

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

presented by
Diplom-Physiker Jan Keller
born in Leipzig

Oral examination: October 18th, 2006

Optimal de-excitation patterns for RESOLFT-Microscopy

Referees: Prof. Dr. Stefan W. Hell
Prof. Dr. Josef Bille

Abstract

RESOLFT microscopy has been the first method that is capable of non-invasively resolving three dimensional structures with real subdiffraction resolution using visible light. It exploits the strong nonlinear saturation of a reversible optical transition. A focal intensity pattern is essential that de-excites or de-activates dyes outside the remaining ultrasharp effective focal spot. For a given amount of available power, the steepest applied de-excitation pattern will yield the highest resolution. In this thesis, for the first time, a comprehensive search, optimization and characterization of de-excitation patterns is performed. The microscope's pupil function is decomposed into orthonormal polynomials which allows the restriction of the space of pupil functions so that boundary conditions are fulfilled. The chosen global optimization algorithm converges reasonably well to pupil functions that can be idealized further to simple shapes. Optimal pupil functions are found according to assumptions made about the practical limitations. The optimization identified a novel, superior de-excitation pattern for circularly polarized light. Its experimental application has led to a hitherto unrivaled lateral resolution of down to 20 nm in biological systems. It is shown that an efficient resolution increase in all spatial directions is only possible by incoherent combinations of de-excitation beams. The optimal choice for current experimental conditions is identified. Finally, a new concept for fast acquisition of high resolution images is developed that is based on the simultaneous creation of compact arrays of sub-diffraction sized fluorescence spots in the sample. An optical setup that can generate the corresponding complex pupil functions is detailed.

Zusammenfassung

RESOLFT Mikroskopie war die erste Methode, welche nichtinvasiv dreidimensionale Strukturen mit Hilfe sichtbaren Lichts deutlich unterhalb des Beugungslimits auflösen konnte. Dabei wird die starke nichtlineare Sättigung von reversiblen optischen Übergängen ausgenutzt. Eine fokale Intensitätsverteilung wird benötigt, um Farbstoffe außerhalb eines verbleibenden, sehr scharfen effektiven Fokus abzuregen oder zu deaktivieren. Bei gegebener Stärke der vorhandenen Leistung wird mit der steilsten Intensitätsverteilungen die höchste Auflösung erzielt. In dieser Arbeit wird zum ersten Mal eine ausführliche Suche, Optimierung und Charakterisierung von Intensitätsverteilungen durchgeführt. Die Pupillenfunktion des Mikroskops wird dabei in orthonormale Polynome zerlegt. Dadurch kann der Raum all dieser Funktionen den Randbedingungen entsprechend eingeschränkt werden. Der ausgewählte globale Optimierungsalgorithmus konvergiert so weit, dass die erhaltenen Pupillenfunktionen zu einfachen Formen idealisiert werden können. Für verschiedene Annahmen über limitierende Faktoren wurden optimale Pupillenfunktionen gefunden. Die Optimierung ergab eine neue, verbesserte Intensitätsverteilung, die zirkular polarisiertes Licht benutzt. Ihre experimentelle Realisierung hat zu einer bis dahin unerreichten Auflösung von bis zu 20 nm in biologischen Systemen geführt. Es wird gezeigt, dass eine effiziente Erhöhung der Auflösung in allen Raumrichtungen nur durch inkohärente Kombinationen von Intensitätsverteilungen möglich ist. Die optimale Wahl unter derzeitigen experimentellen Bedingungen wurde gefunden. Desweiteren wurde ein neues Konzept zur schnellen Akquisition hochaufgelöster Bilder entwickelt, welches auf der gleichzeitigen Erzeugung kompakter Anordnungen von fluoreszierenden Foki unterhalb des Beugungslimits beruht. Ein optischer Aufbau, welcher die dazugehörigen komplexen Pupillenfunktionen erzeugen kann, wird detailliert beschrieben.

Abbreviations

\otimes	convolution of two functions
*	cross-correlation of two functions
1D	one-dimensional
2D	two-dimensional
3D	three-dimensional
4Pi-microscopy	microscopy using two opposing lenses in a coherent way
cw	continuous wave
Exc	excitation
Eff	effective
Em	emission
d_{ER}	distance from focus that corresponds to the expected resolution
Det	detection
FoM	figure-of-merit
FWHM	full width at half maximum
GSD	ground state depletion
MMM	multifocal multi-photon microscopy
NA	numerical aperture of a lens ($NA = n \sin \alpha$)
OTF	optical transfer function, Fourier transform of the PSF
(PAL-)SLM	(parallel-aligned nematic liquid crystal) spatial light modulator
PSF	point spread function
I-PSF	intensity point spread function
A-PSF	amplitude point spread function
RESOLFT	reversible saturable optical (fluorescence) transitions
SNOM	scanning near-field optical microscopy
SNR	signal to noise ratio
STED	stimulated emission depletion
TIRF	total internal reflection fluorescence
OPO	optical parametric oscillator
UV	ultraviolet (light)

standard conditions aplanatic lens,
 $\lambda_{Exc} = \lambda_{Det} = 500\text{nm}$, $\lambda_{RESOLFT} = 700\text{nm}$
 $n_W = 1.333$, $\sin(\alpha) = 0.9$

Contents

1	Introduction	1
2	Theoretical Foundations	4
2.1	Image Formation	4
2.2	Resolution of a Light Scanning Microscope	5
2.3	Introduction to RESOLFT-type Microscopes	8
2.4	Vectorial Image Formation in a RESOLFT Microscope	13
2.5	Engineering Focal Intensity Distributions	16
2.5.1	Calculation of the Focal Field	16
2.5.2	The Optimization Space and its Basis	19
2.5.3	Constructing Subspaces Providing Strict Intensity Zeros	22
3	Optimal Single Focus De-Excitation Patterns	25
3.1	Optimization Algorithm	26
3.2	Optimization Results	28
3.2.1	Limited Amplitude (A)	28
3.2.2	Limited Power (B)	33
3.2.3	Limited focal intensity (C)	38
3.3	Ideal 3D De-Excitation Patterns	39
3.4	Highest Possible Resolution Increase	41
3.5	Experiments	44
3.6	Z-Oriented Molecules	46
4	Multifocal RESOLFT Microscopy	50
4.1	Fast Scanning	50
4.2	Multifocal Arrangements	52
4.3	Optimization Algorithm	53
4.4	Results	54
4.4.1	Expected Resolution	59
4.5	Generation of General Complex Apodizations	61
5	Conclusion and outlook	64
	Bibliography	66

A	Appendix	74
A.1	Zernike Polynomials	74
A.1.1	Completeness and Orthogonality	74
A.1.2	Efficient Calculation of Zernike Polynomials	74
A.2	Singular Value Decomposition	75
A.3	Global Optimization Method	76
A.4	Coherent 3D De-Excitation Patterns	78
A.5	Additional Optimization Results	79
A.6	De-Excitation Patterns for 4Pi-Microscopes	84

“I prefer knowing the cause of a single thing to being king of Persia.”

Democritos

1 Introduction

The application of optics in science is a history of groundbreaking insights. Optics is applied across many different disciplines. E.g., some years ago the Deep Field project with the Hubble telescope [1] provided an unprecedented sharp view of a part of the sky with at least 1500 galaxies at various stages of evolution. At the other end of the length scale, microscopy advanced as well. Already in 1839, optics revealed parts of the microcosmos when Schwann and Schleiden developed a cell theory by using the microscope [2]. After years of development, the utilization of optics in modern science is represented by works such as the disclosure of the embryonic development of zebrafish on a microscopic scale [3] - one of the works awarded with the Nobel prize. Also, the distribution of various proteins could be resolved at the subcellular scale [4]. Enhancing the resolution of light microscopes further would be a big step toward a more fundamental understanding of nature which is the ultimate goal of all work presented in this dissertation.

Far field light microscopy non-invasively delivers three dimensional images of living cells. Highly specific fluorescent markers allow functional imaging because different dyes can be attached to different structures in the cell, specific molecules or even certain sites of macromolecules [5]. The development of more versatile, functionalized markers continues [6]. However, it is known from Abbe [7] that the wave nature of light restricts the obtainable resolution. Spatial frequencies above a certain cut-off are not transmitted by an objective lens - placing a lower bound on the obtainable spot size of focused light. This is known as the diffraction barrier.

Although Abbe's law is universal, developments have arisen in the last 30 years that were able to push the resolution of light microscopes beyond that of conventional far field microscopes. The advent of confocal laser scanning microscopes in 1978 [8] marks an extension of the transmitted frequency band and the onset of true three dimensional fluorescence imaging. Nevertheless, the smallest achievable spot sizes were still confined to $\lambda/4$ by diffraction, a limit unsurpassable by any conventional far field microscope.

Abbe's law states that the resolution scales with the applied wavelength and inversely with the numerical aperture of the lens. One path to higher resolution would therefore be the utilization of smaller wavelengths. For confocal microscopy, the shortest compatible wavelength for live cell imaging is around 400 nm with near UV-light [9]. However, X-Ray and Electron microscopy are using smaller wavelengths. It has been shown that soft X-ray microscopes in the 1-5 nm wavelength region can visualize structures of 30 nm size in biological samples [10, 11, 12]. It is possible to image frozen, hydrated samples, but this method relies on the availability of strong radiation sources and on appropriate contrast mechanisms restricting its applicability. Electron microscopy allows the resolution of single macromolecules because the typical de Broglie wavelength of an electron is about 10^{-3} nm,

but electron microscope samples must undergo an extensive fixation process before imaging that can lead to changes in the cellular substructure. Furthermore, it is restricted to the recording of snapshots in the time domain due to fixation and can only image thin slices due to the limited penetration depth of electrons in tissue.

The other way, the maximization of the aperture, led to the introduction of 4Pi-microscopy [13] which is based on coherent excitation and/or detection through two opposing objective lenses. It delivers a 4-7 fold increase of axial resolution over confocal microscopy [14].

Under favorable conditions, the usage of structured illumination [15] can reduce the image acquisition time but it does not lead to an increase in spatial resolution. Other parallelization techniques include the use of microlens arrays [16]. These methods mainly render high-resolution microscopy more suitable for live cell imaging.

Several concepts allowing resolution beyond the diffraction barrier were introduced. A technique abandoning the far field altogether is scanning near field optical microscopy [17, 18] where a small spot is created near a sub-wavelength sized aperture. In the near field, the area of illumination is not defined by the wavelength but only by the aperture opening. But since these microscopes act as 'optical stethoscopes' they are bound to imaging surfaces and they are prone to artefacts.

It was soon realized that nonlinear light-matter interactions are a very convenient way to fundamentally break the diffraction barrier. The strong nonlinear responses of marker molecules to the distribution of light near the microscope's focus is exploited to add new frequencies to the transmitted band. Multi-photon and especially two-photon excitation [19] is widely used and it turned out to be beneficial when imaging deep in tissue or to realize parallelized spot-scanning microscopy [20]. However, the necessary energy subdivision into multiple photons used for excitation prevents any resolution increase [21]. The use of second-harmonic generation microscopy [22], third-harmonic generation microscopy [23], and coherent anti-Stokes Raman scattering microscopy [24, 25] do not achieve a pronounced increase in resolution for similar reasons. The use of entangled-photons [26] was also suggested but has not been implemented yet.

So far, the only method that utilizes fluorescence markers and visible light and is capable of non-invasively resolving three dimensional structures with real subdiffraction resolution is RESOLFT microscopy. It is in fact a whole family of approaches. The common idea is to establish the nonlinear dependence of the molecular response by using saturation of reversible optical (fluorescence) transitions [27, 28, 29, 30]. Then, although the focal intensity distribution is still diffraction limited, the effective spot of molecules in a specific spectroscopic state can be squeezed down by increasing the applied power and driving the transition to higher saturation levels. Several approaches have been published [31, 32, 33, 34, 35]: among them are ground state depletion (GSD), stimulated emission depletion (STED), repeated excitation and the usage of switchable proteins. So far, STED microscopy increased the resolution to either $\lambda/25$ [36] laterally or $\lambda/23$ [37] along the optic axis in a combination with 4Pi-microscopy.

Commonly, an excited or photo-activated state is imaged. Any kind of saturable, optical transition is suitable in RESOLFT microscopy, e.g. saturating the excitation can lead to increased resolution [80]. However, the application of saturation in excitation or activation

results in images which are prone to noise. Therefore, this work considers only the types of RESOLFT microscopy that rely on the application of de-excitation or -activation light. For these types, the experimental realization of a reshaped focal intensity distribution featuring an isolated intensity zero is essential for a successful implementation. A de-excitation or -activation distribution with a strict intensity zero is used to create increasingly small spots of molecules retaining the ability to emit fluorescence photons when the power of the de-excitation pattern is raised. The applicable power is practically limited, e.g. by the onset of photo-destructive processes, thus severely restricting the potential prominent increase in resolution. It is therefore a deciding factor to find the best (narrowest) de-excitation pattern as it will result in the highest resolution for a given limited amount of applicable power.

Although several different de-excitation patterns were already used in practical applications, a systematic survey was never conducted. The goal of this work is to find optimal de-excitation patterns for RESOLFT microscopy under common conditions. Their application helps to overcome the diffraction barrier by RESOLFT microscopy as best as possible.

For this purpose the following problems are addressed:

1) A comprehensive framework for PSF engineering is laid out for RESOLFT microscopes. By decomposing the pupil function into polynomials and applying algebraic methods, the space of possible solutions is constrained according to applicable boundary conditions. An optimization algorithm is designed that delivers the pupil function resulting in the de-excitation distribution with the most useful shape.

2) Different optimization results are found according to assumptions made about the experimental conditions and sample parameters. Patterns with one intensity node are optimized and the corresponding pupil functions are detailed. A strategy for an efficient three dimensional resolution increase is presented.

3) The prospects of parallelizing RESOLFT microscopy with de-excitation patterns featuring multiple isolated intensity zeros are explored. An arrangement of intensity nodes is proposed that allows the generation of efficient de-excitation patterns around them. The energy-efficiency and the limits of this parallelization technique are then analyzed.

2 Theoretical Foundations

2.1 Image Formation

In fluorescence microscopy, the image is usually a linear mapping of the true object deteriorated by noise. This means that the whole image can be described as a sum of images of object parts. In practice, one obtains a blurred image of the object lacking certain features of the original object distribution. High spatial frequency information is lost irrevocably during the imaging process due to its bandwidth-limitation. In spot-scanning fluorescence microscopy, two processes determine the image formation: excitation and detection. The diffraction limit prevents the excitation of only a single point-like spot in the sample. Instead, for plane wave illumination through a lens with a high numerical aperture, the intensity distribution near the focus features a main peak whose width is in the order of $\lambda/2$. This distribution is termed intensity point spread function (I-PSF) of the excitation and is denoted by $h_{\text{Exc}}(\mathbf{r})$. The area of the main peak in the focal plane is called Airy disk. In the dipole approximation, the excitation rate of a fluorophore is proportional to the absolute square of the electric field. If the molecular response does not incorporate saturation effects the resulting excitation will be proportional to its rate. The achievable excitation level is then proportional to $h_{\text{Exc}}(\mathbf{r})$. Resolving optics can also be applied during detection. The emission of the individual fluorophores does not depend on the phase information of the excitation light. Therefore, the emission light of different fluorophores adds up incoherently on the detector. Commonly, the same lens is used for both, excitation and collection of the emitted fluorescence, which is then uncoupled from the excitation path and projected onto a photo-detector. Since the intensity of the fluorescence emission is measured, the fluorescence collection process gives

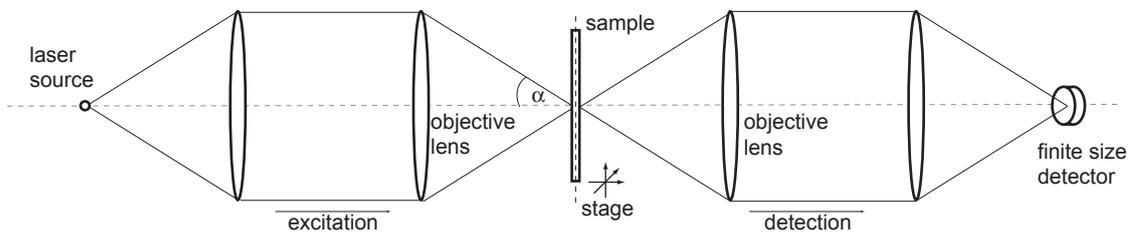


Figure 2.1: Scheme of a confocal laser scanning microscope system as proposed by T. Wilson [8]. A point-like laser source is imaged by a lens with high-numerical aperture into the sample which is consequently imaged by a lens with high-numerical aperture on a detector of finite size. In practical implementations, only one objective lens is used to excite and detect the probe simultaneously. Excitation and fluorescence light are then separated by appropriate filters.

rise to another I-PSF denoted by $h_{\text{Det}}(\mathbf{r})$ that is proportional to the probability of a photon emitted at \mathbf{r} to be detected by a point-like detector. When using the same lens $h_{\text{Exc}}(\mathbf{r})$ and $h_{\text{Det}}(\mathbf{r})$ differ only by a scaling factor that corresponds to the difference in the wavelengths of excitation and emission (Stokes-shift). The emission is usually detected within a spectral window and the effective integration area $D(\mathbf{r})$ on the detector is often restricted by a pinhole approximately the size of the projected Airy disk.

The essential composition of a laser scanning confocal microscope is shown in figure 2.1. Since only part of the sample is illuminated at a time, some form of a scanning system is inevitable. In this process, the sample is mounted on a mobile stage or the light beam's wavefront can be manipulated so that the focus slides over the sample, or both. The probability distribution that a signal from a dye molecule at position \mathbf{r} is detected (if the scanning position is at \mathbf{r}_0) is called the effective PSF $h_{\text{Eff}}(\mathbf{r})$. It is given by the product of the probability that the molecule at \mathbf{r} is excited and the probability that its emission is detected somewhere within the detector area: $h_{\text{Eff}} = h_{\text{Exc}}\text{CEF}$. The collection efficiency function (CEF) describes how efficiently fluorescence photons from a molecule at position \mathbf{r} are collected by the detector and is given by: $h_{\text{Det}} \otimes D$. Under common conditions, the shape of the PSFs for excitation and detection is not altered by scanning. The effective PSF is then only dependent on the coordinate $\mathbf{r} - \mathbf{r}_0$ and is said to be space-invariant. The image $I(\mathbf{r}_0)$ is then given by:

$$I = h_{\text{Eff}} \otimes G, \quad (2.1)$$

where G is the object, i.e. the distribution of fluorophores in space. This convolution is nothing else but a 'smearing out' of the object's distribution with the effective PSF. The formation of I is a linear transformation of the object distribution G . But the image is usually also corrupted by noise. The photon shot noise is directly correlated with the signal strength since the measured number of photoelectrons is a random number with a Poisson probability distribution and the mean value I .

2.2 Resolution of a Light Scanning Microscope

Resolution is intuitively defined as the smallest object feature still visible in a microscope's image. Clearly, resolution must therefore depend on the object, the shape of the effective PSF and the noise level because all these parameters determine the image quality. Due to the inherent random nature of the image formation process, noise cannot be eliminated completely physically or by mathematical postprocessing and even the simplest of all characteristics, which is the overall number of fluorophores, cannot be estimated with certainty. In fact, an object's characteristics can only be estimated and resolution depends on the amount of error which is still regarded as tolerable. Therefore, in a more general approach, resolution is the amount of information about the true object distribution that is accessible through the image. This definition is commonly used in the context of image deconvolution [38].

Nevertheless, several criteria for two-point separability were proposed. Among the most famous of them are the Rayleigh- and Sparrow-criteria. They both give a value of the minimal distance between two individual points in the focal plane that can still be resolved. The

minimal resolvable distance is defined as half the diameter of an Airy disk by the Rayleigh criterion and the smallest distance with an intensity minimum between neighboring objects is described by the Sparrow criterion. The different criteria are realistic under different noise conditions and can be used as first estimates of the resolution in practical applications. Indeed, with certain a-priori knowledge, the minimal distance can even be infinitely small if the photon counts are increased and certain demands on pixelation and background noise are fulfilled (see for example [39]). In this case, one needs to know beforehand that there are exactly two objects of known shape in the focal plane. However, in the general case only marginal a-priori knowledge can be included in the analysis. The one most widely used is the positivity constraint prohibiting the prediction of negative concentrations of dye molecules. Despite the diversity of existing resolution criteria, all have in common that certain characteristics if not the whole shape of the effective PSF h_{EFF} play an important role in determining the achievable resolution. For confocal microscopes which have the highest resolution without relying on nonlinear effects, h_{EFF} is given mainly by the multiplication of I-PSFs, one of the excitation and one of the detection. Due to the wave nature of light it is impossible to reduce the width of peaks in microscope's PSFs significantly beyond $\lambda/2$. More exactly, the full-width-at-half-maximum (FWHM) of an I-PSF is quite precisely given by [40]:

$$\begin{aligned}\Delta x, \Delta y &= \frac{\lambda}{2n \sin \alpha}, \\ \Delta z &= \frac{\lambda}{2n \sin^2(\alpha/2)},\end{aligned}\tag{2.2}$$

with λ , n , and α denoting the wavelength, the refractive index, and the semi-aperture angle of the objective lens, respectively. Δx and Δy are the FWHMs in the focal plane while Δz is along the optic axis, the direction of light propagation. The FWHM already gives a good estimate of the microscope's resolution. High resolution is thereby equivalent to a small FWHM. Theoretically, the FWHM can be made smaller by employing shorter wavelengths and/or larger focusing angles but these strategies already faced their ultimate limits in the past: Modern immersion lenses provide a maximum focusing angle of $\alpha \sim 70^\circ$ and the shortest compatible wavelength for imaging live cells is around 400 nm with near UV-light [41, 9].

In a more rigid examination of resolution, all details of h_{EFF} should be included as well as possible noise levels. Some resolution measures that depend on h_{EFF} shall be introduced here. They will be used later in this work to estimate the possible resolution enhancement of RESOLFT microscopes. Commonly, the focal volume V_F is defined by:

$$V_F = \int_{\mathbb{R}^3} h_{\text{EFF}}(\mathbf{r}) dV,\tag{2.3}$$

where we assume that h_{EFF} is normalized to $\max_{\mathbb{R}^3}(h_{\text{EFF}}) = 1$. This measure is useful if h_{EFF} describes a single main peak whose width is made smaller. Smaller FWHMs correspond to a smaller focal volume and higher resolution. Often, low values of h_{EFF} are insignificant

and lost in shot-noise. Therefore, we define for sparse samples the focal volume \tilde{V}_F which describes the volume within which the main peak is confined and it is given by:

$$\tilde{V}_F^\epsilon = \int_{\mathbb{R}^3} \chi_F(r) dV \quad \text{with} \quad \chi_F(r) = \begin{cases} 1 & h_{\text{Eff}}(r) > \epsilon \\ 0 & \text{otherwise} \end{cases}. \quad (2.4)$$

However, these focal volumes are not sufficient to completely characterize resolution since different PSFs can have identical measures.

A convenient way to gain insight in information loss is to describe imaging in frequency space. The Fourier transform of the image I is given by:

$$\hat{I} = \hat{h}_{\text{Eff}} \cdot \hat{G} = o_{\text{Eff}} \cdot \hat{G}, \quad (2.5)$$

where the hat signifies a 3D Fourier transform and o_{Eff} is the optical transfer function (OTF). The OTF describes the strength with which the spatial frequencies are transferred from the object to the image. It is given by the Fourier transform of the PSF and has usually a limited region where its modulus is nonzero (a limited support). An important characteristic is, that spatial frequencies where $|o_{\text{Eff}}(k)|$ is zero are completely absent in \hat{I} . Object features whose corresponding spatial frequencies are outside the OTF support are lost in the imaging process. In imaging based on linear molecular responses, the extent of the OTF support is determined by the wavelengths of excitation and emission light and is also known as Abbe's diffraction limit. The highest possible transmitted frequency is $2(k_{\text{Exc}} + k_{\text{Em}})$ [14] where k_{Exc} and k_{Em} are the wave vectors of excitation and emission, respectively. Therefore, with visible light and water immersion even the best far field microscopes relying on linear effects have an ultimate resolution limit of about 100 nm in all directions.

Vectorial transfer functions for large angle focusing lenses were calculated in [42, 43] and an analysis of the shape of the confocal OTF including the effects of pinholes is given in [44]. The zero frequency modulus in the OTF marks its maximum because the PSF is a non-negative function. Toward the rim of the support, the OTF's modulus falls off smoothly. Without the presence of noise, it is completely valid to state that the extent of the OTF support is an exact measure of resolution. Frequencies inside the support are transferred and are present in the image whereas information corresponding to spatial frequencies outside the support is lost. Modulating the OTF without extending its support does not increase the resolution in the above sense. Nevertheless, many propositions were made to change the wavefront for excitation or detection in order to modulate the characteristic distribution of the PSF. The most prominent concept stems from Toraldo [45] but many other filters based on this concept were proposed [46, 47, 48, 49, 50, 51]. They are mostly corresponding to the weakening of small frequencies in the OTF. Therefore, these cannot contribute as before in the focusing process and the FWHM approaches the limit of a standing wave of $\lambda/4$. On the other hand, large sidelobes will be generated due to the missing small frequencies.

In the case of a considerable amount of noise, object features can only be determined with a certain error. Additionally to the spatial frequencies that were lost due to the finite extent of the OTF, even more spatial frequencies will be 'quasi' swamped by noise in the

image. So, noise is responsible for further information loss. The main source of error, photon noise, has a broad frequency spectrum. It can be very well described by an almost even background in the Fourier transform of the image. Correlations of the photon noise with the object's structure are present but do not strongly modulate the noise spectra in many cases. Therefore, mainly the frequencies that are transferred weakly, i.e. where the OTF features a small modulus, are compromised by relatively large noise components. To eliminate the detrimental effects of noise, usually frequencies that are transferred only weakly are suppressed before deconvolving the remaining image. A measure of the resolution under such circumstances could be the extent of the OTF where its modulus is greater than a value which is dependent on the noise level. The ϵ extent of the OTF is defined by:

$$S_F^\epsilon = \int_{C_\chi} \chi_o(k) d^3k \quad \text{with} \quad \chi_o(k) = \begin{cases} 1 & |o_{\text{EFF}}(k)| > \epsilon \\ 0 & \text{otherwise} \end{cases}, \quad (2.6)$$

with C_χ denoting the connected area of $\chi_o(k) = 1$ in \mathbb{R}^3 . In the general case, the extent of the support should be as high as possible while, simultaneously, the transmitting strength should be distributed as evenly as possible in order to not discriminate any spatial frequencies. The most promising way so far is to enlarge the support area by incorporating nonlinear effects as in RESOLFT-type microscopy which is introduced in the next section.

2.3 Introduction to RESOLFT-type Microscopes

RESOLFT microscopy can be seen as a family of approaches that utilize reversible saturable optical (fluorescence) transitions. The common formalism is outlined for example in [27, 28, 29, 30] and shall be reviewed here.

As it was outlined in the introduction, the creation of a nonlinear response of marker molecules on the local intensity is the key to improving spatial resolution beyond the diffraction limit. A two-state model is the simplest system that can be used in RESOLFT microscopes. Let us imagine a fluorescent molecule with two distinct states A and B . They can be distinct in their fluorescence, absorption, orientation, or in their ability to undergo a (photo)chemical reaction. The transition $A \rightarrow B$ should be light-driven at a rate $k_{AB} = \sigma_{AB}I$, with σ_{AB} denoting the molecular cross section of the transition $A \rightarrow B$, I denoting the photon flux per area, and λ_{AB} the corresponding wavelength. The reverse transition $B \rightarrow A$ may also be light-driven, thermal, or caused differently with a rate k_{BA} . The kinetics of the molecular states can be described by rate equations. With the normalized population probability $N_{A,B}$ of each state one obtains:

$$dN_A/dt = -dN_B/dt = k_{BA}N_B - k_{AB}N_A. \quad (2.7)$$

We assume that in the beginning the molecule is in state A . For the transition $A \rightarrow B$, we want to consider continuous wave (cw) and pulsed irradiation. For cw irradiance the system reaches an equilibrium at $t \gg (k_{AB} + k_{BA})^{-1}$ and the population of state A is then:

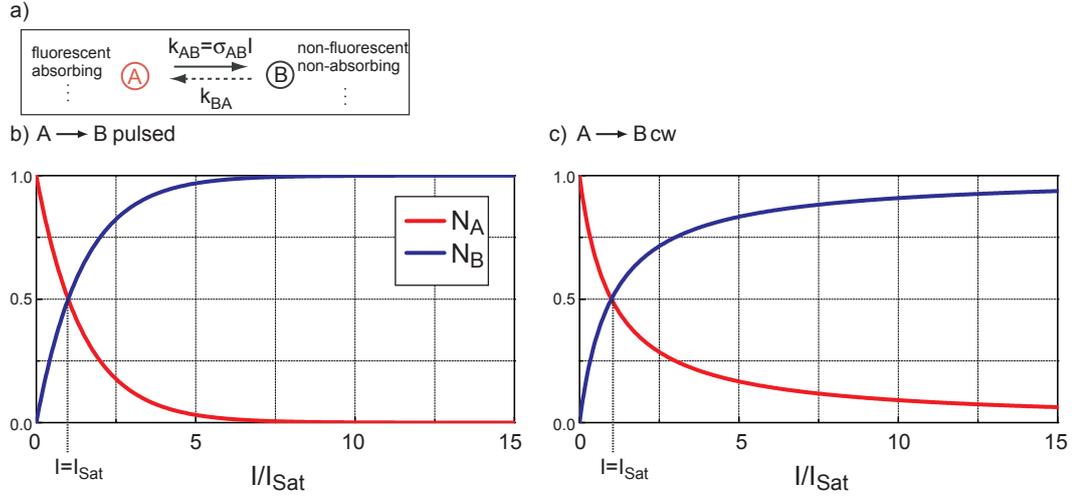


Figure 2.2: a) Simple two-state model with one fluorescent, absorbing, etc. state. The transition $A \rightarrow B$ is driven by light and in the beginning, the system is supposed to be in state A . The dependence of the populations N_A and N_B for state A and B on the applied light intensity is shown for pulsed irradiance b) and for cw irradiance c). In both cases, $N_A = N_B = 1/2$ for $I = I_{\text{Sat}}$ by definition and for large intensities the system can be shifted to state B almost completely.

$$N_A^\infty = k_{BA}/(\sigma_{AB}I + k_{BA}). \quad (2.8)$$

The probability of the molecules to be in A or B therefore depends on I . The intensity where we have equal probability $N_A^\infty = 1/2$ is defined as saturation intensity I_{Sat} . At $I_{\text{Sat}} = k_{BA}/\sigma_{AB}$, half of the molecules are in state B . Increasing I further renders $k_{AB} \gg k_{BA}$, so that the molecule can be virtually shifted to B ($N_A^\infty \rightarrow 0$).

The utilization of pulsed transition light also leads to a saturation at large transition intensities. Assuming the application of a short rectangular pulse with duration τ and wavelength λ_{AB} , the probability of finding the molecule in A directly after the pulse is approximately given by:

$$N_A = \exp(-\sigma_{AB}\tau I), \quad (2.9)$$

for $k_{AB} \gg k_{BA}$. By increasing I the molecule can be virtually shifted to B and the saturation intensity is $I_{\text{Sat}} = \ln 2/(\sigma_{AB}\tau)$. The dependence of the populations of the states on the applied light intensity is shown in figure 2.2. It is highly nonlinear and the largest variation ($dN_{A,B}/dI$) is at $I = 0$. Another common aspect is the asymptotic behavior for large intensities. The state A is driven to a complete de-population in a saturated manner. Accordingly, the molecule is saturated in state B at large intensities.

The RESOLFT concept exploits these nonlinearities. Furthermore, it relies on the preparation of a spatial varying irradiance $I(\mathbf{r})$. Let the intensity distribution $I(\mathbf{r})$ be given by:

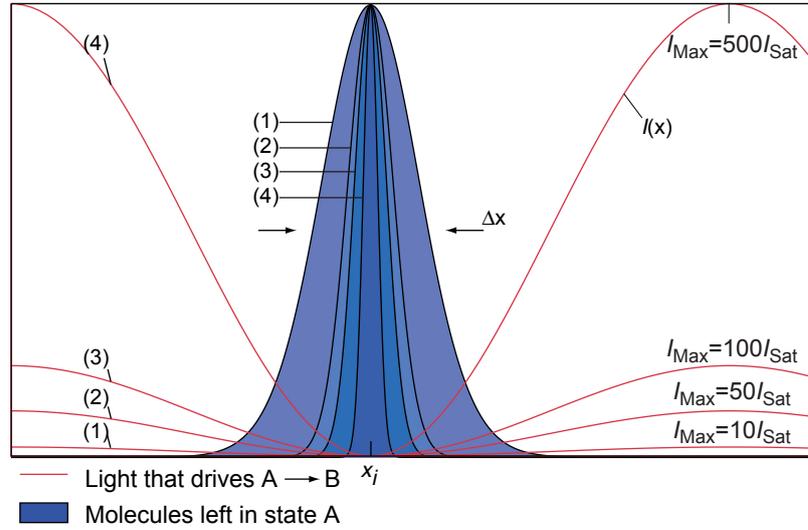


Figure 2.3: Shrinking the spot size of molecules in a specific state by increasing I_{Max} . The profiles 1-4 show the spatial region in which the label is allowed to stay in state A , if a standing wave of light with peak intensities $I_{\text{Max}} = 10, 50, 100$, and 500 times I_{Sat} and with a node at x_i is applied. While I_{Max} is increased the necessary transition intensity to shift the majority of molecules to state B is reached closer to x_i leaving the remaining spot of molecules in state A with an ever smaller width Δx . The width of the spot is not diffraction limited anymore but depends solely on $I_{\text{Max}}/I_{\text{Sat}}$. This concept is generalized to three dimensions without significant changes.

$$I(\mathbf{r}) = I_{\text{Max}} f(\mathbf{r}) \quad 0 \leq f(\mathbf{r}) \leq 1, \quad (2.10)$$

where $f(\mathbf{r})$ is the normalized distribution of a focused light beam, and I_{Max} is the maximal intensity which is proportional to the applied beam power. Consequently, the population of the molecular states will also exhibit a spatial dependence. The change in the population of state A or B is then given by equation (2.7) including a spatially dependent rate: $k_{\text{AB}} = \sigma_{\text{AB}} I(\mathbf{r})$. Initially, all molecules are supposed to be in state A . The dependence of the populations on $I(\mathbf{r})$ is shown in figure 2.2. Let us concentrate on the small area around a single isolated zero of $I(\mathbf{r})$. At the position of the intensity zero, the transition $A \rightarrow B$ will not take place and the molecules at this place will stay in state A . The variations in $f(\mathbf{r})$ are limited due to the diffraction limit, but nevertheless $I(\mathbf{r})$ can be elevated locally to any level by adjusting I_{Max} accordingly. Therefore, the intensity can always be elevated well above I_{Sat} in the close proximity of the intensity zero. There, the population of state A will be quenched well below $1/2$ and we have a pronounced difference in the populations of state A and B within a very small spatial distance. In principle, this condition can be met in an arbitrary small neighborhood of the intensity zero by making the power of the light beam sufficiently large. The possibility to associate the power of the light beam with the extent of the small spot of molecules in a specific state is a distinct property of RESOLFT microscopes.

In figure 2.3, the spatial distribution of molecules left in state A is shown for illumination

with a standing wave and light intensities that exceed I_{Sat} up to a factor of 500. It can be seen that the area where the molecules can reside in A is squeezed down to a spot that is much narrower than the transition light distribution. Theoretically, the spot can be squeezed down to the scale of a single dye molecule. The asymptotical behavior of the spot size on the transition light power can be shown by developing $I(\mathbf{r})$ around an isolated intensity zero ($I(0) = 0$, $I(\mathbf{r}) = I(r)$ for $r < r_0$) in a Taylor series. $I(r)$ is a nonnegative function and therefore the lowest order will be r^2 . Presumed that the width of the resulting spot is already made so small that higher order terms are negligible, one can write a particularly simple proportionality relation of $I(r)$:

$$I(r) \propto I_{\text{Max}} r^2 \quad r < r_0, \quad (2.11)$$

if r_0 is sufficiently small. The proportionality factor depends on $f(r)$. The spot size Δr marks a distance from the intensity zero where the population of state A equals 0.5 which is equivalent to an achieved de-excitation intensity level of I_{Sat} . The following relation holds:

$$\Delta r \propto \frac{1}{\sqrt{\xi}} \quad (2.12)$$

And the saturation factor ξ is defined as:

$$\xi = I_{\text{Max}}/I_{\text{Sat}}. \quad (2.13)$$

Consequently, the volume of the spot which is the product of the spot sizes Δr in all dimensions is proportional to $\xi^{-d/2}$ for a d dimensional system and an isolated intensity zero. Equation (2.12) describes the asymptotic dependence of the FWHM on the power of the light beam. The spot size must already be squeezed down to a size where the behavior of $I(\mathbf{r})$ around the intensity zero can be approximated by a parabola. The shape of the resulting state A distribution for $r < r_0$ is given by inserting equation (2.11) into equations (2.8) and (2.9). Consequently, the shape of the spot will approach a Lorentz function for using cw light and a Gaussian function for using pulsed light asymptotically at high light power.

Under experimental conditions, the intensity zero is often not perfect and a local minimum with a residual intensity ϵI_{Max} remains. Under these circumstances, the highest populations of molecules in state A for the models in this section are $k_{\text{BA}}/(\sigma_{\text{AB}}\epsilon I_{\text{Max}} + k_{\text{BA}})$ and $\exp(-\sigma_{\text{AB}}\tau\epsilon I_{\text{Max}})$ for cw and pulsed light transitions, respectively. This leads to a lowered signal-to-noise ratio with a significant effect on the achievable resolution. Apart from that, the spot size is less efficiently squeezed down for cw illumination.

In practice, one often deals with incomplete transitions $A \rightarrow B$ [37, 52] (e.g. due to coupled re-pumping). Together with an imperfect intensity zero, the spot size would then be squeezed down at first and be broadened again at higher intensities. These effects can be observed in many present experiments [53, 54]. However, for molecular models with higher complexities, e.g. incorporating additional n-photon absorptions, re-excitation and/or (triplet) states, the prediction of the consequences of a residual intensity that would scale with the applied power is not simple and depends on the parameters. In general, it can be

said that any process that decreases the population of state A at the minimum of the intensity distribution while not reducing the population in a small area around the minimum much stronger worsens the achievable resolution. Therefore, ways to prepare a perfect intensity zero are of great importance.

The first and so far most successful implementation of the RESOLFT concept is stimulated emission depletion (STED) microscopy [31, 55]. In STED microscopy, molecules are excited to the fluorescent state S_1 (state A) by an excitation pulse and are immediately transferred back to the molecular ground state S_0 (state B) through stimulated emission by a second light pulse with high intensity. In the simplest realization, the second light pulse is manipulated to form a doughnut distribution with its node at the center of the focused excitation spot. By stimulated emission which is a single-photon phenomenon that has similar cross-sections as single-photon absorption ($\sigma \approx 10^{-16} - 10^{-18} \text{cm}^2$), fluorescence emission can be prevented everywhere except at the center of excitation. Here, STED competes with the spontaneous fluorescence decay of $S_1 \rightarrow S_0$ ($k_{\text{fl}} \approx (1 \text{ns})^{-1}$). The saturation intensity I_{Sat} can be approximated by: k_{fl}/σ . It amounts to $10^{25} - 10^{27}$ photons cm^{-2}s , that is, several tens of MW/cm^2 of peak intensity for 200 ps pulses. Saturated depletion of the excited state then leads to a subdiffraction sized fluorescence spot that can be squeezed down according to equation (2.12). The experimental realization of the de-excitation patterns in STED microscopy was performed by the introduction of phase masks in the beam path that retarded the wavefront locally. The first pronounced sub-diffraction spots with a STED microscope were reported in 2000 by Klar et al. [52]. Using a phase mask that shifted the wavefront by $\lambda/2$ on a circular area, they showed a simultaneous 2-fold improvement in the lateral direction and a 6-fold reduction in the axial direction resulting in a nearly spherical fluorescence spot with a diameter of 90-110 nm. This resolution increase stems from the fact that a field distribution was created that surrounds the focus in all three dimensions. In 2001, a phase mask with a semi-circular wavefront retardation of $\lambda/2$ was used to squeeze the focal spot especially in one lateral direction [56]. Also the incoherent, simultaneous use of two STED beams with different de-excitation patterns was demonstrated. In 2002, Dyba et al. [37] showed that the utilization of STED in 4Pi-microscopes squeezes the focal spot in the axial direction considerably. A FWHM of 33 nm corresponding to $\lambda/23$ was reported, and a phase mask with a shift of about $\lambda/4$ on a circular area was used to additionally cover the sidelobes of the 4Pi focal spot. In 2003, Westphal et al. [36] used single molecules to show a resolution of 28 nm in one lateral direction. As mentioned in the introduction, the goal of this thesis is to find optimal de-excitation patterns. And in fact, we will see that recently, experiments using results of this thesis have delivered a 2D resolution of down to 20 nm in biological systems [57, 58].

The RESOLFT concept is able to reach arbitrary resolution in principle. However, the need of an optimal doughnut distribution that features an intensity zero and can be implemented in a RESOLFT microscope is made clear by noting the practical limitations. E.g., a high usefulness of the dye for the marking process as well as a high photo-stability are necessary. Ultimately, the applicable power is restricted either by a limited supply of available laser power or by the onset of photo-destructive processes in the sample. Therefore,

the reduction of the spot size, as predicted by equation (2.12), is limited. Apart from optimizing the photo-physics of the dye, the highest resolution will therefore only be reached with optimal doughnut distributions. The results from such an optimization would not only be relatively independent from all other characteristics like the dye in use but would also be applicable in addition to all other possible optimizations. Therefore, a high generality and cross sectional applicability of optimized doughnut distributions is expected.

Several requirements for the ideal focal distribution of a focused light beam can be stated as follows:

1. strict intensity zero at the doughnut center,
2. largest possible intensity close by the center ($< \lambda/2$),
3. as uniform as possible around the center.

A focal distribution of light fulfilling these conditions shall be called an ideal doughnut distribution. Retrieving the ideal doughnut distribution is an optimization problem. Therefore, a search in the set of feasible doughnut distributions must be performed. In order to perform the search, an understanding of the formation of doughnut distributions is needed. In section 2.5, the formation of doughnut distributions is outlined and a framework for the efficient search in these distributions is derived. The quality of a resulting de-excitation pattern is estimated by a figure-of-merit (FoM) that assigns a real value to each possible doughnut distribution. The actual choice for the FoM is presented in chapter 3 and 4 where the optimizations for single as well as for multiple doughnut distributions are performed. To estimate the resolution that can be obtained with a doughnut distribution, we need to know how images are formed in RESOLFT microscopes. Especially vectorial effects can have a prominent influence on the achievable resolution. The consideration of these effects in the image formation of RESOLFT microscopes is done in the next section.

2.4 Vectorial Image Formation in a RESOLFT Microscope

To consider the vectorial effects in the image formation of RESOLFT-type microscopes, the orientational distribution of molecules has to be taken into account. The orientation is determined by the molecular transition moment. Sometimes the orientations of the absorption and the emission dipole moments are not equal [59]. However, the most common situation is a parallel orientation which is assumed here. Their direction is then given by \mathbf{n}_D which is determined in an angular coordinate system by:

$$\mathbf{n}_D(\theta, \phi) = \begin{pmatrix} \sin \theta \cos \phi \\ \sin \theta \sin \phi \\ \cos \theta \end{pmatrix}, \quad 0 \leq \theta \leq \pi, \quad 0 < \phi \leq 2\pi. \quad (2.14)$$

Let us define the molecular orientational density $D(\theta, \phi, \mathbf{r})$ at position \mathbf{r} . The number of molecules $G(\mathbf{r})$ at \mathbf{r} is then given by:

$$G(\mathbf{r}) = \frac{1}{4\pi} \int_0^\pi \int_0^{2\pi} D(\theta, \phi, \mathbf{r}) \sin \theta \, d\theta d\phi. \quad (2.15)$$

We are interested in the resulting fluorescence after the application of excitation and de-excitation light. Therefore, the orientational density of excited molecules at position \mathbf{r} is introduced. It is denoted by $D_{\text{Exc}}(\theta, \phi, \mathbf{r})$. The molecular response could be rather complex, but we restrict us here to a common case of RESOLFT microscopy where the excitation is solely used to prepare the state A and both, the excitation and the nonlinear de-excitation are applied as short pulses which are not overlapping in time. In the dipole approximation, the transition rates are proportional to the projection of the electric field vectors on the dipole transition moment $|\mathbf{E} \cdot \mathbf{n}_D|^2$. The components of the electric field in the focal region of a high NA lens $\mathbf{E} = (E_i)$ with $i = x, y, z$ can be described by an amplitude point spread function (A-PSF) which was calculated for example in [60]. The vectorial electric field distribution shall be given by $\mathbf{E}_{\text{Exc}}(\mathbf{r})$ for the excitation and by $\mathbf{E}_{\text{NL}}(\mathbf{r})$ for the nonlinear de-excitation. The excitation shall depend linearly on the excitation rate and the nonlinear dependence of the de-excitation rate shall be given by F_{NL} according to equation (2.9). The effective orientational density of excited molecules is then mainly given by the product of $D(\theta, \phi, \mathbf{r})$, the excitation rate, and the nonlinear dependence F_{NL} on the de-excitation rate. One obtains:

$$D_{\text{Exc}}(\theta, \phi, \mathbf{r}) = D(\theta, \phi, \mathbf{r}) C |\mathbf{E}_{\text{Exc}}(\mathbf{r}) \cdot \mathbf{n}_D|^2 F_{\text{NL}}(|\mathbf{E}_{\text{NL}}(\mathbf{r}) \cdot \mathbf{n}_D|^2), \quad (2.16)$$

with a proportionality constant C . The number of excited molecules $G_{\text{Exc}}(\mathbf{r})$ is calculated analogously to $G(\mathbf{r})$ in equation (2.15). In general, $G_{\text{Exc}}(\mathbf{r})$ cannot be expressed as a product of $G(\mathbf{r})$ with an effective PSF which solely depends on $\mathbf{E}_{\text{Exc}}(\mathbf{r})$ and $\mathbf{E}_{\text{NL}}(\mathbf{r})$. The integration over all directions is sensitive to spatial inhomogeneities of $D(\theta, \phi, \mathbf{r})$. Molecules oriented parallel to the polarization of the excitation light are excited very well whereas other molecules are less efficiently excited. It is worthwhile to note that the contribution of molecules of each specific orientation to the whole image is the convolution of the spatial distribution of molecules of this orientation and an effective PSF that is specific for this orientation. In order to maintain this property in the whole image which is the sum of all these contributions, the molecules' orientational distribution must be separable into a spatial molecular distribution $G(\mathbf{r})$ and a probability density distribution for the molecules' orientations $D_p(\theta, \phi)$.

$$D(\theta, \phi, \mathbf{r}) = G(\mathbf{r}) D_p(\theta, \phi) \quad \text{with} \quad \frac{1}{4\pi} \int_0^\pi \int_0^{2\pi} D_p(\theta, \phi) \sin \theta \, d\theta d\phi = 1 \quad (2.17)$$

Then the distribution of molecules' orientations is spatially invariant. In most practical cases, a random orientation of the molecules will be present which is denoted by $D_p = 1$. Analogously to equation (2.16), the subpopulation of $D_p(\theta, \phi)$ which is effectively excited

can be calculated. In order to derive an expression for an effective PSF in this case, the detection process has to be considered. The efficiency with which emitted photons of the molecules are detected by a photo detector which is located behind a confocal pinhole is given by the collection efficiency function $CEF(\theta, \phi, \mathbf{r})$ [61]. The effective PSF for the case of a spatial invariance of the molecules' orientational distribution is given by the integral over all possible molecular orientations (θ, ϕ) in the product of $D_p(\theta, \phi)$, the de-excitation factor F_{NL} and the corresponding collection efficiencies.

$$h_{\text{Eff}} = \frac{1}{4\pi} \int_0^\pi \int_0^{2\pi} D_p |\mathbf{E}_{\text{Exc}}(\mathbf{r}) \cdot \mathbf{n}_D|^2 F_{NL} (|E_{NL}(\mathbf{r}) \cdot \mathbf{n}_D|^2) CEF(\theta, \phi, \mathbf{r}) \sin \theta d\theta d\phi \quad (2.18)$$

The whole image I is then given by the convolution:

$$I = h_{\text{Eff}} \otimes G. \quad (2.19)$$

This describes the image formation in the case that all molecules are immobile. Please note that the effective PSF is still dependent on the orientational distribution D_p .

In practice, some sort of rotational diffusion can usually be observed. It can take place on the sub nanosecond timescale [62]. Therefore, additional expressions for h_{Eff} for slow (R_{Slow}) and for very fast, unhindered (R_{Fast}) rotation regimes will be given. Starting with an uniformly oriented sample, due to excitation and de-excitation an oriented excited subsample is prepared. Nevertheless, the orientation of this subsample is eventually lost by rotational diffusion. In the case R_{Slow} , it is assumed that this loss takes place on the (slow) ns timescale but still before the emission. Even slower rotation would lead to the static case which has been analyzed above. In the case R_{Fast} , very fast rotation shall occur which destroys the orientational preferences already in the interval between excitation and de-excitation pulses. The effective PSFs are then given by:

$$\begin{aligned} h_{\text{Eff}} &= h_{\text{Exc}} h_{\text{Det}}, \\ h_{\text{Exc}, R_{\text{Slow}}} &= \frac{1}{4\pi} \int_0^\pi \int_0^{2\pi} C |\mathbf{E}_{\text{Exc}}(\mathbf{r}) \cdot \mathbf{n}_D|^2 F_{NL} (|E_{NL}(\mathbf{r}) \cdot \mathbf{n}_D|^2) \sin \theta d\theta d\phi, \\ h_{\text{Exc}, R_{\text{Fast}}} &= C |E_{\text{Exc}}|^2 \frac{1}{4\pi} \int_0^\pi \int_0^{2\pi} F_{NL} (|E_{NL}(\mathbf{r}) \cdot \mathbf{n}_D|^2) \sin \theta d\theta d\phi \text{ and} \\ h_{\text{Det}} &= \frac{1}{4\pi} \int_0^\pi \int_0^{2\pi} CEF(\theta, \phi, \mathbf{r}) \sin \theta d\theta d\phi. \end{aligned} \quad (2.20)$$

h_{Exc} describes the effective excitation PSF after application of excitation and de-excitation pulse. In the case of R_{Fast} , the influence of the excitation pulse has a particularly simple form.

In this dissertation, highly nonlinear de-excitation effects F_{NL} are investigated. Major distortions of the shape of the effective PSF are possible with these effects. Nevertheless, if the orientational distribution of the molecules in the sample is spatially variable, the total image will not be the convolution of the spatial molecules' distribution with an effective PSF anymore. In particular, no de-excitation effect would occur if the molecules' orientation would be perpendicular to the polarization of E_{NL} . This problem can be overcome, if all directions and components are treated equally in the excitation, the de-excitation, and the detection. The image is then independent of the orientational distribution of the molecules. Therefore, an additional requirement for the ideal doughnut distribution shall be that at least the de-excitation is as independent as possible of the dye's orientation.

The common observation for de-excitation fields that do not cover all polarization directions are remaining tails in the effective PSF. They consist of dye molecules that were excited but could not be de-excited efficiently. Vectorial effects in the image formation can also lead e.g. to the breaking of the axial symmetry in the effective PSF as it can be seen in figure 2(f) in [54]. However, this effect could hardly be observed in experimental data.

In section 2.3, the saturation factor ξ was defined as the ratio of the maximal intensity and the saturation intensity (where half of the fluorophores are quenched). Therefore, locally, a maximal quenching effect down to $2^{-\xi}$ can be reached if orientational effects are not regarded. The transition of the scalar model of section 2.3 to a vectorial model leads to the notion that only under certain, optimal conditions a suppression effect of $2^{-\xi}$ can be reached. The necessary conditions are that the dye molecules are strictly oriented in only one direction and the de-excitation field amplitude E_{NL} is parallel to that direction and has no complex parts. Otherwise, if for example circular polarized light is used, only a suppression factor of $2^{-\xi/2}$ can be reached. And therefore only at a saturation factor of 2, half of the molecules will be quenched. For a random distribution of molecules and circular polarized de-excitation light a saturation factor of ≈ 2.82 is needed to de-excite half of all molecules.

2.5 Engineering Focal Intensity Distributions

Doughnut intensity distributions are characterized by complex vectorial pupil functions. They represent the state of the incoming wavefront at the objective lens back aperture. An optimization in the set of feasible doughnut distribution and corresponding pupil functions will be performed in chapters 3 and 4. In order to efficiently search for the ideal doughnut distribution, we will present an effective method to calculate the focal field, given a complex vectorial pupil function. Most importantly, a method to ensure strict intensity zeros is developed.

2.5.1 Calculation of the Focal Field

To calculate the electromagnetic field \mathbf{E} near the focus \mathbf{O} of a lens with high numerical aperture, the coordinate systems and vectors shown in figure 2.4 are introduced. The electromagnetic field near the focus $\mathbf{E}(P_2)$ is then expressed by [63]:

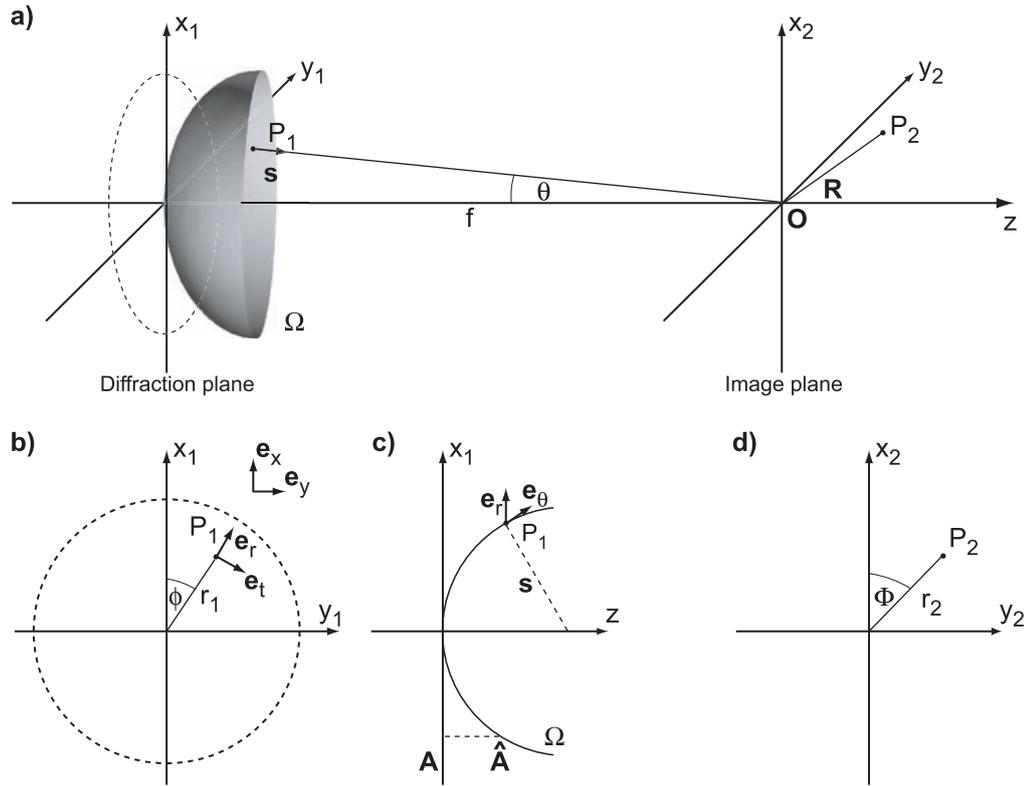


Figure 2.4: Coordinate systems used for describing the diffraction. z is the optic axis with the geometrical focus O at its origin. \mathbf{R} is a vector to the observation point P_2 which is near the focus. At the distance of the focal length f from the focus is the spherical cap Ω . Rays connecting points P_1 on Ω and the focus have an angle θ with the optic axis. In the diffraction plane $z = -f$, the back aperture of the objective lens (dashed circle) is shown. ϕ is the azimuthal angle in the diffraction plane. The complex vectorial pupil function $\mathbf{A}(r_1, \phi)$ which describes the state of the incoming light beam gives rise to a complex vectorial light distribution $\hat{\mathbf{A}}$ on Ω . In the decomposition of \mathbf{A} and $\hat{\mathbf{A}}$ into radial and tangential unit vectors: \mathbf{e}_r and \mathbf{e}_t , the component parallel to \mathbf{e}_t is unaffected by diffraction while the component parallel to \mathbf{e}_r changes its direction to \mathbf{e}_θ . Finally, Φ is the azimuthal angle in the $z = 0$ plane.

$$\mathbf{E}(P_2) = \frac{i}{f^2 \lambda} \int_{\Omega} \int \hat{\mathbf{A}}(P_1) \exp(-i\mathbf{s} \cdot \mathbf{R}) d\Omega. \quad (2.21)$$

This approximation is called the Debye approximation. The field near the focus is generated by a superposition of plane waves of different propagation directions \mathbf{s} filling the aperture Ω which represents a cap of a sphere centered on the focus with radius f . $\hat{\mathbf{A}}$ describes the complex polarization state of the electromagnetic field on Ω after diffraction. Therewith, the polarization of the beam as well as its amplitude or phase are completely determined. Allowing arbitrary transversal polarizations to be included in $\hat{\mathbf{A}}$ permits to explore the full vectorial characteristics of the electromagnetic field near the focus.

A spherical coordinate system is used to represent points P_1 on the spherical surface Ω :

$$\begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} = f \begin{pmatrix} \sin \theta \cos \phi \\ \sin \theta \sin \phi \\ \cos \theta \end{pmatrix}. \quad (2.22)$$

Regarding points P_2 near the focus, a cylindrical coordinate system originating at \mathbf{O} is introduced. The position of the point P_2 representing the position vector \mathbf{R} is:

$$\begin{pmatrix} x_2 \\ y_2 \\ z_2 \end{pmatrix} = \begin{pmatrix} r_2 \cos \Phi \\ r_2 \sin \Phi \\ z_2 \end{pmatrix}. \quad (2.23)$$

Then the following transformation can be used for the solid integration angle: $d\Omega = \sin \theta d\theta d\phi$. The unit vector \mathbf{s} is expressed as $(\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta)$. Therefore, $\mathbf{s} \cdot \mathbf{R}$ is:

$$\mathbf{s} \cdot \mathbf{R} = r_2 \sin \theta \cos(\phi - \Phi) + z_2 \cos \theta. \quad (2.24)$$

Now we want to replace $\hat{\mathbf{A}}$ by the components of the electromagnetic field in the diffraction plane ($z_1 = -f$). The distribution of the complex polarization states in the diffraction plane $\mathbf{A}(r_1, \phi)$ is termed pupil function. It depends on the state of the incoming light beam before diffraction. Here, it is described as a spatial variable amplitude and phase distribution with independent components for x - and y -polarization:

$$\mathbf{A}(r_1, \phi) = A_x(r_1, \phi)\mathbf{e}_x + A_y(r_1, \phi)\mathbf{e}_y. \quad (2.25)$$

It can also be expressed as a linear combination of a radial and a tangential unit vector:

$$\mathbf{A}(r_1, \phi) = A_r(r_1, \phi)\mathbf{e}_r + A_t(r_1, \phi)\mathbf{e}_t \quad (2.26)$$

with

$$\begin{pmatrix} \mathbf{e}_r \\ \mathbf{e}_t \end{pmatrix} = \begin{pmatrix} \cos \phi & \sin \phi \\ -\sin \phi & \cos \phi \end{pmatrix} \begin{pmatrix} \mathbf{e}_x \\ \mathbf{e}_y \end{pmatrix}. \quad (2.27)$$

The complex vectorial light distribution $\hat{\mathbf{A}}(\theta, \phi)$ on the cap Ω can be expressed by its counterpart in the diffraction plane $\mathbf{A}(r_1, \phi)$. The polar coordinate system (r_1, ϕ) in the diffraction plane can be transformed to the spherical coordinates of Ω by:

$$(r_1, \phi) = (f \sin \theta, \phi). \quad (2.28)$$

An aplanatic lens obeying Abbe's sine condition is assumed. There, rays enter and leave the lens at the same distances from the optic axis and an apodization factor $\sqrt{\cos \theta}$ is introduced because of energy conservation. Furthermore, the electromagnetic light component parallel to \mathbf{e}_r is refracted toward the focus at high angles θ . The new direction \mathbf{e}_θ is given by: $\cos \theta \mathbf{e}_r + \sin \theta \mathbf{e}_z$. The tangential component remains unchanged as well as the position of the according field contributions in the (x_1, y_1) plane. The field on the spherical surface $\hat{\mathbf{A}}$ is then given by:

$$\hat{\mathbf{A}}(\theta, \phi) = \sqrt{\cos \theta}((\mathbf{A} \cdot \mathbf{e}_r)\mathbf{e}_\theta + (\mathbf{A} \cdot \mathbf{e}_t) \cdot \mathbf{e}_t) \quad (2.29)$$

and for the cartesian components of $\hat{\mathbf{A}}(\theta, \phi)$ one gets:

$$\hat{\mathbf{A}}(\theta, \phi) = \sqrt{\cos \theta}(A_x(\theta, \phi)\mathbf{B}(\theta, \phi) + A_y(\theta, \phi)\mathbf{M}\mathbf{B}(\theta, \phi - \pi/2)), \quad (2.30)$$

with the cartesian unit vector:

$$\mathbf{B}(\theta, \phi) = \begin{pmatrix} \cos \theta + (1 - \cos \theta) \sin^2 \phi \\ (\cos \theta - 1) \sin \phi \cos \phi \\ \sin \theta \cos \phi \end{pmatrix}, \quad (2.31)$$

and a matrix:

$$\mathbf{M} = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad (2.32)$$

Equations (2.21), (2.24) and (2.30) give an expression for the Cartesian components of the electric field $\mathbf{E}(P_2)$ near the focus:

$$\mathbf{E}(P_2) = iC \int_0^\alpha \int_0^{2\pi} \hat{\mathbf{A}}(\theta, \phi) \exp(ikn(z_2 \cos \theta + r_2 \sin \theta) \cos(\phi - \Phi)) \sin \theta d\theta d\phi. \quad (2.33)$$

Here a normalization constant C is introduced. In the case of linearly polarized light along the x-axis ($A_x = 1$ and $A_y = 0$), this expression reduces to equation (2.26) in Richards and Wolf [60]. In the next section, we want to introduce a complete and orthonormal basis of the pupil function space.

2.5.2 The Optimization Space and its Basis

The electromagnetic field at the back aperture of the objective lens is completely described by a vectorial complex field distribution $\mathbf{A}(r_1, \phi), \mathbb{R}^2 \rightarrow \mathbb{C}^2$. The variables (r_1, ϕ) are polar coordinates in the back aperture plane. It is assumed that the objective lens has a finite circular extent (see figure 2.4(b)). Therefore, only the values of \mathbf{A} inside the extent of the aperture are interesting for the following discussion. The radial coordinate r_1 can be normalized by:

$$r = r_1/(f \sin \alpha) \text{ and } 0 \leq r \leq 1. \quad (2.34)$$

Equation (2.25) shows the composition of \mathbf{A} by completely independent cartesian components. Therewith, spatially varying polarization states can be expressed. For simplicity, a decomposition of the pupil function \mathbf{A} will be performed for a spatially homogeneous polarization only. With a normalized polarization vector \mathbf{P} one gets:

i	n	m	$Z_{mn,e/o}(r, \phi)$	Aberration term
1	0	0	1	bias
2	1	1	$r \cos(\phi)$	tilt x
3	1	1	$r \sin(\phi)$	tilt y
4	2	0	$2r^2 - 1$	defocus
5	2	2	$r^2 \cos(2\phi)$	astigmatism x
6	2	2	$r^2 \sin(2\phi)$	astigmatism y
7	3	1	$(3r^3 - 2r) \cos(\phi)$	coma x
8	3	1	$(3r^3 - 2r) \sin(\phi)$	coma y
9	3	3	$r^3 \cos(3\phi)$	
10	3	3	$r^3 \sin(3\phi)$	
11	4	0	$6r^4 - 6r^2 + 1$	spherical (1st order)
12	4	2	$(4r^4 - 3r^2) \cos(2\phi)$	
13	4	2	$(4r^4 - 3r^2) \sin(2\phi)$	
14	4	4	$r^4 \cos(4\phi)$	
15	4	4	$r^4 \sin(4\phi)$	

Table 2.1: The first 15 Zernike polynomials $Z_n^{\pm m}(r, \phi)$ with the associated primary wavefront aberrations.

$$\mathbf{A}(r, \phi) = \mathbf{P}\tilde{A}(r, \phi), \quad (2.35)$$

with a complex amplitude and phase distribution $\tilde{A}(r, \phi)$. The following decomposition of $\tilde{A}(r, \phi)$ into Zernike polynomials $Z_n^{\pm m}(r, \phi)$ is possible:

$$\tilde{A}(r, \phi) = \sum_{n,m} v_n^{\pm m} Z_n^{\pm m}(r, \phi), \quad (2.36)$$

with complex coefficients $v_n^{\pm m}$ and integers $n \geq 0$, $n \geq m \geq 0$ and $n - m$ even. The Zernike polynomials [64, 65] form a complete set of real valued orthogonal functions on the unit disk (also see appendix A.1.1) and play a fundamental role in the diffraction theory of optical aberrations. They were introduced by Zernike [66] in 1934 and used by Nijboer [67]. Frequently, a Zernike polynomial present in the decomposition of the wavefront is referred to as Zernike mode with $v_n^{\pm m}$ denoting the strength of a particular mode [68]. Normally higher order Zernike modes are found only in marginal fractions in a wavefront simplifying the physical interpretation of low order Zernike modes. Some modes are directly related to classical aberrations (see table 2.1). The Zernike polynomials are divided in even and odd parts:

$$\begin{aligned} Z_n^{+m}(r, \phi) &= R_n^m(r) \cos(m\phi) \\ Z_n^{-m}(r, \phi) &= R_n^m(r) \sin(m\phi). \end{aligned} \quad (2.37)$$

The radial functions $R_n^m(r)$ are defined by:

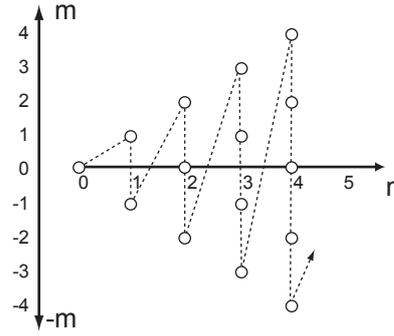


Figure 2.5: Re-numbering of the Zernike polynomials. Each circle represents a permissible pair of indices (m, n) . The new index i runs according to the dotted line.

$$R_{mn}(r) = \sum_{l=0}^{(n-m)/2} \frac{(-1)^l (n-l)!}{l! [(n+m)/2 - l]! [(n-m)/2 - l]!} r^{n-2l}. \quad (2.38)$$

However, it is more convenient to renormalize and re-number the Zernike polynomials $Z_i(r, \phi)$ obtaining orthonormal polynomials with a single index only (see figure 2.5). The functions $\tilde{Z}_i(r, \phi)$ are defined as:

$$\begin{aligned} \tilde{Z}_i(r, \phi) &= \sqrt{2(n+1)/\pi} Z_n^{\pm m}(r, \phi), & m \neq 0, \\ \tilde{Z}_i(r, \phi) &= \sqrt{(n+1)/2/\pi} Z_n^0(r, \phi), & m = 0. \end{aligned} \quad (2.39)$$

It is straightforward to prove the orthonormality relation:

$$\int_0^1 \int_0^{2\pi} \tilde{Z}_i(r, \phi) \tilde{Z}_j(r, \phi) r dr d\phi = \delta_{ij}. \quad (2.40)$$

An approximate decomposition of the pupil function $\tilde{A}(r, \phi)$ in a finite number n_Z of polynomials $\tilde{Z}(r, \phi)$ is then given by:

$$\tilde{A}(r, \phi) = \sum_{i=1}^{n_Z} c_i \tilde{Z}_i(r, \phi), \quad c_i \in \mathbb{C}. \quad (2.41)$$

Finer details can always be included by increasing the number of polynomials \tilde{Z} in the decomposition. It is completely described by the coefficients c_i . The incident laser power P is approximated by:

$$P \sim \int_0^1 \int_0^{2\pi} \tilde{A}(r, \phi)^2 r dr d\phi = \sum_{i=1}^{n_Z} c_i^2. \quad (2.42)$$

A particularly efficient computation of many Zernike polynomials over the aperture using recurrence relations is outlined in appendix A.1.2.

2.5.3 Constructing Subspaces Providing Strict Intensity Zeros

We start by inserting the decomposition of the pupil function (equations (2.35) and (2.41)) into the diffraction integrals (equations (2.30) and (2.33)). The right hand side of equation (2.33) shall be denoted by $\mathbf{K}(\tilde{A}, \mathbf{r}_2)$. Due to the linearity of the integral, the summation over our polynomials \tilde{Z}_j can be brought outside the integral. Equation (2.33) can then be written as:

$$E_i(\mathbf{r}_2) = \sum_{j=1}^{n_Z} c_j \mathbf{K}(\tilde{Z}_j, \mathbf{r}_2) \cdot \mathbf{e}_i. \quad (2.43)$$

It describes the electric field at position \mathbf{r}_2 as a sum of contributions from electric fields that would have been produced if pupil functions consisting of single polynomials \tilde{Z}_j would have been present at the back aperture. Now we want to evaluate equation (2.43) on a grid near the focus. The optimization method will be able to engineer the PSF at these grid positions then. The grid points are chosen such that the zero intensity points are part of the grid. A single index l is sufficient to enumerate the grid point positions as well as the cartesian components of the electric field. A convenient notation is the field vector Y_l defined as:

$$Y_l = E_i(\mathbf{r}_k) \quad (2.44)$$

and the matrix elements m_{lj} :

$$m_{lj} = \mathbf{K}(\tilde{Z}_j, \mathbf{r}_k) \cdot \mathbf{e}_i, \quad (2.45)$$

where $l = 3k + i$ and $i = 0, 1, 2$ stand for the cartesian components x, y, z .

For n_r interesting positions, at which the diffraction integral will be evaluated, we have $3n_r$ values of Y_l and $3n_r n_Z$ values of m_{lj} . The diffraction integral is now written as:

$$Y_l = \sum_{j=1}^{n_Z} c_j m_{lj} \quad (2.46)$$

or as matrix equation:

$$\mathbf{Y} = \mathbf{M}\mathbf{C}, \quad (2.47)$$

with the vector $\mathbf{Y} = (Y_l) \in \mathbb{C}^{3n_r}$, vector $\mathbf{C} = (c_j) \in \mathbb{C}^{n_Z}$ and matrix $\mathbf{M} = (m_{lj}) (3n_r \times n_Z)$. The matrix \mathbf{M} is a linear mapping between the pupil functions and the focal field at certain positions. If \mathbf{M} is pre-calculated the field vector Y_l can be computed efficiently for many coefficient vectors \mathbf{C} . The physical challenge stated in section 2.3 is to find the element of the vector space $\mathbf{C} \in \mathbb{C}^{n_Z}$ within an optimization that yields a certain favorable vector \mathbf{Y} .

It is not clear a-priori what the most favorable \mathbf{Y} is for our purposes. Therefore, a FoM that depends nonlinearly on \mathbf{Y} will be chosen. It maps the goodness of \mathbf{Y} with a real number so that more favorable focal fields have a higher value and the optimization is to find a \mathbf{C} that maximizes this FoM.

One of the requirements for the ideal doughnut distribution were strict boundary conditions, the most important being strict intensity zeros at one or more positions. While a high penalty in the FoM for violations of these boundary conditions would in principle also be viable, it turned out that, under practical conditions, an efficient algorithm must restrict the optimization to a subspace of \mathbb{C}^{n_Z} which fulfills the zero intensity conditions directly. In order to enforce an intensity zero at one grid position, all components of the electric field at this position must be set to zero. Suppose that $Y_l = 0$ for $l = l_1, \dots, l_m$ and let us assume that the number of polynomials n_Z is chosen larger than the number of components of \mathbf{Y} that are set to zero. Furthermore, let $\tilde{\mathbf{M}}$ be the matrix consisting of the l_i -th row vectors of \mathbf{M} . In order to construct the subspace with $Y_{l_1} = \dots = Y_{l_m} = 0$, the solution space of the following equation must be found:

$$\mathbf{0} = \tilde{\mathbf{M}}\mathbf{C}. \quad (2.48)$$

This equations system will have more unknowns (c_j) than equations and a unique solution is not expected. The nullspace of matrix $\tilde{\mathbf{M}}$ can be constructed with a singular value decomposition which gives an orthonormal basis for the nullspace [69, 70]. This decomposition is uniquely determined [69] and an appropriate algorithm is given e.g. in [70]. A variant of this algorithm that works with complex variables and coefficients has been used in this thesis. Details about the singular value decomposition and the structure of the orthonormal basis $\text{span}\{\mathbf{v}_{n+1}, \dots, \mathbf{v}_{n_Z}\}$ of the nullspace of $\tilde{\mathbf{M}}$ are given in appendix A.2. The number of basis vectors of the nullspace $n_Z - n$ depends on the structure of $\tilde{\mathbf{M}}$. Our purpose was to restrict the optimization of the focal field to pupil functions that preserve intensity zeros at certain positions. This is satisfied as long as one is only considering coefficient vectors \mathbf{C} that are within the nullspace of $\tilde{\mathbf{M}}$. The coefficient vectors can then be expressed as:

$$\mathbf{C} = \sum_{i=n+1}^{n_Z} \alpha_i \mathbf{v}_i, \quad (2.49)$$

with complex coefficients α_i . Substitution of the last equation into equation (2.46) gives:

$$Y_l = \sum_{j=1}^{n_Z} \sum_{i=n+1}^{n_Z} \alpha_i v_{ij} m_{lj}. \quad (2.50)$$

The optimization is now conducted in the subspace $\{\alpha_{n+1} \mathbf{v}_{n+1}, \dots, \alpha_{n_Z} \mathbf{v}_{n_Z}\}$ of \mathbb{C}^{n_Z} . With it, the strict intensity zero at the focus will be enforced for the optimization in chapter 3 as well as the multiple intensity zeros at the multiple foci investigated in chapter 4.

With this method specific amplitudes of the electromagnetic field at specific positions can be enforced routinely and efficiently. However, the enforcement of specific intensity val-

ues apart from zero cannot be achieved with this algorithm because then nonlinear conditions on the field components would occur.

3 Optimal Single Focus De-Excitation Patterns

RESOLFT-type microscopes rely on the creation of very small spatial regions of excited molecules. These small regions result from excitation of comparably large areas and de-excitation in a saturated way close to the focus. A high-resolution image is obtained by scanning the effective excitation pattern through the sample. In the simplest version, only one very small spot is created and moved across the sample. Although scanning was initially developed for electron microscopes [71], it is now a widely used, common technique [72]. Several doughnut-shaped patterns for single spot scanning have been presented [52, 56, 37, 36] with STED microscopes. However, a systematic and comprehensive survey of de-excitation patterns was never conducted and is presented here for the first time. The aim was to identify the necessary wavefront manipulations that lead to the most favorable focused de-excitation light beam which will yield the highest resolution under given experimental conditions. This was achieved by a rigorous optimization of the de-excitation patterns with a single intensity zero. The requirements defining the ideal doughnut intensity distribution were stated in section 2.3. They are used to define a figure-of-merit (FoM) in an optimization algorithm that is able to search efficiently in the space of suitable pupil functions. Because different applications require high resolution in one, two or three directions simultaneously, doughnut shapes of different dimensionality are considered in the analysis (see figure 3.1).

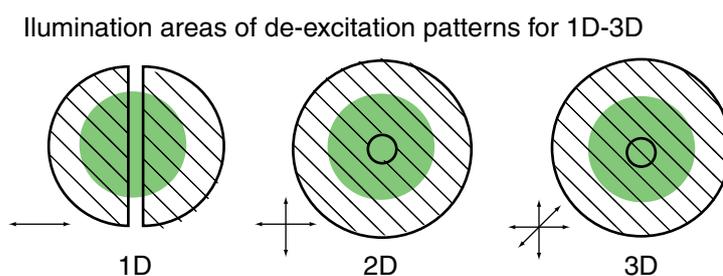


Figure 3.1: De-excitation areas for 1D, 2D and 3D patterns. The green area depicts the excited area. The molecules get de-excited at the hatched areas. The excited area will be squeezed in 1, 2 or 3 directions correspondingly.

3.1 Optimization Algorithm

The requirements for an ideal doughnut intensity distribution are based on the notion that some form of limitation on the applicable RESOLFT power is always present. Otherwise, spot sizes could be made arbitrarily small in any RESOLFT microscope that features a kind of isolated intensity zero. In practice, this is not possible and the following restrictions are common:

- (A) The maximum amplitude in the aperture is limited. This condition is equivalent to a limitation of the available laser power if shaping of the aperture's amplitude distribution is accomplished by transmission filters. Most recent experiments are restricted in this way.
- (B) The total power the sample can sustain without damage is limited. This can be caused by photobleaching, trapping effects, thermal instability or other detrimental effects on the dye's distribution or functionality. In many recent STED experiments, photobleaching is clearly existent and limits the resolving power.
- (C) The maximum local intensity the sample can sustain without damage is limited. Possible causes include nonlinear photobleaching [73, 74, 75], dielectric breakthrough or transient heating.

The goal of the optimization is to find the pupil function $\mathbf{A}(r, \phi)$ that features a strict intensity zero at the geometrical focus while generating the steepest doughnut intensity distribution around the focus. In addition, this pattern shall be as uniform as possible in the spatial directions as well as in the polarization directions. The presence of completely monochromatic and coherent laser light as source for the de-excitation patterns is assumed. The optimization is restricted to a constant polarization \mathbf{P} throughout the pupil function because more complex polarization schemes are experimentally not easily accessible. We considered linear and circular polarization:

$$\begin{aligned} \mathbf{P}_l &= \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \\ \mathbf{P}_c &= \frac{1}{\sqrt{2}} \begin{pmatrix} \pm 1 \\ i \end{pmatrix}. \end{aligned} \tag{3.1}$$

In section 2.5.2, the decomposition of $\tilde{A}(r, \phi)$ in a set of orthonormal polynomials \tilde{Z}_i was shown. Together with \mathbf{P} , the vectorial pupil function $\mathbf{A}(r, \phi)$ is formed. It determines the doughnut intensity distribution completely. The corresponding focal field can be calculated by equation (2.33).

For the optimization of a single doughnut focus the first 90-120 polynomials \tilde{Z}_i were chosen (corresponding to a maximal order $n=12-14$). The corresponding multi-dimensional function space \mathcal{S} includes good approximations for functions that vary only with a low order of the

cartesian coordinates corresponding to a certain restricted oscillation speed. In section 2.5.3, the construction of a subspace of the pupil functions was shown that yielded strict intensity zeros at arbitrary positions. It turns out that this step can be done manually in the case of only one intensity zero located at the origin. Many functions \tilde{Z}_i already feature an intensity zero at the focus due to inherent symmetries. All others have only one or two nonzero components. It is straightforward to manually remove three basis functions and replace all others by linear combinations with one of them so that an intensity zero at the origin is always formed.

Coefficient vectors $\mathbf{C} = (c_i)$ are used to represent the pupil functions $\tilde{A}(r, \phi)$ (see section 2.5.3). The pupil functions are normalized in every iteration of the optimization algorithm in order to fulfill the conditions corresponding to regimes (A), (B) and (C). That means that for (A) the maximal amplitude and for (B) the average amplitude of the pupil function must be held constant while for (C) the maximal local intensity of the doughnut intensity distribution was fixed. A global optimization algorithm was designed to maximize a FoM that assigned a real number to each pupil function. The purpose of the FoM is to measure the potential of the pupil function for resolution increase. The measure that is most directly linked with the steepness of doughnut intensity distributions is the minimum of the second derivatives of the intensity at the focus over all radial directions. A large value corresponds to a steep rise everywhere around the focus. However, in doing so, the remaining shape of the doughnut distribution is not considered. Therefore, directly maximizing the intensity around the focus turns out to be a more successful strategy. Points of interest P_{ER} are placed around the focus at distances d_{ER} corresponding to the expected resolution of the system. The FoM is denoted by M and is calculated as the minimal intensity on these points. While the minimal intensity on a circle or a spherical surface would be the canonical choice for 2D or 3D resolution increase (left column of figure 3.2), a reasonable FoM is already achieved by choosing a sufficient number of points on this surfaces. They are strategically placed at distances to each other that are smaller than the wavelength. Then, due to the diffraction limit, the local intensity does not vary much between them. In the 1D-cases, two points of interest which are located on an axis are the canonical choice.

The influence of d_{ER} on the result of the optimization turned out to be only minor in the range of $\lambda/50 - \lambda/5$ which covers the whole field of expected resolutions. A value of about 100 nm was chosen for the optimization algorithm because it resulted in the best convergence of the whole pattern. With a constant polarization vector \mathbf{P}_l or \mathbf{P}_c , an isotropic coverage of all polarization directions cannot be expected. Even though depolarization occurs at high numerical apertures, most of the light beam's energy will remain in the original polarization directions. Therefore, only certain components of the intensity $I_j = |\mathbf{E}_j|^2$ with $j = x, y, z$ were regarded in the FoM. The FoM using linearly polarized light M_l and circularly polarized light M_c is given by:

$$\begin{aligned} M_l &= \min I_x(r_i \in P_{\text{ER}}) \\ M_c &= \min I_x(r_i), I_y(r_i) \quad r_i \in P_{\text{ER}}. \end{aligned} \quad (3.2)$$

The global optimization has been performed for the regimes (A) and (B). It is a combination of two methods. A Metropolis algorithm with periodically changing control parameter

which is adopted from statistical physics was used to optimize the FoM globally [70]. In between, simplex searches were inserted for the efficient search for local maxima of the FoM.

A good estimation of the global maximum can only be expected if the algorithm is performed for a large number of steps. The total number of Metropolis steps was $5 \cdot 10^5$ for each set of parameters. Every 10^4 Metropolis steps, one simplex search was inserted. The control parameter was altered periodically. Details of the global optimization algorithm are given in appendix A.3.

3.2 Optimization Results

In this section, the resulting pupil functions $\tilde{A}(r, \phi)$ that maximize the FoMs M_l and M_c are presented. The regimes (A), (B) and (C) are investigated and optimized pupil functions are found. The convergence speed of the global optimization slows down near the global optimum but relatively clear shapes could be identified after a reasonable optimization runtime. They allow to successfully idealize the pupil functions toward optimal resolution increase for all regimes. Neither of the pupil functions produces a narrow and bright doughnut distribution in the I_z intensity component. This will be addressed in section 3.6.

3.2.1 Limited Amplitude (A)

The results for the popular case (A) (the maximal amplitude of the pupil function is limited) are shown in figure 3.2 and figure 3.3 for linearly and circularly polarized light. While their symmetry is still broken at this stage of the optimization, clear shapes can be identified in the phase and amplitude distributions. The phase distributions mainly consist of two domains which have an average phase difference of π . The boundaries of these phase domains are of relatively simple shape, mainly circular or straight lines. The amplitude distribution is quite flat except at the phase domain boundaries. The average intensity amounts to 60%-70% of the maximum.

In the case of circular polarization, pupil functions for doughnut distributions along the x- and y-axis are not shown because in these cases, the algorithm converged to the pupil function of the lateral doughnut as well. However, for linear polarization, different pupil functions are observed for doughnut distributions in the x- and y-direction. E.g., simply rotating the pupil function for a doughnut on the y-axis by 90° without rotating the polarization would have resulted in a pronounced amount of focal intensity in the z-component of the electric field destroying the intensity zero condition. The phase distribution of figure 3.2(b) seems to be a suitable combination of the phase distributions of c) and d) in the same figure. For 3D doughnuts, it does not seem to make a difference whether linear or circular polarization is used. The phase and amplitude distributions for the doughnut distribution along the optic axis is very similar to the results for the 3D doughnut distribution. The latter shows more inhomogeneities which is due to a less efficient convergence caused by the higher complexity of the FoM. The resulting pupil function for a lateral doughnut intensity distribution

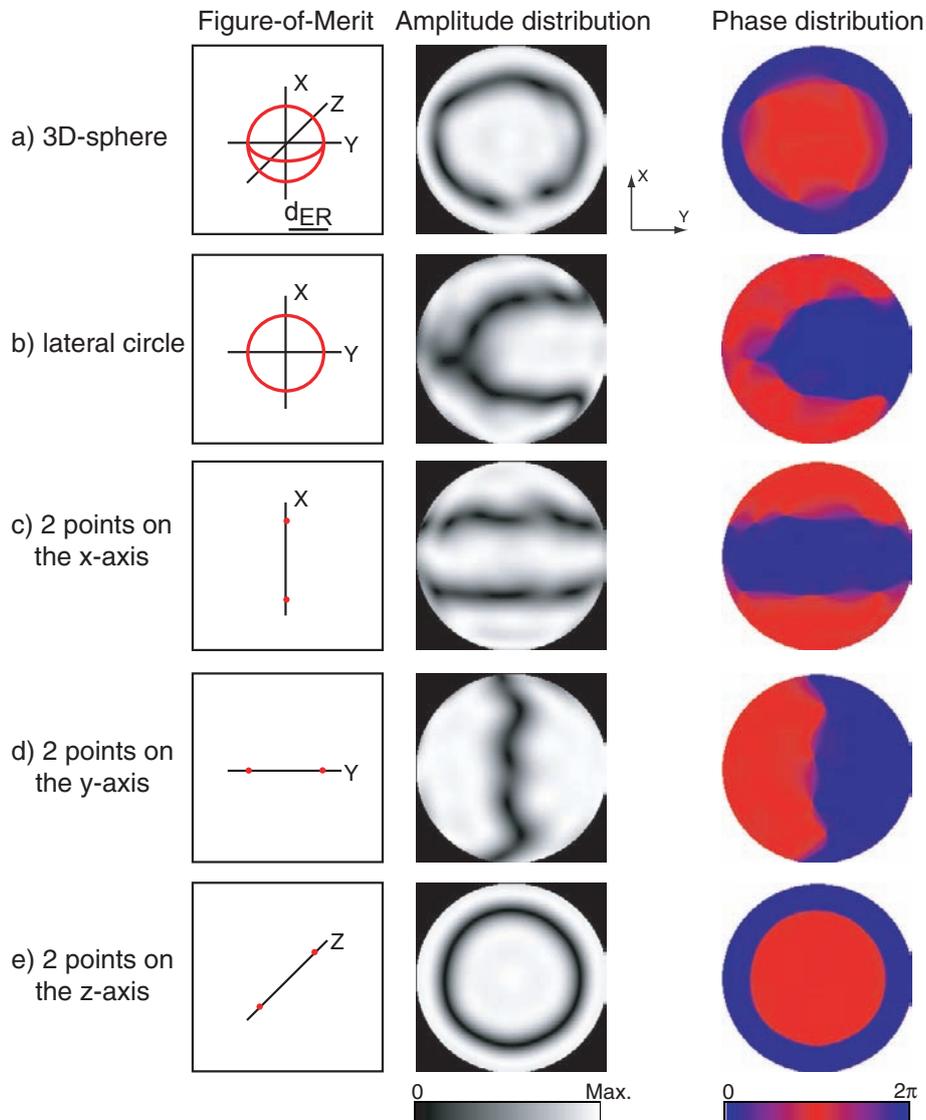


Figure 3.2: Results of the global optimization algorithm for linearly polarized light and restriction (A). In the left column, the positions of the points of interest are shown for all investigated de-excitation patterns. The center column shows the resulting amplitude distributions of $\hat{A}(r, \phi)$ over the pupil area which corresponds to the unit disk. In the right column, the phase distributions of $\hat{A}(r, \phi)$ are shown.

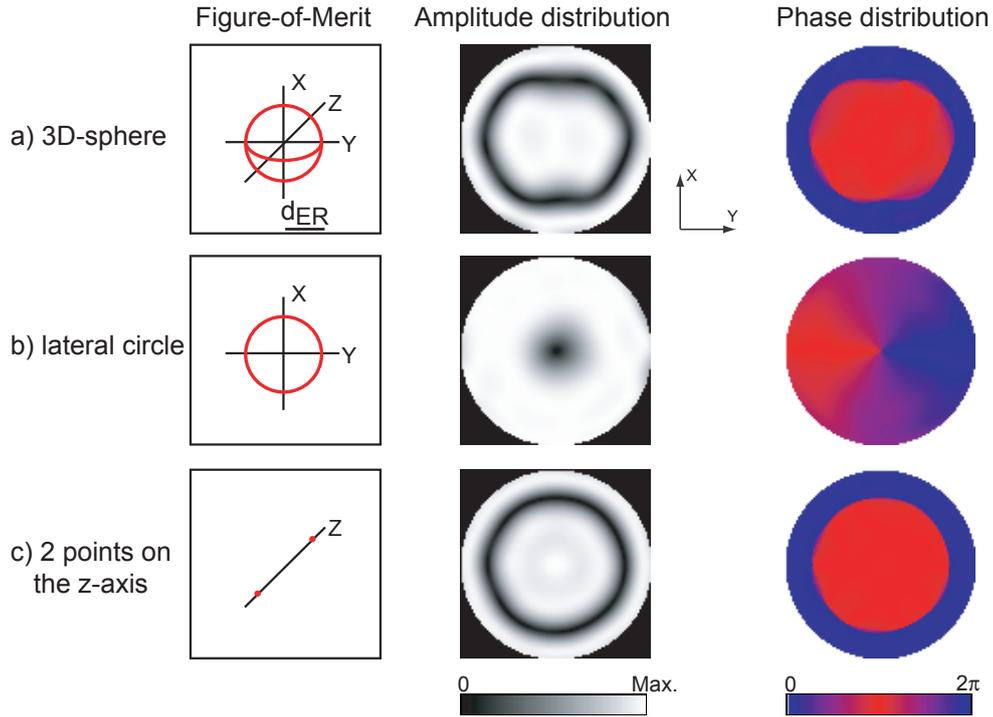


Figure 3.3: Results of the global optimization algorithm for circularly polarized light and restriction (A). Three types of different doughnut distributions are investigated. The resulting pupil functions $\tilde{A}(r, \phi)$ are shown as amplitude and phase distribution over the unit disk.

using circularly polarized light (shown in figure 3.3(b)) is exceptional. The corresponding phase distribution does not consist of two phase domains but it resembles an angular phase ramp that runs linearly from 0 to 2π . The PSFs produced by these pupil functions are shown in appendix A.5. All patterns for circularly polarized light show comparable behavior in the I_x and I_y components. Inhomogeneities and asymmetries are clearly present. Therefore, it is reasonable to assume that the algorithm has not reached the optimum yet and idealized pupil functions will likely improve the situation further. The zero amplitude regions where the phase distributions have a discontinuity are an inherent property of any smooth pupil function which is approximating a discontinuous phase distribution. The width of the valleys seems to be limited only by the order of the used polynomials. It can be assumed that the optimization algorithm converged toward the best approximation of a constant amplitude distribution. Therefore, phase-only pupil functions are used in the idealization. These maximize the transmitted power. The following simplifications of the phase distributions of figures 3.2 and 3.3 are made in the idealization:

1. Simple phase domains with only one value each (0 or π) will be used.
2. The boundaries of these domains will consist of circles, semi-circles or straight lines.
3. An exception is the lateral doughnut distribution for circularly polarized light. The

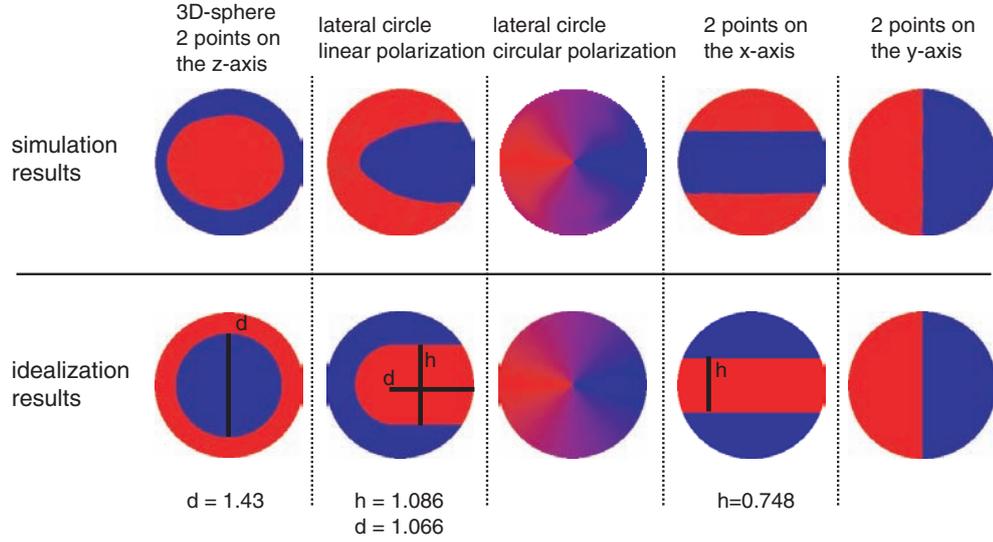


Figure 3.4: Phase distributions of the idealized phase-only pupil functions in comparison with the optimization results. The optimal values of the parameters are specified.

phase of the corresponding pupil function will increase linearly with ϕ .

4. Where applicable, the phase domains will be parametrized and the parameters will be chosen to ensure the zero and maximize the FoM.

The chosen idealized phase distributions are the obvious assignment to simple forms. The pupil functions of these idealized distributions are then given by:

$$\begin{aligned}
 \tilde{A}_{3D,cl}(r, \phi) &= \begin{cases} 1 & , r > d/2 \\ -1 & , else \end{cases} \\
 \tilde{A}_{Lat,l}(r, \phi) &= \begin{cases} 1 & , |x| > h/2 \quad \text{or} \quad y > (1-d) - \sqrt{(h/2)^2 - x^2} \\ -1 & , else \end{cases} \\
 \tilde{A}_{X,l}(r, \phi) &= \begin{cases} 1 & , |x| < h/2 \\ -1 & , else \end{cases} \\
 \tilde{A}_{Y,l}(r, \phi) &= \begin{cases} 1 & , \phi < \pi \\ -1 & , else \end{cases} \\
 \tilde{A}_{Lat,c}(r, \phi) &= \exp(i\phi)
 \end{aligned} \tag{3.3}$$

The subscript c or l describes which polarization (circular or linear) the pupil function is designed for. The optimal values of d , h and the resulting phase distributions are given in

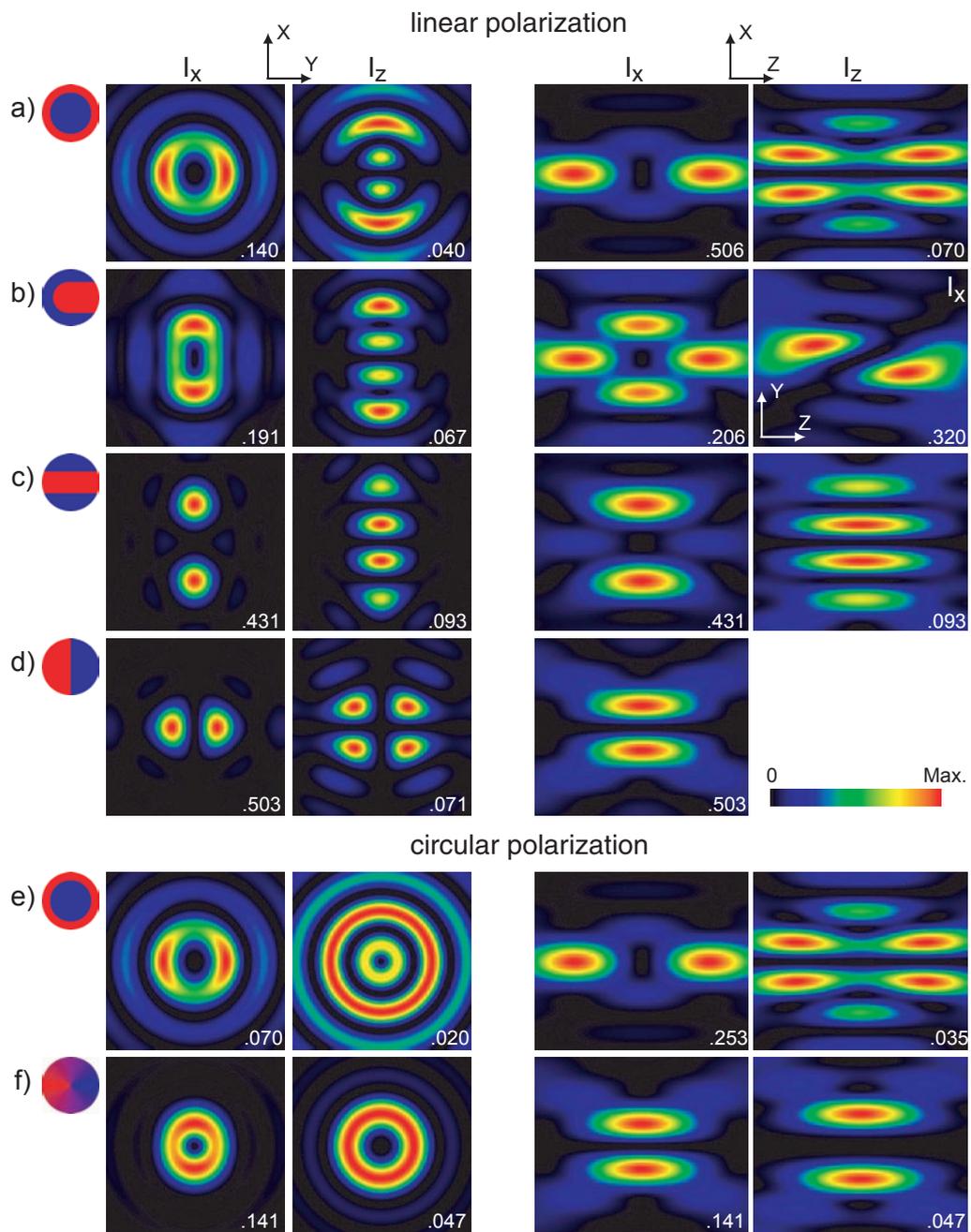


Figure 3.5: YX- and ZX- cuts of the PSFs that are produced by the phase distributions shown in figure 3.4. The I_x and I_z intensity component distributions in the most important planes are shown. The I_y component is generally very weak for linearly polarized light (with the polarization parallel to the x-axis). For circularly polarized light, the I_y component distribution is equal to the I_x component (only the image is rotated by 90°). The values in the lower right corner of each image describes the normalized maximal intensity. For this normalization, a unity pupil function would have yielded 1 as the maximal intensity at the focus.

figure 3.4. With the given values, a residual focal intensity of less than 10^{-6} for a normalized PSF has been achieved and can be reduced further taking a higher precision in the parameters into account. The PSFs that correspond to the pupil functions of equation (3.3) are shown in figure 3.5. While the homogeneity expectedly increased, a large similarity with the results of the global optimization is obvious. The corresponding FoMs have been increased through idealization as expected. The resulting relative values are given in table A.5. These support the assumption that the pupil functions \tilde{A}_i constitute the limit the optimization algorithm was converging to. Therefore, it can be stated that phase-only pupil functions must always be used if the maximal amplitude in the pupil function is limited.

For all obtained de-excitation patterns in this section, the pupil function $\tilde{A}_{Y,l}$ generates the steepest doughnut intensity distribution. However, it is restricted to the I_x component and the y-axis. The de-excitation pattern, created by $\tilde{A}_{Lat,c}$ is similarly steep if compared at normalized maximal intensity. It covers the xy-plane with both, I_x and I_y intensity components. Most other obtained de-excitation patterns are significantly less steep.

3.2.2 Limited Power (B)

The results of the global optimization for the case (B) (power in the sample is limited) are shown in figure 3.6 and figure 3.7 for linearly and circularly polarized light, respectively. The phase distributions resemble their counterparts of restriction (A) very closely and likewise, are divided into two binary phase domains. In contrast, minimal amplitude is found at the phase domain boundaries. Relatively smooth amplitude profiles were obtained with peaks that are located at the centers of their respective phase domains. In the case of the angular phase ramp, the main amplitude is located at the edge of the pupil area.

In general, it can be stated that all parts of the aperture must be illuminated to some extent in order to avoid distributing the power onto a large area in the focal plane. On the other hand, some parts of the pupil area can be overweighted. E.g., its edge which is responsible for the creation of high lateral frequencies in the PSF. The results of the global optimization exhibit similar phase behavior as for (A) and therefore the same shapes are used in the idealization (see equation (3.3)). The idealization of the amplitude distributions has several goals. By implementing the symmetries identified in the former results (radial, x-axis, y-axis) and incorporating smooth amplitude profiles, the FoM will be increased further. Furthermore, more homogeneous amplitude distributions will be easier to implement in practical applications. Functions are chosen that resemble the amplitude profiles of the global optimization assuming that the optimal amplitude profile is not too different from the optimization results. The amplitude distributions are parametrized with a small number of parameters. The analysis was performed only for the most important doughnut distributions: the 3D de-excitation pattern, the pattern along the y-axis with linearly polarized light, and the lateral doughnut with circularly polarized light. For the 3D de-excitation pattern, the following set of pupil functions has been investigated:

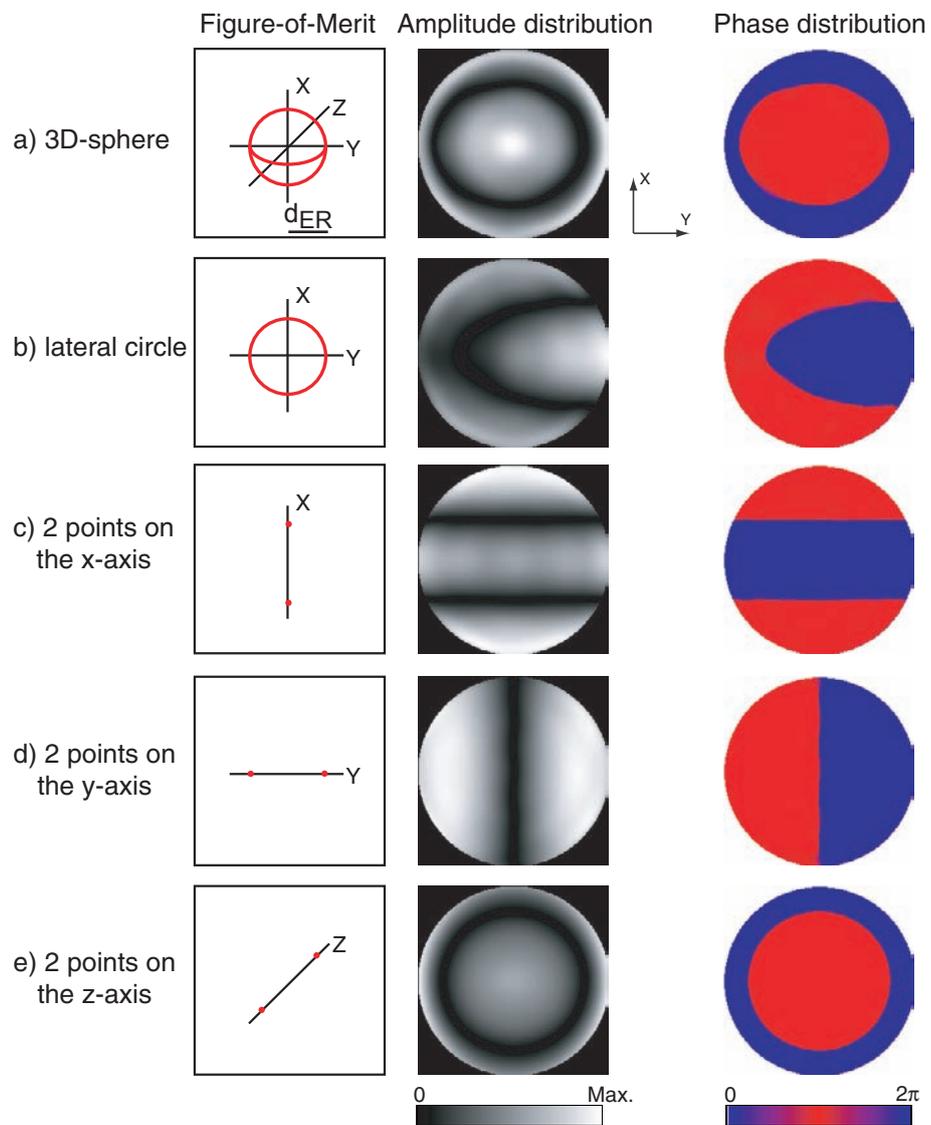


Figure 3.6: Results of the global optimization algorithm for linearly polarized light and restriction (B). The composition of the figure is analogous to figure 3.2.

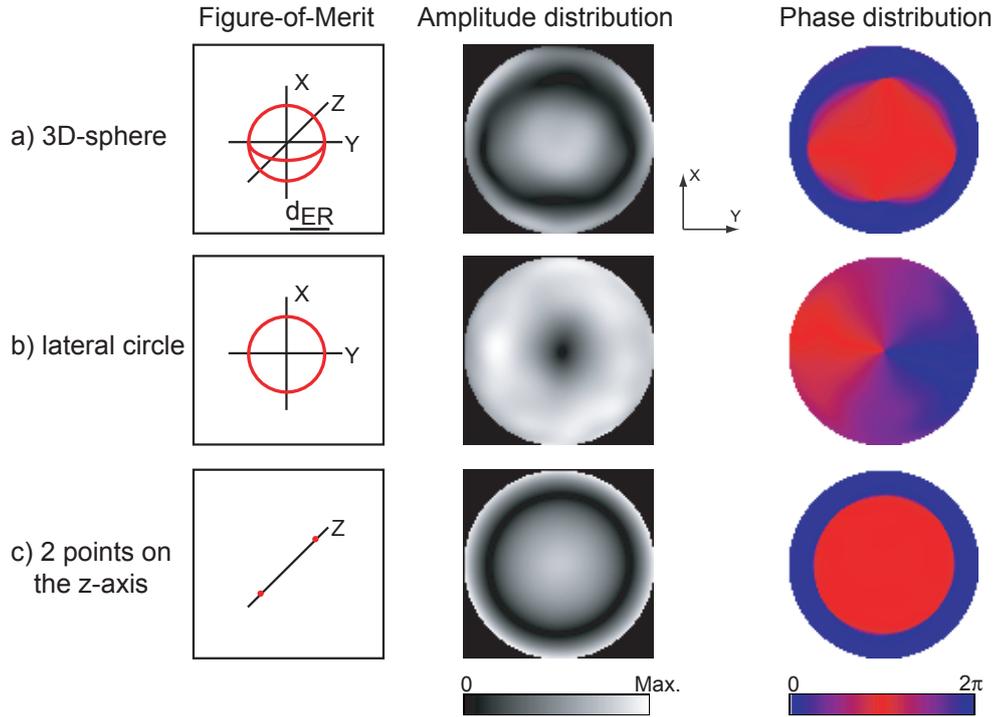


Figure 3.7: Results of the global optimization algorithm for circularly polarized light and restriction (B). The composition of the figure is analogous to figure 3.3.

$$\hat{A}_{3D,c}(r, \phi) = \begin{cases} c_\alpha (r_\alpha - r)^\alpha & r < r_\alpha \\ -c_\beta (r - r_\beta)^\beta & r \geq r_\beta \\ 0 & \text{otherwise} \end{cases} \quad (3.4)$$

with $0 < r_\alpha < r_\beta < 1$ and $\alpha, \beta > 0$. It was chosen, because the maximal amplitude of the global optimization is located at the center and the edge of the pupil area. Different slope and curvature in the area between the maxima can be achieved. Even a dark ring between the amplitude maxima is permitted to incorporate large overweighting of the center and the edge. The parameters c_α and c_β are dependent on α and β to ensure the focal intensity zero and the power normalization. The highest FoM was found for $r_\alpha = r_\beta = 0.71$ and $\alpha \sim 0.57412$, $\beta \sim 1.58125$, $c_\alpha \sim 2.1417$, $c_\beta \sim 14.3287$.

For the doughnut intensity distribution along the y-axis the investigated set of pupil functions was:

$$\hat{A}_{Y,1}(r, \phi) = \sqrt{\frac{\sqrt{\pi}\Gamma(a+1)(a+1)}{\Gamma(a+1/2)}} y^\alpha \begin{cases} 1 & \phi < \pi \\ -1 & \text{otherwise} \end{cases} \quad (3.5)$$

with parameter $\alpha \geq 0$. The positions $y = \pm 1$ possess maximal amplitude and a mirror symmetry about the x-axis has been implemented. Toward the center, variable slopes and

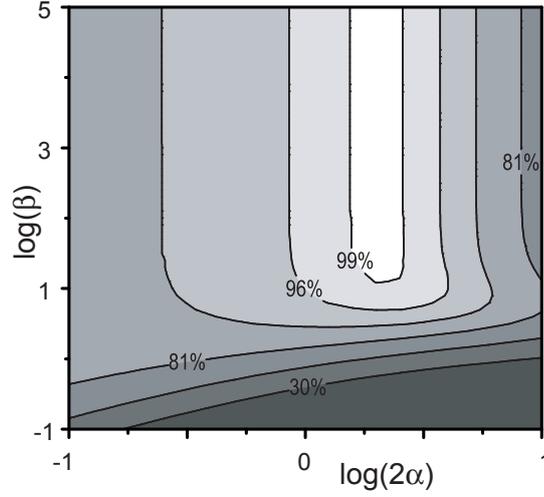


Figure 3.8: Contour plot of the FoM for parameters α and β of $\hat{A}_{\text{Lat,c}}^b$. The largest values of the FoM corresponding to bright areas in the plot are obtained for $\alpha \sim 1$ and $\beta \rightarrow \infty$.

curvatures are allowed. The highest FoM was found for $\alpha \sim 1.288$. The best α depends slightly on the chosen value of d_{ER} in the calculation of the FoM.

For the lateral doughnut intensity distribution using circularly polarized light the following sets of pupil functions were used:

$$\begin{aligned}
 \hat{A}_{\text{Lat,c}}^a(r, \phi) &= \sqrt{\frac{\alpha + 1}{\pi}} r^\alpha \\
 \hat{A}_{\text{Lat,c}}^b(r, \phi) &= \sqrt{\frac{\beta}{2\pi \mathcal{B}(2\alpha/\beta + 1, 2(\alpha + 1)/\beta)}} r^\alpha (1 - r^\beta)^{\alpha/\beta} \\
 \hat{A}_{\text{Lat,c}}^c(r, \phi) &= C e^{-\frac{(1-r)^2}{2\beta^2}} r^\alpha
 \end{aligned} \tag{3.6}$$

The first function has its maximal amplitude at the edge of the aperture and falls off toward the center with a variable slope and curvature. In order to allow for a zero amplitude at the edge too, a second function was introduced. It describes an adjustable annulus of illumination with its maximum at $r = 2^{-\beta/2}$. The third function generalizes the behavior of the first function by introducing plateau areas in the radial amplitude profile. The constant C is used for power normalization. It turned out that the third function is inferior to the others. The optimal parameters for the second function are assumed for $\beta \rightarrow \infty$ (see figure 3.8). In this limit, the second function converges to the first which was therefore chosen for further analysis. The optimal parameter α was estimated to be ~ 1.023 . For all optimized amplitude distributions, the optimal parameter values will vary to some extent with the focusing power of the lens.

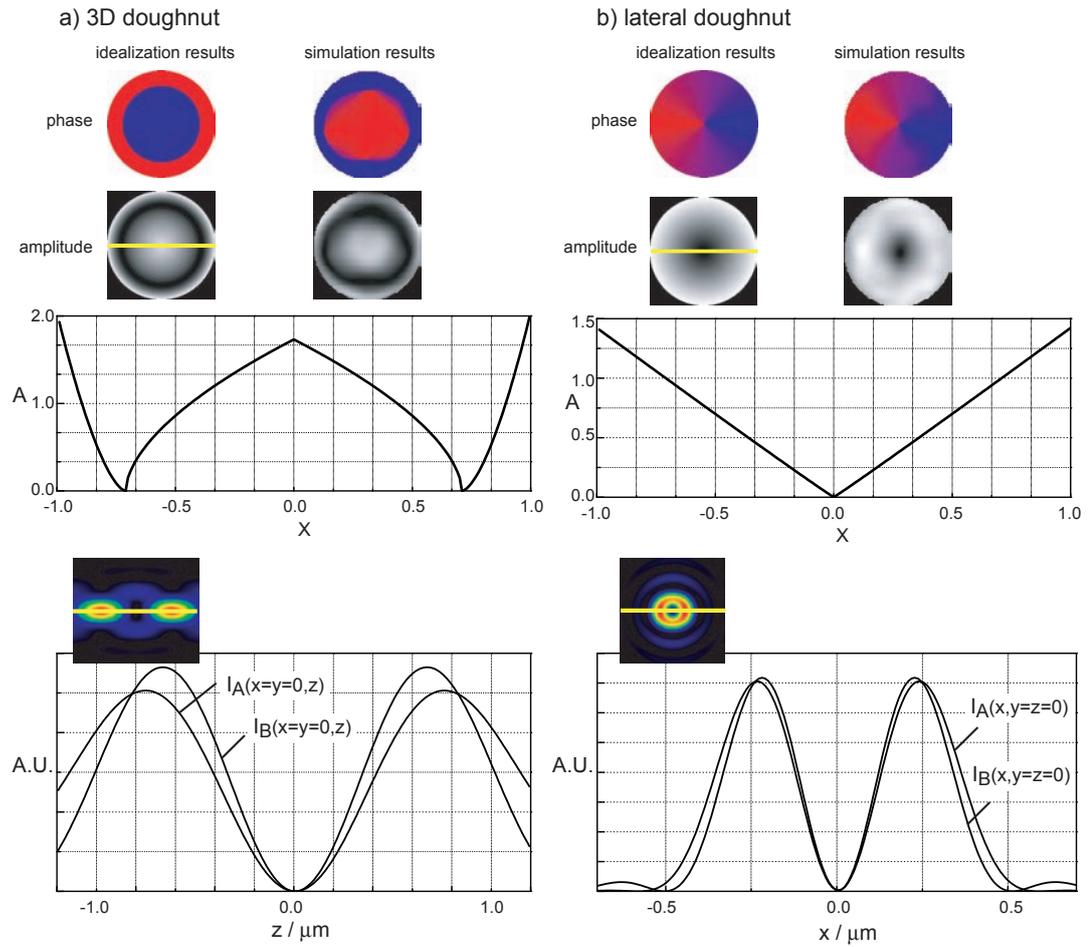


Figure 3.9: Characterization of the optimal pupil functions with restricted power in the sample. In a), the results for the 3D doughnut distribution are shown. The phase and amplitude distributions according to the idealized pupil function $\hat{A}_{3D,c}$ with parameters as in the text are shown together with the results from the global optimization. A profile through the idealized amplitude distribution is shown. In b), the results for the lateral doughnut distribution are shown. Here, the idealized pupil function $\hat{A}_{Lat,c}^a$ with parameters as in the text is shown together with its global optimization results. A profile through the amplitude distribution is also shown. The graphs at the bottom show intensity profiles through the PSF corresponding to these pupil functions, cuts of the PSF are shown in the inlays. In each graph, $I_A(x, y, z)$ is the distribution for the corresponding pupil functions of regime (A). $I_B(x, y, z)$ corresponds to the pupil functions shown above and shows improved steepness.

In figure 3.9, a comparison of the pupil functions \hat{A} and their phase-only counterparts \tilde{A} at equal power is shown for the 3D and the lateral optimized doughnut distributions. The improvements in the FoM (see table A.5) are in the range of 10% - 30%.

3.2.3 Limited focal intensity (C)

In this regime, only the maximal local intensity in the de-excitation pattern is restricted. It is known from the analysis of sharp single peaks whose maximal local intensity is limited that the optimal pupil function for this case will not utilize the whole aperture but only very small areas of it [76]. Analogously, the suitability of pupil functions, whose illumination area is restricted, is investigated here. Therefore, the FoM of modified pupil functions originating from equation (3.3) is calculated. Their amplitude is set to zero in variable areas starting from the phase domain boundaries (for $\tilde{A}_{\text{Lat},c}$ the center of the pupil area is taken). In general, the sparse illumination of the aperture sharpens spots but also creates relatively high secondary lobes and distributes energy over a large area in the sample. To normalize different de-excitation patterns to a constant maximal local intensity, a laborious and computationally expensive search in a large area around the focus seems necessary. However, the maximal intensity is always assumed at the first intensity annulus around the focus and its position is shifted toward the focus for decreased illuminated areas. Because of the symmetries of the resulting de-excitation patterns, it was sufficient for the estimation of the maximal local amplitude to examine a certain interval on some radial rays starting from the focus. A simple bracketing algorithm was able to track the maximal intensity. The investigated pupil functions were:

$$\begin{aligned}
 \bar{A}_{3D,c}(r, \phi) &= \begin{cases} C_a & r < r_a \\ -C_b & r > r_b \\ 0 & r_a \leq r \leq r_b \end{cases} \\
 \bar{A}_{Y,1}(r, \phi) &= \begin{cases} 1 & y > y_a \\ -1 & y < -y_a \\ 0 & \text{otherwise} \end{cases} \\
 \bar{A}_{\text{Lat},c}(r, \phi) &= \begin{cases} \exp(i\phi) & r > r_a \\ 0 & \text{otherwise} \end{cases}
 \end{aligned} \tag{3.7}$$

The parameters in these functions determine the extent of the illuminated area. The shape of the illuminated area was chosen so that the intensity zero was enforced. In $\bar{A}_{3D,c}$, r_a and r_b were allowed to change independently and constants C_a and C_b were chosen accordingly. The pupil functions \bar{A}_i represent de-excitation patterns that look very similar to the patterns of figure 3.5. However, less intensity in the first annulus is observed whose maximum is closer to the focus. Furthermore, higher secondary sidelobes are present. The smallest investigated illumination area has been 1% of the pupil. Within this range, the maximization of the FoM has yielded maximal values of the FoM for parameters of \bar{A}_i that are located at the borders of the possible ranges and the illuminated area is always minimized. The remaining illuminated areas correspond to the creation of very high spatial frequencies in the de-excitation patterns. Only step-like amplitude distributions have been investigated here, but they support the view that the illuminated area is generally minimized. The exclu-

sive use of the highest possible spatial frequencies to form a very steep doughnut distribution with limited maximal intensity is not unexpected. E.g., illumination of very thin rings leads to focal peaks with smallest width [76]. Therefore, the optimized pupil functions with minimum illuminated area are likely to present the optimal choice for regime (C) and they are given by:

$$\begin{aligned}\bar{A}_{3D,c}(r, \phi) &= \begin{cases} C & r < \epsilon \\ -1 & r > 1 - \epsilon \\ 0 & \text{otherwise} \end{cases} \\ \bar{A}_{Y,1}(r, \phi) &= \begin{cases} 1 & y > 1 - \epsilon \\ -1 & y < -1 + \epsilon \\ 0 & \text{otherwise} \end{cases} \\ \bar{A}_{Lat,c}(r, \phi) &= \begin{cases} \exp(i\phi) & r > 1 - \epsilon \\ 0 & \text{otherwise} \end{cases}\end{aligned}\quad (3.8)$$

with $\epsilon \rightarrow 0$ and a suitable constant C . Note however that for an illuminated area of 1% in $\bar{A}_{Lat,c}$ ($\epsilon = 0.005$) about 60 times more power is present in the sample to achieve the same FoM as with $\hat{A}_{Lat,c}^a$. Therefore in practice, one would start with pupil functions \hat{A}_i and steepen these when local intensity starts to be the limiting factor.

3.3 Ideal 3D De-Excitation Patterns

The doughnut distribution along the y-axis corresponding to $\tilde{A}_{Y,1}$ using linearly polarized light might be the pattern of choice in special applications. It features the relatively highest FoM of all obtained de-excitation patterns but it is restricted to one lateral direction and one polarization component. For general applications, a complete coverage of all spatial directions and all intensity components is needed. The results from section 3.2 seem however to indicate that a narrow de-excitation pattern for all directions and all intensity components cannot be created coherently.

More justification for this result will be given here. One reason is the complementary behavior of the I_x , I_y component and the I_z component. The vectorial dependence of the diffraction integral (2.31) is given completely by $\mathbf{B}(\theta, \phi)$ (see equation (2.31)) which is modulated with $\sin^2 \phi$ in the x-component and with $\cos \phi$ in the z-component. This inhibits the simultaneous development of a strong I_x and I_z component at the same position. E.g., for $\tilde{A}_{Lat,c}$ narrow rings are formed in the I_x and I_y component and a broader, less bright ring in the I_z component. Another reason is, that de-excitation patterns with three steep directions cannot be created coherently. An intuitive explanation for this is given in appendix A.4.

Therefore, to efficiently create a 3D de-excitation pattern, incoherent combinations of different patterns must be used. This means the incoherent application of several de-excitation

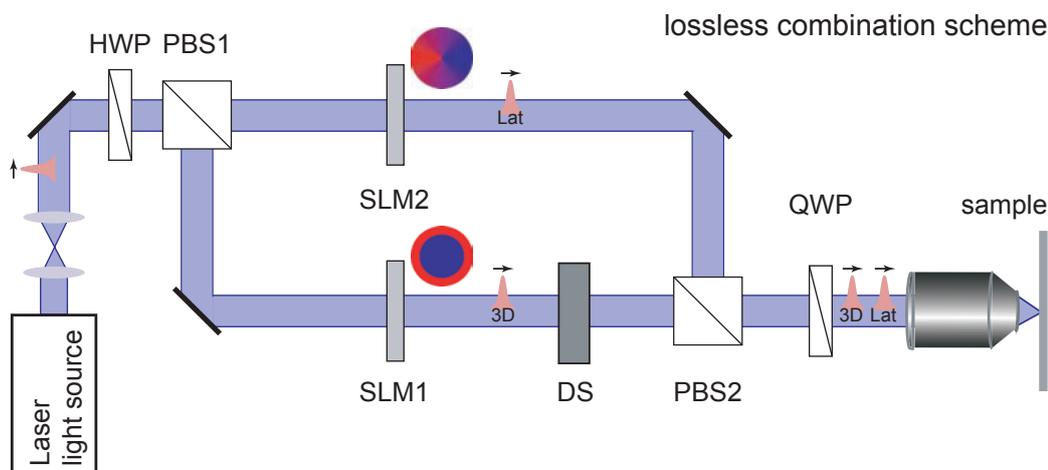


Figure 3.10: Proposed setup for the incoherent use of two doughnut distributions using circularly polarized light. A laser source provides pulsed, linearly polarized, monochromatic light. The beam is divided by a polarizing beam splitter (PBS1) in two arms. In each arm a phase and amplitude distribution can be imprinted on the beam by a spatial light modulator (SLM1 and SLM2). Here, $\tilde{A}_{3D,cl}$ and $\tilde{A}_{Lat,c}$ is used. The pulse in one arm is delayed (DS) by more than the coherence length of the light pulses. At another polarizing beam splitter (PBS 2), the two pulses are then combined losslessly. The two pulses will have orthogonal polarization after the combination. Applying a quarter wave plate (QWP) results in a left- and a right-handed circularly polarized beam. Left-handed circular polarization must be ensured for $\tilde{A}_{Lat,c}$. Both pulses are focused on the sample were they arrive consecutively. The half wave plate (HWP) is used to adjust the power distribution of the two pulses.

patterns on the sample before detections takes place. The number of involved patterns determines the experimental feasibility, e.g. by the need to adjust the intensity zeros of all participating doughnut distributions to one common spatial position. Therefore, the smallest possible number of different patterns must be used. However, the analysis of section 2.4 showed that all components of the focal de-excitation field should cover the space near the focus to efficiently de-excite excited molecules. Here, the combination of $\tilde{A}_{Lat,c}$ and $\tilde{A}_{3D,cl}$ and the use of circularly polarized light is proposed. It has the advantage to efficiently cover the whole area around the focus in two intensity components with only two different de-excitation patterns, a number not reached by any other suitable combination. Combining the optimal de-excitation pattern in the lateral and in the axial direction makes optimal use of the available patterns. For the case of combining two patterns, this combination allows the highest 3D resolution. Depending on the present conditions, the total power must be optimally distributed on both de-excitation patterns. A setup that can efficiently realize this combination is proposed in figure 3.10. Working with only two de-excitation patterns allows the lossless combination of light beams with perpendicular polarization with a polarizing beam splitter. The necessary circular polarization is then created with a quarter wave plate. The combination of more de-excitation patterns would likely need an efficient combination of many light beams. The use of dichroic mirrors and de-excitation light beams of different wavelengths could be beneficial then, provided that the RESOLFT transition possesses a

sufficiently broad spectral width.

4Pi-microscopy enlarges the aperture and features axial resolution improvement. The increased focusing capabilities of 4Pi-microscopes can also be exploited in the RESOLFT concept and increased resolution in the axial direction can be expected. This situation is briefly regarded in appendix A.6 and some useful doughnut distributions are given.

3.4 Highest Possible Resolution Increase

In general applications, the best resolution is delivered by focal spots featuring the smallest focal volumes and the highest extents of their OTF. In section 2.2, these two resolution measures have been defined and justified. They are calculated according to equations (2.3), (2.4), and (2.6) and will be used to estimate the achievable resolution of the ideal 3D de-excitation pattern that has been introduced in the last section. The resolution of the incoherent combination of patterns shall be investigated in comparison to each of its de-excitation patterns, corresponding to $\tilde{A}_{\text{Lat,c}}$ and $\tilde{A}_{\text{3D,cl}}$. Circularly polarized light is used for both, excitation and de-excitation. Furthermore, the pulsed RESOLFT model from section 2.3 is implemented and the effective PSFs are calculated according to section 2.4. In the case of an incoherent combination of several de-excitation patterns, the total de-excitation effect must be calculated as the product of de-excitation effects for each pattern of the combination. The effective PSFs are computed according to equation (2.20) for two different conditions (R_{Slow}) and (R_{Fast}). Under condition (R_{Slow}), molecules are assumed to remain stationary and for (R_{Fast}), molecules can rotate freely between excitation and de-excitation. We assume a constant isotropic orientational distribution of the molecular dipole moments.

The resolution measures have been calculated in dependence of the normalized saturation factor F_{Sat} which is defined by:

$$F_{\text{Sat}} = \xi/g, \quad (3.9)$$

with the saturation factor ξ and a normalization factor g which is defined as the ratio of the maximal intensities of the de-excitation pattern and of an unaberrated PSF at equal power. Therefore, the normalized saturation factor of a de-excitation patterns is determined by its power and the meaning of the value of F_{Sat} is that an unaberrated PSF ($g = 1$) with the same power would feature such a ξ at its focus. Initially, the optimal power distribution into the de-excitation patterns of the incoherent combination has been investigated. Simulations have revealed that for optimal resolution increase $\sim 70\%$ of the available power must be distributed in the de-excitation pattern corresponding to $\tilde{A}_{\text{Lat,c}}$ and $\sim 30\%$ in the other pattern. This result is largely independent of the applied saturation factor (for $F_{\text{Sat}} > 50$) and the assumed rotational regime and can therefore be seen as an universal choice for the optimal incoherent combination of the given patterns.

Figure 3.11 shows the possible resolution increases relative to a confocal microscope. For the increase in the OTF extent, a maximal resolution of 24 times the confocal value is reached with the incoherent combination, and for the inverse focal volume, a maximal factor of about 32 is reached. Indeed, the incoherent combination is superior to its single patterns

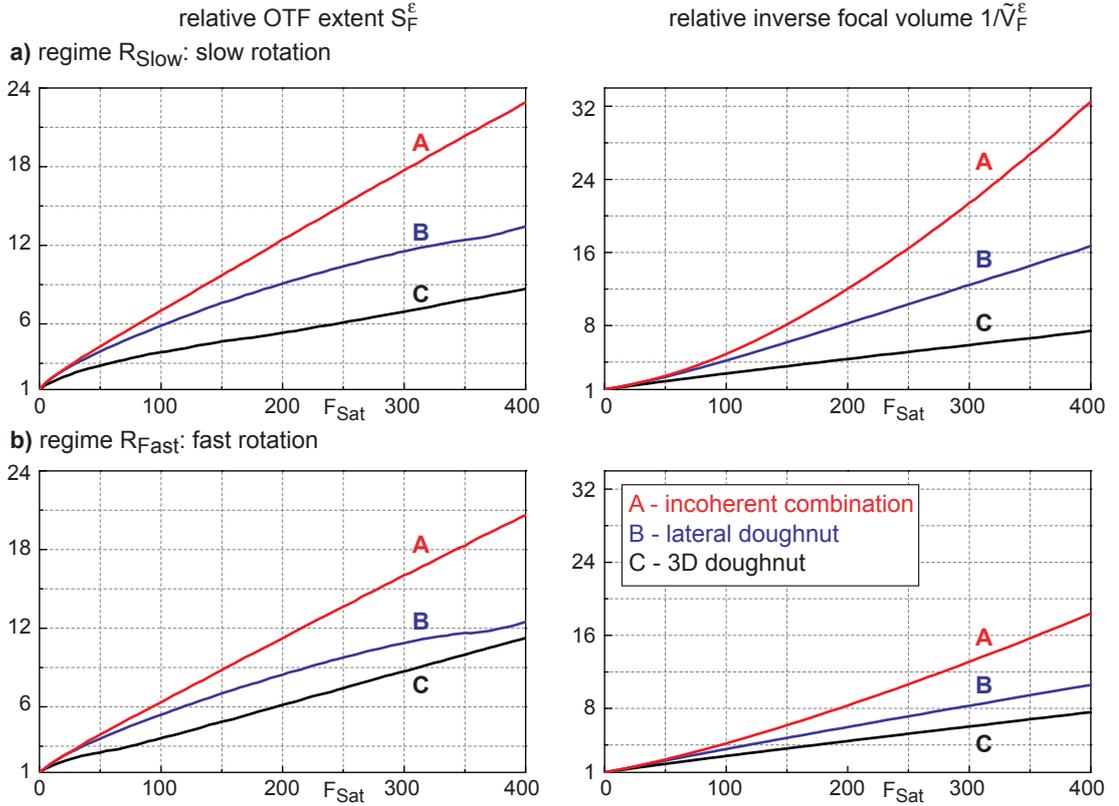


Figure 3.11: Resolution increase as a function of the normalized saturation factor F_{Sat} . The increase of the OTF extent relative to its confocal value is shown in the left column. A very small cut-off ϵ was chosen which was 6 orders of magnitude lower than the zero frequency modulus (at $F_{\text{Sat}} = 0$) but still 1-2 orders above the numerical noise level. The right column shows the relative increase in the inverse focal volume for parameter $\epsilon = 0.05$. In a), the slow rotation regime is assumed (R_{Slow}) and in b), the fast rotation regime is assumed (R_{Fast}). The de-excitation pattern A corresponds to the optimal incoherent combination which consists of a relative amount of 70% of pattern B and 30% of pattern C. The corresponding pupil function for pattern B is $\tilde{A}_{\text{Lat,c}}$ and for C it is $\tilde{A}_{\text{3D,c1}}$.

for all cases. The single lateral de-excitation pattern performs better than the single 3D de-excitation pattern. At large saturation factors, the curves seem to converge toward a simple power law behavior. The asymptotic rise of the resolution measures can then be described by a single exponent. As expected, the incoherent combination shows significantly larger exponents than the single patterns which is a direct consequence of the optimal coverage of all spatial directions with steep de-excitation patterns. In general, the inverse focal volume has been increased further than the OTF extent at equal F_{Sat} . A reason could be the higher sensitivity of the OTF extent on tails in h_{EFF} which increasingly determine the whole shape of h_{EFF} at larger F_{Sat} . Furthermore, it can be seen that the regime (R_{Fast}) results in less a increased resolution. Fast rotational diffusion equalizes the orientational distribution of excited molecules already before de-excitation. Consequently, a higher fraction of molecules will then be

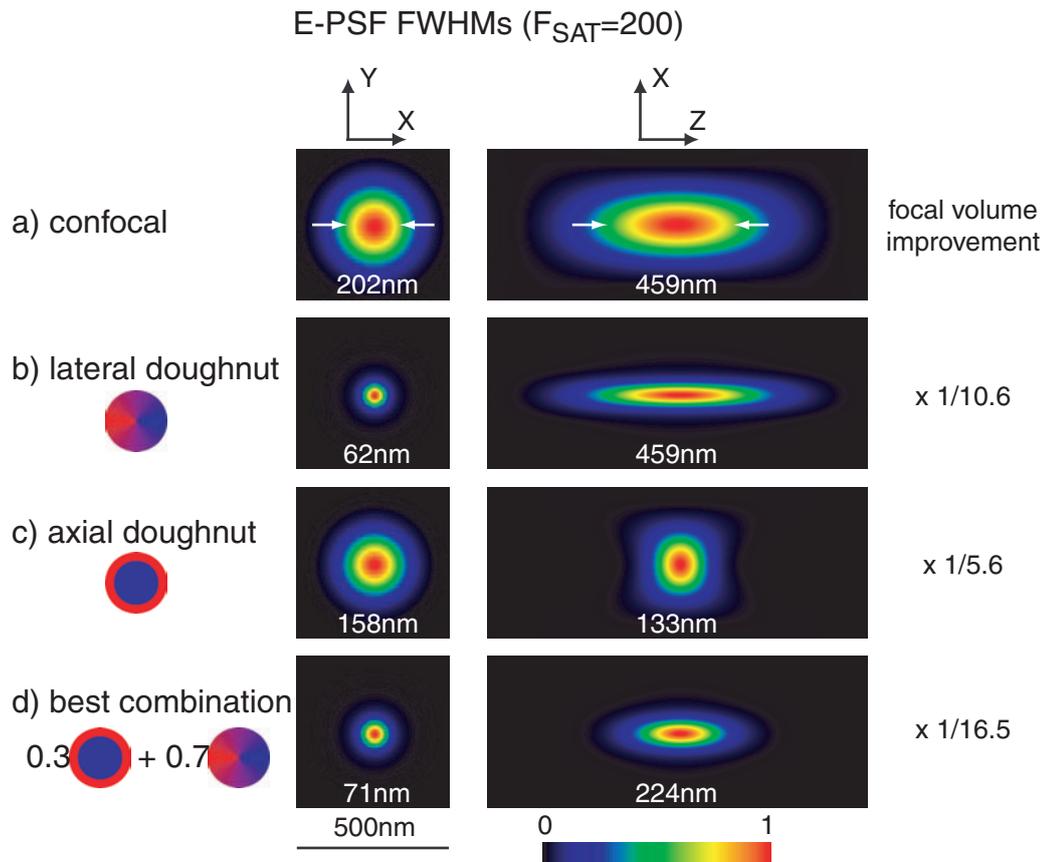


Figure 3.12: Effective PSF xy - and xz -cuts for a) a confocal PSF, b) a RESOLFT-PSF with the lateral de-excitation pattern, c) a RESOLFT-PSF with the axial de-excitation pattern, and d) the optimal incoherent combination. The expected widths are given for our standard conditions which are defined in the list of abbreviations. $F_{\text{Sat}} = 200$ was chosen. The reduction in the focal volume (here product of widths) is also calculated.

z -oriented and, therefore, less efficiently de-excited. Especially at large saturation factors, z -oriented molecules which are not efficiently de-excited here will create pronounced tails in h_{Eff} deteriorating the possible resolution increase. To overcome this, de-excitation patterns for z -oriented molecules are introduced in section 3.6.

The shapes of some effective PSFs are shown in figure 3.12 for $F_{\text{Sat}} = 200$. The lateral de-excitation pattern squeezes the main spot of h_{Eff} solely in the lateral directions which generates a large asymmetry in the resulting spot. However, it features the smallest possible widths. The axial/3D de-excitation pattern produces a nearly isotropical spot but at the cost of a comparably small overall resolution increase. The incoherent combination does neither feature the smallest widths nor an isotropical spot but exhibits the largest reduction of the focal volume. At $F_{\text{Sat}} = 200$, the highest possible reduction in the focal volume is about 16 times for the incoherent combination.

3.5 Experiments

The imaging of synaptic vesicles requires a very high resolution especially in the focal plane because mainly the distribution of vesicles near the presynaptic membrane and the membrane itself is of interest. For this purpose, experiments using STED microscopes seemed promising because a high resolution in one direction was already achieved [36]. In this case, the successful implementation required a very steep de-excitation pattern in the xy -directions. Randomly oriented molecules had to be de-excited efficiently. Furthermore, the available laser power was limited. The optimized de-excitation pattern corresponding to $\tilde{A}_{\text{Lat},c}$ (see equation (3.3)) which was found in this thesis represented the best choice for this case and was therefore employed.

The setup shown in figure 3.13¹ describes a STED microscope that can generate the laterally optimal de-excitation pattern. It featured a numerical aperture of 1.4. A green-emitting dye was used which could be excited at 470 nm and de-excited at 615 nm. The realization of

¹Figures 3.13, 3.14 and 3.15 slightly altered and reprinted with kind permission.

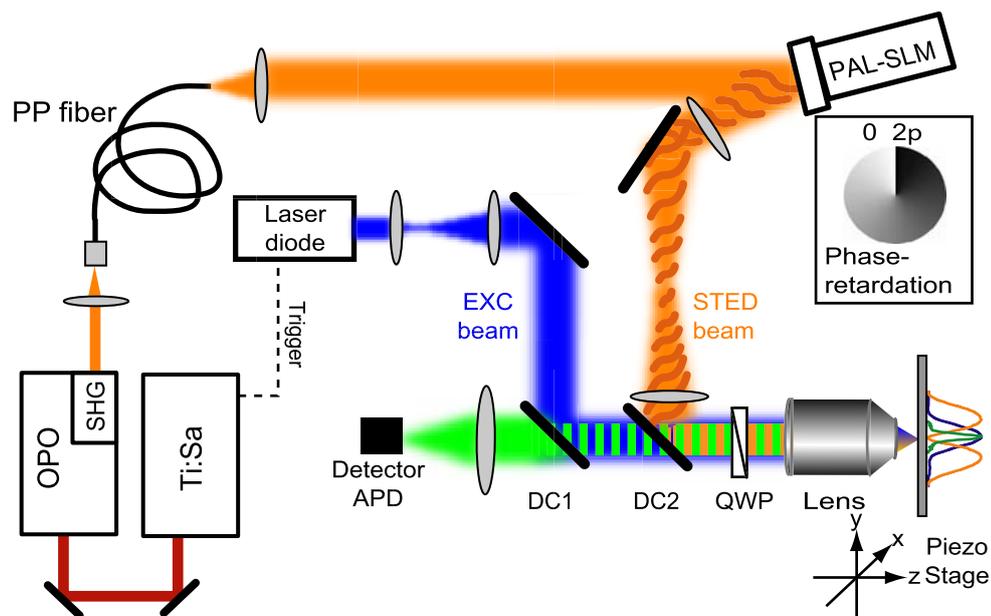


Figure 3.13: STED microscope. Fluorescence excitation was accomplished by a pulsed laser diode. Tunable STED light was provided by an optical parametric oscillator pumped by a mode-locked Ti:Sapphire laser. The pulses delivered from the OPO were stretched to a duration of about 150-250 ps with a polarization preserving fiber. The wavefront conversion into a doughnut shaped beam was accomplished by means of a spatial light modulator (PAL-SLM) which was imaged onto the back aperture of the objective lens. Both, the excitation (Exc) and the STED beam were coupled into the objective lens using dichroic beam splitters ($DC_{1,2}$). The polarization was turned from linearly polarized light into circularly polarized light by an achromatic quarterwave plate (QWP). The fluorescence signal was collected by the same lens and detected by an avalanche photodiode (APD).

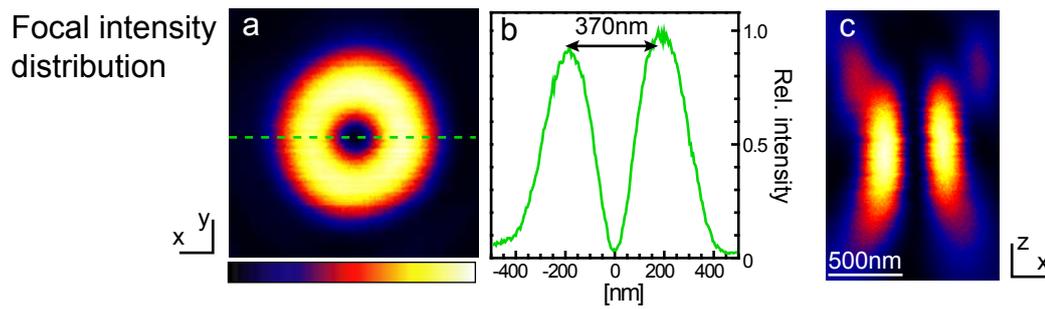


Figure 3.14: Verification of the focal de-excitation pattern by imaging of a small gold bead with a diameter of about 40 nm. a) Doughnut shape intensity distribution within the focal plane and b) its central profile. c) Focal intensity distribution along the optic axis.

the phase-only pupil function $\tilde{A}_{\text{Lat},c}$ was accomplished by the use of a SLM which has the advantage of being able to control the wavefront by a large number of free parameters. A measurement of the resulting focal doughnut intensity distribution is shown in figure 3.14. It resembles nicely the pattern which is expected theoretically but a residual amount of intensity in the focal minimum was found. It amounts to 1%-4% of the maximal intensity on the doughnut rim. Also, the doughnut's annulus is slightly inhomogeneous. Aberrations in the optical setup caused the axial shift of the maxima of both peaks in figure 3.14(c). Commonly, the illumination on the back aperture is falling off slightly in the radial direction under experimental conditions which results in a somewhat broader de-excitation distribution compared to the patterns shown in this dissertation but it should not have detrimental effects on the purity of the focal intensity zero.

To compare the resolution with the confocal microscope, a fluorescence spot was measured. The image of fluorescent beads with a diameter of 40 nm yielded a FWHM of 195 nm. In the STED microscope a FWHM of about 66 nm was measured, corresponding to a ninefold reduction in the effective focal area. The images of these small objects were thinner at the top than comparable gaussians, implying that the resolution would be better than a gaussian of the same width. Simulations revealed that given the encountered noise conditions, the STED microscope would allow the separation of point objects 45 nm apart in the focal plane.

With this formerly unmatched lateral resolution synaptic vesicles within nerve terminals could be resolved and it was possible to differentiate between alternative recycling processes of synaptic vesicle proteins: Either vesicle proteins could be spread to some extent across the presynaptic membrane upon fusion and sorting of these proteins would have to occur to regenerate vesicles or vesicles constituents could remain together after exocytosis. It could be shown that the protein synaptotagmin I from the vesicle membrane remains clustered in isolated patches on the presynaptic membrane [77]. A comparison of a confocal image and a STED image with superior resolution is shown in figure 3.15.

Recently, using a very similar system the lateral resolution could be pushed even further [58]. With higher power and avoiding bleaching an impressive FWHM of the PSF of 15-20 nm has been achieved.

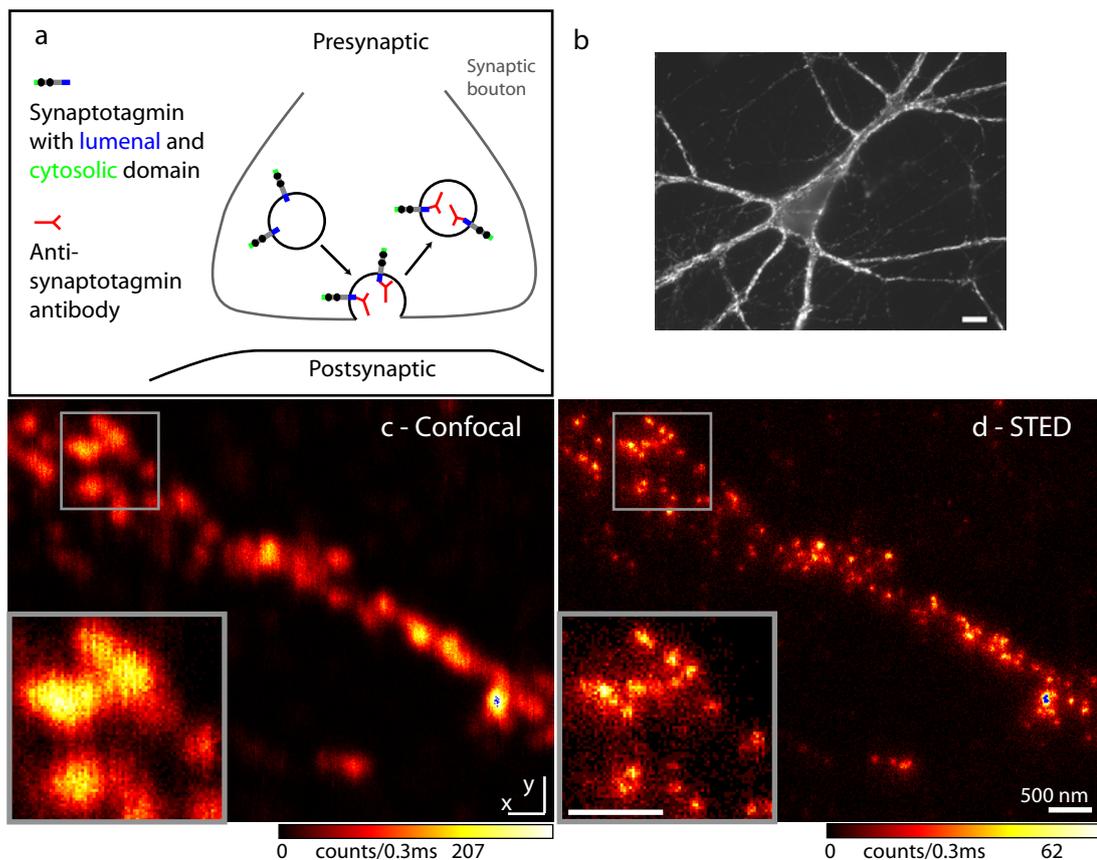


Figure 3.15: Resolution of synaptic vesicles of cultured neurons by STED microscopy. a) Synaptotagmin domains of synaptic vesicles can be labeled by specific antibodies at exocytosis; upon endocytosis the labels are internalized. b) Typical fluorescence wide field image of a neuron. The scale bar is $10 \mu\text{m}$. The same preparation was imaged by confocal c) and STED d) microscopy using a lateral doughnut distribution. The resolution is significantly better in the STED image, allowing the separation of individual vesicles.

The experiments show, that the laterally optimal de-excitation pattern can be implemented successfully in practical applications. However, differences between the theoretical and the experimental de-excitation patterns are present. The inhomogeneity of the first intensity annulus results in an inhomogeneous squeezing of the focal spot which has only minor effects on the obtainable resolution. More severe is the remaining intensity in the focal minimum which reduces the overall fluorescence signal.

3.6 Z-Oriented Molecules

The creation of optimal de-excitation patterns in the I_z intensity component has not been tackled in this work until now. They would allow the efficient de-excitation of axially oriented molecules. In section 3.3, it was seen that an efficient de-excitation pattern for all

polarization directions cannot be created coherently. But the use of circularly polarized light leads to de-excitation patterns with I_x and I_y intensity components and would therefore already give a largely orientation independent quenching effect. The incoherent combination of two de-excitation patterns that has been introduced in section 3.3 uses circularly polarized light and does not feature significant I_z components. Indeed, at currently feasible saturation factors, the contribution of inefficiently de-excited z-oriented molecules to the effective PSF's shape is small, but for larger saturation factors (> 100), these molecules will eventually result in pronounced broad tails outweighing the sharp central peak in the effective PSF. To overcome this, optimal de-excitation patterns for z-oriented molecules could be added to incoherent combinations.

The use of radially polarized light seems promising because it efficiently creates a prominent I_z component with high NA objective lenses [78]. It turned out, that the phase distributions from section 3.2.1 together with radially polarized light yield efficient de-excitation patterns whose quality is comparable with the results of section 3.2.1 for circularly polarized light. Using the phase distribution of $\tilde{A}_{\text{Lat},c}$ together with radially polarized light results in a steep lateral I_z component doughnut distribution but unfortunately, there are also nonzero I_x and I_y components present at the focus. The intensity zero at the focus can be efficiently restored by generating a focal field which counteracts these focal I_x and I_y components while featuring a minimal I_z component (to avoid distortions of the steep I_z component distribution). Using tangentially polarized light does not produce any I_z component at all as can be seen by evaluating equation (2.33). Therefore, the use of tangentially polarized light is proposed for the second focal field. The amplitude of the pupil function for the second focal field must be adjusted so that the sum of both fields gives a focal zero. Therefore, we define the vectorial pupil function $\mathbf{A}_{\text{Lat},z}$ by:

$$\mathbf{A}_{\text{Lat},z}(r, \phi) = \left[f(r) \begin{pmatrix} \cos \phi \\ \sin \phi \end{pmatrix} + ig(r) \begin{pmatrix} \sin \phi \\ -\cos \phi \end{pmatrix} \right] \exp(i\phi), \quad (3.10)$$

with suitable radial functions $f(r)$ and $g(r)$. A simple choice for these functions which gives a focal zero already for every contribution of two opposing points on the pupil area is found by:

$$\mathbf{A}_{\text{Lat},z}^a : f(r) = 1, g(r) = \sqrt{1 - r^2 \sin^2 \alpha} \quad (3.11)$$

which shows a variable amplitude of the tangentially polarized part of the pupil function. This pupil function will result in a lateral de-excitation pattern mainly in the I_z component. However, constant amplitude would simplify the preparation of these doughnut distributions. Therefore, another pupil function is defined by:

$$\mathbf{A}_{\text{Lat},z}^b : f(r) = 1, g(r) = C, \quad (3.12)$$

with a suitable constant C . The constant C is determined by inserting $\mathbf{A}_{\text{Lat},z}^b$ into equation (2.33) and solving the resulting equation for vanishing focal intensity. C is determined by:

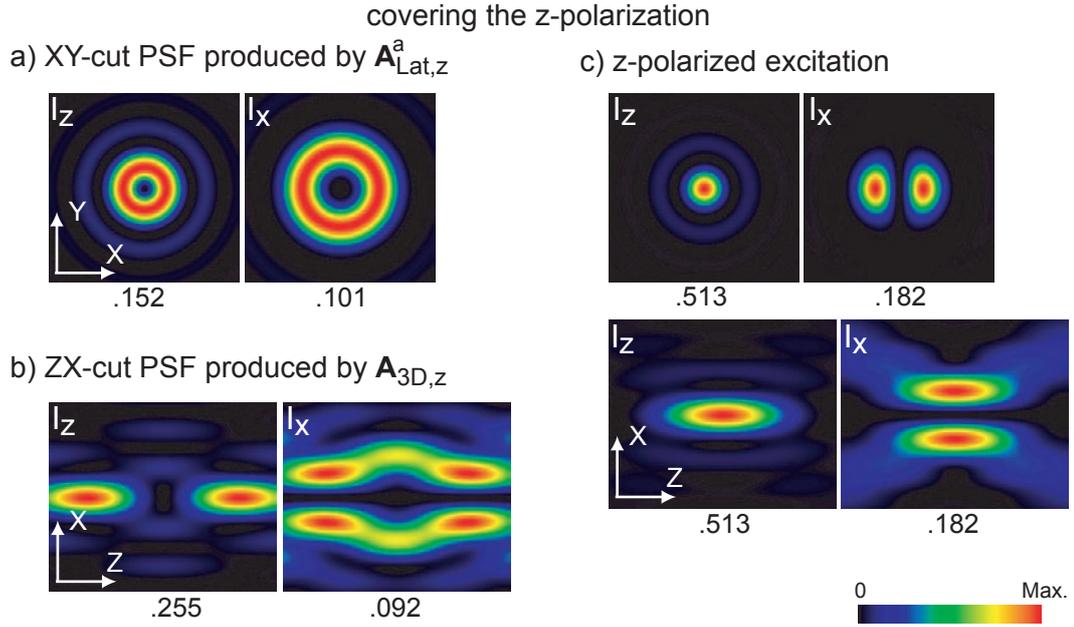


Figure 3.16: Proposed de-excitation patterns for z-polarized RESOLFT microscopy. In a), a XY-cut of the pattern produced by $\mathbf{A}_{\text{Lat},z}^a$ is shown. Its narrowness is comparable with figure 3.5(f). The ZX-cut produced by $\mathbf{A}_{3D,z}$ is shown in b). It features a large I_z component on the optic axis and can be compared to figure 3.5(a). In (c), illumination with radially polarized light is assumed. The resulting large I_z component is maximal at the focus and can be used for the efficient excitation of z-oriented molecules.

$$C = \int_0^\alpha (\cos \theta)^{3/2} d\theta / \int_0^\alpha (\cos \theta)^{1/2} d\theta. \quad (3.13)$$

And for $\sin \alpha = 0.9$ one obtains $C \sim 0.8235$. The polarization of the presented vectorial pupil functions is equivalent to an elliptical polarization with variable principal axes.

We want to create a steep de-excitation pattern in the I_z component also along the optic axis. Therefore, the phase distribution of the pupil function $\tilde{\mathbf{A}}_{3D,cl}$ was combined with with radially polarized light. The resulting vectorial pupil function is given by:

$$\mathbf{A}_{3D,z}(r, \phi) = \begin{pmatrix} \cos \phi \\ \sin \phi \end{pmatrix} \begin{cases} 1 & , r > d/2 \\ -1 & , else \end{cases}. \quad (3.14)$$

The constant d is determined to provide an focal intensity zero. For $\sin(\alpha) = 0.9$ one obtains $d \sim 1.6206$. The resulting doughnut distributions are shown in figure 3.16 together with a PSF that efficiently excites axially oriented molecules [78].

These results show that the RESOLFT concept can readily be applied to axially oriented molecules. The steepness of the doughnut distributions in the I_z components is similar to the doughnut distributions in the I_x and I_y components ($\sin \alpha = 0.9$). Hence, even if the presented pupil functions here are not the results of a direct optimization they are likely to present a very good estimate of the optimal doughnut distributions for the I_z component.

A systematic exploration of possible functions $f(r)$ and $g(r)$ in equation (3.10) could improve the FoM further. The combined, incoherent use of these de-excitation patterns with the combination of section 3.3 will be necessary at large saturation factors. A prominent difference to the former results is the polarization of the pupil distribution. The realization of RESOLFT microscopy with axially oriented molecules relies therefore on the accurate creation of radially and tangentially polarized light [79].

4 Multifocal RESOLFT Microscopy

4.1 Fast Scanning

So far, STED microscopes have yielded the highest achievable spatial resolution in far field fluorescence microscopy. While the size of the focal spot could be squeezed substantially within the last years, the time necessary to acquire such highly resolved images has increased due to two main reasons: The reduced focal volume leads to less fluorescence photons per scanning position and to keep a sufficient SNR, the pixel dwelltime has to be increased accordingly. Furthermore, more scanning steps are needed to cover the whole observation volume.

However, speed is a critical parameter in many real-world applications. For example, transport processes can be investigated in real-time by microscopy [81]. An overview about the imaging speed of different types of fluorescence far field microscopes is given in table 4.1.

In any scanning microscope, there are two complementary ways to increase the image acquisition speed:

1. Increasing the scanning speed or
2. Parallelizing the imaging process.

The maximal scanning speed is primarily limited by the smallest possible pixel dwell-times whose lower bound is ultimately determined by the maximal feasible excitation level. This is in turn limited by the onset of photobleaching or saturation with modern microscopes already operating at this limit. Parallelization must be used to increase the imaging speed further. The types in table 4.1 that are marked by * include some kind of parallelization. In parallelized microscopy, the illuminated volume of the sample is increased by generating more than one focus per scanning step, therefore increasing the photon flux from the sample. However, to keep a resolution comparable to the confocal, the sample may be illuminated only partly which in turn makes some sort of scanning necessary. For example, traditional real-time 3D microscopy is accomplished by disk scanning confocal microscopy with the use of an array of pinholes arranged in a Nipkow-type pattern [82] which creates many foci simultaneously in the sample. Another realization with pulsed two-photon microscopy is the multifocal multi-photon microscopy (MMM) [83]. The focal spots are commonly well separated by many wavelengths.

Another method, structured illumination microscopy [15], is a concept that features an even higher degree of parallelization while using the focusing power of a lens in only one

<i>type of microscope</i>	<i>lateral resolution</i>	<i>axial resolution</i>	<i>framerate</i>
<i>linear response</i>			
widefield*	$< \lambda/2$	—	real-time
patterned illumination*	$< \lambda/2$	$\sim \lambda$	fast
confocal	$< \lambda/2$	$\sim \lambda$	slow
direct view confocal*	$< \lambda/2$	$\sim \lambda$	fast
4Pi	$< \lambda/2$	$\lambda/7$	slow
MMM-4Pi*	$< \lambda/2$	$\lambda/7$	fast
<i>nonlinear response</i>			
saturated pattern illumination*	$\lambda/10$	—	moderate
STED	$\lambda/25$	$\lambda/5$	slow
STED-4Pi	$\lambda/4$	$\lambda/23$	slow

Table 4.1: Spatial resolution and framerate values for different types of far field fluorescence microscopy. For nonlinear responses the highest achieved resolutions are presented. Widefield microscopy gives real-time framerates and all types of microscopy featuring any kind of parallelization (*) show increased framerates in comparison to single spot scanning microscopes. Saturated pattern illumination [80] involves considerably more scanning steps than direct view confocal microscopy and has therefore a less increased framerate.

direction at a time. However, it relies on accurate mathematical post-processing to obtain an image.

Our approach is based on the generation of arrays of foci similar to MMM. We want to present an array of doughnut distribution foci which, if combined with a corresponding array of excitation foci, can be used to realize fast RESOLFT microscopy. We apply a coherent pattern consisting of many (4 - 100) doughnut intensity distributions. We want to exploit the crosstalk that occurs in the coherent creation of a foci array to improve the quality of the foci pattern. E.g., the illumination power for a given FoM could be reduced by mutual sharing of doughnut structures of neighboring foci. In this situation, higher saturation factors could be achieved in the case that the power that the sample can sustain is limited. A very narrow pattern of foci is aspired here. However, the desired average distance between two neighboring foci shall not be much smaller than the size of the Airy disk for two reasons: the intensity pattern of foci will be diffraction limited itself, so the intensity between closely neighbored foci could not be raised efficiently and second, in the imaging process, many spatial frequencies will not be transmitted - the detection will not resolve neighbored excitation peaks. In the ideal case, the steepness of a single doughnut distribution will be maintained in the case of the parallelized doughnut pattern.

Under incoherent conditions, the application of arrays of many foci was realized by time-multiplexed MMM where each foci generated by a microlens has an individual time-shift [84]. Therefore, crosstalk between neighboring foci is eliminated. While the application of time-multiplexed arrays of de-excitation distributions in RESOLFT microscopes would seem promising too, the preparation of many time-multiplexed doughnut intensity distributions, e.g. with microlens-arrays, will certainly pose its own special problems during the practical

realization. This method is not pursued in this work.

4.2 Multifocal Arrangements

The creation of a pattern of doughnut distributions in the focal plane requires the generation of a relatively complex pupil function. In general, not only the phase but also the amplitude of the pupil function must be variable. In figure 4.1(a), a scheme is shown that produces two foci in the focal plane by two incident plane waves which have different propagation directions. An equivalent approach is to generate the complex pupil function that corresponds to the reunion of the two plane waves at the pupil plane (shown in figure 4.1(b)). We extend this idea to create a whole pattern of several doughnut distribution foci in the focal plane by

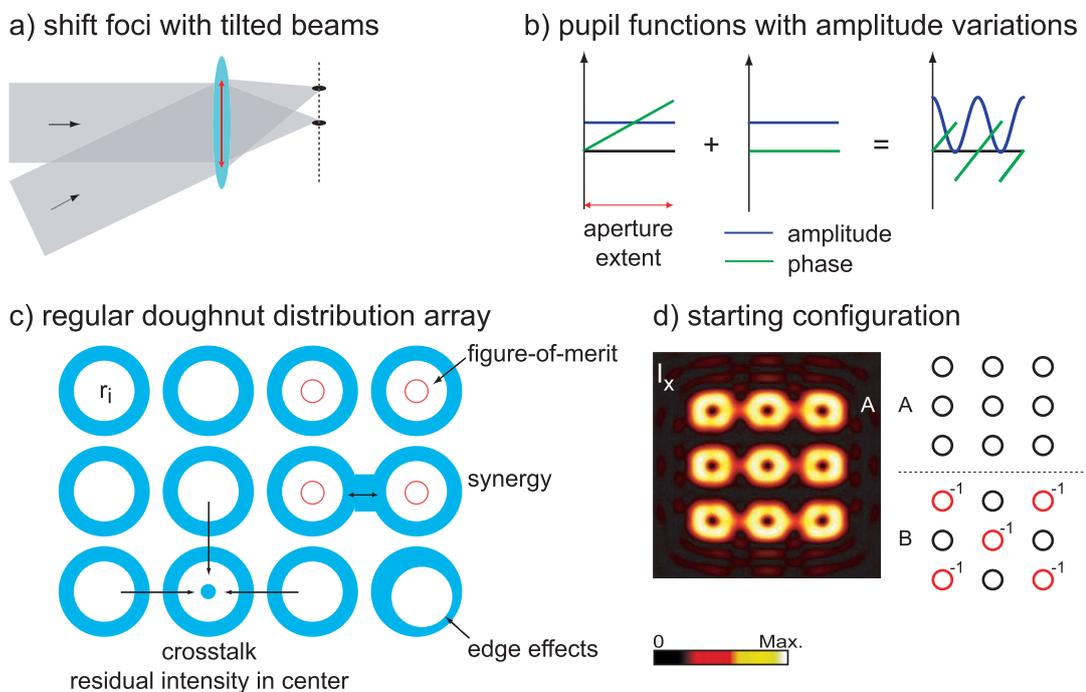


Figure 4.1: Multi-Doughnut arrays. a) Incoming plane waves with different propagation directions will be focused to spatially separated spots. b) This leads to spatially varying amplitude and phase distributions in the pupil area (the red double arrow-headed line in a)). c) The foci of an array of doughnut distributions placed on a quadratical grid are mutually influenced by interference. The intensity zero property of the single doughnut distribution is lost and the edge of the pattern can be under- or overweighted. Another effect is the synergy between doughnut structures of neighbored foci. The pattern in d) depicts the I_x component of the coherent addition of 9 single doughnut distributions which were placed on corresponding grid positions. (A) and (B) show the places of the single doughnut distributions which were used in the starting configurations. In (B), every second doughnut distribution has an additional phase shift of π that results in the subtraction of their focal fields in the resulting pattern.

accordingly generating a suitable complex pupil function. A general observation can be made thereby: the more extended the desired area of foci is, the higher the variations in amplitude and phase that will be necessary in the pupil plane.

Just as in chapter 3, circularly polarized light was used with a constant polarization over the whole pupil area. This allows a relatively homogeneous de-excitation of molecules oriented in the xy -plane. Covering the z -oriented molecules is of minor importance since they are normally a lot less efficiently excited and detected. In principle, the locations of the foci could be distributed over the focal plane in any suitable configuration. Nevertheless, a direct optimization for different configurations with random distances between neighboring foci turned out to be inefficient because of the huge number of free parameters. Furthermore, efficient scanning requires some kind of regular lattice. The symmetries of the electromagnetic field components of a single lateral doughnut (if the real part of the x -component is symmetrical to the x -axis the imaginary part will be symmetrical to the y -axis), support the usage of a quadratical lattice. We confirmed by simulations that other planar lattices indeed lead to less efficient interference effects. Therefore, only quadratical lattices were analyzed in detail. The scanning of the sample will be conducted so that every foci of the pattern is steadily shifted over every position of the sample.

In figure 4.1(d), a superposition of single doughnut distributions ($\tilde{A}_{\text{Lat},c}$ from equation (3.3)) with shifted centers corresponding to the grid points of a quadratical grid is shown. Despite of the crosstalk effects, it can be seen that the local behavior of the doughnut distributions remains similar to the single doughnut distribution case. Two configurations are shown right to the image: one is the coherent addition of single doughnut distributions while the other shifts the phase of each second distribution by π which results in the subtraction of their focal fields in the pattern. Both configurations show a similar intensity behavior around the grid positions but differ in the space between the doughnut distributions. For figure 4.1(d), the distance between neighboring doughnut distributions is in the order of about two Airy disks. The intensity is not zero at the grid positions. Nevertheless, this superpositions can already be seen as an approximation to a multi focus de-excitation array.

4.3 Optimization Algorithm

As in the case of single doughnuts, strict intensity zeros need to be enforced. A method to explicitly construct a subspace of pupil functions which assert strict intensity zeros was introduced in section 2.5.3. The full potential of this method can be exploited here.

We use the basis of pupil functions $\tilde{Z}_i(r, \phi)$ which were introduced in section 2.5.2. Initially, the starting configuration is decomposed into that basis. This allows a fast computation of electric field components (see equation (2.43)). A quite large number of basis functions is taken into account here in order to accurately represent pupil functions and multifocal areas. The necessary number of basis functions is determined by the necessary order of high frequency components of the pupil function in order to cover the whole doughnut array in the focal plane accurately. E.g., the pupil function responsible for two foci which are a distance d apart will approximately oscillate with a period of $2f\lambda/(nd)$. A practical criterion has

been the quality of the decomposition of the starting configuration. E.g., for an array of 100 foci up to 5000 basis vectors were needed. Our starting configurations, the superpositions of single doughnuts on a quadratical lattice, (A) and (B), feature a distribution of residual intensities on the lattice points. After the decomposition, a special solution was therefore constructed that provided exactly the opposing electric field vectors at the grid positions in order to restore the intensity zeros. The sum of both of these pupil functions and an arbitrary vector of the subspace, that leaves the intensity zeros unchanged, represent the optimization space from which the pupil function $\tilde{A}(r, \phi)$ with the highest FoM should be extracted. The FoM M_m was derived from the case of single doughnut optimization (see equation (3.2)). The intensities on a collection of points P_{ER} around each focus were investigated and the global minimum was taken.

$$M_m = \min_{r_i \in P_{\text{ER}}} I_x(r_i), I_y(r_i) \quad (4.1)$$

The points P_{ER} were strategically placed on spheres around each foci (see red circles in figure 4.1(c)). The radii of the spheres (d_{ER}) was chosen equally to its value in the single doughnut optimization (section 3.1). The starting configuration already represented a good configuration and the correction for the crosstalk at the grid position did not have a large influence. Therefore, a local optimization of the FoM seemed to be sufficient. We used the simplex-maximization algorithm [85] which was already applied in chapter 3. Because of the increased number of basis functions and the high number of necessary iterations of the optimization, the simulations were quite time consuming so that a parallel version of the algorithm was developed that ran on four CPUs from several hours to days for each optimization.

4.4 Results

The optimization algorithm was performed for 4, 9, 16, 25, 49 and 100 foci on a quadratical lattice for different values of the lattice constant. Intensity patterns are shown in figure 4.2 for a lattice parameter of 1.4 Airy disks.

They feature a strict intensity zero pattern on every grid position while the main intensity is present in the I_x and I_y components in the area close to the grid positions. Clearly, an optimized multi-doughnut distribution can be created. Nevertheless, the resulting pattern of doughnut distributions is not uniform which is a direct consequence of the limited spatial frequency content of the de-excitation pattern. Therefore, each individual focus in the pattern is not expected to produce equal resolution enhancement. In the figure, doughnut intensity distributions at the edge of the pattern are slightly less narrow than at the center of the pattern. The spread in the widths of the doughnuts is about 5%-20% of the minimum for typical patterns.

The distance of 1.4 Airy disks was not chosen arbitrarily. At this value of the lattice constant a very efficient use of energy was observed. The energy efficiency is described by the parameter η . It is defined as the ratio of the FoM for multifocal patterns and the FoM for a single doughnut with optimized amplitude profile (see section 3.2.2). The power of the

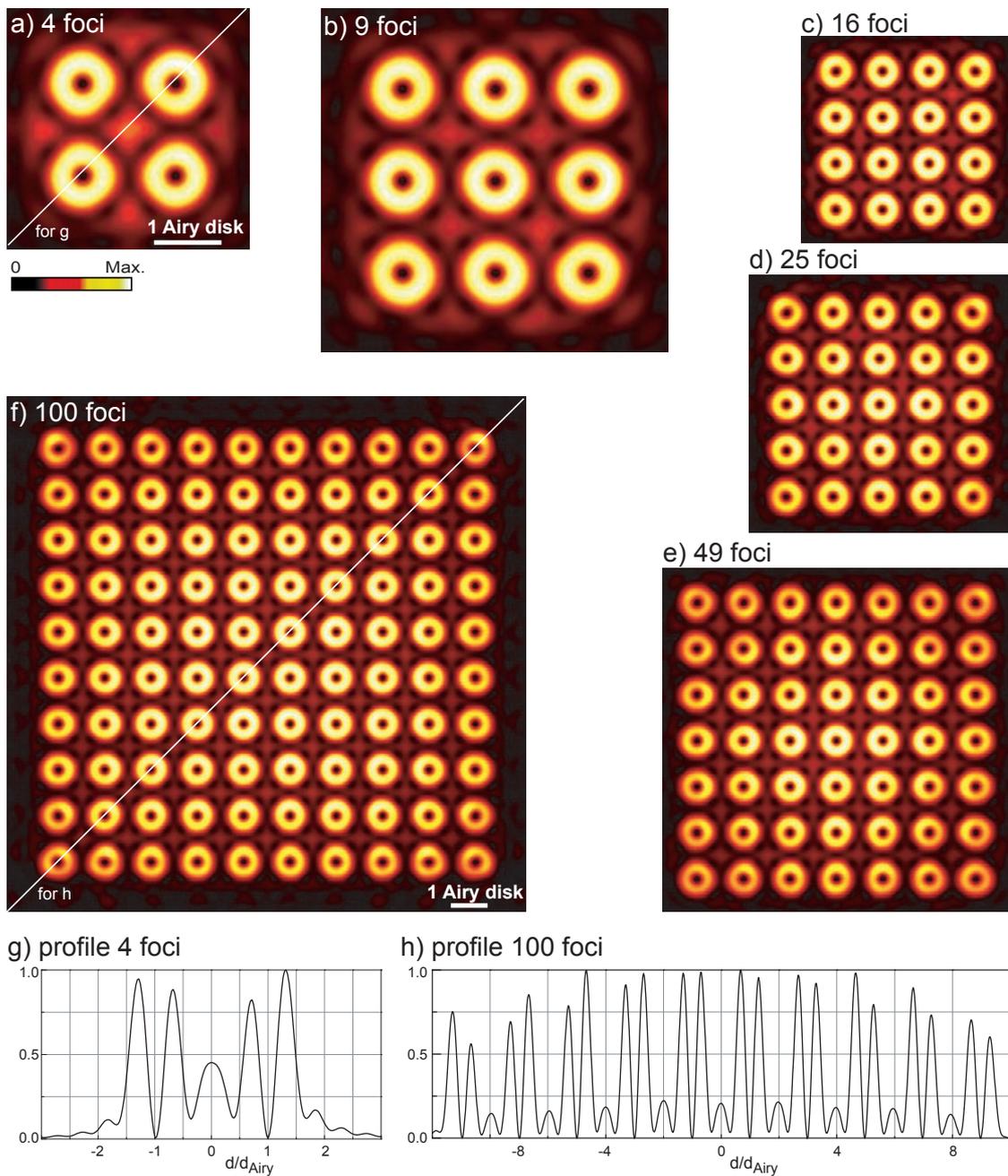


Figure 4.2: Optimal multi doughnut arrays for a) 4, b) 9, c) 16, d) 25, e) 49 and f) 100 doughnut foci with a distance of 1.4 Airy disks between neighboring foci. The resulting intensity distribution in the focal plane is shown. Although, the individual foci look very similar, a slight modulation over the whole pattern can be observed. In g) and h), the diagonal profiles of the patterns a) and f) are shown.

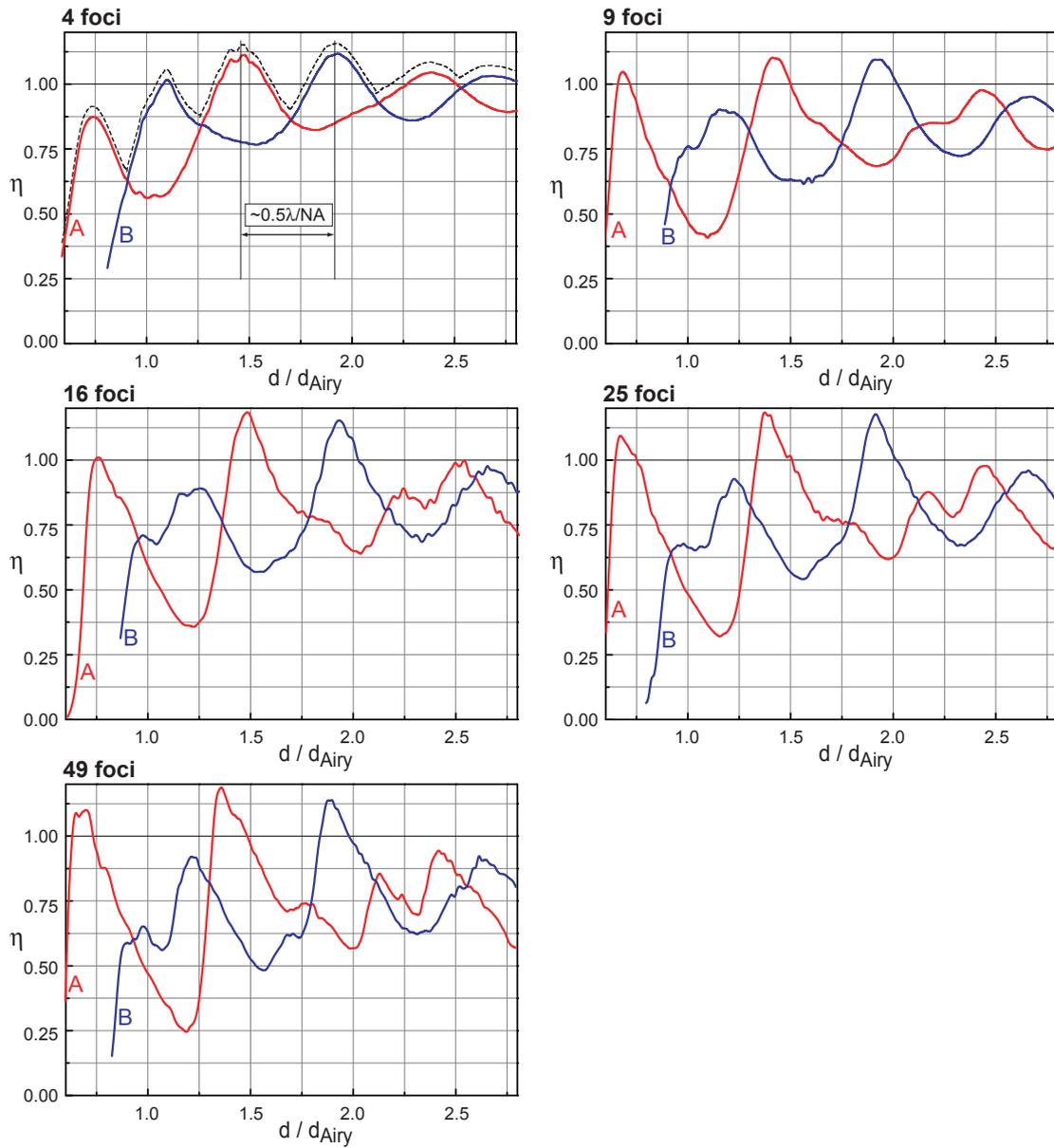


Figure 4.3: Dependence of the energy efficiency η on the lattice parameter and the number of foci. The two curves (A) and (B) represent the two starting configurations from figure 4.1(d). The black dotted line depicts the maximum of curves A and B. It marks the highest possible η within the optimization algorithm.

pattern is normalized to the number of foci it contains. A value of 1 means that a pattern of n foci with n times the power of the single doughnut achieves the same FoM as the optimal single doughnut pattern. A higher value means that the creation of a pattern of n foci with the same FoM as in the single focus case is possible with less than n times the power.

This means that synergistic effects as for example shared use of intensity walls by neighboring foci prevail over negative crosstalk effects.

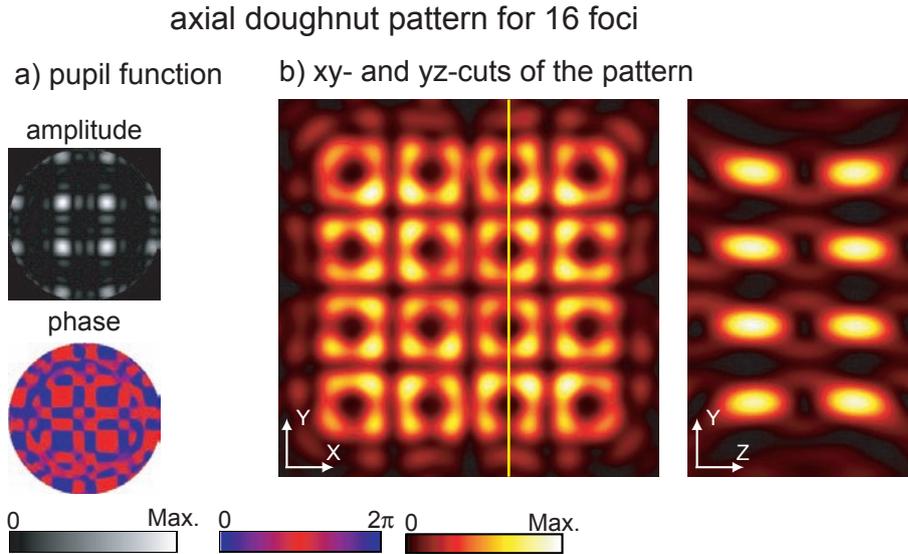


Figure 4.4: Pupil function of the optimized multi doughnut pattern of axial doughnuts a). In b), the intensity distribution in the focal plane is shown. In the yz-image which is taken at the x-position marked by the yellow line, it can be seen that each focus is accompanied by a doughnut along the optic axis. The maximal intensity in the xy-image is 36.8% of the maximal intensity in the zy-image.

In figure 4.3, the maximized energy efficiency η is shown in dependence of the number of foci in the pattern and the value of the lattice constant. We started the optimization algorithm with two different starting configurations (as described in section 4.2). After restoration of the intensity zeros a local optimization of the FoM was performed. Indeed, the results showed efficiencies η higher than one for some values of the lattice constant. However, the values of the lattice constant corresponding to a high η were dependent on the used starting configuration. This is not surprising since both starting configurations correspond to different interference effects in the area between the doughnut foci. For high η , the energy is mostly confined to the grid positions and destructive interference must take place between the doughnut foci. Due to an additional phase factor of π for each second doughnut foci in starting configuration (B), we expect the favorable interference there to take place at lattice constants that are located between the favorable lattice constants for starting configuration (A). In the case of 4 foci, a global optimization was performed additionally (according to appendix A.3) that yielded the maximum of curves A and B in the figure, independently of the initial configuration. The maximum of both curves in the figure depicts the highest η in dependence for each value of the lattice constant that was reached with the optimization algorithm. The distance between consecutive favorable lattice constants is $\sim 0.5\lambda/NA$. The highest achieved efficiency was about 1.2, and it could be reached for all numbers of foci. The two values of the lattice constants that were especially efficient were $\sim 1.4d_{\text{Airy}}$ and $\sim 1.9d_{\text{Airy}}$. Therefore, it is possible to create an array of n doughnut intensity distributions with perfect intensity zeros and 20% less relative power usage. The FoM M_m was designed to depict the smallest intensity on circles around each focus. Therefore, it would be a lower

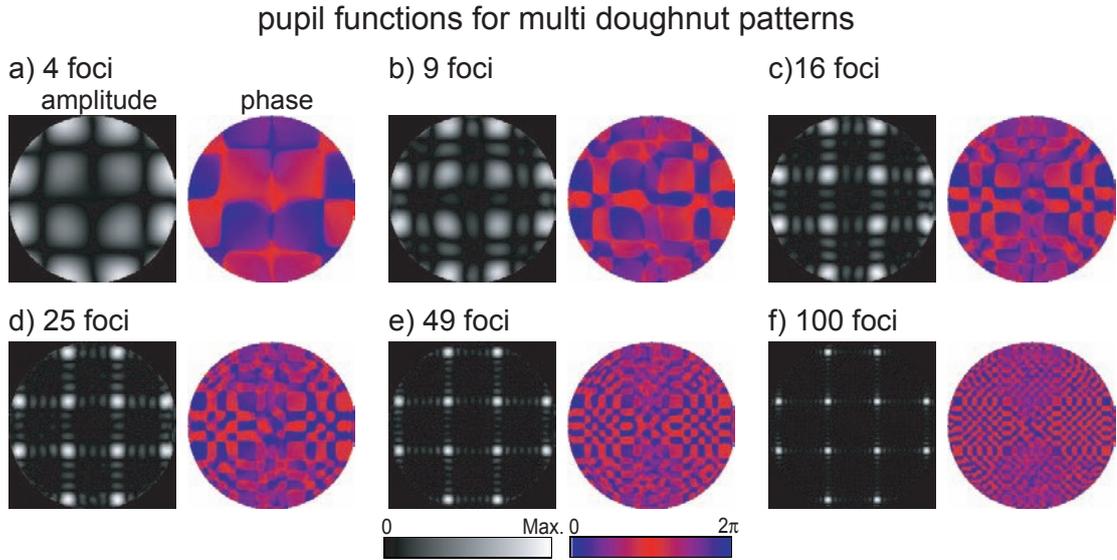


Figure 4.5: Amplitude and phase distributions representing the pupil function that generate the multifocal arrays of figure 4.2. Only a few areas of the pupil area are significantly illuminated. An increasing number of foci in the pattern is accompanied by a sharpening of the illuminated areas. The phase distributions are very complex but for the main peaks in the amplitude, a phase behavior similarly to $\tilde{A}_{\text{Lat},c}$ can be observed.

bound for the FoMs $\{M_c(i)\}$ (see section 3.1) that could be assigned for each doughnut distribution i out of the array. Calculating these single focus FoMs, their minimum would correspond to M_m . At a lattice constant of $\sim 1.4d_{\text{Airy}}$ or $\sim 1.9d_{\text{Airy}}$ the spread in $\{M_c(i)\}$ would be about 10% of M_m .

For very small lattice parameters (<0.7 Airy disks), the method has been very inefficient because a large amount of crosstalk has to be taken into account in enforcing the intensity zeros. Consequently, the resulting configuration is far away from any optimal FoM. Additionally, it is known that if intensity zeros are positioned too close to each other the intensity between them cannot be raised considerably while the main energy is directed outside of the area of interest (e.g. see Toraldos work [45]). For lattice constants larger than one Airy disk η oscillates around unity. Indeed, for a large distance of neighboring foci, the synergy effects diminish and finally the efficiency should approach unity.

Arrays of axial doughnut distributions were also investigated. Analogously to the procedure for lateral multi doughnut arrays, the doughnut distributions produced with $\tilde{A}_{3D,cl}$ were coherently placed on a quadratical grid as a starting point. The subsequent optimization yielded a favorable pattern of 3D doughnuts where the axial steepness of the pattern at each foci position is comparable with the single intensity zero distribution corresponding to $\tilde{A}_{3D,cl}$. The results are shown for 16 foci in figure 4.4. The amplitude distribution of the corresponding pupil function is very similar to the amplitude distributions in figure 4.5 while the phase distribution shows less variations.

4.4.1 Expected Resolution

To obtain an estimate of the expected resolution, we chose a configuration of 25 lateral doughnut foci with a lattice constant of 1.4 Airy disks. This corresponds to an average distance between neighboring foci of $\sim 1\mu\text{m}$ with $\lambda_{\text{RESOLFT}} = 700\text{nm}$. The excitation pattern was chosen to be an array of 25 foci that were the product of coherent interference of 25 single unaberrated foci shifted so that they were aligned with the doughnut array (figure 4.6(a)). No further optimization of the excitation pattern was made in order to show the compatibility of the creation of the excitation pattern with conventional techniques like microlens-arrays [16]. Around each peak in the excitation pattern, one doughnut of the de-excitation pattern was located (figure 4.6(b)). A saturation factor of $\xi = 50$ which is achievable in practice [58] was assumed. A pattern of sharp peaks describes the distribution of effective excitation and is shown in figure 4.6(c). The FWHM varies between 43 nm at the center of the pattern and 49 nm at the edge.

Depending on the nature of the sample, different detection strategies can be implemented. Very important is the correct assignment of the detected photons to the peaks in the pattern from which they originate. For this purpose, a pinhole array or a CCD camera with appropriately combined pixels can be used in practical applications. In figure 4.6(d), the image of a focused plane-like object is shown. The assignment of most of the detected photons to the foci position of the pattern from which they originate is possible because the peaks in the detection plane are well separated. For thick samples however, it might be necessary to restrict the assignment to small areas around each foci position in the image because a larger background will be created. During the scan, all foci of the multifocal pattern are moved over all parts of the sample. In this way, each pinhole or combined pixel of the CCD camera which is assigned to a foci of the pattern gives rise to a specific effective PSF. These PSFs are slightly inhomogeneous and the obtained data will consist of 25 very similar images of the object, each of these connected with one PSF. For the sake of completeness, let us mention that other image restoration methods using the full spatial fluorescence distribution in the detection plane could be utilized as well [80].

Compared to a single spot scanning RESOLFT microscope, the spatial resolution is retained almost unchanged while the photon flux from the sample has increased by a factor of 25. If the field of view is greater than the area of the multifocal pattern (side length about $5\mu\text{m}$ here), the imaging speed will also increase by a factor of 25. This shows the power of the method. Indeed, a linear increase in imaging speed with the number of used foci can be expected in many situations. However, the stability of the doughnut pattern, particularly of the intensity zeros, has to be maintained. Unwanted aberrations in the optical setup that lead to remaining intensities at the grid points could be corrected conveniently with the same method that is used to correct crosstalk in the doughnut pattern. In particular, it is possible to apply fields that yield nonzero intensities only at one or a few grid points. Therefore each single focus in the pattern could be adjusted independently.

The excitation pattern shows small crosstalk effects mainly present by distortions of the local shapes of the foci including biasing of the intensity maxima toward their neighbors. An optimization of the excitation pattern including the correction of crosstalk effects or

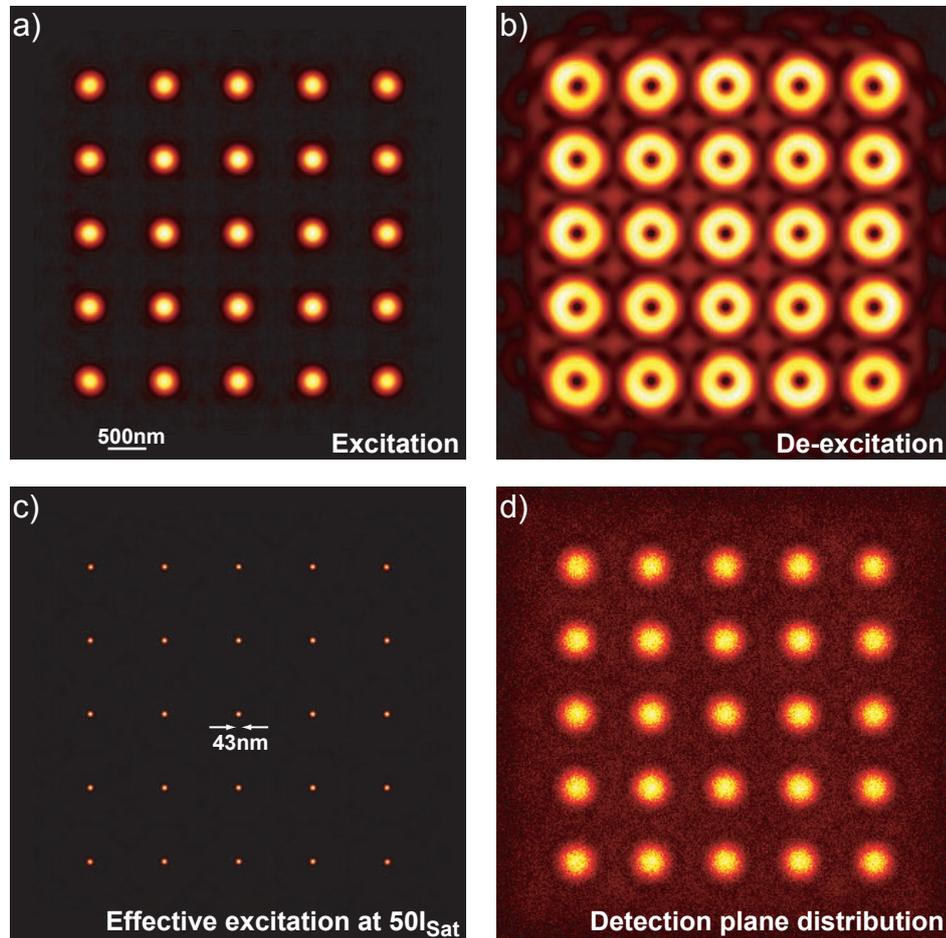


Figure 4.6: Multi foci pattern for a) excitation, b) de-excitation doughnuts, c) effective excitation and d) a possible intensity distribution in the detection plane for a focused plane-like object. The FWHM of the 25 spots in the excitation pattern is about 219 nm. The excitation pattern is not optimized and the brightness of the single spots varies about 17%. Panel c) shows the effective excitation pattern for $50I_{Sat}$ calculated for STED microscopes. The FWHMs varies from 43 nm (center peak) to 49 nm (corner peak). The brightness of the single spots varies about 24%. The side length of the array of foci is about $5\mu\text{m}$. The intensity distribution in the detection plane has a voxel length of 15 nm, 126 counts in the maximum and a background noise of 5 counts on average. This reflects a common situation in high resolution microscopy.

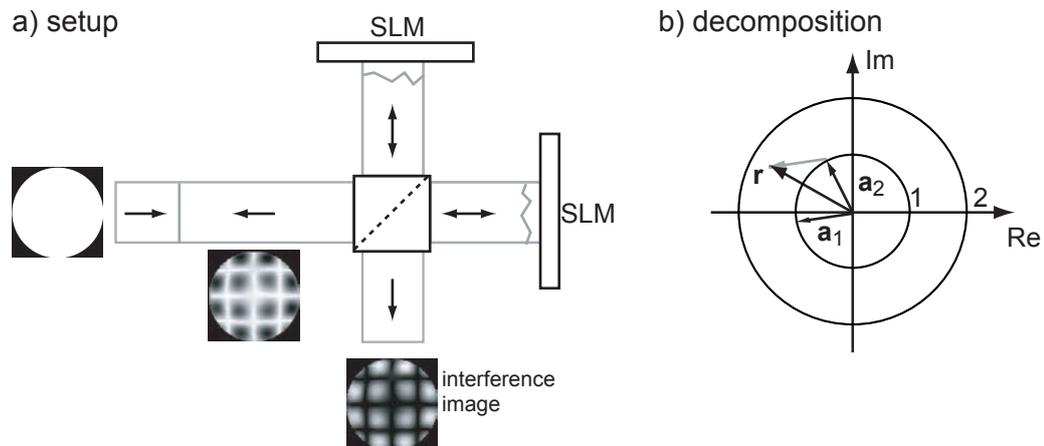


Figure 4.7: a) Proposed setup for creating arbitrary complex pupil functions. The incoming light beam, a plane wave, is subject to a passage of a spatial variable interferometer. In both arms of the interferometer, a spatial light modulator (SLM) distorts the wavefront. Provided that the optical path lengths can be aligned sufficiently, already one SLM could be used sharing its active area for both arms of the interferometer. The interference of both manipulated wavefronts shows amplitude and phase variations. Part of the light is sent back to the origin. In b), the decomposition of a vector r with arbitrary amplitude (amplitude is normalized to be within circle 2) in two vectors a_i with fixed amplitude (circle 1) is shown. Together, all vectors in the diagram form a parallelogram. The decomposition can always be done and is unique except of the order of the two vectors.

specific additional suppression of intensity between the foci would be easily feasible with the algorithm presented in this chapter.

The effects of additional gaussian noise (spatially and timely uncorrelated random phase and amplitude deviations) on the wavefront presented at the objective lens back aperture were regarded briefly. It can be seen that the allowable strength of uncorrelated phase and amplitude noise scales with $n^{-1/2}$ where n is the number of foci. Furthermore, the influence of pixelation and discretization of a SLM creating the doughnut pattern was simulated. Realizing the necessary complex wavefronts by a SLM with 256 discrete phase steps and at least 500×500 pixels in the active area has only negligible effects on the purity of the intensity zeros for foci patterns with no more than 100 foci.

4.5 Generation of General Complex Apodizations

Some of the highly complex pupil functions that are used to produce the doughnut patterns are shown in figure 4.5. A common characteristic of these pupil plane distributions is the very heterogeneous illumination of the aperture area. The successful experimental realization of multifocal RESOLFT microscopes will require a very careful preparation of these complex amplitude and phase apodizations. Particularly the edge areas of the aperture are of importance since they are mainly responsible for creating small lateral structures. Furthermore, other intensity distributions suitable for parallelization like lines or multifocal patterns

for the excitation require complex pupil functions as well.

Some methods for creating a complex amplitude and phase distribution were already introduced. E.g., a reconfigurable array of optical tweezers was realized [86, 87]. However, the originally used generalized phase contrast method could only generate amplitude profiles. Another method was developed by Wilson [15, 88]. Both of these methods rely on phase or amplitude filters in the Fourier transform of a phase-only distorted wavefront to produce a complex distribution. But the prediction of the accuracy of the generation of our complex pupil functions is difficult since higher order contributions are usually neglected. While a combination of these techniques might well give an efficient and accurate method to produce the desired pupil functions, another way that mainly relies on spatially variable interferometric beam shaping shall be proposed here. A careful local control of the wavefront can be realized with a spatial light modulator (SLM). Commonly, it can be programmed and therefore possess the possibility to improve the result of a wavefront manipulation with feedback loops. An optical setup composed of two SLMs (each one capable of manipulating the wavefront of one beam) is shown in figure 4.7(a). Upon impinging the beam splitter for the second time, the outgoing light beam will consist of a superposition of the manipulated wavefronts of each SLM. The amplitude of this beam can vary from zero (destructive interference of both SLM contributions) to the original value (constructive interference). Figure 4.7(b) illustrates the necessary calculations to obtain the SLM phase distortion functions for a given desired amplitude and phase function on the exit of this setup.

The procedure determines the state of the outgoing beam for each position by the corresponding phase distortions of both interferometer arms. On the backward going direction of the beam splitter, another spatial amplitude and phase distribution is created additionally. It is complementary to the distribution of interest so that the sum of both equals the original incoming beam distribution. The desired amplitude distribution can be seen as 'carved out' of the amplitude distribution of the incoming light beam. Therefore, this technique to generate arbitrary amplitude and phase distribution exhibits a loss in available power that can be severe for amplitude distributions with large sparse peaks. This technique would rely on a abundant supply of available laser power. Defining the transmission as the relation of power in the outgoing beam and power of the incoming beam, the energy efficiency of the generation of the complex apodizations can be quantified. The middle column of table 4.2 shows that the transmission drops below 10% already for more than 9 foci. This is due to the needed distribution of sharp peaks in the pupil function as shown in figure 4.5. An important observation is that the Fourier transform of such sparsely illuminated pupil functions can be created efficiently with our method. This can be exploited by creating the Fourier transform of the pupil function of interest and focusing this pattern onto the back aperture of the objective lens with a low NA lens. However, while the pupil function has a limited extent, its Fourier transform is not band-limited. The setup described in figure 4.7 is restricted to the creation of complex distributions with limited spatial extent. Therefore, only band-limited approximated Fourier transforms of the pupil function could be generated. Appropriate band-limited Fourier transforms of functions that approximate the desired pupil function were found with a modified Gerchberg-Saxton algorithm [89]. An important parameter was the cut-off frequency of the band-limited approximation. The highest efficiency was obtained for the smallest possible

number of foci	transmission in the pupil plane	maximal transmission in the Fourier plane
1	1.0	-
4	0.145	0.12
9	0.104	0.20
16	0.064	0.148
25	0.044	0.159
49	0.027	0.154

Table 4.2: Average transmission in the pupil plane and in the Fourier plane after application of the optimization algorithm. The middle column describes the average transmission of pupil functions from figure 4.5 which corresponds to the energy efficiency of the creation of the pupil function if the setup of figure 4.7 is used directly. For a large number of foci, the efficiency is very low. Using the algorithm presented in the text and introducing an additional Fourier transform in the setup, the efficiencies (right column) can be raised to an average level of 14% (right column). The variance in the values is probably caused by the particular mode of operation of the optimization algorithm.

cut-off frequency where the algorithm could still deliver accurate results.

The algorithm was designed as follows: In every iteration step an estimate of the pupil function is Fourier transformed to yield an estimate in the Fourier plane. The band-limit condition is applied by setting all values above the cut-off frequency to zero and the inverse Fourier transform is applied. In the pupil area, the estimate is set to the desired complex apodization. Values outside the pupil function area could vary arbitrarily. The obtained maximal transmissions are also shown in table 4.2. It can be seen that the energy efficiency can always be held above 14% by either generating the pupil plane apodization directly or by generating a complex distribution whose Fourier transform includes the desired apodization. Systematic errors are introduced due to the restrictions in the Fourier plane. However, the purity of the intensity zeros of the multi doughnut grid stayed below less than 1% of the intensity 50 nm away from the foci, a value below common experimental values for single doughnuts.

5 Conclusion and outlook

A comprehensive search, optimization and characterization of de-excitation patterns for RESOLFT microscopy has been performed.

A framework for the description of a RESOLFT microscope's focused de-excitation beam was developed. In order to rigorously search in the space of pupil functions, they were decomposed into orthonormal polynomials. However, an efficient search for optimal de-excitation patterns only became feasible by constructing a subspace that enforces intensity zeros at the focal positions. They are essential in RESOLFT microscopy. An algorithm adopted from statistical physics allowed the search for global optima.

If the maximal amplitude of the pupil function is limited the algorithm converges to phase-only pupil functions. When other restrictions were regarded, more sophisticated patterns with variable amplitude emerged.

Some of the phase masks [90, 56] that were used in experiments before correspond to pupil functions that were also identified in our optimization encouraging their ongoing application. However, the optimization identified a novel, superior lateral doughnut distribution for circularly polarized light. Its experimental implementation has led to a hitherto unrivaled lateral resolution of down to 20 nm [57, 58] in biological systems.

It is conclusively shown in this work that a single de-excitation pattern cannot efficiently cover all intensity components and all directions around an intensity zero. Expanding the search to optimal incoherent combinations of doughnut distributions, a combination was found that can improve the resolution substantially over a single de-excitation pattern. It was composed of a doughnut distribution in the focal plane as well as a doughnut distribution along the optic axis, both utilizing circularly polarized light. This ensures the efficient coverage of two intensity components in all three directions with moderate experimental effort. Based on the previous findings, optimal de-excitation patterns for the remaining intensity component were proposed. The experimental realization of such incoherent combinations, the implementation of pupil functions with variable amplitude, and the usage of de-excitation patterns for axially oriented molecules is intended in the near future.

In order to utilize the full potential of RESOLFT microscopy, a new parallelization concept was developed that would allow the fast acquisition of high resolution images. Other than existing multifocal parallelization techniques our technique is based on the coherent creation of compact arrays of sub-diffraction sized fluorescence spots in the sample. Rather complex pupil functions involving strong amplitude variations are necessary to create these distributions. A calculation and optimization scheme for such pupil functions has been outlined and the most favorable lattice parameters for multifocal arrangements on quadratical lattices were identified. Arrays with up to 100 foci, created by this technique, are presented and an optical setup which could generate such complex pupil functions is detailed. Its prac-

tical realization is still pending. Such a setup will offer several advantages over other forms of fast RESOLFT microscopy, e.g. using line patterns. Most importantly, it does not require mathematical post-processing. The generation of even larger patterns is possible using the technique outlined here but is unrealistic at the moment due to restricted available laser power. It was therefore not pursued.

Recently, interest in the focusing properties of partially coherent beams has increased [91, 92, 93]. While the use of coherent light was assumed throughout this dissertation, it might be necessary to account for the degree of coherence that can be realized in practice. Calculations show that under certain conditions the quality of the intensity zeros can be compromised [91].

Even for high degrees of coherence, special experimental effort is necessary to provide strict intensity zeros for all utilized de-excitation patterns. To accomplish this, active optics is already routinely applied. The optimization framework laid out in this dissertation could constitute the foundation to extend the current static correction scheme [94] to a closed-loop optimization scheme of de-excitation patterns.

RESOLFT microscopy in general poses new challenges for image restoration. The way higher spatial frequencies are created depends on the molecules' orientational density in the sample. For inhomogeneous distributions, the effective PSF can be space-variant. Furthermore, the presented multifocal de-excitation patterns show an intrinsic variation in the pattern of effective foci also resulting in variable PSFs for some scanning schemes. Therefore, the development of new deconvolution techniques, which could also profit from the theoretical framework presented in this dissertation, will ultimately be required.

In this dissertation, the field of possible de-excitation patterns was explored in depth. Combined with the recent advances in the photophysics of the RESOLFT concept the application of the results has led to an unprecedented resolution of far field light microscopy. The resolving power of far field fluorescence microscopes has now reached the macromolecular scale which promises many new, exciting applications.

Bibliography

- [1] R. E. Williams, B. Blacker, M. Dickinson, W. V. Dixon, H. C. Ferguson, A. S. Fruchter, M. Giavalisco, R. L. Gilliland, I. Heyer, R. Katsanis, Z. Levay, R. A. Lucas, D. B. McElroy, L. Petro, M. Postman, H. M. Adorf, and R. N. Hook. Hubble deep field: Observations, data reduction, and galaxy photometry. *Astronomical Journal*, 112(4):1335, 1996.
- [2] M. Schleiden, T. Schwann, and M. Schultze. *Klassische Schriften zur Zellenlehre*. Harri Deutsch, Frankfurt am Main, 2003.
- [3] J. Odenthal and C. Nusslein-Volhard. Fork head domain genes in zebrafish. *Development Genes and Evolution*, 208(5):245–258, 1998.
- [4] W. K. Huh, J. V. Falvo, L. C. Gerke, A. S. Carroll, R. W. Howson, J. S. Weissman, and E. K. O’Shea. Global analysis of protein localization in budding yeast. *Nature*, 425(6959):686–691, 2003.
- [5] G. J. Cotton and T. W. Muir. Generation of a dual-labeled fluorescence biosensor for crk-ii phosphorylation using solid-phase expressed protein ligation. *Chemistry and Biology*, 7(4):253–261, 2000.
- [6] J. Lippincott-Schwartz and G. H. Patterson. Development and use of fluorescent protein markers in living cells. *Science*, 300(5616):87–91, 2003.
- [7] E. Abbe. Beiträge zur Theorie des Mikroskops und der mikroskopischen Wahrnehmung. *Arch. f. Mikroskop. Anat.*, 9:413–420, 1873.
- [8] T. Wilson and C.J.R. Sheppard. *Theory and Practice of Scanning Optical Microscopy*. Academic Press, New York, 1984.
- [9] K. König, Y. Tadir, P. Patrizio, M. W. Berns, and B. J. Tromberg. Effects of ultraviolet exposure and near infrared laser tweezers on human spermatozoa. *Human Reproduction*, 11(10):2162–2164, 1996.
- [10] J. Kirz, C. Jacobsen, and M. Howells. Soft-x-ray microscopes and their biological applications. *Quarterly Reviews of Biophysics*, 28(1):33–130, 1995.
- [11] G. Schneider. Cryo x-ray microscopy with high spatial resolution in amplitude and phase contrast. *Ultramicroscopy*, 75(2):85–104, 1998.

- [12] D. Shapiro, P. Thibault, T. Beetz, V. Elser, M. Howells, C. Jacobsen, J. Kirz, E. Lima, H. Miao, A. M. Neiman, and D. Sayre. Biological imaging by soft x-ray diffraction microscopy. *Proceedings of the National Academy of Sciences of the United States of America*, 102(43):15343–15346, 2005.
- [13] S. W. Hell. European patent (ep 0 491 281 b1): Doppelkonfokales Rastermikroskop. 1990.
- [14] H. Gugel, J. Bewersdorf, S. Jakobs, J. Engelhardt, R. Storz, and S. W. Hell. Cooperative 4pi excitation and detection yields sevenfold sharper optical sections in live-cell microscopy. *Biophysical Journal*, 87(6):4146–4152, 2004.
- [15] M. A. A. Neil, R. Juskaitis, and T. Wilson. Method of obtaining optical sectioning by using structured light in a conventional microscope. *Optics Letters*, 22(24):1905–1907, 1997.
- [16] A. Egner, V. Andresen, and S. W. Hell. Comparison of the axial resolution of practical nipkow-disk confocal fluorescence microscopy with that of multifocal multiphoton microscopy: theory and experiment. *Journal of Microscopy-Oxford*, 206:24–32, 2002. Part 1.
- [17] D. W. Pohl, W. Denk, and M. Lanz. Optical stethoscopy - image recording with resolution $\lambda/20$. *Applied Physics Letters*, 44(7):651–653, 1984.
- [18] D. Axelrod. Total internal-reflection fluorescence microscopy. *Methods in Cell Biology*, 30:245–270, 1989.
- [19] W. Denk, J. H. Strickler, and W. W. Webb. 2-photon laser scanning fluorescence microscopy. *Science*, 248(4951):73–76, 1990.
- [20] J. Bewersdorf, R. Pick, and S. W. Hell. Multifocal multiphoton microscopy. *Optics Letters*, 23(9):655–657, 1998.
- [21] A. Schönle, J. Keller, B. Harke, and S.W. Hell. Diffraction unlimited far-field fluorescence microscopy. In M. Masters and P. So, editors, *Handbook of Biological Nonlinear Optical Microscopy*. Oxford University Press, Oxford, 2007.
- [22] L. Moreaux, O. Sandre, and J. Mertz. Membrane imaging by second-harmonic generation microscopy. *Journal of the Optical Society of America B-Optical Physics*, 17(10):1685–1694, 2000.
- [23] Y. Barad, H. Eisenberg, M. Horowitz, and Y. Silberberg. Nonlinear scanning laser microscopy by third harmonic generation. *Applied Physics Letters*, 70(8):922–924, 1997.

- [24] J. X. Cheng, Y. K. Jia, G. F. Zheng, and X. S. Xie. Laser-scanning coherent anti-stokes raman scattering microscopy and applications to cell biology. *Biophysical Journal*, 83(1):502–509, 2002.
- [25] T. Ichimura, N. Hayazawa, M. Hashimoto, Y. Inouye, and S. Kawata. Tip-enhanced coherent anti-stokes raman scattering for vibrational nanoimaging. *Physical Review Letters*, 92(22), 2004.
- [26] M.C. Teich and B.E.A. Saleh. Entangled-photon microscopy. *Ceskoslovensky casopis pro fyziku*, 47:3–8, 1997.
- [27] S. W. Hell. Toward fluorescence nanoscopy. *Nature Biotechnol.*, 21(11):1347–1355, 2003.
- [28] S. W. Hell. Strategy for far-field optical imaging and writing without diffraction limit. *Physics Letters A*, 326(1-2):140–145, 2004.
- [29] S. W. Hell, M. Dyba, and S. Jakobs. Concepts for nanoscale resolution in fluorescence microscopy. *Curr. Opin. Neurobio.*, 14(5):599–609, 2004.
- [30] S. W. Hell, S. Jakobs, and L. Kastrup. Imaging and writing at the nanoscale with focused visible light through saturable optical transitions. *Appl. Phys. A*, 77:859–860, 2003.
- [31] S. W. Hell and J. Wichmann. Breaking the diffraction resolution limit by stimulated emission: stimulated emission depletion microscopy. *Opt. Lett.*, 19(11):780–782, 1994.
- [32] S. W. Hell and M. Kroug. Ground-state depletion fluorescence microscopy, a concept for breaking the diffraction resolution limit. *Appl. Phys. B*, 60:495–497, 1995.
- [33] 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(49):17565–17569, 2005.
- [34] A. Schonle and S. W. Hell. Far-field fluorescence microscopy with repetitive excitation. *European Physical Journal D*, 6(3):283–290, 1999.
- [35] A. Schonle, P. E. Hanninen, and S. W. Hell. Nonlinear fluorescence through intermolecular energy transfer and resolution increase in fluorescence microscopy. *Annalen Der Physik*, 8(2):115–133, 1999.
- [36] V. Westphal, L. Kastrup, and S. W. Hell. Lateral resolution of 28nm ($\lambda/25$) in far-field fluorescence microscopy. *Appl. Phys. B*, 77(4):377–380, 2003.
- [37] M. Dyba and S. W. Hell. Focal spots of size $\lambda/23$ open up far-field fluorescence microscopy at 33 nm axial resolution. *Phys. Rev. Lett.*, 88:163901, 2002.

- [38] M. Bertero and P. Boccacci. Introduction to inverse problems in imaging. page 351. Instituts of Physics Publishing, 1998.
- [39] S. Ram, E. S. Ward, and R. J. Ober. Beyond rayleigh's criterion: A resolution measure with application to single-molecule microscopy. *Proceedings of the National Academy of Sciences of the United States of America*, 103(12):4457–4462, 2006.
- [40] M. Born and E. Wolf. *Principles of Optics*. Cambridge University Press, Cambridge, seventh (expanded) edition, 1999.
- [41] K. Konig, P. T. C. So, W. W. Mantulin, B. J. Tromberg, and E. Gratton. Two-photon excited lifetime imaging of autofluorescence in cells during uva and nir photostress. *Journal of Microscopy-Oxford*, 183:197–204, 1996. Part 3.
- [42] A. Schonle and S. W. Hell. Calculation of vectorial three-dimensional transfer functions in large-angle focusing systems. *Journal of the Optical Society of America a-Optics Image Science and Vision*, 19(10):2121–2126, 2002.
- [43] C. J. R. Sheppard, M. Gu, Y. Kawata, and S. Kawata. 3-dimensional transfer-functions for high-aperture systems. *Journal of the Optical Society of America a-Optics Image Science and Vision*, 11(2):593–598, 1994.
- [44] M. Gu and C. J. R. Sheppard. Confocal fluorescent microscopy with a finite-sized circular detector. *Journal of the Optical Society of America a-Optics Image Science and Vision*, 9(1):151–153, 1992.
- [45] G. Toraldo di Francia. Supergain antennas and optical resolving power. *Nuovo Cimento Suppl.*, 9:426–435, 1952.
- [46] V. F. Canales, P. J. Valle, J. E. Oti, and M. P. Cagigal. Variable resolution with pupil masks. *Optics Communications*, 257(2):247–254, 2006.
- [47] H. X. Luo and C. H. Zhou. Comparison of superresolution effects with annular phase and amplitude filters. *Applied Optics*, 43(34):6242–6247, 2004.
- [48] M. Martinez-Corral, P. Andres, C. J. Zapata-Rodriguez, and M. Kowalczyk. Three-dimensional superresolution by annular binary filters. *Optics Communications*, 165(4-6):267–278, 1999.
- [49] M. Martinez-Corral, M. T. Caballero, E. H. K. Stelzer, and J. Swoger. Tailoring the axial shape of the point spread function using the toraldo concept. *Optics Express*, 10(1):98–103, 2002.
- [50] M. Martinez-Corral, C. Ibanez-Lopez, and G. Saavedra. Axial gain resolution in optical sectioning fluorescence microscopy by shaded-ring filters. *Optics Express*, 11(15):1740–1745, 2003.

- [51] M. Y. Yun, L. R. Liu, J. F. Sun, and D. A. Liu. Three-dimensional superresolution by three-zone complex pupil filters. *Journal of the Optical Society of America a-Optics Image Science and Vision*, 22(2):272–277, 2005.
- [52] T. A. Klar, S. Jakobs, M. Dyba, A. Egner, and S. W. Hell. Fluorescence microscopy with diffraction resolution limit broken by stimulated emission. *Proc. Natl. Acad. Sci. USA*, 97:8206–8210, 2000.
- [53] L. Kastrup, H. Blom, C. Eggeling, and S. W. Hell. Fluorescence fluctuation spectroscopy in subdiffraction focal volumes. *Physical Review Letters*, 94(17), 2005.
- [54] M. Dyba, J. Keller, and S. W. Hell. Phase filter enhanced sted-4pi fluorescence microscopy: theory and experiment. *New Journal of Physics*, 7, 2005.
- [55] S. W. Hell. Increasing the resolution of far-field fluorescence light microscopy by point-spread-function engineering. In J.R. Lakowicz, editor, *Topics in Fluorescence Spectroscopy*, volume 5, pages 361–422. Plenum Press, New York, 1997.
- [56] T. A. Klar, 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.
- [57] K. I. Willig, S. O. Rizzoli, V. Westphal, R. Jahn, and S. W. Hell. Sted microscopy reveals that synaptotagmin remains clustered after synaptic vesicle exocytosis. *Nature*, 440(7086):935–939, 2006.
- [58] G. Donnert, J. Keller, R. Medda, A. Andrei, S. O. Rizzoli, R. Lüthmann, R. Jahn, C. Eggeling, and S. W. Hell. Macromolecular-scale resolution in biological fluorescence microscopy. *Proc. Natl. Acad. Sci. USA*, 103(31):11440–11445, 2006.
- [59] A. A. Heikal, S. T. Hess, G. S. Baird, R. Y. Tsien, and W. W. Webb. Molecular spectroscopy and dynamics of intrinsically fluorescent proteins: Coral red (dsred) and yellow (citrine). *Proceedings of the National Academy of Sciences of the United States of America*, 97(22):11996–12001, 2000.
- [60] B. Richards and E. Wolf. Electromagnetic diffraction in optical systems ii. structure of the image field in an aplanatic system. *Proc. R. Soc. Lond. A*, 253:358–379, 1959.
- [61] J. Enderlein. Theoretical study of detection of a dipole emitter through an objective with high numerical aperture. *Optics Letters*, 25(9):634–636, 2000.
- [62] G. S. Harms, M. Sonnleitner, G. J. Schutz, H. J. Gruber, and T. Schmidt. Single-molecule anisotropy imaging. *Biophysical Journal*, 77(5):2864–2870, 1999.
- [63] M. Gu. *Advanced Optical Imaging*. Springer, Berlin, 2000.

- [64] A. B. Bhatia and E. Wolf. On the circle polynomials of zernike and related orthogonal sets. *Proceedings of the Cambridge Philosophical Society*, 50(1):40–48, 1954.
- [65] W. J. Tango. Circle polynomials of zernike and their application in optics. *Applied Physics*, 13(4):327–332, 1977.
- [66] F. Zernike. Inflection theory of the cutting method and its improved form, the phase contrast method. *Physica*, 1:689–704, 1934.
- [67] B. R. A. Nijboer. The diffraction theory of optical aberrations .2. diffraction pattern in the presence of small aberrations. *Physica*, 13(10):605–620, 1947.
- [68] M. Schwertner, M. J. Booth, M. A. A. Neil, and T. Wilson. Measurement of specimen-induced aberrations of biological samples using phase stepping interferometry. *Journal of Microscopy-Oxford*, 213:11–19, 2004. Part 1.
- [69] G. H. Golub and C. F. Van Loan. *Matrix Computations*. Johns Hopkins University Press, Baltimore, 1996.
- [70] W. H. Press and et. al. *Numerical Recipes in C: the art of scientific computing*. Cambridge University Press, 1992.
- [71] M. Knoll. Charge potential and secondary emissions of electron irradiated bodies. *Physikalische Zeitschrift*, 36:861–869, 1935.
- [72] J. Bloem, M. Veninga, and J. Shepherd. Fully-automatic determination of soil bacterium numbers, cell volumes, and frequencies of dividing cells by confocal laser-scanning microscopy and image-analysis. *Applied and Environmental Microbiology*, 61(3):926–936, 1995.
- [73] C. Eggeling, A. Volkmer, and C. A. M. Seidel. Molecular photobleaching kinetics of rhodamine 6g by one- and two-photon induced confocal fluorescence microscopy. *Chemphyschem*, 6(5):791–804, 2005.
- [74] G. H. Patterson, S. M. Knobel, W. D. Sharif, S. R. Kain, and D. W. Piston. Use of the green fluorescent protein and its mutants in quantitative fluorescence microscopy. *Biophysical Journal*, 73(5):2782–2790, 1997.
- [75] G. H. Patterson and D. W. Piston. Photobleaching in two-photon excitation microscopy. *Biophysical Journal*, 78(4):2159–2162, 2000.
- [76] C. J. R. Sheppard, G. Calvert, and M. Wheatland. Focal distribution for superresolving toraldo filters. *Journal of the Optical Society of America a-Optics Image Science and Vision*, 15(4):849–856, 1998.
- [77] K. I. Willig. *STED microscopy in the visible range*. Dissertation, Ruperto-Carola University, 2006.

- [78] S. Quabis, R. Dorn, M. Eberler, O. Glockl, and G. Leuchs. Focusing light to a tighter spot. *Optics Communications*, 179(1-6):1–7, 2000.
- [79] L. Novotny, M. R. Beversluis, K. S. Youngworth, and T. G. Brown. Longitudinal field modes probed by single molecules. *Physical Review Letters*, 86(23):5251–5254, 2001.
- [80] M. G. L. Gustafsson. Nonlinear structured-illumination microscopy: Wide-field fluorescence imaging with theoretically unlimited resolution. *Proceedings of the National Academy of Sciences of the United States of America*, 102(37):13081–13086, 2005.
- [81] D. P. Bratu, B. J. Cha, M. M. Mhlanga, F. R. Kramer, and S. Tyagi. Visualizing the distribution and transport of mrnas in living cells. *Proceedings of the National Academy of Sciences of the United States of America*, 100(23):13308–13313, 2003.
- [82] G. Q. Xiao, T. R. Corle, and G. S. Kino. Real-time confocal scanning optical microscope. *Applied Physics Letters*, 53(8):716–718, 1988.
- [83] M. Straub and S. W. Hell. Multifocal multiphoton microscopy: a fast and efficient tool for 3-d fluorescence imaging. *Bioimaging*, 6(4):177–185, 1998.
- [84] A. Egner and S. W. Hell. Time multiplexing and parallelization in multifocal multiphoton microscopy. *Journal of the Optical Society of America a-Optics Image Science and Vision*, 17(7):1192–1201, 2000.
- [85] J. A. Nelder and R. Mead. A simplex-method for function minimization. *Computer Journal*, 7(4):308–313, 1965.
- [86] V. R. Daria, P. J. Rodrigo, and J. Gluckstad. Dynamic array of dark optical traps. *Applied Physics Letters*, 84(3):323–325, 2004.
- [87] P. J. Rodrigo, I. R. Perch-Nielsen, and J. Gluckstad. High-speed phase modulation using the rpc method with a digital micromirror-array device. *Optics Express*, 14(12):5588–5593, 2006.
- [88] M. A. A. Neil, F. Massoumian, R. Juskaitis, and T. Wilson. Method for the generation of arbitrary complex vector wave fronts. *Optics Letters*, 27(21):1929–1931, 2002.
- [89] R. W. Gerchberg and W. O. Saxton. Practical algorithm for determination of phase from image and diffraction plane pictures. *Optik*, 35(2):237, 1972.
- [90] T. A. Klar. *Progress in Stimulated Emission Depletion Microscopy*. Doctoral thesis, Rupertus Carola University of Heidelberg, 2001.
- [91] J. X. Pu, X. Y. Liu, and S. Nemoto. Partially coherent bottle beams. *Optics Communications*, 252(1-3):7–11, 2005.

- [92] J. X. Pu, S. Nemoto, and X. Y. Liu. Beam shaping of focused partially coherent beams by use of the spatial coherence effect. *Applied Optics*, 43(28):5281–5286, 2004.
- [93] E. Wolf. Unified theory of coherence and polarization of random electromagnetic beams. *Physics Letters A*, 312(5-6):263–267, 2003.
- [94] E. Engler. in preparation.
- [95] A. Prata and W. V. T. Rusch. Algorithm for computation of zernike polynomials expansion coefficients. *Applied Optics*, 28(4):749–754, 1989.
- [96] J. Y. Lee and M. Y. Choi. Optimization by multicanonical annealing and the traveling salesman problem. *Physical Review E*, 50(2):R651–R654, 1994. Part A.
- [97] D. P. Landau and K. Binder. *A Guide to Monte Carlo Simulations in Statistical Physics*. Cambridge University Press, Cambridge, 2000.
- [98] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, 21(6):1087–1092, 1953.
- [99] J. Bewersdorf. *4Pi-konfokale Fluoreszenzmikroskopie mit 1-Photonen Anregung*. Dissertation, Ruprecht-Karls University, 2002.
- [100] N. Bokor and N. Davidson. Generation of a hollow dark spherical spot by 4 pi focusing of a radially polarized laguerre-gaussian beam. *Optics Letters*, 31(2):149–151, 2006.

A Appendix

A.1 Zernike Polynomials

A.1.1 Completeness and Orthogonality

The radial functions $R_n^m(r)$ of the Zernike polynomials are defined in equation (2.38). They are normalized to $R_n^m(1) = 1$. It is shown in [64] that this set of functions is unique in the sense that it a) is orthogonal for the interior of the unit circle, b) contains only polynomials that are invariant in form with respect to rotations of axes about the origin $x = y = 0$ and c) contains one polynomial for each permissible pair of values of n and m . The orthogonality relation for the radial functions $R_n^m(r)$ is:

$$\int_0^1 R_{mn}(r)R_{m'n'}(r)r dr = \frac{1}{2(n+1)}\delta_{nn'}. \quad (\text{A.1})$$

It can be derived regarding the $R_n^m(r)$ as functions obtained by orthogonalizing the sequence r^m, r^{m+2}, \dots with the weighting factor r over the range $[0, 1]$ [64]. The set of functions Z_n^m (defined in equation (2.37)) contains $(n+1)(n+2)/2$ linearly independent polynomials of degree $\leq n$. Therefore, every monomial $x^i y^j$ ($i \geq 0, j \geq 0$) and consequently every polynomial can be expressed as a linear combination of a finite number of Z_n^m 's. By Weierstrass's theorem on approximations by polynomials it then follows that the set of Zernike functions is also complete. The orthogonality of $Z_n^m(r, \phi)$ is expressed by:

$$\int_0^1 \int_0^{2\pi} Z_n^{\pm m}(r, \phi)Z_{n'}^{\pm m}(r, \phi)r dr d\phi = \delta_{mm'}\delta_{nn'} \cdot \begin{cases} \frac{2\pi}{\sqrt{2(n+1)}} & m = 0 \\ \frac{\pi}{\sqrt{2(n+1)}} & m > 0 \end{cases} \quad (\text{A.2})$$

A.1.2 Efficient Calculation of Zernike Polynomials

For the efficient computation of many Zernike polynomials over the aperture the following recurrence relations were used [95]:

$$\begin{aligned}
R_{mm}(r) &= r^m \\
R_{m,m+2}(r) &= R_{mm}(r)((m+2)r^2 - m - 1) \\
R_{m,n+2}(r) &= \frac{n+2}{(n+2)^2 - m^2} \\
&\quad \left\{ \left[4(n+1)r^2 - \frac{(n+m)^2}{n} - \frac{(n-m-2)^2}{n+2} \right] R_{mn}(r) - \frac{n^2 - m^2}{n} R_{m,n-2}(r) \right\}
\end{aligned} \tag{A.3}$$

Using these relations recursively, the radial function $R_n^m(r)$ can be computed very efficiently. Indeed, the computational effort is lower than evaluating the nonrecursive equation (2.38) directly except for low indices (n, m) . The direct evaluation also suffers from a poor machine precision from certain indices (n, m) on. Furthermore, while computing a specific radial function at a position r , the values of many radial functions with lower indices are also computed for this position. Therefore, if all radial functions up to a certain index pair (n, m) and for a specific position r are needed, the calculation is particularly fast if the functions are computed recursively using a table which keeps track of former results. Together with another table which keeps the values of $\sin(m\phi)$ and $\cos(m\phi)$, the integration over the polynomials \tilde{Z}_j , which are derived from Zernike polynomials, can be performed efficiently in equation (2.45).

A.2 Singular Value Decomposition

During the singular value decomposition, a $(m \times n_Z)$ matrix $\tilde{\mathbf{M}}^1$ is factorized in the form:

$$\tilde{\mathbf{M}} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T, \tag{A.4}$$

where \mathbf{U} is a $(m \times m)$ unitary matrix, the matrix $\mathbf{\Sigma}$ is $(m \times n_Z)$ with nonnegative numbers on the diagonal and zeros off the diagonal, and \mathbf{V}^T is the transpose of a $(n_Z \times n_Z)$ unitary matrix. The square of the diagonal elements of $\mathbf{\Sigma}$ are the eigenvalues of $\tilde{\mathbf{M}}^T\tilde{\mathbf{M}}$ in descending order. The latter matrix is symmetrical and possesses only nonnegative eigenvalues. The elements on the main diagonal of $\mathbf{\Sigma}$ are called singular values σ_i . The factorization of equation (A.4) can always be done, even for singular matrices and the singular values are uniquely determined [69]. If the column vectors of \mathbf{U} are denoted by \mathbf{u}_i and the column vectors of \mathbf{V} by \mathbf{v}_i , the factorization of $\tilde{\mathbf{M}}$ allows us to write:

$$\begin{aligned}
\tilde{\mathbf{M}}\mathbf{v}_i &= \sigma_i\mathbf{u}_i, \\
\tilde{\mathbf{M}}^T\mathbf{u}_i &= \sigma_i\mathbf{v}_i.
\end{aligned} \tag{A.5}$$

¹The notation of variables in this section is conform with section 2.5.3.

Suppose that the highest index of $\sigma_i \neq 0$ is n . Then $\text{span}\{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ is equal to the range of the matrix $\tilde{\mathbf{M}}$ and $\text{span}\{\mathbf{v}_{n+1}, \dots, \mathbf{v}_{n_Z}\}$ is equal to the nullspace of matrix $\tilde{\mathbf{M}}$. The columns of \mathbf{V} are the normed eigenvectors of $\tilde{\mathbf{M}}^T \tilde{\mathbf{M}}$ [69, 70]. Therefore, $\text{span}\{\mathbf{v}_{n+1}, \dots, \mathbf{v}_{n_Z}\}$ represents an orthonormalized basis of the nullspace of $\tilde{\mathbf{M}}$. An algorithm implementing the singular value decomposition is given in [70].

A.3 Global Optimization Method

The Monte-Carlo method described here is suitable for global optimization problems in many dimensions with many local extrema. Sometimes also referred to simulated annealing, it was also applied e.g. to the famous *traveling salesman problem* [96].

The essence of the method is in an analogy to thermodynamics. At high temperatures, the molecules of a liquid can move freely with respect to one another. When the liquid is cooled down, thermal mobility is lost. During this process the molecules form a macroscopic crystalline order which represents a minimal energy state of the system. An important factor is the speed of cooling. Slow cooling allows the molecules to redistribute while loosing their mobility and a global energy minimum can be achieved. The inherent natural minimization algorithm is based on the Boltzmann probability distribution:

$$P(E) \sim \exp(-E/k_B T), \quad (\text{A.6})$$

describing a system in thermal equilibrium at temperature T realizing an exponential distribution for its energy states E . For very low temperatures, the system is preferably in low energy states but even then there is a small probability to be in a high energy state. Therefore, there is always a chance to escape from a local energy minimum going through higher energy levels to a more favorable one.

These global optimization properties are exploited in the algorithm used in this thesis. It resembles the Metropolis algorithm [97, 98] (1953). In this algorithm, the states of a system are changed with small random deviations. In this way, the evolution of the system has similarities with a random walk or the diffusive path of a particle in solution. However, the system is assumed to change its configuration according to a Boltzmann probability distribution. That means, changes are always accepted if the energy is lowered whereas energy increases are only accepted with an exponential probability distribution.

To further clarify the mode of operation of the algorithm, some terms need to be defined. The configuration of the system here describes a possible pupil function $\mathbf{A}(r, \phi)$ which is represented by a vector of coefficients \mathbf{C} that form the decomposition of the pupil functions in polynomials (see section 2.5.2). A small random change in the configuration is generated by shifting its corresponding coefficient vector by an amount small compared to its length in a random direction. The figure-of-merit (FoM), that maps a measure of the quality of the created PSF to each configuration is equivalent to the negative energy in the concept outlined above. Therefore, the FoM gets maximized if the energy is minimized. The positive control parameter β is connected to the inverse temperature ($\sim 1/T$). The implementation of our algorithm can then be described as follows:

- (1) Choose an initial random configuration described by the coefficient vector \mathbf{C}
- (2) Choose a random small change in the configuration space $\Delta\mathbf{C}$
- (3) Calculate the change in the FoM (ΔM) resulting from changing the actual configuration from \mathbf{C} to $\mathbf{C} + \Delta\mathbf{C}$
- (4) Generate a random number r which is equally distributed in $[0, 1]$
- (5) If $r < \exp(\beta\Delta M)$ accept the change of the configuration to $\mathbf{C} + \Delta\mathbf{C}$ otherwise not
- (6) Change the control parameter β by a small amount so that it changes periodically over the number of iterations
- (7) Keep track of the highest reached FoM M
- (8) If β is maximal perform a search for local extrema with the simplex method
- (9) Repeat the algorithm from step (2)

The outcome of the algorithm is a configuration with a maximized FoM if a large number of iterations is done. Please note that changes that increase the FoMs are always accepted in step (5). Therefore, the system is always transiting to higher FoMs but only sometimes to lower FoMs.

In simulated annealing, it is generally recommended to slowly decrease the temperature (increase β) and to restart the system after a period of decreased temperature. In this algorithm, this is automatically incorporated by periodically changing β . Therefore, at sufficiently low values of β , almost all changes will be accepted in step (5) and the system will perform an almost free random walk which results in a new state unaware of its history. Furthermore, many periods of high values of β are obtained within one optimization run. High values of β lead to a consequent evolution of the system toward the nearest local minimum. Since the random changes Δx do not very accurately lead to the next local or global extrema another local optimization (after a period of high values of β) was performed. The downhill simplex method by Nelder and Mead [85] was chosen. Like the whole algorithm, it only requires evaluations of the FoM, not of derivatives of it. This requires more functional evaluations as e.g. in the method of steepest descent but it is rather robust. A simplex is a geometrical figure consisting of $N + 1$ vertices in N dimensions. The algorithm is able to finally position all the vertices of the simplex close around a local extremum. Finally, the absolute value of the differences in the FoMs between all simplex vertices is smaller than a given tolerance ϵ and the algorithm stops. This method is robust in finding a local extrema if it is started with vertex-positions not far away from an extremum.

Some arbitrary but performance sensitive choices have to be made in the algorithm in steps (2), (6) and (8). Too small random changes Δx make the algorithm clearly inefficient since the number of function evaluations is increased without increasing the convergence speed. Steps larger than the diameter of the typical surrounding of a local extremum will

very often leap across them without getting knowledge of these extrema. A value of the step size that seemed to optimize the performance of the algorithm was chosen. The periodical change of β on the number of iterations t was set to $\beta = a - b \sin(t/T) \geq 0$ with suitable constants a and b . T was set to 10^4 . In this way, the control parameter was first lowered which agrees with a pure random walk. Consequently, the acceptance ratio, which is the ratio of accepted changes in step (5) to the overall number of iterations, was almost 1 during the random walk at the beginning. The control parameter is then increased again finally leading to an acceptance ratio of almost 0. This is similar to a thermodynamic system in its frozen ground state. The total number of iterations in the algorithm is also an arbitrary choice mainly limited by computer time. Most optimization runs took about 1 day on a 2-3 GHz single CPU. At the end of the runtime of the algorithm, only small improvements were observed compared to the initial increase in the FoM. A total acceptance ratio of 0.4 to 0.6 has been observed in the optimizations which indicates a satisfying functioning of the algorithm with 'cold' and 'hot' periods.

A.4 Coherent 3D De-Excitation Patterns

In section 3.3, it was mentioned that de-excitation patterns that exhibit an intensity distribution which is very steep in all three spatial directions for one vectorial component of the intensity cannot be created coherently. This fact is explained by considering a scalar situation. The electric field E can be described by a continuous band limited complex function in the focal area. Therefore, we have one real and one imaginary part of E at every position in space. Surely, the narrowest de-excitation intensity distribution in one direction would be a standing wave. It should feature a focal node in order to enforce the focal intensity zero. We can assume that the standing wave features only values of E in the real part or the imaginary part. In this part, its spatial dependence would be a sinusoidal function with opposite signs at opposite sides of the focus. One would be tempted to use standing waves in every direction so that a three dimensional intensity distribution is formed. However, considering only one part (real or imaginary) of E , opposite signs on opposite sides for every direction cannot be achieved for a smooth function. A whole plane of zero values in this part of E will remain separating the areas with positive and negative sign. Even if the other part of E will feature a different plane with zero values, the whole space around the focus can never be covered completely. The regions without intensity in either of the parts will be the intersection of two planes (both have the focus as common point) which is at least a line. It follows, that a narrow distribution can only be reached in two directions.

However, all three spatial directions can be covered completely, but then opposite signs on opposite sides must be avoided for at least one direction. This can be done by adding a constant offset to a standing wave which can result in either strictly positive or negative values on opposite sides of the focus. However, the narrowness of the corresponding intensity holes will be approximately doubled in comparison to the former case and in general, the lowest order in a radial coordinate of the intensity around the focus will be the fourth instead of the second.

type of de-excitation pattern	practical restriction	global optimization	idealization
<i>linearly polarized light</i>			
along the x-axis	A	0.019	0.032
	B	0.039	
along the y-axis	A	1.557	1.844
	B	2.516	2.791
in the lateral plane	A	0.012	0.021
	B	0.025	
<i>linearly polarized light and circularly polarized light</i>			
along the z-axis	A	0.157	0.216
	B	0.278	0.296
3D	A	0.009	0.012
	B	0.016	
<i>circularly polarized light</i>			
in the lateral plane	A	0.527	1
	B	0.761	1.128

Table A.1: FoMs for the single focus de-excitation patterns and for a distance from the focus of about 50 nm. The values for circularly polarized light were doubled allowing direct comparison with the patterns for linearly polarized light because circularly polarized light covers two intensity components at the same time. The value of the FoM for the idealization of the lateral doughnut in the restriction regime (A) (using circularly polarized light) was normalized to 1. The optimal doughnut distributions for the z-axis and for 3D are very similar for both polarizations and are presented only once. Idealizations were made for the most important doughnut distributions. Restriction regime (A) was idealized by phase-only pupil functions while regime (B) included a variable amplitude pupil function. The de-excitation pattern along the x-axis, in the lateral plane (both with linearly polarized light) and the 3D de-excitation pattern show small FoMs.

A.5 Additional Optimization Results

In this section, additional optimization results are shown which allow to obtain more detailed knowledge about the mode of operation and the results of the optimization process. In table A.5, the normalized FoMs of the results of the global optimization and the idealizations in chapter 3 are shown. They correspond to the pupil functions shown in figures 3.2 - 3.9. The corresponding PSFs for the idealization are shown in chapter 3 and the PSFs for the results of the global optimization are shown here in figures A.1, A.2 and A.3. The results before idealization have a significantly lower FoM than for the idealized pupil functions. Inhomogeneities are also visible but their shapes are nevertheless quite similar.

Finally, figure A.4 shows the intensity components of several multifocal de-excitation patterns from chapter 4 separately. As expected, using circularly polarized light, the I_x and I_y component are equal up to a rotation of 90° about the z-axis while the I_z component shows broader and weaker patterns.

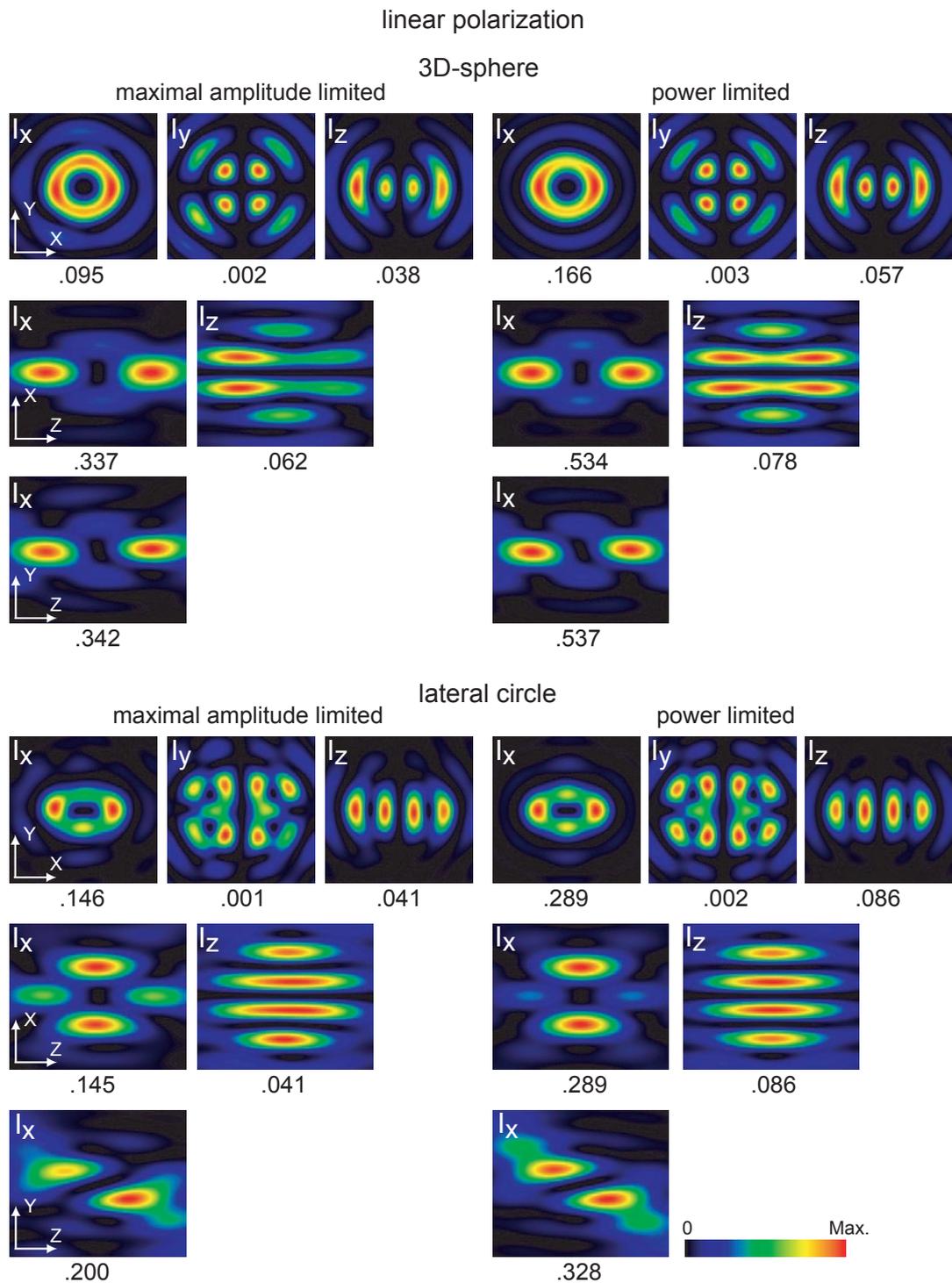


Figure A.1: Characterization of de-excitation patterns corresponding to the pupil functions in figures 3.2 and 3.6. Linearly polarized light is used. Intensity component distributions are shown for relevant 2D cuts of the effective PSF. The 3D de-excitation patterns are shown in a) and the lateral de-excitation patterns are shown in b) for both regimes, (A) and (B).

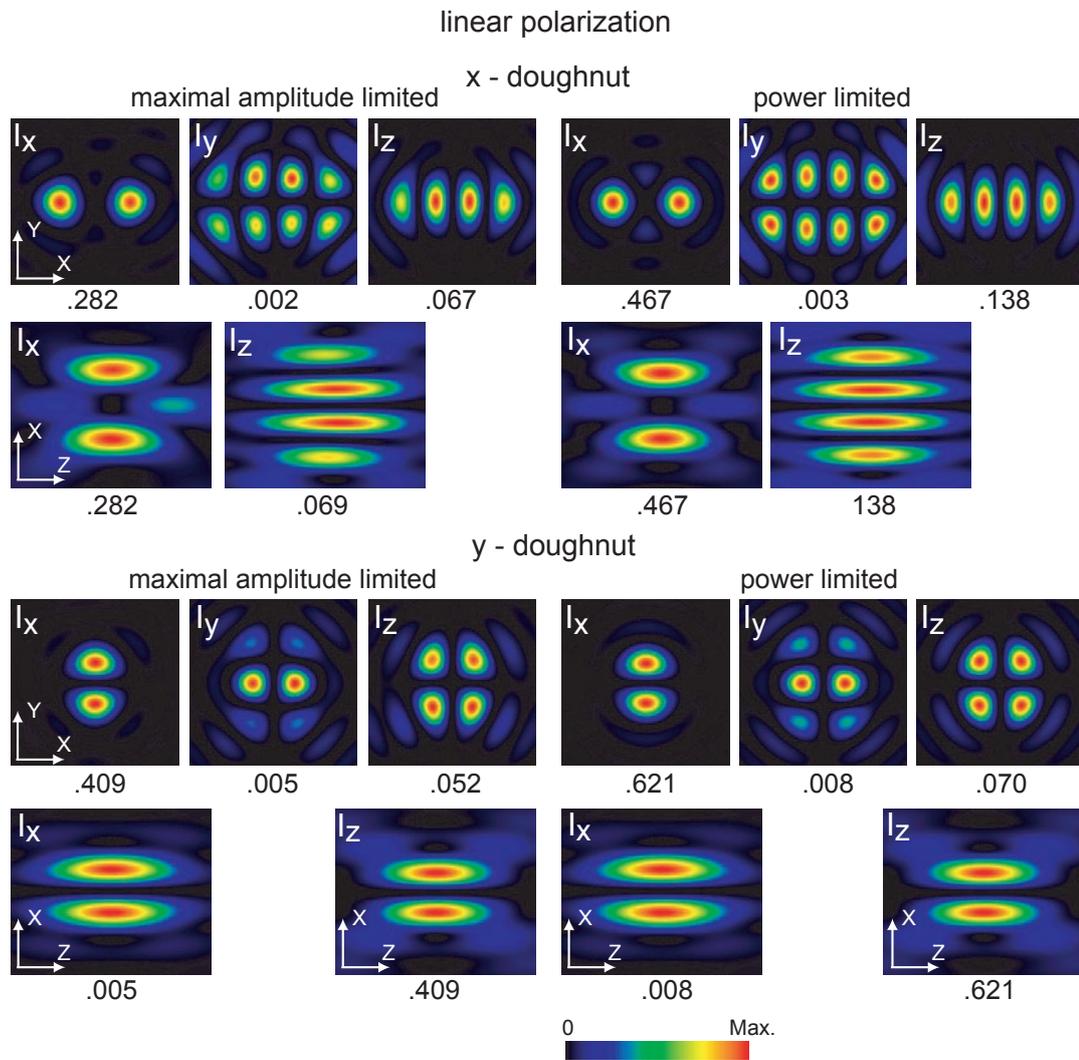


Figure A.2: Extension to figure A.1. In a), the de-excitation pattern along the x-axis is shown and in b), the de-excitation pattern along the y-axis is shown.

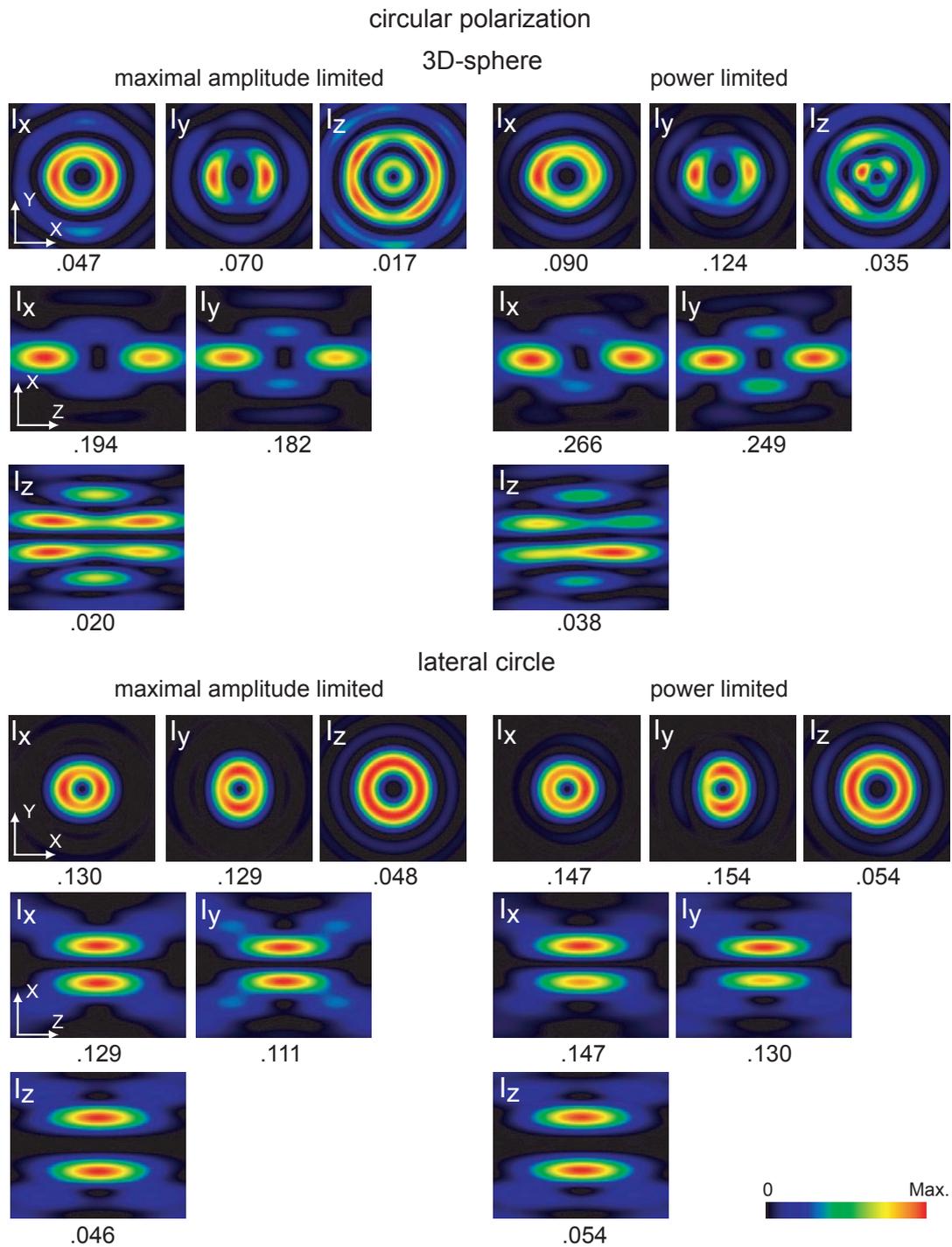


Figure A.3: Characterization of de-excitation patterns corresponding to the pupil functions in figures 3.3 and 3.7. Circularly polarized light is used. Intensity component distributions are shown for relevant 2D cuts of the effective PSF. The 3D de-excitation patterns are shown in a) and the lateral de-excitation patterns are shown in b) for both regimes, (A) and (B).

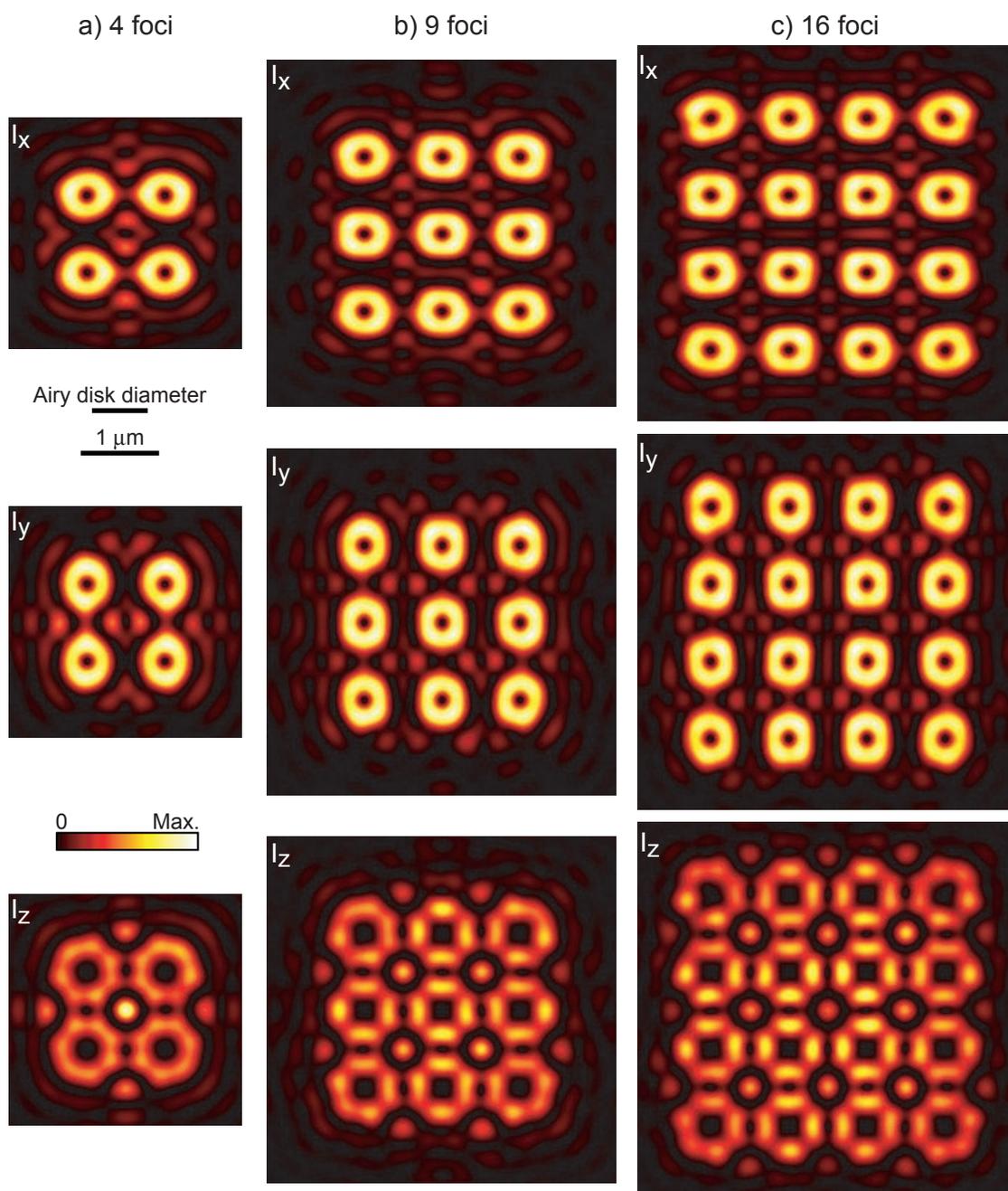


Figure A.4: Intensity component distributions for multifocal RESOLFT microscopy. Using circularly polarized light, the I_y component distributions is given by a rotation of the I_x component distribution. The cases of 4, 9 and 16 foci are shown.

A.6 De-Excitation Patterns for 4Pi-Microscopes

4Pi-microscopy is based on the coherent excitation and/or detection by two opposing objective lenses [13] (also see [99, 21]). Therefore, a focusing angle is synthesized that comes closer to the solid angle of 4π . In 4Pi-microscopy, the axial resolution can be improved 4-7 fold over the resolution of confocal microscopy [14]. Due to the fact that the achievable semiaperture angle α is considerably smaller than 90° , the main maximum is accompanied by 1-2 axial sidelobes at either side.

In a combination with STED, the enlarged aperture was used for the creation of axial, steep de-excitation patterns and featured superior axial resolution [37, 54]. A resolution of $\lambda/33$ along the optic axis could be obtained with infrared light $\lambda_{\text{STED}} = 760\text{nm}$. In the de-excitation pattern, the intensity zero at the focus was formed by destructive interference of the two interfering wavefronts at this point. When using two unaberrated wavefronts with a phase difference of π resulting in destructive interference at the focus, the intensity mainly forms a doughnut distribution along the optic axis. The steepness of this doughnut intensity distribution is however increased if compared to the case of a single objective lens. While with $\tilde{A}_{3D,cl}$ and a single objective lens only a lowest order of z^4 close around the focus can be reached, the de-excitation pattern used in the 4Pi-STED experiment features a lowest order of z^2 . Therefore, the axial direction in 4Pi-RESOLFT microscopy can be squeezed as efficient as the lateral directions in single objective lens RESOLFT microscopy. The used de-excitation pattern describes the best approximation to a standing wave in the axial direction and further improvement in the steepness is only expected including a variable amplitude of

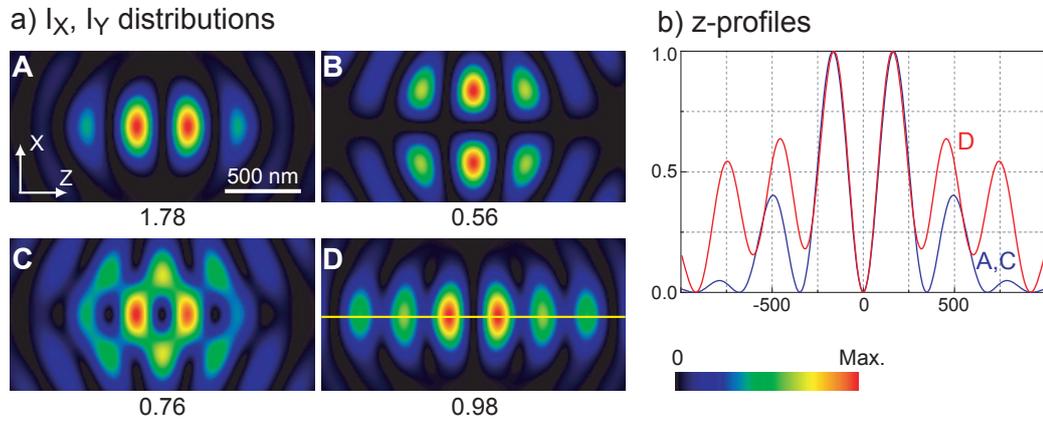


Figure A.5: a) zx-cuts showing the I_x (or I_y) components of de-excitation patterns for 4Pi-microscopy. A: $\tilde{A}_{A,4Pi-1,2}$, B: $\tilde{A}_{B,4Pi-1,2}$, C: $\tilde{A}_{C,4Pi-1,2}$ with $\alpha = 5/18\pi$ and D: $\tilde{A}_{D,4Pi-1,2}$ with $r_D = 0.8$ and $\alpha = \pi/2$. The maximal intensities are drawn below the images. For this normalization, the constructive interference of unaberrated wavefronts would feature a maximal, focal intensity of 2. In the lateral direction, image B is slightly narrower than image C. The corresponding zy-cut for B would yield an identical image while C would not show any intensity along the y-direction. The z-profiles shown in b) are normalized to equal maximum. The profile from D, which does not feature zero intensity points on the optic axis near the focus, is comparable to A and C.

the pupil functions of both lenses.

In this section, some useful doughnut distributions that are suitable for the combination of 4Pi- and RESOLFT microscopy are given. They are mainly motivated by the results of the optimization in chapter 3.

It is known that 4Pi-microscopy cannot extend the band of transmitted frequencies in the lateral direction over the case of a single objective lens [21]. Therefore, the results of this thesis can be directly transferred to this case. The implementation of the pupil function $\tilde{A}_{\text{Lat},c}$ at both objective lenses will however feature the advantage that the local intensity in the focal plane will be twice as large as in the case of illumination through only one lens with equal power, an effect well known for 4Pi-microscopes.

It was shown in this work that a three dimensional doughnut distribution in one intensity component cannot be created coherently (see appendix A.4). Therefore, the incoherent combination of the lateral doughnut distribution created by $\tilde{A}_{\text{Lat},c}$ and the axial doughnut distribution from [37], both using circularly polarized light, seems promising in 4Pi-RESOLFT microscopy.

However, efficient distributions in the zx - or zy -directions are also possible which allow another incoherent combination of two patterns that delivers optimal 3D resolution.

The de-excitation patterns, which are introduced in the following, are characterized by a pair of pupil functions describing the state of the incoming light beam at the two objective lenses before focusing. The notation of this pairs will be $\tilde{A}_{4\text{Pi}-1,2}$.

The pupil function pair used in [54] is defined by:

$$\begin{aligned}\tilde{A}_{\text{A},4\text{Pi}-1}(r, \phi) &= 1, \\ \tilde{A}_{\text{A},4\text{Pi}-2}(r, \phi) &= -1.\end{aligned}\tag{A.7}$$

It describes the steep de-excitation pattern along the optic axis. The pupil function pair used for laterally steep de-excitation patterns is given by:

$$\begin{aligned}\tilde{A}_{\text{B},4\text{Pi}-1}(r, \phi) &= \exp(i\phi), \\ \tilde{A}_{\text{B},4\text{Pi}-2}(r, \phi) &= \exp(i\phi),\end{aligned}\tag{A.8}$$

which corresponds to the use of $\tilde{A}_{\text{Lat},c}$ at both lenses and constructive interference in the focal plane. The de-excitation patterns for the zx -direction is defined by the pair:

$$\begin{aligned}\tilde{A}_{\text{C},4\text{Pi}-1}(r, \phi) &= \exp(i \operatorname{sgn}(r \cos \phi)\alpha), \\ \tilde{A}_{\text{C},4\text{Pi}-2}(r, \phi) &= -\exp(-i \operatorname{sgn}(r \cos \phi)\alpha),\end{aligned}\tag{A.9}$$

with a parameter α . α can be used to balance the distribution of power in the two directions. $\alpha = 0$ corresponds to the use of $\tilde{A}_{\text{A},4\text{Pi}-1,2}$ and $\alpha = \pi/2$ corresponds to the use of $\tilde{A}_{\text{Y},1}$ in constructive interference mode. $\alpha = \pi/4$ marks an equal power distribution into both modes. The zy -directions can be covered easily by rotating the system by 90° around the optic axis. In [54], a modified pupil function pair was given that could smear out the modulation of the de-excitation pattern along the optic axis. This allowed additionally the efficient suppression of the excitation sidelobes whose height is an important factor in 4Pi-microscopy [13]. A generalization of the used pupil functions for this case is defined by:

$$\begin{aligned}\tilde{A}_{D,4\text{Pi}-1}(r, \phi) &= \begin{cases} 1 & r \leq r_D \\ \exp(i\alpha) & r > r_D \end{cases}, \\ \tilde{A}_{D,4\text{Pi}-2}(r, \phi) &= -\tilde{A}_{D,4\text{Pi}-1}(r, \phi).\end{aligned}\tag{A.10}$$

An axial de-excitation pattern with increased height at the positions of the excitation sidelobes can be reached. Different trade-offs between axial doughnut narrowness and side-lobe reduction can be achieved by adjusting the parameters r_D and α . Especially, $\alpha = \pi/2$ seems promising.

The resulting doughnut distributions of the pupil function pairs which are defined in this section are shown in figure A.5. They were calculated for a 4Pi-setup with a triangular cavity [99]. However, the necessary changes for the implementation in a rectangular cavity are straightforward. The pupil function pair $\tilde{A}_{C,4\text{Pi}-1,2}$ is indeed very efficient in the zx -plane and the pupil functions $\tilde{A}_{D,4\text{Pi}-1,2}$ with the chosen parameter values are appropriate to suppress the 4Pi-sidelobes while showing only a moderate decrease in the axial steepness at the focus. For normalized maximal intensity in the sample, the doughnut distribution of the last pupil function pair does not show any loss in axial steepness. In this section, phase-only pupil functions were given. However, a moderate, further increase in the steepness in the patterns can be expected by optimizing the amplitude similar to the optimization in section 3.2.2.

Acknowledgment

I would like to thank all people who have directly or indirectly contributed to the success of this thesis.

Prof. Dr. Stefan W. Hell for proposing this fascinating project, introducing me to the subject and providing excellent working conditions. I am also thankful for valuable discussions and his constant encouragement to follow new ideas.

Prof. Dr. Josef Bille for his interest and for agreeing to be the second referee.

Dr. Andreas Schönle for his excellent advice. Numerous discussions and his valuable ideas have significantly contributed to the success of this thesis. I appreciated his expertise in computer programming and value his careful proof-reading of the manuscript. His software 'Inspector' helped in the analysis of the results.

Dr. Katrin I. Willig for her constant feedback about the experimental applicability of my results. I always enjoyed our cooperation between lab and blackboard.

Claas v. Middendorff, Arnold Giske, Michael Hofmann, and Dr. Mariano Bossi for proof-reading the manuscript and giving very helpful comments.

Hartmut Seebesse for helping with 3D graphics.

Peggy Findeisen for proof-reading the manuscript and especially for her moral support.

The whole NanoBiophotonics department including former members and guests and the roommates from my WG for being what they are - good company.

Last but not least, I want to thank my parents for making my education possible and always being there for me.

List of publications

Parts of this thesis have been published as follows

1. G. Donnert, J. Keller, R. Medda, A. Andrei, S.O. Rizzoli, R. Lührmann, R. Jahn, C. Eggeling, and S.W. Hell. Macromolecular-scale resolution in biological fluorescence microscopy. *PNAS* 103(31): 11440-11445, 2006.
2. K.I. Willig, J. Keller, M. Bossi, and S.W. Hell. STED microscopy resolves nanoparticle assemblies. *New. J. Phys.* 8: 106, 2006.
3. M. Dyba, J. Keller, and S.W. Hell. Phase filter enhanced STED-4Pi fluorescence microscopy: theory and experiment *New. J. Phys.* 7: 134, 2005.
4. A. Schönle, J. Keller, B. Harke, and S.W. Hell. Diffraction Unlimited Far-Field Fluorescence Microscopy. In *Handbook of Biological Nonlinear Optical Microscopy*, B. Masters and P. So, editors, Oxford University Press. (in press) 2007