresholding with $T_{\text{tot}} \geq 2T$. Otherwise, pure polarizations will be over-represented.

Analogous to all estimation problems treated so far it is of considerable interest to study the information content of the measured data and the performance of employed estimators. The discussion is restricted to two dimensions, projection effects from the third unobserved dimension z are assumed to have been treated as in the first paragraph. Thus, the signal in the detection plane, N, can assumed to be known. The numbers of photo-electrons N_x, N_y in the polarization channels then follow a double Poisson distribution, compare Equation (1.9), like

$$p(N_x, N_y|\varphi, N, \bar{b}) = \prod_{i=x,y} p_{\text{poi}}\left(N_i|\bar{N}_i\right) \tag{4.43}$$

with the mean values

$$\begin{aligned} \bar{N}_x(\varphi) &= N\cos^2(\varphi) + \bar{b} \\ \bar{N}_y(\varphi) &= N\sin^2(\varphi) + \bar{b} \end{aligned} \tag{4.44}$$

and the mean background given by $\bar{b}$. This distribution models the photo-detection process[3]. The Fisher information $I(\varphi)$ associated with the angle φ according to Equation (3.9) reads

$$I(\varphi) = \frac{\bar{N}_x'(\varphi)^2}{\bar{N}_x(\varphi)} + \frac{\bar{N}_y'(\varphi)^2}{\bar{N}_y(\varphi)}. \tag{4.45}$$

Hence, the information measure is given by

$$I(\varphi) = 4N^2\sin^2(\varphi)\cos^2(\varphi)\left\{\frac{1}{N\sin^2(\varphi) + \bar{b}} + \frac{1}{N\cos^2(\varphi) + \bar{b}}\right\}. \tag{4.46}$$

Figure (4.13) contributes a plot of $I(\varphi)$ and the associated Cramer-Rao bound $1/\sqrt{I(\varphi)}$ for different signal photon numbers $N = 25, 50, 100, 150$ and a background of $\bar{b} = 27$. This is a typical value taking into account a characteristic mean noise value of one per pixel and assuming the single-molecule image to extend over a pixel field of 3×3 pixels. Both, the Fisher information measure $I(\varphi)$ and the size of the interval centered at $\varphi = 45°$ over which the Cramer-Rao bound is essentially constant are increasing with increasing photon

[3] Again, a multinomial image model can be discussed on the same grounds as in connection with Equation (3.11).

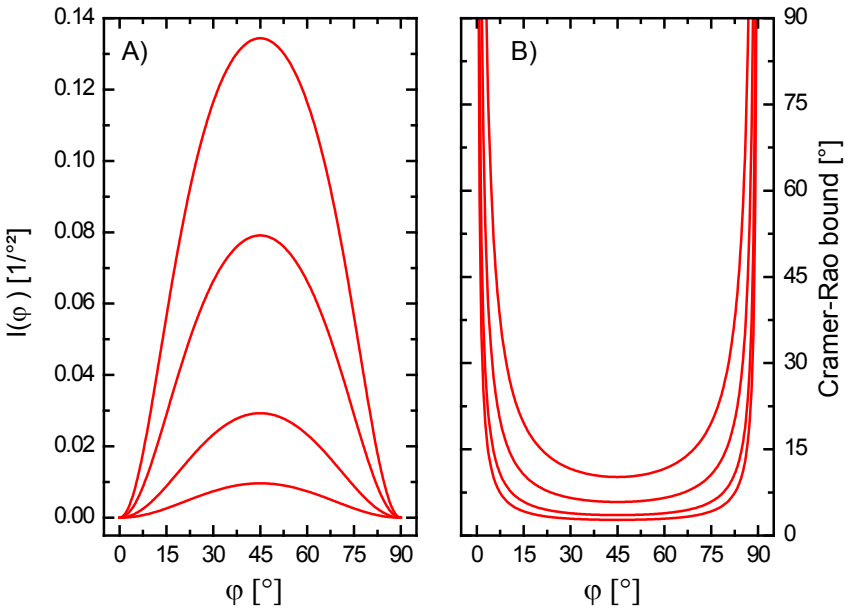

Figure 4.13: Fisher information $I(\varphi)$ (A) and the associated Cramer-Rao bound $1/\sqrt{I(\varphi)}$ on the lower error of unbiased estimators of the polarization angle (B) for different photon numbers $N = 25, 50, 100, 150$ and a background value $\bar{b} = 27$ per channel. In the left image photon numbers are increasing from the bottom to the top whereas on the right side it is vice versa. The background has been chosen in such a way to account for typical single-pixel noise values and typical numbers of CCD camera pixels over which the single-molecule image extends. It is seen that the error in any unbiased determination of φ is at the least on the order of a few degrees with a constant region around $\varphi = 45°$ increasing in size proportionally to the number of signal photons. The pure polarizations at $\varphi = 0°, 90°$ cannot be determined without bias and with finite error at the same time since both signals $\bar{N}_x(\varphi)$ and $\bar{N}_y(\varphi)$ have zero derivative at these points and the Fisher information consequently is zero, compare Figure (4.10, B).

number N. The bound for the minimum standard deviation of any unbiased estimator is of the order of a few degrees for polarization values on the plateau but diverges as φ approaches $0°$ or $90°$, respectively. Nonetheless, biased estimators may exist that feature errors smaller than the unbiased Cramer-Rao bound. This has already been clear from an inspection of the generalized Cramer-Rao inequality for biased estimators, Equation (3.22), where the derivative of the bias with respect to the parameter increasingly eliminates the influence of Fisher information on the minimum error the more it approaches -1. Fluctuation errors may be squeezed down at the expense of rising systematic errors. Or in other words, interval estimation may be superior to point estimation as far as the estimation error is concerned.

A Monte-Carlo simulation of the process of determining the polarization state confirms this finding. $5 \cdot 10^4$ different realizations of N_x and N_y were drawn from the distribution of Equation (4.43) for each value of φ. The polarization estimator $\hat{\varphi}$ was taken as

$$\hat{\varphi} = \arctan\left\{\sqrt{\frac{N_y - \bar{b}}{N_x - \bar{b}}}\right\}. \tag{4.47}$$

The mean background noise $\bar{b}$ was assumed to be known and has been subtracted. Samples

leading to a negative or a diverging radicand were discarded.

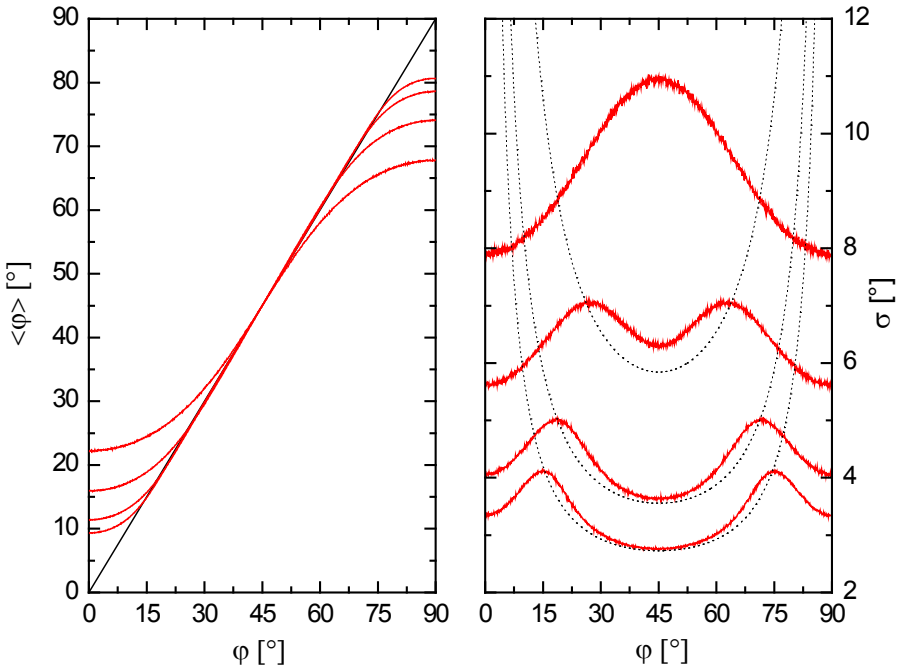

Figure 4.14: Results of a Monte-Carlo simulation modeling the estimation of the state of polarization φ from double-channel intensities. The expectation value $\langle\hat{\varphi}\rangle$ (A) and the resulting standard deviation $\sigma(\varphi)$ (B) were estimated from $1 \cdot 10^5$ random samples at each value of φ. The signal photon numbers were taken as $N = 25, 50, 100, 150$ and the noise as $\bar{b} = 27$. The bias becomes stronger the more φ moves away from $45°$ and decreases with increasing signal N as does the standard deviation. As a comparison, the Cramer-Rao bounds for the unbiased estimator of Figure (4.13) have been added (dotted black). In the region around $45°$ where the bias is nearly zero the error $\sigma(\varphi)$ is bound from below by the unbiased Cramer-Rao bound. Closer to the pure polarizations $\varphi = 0°, 90°$ the error $\sigma(\varphi)$ undercuts this bound while the estimated mean polarization shows a deviation from the line through the origin. Directly at $\varphi = 0°, 90°$ the unbiased Cramer-Rao bound diverges while the derivative of the bias approaches -1. These two processes lead to a finite error $\sigma(\varphi = 0°, 90°)$.

Figure (4.14) delivers the resulting mean outcome $\langle\hat{\varphi}\rangle$ and the standard deviation $\sigma(\varphi)$ that were estimated from the statistical samples for $1 \cdot 10^3$ equally spaced $\varphi-$values. Both, the bias and the standard deviation decrease with increasing signal N. The error is bound from below by the unbiased Cramer-Rao bound in regions where the bias is nearly zero. The estimator approaches the Cramer-Rao bound from above for increasing signal photon number. In regions closer to $\varphi = 0°, 90°$ the bias increases while the error falls below the unbiased Cramer-Rao bound. Directly at $\varphi = 0°, 90°$ the error $\sigma(\varphi = 0°, 90°)$ is finite although the unbiased Cramer-Rao bound diverges. These two effects are compatible since the bias approaches a derivative -1 at this point, the mean of the estimator is horizontal there. This can also be seen from Equation (3.22).

5 Deconvolution of PALMIRA Images

In the following a deconvolution theory for PALMIRA images based on local Gaussian point-spread-functions is formulated. Such a procedure is desirable for two reasons. On the one hand, it circumvents the loss of information inherent in averaging over the photon number when sorting events into the same histogram irrespectively of this degree of freedom. On the other hand, deconvolution narrows the spatial distributions of events and therefore allows to work with a smaller threshold while still yielding the same resolution. This is advantageous because less events that are in principle recognizable have to be discarded. The theory is applied to stochastically generated images of two equally bright particles and experimental data of stained microtubuli.

5.1 Deconvolution with Local Point-Spread-Functions

It has been shown in Equation (2.17) that the imaging process for the mean image consists in the sum over partial imaging processes at fixed photon number N,

$$b(\mathbf{r}) = \sum_N b(\mathbf{r}|N), \tag{5.1}$$

with the image at fixed photon number given by

$$b(\mathbf{r}|N) = h(\mathbf{r}|N, \bar{N}, T) * o(\mathbf{r}) \tag{5.2}$$

and the local PSFs by

$$h(\mathbf{r}|N, \bar{N}, T) = p(\mathbf{r}|N)p(N|\bar{N}, T). \tag{5.3}$$

Here, $p(\mathbf{r}|N)$ is the Gaussian of Equation (2.5) and $p(N|\bar{N}, T)$ the truncated geometric distribution of Equation (2.8). The local PSF is just the localization estimation distribution at fixed photon number weighted by the probability of such a photon number to occur. This is important to properly account for the fact that events with higher photon number are of higher localization quality but show up much less frequently. The factor $p(N|\bar{N}, T)$ will enhance their contribution in the right way. Without this prior frequency factor the events with lower photon number would be equipped with too high an importance.

The imaging process for the mean image, Equation (5.1), entails a loss of information since the photon number as a quality-index for the position estimation is discarded completely. This has already been obvious in Chapter (2.1) where the expected mean PSF has been calculated. The construction of the mean PSF by superposition of fixed-photon number localization PSFs weighted by the geometric distribution of the photon number as in Equation (2.9) is reminiscent of the construction of the density operator as a statistical mixture

of pure quantum states [94]. There, the associated mixing entropy measures the similarity of the process under consideration to a pure state. Here, the equivalent of the pure state is fixed-photon number localization and the statistical mixture is characterized by the geometric distribution. The mixing entropy for the PALM-PSF follows as

$$\begin{aligned} S(\bar{N}) &= -\sum_{N=T}^{\infty} p(N|\bar{N},T)\ln p(N|\bar{N},T) \\ &= (\bar{N}+1)\ln(\bar{N}+1) - \bar{N}\ln\bar{N} \end{aligned} \tag{5.4}$$

and does not depend on the photon threshold T due to the self-similarity of the truncated geometric distribution in Equation (2.8) at all scales. The relative mixing strength is always the same irrespective of the minimum photon number taken into account.

Figure (5.1) delivers a plot of the entropy $S(\bar{N},T)$ as a function of the mean photon number $\bar{N}$. The higher the mean photo number $\bar{N}$ the more the geometrical distribution

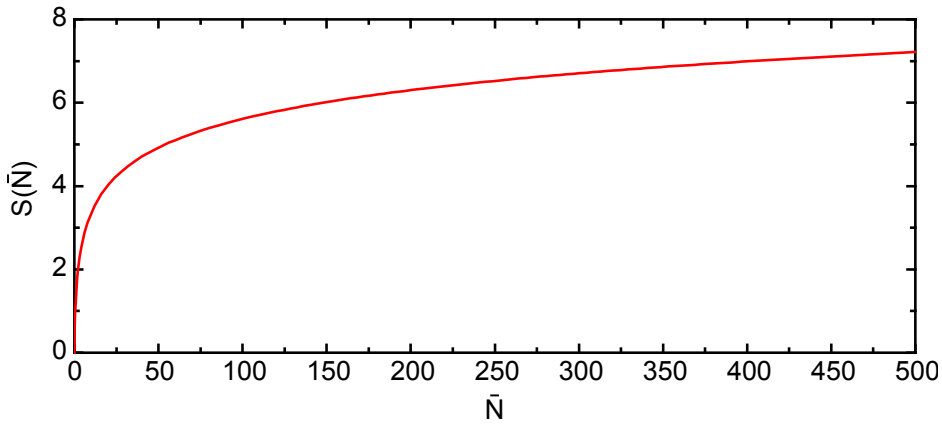

Figure 5.1: Mixing entropy $S(\bar{N})$ as a function of the mean photon number $\bar{N}$ for the mean PSF as constructed in Equations (5.2) or (2.9). The summation over fixed-photon number localization distributions entails a loss of information about the photon number. The more different the localization distributions contributing to the full PSF are the higher is this loss of information. The truncated geometric distribution of Equation (2.8) decides on which fixed-photon number localization distributions contribute to the image. Therefore, its associated entropy, quantifying the similarity of the mean PALMIRA PSF to a fixed-photon number localization PSF, provides a quantitative measure of the information loss inherent in displaying localization events of different accuracy in the same image. Deconvolution can circumvent this loss of information.

stretches out in space and the less similar is the PALMIRA imaging process to a fixed-photon number localization. The lower the photon number the narrower is the distribution and the less different fixed-photon number localization processes contribute to the final mean image. The loss of information grows strongly for photon numbers $\bar{N} < 100$ and then progressively stabilizes. Even at moderate photon numbers like $\bar{N} = 50$ for the protein Dronpa [90] the mixing process has a strong effect and it seems to be worth to seek for methods circumventing it. Linear deconvolution of multiple images with local PSFs is the simplest approach and will be developed in the following.

In the subsequent, functions in Fourier space will be denoted by $\hat{f}(\mathbf{k}) = FT(f(\mathbf{r}))$. A Fourier transformation of Equation (5.2), application of the convolution theorem, multiplication by $\hat{h}^*(\mathbf{k}|N, \bar{N}, T)$ and summation over N yields

$$\sum_N \hat{b}(\mathbf{k}|N)\hat{h}^*(\mathbf{k}|N, \bar{N}, T) = \hat{o}(\mathbf{k}) \sum_{N'} \left|\hat{h}(\mathbf{k}|N', \bar{N}, T)\right|^2 . \tag{5.5}$$

Backward Fourier transformation and use of the convolution theorem leads to the estimated object $\bar{o}(\mathbf{r})$

$$\bar{o}(\mathbf{r}) = \sum_N b(\mathbf{r}|N) * H(\mathbf{r}|N, \bar{N}, T) \tag{5.6}$$

with the deconvolution function

$$H(\mathbf{r}|N, \bar{N}, T) := FT^{-1}\left\{ \frac{\hat{h}^*(\mathbf{k}|N, \bar{N}, T)}{\sum_{N'} \left|\hat{h}(\mathbf{k}|N', \bar{N}, T)\right|^2 + \lambda} \right\} . \tag{5.7}$$

Here, the real-valued regularization parameter λ has been introduced in order to avoid division by zero. Equation (5.7) defines a variant of the Wiener filter [95]. Figure (5.2) shows a one-dimensional plot of this function for a FWHM of the underlying detection PSF $F = 250$ nm, the local photon number $N = 100$, a mean photon number $\bar{N} = 250$, a threshold $T = 50$ and a regularization parameter of $\lambda = 1 \cdot 10^{-9}$. The function features a strong

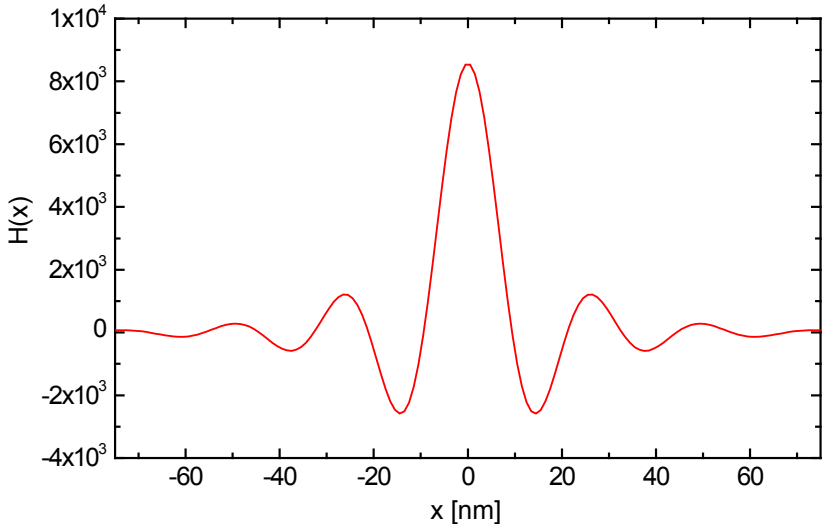

Figure 5.2: The deconvolution wavelet $H(x; y = 0|N, \bar{N}, T)$ according to Equation (5.7) for a photon number $N = 100$, a mean photon number $\bar{N} = 250$, a threshold $T = 50$ and a regularization parameter of $\lambda = 1 \cdot 10^{-9}$. The FWHM of the underlying optical PSF is $F = 250$ nm. The central peak is narrower than the associated localization accuracy. The negative side-lobes have the effect of contracting the function with which it is convolved. Statistical images have to feature a sufficiently high degree of sampling when deconvolution with this function is to be meaningful in a probabilistic sense.

similarity with a $\sin(x)/x$-function. The central peak amplifies the image directly at the retrieved molecule positions. The negative oscillations beneath this peak lead to a narrowing of the distribution of neighboring events. The introduction of the regularization parameter λ is necessary in order not to amplify background noise too strongly but entails a systematic alteration of the imaging process. The resulting imaging process might described as follows. Combining Equations (5.2) and (5.6) the estimated object $\bar{o}(\mathbf{r})$ is related to the real object $o(\mathbf{r})$ in the following way

$$\bar{o}(\mathbf{r}) = FT^{-1}\left\{\frac{1}{1+\lambda/\sum_{N'}\left|\hat{h}(\mathbf{k}|N',\bar{N},T)\right|^2}\right\} * o(\mathbf{r}). \tag{5.8}$$

The images $b(\mathbf{r}|N)$ are composed of M_N localization events at positions $\bar{\mathbf{r}}_j^N$ with $j = 1..M_N$. That is, $b(\mathbf{r}|N)$ is given by

$$b(\mathbf{r}|N) = \sum_{j=1}^{M_N} \delta\left(\mathbf{r} - \bar{\mathbf{r}}_j^N\right). \tag{5.9}$$

Thus, from Equation (5.6) it is obtained that the deconvolution process consists merely of a sum over the deconvolution wavelet $H(\mathbf{r}, N|\bar{N}, T)$ centered at the estimated molecule positions

$$\bar{o}(\mathbf{r}) = \sum_N \sum_{j=1}^{M_N} H(\mathbf{r} - \bar{\mathbf{r}}_j^N, N|\bar{N}, T). \tag{5.10}$$

The calculation of the deconvolution wavelet $H(\mathbf{r}|N, \bar{N}, T)$ of Equation (5.7) poses a certain computational difficulty since the Fourier-transformed Gaussians $\hat{h}(\mathbf{k}|N', \bar{N}, T)$ have to be known for all photon numbers N' in order to calculate $H(\mathbf{r}|N, \bar{N}, T)$ for just one photon value N. Therefore, an algorithm would proceed in the following way:

1. sort the events according to their photon numbers (photon lookup-table),
2. calculate the denominator in Equation (5.7) with only the observed photon numbers contributing a Gaussian to the sum, i.e. estimate the prior photon distribution from the experimental data,
3. perform the deconvolution sum of Equation (5.10) piecewise for all events of the same photon number. Repeat this step until all photon numbers are covered.

In step (2) it is reasonable to use the observed distribution of photons instead of a theoretical geometric distribution for several reasons. Firstly, the point of interest here is the experimental occurrence of different photon numbers and not the similarity of their distribution to a theoretical geometric one. Using a theoretical distribution in Equation (5.7) would mean that the sum in the denominator would contain many more different Gaussians than were observed in the experiment. The degree of completeness of the sampling of the photon distribution is not to be captured here. Secondly, the use of a theoretic distribution would imply an error-free knowledge of the mean photon number $\bar{N}$. Thirdly, numerical calculation times are shorter here since not every photon number has to be summed over.

5.2 Practical Examples

In the following sections the deconvolution strategy of the foregoing paragraph is applied to stochastically simulated PALMIRA images of two equally bright objects of different distances and to experimental images of stained micro-tubuli.

5.2.1 Deconvolution of Equally Bright Objects

In a first instance, the performance of the deconvolution theory is studied for two equally bright objects at positions $y = \pm d/2$ for different values of the inter-molecule distance d. This is of course an idealization since real images will inevitably contain objects which might on average be optically equally bright, i.e. they have the same mean photon number $\bar{N}$, but contribute a different number of switching cycles M during their lifetimes before photo-bleaching and not being seen any more. Nevertheless, two molecules yielding the same number of position estimates is the most simple model for two objects to be discerned in the field of PALMIRA microscopy. This model is closely analogous to the one used by Sparrow in classical optics [96].

Figure (5.3) shows typical stochastic images for different distances $d = 25, 20, 17.5, 15, 12.5, 10$ nm and a number of position estimates $M = 500$ sufficiently high to obtain stable images. The deconvolution parameter λ has been set individually in such a way to push inevitable negative values of the deconvolved image to approximately 10% of the maximum

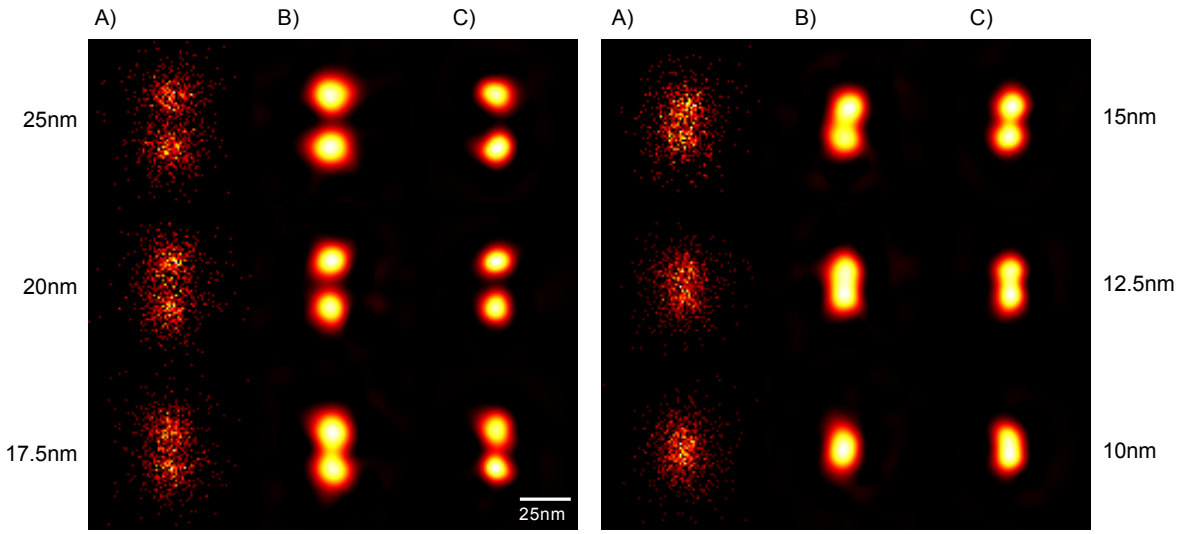

Figure 5.3: Typical stochastic realizations of the conventional image (A), the deconvolution using the mean PSF of Equation (2.13) (B) and the point-wise-deconvolved image (C) for two equally bright objects at different vertical distances d. The number of position estimates per object is $M = 500$, the mean photon number $\bar{N} = 250$, the threshold $T = 50$ and the FWHM of the detection PSF $\omega = 250$ nm. The point-wise deconvolution always features results superior to or equally good as the deconvolution with the mean PSF. For $d = 15$ nm the deconvolved image (C) clearly displays two objects which cannot be discerned in the mean image (A) and which can hardly be distinguished in (B).

value. It is important to notice three points: (i) The point-wise deconvolution scheme (C) is always of equal or superior quality as the deconvolution with a mean PSF (B). (ii) At a distance $d = 15$ nm the two objects are clearly discernible in the deconvolved image (C) while they are undistinguishable in the conventional PALMIRA image (A). Here, the distance d approaches the approximate FWHM of the PALMIRA-PSF ~ 16 nm. (iii) From a certain maximum distance on even the deconvolved image does not allow to discern the objects (iii). The first point demonstrates that deconvolution in PALMIRA microscopy is more than a post-experimental smoothing of the data since it avoids an information loss.

5.2.2 Deconvolution of Experimental Data

When applying linear deconvolution to real data care has to be taken that the number of position samples per molecule is high enough for the procedure to be statistically meaningful. The theory is applied to the same experimental data originating from rsFastLime-stained mircotubuli [7] that was used to evaluate the object recognition theory in Figure (4.4). At this occasion, the physical parameters of the measurement have already been stated. Figure (5.4) presents the results. The regularization parameter has been chosen as $\lambda = 1 \cdot 10^{-9}$ in order to prevent negative values from becoming stronger than 10% of the maximum value. Several single peaks at distances ~ 30 nm are revealed by the deconvolution algorithm.

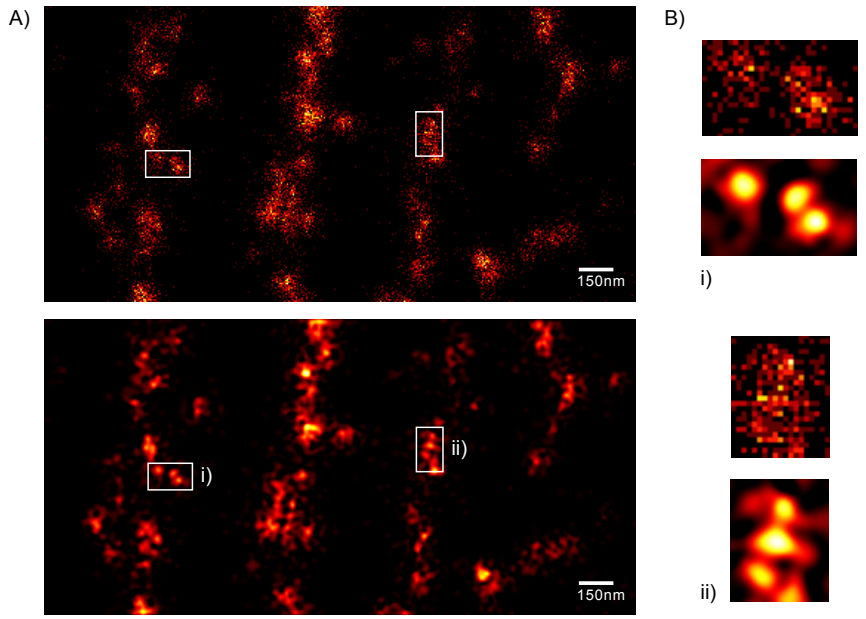

Figure 5.4: Linear deconvolution of experimental PALMIRA data using local PSFs. (A) shows a $2\mu m \times 1\mu m$ image of rsFastLime-stained microtubuli (top) and its deconvolved counterpart (bottom). (B) shows two regions in which the deconvolution reveals details that are hidden in the mean conventional image. The peaks which can be identified have a minimum distance of 30 nm. This is roughly twice the FWHM of the mean PALMIRA-PSF ~ 26 nm like in the theoretical study with two equally bright objects. The number of position samples per object is about 100.

Part II

Stochastic Studies in Fluorescence Correlation Spectroscopy

6 Preliminaries

In this chapter a short introduction to mathematical aspects of fluorescence correlation spectroscopy and to the way of its stochastic simulation is provided. A concise expression for the calculation of the correlation curve is given and relevant parameters entering the calculation are discussed. Next, the algorithm for the Monte-Carlo generation of fluorescence time traces is stated and the design of the estimation procedure for the correlation curve is presented.

6.1 Fluorescence Correlation Spectroscopy

The un-normalized fluorescence correlation curve $g(\tau)$ in FCS measurements is defined by

$$g(\tau) = \langle q(t)q(t+\tau)\rangle \tag{6.1}$$

with $q(t)$ being the fluorescence signal recorded at t. The normalized variant is

$$g(\tau) = \frac{\langle q(t)q(t+\tau)\rangle}{\langle q^2(t)\rangle}. \tag{6.2}$$

The brackets indicate a temporal average. Theoretically, $g(\tau)$ of Equation (6.1) has to be calculated by the following double integral

$$g(\tau) = \sum_{ij} \int\int dV dV' q_j(\mathbf{r}')q_i(\mathbf{r})c_i(\mathbf{r})p_{i\to j}(\mathbf{r},\mathbf{r}',\tau). \tag{6.3}$$

The numbers i and j are the particle type indices. The summation and the integration extend over both the initial and the final configurations $(\mathbf{r}', i)$ and $(\mathbf{r}, j)$, respectively. Here, $q_i(\mathbf{r})$ is the fluorescence brightness of molecule type i at position $\mathbf{r}$. $p_{i\to j}(\mathbf{r},\mathbf{r}',\tau)$ is the propagator and describes the differential probability to be in the state $(\mathbf{r}', j)$ at $t = t_0 + \tau$ if the starting state at $t = t_0$ was $(\mathbf{r}, i)$. $p_{i\to j}(\mathbf{r},\mathbf{r}',\tau)$ encodes the different paths along which the system can evolve. $c_i(\mathbf{r})$ is the concentration of species i. There are several cases in which Equation (6.3) might be simplified considerably. It is assumed that particles are independent from each other.

(i) Assuming a single freely diffusing species Equation (6.3) reduces to

$$g(\tau) = \int\int dV dV' q(\mathbf{r}')q(\mathbf{r})c(\mathbf{r})p(\mathbf{r},\mathbf{r}',\tau) \tag{6.4}$$

with the diffusion propagator [97]

$$p(\mathbf{r},\mathbf{r}',\tau) = \frac{1}{(2\pi)^{3/2}(2D\tau)^{3/2}} \exp\left\{-\frac{(\mathbf{r}-\mathbf{r}')^2}{4D\tau}\right\}. \tag{6.5}$$

Here, the propagator just describes the distribution of diffusion steps $\mathbf{r}' \rightarrow \mathbf{r}$ with a diffusion constant D and a time step of τ. This is a Gaussian with mean $\boldsymbol{\mu} = 0$ and variance $\sigma^2 = 2D\tau$. If the brightness $q(\mathbf{r})$ has also a Gaussian shape,

$$q(\mathbf{r}) = q \exp\left(-\frac{x^2+y^2}{2\omega_{xy}^2} - \frac{z^2}{2\omega_z^2}\right), \tag{6.6}$$

the integral yields

$$g_{\text{D}}(\tau) = Nq^2 \left(1+\tau/\tau_{\text{D}}\right)^{-1} \left(1+\epsilon\tau/\tau_{\text{D}}\right)^{-1/2} \tag{6.7}$$

for the un-normalized correlation curve and

$$g_{\text{D}}(\tau) = \frac{1}{N} \left(1+\tau/\tau_{\text{D}}\right)^{-1} \left(1+\epsilon\tau/\tau_{\text{D}}\right)^{-1/2} \tag{6.8}$$

for the normalized one. N is the average number of molecules in the detection volume V_{det},

$$V_{\text{det}} = \pi^{3/2}\omega_{xy}^2\omega_z, \tag{6.9}$$

and τ_{D} is the time a molecule takes to diffuse through the focus laterally and given by

$$\tau_{\text{D}} = \omega_{xy}^2/4D \tag{6.10}$$

with D being the diffusion constant. The constant $\epsilon = \omega_{xy}^2/\omega_z^2$ accounts for the elongation of the detection volume along the optic axis.

(ii) For a stationary molecule that alternates between a state 1 with brightness $q_1(\mathbf{r}) = q(\mathbf{r})$ and and a dark state 2 with $q_2(\mathbf{r}) = 0$ it is obtained from Equation (6.3)

$$g(\tau) = \int dV q^2(\mathbf{r}) c_1(\mathbf{r}) p_{1\rightarrow 1}(\mathbf{r},\tau) \tag{6.11}$$

because due to the stationarity the propagator factorizes into

$$p_{1\rightarrow 1}(\mathbf{r},\mathbf{r}',\tau) = p_{1\rightarrow 1}(\mathbf{r},\tau)\delta(\mathbf{r}-\mathbf{r}'). \tag{6.12}$$

If the alternation rates are given as $k_{12}(\mathbf{r})$ and $k_{21}(\mathbf{r})$ the propagator reads

$$p_{1\rightarrow 1}(\mathbf{r},\tau) = \frac{k_{21}(\mathbf{r})}{k_{21}(\mathbf{r})+k_{12}(\mathbf{r})} \left(\frac{k_{12}(\mathbf{r})}{k_{21}(\mathbf{r})} \exp\left[-(k_{12}(\mathbf{r})+k_{21}(\mathbf{r}))\tau\right] + 1\right). \tag{6.13}$$

This formula is found by solving the rate equations for a two-level system. Under the initial condition to reside in state 1 the probability to remain in this state after a time τ is given by $p_{1\rightarrow 1}(\mathbf{r},\tau)$.

The propagators $p(\mathbf{r}',\mathbf{r},\tau)$ for diffusion of Equation (6.5) and $p_{1\rightarrow 1}(\mathbf{r},\tau)$ for two-level type conversion of Equation (6.13) are the key distributions from which stochastic samples are drawn during a Monte-Carlo simulation.

6.2 Algorithms for the Monte-Carlo Simulation

A Monte-Carlo experiment of fluorescence correlation spectroscopy mainly involves two steps: the generation of the time trace $q(t)$ and the estimation of the correlation curve $g(\tau)$ from the time trace. Both points are described in the following.

6.2.1 Generation of Time Traces

A Monte-Carlo simulation of the fluorescence time trace $q(t)$ progresses as follows. At first, a uniformly distributed initial distribution of particles $\{\mathbf{r}_{i,\mathrm{R}}\}$ is generated. Next, these particles are relocated by drawing Gaussian random numbers of the propagator in Equation (6.5). Then, molecule type transitions are performed by drawing random numbers from the reaction propagator in Equation (6.13). Finally, the expectation value for the fluorescence signal q is read out for every particle and a Poisson random number is drawn to account for the photo-detection process. Figure (6.1) displays the flow diagram for one time step of the Monte-Carlo simulation. The process is repeated until the required trace length is reached.

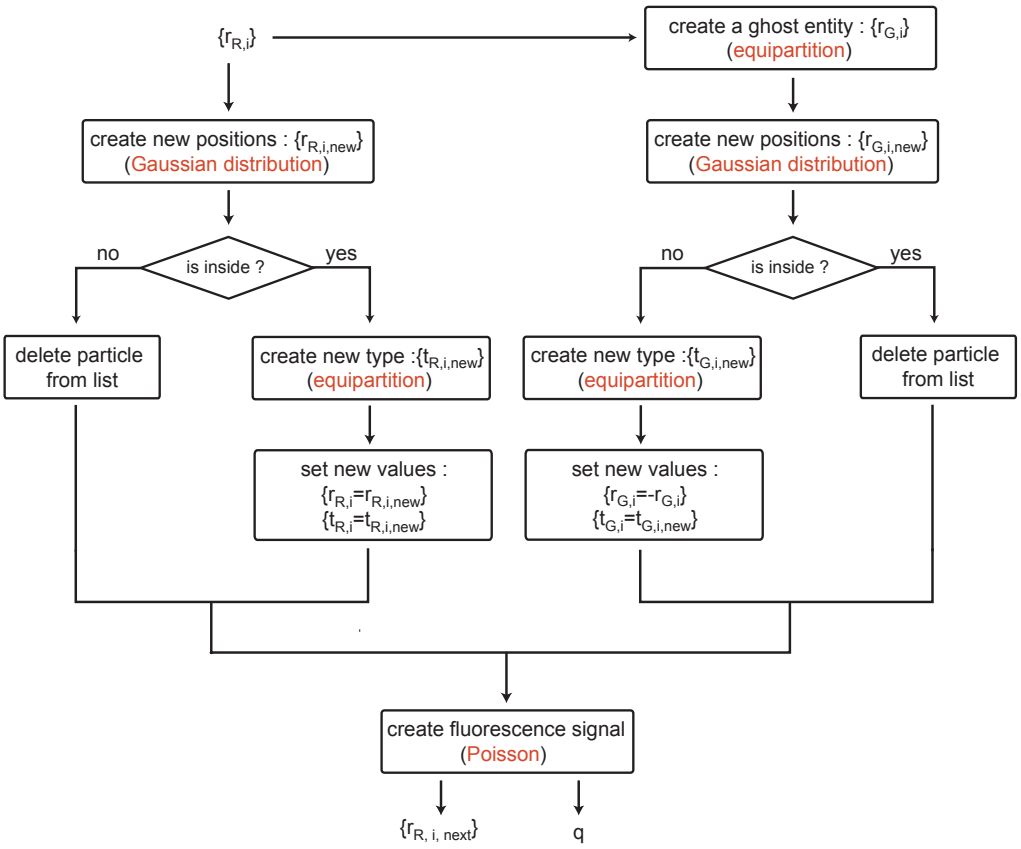

Figure 6.1: Flow diagram for one time step of the generation of the fluorescence intensity time series by a Monte-Carlo simulation. For stochastic steps the related probability distribution is highlighted in red. Essentially, the time evolution of two entities of particles has to be followed stochastically: that of the real fluorescing particles with initial positions $\{\mathbf{r}_{i,\mathrm{R}}\}$ and that of a ghost entity $\{\mathbf{r}_{i,\mathrm{G}}\}$ of similar size to account for boundary effects related to the finite size of the simulation volume. The ghost entity has to be newly initialized in each step. Then, candidates for the new positions $\{\mathbf{r}_{i,\mathrm{R,new}}; \mathbf{r}_{i,\mathrm{G,new}}\}$ are stochastically drawn. Real particles diffusing out and ghost particles staying inside the simulation volume are discarded. Next, it is determined whether a particle type transformation takes place. Accepted ghost particles are restored at the inverse of their initial position $-\mathbf{r}_{i,\mathrm{G}}$. Finally, the fluorescence signal q is read out at every particle position. The procedure models properly the random flux across the simulation volume boundary. Periodic boundary conditions would lead to an artefact in the correlation curve at the lag time that corresponds to the mean diffusion time through the simulation volume.

The most important aspect is to account properly for the boundary conditions. Using periodic boundary conditions where a particle leaving the simulation volume at one side enters it again at the opposite side is possible but leads to a pronounced artefact in the correlation curve. A clear signature will be seen on the time scale on which a particle traverses the whole simulation support. Important details of the correlation curve may be disturbed by this artefact. The problem can be avoided by making the stochastic particle flux across the simulation support boundary a part of the simulation [98]. Particles leaving the simulation volume disappear while new particles randomly show up inside the simulation volume in order to keep the number of active particles on average stable. The random appearance of new particles is achieved by newly creating a 'ghost' entity of particles in each step that has the same mean number as the real particles. This ghost entity is relocated in the same way as the ordinary active particles. The difference is that ghost particles not leaving the simulation volume are discarded while those leaving the volume are restored at minus their initial positions. In other words, those ghost particles leaving the volume supply the newly appearing real particles.

The influx boundary method may be justified mathematically as follows. Let the simulation volume be denoted by V. Let the probability for a particle to reside at $\mathbf{r}$ be constant and given by $p_{\text{s}}(\mathbf{r}) = p_0$. Thus, a uniform distribution is assumed. Then, the probability $p_{\text{in}}(\mathbf{r})$ that a particle from the outside V^c will show up at $\mathbf{r}$ is given by

$$p_{\text{in}}(\mathbf{r}) = \int_{V^c} p(\mathbf{r}, \mathbf{r}', \tau) p_0 d\mathbf{r}' \tag{6.14}$$

with $p(\mathbf{r}, \mathbf{r}', \tau) = p(\mathbf{r} - \mathbf{r}', \tau)$ from Equation (6.5). Here, the integration extends over all positions $\mathbf{r}'$ in the complement V^c that could contribute to the process. Changing the integration variable $\mathbf{r}' \rightarrow -\mathbf{r}'$, taking into account that $p(\mathbf{r}, \mathbf{r}', \tau)$ depends on $\mathbf{r} - \mathbf{r}'$ only and assuming a simulation volume V that is symmetric around the origin it is obtained

$$p_{\text{in}}(\mathbf{r}) = \int_{V^c} p(\mathbf{r} + \mathbf{r}', \tau) p_0 dr'. \tag{6.15}$$

If the probability that a particle is allowed to be put at $\mathbf{r}$ is also given by $p_{\text{a}}(\mathbf{r}) = p_0$ the probability $p_{\text{out}}(\mathbf{r})$ that a particle at $\mathbf{r}$ leaves the volume is defined as

$$p_{\text{out}}(\mathbf{r}) = \int_{V^c} p(\mathbf{r}' - \mathbf{r}, \tau) p_0 d\mathbf{r}' \tag{6.16}$$

Comparing Equations (6.15) and (6.16) it follows

$$p_{\text{in}}(\mathbf{r}) = p_{\text{out}}(-\mathbf{r}). \tag{6.17}$$

The process of getting a new particle at $\mathbf{r}$ by diffusion from the outer volume equals the probability that a particle from the inside at $-\mathbf{r}$ disappears. In the argumentation it has been crucial that the probabilities $p_{\text{s}}(\mathbf{r})$ to be removable from a certain point and $p_{\text{a}}(\mathbf{r})$ to be accepted at a certain point are equal. Furthermore, the simulation volume V has to be symmetric around $\mathbf{r} = (0, 0, 0)$.

A typical FCS simulation requires a variety of parameters ranging from diffusion constants across photo-chemical rates to full point-spread functions to be supplied. Table (6.1) assembles these in one place.

Name	Parameter	Typical value
D	diffusion constant	$3 \cdot 10^{-6}\text{cm}^2/\text{s}$
c	particle concentration	$15\ \mu\text{m}^{-3}$
V	simulation volume	$2\ \mu\text{m} \times 2\ \mu\text{m} \times 2\ \mu\text{m}$
Δt	time step	$0.1\ \mu\text{s}$
s	simulation loops	$1 \cdot 10^7$
k_{isc}	inter-system crossing rate	$1.1\ \mu\text{s}^{-1}$
k_{T}	triplet relaxation rate	$0.49\ \mu\text{s}^{-1}$
I_0	excitation peak intensity	1 MW/cm^2
σ	excitation cross section	$5 \cdot 10^{-20}$ m^2
q_0	molecular brightness	$1 \cdot 10^4$ MHz
$I(\mathbf{r}), h_{\text{det}}(\mathbf{r})$	excitation, detection profile	-

Table 6.1: The input functions and parameters of the Monte-Carlo simulation. Typical values are provided in the third column. In many simulations the molecular brightness q_0 is much higher than for real fluorophores. This leads to fast convergence of the correlation curve and is used when fluctuation properties of the curve are to be suppressed.

6.2.2 The Estimator for the Multi-τ-Correlator

Once, the fluorescence signal time series $\{q_i\}, i = 0..M-1$ is given the autocorrelation of Equation (6.2) is to be estimated. In general, the correlation curves measured by FCS display a maximum close to $\tau = 0$ and decay off to zero in a characteristic way for larger τ. The main interesting features can be found at comparatively low lag times τ whereas the variations of the correlation curve become more and more less pronounced for higher lag times. For a linear correlator relying on FFT the correlator would have to possess as many bins as the time trace is long in order to cover the largest lag time possible. Coarse-graining procedures work with multiple lag times and achieve higher lag times with less numbers of bins at the expense of averaging in higher bins. In such a way it is possible to cover extremely large lag time with a limited number of bins. Furthermore, intensity fluctuations get averaged out at lag times large compared to the timescale of these fluctuations. Such a quasi-logarithmic scheme has been proposed in the literature [43, 44]. Hereby, the authors also introduce the symmetric normalization scheme that achieves a reduction of the estimator noise by suppression of contributions of boundary effects originating from the beginning $q_0..q_\tau$ and the end $q_\tau..q_{M-1}$ of the time trace. Nevertheless, the published work gives a concise mathematical formulation only for linear correlation. A precise mathematical formulation of multiple lag time correlation is given in the following.

Firstly, the relevant parameters are defined and the resulting lag times are calculated. The multi-τ-correlator is characterized by specifying the following variables

1. N : the number of octaves,
2. $\{k_i\}, i = 0..N-1$: the number of bins per octave i,
3. $\{f_i\}, i = 0..N-2$: the bin width augmentation factor at the end of octave i.

As a convention the octaves will be numerated by $i = 0..N-1$ and the bin $j = 0..k_i - 1$ of octave i by (i, j). Then, the width ω_i of the elementary bin of octave i is given by

$$\omega_i := \omega_0 \prod_{l=0}^{i-1} f_l. \tag{6.18}$$

It has simply been used that each transition into a new octave is accompanied by an augmentation factor. Once, the bin widths are known they might be summed up to yield the absolute time position T_j^i of a multi-τ-bin at (i, j) as

$$T_j^i := \sum_{l=0}^{i-1} \omega_l k_l + j\omega_i. \tag{6.19}$$

To get the right result for $i = 0$ in Equations (6.18) and (6.19) the convention is introduced that backward-running indices in sums and products yield the neutral element,

$$\prod_{l=0}^{-1} \ldots = 1 \tag{6.20}$$

$$\sum_{l=0}^{-1} \ldots = 0. \tag{6.21}$$

Figure (6.2) sketches the structure of a simple multiple lag time correlator with three similar octaves $N = 3$, augmentation factors $f = 2$ and three bins per octave $k = 3$.

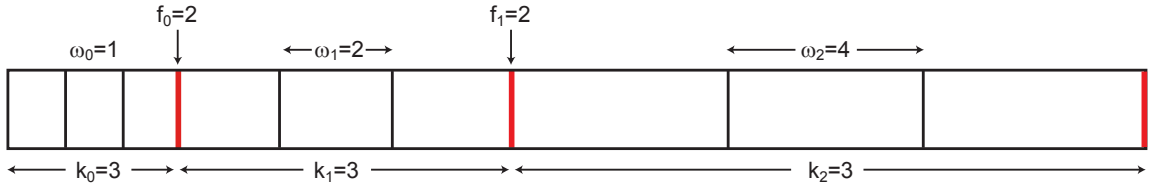

Figure 6.2: Binning scheme for a multi-τ-correlator. The characteristic parameters of this special correlator are the number of octaves $N = 3$, the numbers of bins per octave $\{k_i = 3\}, i = 0, 1, 2$ and the octave factors $\{f_i = 2\}, i = 0, 1$. The widths of the elementary bins of the octaves then follow as $\{\omega_i\}, i = 0, 1, 2$ given by Equation (6.18). The end of the octaves is highlighted (red).

Secondly, the un-normalized correlation function is calculated. The time trace is put piecewise into the correlator. Thereby, a compressed time trace is built up. Each bin of the compressed time trace experiences a correlation after a certain number $m =$ of time trace pieces has been supplied. m shall be called the filling number in the following. For instance, with $\omega_0 = 1$ (no averaging in zeroth octave) the first bin $(0, 0)$ is correlated at filling numbers $m = 1, 2, .., M$ for at total M times, the second bin $(0, 1)$ at filling numbers $m = 2, .., M$ for at total $M - 1$ times since $T_1^0 = 1$ and $T_0^0 = 0$ and so on. This can be formalized further. Given a trace of length M the bin (i, j) will be correlated at the filling numbers m

$$m \in \mathbb{K}_j^i(M) := \left\{T_{j+1}^i, T_{j+2}^i, .., M\right\} \tag{6.22}$$

with the number of correlations $K_j^i(M)$ being the integer division

$$K_j^i(M) := \frac{M - T_j^i}{\omega_i}. \tag{6.23}$$

$K_j^i(M)$ is nothing else than the cardinal number of the set of correlation times $\mathbb{K}_j^i(M)$.

The compressed time trace $S_j^i(m)$ for bin (i, j) assuming that an elementary trace of length m has been supplied to the correlator reads with brackets denoting indexing, $q[i] = q_i$,

$$S_j^i(m) := \sum_{l=0}^{\omega_i - 1} q\left[m - 1 - T_j^i - l\right]. \tag{6.24}$$

Centered at the base time T_j^i the average is formed over ω_i elementary time bins. Not surprisingly, the first bin $(0, 0)$ always holds the 'freshest' elementary bin $S_0^0(m) = q_{m-1}$.

Each bin of the compressed trace is correlated with an earlier bin of the same compressed trace. The earlier time trace is separately encoded in the so-called zero bin $Z^i(m)$ for the octave i and again depends on the filling number m of the correlator. The content in the zero bin transverses through the correlator and becomes the first bin of the *i*th octave of the compressed trace exactly T_0^i steps later. Hence, it follows

$$Z^i(m) = S_0^i(m + T_0^i) \tag{6.25}$$

and it can inferred from Equation (6.24) that

$$Z^i(m) = \sum_{l=0}^{\omega_i - 1} q\left[m - 1 - l\right]. \tag{6.26}$$

The content of the un-normalized correlation function $C_j^i(M)$ after the full time trace has been supplied then reads

$$C_j^i(M) = \sum_{m \in \mathbb{K}_j^i(M)} S_j^i(m) Z^i(m). \tag{6.27}$$

Here, Equations (6.22), (6.24) and (6.27) have been used. Each bin of the compressed trace is correlated with its corresponding zero bin. Summation is over all filling numbers at which correlation takes place.

Thirdly, normalization is considered. To perform a symmetric normalization the sum of the counts having passed bin (i, j) as well as the sum of counts having passed the zero bin of octave i since bin (i, j) had been correlated for the first time has to be known.
(i) The sum of counts passed through bin (i, j), $R_j^i(M)$, is given by

$$\begin{aligned} R_j^i(M) &= \sum_{m \in \mathbb{K}_j^i(M)} S_j^i(n) \\ &= \sum_m \sum_{l=0} q\left[m - 1 - T_j^i - l\right] \\ &= q\left[0\right] + q\left[1\right] + ... + q\left[M - 1 - T_j^i\right] \\ &= \sum_{l=0}^{M-1-T_j^i} q_l. \end{aligned} \tag{6.28}$$

(ii) The sum of counts having passed the zero bin, $L_j^i(M)$, is given by

$$\begin{aligned} L_j^i(M) &= \sum_{m\in\mathbb{K}_j^i(M)} Z^i(m) \\ &= \sum_m \sum_{l=0} q\left[m-1-l\right] \\ &= q\left[T_j^i\right] + q\left[T_j^i+1\right] + ... + q\left[M-1\right] \\ &= \sum_{l=T_j^i}^{M-1} q_l. \end{aligned} \tag{6.29}$$

These are just the sums of the individual factors in Equation (6.27). To avoid redundancy one could take advantage of the fact that the zero bin is the same for every bin in an octave provided it already takes part in the correlation. Therefore, it is a computational advantage to 'freeze' the counts having passed the zero bin of octave i until bin (i,j) had been correlated for the first time. The zero bin $Z^i(M)$ is created for the first time when the zeroth bin $(i,j=0)$ is correlated for the first time, i.e. at T_1^i. The bin (i,j) is correlated for the first time at T_{j+1}^i. Hence, the freeze variable should be defined as

$$\begin{aligned} F_j^i &:= \sum_{m=T_1^i,T_1^i+\omega_i,..}^{T_j^i} Z^i(m) \\ &= \sum_m \sum_{l=0}^{\omega_i-1} q[m-1-l] \\ &= q\left[T_0^i\right] + q\left[T_0^i+1\right] + ... + q\left[T_j^i-1\right] \\ &= \sum_{l=T_0^i}^{T_j^i-1} q_l. \end{aligned} \tag{6.30}$$

F_j^i does not depend any more on the trace-length M once every bin of the correlator has been filled for the first time, $M > T_j^i$. If F_j^i is subtracted from the sum of values having passed the zero bin $Z_{\text{ges}}^i(M)$ with

$$\begin{aligned} Z_{\text{ges}}^i(M) &:= \sum_{m=T_0^i,T_0^i+\omega_i,..}^{M} Z^i(m) \\ &= \sum_{l=T_0^i}^{M-1} q_l, \end{aligned} \tag{6.31}$$

Equation (6.29) is recovered immediately. Compared to Equation (6.29) it is less consuming to keep $Z_{\text{ges}}^i(M)$ up to date once F_j^i has been established since it does not depend on j.

Finally the symmetrically normalized correlation from Equations (6.27), (6.28), (6.30) and (6.31) is computed as

$$G_j^i(M) = \frac{C_j^i(M)}{R_j^i(M)\left\{Z_{\text{ges}}^i(M) - F_j^i\right\}}. \tag{6.32}$$

In the subsequent the special example of the zeroth octave is discussed to further shed light on the formalism. Here one has $\omega_0 = 1, T_j^0 = j, j = 0...k_0 - 1$. Hence, the correlation factors, Equations (6.28) and (6.29), are given by

$$\begin{aligned} S_j^0(M) &= q\left[M - 1 - j\right] \\ Z^0(M) &= q\left[M - 1\right]. \end{aligned} \tag{6.33}$$

The correlation times of Equation (6.22) are given by

$$\mathbb{K}_j^0(M) = \left\{T_{j+1}^0, T_{j+1}^0 + 1, ..., M\right\} = \left\{j + 1, ..., M\right\}, \tag{6.34}$$

and the number of summands of Equation (6.23) is

$$K_j^0(M) = M - T_j^0 = M - j. \tag{6.35}$$

The correlation curve of Equation (6.27) becomes

$$C_j^0(M) = \sum_{m=j+1}^{M} S_j^0(m) Z^0(m) \tag{6.36}$$

$$= \sum_{m=j+1}^{M} q\left[m - 1 - j\right] q\left[m - 1\right] \tag{6.37}$$

$$= \sum_{m=0}^{M-1-j} q_m q_{m+j} \tag{6.38}$$

The normalization constants, Equations (6.28) and (6.29), are

$$\begin{aligned} R_j^0(M) &= \sum_{l=0}^{M-1-j} q_l \\ L_j^0(M) &= \sum_{l=j}^{M-1} q_l. \end{aligned} \tag{6.39}$$

As it should be, these are the results for a linear correlator.

The implementation of the correlator is most adequately done with a shift register. Thereby, the basic entity is the octave which is equipped with a trigger algorithm that manages the compressed trace, the correlation and the normalization sums. The octave has to be provided with a seed value q and the zero sum Z^{i-1} of the preceding octave. At initialization $Z^{-1} = q$ is taken. This is the linear correlation. The trigger function of the first octave is

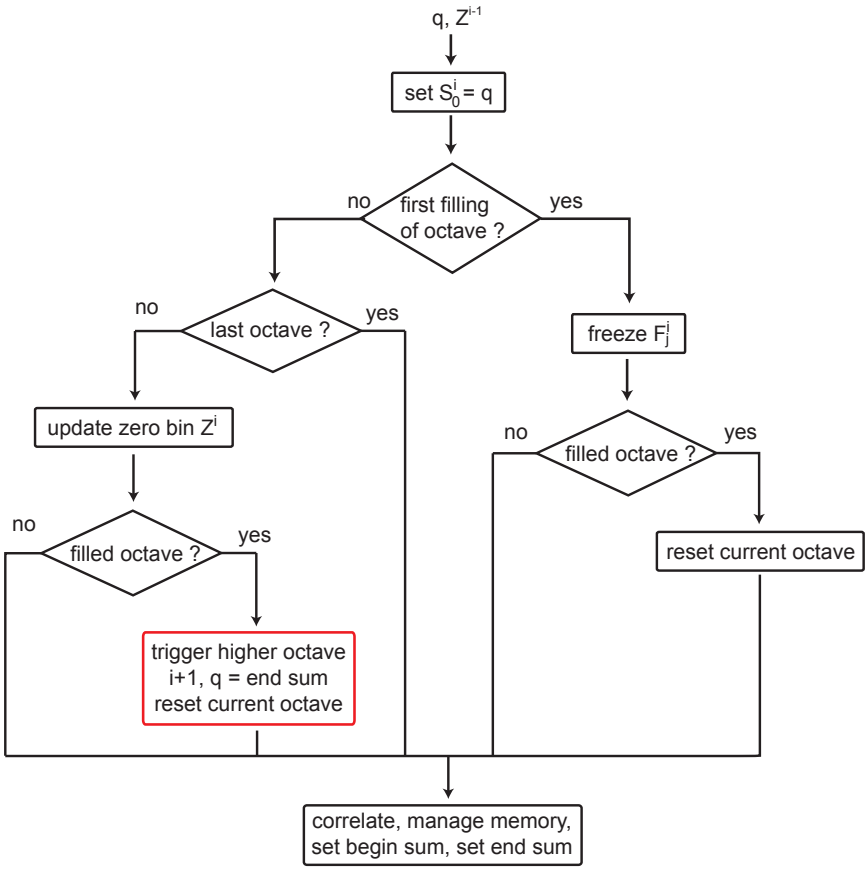

Figure 6.3: Flow diagram for triggering the octave i of a multi-τ-correlator. The algorithm is a shift register calling recursively the trigger routine for higher octaves $> i$ until the first octave $i_{\max}$ that is not yet ready for correlation is reached. The recursive call is highlighted (red). While ascending to higher octaves the compressed time trace is built up. Then, the algorithm descends back from $i_{\max}$ meanwhile performing the correlation. Bookkeeping the signal that flows into and out of the octave ('begin sum' and 'end sum') is necessary. Just after correlation of an octave its waiting time until the next correlation has to be reset. Apart from the standard recursive call with subsequent correlation several special cases have to be covered. (i) The maximum nesting depth is reached finally when arriving at the last octave. (ii) When octaves get filled for the first time the freeze variable has to be set.

called and the path descends down into higher octaves as long as the octaves are ready to be correlated. Thereby, it has to be monitored how large a signal is entering the octave (begin sum) and exiting the octave (end sum). Higher octaves, are called with the end sum of the next lower octave as a seed. Then, the correlations are calculated backwards from higher to lower octaves. In other words, the compressed time trace is built up when ascending from the first octave to lower octaves and the correlation is done when descending again. Figure (6.3) shows the corresponding flow diagram. In addition the standard pathway calling recursively higher octaves care has to be taken when arriving at the last possible octave. Furthermore, when filling an octave for the first time the freeze variable has to be kept. In the end, each correlated octave is shifted one bin to the right since the last bin flows to the next octave. This provides space for a new value to be supplied from the outside data series.

7 Fluorescence Correlation Spectroscopy with STED Focal Volumes

Recently, experimental work has been presented that aims at a combination of FCS and focal volume reduction by the STED method [12, 13]. With decreasing focal volume the relative intensity fluctuations due to single molecules diffusing out of the focus increase. The smaller the focal volume the smaller the mean intensity over which the fluctuations are superimposed. This follows from the fact that the mean intensity signal stems from the whole fluorophore ensemble in the observance volume. Hence, by using smaller volumes higher maximum fluorophore concentrations can be supported before reaching the detection limit. Further, nanoscale details in molecular pathways such as trapping are revealed that are otherwise averaged out in conventional detection.

In the following the signatures of the volume reduction as observed by FCS are studied. Firstly, the STED measurement scheme is outlined. The construction of the effective PSF is discussed and properties of the correlation curve that can already be inferred analytically are discussed. Then, results of a Monte-Carlo simulation of an FCS measurement with free particles diffusing in a STED-focus are presented and compared to experimental results.

7.1 Motivation and Analytical Study of the Focal Volume

STED microscopy [19, 5, 18] was first implemented as an extension of confocal microscopy [99]. Similarly to confocal microscopy, the fluorophores are excited with a laser beam of a diffraction-limited profile $h_{\text{exc}}(\mathbf{r})$. Next, the molecules are de-excited by a second doughnut-shaped beam of a profile $h_{\text{STED}}(\mathbf{r})$ with a more red-shifted wavelength than the excitation beam in order to assure de-excitation by stimulated emission and to avoid re-excitation. The molecules are de-excited over the whole periphery of the beam $h_{\text{STED}}(\mathbf{r})$ and stay unaffected at the center where $h_{\text{STED}}(\mathbf{r} = 0) = 0$. Although $h_{\text{STED}}(\mathbf{r})$ is diffraction-limited like every other far-field optical pattern the resulting de-excitation probability $\eta(\mathbf{r})$ may be much narrower than $h_{\text{exc}}(\mathbf{r})$. This originates from the strong non-linearity inherent in the rate equation kinetics of stimulated emission. Figure (7.1)[1] displays a typical STED setup.

In order to calculate the de-excitation probability $\eta(\mathbf{r}, I_{\text{STED}})$ the system of rate equations describing the photo-kinetics of the dye has to be solved. For pulsed de-excitation the de-

[1] The figure is courtesy of B. Harke, MPI f. Biophysical Chemistry, Göttingen.

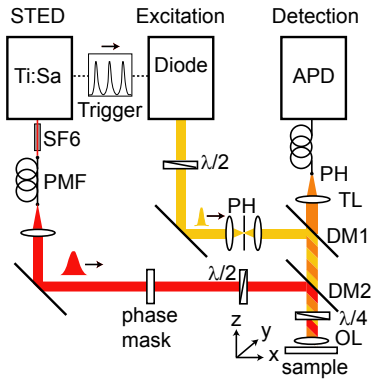

Figure 7.1: A typical STED setup. At a dichroic mirror (DM2) the excitation beam $h_{\text{exc}}(\mathbf{r})$ (yellow) is overlayed with a doughnut-shaped de-excitation beam $h_{\text{STED}}(\mathbf{r})$ (red). Both pulses are electronically synchronized by a trigger feedback loop. The doughnut shape is achieved by directing the STED beam on a phase mask. The overlayed beams are focused in the sample by an objective lens (OL). The fluorescence light (orange) is coupled out at another dichroic mirror (DM1) and recorded on an APD. The sub-resolution performance of STED relies decisively on the non-linearity of the photo-kinetics of the dye in conjunction with the fact that the STED beam $h_{\text{STED}}(\mathbf{r})$ features a zero value where the excitation beam $h_{\text{exc}}(\mathbf{r})$ has a maximum.

pletion factor $\eta(\mathbf{r}, I_{\text{STED}})$, i.e. the probability to be be de-excited by stimulated emission at position $\mathbf{r}$, is obtained as

$$\eta(\mathbf{r}, I_{\text{STED}}) = \exp\left[-\frac{I_{\text{STED}}}{I_{\text{sat}}} h_{\text{STED}}(\mathbf{r})\right]. \tag{7.1}$$

Here, the saturation intensity I_{sat} is characteristic of the dye and depends on the de-excitation cross-section, the rotational properties of the molecule, the polarization, the pulse length, the pulse duration and the slope of the STED beam [23]. The function $h_{\text{STED}}(\mathbf{r})$ can be calculated by vectorial diffraction theory using characteristic phase mask apodization functions [21]. Assuming the excitation, de-excitation and detection process to be independent the effective PSF $h(\mathbf{r})$ of the STED microscope reads [100]

$$h(\mathbf{r}, I_{\text{STED}}) = \eta(\mathbf{r}, I_{\text{STED}}) h_{\text{exc}}(\mathbf{r}) \left[h_{\text{det}}(\mathbf{r}) * \Theta(\mathbf{r})\right] \tag{7.2}$$

This is just the confocal imaging equation [59] scaled by the depletion factor $\eta(\mathbf{r})$. $\Theta(\mathbf{r})$ is the transmission function of the confocal pinhole placed into the detection path described by $h_{\text{det}}(\mathbf{r})$.

A first idea of the behavior of the normalized correlation curve $g(\tau, I_{\text{STED}})$ as a function of the STED intensity I_{STED} can be gained by considering the zero value of the curve $g_0(I_{\text{STED}}) = g(\tau = 0, I_{\text{STED}})$ separately. From Equations (6.1) and (6.3) it follows that

$$g_0(I_{\text{STED}}) = \frac{1}{CV(I_{\text{STED}})} \tag{7.3}$$

with the concentration C and the effective focal volume $V(I_{\text{STED}})$ defined by

$$V(I_{\text{STED}}) = \frac{\left[\int h(\mathbf{r}, I_{\text{STED}}) d^3\mathbf{r}\right]^2}{\int h^2(\mathbf{r}, I_{\text{STED}}) d^3\mathbf{r}}. \tag{7.4}$$

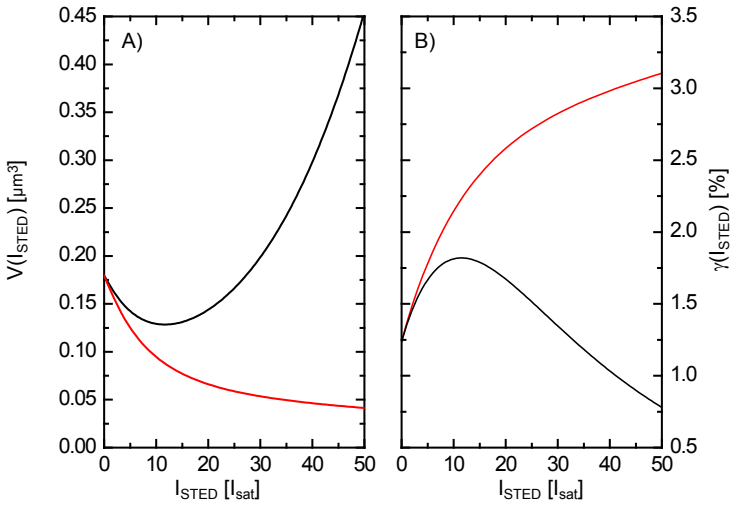

Figure 7.2: Effective volume $V(I_{\text{STED}})$ (A) and signal factor $\gamma(I_{\text{STED}})$ (B) as a function of the STED intensity I_{STED}. Results using a STED-PSF featuring a perfect zero point at $\mathbf{r} = 0$ (red) are opposed to those of an imperfect STED-PSF with a value at the origin featuring $\sim 2\%$ of the overall maximum (black). For the perfect STED-PSF the effective volume decreases monotonically with increasing STED intensity whereas it firstly decreases, then reaches a minimum and finally diverges for the imperfect STED-PSF. The process of decreasing effective volume is accompanied by an increase of the fraction of the signal stemming from the focal plane. Conversely, an increasing effective volume means more signal from the non-focal regions.

In mathematical terms, the value $g_0(I_{\text{STED}})$ measures the peakedness of the PSF $h(\mathbf{r}, I_{\text{STED}})$, i.e. the degree of sensitivity to squaring of the PSF.

Figure (7.2) presents a plot of the effective volume $V(I_{\text{STED}})$ for both an ideal STED-PSF and a STED-PSF with imperfect zero value at $\mathbf{r} = 0$. Moderate non-ideal zeros are frequently encountered in experiments. The phase mask was taken as a two-ring phase mask that consists of a circle of phase retardation π surrounded by a circular ring of phase retardation 0. The azimuth angle separating the two zones is $\vartheta = 39°$ for the perfect zero value and $\vartheta = 43°$ for the imperfect one. The imperfect zero node features $\sim 2\%$ of the maximum peak intensity of the full STED-PSF. Such a two-ring phase mask is commonly used in STED microscopy to achieve axial resolution increase [21]. The wavelengths of the PSFs are $\lambda_{\text{exc}} = 640$ nm, $\lambda_{\text{det}} = 673$ nm, $\lambda_{\text{STED}} = 780$ nm. The parameters are typical of the fluorophore MR121 [101]. The pinhole has radius $r = 25$ μm. The magnification in the detection path is $M = 100$. Obviously, the focal volume decreases steadily with increasing STED power with a perfect STED-PSF as opposed to the imperfect STED-PSF where the volume first decreases, then reaches a stable value and in the end grows without bounds. These processes are accompanied by a characteristic behavior of the fraction $\gamma(I_{\text{STED}})$ of signal originating from a slice of thickness 10 nm centered at the focal plane V_{fpl} relative to the total signal,

$$\gamma(I_{\text{STED}}) = \frac{\int_{V_{\text{fpl}}} h(\mathbf{r}, I_{\text{STED}}) d^3\mathbf{r}}{\int_V h(\mathbf{r}, I_{\text{STED}}) d^3\mathbf{r}}. \tag{7.5}$$

An increase in the focal volume is always linked to an increase of the off-plane signal fraction. A decreasing effective volume implies a larger amount of signal from the focal plane.

7.2 Phenomenology of STED-FCS Measurements

In the foregoing paragraph it has been found that the correlation curve of STED-FCS at $\tau = 0$ is fully determined by the effective focal volume of Equation (7.4). Here, a Monte-Carlo simulation is presented that delivers the correlation curve for different STED intensities I_{STED} and non-zero values of the lag time τ. This is of interest because in principle STED could be responsible for qualitative changes of the correlation curve that are not captured by the STED-intensity dependence of the amplitude. For instance, it could be possible that features will show up at certain non-zero lag times that equal the mean diffusion time from the central zero value to some adjacent secondary zero node of the STED beam. The Monte-Carlo simulations should reveal such effects.

Figure (7.3) presents the results using a STED-PSF featuring a perfect zero node. The wavelengths are the same as in the preceding paragraph. The diffusion constant is $D = 2.8 \cdot 10^6$ cm^2/s, a typical value for molecules in water. The time step is $\Delta t = 0.1$ μs leading to a mean relocation of ~ 8 nm. The number of simulation steps is $s = 3 \cdot 10^7$, the brightness is $q = 1 \cdot 10^3$. The molecule concentration amounts to $C = 15/\ \mu\text{m}^3$ with a support $2\ \mu\text{m} \times 2\ \mu\text{m} \times 2\ \mu\text{m}$. The growth behavior of the correlation curve and the decrease of the mean diffusion time through the focal volume are consistent with a monotonically decreasing effective volume with increasing STED intensity. As it should be, the amplitude follows directly the inverse effective volume of Figure (7.2). The curves are decreasing monotonically.

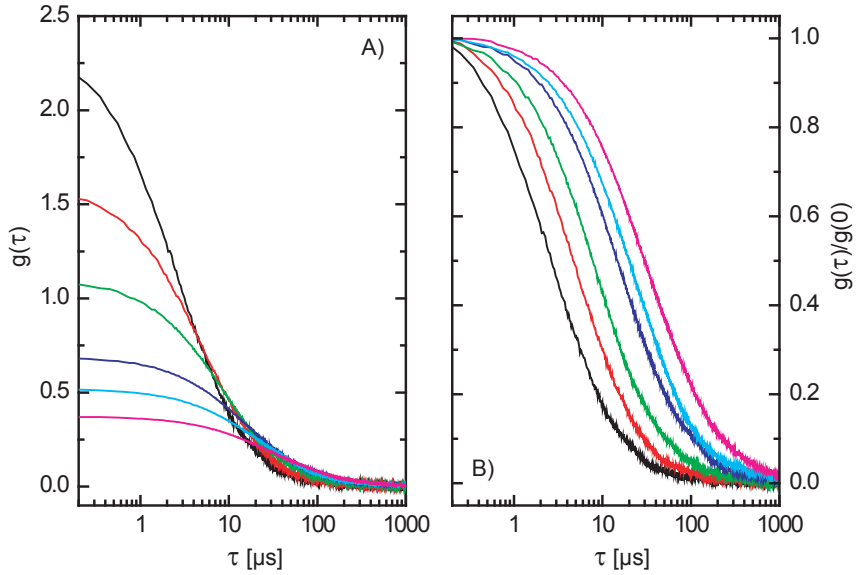

Figure 7.3: Normalized correlation curves $g(\tau, I_{\text{STED}})$ for STED-FCS generated by a Monte-Carlo simulation with a STED-PSF possessing a perfect zero point. The original correlation curves (A) and those normalized to unity (B) are shown. The STED-intensity covers the following values : $I_{\text{STED}} = 0 I_{\text{sat}}$ (magenta), $I_{\text{STED}} = 5 I_{\text{sat}}$ (cyan) $I_{\text{STED}} = 10 I_{\text{sat}}$ (blue), $I_{\text{STED}} = 25 I_{\text{sat}}$ (green), $I_{\text{STED}} = 50 I_{\text{sat}}$ (red) and $I_{\text{STED}} = 100 I_{\text{sat}}$ (black). On the left the growth in amplitude is fully consistent with the behavior of the effective volume in Figure (7.2). On the right it is seen how raising the STED intensity gradually lowers the mean diffusion time through the focal volume.

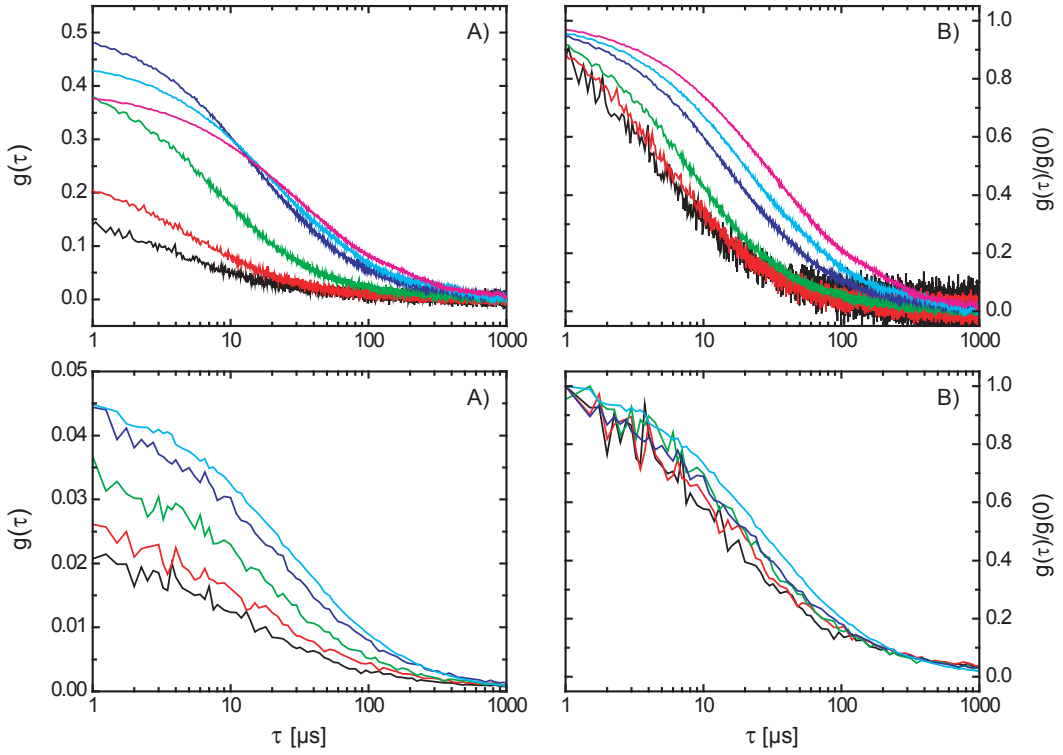

Figure 7.4: Normalized correlation curves $g(\tau, I_{\text{STED}})$ generated by a Monte-Carlo simulation with a STED-PSF featuring an imperfect zero point (top row) and high-quality experimental STED-FCS measurements on the fluorophore MR121 (bottom row). The standard correlation curves (A) and the curves normalized to unity are shown (B). For the calculations, the STED-intensity assumes the following values : $I_{\text{STED}} = 0I_{\text{sat}}$ (magenta), $I_{\text{STED}} = 5I_{\text{sat}}$ (cyan) $I_{\text{STED}} = 10I_{\text{sat}}$ (blue), $I_{\text{STED}} = 25I_{\text{sat}}$ (green), $I_{\text{STED}} = 40I_{\text{sat}}$ (red) and $I_{\text{STED}} = 50I_{\text{sat}}$ (black). For the experiment, the time-averaged power $\bar{P}$ of the STED-pulse was increased from 0 mW to 35 mW in equal steps (cyan, blue, green, red, black). Obviously, even the most sophisticated experimental measurements are influenced by off-axis fluorescence. A purer STED effect might probably achieved by confining the experiment to two dimensions where fluorescence originating from non-focal regions is suppressed [13].

Figure (7.4) gives the corresponding Monte-Carlo simulation for a STED-PSF featuring a non-ideal zero point together with experimental curves measured on the fluorophore MR121[2]. In the experiment the same phase mask geometry as for the calculations was used. The signal-to-noise ratio of the correlation curves decreases with increasing STED-power since the recorded signal decreases due to fluorescence depletion. This is stronger than in Figure (7.3) because the imperfect zero node does lead to additional depletion. The experimental measurements and the results of simulations coincide as far as the decrease of the diffusion time with increasing STED power is concerned. The volume effect is reproduced qualitatively. For 3D STED-FCS measurements with a circular π-phase mask the volume reduction effect is anti-proportional to the fraction of off-axis fluorescence.

[2]The measurement is courtesy of L. Kastrup, MPI f. Biophysical Chemistry, Göttingen.

8 Triplet State Dynamics

In practice, molecules often do not fluoresce at constant brightness but undergo state changes that affect their detectability such as bleaching or inter-system crossing into a dark triplet state. FCS is frequently used to determine the rates of these processes. The discussion is restricted to triplet state dynamics in the following but the results are equally applicable to any such process. The standard analysis of correlation curves assumes position-independent inter-system-crossing and triplet rates between the singlet and the triplet manifold, respectively. In reality, this assumption is violated because inter-system crossing rates are strongly dependent on the excitation profile. Here, a fit model is developed that concisely takes into account the spatial variations in the rates.

At first, the relevant photo-physical level schemes are introduced. The probabilities to reside in the singlet state S_1 are discussed for continuous-wave and pulsed excitation. A Monte-Carlo simulation is presented that clarifies the effect of spatially dependent rates as opposed to constant rates. Then, an analytical expression for the correlation curve for diffusion and position-dependent inter-system crossing rate is given. Stochastically simulated data is fitted with this curve and the resulting outcome is compared to the result of the standard procedure.

8.1 Photochemical Schemes and their Phenomenology in FCS

The triplet dynamics of typical fluorophores like Rh6G or Rh110 have been studied in a number of experiments [14, 102]. From that it is known that the quantum-mechanical energy structure might by described by a simple three-level scheme involving a singlet ground state S_0, an excited state S_1 and a triplet state T. The Jablonski diagram for this model is shown in Figure (8.1, A). The fluorescence signals from the three states are $q_0 = q_T = 0$ for the ground state and the triplet state, respectively, and

$$q_1(\mathbf{r}) = q h_{\text{det}}(\mathbf{r}) P_1(\mathbf{r}) \tag{8.1}$$

for the excited state S_1. q measures the molecular brightness, $h_{\text{det}}(\mathbf{r})$ is the collection efficiency function modeling the detection process and $P_1(\mathbf{r})$ describes the probability to reside in the excited state S_1 and is related to the excitation intensity $I(\mathbf{r})$ in a way characteristic for the excitation mode. The signal $q_1(\mathbf{r})$ depends on the position of the molecule. The excitation intensity $I(\mathbf{r})$ is position-dependent. The excitation cross-section and the fluorescence rate are given by σ and k_{10}, respectively. The inter-system crossing rate k_{T} and the triplet relaxation rate k_{isc} are physical parameters of the molecule.

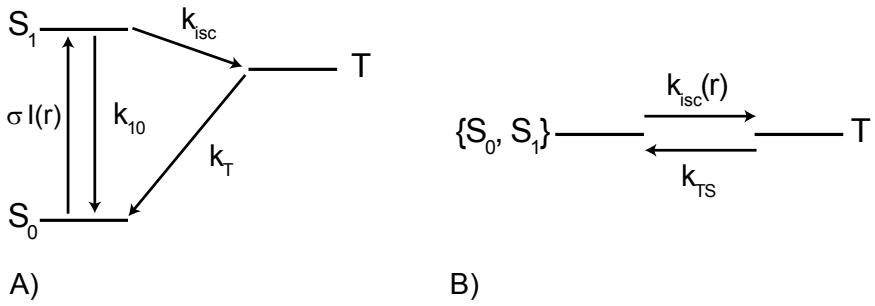

Figure 8.1: Photo-physical three-level scheme valid for typical fluorophores like Rh110 or Rh6G (A). The singlet system is driven with a position-dependent excitation rate $\sigma I(\mathbf{r})$ and the S_1-state has a lifetime of k_{10}^{-1}. The inter-system crossing rates k_{T} and k_{isc} are constant in this model. The population $P_1(\mathbf{r})$ of the state S_1 is position-dependent due to the variation of the excitation intensity $I(\mathbf{r})$. The model on the left is equivalent to a two-level model with position-dependent inter-system crossing rate $k_{\mathrm{isc}}(\mathbf{r})$ (B).

The three-level model in Figure (8.1, A) is equivalent to the two-level model of Figure (8.1, B) with a position-dependent inter-system crossing rate given by

$$k_{\mathrm{isc}}(\mathbf{r}) = k_{\mathrm{isc}} P_1(\mathbf{r}). \tag{8.2}$$

The effective inter-system crossing rate $k_{\mathrm{isc}}(\mathbf{r})$ is position-dependent because the triplet state can only be populated through the path $S_1 \rightarrow T$. In this way, the position-dependency of $P_1(\mathbf{r})$ is propagated into that of $k_{\mathrm{isc}}(\mathbf{r})$. For the case of the three-level model above the triplet relaxation rate k_{T} stays constant. However, this parameter would also depend on the position if triplet relaxation were stimulated by the excitation light $I(\mathbf{r})$. For simplicity, a constant k_{T} is used for all further simulations but all analytical calculations explicitly mention possible position dependencies in order not to restrict the discussion unnecessarily.

The probability for the molecule to reside in S_1 depends on the excitation mode. In the simple case of continuous wave excitation the probability $P_1(\mathbf{r})$ in the steady-state reads

$$P_1(\mathbf{r}) = \frac{\sigma I(\mathbf{r})}{\sigma I(\mathbf{r}) + k_{10}}. \tag{8.3}$$

This is fully analogous to the two-level molecule-type conversion probability of Equation (6.13). For excitation with rectangular pulses of length T_p and repetition period $T_r > T_p$ the mean S_1-probability has to be replaced by the formula [103]

$$\begin{aligned} P_1(\mathbf{r}) \;=\; & \frac{\sigma I(\mathbf{r})}{\sigma I(\mathbf{r}) + k_{10}} \frac{T_p}{T_r} + \\ & \frac{[\sigma I(\mathbf{r})]^2}{[\sigma I(\mathbf{r}) + k_{10}]^2} \frac{\left[e^{-k_{10}(T_r - T_p)} - 1\right]\left[e^{-(\sigma I(\mathbf{r}) + k_{10})T_p} - 1\right]}{[k_{10} T_r]\left[1 - e^{-k_{10}T_r - k_{01}T_p}\right]}. \end{aligned} \tag{8.4}$$

Here, the time-dependent rate equations have been solved piece-wise on the excitation interval $t \in [0, T_p]$ and the relaxation interval $t \in [T_p, T_r]$ for every $\mathbf{r}$. Then, the mean probability has been calculated by temporal integration. Clearly, Equation (8.4) reduces to Equation

(8.3) for $T_p = T_r$. For $T_p < T_r$ the formula of Equation (8.4) describes to which degree the molecule gets excited during the pulse interval $[0, T_p]$ and how complete the relaxation takes place after the pulse in the interval $[T_p, T_r]$.

Once, the excitation probability $P_1(\mathbf{r})$ is known a Monte-Carlo simulation can be performed to contrast the position-dependent model of Figure (8.1, B) with rate $k_{\text{isc}}(\mathbf{r})$ with the standard model having a position-independent rate k_{isc}. Figure (8.2) presents the results

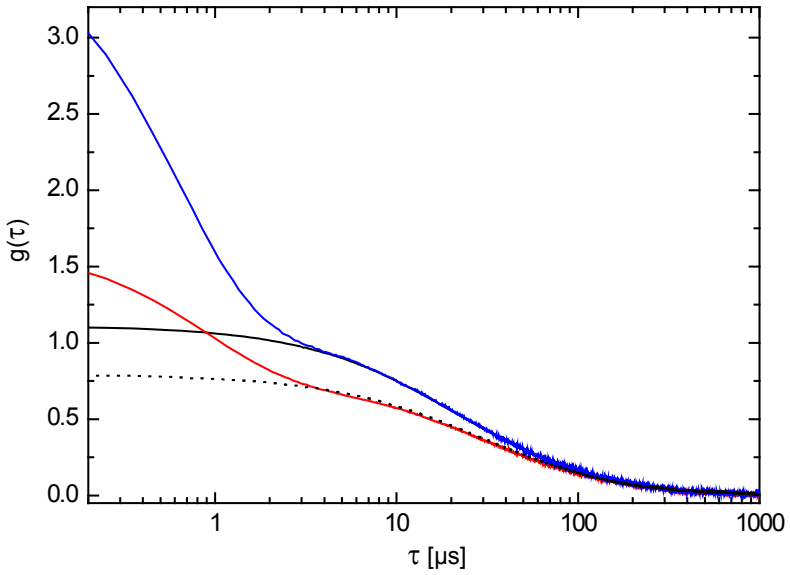

Figure 8.2: Correlation curves $g(\tau)$ estimated by a Monte-Carlo simulation for no triplet (black), space-independent inter-system crossing (blue) and space-dependent inter-system crossing (red). The excitation mode is continuous wave and the peak rates are $k_{\text{isc}} = 1.1\ \mu\text{s}^{-1}$ and $k_{\text{T}} = 0.49\ \mu\text{s}^{-1}$. The diffusion constant is $D = 2.8 \cdot 10^6\ \text{cm}^2/\text{s}$ as for Rh6G in water. The curve for the steady-state effective signal of Equation (8.5) is shown as a comparison (dotted black). The two-component shape of the correlation curve - triplet term at low τ, diffusion term at higher τ - is visible for both the space-dependent and the space-independent case. Nevertheless, it is clear that a fit of the red curve with the standard model assuming spatially constant rates would underestimate the value of k_{isc} and k_{T}.

for the case of no triplet, constant inter-system crossing rate and space-dependent inter-system crossing rate according to Equation (8.2). The excitation mode is continuous wave with intensities far away from saturation, $\sigma I_0 = k_{10}/10$. The excitation wavelength reads $\lambda_{\text{exc}} = 488$ nm. The peak rates are $k_{\text{isc}} = 1.1\ \mu\text{s}^{-1}$ and $k_{\text{T}} = 0.49\ \mu\text{s}^{-1}$ as has been measured for Rh6G [14]. The time step is $\Delta t = 0.1\ \mu$s. This leads to a mean diffusion relocation step of ~ 8 nm. The brightness q is set to $1 \cdot 10^3$ to speed up the convergence of the simulation. After $3 \cdot 10^7$ steps the correlation curves were practically noise-free. The correlator has octaves with $16, 8, ..$ bins at a relative factor of 2. The correlation curves with triplet transitions feature a two-component shape irrespectively of whether the rate k_{isc} is space-dependent or not: the triplet term dominates at low τ, the diffusion term at higher τ. However, the standard fit using space-independent rates of the curve for the space-dependent rates would strongly underestimate the rates. Not surprisingly, the diffusion part of this curve might be recovered

by a triplet-free simulation using the following effective signal $q_{\text{eff}}(\mathbf{r})$,

$$q_{\text{eff}}(\mathbf{r}) = \frac{k_{\text{T}}}{k_{\text{T}} + k_{\text{isc}} P_1(\mathbf{r})} q_1(\mathbf{r}). \tag{8.5}$$

This is just the triplet-free signal $q_1(\mathbf{r})$ of Equation (8.1) scaled by the steady-state probability to reside in the singlet system.

8.2 An Improved Model for Triplet Effects

An explicit fit model is required in order to be able to evaluate measured correlation curves. The standard argumentation progresses as follows. Assuming (i) a diffusion constant D unaffected by the state change, (ii) no fluorescence detected from a molecule in the triplet system and (iii) spatially constant rates k_{T} and k_{isc} it is straightforward to show that the correlation curve $g(\tau)$ calculated from Equation (6.3) is given by [38]

$$g(\tau) = g_{\text{D}}(\tau) g_{\text{T}}(\tau) \tag{8.6}$$

with the mean particle number N, the triplet term,

$$g_{\text{T}}(\tau) = 1 + A \exp(-\tau/\tau_{\text{T}}), \tag{8.7}$$

and where the parameters

$$A = \frac{k_{\text{isc}}}{k_{\text{T}}} \quad \text{and} \quad \tau_{\text{T}} = \frac{1}{k_{\text{isc}} + k_{\text{T}}}. \tag{8.8}$$

are constants describing the amplitude and the decay time of the triplet part of the correlation curve. $g_{\text{D}}(\tau)$ is the correlation curve for ordinary diffusion of Equation (6.8).

Assumption (iii) is not met in real experiments. Usually, the error originating from a violation of this assumption is heuristically corrected by using half the maximum intensity $I_0/2$ to relate the cross-section σ, the peak intensity I_0 and k_{isc} yielding

$$g(\tau) = g_{\text{D}}(\tau) \left[1 + A(I_0/2) \exp\left(-\frac{\tau}{\tau_{\text{T}}(I_0/2)} \right) \right]. \tag{8.9}$$

To describe the situation more adequately, rates with spatial dependence have to be introduced into the model. Assuming that the diffusion time τ_{D} is much larger than the triplet timescale $\max\{\tau_{\text{T}}(\mathbf{r})\}$ the molecules can be considered quasi-stationary in the time regime of triplet dynamics. Thus, triplet dynamics can be studied independently from the diffusion dynamics. Such an assumption is reasonable as long as the triplet rate $k_{\text{T}}(\mathbf{r})$ has a sufficiently high constant component such that $k_{\text{isc}}(\mathbf{r}) = 0$ at any point $\mathbf{r}$ does not lead to a divergence of τ_{T}. The correlation curve for a stationary two-component system one of which is dark has already been stated in Chapter (6.1). Defining the function $P_S(\mathbf{r})$ as the steady-state probability to reside in the singlet system

$$P_S(\mathbf{r}) = \frac{k_{\text{T}}(\mathbf{r})}{k_{\text{isc}}(\mathbf{r}) + k_{\text{T}}(\mathbf{r})} \tag{8.10}$$

Equation (6.11) transforms into

$$g(\tau) = c_{\text{tot}} q^2 \int h_{\text{det}}^2(\mathbf{r}) P_1^2(I(\mathbf{r})) P_S^2(I(\mathbf{r})) g_{\text{T}}(\tau, I(\mathbf{r})) d^3\mathbf{r} \tag{8.11}$$

The function $g_{\text{T}}(\tau, \mathbf{r})$ is the individual triplet correlation curve of Equation (8.7) that depends on the lag time τ and on the position through the intensity-dependent rates. The local brightness $q(\mathbf{r})$ is given by Equation (8.1) and is essentially described by the product of the probability to be excited $P_1(\mathbf{r})$ times the probability to be detected $h_{\text{det}}(\mathbf{r})$. It has been used that the concentration of molecules in the singlet system $c_s(\mathbf{r})$ is related to the total concentration c_{tot} in a simple way

$$c_s(\mathbf{r}) = c_{\text{tot}} P_S(I(\mathbf{r})). \tag{8.12}$$

Remarkably, the integrand in Equation (8.11) does depend explicitly on the position only through the detection probability $h_{\text{det}}(\mathbf{r})$. All other spatial dependencies are mediated through the spatial variation of the excitation profile $I(\mathbf{r})$. Thus, it is advantageous to rewrite this equation in a form that replaces the full three-dimensional space-integral by a one-dimensional integral over the relative brightness $x \in [0, 1]$,

$$g_{\text{T}}(\tau) = \int m^{(2)}(x) P_S(xI_0)^2 P_1(xI_0)^2 g_{\text{T}}(\tau, xI_0) dx, \tag{8.13}$$

with I_0 being the peak excitation intensity and the weight given by

$$m^{(i)}(x) = \int d^3r\, \delta\left(I(\mathbf{r})/I_0 - x\right) \left[h_{\text{det}}(\mathbf{r})\right]^i. \tag{8.14}$$

The function $m^{(i)}(x)$ is simply a measure of the size of the volume of constant relative brightness x conditioned on the detectability of signals originating from positions in this volume. It only depends on the detection efficiency function $h_{\text{det}}(\mathbf{r})$ and the form of the excitation pattern $I(\mathbf{r})/I_0$, not on any of the other parameters. $m^{(i)}(x)$ is a purely geometric factor and can therefore be pre-calculated before fitting the model reducing numerical effort from a three-dimensional to a one-dimensional integral.

Figure (8.3) displays $m^{(1)}(x)$ and $m^{(2)}(x)$ for typical excitation and detection PSFs with $\lambda_{\text{exc}} = 488$ nm, $\lambda_{\text{det}} = 515$ nm, confocal pinhole radius $r = 10\,\mu$m, magnification $M = 35$, lens refractive index $n = 1.33$ (oil immersion objective) and half aperture angle $\alpha = 64.5°$. Both functions show clear signatures of the oscillatory structure of the underlying PSF. Lower values of x feature a higher weight. The function $m^{(2)}(x)$ is roughly one order of magnitude smaller than the function $m^{(1)}(x)$. This results from the squaring of $h_{\text{det}}(\mathbf{r})$ in the integrand.

Using these functions $m^{(i)}(x)$ the normalized correlation curve turns out as

$$g(\tau) = \frac{\eta}{c_{\text{tot}}} \left(1 + \tau/\tau_{\text{D}}\right)^{-1} \left(1 + \epsilon\tau/\tau_{\text{D}}\right)^{-1/2} g_{\text{T}}(\tau) \tag{8.15}$$

with the diffusion term of Equation (6.8), $g_{\text{T}}(\tau)$ from Equation (8.13) and the normalization factor η given by

$$\eta = \frac{\left[\int m^{(1)}(x) p_1(xI_0) dx\right]^2}{\left[\int m^{(2)}(x) p_1(xI_0)^2 dx\right] \left[\int m^{(1)}(x) p_S(xI_0) p_1(xI_0) dx\right]^2} \tag{8.16}$$

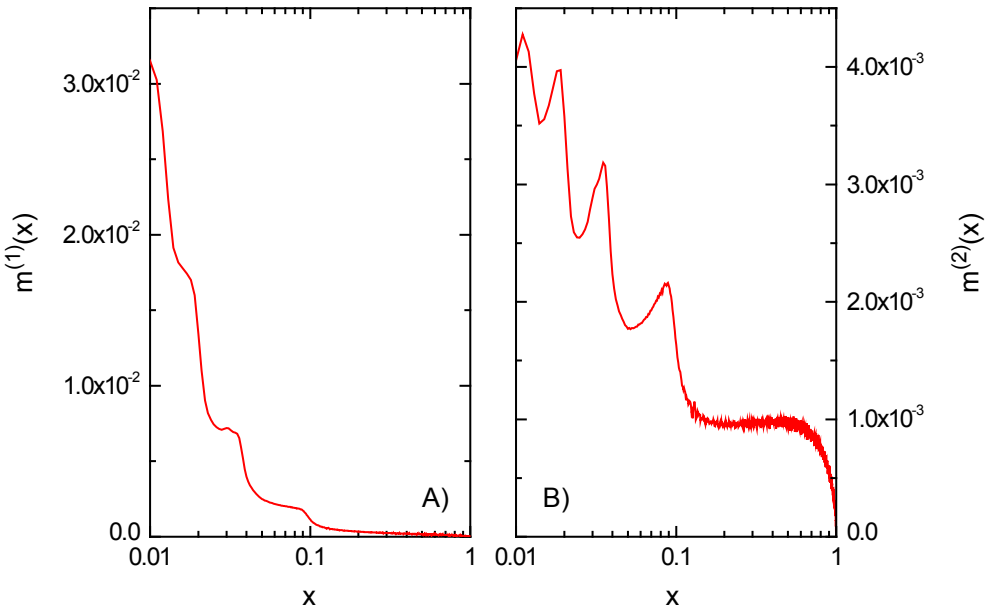

Figure 8.3: Measure $m^{(i)}(x)$ of the size of volumes of constant relative brightness x for the detection order parameter $i = 1$ (A) and $i = 2$ (B). Naturally, values of lower relative brightness x are equipped with higher weights. Both functions show clear signatures of the oscillations of the associated PSFs. $m^{(2)}(x)$ is about one order of magnitude smaller than $m^{(1)}(x)$ due to the squaring of $h_{\text{det}}(\mathbf{r})$ in the integrand of Equation (8.14).

so that Equation (8.15) reduces properly to Equation (8.6) for the case of spatially constant rates. η^{-1} plays the role of an effective volume. In the next section the performance of the improved model is contrasted with that of the standard model.

8.3 Results of Computations

In this section the evaluation of correlation curves obtained by Monte-Carlo simulations with the classical fit function of Equation (8.9) and the space-averaged fit function of Equation (8.15) are compared. Both, cw and pulsed excitation schemes leading to the S_1-population distributions of Equations (8.3) and (8.4) are considered.

Figure (8.4) presents the results. The dye is modeled with intersystem crossing rate $k_{\text{isc}} = 1\ \mu\text{s}^{-1}$, triplet rate $k_{\text{T}} = 0.2\ \mu\text{s}^{-1}$, excitation cross section $\sigma = 2.6 \cdot 10^{-20}$ m^2 and S_1-lifetime $k_{10}^{-1} = 3.6$ ns. This is typical of Rh110 [102]. In addition to the chemical kinetics the dye exerts Brownian motion with a diffusion constant $D = 3 \cdot 10^{-6}$ cm^2/s. The simulation run contained $3 \cdot 10^7$ relocation and intersystem crossing steps of length $\Delta\tau = 0.1\ \mu$s. Since

$$\Delta\tau \ll 1/(k_{\text{T}} + k_{\text{isc}}) \tag{8.17}$$

it is reasonable to consider only one-fold inter-conversion processes during a time step. If the triplet rates would be much higher one would have to chose a smaller time step $\Delta\tau$ or would have to include multiple inter-conversion terms. The simulated fluorescence time trace is correlated using a multiple-τ-correlator with 12 octaves of bin lengths $16, 8, 8, ..$ and

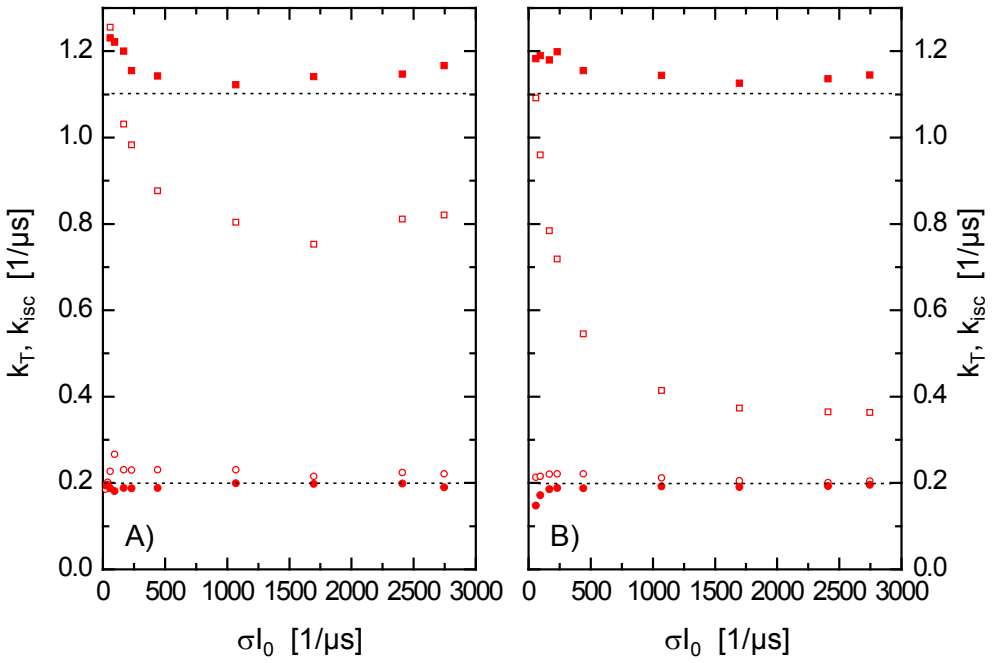

Figure 8.4: Mean fit results for the intersystem crossing rate k_{isc} and the triplet rate k_{T} using simulated correlation curves for continuous wave excitation (A) and pulsed excitation (B) at different peak excitation rates σI_0. The results of the standard fit model of Equation (8.9) are denoted by open symbols, those of the more sophisticated model of Equation (8.15) by the painted ones. The original rates $k_{\text{isc}} = 1.1\ \mu\text{s}^{-1}$ and $k_{\text{T}} = 0.2\ \mu\text{s}^{-1}$ assumed in the calculation are indicated by horizontal lines. The excitation rates k_{01} correspond to mean excitation powers up to 2.8 mW. The advanced method features a systematic error $< 10\%$. The standard method shows an underestimation of k_{isc} that grows with the intensity I_0. The deviation can reach up to 60% for this case. The bias is even worse for pulsed excitation as compared to cw excitation. The initially constant rate k_{T} is retrieved properly with both methods.

octave factors 2. As usual, the brightness $q_0 = 1 \cdot 10^3$ was chosen high enough to yield a stable quasi-noise-free simulation result. This is reasonable since the interest lays on the systematic bias and not on the fluctuations here. The final fit of the data was done on 10 independent correlation curves per intensity using the Gaussian least-squares method with equal weights [104]. The zero point $g(\tau = 0)$ of the correlation curve was excluded from the fit since the photon shot noise does not obey the assumption of symmetric errors of the Gaussian likelihood. The relative errors of the mean fit outcomes for the rates were smaller than 5% in all cases. Clearly, both methods achieve a correct retrieval of the constant rate k_{T}. The advanced model returns a more correct value of k_{isc} as the standard model. This effect is more pronounced for higher intensities. Systematic errors of the classical model are up to 30% for cw excitation and up to 60% for pulsed excitation. The systematic error of the more sophisticated model is $< 10\%$ for both excitation modes.

In the standard fit the inter-system crossing rate k_{isc} could only be retrieved properly in the limit of vanishing excitation intensity $I_0 \to 0$. However, for low excitation intensities the triplet effect is weak and measurement errors are high. This problem is circumvented by the new method which allows bias-free extraction of the rates at high excitation intensities.

9 Statistical Properties of the Correlation Curve

To retrieve parameters of interest the measured FCS correlation curve has to be fitted with a theoretical model like in the last chapter. Mostly, this is done by a classical least-squares fit. This assumes that the errors of the measured correlation curve are independently Gaussian-distributed. The signal-to-noise ratio of the correlation curve at fixed lag time, i.e. the variance var $[g(\tau)]$ has been studied theoretically [105, 47, 48, 106] as well as experimentally [107] for decades. Only in the last few years correlations between values of the correlation curve at different lag time values have been studied. However, the argumentation is partially refined to the estimation of the relaxation time spectrum in dynamic light scattering [49] or does not address the question to which accuracy parameters of interest can be retrieved [51]. A multivariate fit is proposed in [50]. Nevertheless, the whole work done so far assumes the correlation curve to be a single exponential $g(\tau) \sim \exp(-\tau/\tau_{\mathrm{D}})$ and does not account for the special situation of FCS. In this chapter, it is assessed whether the assumption of a diagonal covariance of the correlation curve can be maintained for FCS correlation curves.

The first section investigates the covariance matrix of the correlation curve for the simple model of free diffusion in a Gaussian focal volume. The second section develops a multivariate least-squares procedure for fitting correlation curves and elaborates on the influence of the non-vanishing covariance on the fitting of the diffusion time and the particle number.

9.1 Covariance of the Correlation Curve

Let the correlation curve at N discrete lag times $\{\tau_i\}, i = 1..N$ be given as $\{g_i\}, i = 1..N$. Then the covariance Σ_{ij} between the values g_i and g_j is defined as [86]

$$\Sigma_{ij} = \langle (g_i - \langle g_i \rangle)\,(g_j - \langle g_j \rangle) \rangle\,. \tag{9.1}$$

The brackets indicate the ensemble average over different stochastic realizations of the correlation curve. The correlation coefficient C_{ij} of the curve $\{g_i\}$ is defined as the normalized covariance

$$C_{ij} = \frac{\Sigma_{ij}}{\sigma_i \sigma_j} \tag{9.2}$$

with $\sigma_i^2 = \Sigma_{ii}$ being the ordinary variance of the random variable g_i. C_{ij} measures the correlation independently of the strength of individual fluctuations whereas Σ_{ij} captures both. Consequently, one has $C_{ii} = 1$. It can be shown that

$$-1 \leq C_{ij} \leq 1. \tag{9.3}$$

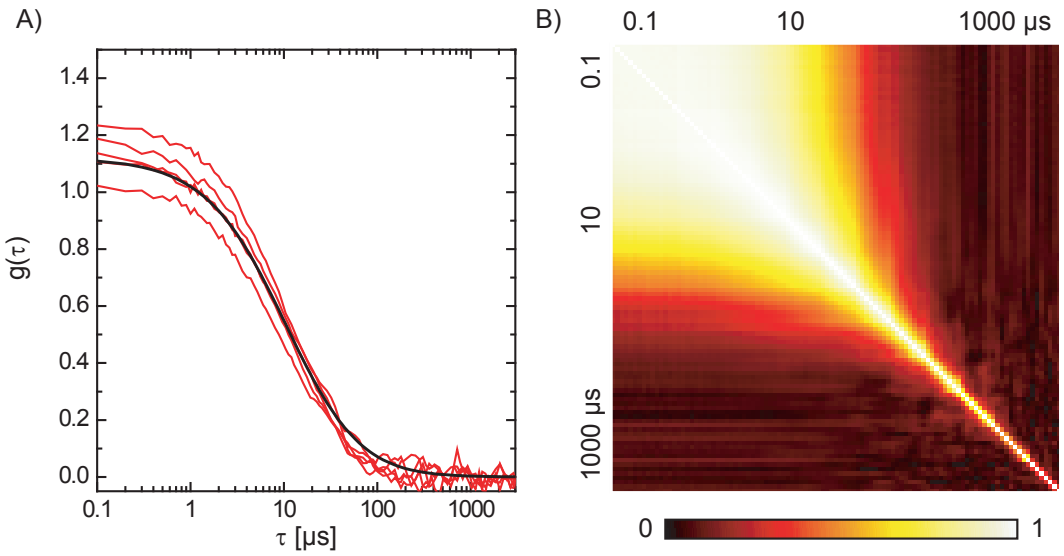

Figure 9.1: Some typical correlation curves (A) out of the ensemble of size $K = 2048$ that has been generated by a Monte-Carlo simulation to estimate the correlation coefficient $\{C_{ij}\}$ (B). The particles are diffusing freely in a Gaussian volume of full-widths-at-half-maximum $F_{x,y} = 190$ nm and $F_z = 470$ nm. The diffusion time amounts to $\tau_{\mathrm{D}}^{x,y} = 11$ μs and $\tau_{\mathrm{D}}^{z} = 66$ μs. The different realizations (red) of the correlation curve fluctuate around the exact result (black) given in Equation (6.8). There are strong positive correlations > 0.5 for both lag times < 30 μs while the correlations decay off to zero for at least one of the lag times being > 30 μs. This is due to the progressive smoothing of fluctuations at larger lag times. The octave transitions are visible as slight oscillations along the line (i, i).

Negative values indicate anti-correlation and positive values correlation. In the former case, two values tend to fluctuate in opposite directions while in the latter case they deviate in the same direction.

If a set of K of realizations of the correlation curve is provided as $\{g_i^k\}, i = 1..N, k = 1..K$ the covariance matrix Σ_{ij} may be estimated with the standard estimator

$$\hat{\Sigma}_{ij} = \frac{1}{K-1} \sum_{k=1}^{K} \left(g_i^k - \bar{g}_i\right) \left(g_j^k - \bar{g}_j\right). \tag{9.4}$$

Here, the ordinary mean $\bar{g}_i$ is as usual given by

$$\bar{g}_i = \frac{1}{K} \sum_{k=1}^{K} g_i^k. \tag{9.5}$$

Figure (9.1) presents the results of estimating the correlation coefficient using Equation (9.4) from an ensemble of correlation curves generated by a Monte-Carlo simulation for freely diffusing molecules. In this case, the expected mean correlation curve is given by the product of the hyperbolic functions of Equation (6.8). In the calculation, the focal volume has assumed to be the Gaussian of Equation (6.6) with $\omega_{xy} = 80$ nm and $\omega_z = 200$ nm. The diffusion constant is $D = 6 \cdot 10^{-6}$ cm^2/s and the time step $\Delta t = 0.1$ μs. Thus, the

mean diffusion relocation step amounts to 11 nm. The diffusion times are $\tau_{\mathrm{D}}^{x,y} = 11\ \mu$s and $\tau_{\mathrm{D}}^{z} = 66\ \mu$s. The fluorophore concentration is $C = 15\ \mu\mathrm{m}^{-3}$ yielding an amplitude $A = 1.12$ of the normalized correlation curve. The simulation volume is $2\ \mu\mathrm{m} \times 2\ \mu\mathrm{m} \times 4\ \mu\mathrm{m}$ with influx boundary conditions. A multi-τ-correlator featuring 12 octaves at widths $16, 8, 8, ...$ and augmentation factors 2 is used. Thus, the last bin has lag time $\tau = 3072\ \mu$s. The brightness of the molecule was taken as $q_0 = 50$ kHz [11] and the number of simulation steps was $s = 6 \cdot 10^7$, i.e. 6 s. The ensemble size is $K = 2048$.

The correlation curve features strong positive correlations if both lag times are $< 30\ \mu$s. For at least one lag time increasing above $30\ \mu$s the correlation approaches zero. Low-lag-time bins are highly correlated while higher-lag-time bins are not. This is consistent with the fact that fluctuations are averaged more strongly at larger lag times in the multi-τ-correlation scheme. Along the line through origin (i, i) there persist signatures of the transition between the correlator octaves.

9.2 Multivariate Gaussian Fit Method

Once it has been established that the covariance matrix Σ of the correlation curve contains significant off-axis elements it is necessary to investigate to which degree the correlation has an effect on the results of different fit procedures. Here, standard least-squares is judged against least-squares with correlated errors. Firstly, the error distribution is introduced. Then, the effect of the correlation on fitting synthetic data following exactly this distribution is studied. In a third step, it is assessed whether the assumption of a multivariate Gaussian is justified for the error distribution of FCS correlation curves and the performance of both fit schemes is evaluated on the realistic data.

9.2.1 The Statistical Models

For a maximum-likelihood fit a model of the data conditional on the parameters is needed. Here, the parameters are the diffusion time τ_{D} and the amplitude of the correlation curve A in Equation (6.8). The data is a measured vector $\mathbf{g}$ of length N describing the points of the experimental correlation curve. Assuming that the error fluctuates around the exact result in a Gaussian way with covariance Σ the distribution of $\mathbf{g}$ reads [86]

$$p(\mathbf{g}|\boldsymbol{\mu}, \Sigma) = \frac{1}{(2\pi)^{N/2}\sqrt{\det \Sigma}} \exp\left\{-\frac{1}{2}\left(\mathbf{g} - \boldsymbol{\mu}\right)^T \Sigma^{-1} \left(\mathbf{g} - \boldsymbol{\mu}\right)\right\} \tag{9.6}$$

Here, the mean $\boldsymbol{\mu}(\tau_{\mathrm{D}}, A)$ is the exact correlation curve and depends on the parameters. For standard non-correlated least-squares Σ is diagonal and Equation (9.6) becomes a product of independent Gaussians. The likelihood $L_{\mathrm{mvr}}(\mathbf{g}|\boldsymbol{\mu}, \Sigma)$ associated with Equation (9.6) reads with constant summands omitted

$$L_{\mathrm{mvr}}(\mathbf{g}|\boldsymbol{\mu}, \Sigma) = -\frac{1}{2}\left(\boldsymbol{g} - \boldsymbol{\mu}\right)^T \Sigma^{-1} \left(\boldsymbol{g} - \boldsymbol{\mu}\right). \tag{9.7}$$

The negative inverse of the covariance matrix Σ is the metric under which the likelihood becomes the euclidean length in the space of data values. This should be contrasted with the

likelihood $L_{\text{std}}(\mathbf{g}|\boldsymbol{\mu}, \boldsymbol{\sigma})$ for non-correlated Gaussian errors of variances $\sigma_i^2 = \Sigma_{ii}$

$$L_{\text{std}}(\mathbf{g}|\boldsymbol{\mu}, \boldsymbol{\sigma}) = -\frac{1}{2}\sum_{i=1}^{N}\left[\frac{g_i - \mu_i}{\sigma_i}\right]^2. \tag{9.8}$$

Fit results stemming from maximizing the likelihoods $L_{\text{mvr}}(\mathbf{g}|\boldsymbol{\mu}, \Sigma)$ and $L_{\text{std}}(\mathbf{g}|\boldsymbol{\mu}, \boldsymbol{\sigma})$, respectively, will be compared in the subsequent.

9.2.2 Significance of the Effect of Off-Axis Correlations

It is not a priori clear that FCS data follows a multivariate Gaussian. To study whether the correlations can in principle have a considerable impact it is advantageous to work with data of a known type of distribution. Thus, a first artificial ensemble of correlation curves was generated by degrading the exact curve with multivariate Gaussian noise according to Equation (9.6). The covariance matrix Σ was prescribed and was estimated from the same ensemble of real FCS curves that provided the correlation matrix in Figure (9.1, B). The histograms of relative frequencies for the fit results are shown in Figure (9.2). One million correlation curves have been fitted by a simplex maximization of the likelihood functions

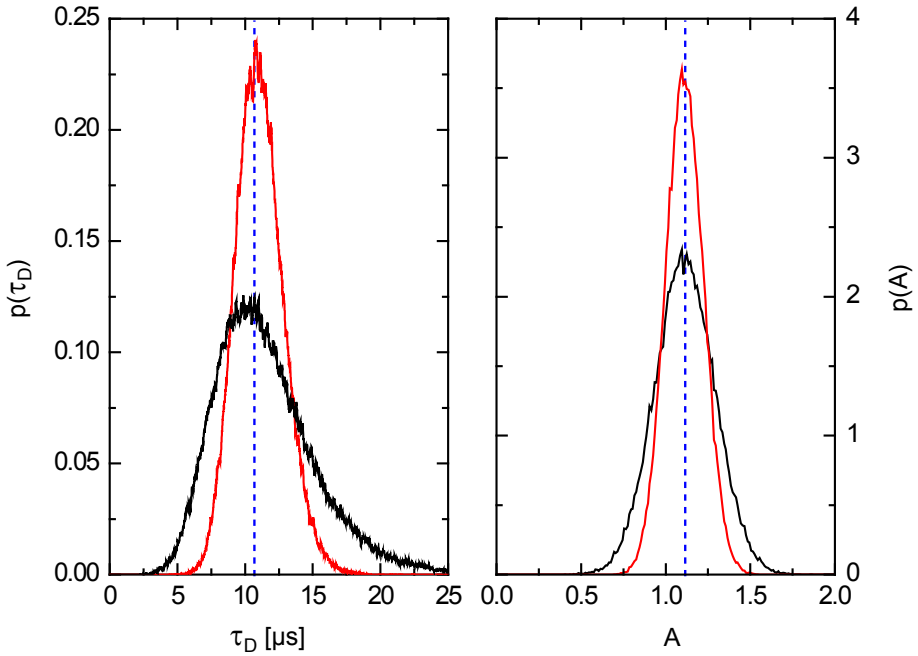

Figure 9.2: Relative frequencies of fit results $p(\tau_{\text{D}})$ and $p(A)$ for FCS-curves suffering from multivariate Gaussian noise. The covariance matrix Σ has been estimated from the same ensemble of correlation curves as the correlation matrix in Figure (9.1, B). $1 \cdot 10^6$ correlation curves have been fitted with the multivariate least-squares method (red) and the standard least-squares method (black). For that, the likelihood functions of Equations (9.7) and (9.8) have been maximized. The true values are indicated as vertical lines (blue). Clearly, the Gaussian fit neglecting the correlations but taking into account properly the single-point variances yields a less credible result. This is more pronounced for the diffusion time than for the amplitude.

$L_{\text{mvr}}(\mathbf{g}|\boldsymbol{\mu}, \Sigma)$ and $L_{\text{std}}(\mathbf{g}|\boldsymbol{\mu}, \boldsymbol{\sigma})$. The fit has been initialized with the true values and has been terminated if the change in the least-squares sum normalized to unity (the χ^2-value) has been smaller than 0.05 for 20 times. In general, this was the case after ~ 25 iterations. The different fit results for the diffusion time and the amplitude have been tabulated according to their occurrence. Obviously, the correlation has a significant influence. The results of the multivariate fit are always narrower than those for the standard least-squares procedure that neglects the off-axis correlations but takes properly into account the individual variances. Furthermore, the distributions are symmetric for multivariate fitting whereas the standard fit result displays a slight asymmetry of the distribution of the diffusion time. Also, the normalized χ^2-values indicate a higher level of credibility for the multivariate fit than for the least-squares fit. This will be discussed in more detail in the following section.

9.2.3 Performance on Typical Stochastic Data

Before fitting real FCS correlation curves it is necessary to investigate whether their errors are of multivariate Gaussian type. It is possible to de-correlate data values that follow a multivariate Gaussian distribution [86]. The transformed data values $\tilde{\boldsymbol{g}}$ defined by

$$\tilde{\boldsymbol{g}} = \Sigma^{-1/2} \left(\mathbf{g} - \boldsymbol{\mu} \right) \tag{9.9}$$

are non-correlated and independently normally distributed with mean 0 and variance 1. Here, $\Sigma^{1/2}$ denotes the Cholesky decomposition of the covariance matrix Σ. The distribution of the de-correlated values can be examined by a probability-probability plot [108]. Thereby, the cumulative densities $P_{\text{exp}}(\tilde{g})$ and $P_{\text{theo}}(\tilde{g})$ for the experimental fluctuation and the theoretical expectation are plotted in the same diagram. The de-correlated fluctuation value $\tilde{g} \in [-\infty, \infty]$ is the arc-length parameter $\tilde{g}$ of the resulting curve $[P_{\text{theo}}(\tilde{g}), P_{\text{exp}}(\tilde{g})]$. For the case of the Gaussian with mean 0 and variance 1 the theoretical cumulative density is given by the error integral [52]

$$P_{\text{theo}}(\tilde{g}) = \int_{\tilde{g}}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-s^2/2\right) ds. \tag{9.10}$$

The experimental cumulative density $P_{\text{exp}}(\tilde{g})$ is estimated from the values $\tilde{g}_i$ by first sorting them in ascending order and then counting the relative frequencies of values smaller than $\tilde{g}$. Here, it is of importance to notice that after the transformation according to Equation (9.9) all fluctuations have the same distribution $N(0, 1)$, i.e. each curve yields N Gaussian deviates. It can be shown that the slope of the PP-curve is proportional to the quotient σ_1/σ_2 when testing a Gaussian against another Gaussian with different mean values μ_1, μ_2 and variances σ_1, σ_2. The difference in the mean values affects merely the edge regions of the curve.

Figure (9.3) displays the PP-plot for fluctuations stemming from several correlations curves. The curves used for the plot are different from the curves employed to estimate the covariance matrix and therefore statistically independent of the latter. Obviously, no severe deviations from a straight line can be recognized. The assumption of multivariate Gaussian noise on FCS correlation curves is reasonable.

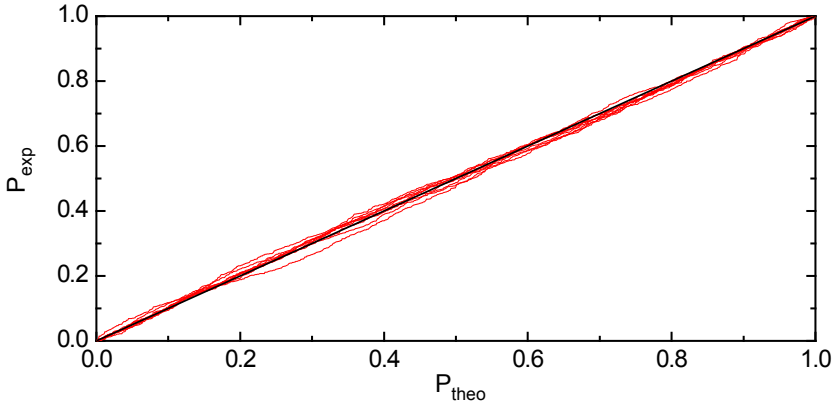

Figure 9.3: Probability-probability plots for some realizations out of an ensemble of FCS correlations curves. The data has been transformed according to Equation (9.9) and its experimental cumulative density $P_{\text{exp}}(\tilde{g})$ is plotted against the theoretical cumulative density $P_{\text{theo}}(\tilde{g})$ for a Gaussian of mean zero and variance one. The de-correlated fluctuation $\tilde{g} \in [-\infty, \infty]$ is the arc-length parameter on the curve. The extreme fluctuations contribute to the end points of the curves while moderate values affect the center. If the de-correlated data is independently Gaussian distributed the PP-plot has to yield a line through the origin. This seems to be the case. The data traces used for the plot are independent from the traces which were used for estimation of the covariance matrix Σ.

Using the de-correlated data values $\tilde{\mathbf{g}}$ it follows from Equations (9.6) and (9.9) that the likelihood $L_{\text{mvr}}(\tilde{\mathbf{g}})$ is given by

$$L_{\text{mvr}}(\tilde{\mathbf{g}}) = -\frac{1}{2}\tilde{\mathbf{g}}^T\tilde{\mathbf{g}} = -\frac{1}{2}\sum_{i=1}^{N}\tilde{g}_i^2. \tag{9.11}$$

Since $-2L_{\text{mvr}}(\tilde{\mathbf{g}})$ is a sum of N squares of normal $\sim N(0,1)$ variables it should follow a χ^2_N-distribution with N degrees of freedom [86]. The χ^2_N-distribution has a mean value of N. Usually, the χ^2-value used as a goodness-of-fit parameter is defined as the normalized sum of squared deviations, i.e. as

$$\chi^2 = \frac{1}{N}\sum_{i=1}^{N}\tilde{g}_i^2. \tag{9.12}$$

This value should be distributed around 1. A mean value smaller than 1 indicates that the fit is too good compared to what the knowledge of the error statistics lets expect.

Figure (9.4) presents the distributions of the fit results for τ_{D} and A according to the two methods for realistic FCS correlation curves. The $K = 2048$ correlation curves have been generated by a Monte-Carlo calculation analogous to the one used for the estimation of the covariance matrix so that the covariance is given by Σ but the estimator $\hat{\Sigma}$ is statistically independent from the correlation curves. In general, the multivariate fit is always of the same or higher quality as the standard least-squares fit. The diffusion time is retrieved by the standard least-squares algorithm with a systematic error of -2 μs while the systematic error is -1 μs for the multivariate fit. The amplitudes are retrieved at equal quality but with

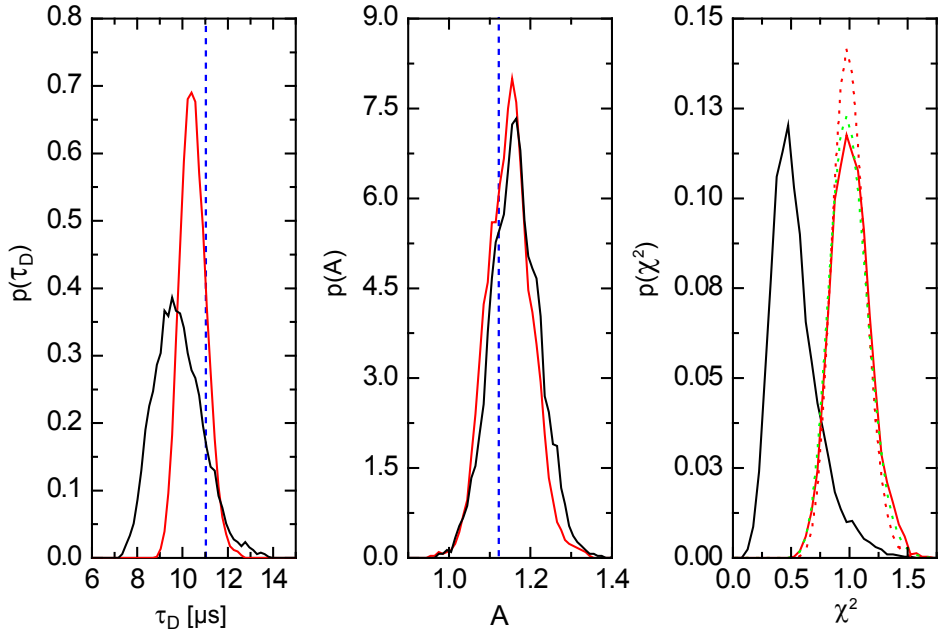

Figure 9.4: Relative frequencies $p(\tau_D)$ and $p(A)$ of the fit results for FCS correlation curves suffering from diffusion-noise and photo-detection shot noise. The results of the standard least-squares fit (black) are judged against those of the multivariate least-squares method (red). The original values are indicated (blue). The distribution of the experimental normalized χ^2-residuum and the theoretical χ^2-distribution function for $N = 102$ (dotted red) and $N = 75$ (dotted green) are shown. As in Figure (9.2) with the theoretical noise the multivariate fit retrieves the diffusion time with less spread than the least-squares fit. The amplitude fit is of the same quality for both methods. Both fit procedures tend to slightly underestimate the diffusion time and to overestimate the amplitude. The normalized χ^2-value is distributed around 1 as it should be for the multivariate fit and by a factor 0.5 too low on average for the standard least-squares. This favors the credibility of the multivariate fit. However, there seem still to be hidden correlations in the data since for the number of degrees of freedom $N = 75$, which is lower than the expected $N = 102$, the theoretical χ^2_N-distribution better fits the χ^2-residuum.

a small bias of $+0.05$. It makes sense that an overestimation of A is accompanied by an underestimation of τ_D because for higher amplitude the correlation curve has to decay off stronger to reach the same final value. The distribution of the χ^2-residua is reported together with theoretical χ^2_N-distributions for $N = 75$ and $N = 102$ degrees of freedom. The latter is just the number of data points, i.e., the length of the correlator $N = 104$, minus the number of fit parameters. The standard least-squares fit features a mean $\chi^2 = 0.5$ which is too low to be acceptable. The multivariate fit yields the expected mean $\chi^2 = 1$. However, the experimental distribution of χ^2-values is better described by a χ^2_N-distribution with a number of degrees of freedom $N = 75$ that is lower by about 25 than the expected degree of freedom $N = 102$. Hidden correlations seem to exist which are not captured adequately by second order covariance. It is astonishing that the PP-plot nevertheless has yielded such a good agreement with a multivariate Gaussian for which higher-order correlations are entirely determined by second-order correlations. The multivariate fit delivers results superior to the standard least-squares procedure that is commonly used in the analysis of correlation curves.

Conclusion

The gradual elimination of ensemble averaging from fluorescence imaging by focal volume reduction has led to an increasing importance of statistical concepts in this field. In this thesis analytical and numerical stochastic methods have been used to model mathematically the imaging process of PALMIRA microscopy for the first time and to develop advanced methods for the computation and evaluation of fluorescence time traces in complicated focal geometries. Thereby, more adequate data evaluation procedures and the quantitative description of several effects which had been described merely qualitatively have been gained.

It has been found that the point-spread-function of PALMIRA can be expressed with the help of Lerch's ϕ-transcendent in closed form and that the imaging equation is given as a convolution like in confocal microscopy as long as molecules are imaged independently and as long as the localization accuracies are translation invariant. The resolution, the temporal stretch of the readout and the level of confidence of an image have been expressed using simple formulas as a function of the mean photon number, the photon cut-off and the sampling degree. A variety of different position estimation schemes - multi-channel astigmatic, 4pi-, defocus and multi-point imaging - have been investigated using the tool of Fisher information and Cramer-Rao bounds in addition to numerical calculations. The recognition of objects has been analyzed quantitatively with respect to the spatial variation of sampling frequencies due to thresholding effects, the distribution of detected photon numbers in the presence of background noise, the occurrence of multi-molecule events with overlapping optical patterns and the discernability of objects according to their color and polarization. A deconvolution procedure taking into account the event photon number as a quality index for the position estimate has been introduced, tested on synthetic and experimental data and shown superior to a simple occurence histogram neglecting the photon degree of freedom.

Furthermore, it has been confirmed that the strong focal volume reduction achieved by stimulated emission using a doughnut-shaped depletion beam (STED) can be detected conveniently by fluorescence correlation spectroscopy. Analytical expressions as well as Monte-Carlo simulations with varying STED intensity have yielded that the reduction effect is proportional to the amount of signal originating from the focal plane and that un-correlated off-focus noise has to be minimized when a STED effect as pure as possible is to be observed. An improved fit model for triplet state dynamics relying on a concise treatment of the spatial dependency of the inter-system crossing rate has been shown to be superior to the standard model using space-independent rates and a mean excitation intensity. Moreover, a multivariate noise model taking into account the non-vanishing covariance of correlation curves has been formulated. The quality of the associated maximum-likelihood fit has turned out to be superior to the common Gaussian least-squares procedure.

A Richards-Wolf Vectorial Diffraction Theory

It is necessary to document properly the way in which the ordinary as well as the astigmatic PSFs employed in this thesis have been calculated. Since high-numerical aperture objectives are employed standard diffraction theory [57] has to be generalized. This has been done in the work of Richards and Wolf [56, 109]. The point where the concrete physical structure of the setup enters the calculation is the apodization function $\Psi(\phi, \theta)$. This function describes the phase and amplitude changes due to the aperture. For a simple lens, $\Psi(\phi, \theta)$ models merely the finiteness of the optical aperture. The apodization function in Equation (3.12) describes an astigmatic aberration. Again different apodization functions have to be used for the phase mask filter creating the doughnut-shaped STED beam [21]. The integral for the electric field $\boldsymbol{E}$ to be evaluated is given by Equation (2.26) in the canonical work of Richards and Wolf [56]

$$\begin{aligned} E_x &= -\frac{iA}{\pi}\int\int \Psi(\theta,\phi)\sqrt{\cos\theta}\sin\theta\left(\cos\theta+(1-\cos\theta)\sin^2\phi\right)e^{ikr\cos\epsilon}d\theta d\phi \\ E_y &= \frac{iA}{\pi}\int\int \Psi(\theta,\phi)\sqrt{\cos\theta}\sin\theta\,(1-\cos\theta)\cos\phi\sin\phi e^{ikr\cos\epsilon}d\theta d\phi \\ E_z &= \frac{iA}{\pi}\int\int \Psi(\theta,\phi)\sqrt{\cos\theta}\sin^2\theta\cos\phi e^{ikr\cos\epsilon}d\theta d\phi \end{aligned} \tag{A.1}$$

with r, ϕ, θ being spherical polar coordinates of the aperture with the polar axis $\theta = 0$ in the z-direction, k the wavenumber and ϵ an angle defined in Equation (2.20) of [56]. The integral stated above has been calculated numerically by Gauss-Legendre integration.

wavelength	$\lambda = 488$ nm or 575 nm
refractive index	$n = 1.515$
semi-aperture angle	$\alpha = 64.5°$
integration support	$[4000\text{ nm}, 4000\text{ nm}, 2000\text{ nm}]$
integration grid	$(201, 201, 401)$
number of Legendre polynomials	80

Table A.1: Typical parameters for the calculation of PSFs by high-numerial aperture vectorial diffraction theory. If not stated otherwise, these parameters have been used for all PSFs employed in this thesis. Detection PSFs are calculated by averaging several PSFs of linear excitation covering the whole solid angle $[0, 2\pi]$ in 24 equal steps.

B Distribution of Event Photon Numbers

This sections aims at deriving the probability $p(N)$, Equation (1.8), to record N photons in a single frame. The parameters are the mean photon number until switching off $\bar{N}$ and the mean number of photo-electrons $\bar{N}_{\text{fl}}$ per frame for a fluorophore that is constantly active. The idea here is to measure the length of the frame T_{fl} in units of excitation cycles of the fluorescence transition $\bar{N}_{\text{fl}}$. In Equation (1.10) is has been shown that these two values are proportional to each other. If the fluorophore is practically infinitely long-lived, $\bar{N} \gg \bar{N}_{\text{fl}}$, the molecule radiates always over the entire time and the distribution of photo-electrons is the Poisson distribution of Equation (1.9). If the fluorophore switches off after a short fraction of the image frame, $\bar{N} \ll \bar{N}_{\text{fl}}$, many frames will see the whole fluorescence cycle of the molecule and the photon distribution is geometric like in Equation (1.5). Figure (B.1) sketches these two cases.

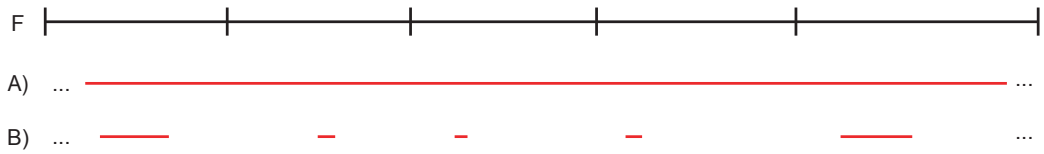

Figure B.1: Different realizations of the distribution of photon number among image frames (F). The two extreme cases are either an switched-on state of the molecule long-lived compared to the frame duration, $\tau_{\text{sw}} \gg T_{\text{fl}}$ (A) or a lifetime of the switched-on state very short to the frame time, $\tau_{\text{sw}} \ll T_{\text{fl}}$ (B). This corresponds to mean photon numbers per switched-on state $\bar{N}$ much larger or smaller than the mean number of photo-electron $\bar{N}_{\text{fl}}$ in a frame, $\bar{N} \gg \bar{N}_{\text{fl}}$ or $\bar{N} \ll \bar{N}_{\text{fl}}$, respectively. For simplicity, it has been assumed that one and only one switching takes place per frame.

If one assumes that one and only one switching takes place per frame the probability $p(N)$ is can be directly written down,

$$p(N) = \left[\frac{\bar{N}}{\bar{N}+1}\right]^N \left\{ p_{\text{poi}}(N|\bar{N}_{\text{fl}}) + \frac{1}{\bar{N}+1}\left(1 - \sum_{n=0}^{N} p_{\text{poi}}(n|\bar{N}_{\text{fl}})\right)\right\} \tag{B.1}$$

because the logical structure is obviously

$$\mathbf{X}^N \wedge \{\mathbf{Y}_0 \vee (\neg\mathbf{X} \wedge \mathbf{Y}_1)\}\,. \tag{B.2}$$

It has simply been used that the joint probability of independent events is the product of the individual probabilities and that the probability for the union of two events $\mathbf{X}$ and $\mathbf{Y}$ obeys

$$P(\mathbf{X} \cup \mathbf{Y}) = P(\mathbf{X}) + P(\mathbf{Y}) - P(\mathbf{X} \cap \mathbf{Y}). \tag{B.3}$$

In other words, logical AND corresponds to multiplication and logical OR to addition. The different events arising are

event	content	probability
X	emission of a photon	$\frac{\bar{N}}{\bar{N}+1}$
$\neg$**X**	bleaching in this emission cycle	$\frac{1}{\bar{N}+1}$
$\mathbf{Y}_0$	the frame is filled	$p_{\text{poi}}(N\|\bar{N}_{\text{fl}})$
$\mathbf{Y}_1$	the frame is not yet filled	$\sum_{n>N} p_{\text{poi}}(n\|\bar{N}_{\text{fl}})$

Figure (B.2) presents a plot of this probability as a function of N for the fixed mean value $\bar{N} = 100$ and different frame lengths $\bar{N}_{\text{fl}}$. It should be mentioned that the case of more than one switching per frame is much more complicated since the number of possible switchings depends on the history of the radiated photon number. Roughly, this can be accounted for with a higher $\bar{N}$ since the individual contributions are added over a frame.

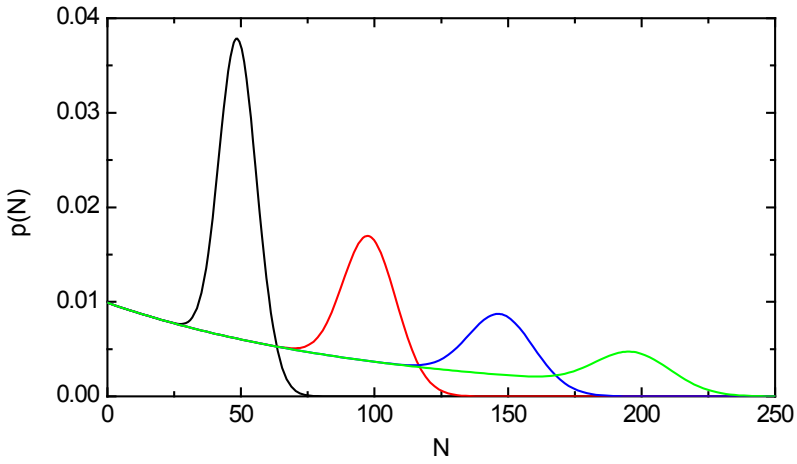

Figure B.2: The distribution of photon numbers per frame $p(N)$ for a mean photon number until switching back of $\bar{N} = 100$ and several mean frame lengths $\bar{N}_{\text{fl}} = 50$ (black), $\bar{N}_{\text{fl}} = 100$ (red), $\bar{N}_{\text{fl}} = 150$ (blue) and $\bar{N}_{\text{fl}} = 200$ (green). Radiation cycles longer than one frame are split up into smaller units with average length $\bar{N}_{\text{fl}}$. This is the more pronounced the smaller $\bar{N}_{\text{fl}}$ becomes.

Abbreviations and Conventions

Names

PSF	point-spread-function
STED	stimulated emission depletion
RESOLFT	reversible saturable optical fluorescence transitions
PALM	photo-activation localization microscopy
PALMIRA	photo-activation localization microscopy with independently running acquisition
CCD	charge-coupled device
APD	avalanche-photo-diode
FWHM	full-width-at-half-maximum
FFT	fast Fourier transform
PP-Plot	probability-probability plot
cw	continuous wave
isc	inter-system crossing

Variables

N	photon number
$\bar{N}$	mean photon number
σ^2	variance
$\mathbf{r}$	position
$\hat{\mathbf{r}}$	position estimator
$\bar{\mathbf{r}}$	estimated position
$g(\tau)$	correlation curve
$I(\mathbf{r})$	intensity profile
I_0	peak intensity
$h_{\text{exc}}(\mathbf{r})$	excitation PSF
$h_{\text{det}}(\mathbf{r})$	detection PSF
$h_{\text{STED}}(\mathbf{r})$	STED-PSF
I_{STED}	STED intensity
I_{sat}	saturation intensity
k_{isc}	inter-system crossing rate
k_{T}	triplet rate
D	diffusion constant
q	molecular brightness

Bibliography

[1] T.M. Porter. *The Rise of Statistical Thinking*. Princeton University Press, Princeton, 1986. ix

[2] S.M. Stigler. *Statistics on the Table : The History of Statistical Concepts and Methods*. Harvard University Press, Cambridge, 2002. ix

[3] Ernst Abbe. Beiträge zur Theorie des Mikroskops und der mikroskopischen Wahrnehmung. *Archiv für Mikroskopische Anatomie*, 9:413–468, 1873. ix, x

[4] S.W. Hell. Far-field optical nanoscopy. *Science*, 316:1153–1158, 2007. ix, xi, xii, 17

[5] S.W. Hell and J. Wichmann. Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy. *Optics Letters*, 19:780ff, 1994. ix, x, 85

[6] M. Hofmann, C. Eggeling, S. Jakobs and S.W. Hell. Breaking the diffraction barrier in fluorescence microscopy at low light intensities by using reversibly photoswitchable proteins. *Proc. Natl. Acad. Sci. USA*, 102:17565–17569, 2005. ix, xi

[7] A. Egner, C. Geisler, C. von Middendorff, H. Bock, D. Wenzel, R. Medda, M. Andresen, A. Stiel, S. Jakobs, C. Eggeling, A. Schönle and S. W. Hell. Fluorescence nanoscopy in whole cells by asynchronous localization of photoswitching emitters. *Biophysical Journal*, 93:3285–3290, 2007. ix, xii, 3, 29, 37, 52, 71

[8] C. Geisler, A. Schönle, C. von Middendorff, H. Bock, C. Eggeling, A. Egner and S.W. Hell. Resolution of $\lambda/10$ in fluorescence microscopy using fast single molecule photo-switching. *Appl. Physics A*, 88:223–226, 2007. ix, xii, 37

[9] C. Geisler, A. Egner, C. v. Middendorff, A. Schönle and S.W. Hell. Three-dimensional sub-diffraction resolution in PALMIRA microscopy. *to be published*, 2008. ix

[10] J. Fölling, M. Bossi and S.W. Hell. Multi-color far-field fluorescence nanoscopy through single-molecule spectroscopy. *submitted*, 2008. ix, 60

[11] R. Rigler and E. Elson. *Fluorescence Correlation Spectroscopy : Theory and Applications*. Springer, Heidelberg, 2001. x, xiii, 99

[12] L. Kastrup, H. Blom, C. Eggeling and S.W. Hell. Fluorescence fluctuation spectroscopy in subdiffraction focal volumes. *Phys. Rev. Lett.*, 94:143903:1–4. x, xiv, 85

[13] C. Eggeling, C. Ringemann, R. Medda, B. Hein, G. Schwarzmann, K. Sandhoff, S. Polyakova, V. Belov, C. von Middendorff, A. Schönle and S.W. Hell. Nanoscale dynamics of single lipid molecules in living cells reveals cholesterol-assisted membrane complexes. *submitted*, 2008. x, xiv, 85, 89

[14] J. Widengren, Ü. Mets and R. Rigler. Fluorescence correlation spectroscopy of triplet states in solution: a theoretical and experimental study. *J. Phys. Chem.*, 99:13368–13379, 1995. x, xiv, 90, 92

[15] M.M. Woolfson. *X-Ray Cristallography*. Cambridge University Press, Cambridge, 1970. x

[16] M. Knoll and E. Ruska. Das Elektronenmikroskop. *Zeitschrift für Physik*, 78:318–339, 1932. x

[17] S.W. Hell and E.H.K. Stelzer. Properties of a 4pi-confocal fluorescence microscope. *J. Opt. Soc. Am. A*, 9:2159–2166, 1992. x, 34

[18] T.A. Klar and S.W. Hell. Subdiffraction resolution in far-field fluorescence microscopy. *Optics Letters*, 24:954f, 1999. x, 85

[19] S.W. Hell. Toward fluorescence nanoscopy. *Nature Biotechnology*, 21:1347–1355, 2003. x, 85

[20] T. Wilson, C.J.R. Sheppard. *Theory and Practice of Scanning Optical Microscopy*. Academic Press, London, 1984. x

[21] J. Keller, A. Schönle and S.W. Hell. Efficient fluorescence inhibition patterns for RESOLFT microscopy. *New J. Phys.*, 9:435:1–10. xi, 86, 87, 107

[22] S.W. Hell and A. Schönle. *Nanoscale Resolution in Far-Field Microscopy, Science of Microscopy*, volume 2, pages 790–835. Springer, Heidelberg, 2007. xi

[23] B. Harke, J. Keller, C.K. Ullal, V. Westphal, A. Schönle and S.W. Hell. Resolution scaling in STED microscopy. *Optics Express*, 16:6:4152–4154, 2008. xi, 17, 86

[24] G. Donnert, J. Keller, R. Medda, M. A. Andrei, S. O. Rizzoli, R. Lührmann, R. Jahn, C. Eggeling and S. W. Hell. Macromolecular-scale resolution in biological fluorescence microscopy. *Proc. Natl. Acad. Sci. USA*, 103:11440–11445, 2006. xi

[25] M. Dyba and S.W. Hell. Photostability of a fluorescent marker under pulsed depletion through stimulated emission. *Appl. Optics*, 42:5123–5129, 2003. xi

[26] S. Bretschneider, C. Eggeling and S.W. Hell. Breaking the diffraction barrier in fluorescence microscopy by optical shelving. *Phys. Rev. Lett.*, 98:218103, 2007. xi

[27] W. Heisenberg. *Die physikalischen Prinzipien der Quantentheorie*, page 17. Hirzel, Leipzig, 1930. xi

[28] X. Qu, D. Wu, L. Mets and N.F. Scherer. Nanometer-localized multiple single-molecule fluorescence microscopy. *Proc. Natl. Acad. Sci. USA*, 101:31:11298–11303, 2004. xi

[29] E. Betzig, G.H. Patterson, R. Sougrat, O.W. Lindwasser, S. Olenych, J.S. Bonifacino, M.W. Davidson, J.Lippinscott-Schwartz and H.F. Hess. Imaging introcellular fluorescent proteins at nanometer resolution. *Science*, 313:1642–1645, 2006. xi, 3, 16

[30] S.T. Hess, Thanu P.K. Girirajan and M.D. Mason. Ultra-high resolution imaging by fluorescent photoactivation localization microscopy (FPALM). *Biophysical Journal*, 91:4258–4272, 2006. xii, 3

[31] M. Rust, M. Bates and X. Zhuang. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nature Methods, Advanced Online Publication*, pages 1–3, 2006. xii, 3

[32] M. Bates, B. Huang, G.T. Dempsey and X. Zhuang. Multicolor super-resolution imaging with photo-switchable fluorescent probes. *Science*, 317:1749–1752, 2007. xii

[33] H. Shroff, C.G. Galbraith, J.A. Galbraith, H. White, J. Gillette, S. Olenych, M.D. Davidson and E. Betzig. Dual-color superresolution imaging of genetically expressed probes within individual adhesion complexes. *Proc. Natl. Acad. Sci. USA*, 104:20308–20313, 2007. xii

[34] B. Huang, W. Wang, M. Bates and X. Zhuang. Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy. *Science*, 319:810–813, 2008. xii, 32

[35] C. Blatter. *Wavelets: A Primer*. Peters, Natick, Mass., 2002. xii

[36] B.J. Berne and R. Pecora. *Dynamic Light Scattering*. Wiley, New York, 1976. xiii

[37] D. Magde, E. Elson and W.W. Webb. Thermodynamic fluctuations in a reacting system - measurement by fluorescence correlation spectroscopy. *Phys. Rev. Lett.*, 29:11:705–708, 1972. xiii

[38] O. Krichevsky and G. Bonnet. Fluorescence correlation spectroscopy : the technique and its applications. *Rep. Prog. Phys.*, 65:251–297, 2002. xiii, 93

[39] D.P. Landauer and K. Binder. *A Guide to Monte-Carlo Simulations in Statistical Physics*. Cambridge University Press, Cambridge, 2000. xiii

[40] W. Krauth. *Statistical Mechanics: Algorithms and Computations*. Oxford University Press, Oxford, 2006. xiii

[41] D.H. Bunfield and L.M. Davis. Monte-Carlo simulation of a single-molecule detection experiment. *Appl. Opt.*, 37:2:2315–2326, 1998. xiii

[42] M.H. de la Torre, R. Forni and G. Chirico. Brownian dynamics simulations of fluorescence fluctuation spectroscopy. *Eur. Biophys. J.*, 30:129–139, 2001. xiii

[43] K. Schätzel, M. Drewel and S. Stimac. Photon correlation measurements at large lag times: improving statistical accuracy. *J. Mod. Opt.*, 35(4):711718, 1987. xiii, 79

[44] K. Schätzel. Noise on photon correlation data: Autocorrelation functions. *Quantum Opt.*, 2:287–305, 1990. xiii, 79

[45] S.T. Hess and W.W. Webb. Focal volume optics and experimental artifacts in confocal fluorescence correlation spectroscopy. *Biophysical Journal*, 83:2300–2317, 2002. xiv

[46] J. Widengren and R. Rigler. Mechanisms of photobleaching investigated by fluorescence correlation spectroscopy. *Bioimaging*, 4:149–157, 1996. xiv

[47] D. Koppel. Statistical accuracy in fluorescence correlation spectroscopy. *Phys. Rev. A*, 10:6:1938–1945, 1974. xiv, 97

[48] H. Qian. On the statistics of fluorescence correlation spectroscopy. *Biophys. Chem.*, 38:49–57, 1990. xiv, 97

[49] D. Maier, M. Marth, J. Honerkamp and J. Weese. Influence of correlated errors on the erstimation of the relaxation time spectrum in dynamic light scattering. *Appl. Opt.*, 38:21:4671–4680, 1999. xiv, 97

[50] A. Lomakin. Fitting the correlation function. *Appl. Opt.*, 40:24:4079–4086, 2001. xiv, 97

[51] A.E. Smart, R.V. Edwards and W.V. Meyer. Quantitative simulation of errors in correlation analysis. *Appl. Opt.*, 40:24:4064–4078, 2001. xiv, 97

[52] W. Feller. *An Introduction to Probability Theory and Its Applications*. Wiley, New York, 1968. xv, 6, 19, 49, 101

[53] H.O. Goergii. *Einführung in die Wahrscheinlichkeitstheorie und Statistik*. de Gruyter, Berlin, 2004. xv, 13, 26

[54] J. Hogbom. Aperture synthesis with a non-regular distribution of interferometer baselines. *Astrophys. J. Suppl. Ser.*, 15:417–426, 1974. 8, 9

[55] R.E. Thompson, D.L. Larson and W.W. Webb. Precise nanometer localization analysis for individual fluorescent probes. *Biophysical Journal*, 82:2775–2783, 2002. 8, 13, 18, 24

[56] B. Richards and E. Wolf. Electromagnetic diffraction in optical systems: Structure of the image field in an aplanatic system. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 253(1274):358–379, 1959. 12, 31, 107

[57] M. Born and E. Wolf. *Principles of Optics*. Cambridge University Press, Cambridge, 2002. 12, 13, 31, 107

[58] B.R. Frieden. *Probability, Statistical Optics, and Data Testing : a Problem Solving Approach*. Springer, Berlin, 2001. 12

[59] M. Gu. *Advanced Optical Imaging Theory*. Springer, Heidelberg, 2000. 13, 86

[60] D.R. Stirzaker and G.R Grimmett. *Probability and Random Processes*. Oxford University Press, Oxford, 2001. 13

[61] D.R. Stirzaker. *Elementary Probability*. Cambridge University Press, Cambridge, 2005. 13

[62] S. Kullback. *Information Theory and Statistics*. Wiley, New York, 1959. 14

[63] A. Erdlyi, W. Magnus, F. Oberhettinger and F.G. Tricomi. *Higher Transcendental Functions*, volume 1, pages 27–31. Krieger, New York, 1981. 15

[64] A. Laurincikas and R. Garunkstis. *The Lerch Zeta-Function*, pages 17–18. Kluwer, Dordrecht, 2002. 15

[65] A. Papoulis. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, New York, 1984. 22

[66] H. Jeffreys and B.S. Jeffreys. *Methods of Mathematical Physics*, pages 463–464. Cambridge University Press, Cambridge, 1988. 22

[67] J. Nowakowski and M. Elbaum. Fundamental limits in estimating light pattern position. *J. Opt. Soc. Am.*, 73:1744–1758, 1983. 24

[68] K. Winick. Cramer-Rao lower bounds on the performance of charge-coupled-device optical position estimators. *J. Opt. Soc. Am. A*, 3:1809–1815, 1986. 24

[69] M.K. Cheezum, W.F. Walker and W.H. Guildford. Quantitative comparison of algorithms for tracking single fluorescent particles. *Biophysical Journal*, 81:2378–2388, 2001. 24, 43

[70] R.J. Ober, S. Ram and E.S. Ward. Localization accuracy in single-molecule microscopy. *Biophysical Journal*, 86:1185–1200, 2004. 24

[71] A. d. Dekker, S. v. Art, A. v.d. Bos and D. v. Dyck. Maximum likelihood estiation of structure parameters from high resolution electron microscopy images : A theoretical framework. *Ultramicroscopy*, 104:83–106, 2005. 24, 25

[72] A. d. Dekker, S. v. Art, A. v.d. Bos and D. v. Dyck. Maximum likelihood estiation of structure parameters from high resolution electron microscopy images : A practical example. *Ultramicroscopy*, 104:107–125, 2005. 24

[73] M. Speidel, A. Jonas and E.-L. Florin. Three-dimensional tracking of fluorescent nanoparticles with subnanometer precision by use of off-focus imaging. *Optics Letters*, 28:69–71, 2003. 24

[74] E. Toprak, H. Balci, B.Blehm and P. Selvin. Three-dimensional particle tracking via bifocal imaging. *Nanoletters*, 7:2043–2045, 2007. 24

[75] H.P. Kao and A.S. Verkman. Tracking of single fluorescent particles in three dimensions: Use of cylindrical optics to encode particle position. *Biophysical Journal*, 67:1291–1300, 1994. 24

[76] F. Aguet, D. Van De Ville and M. Unser. A maximum-likelihood formalism for subresolution axial localization of fluorescent nanoparticles. *Optics Express*, 13:10503–10522, 2005. 24

[77] R. A. Fisher. Theory of statistical estimation. *Proc. Cambridge Philos. Soc.*, 22:700–725, 1925. 26

[78] D.R. Cox and D.V. Hinkley. *Theoretical Statistics*. Chapman and Hall, London, 1974. 26, 42

[79] B.S. Everitt. *Cambridge Dictionary of Statistics*. Cambridge University Press, Cambridge, 2006. 26

[80] T.M. Cover and J.A. Thomas. *Elements of Information Theory*. Wiley, New York, 2005. 26, 42

[81] H. Cramer. *Mathematical Methods of Statistics*. Princeton University Press, Princeton, 1946. 26

[82] C. Rao. Information and the accuracy attainable in the estimation of statistical parameters. *Bull. Calcutta Math. Soc.*, 37:81–89, 1945. 26

[83] J. Fölling, V. Belov, D. Riedel, A. Schönle, A. Egner, C. Eggeling, M. Bossi and S. W. Hell. Fluorescence nanoscopy with optical sectioning by two-photon induced molecular switching using continuous-wave lasers. *ChemPhysChem*, 9:321–326, 2008. 28, 56

[84] M. Nagorni and S.W. Hell. Coherent use of oppsing lenses for axial resolution increase in fluorescence microscopy. i. comparative study of concepts. *J. Opt. Soc. A*, 18:1:36–48, 2001. 34

[85] M. Dyba and S.W. Hell. Focal spots of size $\lambda/23$ open up far-field flourescence microscopy at 33nm axial resolution. *Phys. Rev. Lett.*, 88:163910–1ff, 2002. 34

[86] L. Wasserman. *A Concise Course in Statistical Inference*. Springer, Heidelberg, 2004. 43, 97, 99, 101, 102

[87] J.A. Nelder and R. Mead. A simplex method for function minimization. *Computer Journal*, 7:308313, 1965. 44

[88] E. Zeidler and B. Hunt. *Oxford User's Guide to Mathematics*. Oxford University Press, Oxford, 2004. 49

[89] A.C. Stiel, S. Trowitzsch, G. Weber, M. Andresen, C. Eggeling, S. W. Hell, S. Jakobs and M. C. Wahl. 1.8 Å bright-state structure of the reversibly switchable fluorescent protein dronpa guides the generation of fast switching variants. *Biochemical Journal*, 402:35–42, 2007. 52

[90] R. Ando, H. Mizuno and A. Miyawaki. Regulated fast nucleocytoplasmic shuttling observed by reversible protein highlighting. *Science*, 306:13701373, 2004. 52, 67

[91] S. Ram, E.S. Ward and R.J. Ober. A stochastic analysis of performance limits for optical microscopes. *Multidim. Syst. Sig. Process.*, 17:27–57, 2006. 55

[92] S. Ram, E.S. Ward and R.J. Ober. Beyond rayleigh's criterion: a resolution measure with application to single-molecule microscopy. *Proc. Natl. Acad. Sci. USA*, 103:12:4457–4462, 2006. 55

[93] A. Schönle and S. W. Hell. Fluorescence nanoscopy goes multicolor. *Nature Biotechnology*, 25:11:1234–1235, 2007. 60

[94] E. Merzbacher. *Quantum Mechanics*. Wiley, New York, 1998. 67

[95] E. Blahut. *Theory of Remote Image Processing*, pages 286–293. Cambridge University Press, Cambridge. 68

[96] S.G. Lipson, H.S. Lipson and D.S. Tannhauser. *Optical Physics*. Cambridge University Press, Cambridge, 1995. 70

[97] C.W. Gardiner. *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences*. Springer, Berlin, 2004. 75

[98] A. Schönle. *Point-Spread-Function Engineering in Fluorescence Spectroscopy*. Dissertation, Heidelberg, 2003. 78

[99] J.B. Pawley, ed. *Handbook of Biological Confocal Microscopy*. Springer, New York, 2006. 85

[100] Klar, T. A., E. Engel and S. W. Hell. Breaking Abbe's diffraction resolution limit in fluorescence microscopy with stimulated emission depletion beams of various shapes. *Phys. Rev. E*, 64: 066613:1–9, 2001. 86

[101] S.W. Hell, M. Dyba and S. Jakobs. Concepts for nanoscale resolution in fluorescence microscopy. *Curr. Op. in Neurobio.*, 14:5:599–609, 2004. 87

[102] C. Ringemann, A. Schönle, A. Giske, C. v. Middendorff, S.W. Hell and C. Eggeling. Enhancing fluorescence brightness: Effect of reverse intersystem crossing studied by fluorescence fluctuation spectroscopy. *ChemPhysChem*, 9:612–624, 2008. 90, 95

[103] I. Gregor, D. Patra and J. Enderlein. Optical saturation in fluorescence correlation spectroscopy under cw- and pulsed excitation. *ChemPhysChem*, 5:1–7, 2004. 91

[104] W.H. Press, S.A. Teukolsky, W.T. Vetterling and B.P. Flannery. *Numerical Recipes in C*. Cambridge University Press, Cambridge, 2007. 96

[105] E. Jakeman, E.R. Pike and S. Swain. Statistical accuracy in the digital autocorrelation of photon counting fluctuations. *J. Phys. A: Gen. Phys.*, 4:517–534, 1971. 97

[106] P. Kask, R. Günther and P. Axhausen. Statistical accuracy in fluorescence fluctuation experiments. *Eur. Biophys. J.*, 25:163–169, 1997. 97

[107] K. Starchev, J. Ricka and J. Buffle. Noise on fluorescence correlation spectroscopy. *J. Coll. a. Interf. Sci.*, 233:50–55, 2001. 97

[108] R. Gnandesikan. *Methods for Statistical Analysis of Multivariate Observations*. Wiley, New York, 1977. 101

[109] E. Wolf. Electromagnetic diffraction in optical systems: An integral representation of the image field. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 253(1274):349–357, 1959. 107

⁎

List of publications related to the thesis.

1. A. Egner, C. Geisler, C. v. Middendorff, H. Bock, D. Wenzel, R. Medda, M. Andresen, A. Stiel, S. Jakobs, C. Eggeling, A. Schönle and S. W. Hell. Fluorescence nanoscopy in whole cells by asynchronous localization of photoswitching emitters. *Biophysical Journal* 93:3285-3290, 2007
2. C. Geisler, A Schönle, C. v. Middendorff, H. Bock, C. Eggeling, A. Egner and S.W. Hell. Resolution of $\lambda/10$ in fluorescence microscopy using fast single molecule photo-switching. *Appl. Phys. A* 88: 223226, 2007
3. H. Bock, C. Geisler, C. Wurm, C. v. Middendorff, S. Jakobs, A. Schönle, A. Egner, S.W. Hell and C. Eggeling. Two-color far-field fluorescence nanoscopy based on photoswitchable emitters. *Appl. Phys. B* 88, 161165, 2007
4. C. Ringemann, A. Schönle, A. Giske, C. v. Middendorff, S. W. Hell and C. Eggeling. Enhancing fluorescence brightness: effect of reverse intersystem crossing studied by fluorescence fluctuation spectroscopy. *ChemPhysChem* 9:612 624, 2008
5. C. Geisler, A. Egner, C. v. Middendorff, A.Schönle and S.W. Hell. Three-dimensional sub-diffraction resolution in PALMIRA microscopy. *to be published*, 2008
6. C. v. Middendorff, A. Schönle and S.W. Hell. Translation invariant nanoscale resolution in PALMIRA-4pi microscopy. *to be published*, 2008
7. C. Eggeling, C. Ringemann, R. Medda, B. Hein, G. Schwarzmann, K. Sandhoff, S. Polyakova, V. Belov, C. v. Middendorff, A. Schönle and S.W. Hell. Detecting the nanoscale dynamics of membrane lipids in a living cell reveals cholesterol-assisted complexes. *submitted*, 2008

Acknowledgements

Prof. Stefan W. Hell, Max-Planck-Institute for Biophysical Chemistry, Göttingen, is warmly thanked for constant support and for making me acquainted with this timely topic.
Prof. Cremer, University of Heidelberg, kindly agreed to be the second referee of this thesis.

During this work I have profited from conversations with many people.
I thank Dr. Andreas Schönle and Dr. Jan Keller for many stimulating discussions and for bringing me on the right track in a number of cases.
Dr. Christian Eggeling, Dr. Alexander Egner, Jonas Fölling, Claudia Geisler, Christian Ringemann and Ilaria Testa are warmly acknowledged for having supplied feedback from the lab reality.
Hannes Bock, Dr. Christian Eggeling, Jonas Fölling, Claudia Geisler, Dr. Benjamin Harke, Dr. Lars Kastrup, Dr. Jan Keller, Marcel Lauterbach, Lars Meyer, Brian Rankin, Christian Ringemann, Steffen Sahl, Dr. Andreas Schönle, Dr. Chaitanyah Ullal and Dr. Giuseppe Vicidomini have kindly proofread parts of the text.

The outstanding services of the Otto-Hahn library, represented by Renate Hägele, Reinhard Harbaum and Bernhard Reuse, and the Staats- und Unversitätsbibliothek, Göttingen, have been of great usefulness.
The whole department at the Max-Planck-Institute is thanked for reliably offering an amicable synergy of competitiveness and collegiality.

Göttingen, August 13, 2008