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

Put forward by

Matthias Reuss, Master of Arts
Born in Munich

Oral examination: December 15, 2010

Simpler STED setups

Referees:

Prof. Dr. Stefan W. Hell
Prof. Dr. Rainer H. A. Fink

"L'essentiel est invisible pour les yeux"

Antoine de Saint-Exupéry, Le Petit Prince.

Abstract

This work discusses rugged technical simplifications for STED nanoscopy and methods related to it (RESOLFT). STED is a type of laser scanning microscopy that uses stimulated emission in order to inhibit spontaneous fluorescence from outer parts of the focal spot. This deliberate transient off-switching of fluorophores permits even densely packed fluorescent molecules to be spatially isolated from the rest; and because isolated molecules can be registered sequentially — no matter how close they are — the resolution is increased beyond the diffraction limit.

For switching, STED asks for a separate inhibition beam modified in a way that it suppresses fluorescence, but only in the peripheral regions of the focal spot. The requirements of a second beam path and additional means for beam shaping have kept STED from gaining the currency it deserves, although high resolution is in huge demand, generally.

Here, it is discussed how contemporary STED setups can be simplified radically; in fact up to a point where it is possible to upgrade existing laser scanning systems with a co-aligned STED-beam, simply by adding an element about the size of an optical thin-film filter. As the apparent need for a separate beam path ceases to exist, STED becomes simpler and more reliable. At the same time, the performance is the same as for a traditional STED-setup.

Zusammenfassung

Die vorliegende Arbeit erörtert Möglichkeiten zur technischen Vereinfachung von STED Nanoskopen und verwandten Verfahren (RESOLFT). STED ist eine Weiterentwicklung traditioneller Laserrastermikroskopie und basiert auf der Unterdrückung von spontaner Fluoreszenz im äußeren Bereich des Anregungs-Beugungsflecks mittels stimulierter Emission. Das gezielte transiente Ausschalten von Farbstoff-Molekülen hat zur Folge, dass Objekte von sehr nahe liegenden anderen Objekten räumlich getrennt werden können; das wiederum ermöglicht sequentielles Auslesen dieser Objekte und damit Auflösungen unterhalb der Beugungsgrenze.

Für das Ausschalten wird jedoch ein zweiter Strahl erforderlich, der für die Unterdrückung der Fluoreszenz am Rande der Anregungs-Beugungsflecks verantwortlich zeichnet. Das erfordert einen unabhängigen Strahlengang, da der STED-Strahl so verändert werden muss, dass er nur am Rand wirkt. Nun ist generell der Bedarf an Systemen mit verbesserter Auflösung sehr groß, allerdings stehen oben genannte Erfordernisse der großflächigen Verbreitung von STED im Wege — bis jetzt.

Hier werden nun Möglichkeiten zur radikalen Vereinfachung gegenwärtiger STED-Aufbauten diskutiert. In der Tat wird gezeigt, dass auch bestehende Laserrastermikroskope durch Bereitstellen einer STED-Quelle und den Einbau eines Elements etwa von der Größe eines Dünnschichtfilters zu STED-fähigen Systemen umgerüstet werden können. Weil die Notwendigkeit eines weiteren Strahlengangs entfällt, werden diese Systeme robuster und zuverlässiger, gleichzeitig aber ist ihre Leistungsfähigkeit und die erzielbare Auflösung in keinster Weise vermindert.

Contents

List of Figures

Glossary of symbols and acronyms

$[\rho, \phi]$ Polar coordinates in the back focal plane (pupil plane)
$[h, v]$ Cartesian coordinates in the back focal plane (pupil plane)
α Opening angle of objective lens
Γ Birefringent phase or retardance
$\imath$ Imaginary unit
λ_{exc} Excitation wavelength
λ_F Wavelength of fluorescence
λ_{STED} Wavelength of the STED-beam
$\mathbf{B}$ Magnetic field
$\mathbf{D}$ Displacement field
$\mathbf{E}$ Electric field
$\mathbf{E}_{STED}$ Electric STED-field
$\mathbf{H}$ Magnetization
$\mathbf{J}$ Jones vector
$\mathbf{k}$ Wave vector
$\mathbf{M}$ Jones matrix
$\mathbf{R}$ General rotation matrix
$\mathbf{r} = [x, y, z]$ Cartesian coordinates in the focal plane (x, y) and or along the optical axis (z)
$\mathcal{F}$ Fourier transform
μ Permeability
μ_0 Permeability of free space
φ Phase
θ Orientation of the optic axis of a wave plate
ε Permittivity
ε_0 Permittivity of free space
A Pupil function
c Speed of light
d Full width at half maximum of a point spread function
f Focal length
h Planck's constant, point spread function

h_{det} Detection point spread function
h_{eff} Effective point spread function
h_{exc} Excitation point spread function
h_{STED} STED point spread function
I_{sat} Saturation intensity of a fluorophore
I_{STED} Intensity of STED-beam
k Wave number, rate constant
N Population of an electronic or vibrational state
n Refractive index, placeholder for a positive integer
n_e Extraordinary refractive index
n_o Ordinary refractive index
NA Numerical aperture
S_n^* A vibrational state
S_0 The electronic ground state
S_1 The ground state of the first excited electronic state
S_n An electronic state
t Time
T_n A triplet state
w Vibrational frequency
A.U. Airy unit
APP Amyloid precursor protein
BFP Back focal plane
CLSM Confocal laser scanning microscope
FWHM Full width at half maximum
GFP Green fluorescent protein
GSDIM Ground-state depletion microscopy
LSM Laser scanning microscope
MOM Molecular orientation microscopy
PALM Photo-activated localization microscopy
PBS Polarizing beam splitter
PSF Point Spread Function
RESOLFT Reversible saturated optical fluorescence transitions
STED Stimulated emission depletion (microscopy)
STORM Stochastic optical reconstruction microscopy
SWP Segmented wave plate

1

Introduction

Antonie van Leeuwenhoek is commonly considered to be the father of microbiology. Not by mere coincidence, his name is also tightly linked to the field of light microscopy. It was van Leeuwenhoek who first put the existing microscopes to biological use, investigating erythrocytes, spermatozoa, protozoa, algae and even bacteria [79].

Other important contributors were the Jannsen family, Robert Hooke and Galileo Galilei [14], but it was not until 200 years later that, starting in the second half of the 19th century, Ernst Abbe put optical microscope design on a solid scientific base [2, 3]. Since then, light microscopy has become an essential research tool and can be found in laboratories all over the world; probably by reason that as human beings we strive to not only analyze things indirectly — we want to perceive them, visually, in full realness.

The success of optical microscopy is in spite of diffraction, which inflicts a limit on the resolution of a microscope, as has already been described by Abbe. A detailed analysis will be provided in chapter 2; for now it shall be sufficient to note that any two points in a specimen separated by a distance that falls below the resolution limit cannot be observed as individual points but rather mingle into one single spot. The amount of diffraction and thus the resolving power is known to be a function of, among other things, the wavelength. This was indeed the reason for the development of microscopes that employ electrons with a deBroglie-wavelength far below the wavelengths of light, because they allow the observation of much finer structures.

How fine? The smallest structures of interest to a cell biologist are single molecules such as proteins and lipids, whose size is in the nanometer-range. But cells exhibit activity on many different lengthscales. Individual molecules work flat out together to form more complex machinery and, step by step, cell organelles, and eventually whole cells. Using an electron microscope [61], resolutions in the realm of single molecules are readily achieved [48], however, the destructive effect of accelerated electrons on

living matter [22] together with their small penetration depth effectively hinder the application of electron microscopy to whole, not to mention living cells. Similar arguments and the limitation to two-dimensional surfaces prevent near field techniques such as Near Field Scanning Optical Microscopy (NSOM) [5] or Atomic Force Microscopy (AFM) [8] from being applied to volumetric specimen. Also, X-ray microscopy [59] in principle suffers from similar effects as electron microscopy.

In contrast, visible light is able to penetrate whole and even living cells almost non-destructively. It can be used for observing not only the fixed *structure* but also the actual *live workings* of cellular machinery. The difference this makes must not be underestimated; life is by definition dynamic, and only by observing the dynamics can the full picture be assembled [64]. Unfortunately, diffraction comes into play again and restricts the resolution of optical microscopes, which means we miss out essentially on most of what takes place below ca. 200 nm. In order to fully exploit the potential of light microscopes a way has to be found to circumvent the diffraction limit; a way that can combine the high resolution of electron microscopy with the live-cell and three-dimensional capabilities of optical far-field techniques.

The first device to reach beyond Abbe's (traditional) diffraction limit was the 4Pi microscope [41, 43]. It makes use of two opposing objective lenses instead of one, thereby generating two counter-propagating fields of focused light. The resulting intensity distribution features very fine structures that lead to a four- to sevenfold increase in resolution along the optical axis, compared to conventional confocal microscopy [31, 42]. It is, however, still a diffraction-limited technique, in that the achievable resolution is tightly linked to the interference of light waves that generate the focal spot.

In 1994, it was realized that the exploitation of switchable molecular states uncouples resolution from the phenomenon of interference. The diffraction barrier had been broken for the first time. **S**timulated **E**mission **D**epletion Microscopy (STED) was the first technique that demonstrated that in principle infinite resolution — *nanosopy* — was possible, even in the far-field and with light in the visible range. Since then, several methods such as the generalization RESOLFT [50], numerous PALM (photo-activated localization microscopy) variants [7,24,27,38,49], STORM (stochastic optical reconstruction microscopy) [88], GSDIM (ground state depletion microscopy) [26] as well as saturated structured illumination microscopy [32] and many more have emerged in the wake of STED, all sharing a common denominator: molecules being switched between two or more (meta)stable states.

Nanoscopy has gone a long way since then, and it has shown its potential in several assays [1, 45, 51, 60, 66, 68, 91, 105, 113]. In 2007, a commercial version of a STED setup has become available, further underlining the need for high-resolution imaging that only fluorescence microscopy can offer non-invasively. While a commercial device is the way to go for the enduser inexperienced in optical engineering, it is necessarily much less flexible than a home-build apparatus and lags behind the state-of-the-art by several years. Recent developments and improvements such as continuous wave-STED [111], three-dimensional high-resolution modes [36, 92], newly developed laser sources [109], and so on may not be easily added. This can be a huge drawback in such a dynamic field as nanoscopy. On the other hand, building a nanoscope requires significant knowledge in optics, electronics and microscopy, while the potential user is most likely educated in medicine or biology and is not willing to invest into acquiring these skills. Especially STED is often perceived as being technically complicated and although this is by far not the whole story, it probably discourages potential users even more.

By countering most arguments concerning the complexity of STED, this work enables many more users to get access to a STED-nanoscope. It will be demonstrated how a powerful STED, or for that matter, RESOLFT setup can be laid out much simpler and robust than before; in fact, even a relatively inexperienced user is now able to assemble a STED setup by himself. Indeed, the upgrade of an existing laser scanning microscope to a STED-nanoscope is also straightforward.

The dissertation itself is divided into two parts. First, a general introduction into fluorescence light microscopy and imaging is given. Then, in the main part, devices are presented that enable said simplification of STED-nanoscopes for two- as well as three-dimensional resolution enhancement, and it is shown that they can be employed robustly, yet with very little effort. Additionally, a new technique is reported that allows the user to investigate the spatial orientation of fluorescent molecules.

2

Theory and basic principles

This chapter is intended to give an insight into how images are formed in light microscopy. General principles of optical microscopy will be discussed, before delving into the principles of non-diffraction-limited nanoscopy, in particular RESOLFT and one of its embodiments — STED. Other nanoscopy techniques are covered as well.

2.1 Standard fluorescence light microscopy

Even the simplest forms of life are made up of an uncountable number of molecules. Still, when only distinct types of molecules are considered, a bacterium of type E. Coli consists of tens of thousands of different proteins, enzymes, lipids and nucleic acids [9]. An important step in acquiring an image of a cell's interior is therefore to come up with a way of separating the molecules of interest from this background. One of the great advantages of fluorescence microscopy is specificity [81], the ability to image only the relevant molecules — even when they are buried amidst a pile of similar but uninteresting material. By selectively attaching fluorescent markers (*fluorophores*) to molecules of interest and by imaging those fluorophores, one can in turn investigate the organization of the cellular components that have been labeled simply by concluding that the distribution of fluorophores resembles the underlying structure [15]. Particular types of cellular entities can be selectively marked with fluorophores using antibodies raised against those molecules (*immunofluorescence*). To this end, the antibody can itself be conjugated with a fluorophore; often, however, a second antibody targeting the first one bears the fluorescent marker. This way the signal is increased over the use of a single antibody. Highly selective tagging can also be performed with special molecules; take phalloidin for example, a protein from the death cap that binds to actin and, conjugated with a fluorophore, can be used to label the actin cytoskeleton. Unfortunately, antibodies and especially phalloidin are usually toxic for living cells and require fixation. Another method is to use genetical

engineering in order to fuse the gene encoding for a protein of interest with the gene for a fluorescent protein, GFP for example [98], cloned from the jellyfish. This induces expression of proteins with fluorescent markers readily attached to them, even in living cells. A researcher can now also follow the making of proteins and their fate — the fluorescent gene can even be passed to the next generation during breeding.

Either way, cellular molecules that have been fluorescently labeled can be imaged with high specificity amidst non-labeled and therefore non-fluorescent surroundings — simply because they emit light. Consequently, before taking a look at image formation in general, the phenomenon of fluorescence will be discussed.

2.1.1 Fluorescence

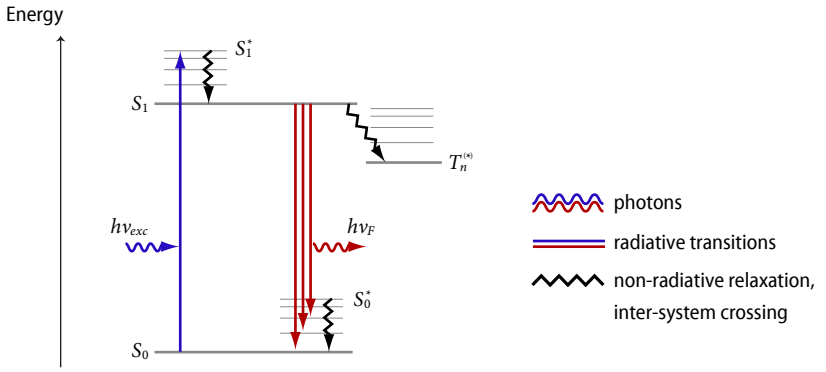

Figure 2.1.: Jablonski-diagram of a fluorophore and the principle of fluorescence.

Fluorescence is the emission of light from a molecule after it has absorbed light with a shorter wavelength. It can be explained semi-classically with the framework of the *Jablonski diagram* [62] (Figure 2.1), which depicts the relevant electronic levels, respectively states, a fluorescent molecule can be found in and transitions between them. Without external perturbances, the molecule is in its lowest energy state or ground state S_0 for most of the time, as stated by the principle of minimum total potential energy. This electronic ground state is usually degenerated into vibrational levels S_n^*. A fluorophore can be brought to the S_1 state by absorption of a photon with wavelength $\lambda_{exc} = c/\nu_{exc} = hc/\Delta E$ which corresponds to the energy gap ΔE between two states, e.g. S_0 and S_1^*. Multi-photon excitation is also possible; here the energy gap is bridged by two or more photons with longer wavelengths simultaneously. In any case, the molecule is transferred to the S_1 state, in this case a higher vibrational level S_1^*

of S_1, depending on the exact value of λ_{exc}. The absorption process itself happens almost instantaneously and within another few picoseconds the molecule relaxes to the basic state of S_1 without radiating, while excess energy is converted to heat. The lifetime of this state is typically three orders of magnitude greater than that of a vibrational state ($\approx$ nanoseconds), so that the following transition, namely fluorescence, originates practically from this state only (*Kasha's rule*). During fluorescence, the molecule falls back into the basic (S_0) or a higher vibrational state S_0^*, releasing energy in the form of a photon with wavelength λ_F. Because energy has been lost during the vibrational relaxations, the emitted photon has less energy than the exciting photon and is therefore red-shifted; in the next section, it will become clear that this *Stokes shift* has important consequences for the use of fluorescent molecules as markers in light microscopy.

Another point worth noting is that λ_F is not clearly defined a priori. Quantum mechanically, the probability for a transition from state A to state B described by two wave functions $\psi_{A/B}$ depends on the vertical overlap of the wave functions: $P_{A\to B} = \langle \psi_B | \boldsymbol{\mu} | \psi_A \rangle$. (Without going in too much detail, the molecular dipole operator $\boldsymbol{\mu} = \sum q_i \mathbf{r}_i$ carries information about the positions $\mathbf{r}_i$ of charges q_i in the molecule, which obviously have a great impact on electronic transitions.) Thus, one can only specify the probability that the fluorophore returns to a particular vibronic state of S_0. This, together with inhomogeneous broadening in solvents, gives the emission and absorption spectra their continuous appearance.

2.1.2 Standard fluorescence microscopy techniques

It has been presented above how, by absorbing a photon of wavelength λ_{exc}, a fluorophore enters an excited state and subsequently emits a redder photon with wavelength λ_F. It has also been mentioned already that fluorophores specifically attached to molecules of interest can help to generate contrast; putting those two together, this section aims to explain how exactly labeled proteins are imaged in a fluorescence microscope.

The simplest conceivable fluorescence light microscope is an apparatus flooding the sample with excitation light from one side while collecting fluorescence from the other side with an objective lens that images the emitted light onto a camera or into the eye of the observer. Notably, the Stokes shift facilitates the separation of residual excitation light from the red-shifted fluorescence signal by means of dichroic filters. Instead of having separate paths, more advanced designs use the same objective lens for both excitation and observation [83]. This has several advantages, such as intrinsic alignment of excitation and emission, and, most importantly: the majority of the excitation light that has not been absorbed continues to travel away from the observer. While those *epifluorescence* microscopes

(*epi-*, greek: upon, above) legitimately claim their spot in life sciences, certain remaining drawbacks can be eliminated with *confocal laser scanning* microscopes.

Because nearly the whole specimen is submerged in excitation light, the image of an epifluorescence microscope is heavily blurred by out-of-focus information. Confocal microscopy was first described in the 1950s [4, 73, 81, 106] and aims at reducing this unwanted stray light and accordingly enables the sectional imaging of thick samples[1]. Here, the excitation beam is *focused* using the objective lens; the intensity in the focal spot — which is actually a focal volume as will soon become clear — is high enough to excite a significant number of fluorescent molecules within. Nevertheless, the focusing cone (light traveling to and from the focal plane) also causes molecules to emit. This out-of-focus signal compounds in thick specimen and causes fine details of features in the focal plane to be lost in background light. This is similar to epifluorescence where the whole specimen baths in excitation, alone, this unwanted light can now be eliminated using *confocality*. For this, fluorescence is imaged onto a diaphragm or *pinhole* placed in a conjugated focal plane. Since light from the focal plane is again at focus (confocal, *con-*, Latin: together) at this plane, it passes straight through the pinhole onto the photodetector. In contrast, light from any other plane is not at focus at the pinhole and thus only a very small portion of this light is able to reach the detector. This spatial filtration successfully excludes out-of-focus emission from the final image and thus enables one to record sections that only include light from the focal plane and close to it. The effect may vary, depending on the signal to noise ratio, specimen thickness and background; a smaller pinhole can increase sectioning, but a pinhole too small will probably yield insufficient signal.

An important consequence is that only fluorescence signal originating directly from inside the focal volume can pass through the conjugate aperture. This means that not only light from other planes, but also light originating from other locations *in* the focal plane is excluded. The remedy is to raster-scan the focal volume across the specimen in order to assemble a full picture of the sample. This stands in contrast to epifluorescence where the complete image is recorded at once and is simply the price to pay for sharper, more detailed images. Devices of this kind are commonly called *confocal laser scanning* (LSM) microscopes.

Confocal microscopy is now a routine technique in laboratories all over the world and permits investigations not possible until a few years ago, due to its three-dimensional sectioning capability combined with the possibility to produce contrast by labeling with fluorescent markers. Nevertheless,

1 Samples with an axial extension much larger than the depth of field are considered thick.

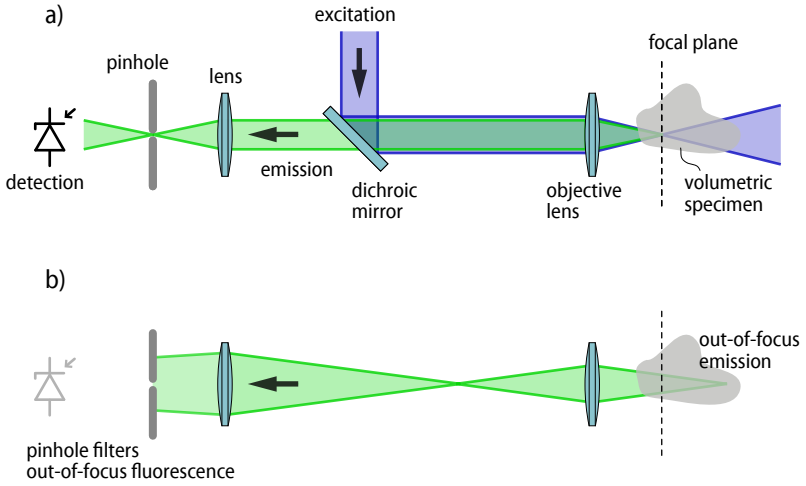

Figure 2.2.: The confocal principle. a) fluorophores are excited by focused light. The emission is imaged onto a pinhole that allows only light originating from the focal plane to pass through. Fluorescence emission from an unfocused region in the specimen is blocked by the pinhole as shown in b).

the resolution of this technique is poor compared to electron microscopy which, on the other hand, lacks life-cell capacity and involves tedious protocols for sample preparation. Thus, high resolution optical schemes such as STED are increasingly called on to investigate cellular workings on the smallest spatial dimensions. Although STED itself is by no means based on confocality, most STED-setups are a derivation of a laser scanning microscope (without the pinhole).

However, before progressing into the discussion of STED and similar techniques, the term *resolution* itself will be illuminated firstly.

2.1.3 The resolution problem in image forming devices

The resolving power of even a perfectly constructed microscope is ultimately limited by diffraction [2, 3]. As rays emitted from an point source in the focal plane of a microscope pass the aperture of the objective lens they are diffracted. This causes a ring-shaped interference structure in the image plane, the *Airy disk*. The resulting three-dimensional intensity distribution is called *point spread function*[1] (PSF), in this case detection PSF

1 Somewhat ambiguous, the cut section of the PSF that lies in the focal plane is called Airy disk, but often also two-dimensional, xy, or lateral point spread function. When talking about the three-dimensional diffraction pattern, the term PSF solely applies.

$h_{det}(\mathbf{r})$, and describes the probability that a photon emitted at $\mathbf{r}$ reaches the detector after having been diffracted by the lens and the rest of the system. In other words, the detection PSF is the impulse response of a microscope.

Not surprisingly, diffraction occurs not only when imaging *from* object space, but also when trying to focus a laser beam *into* object space. Just as the image of a point source gets blurred when viewed through a microscope, focusing of a laser beam in a laser scanning microscope does not yield a perfect spot; it yields a distribution that is spread out according to the excitation PSF h_{exc} of the system. $h_{exc}(\mathbf{r})$ describes the probability that molecules located at $\mathbf{r}$ are in fact excited. As a result, these fluorophores emit fluorescence light, which in turn is detected in accordance with the detection PSF h_{det}. Technically, the resulting effective point spread function [81] is the product of the excitation and the detection PSF:

$$h_{eff}(\mathbf{r}) = h_{exc}(\mathbf{r}) \cdot [h_{det}(\mathbf{r}) \otimes p(\mathbf{r})], \tag{2.1}$$

where, in confocal microscopy, the action of the pinhole is accounted for by the convolution of the detection PSF with the pinhole function

$$p(x, y, z = 0) = \begin{cases} 1 & \text{if } \sqrt{x^2 + y^2} < r_{pinhole}, \\ 0 & \text{otherwise.} \end{cases}$$

$r_{pinhole}$ is the pinhole radius. Again, h_{eff} is the response for a point source. In practice, however, a meaningful specimen is more than a point source. Fortunately, any arbitrary object o can be pictured as the sum of many point sources, or delta functions $\delta(\mathbf{r})$, weighted with a factor o allowing for emitters with different brightness:

$$\int_{\infty} o(\mathbf{r}')\delta(\mathbf{r} - \mathbf{r}')\ \mathrm{d}\mathbf{r}' \tag{2.2}$$

In the image i, every point $\delta(\mathbf{r})$ with brightness o will simply become an h_{eff}, still with brightness o. The fact that every infinitely small point gets treated by the imaging system according to the effective PSF is reflected mathematically by the convolution operation '$\otimes$', which simply speaking puts the point spread function in place of every delta function it acts upon[1]. Thus, the image i of $o(\mathbf{r})$ is

$$i(\mathbf{r}) = \int_{\infty} o(\mathbf{r}')h_{eff}(\mathbf{r} - \mathbf{r}')\ \mathrm{d}\mathbf{r}' \tag{2.3}$$

$$= o(\mathbf{r}) \otimes h_{eff}(\mathbf{r}). \tag{2.4}$$

Equation 2.4, together with Equation 2.1 fully describes the process of image formation mathematically.

1 Fluorescence imaging systems are assumed to have linear properties, that is, the image of a sum of objects is the sum of the individual images of each respective object, up to a weighting factor. In practice, the PSF may depend on, for example, the position in the object plane.

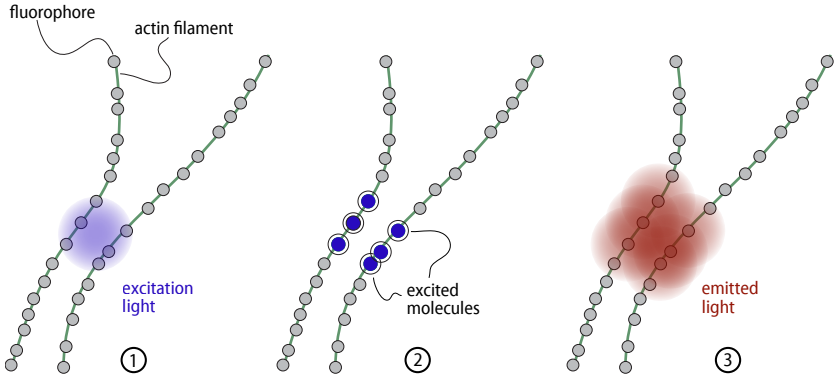

Figure 2.3.: Scanning a sample with a diffraction-limited beam. Features smaller than the PSF can not be separated.

Because the PSF has a finite width, the images of point sources separated in object space can not necessarily be distinguished in image space. The situation is exemplarily illustrated in Figure 2.3 for actin fibers labeled with fluorescent markers. At a specific location during the scan, several molecules are excited by h_{exc} and subsequently emit after excitation. The image of each single molecule gets blurred in agreement with the detection PSF h_{det}. Although in the picture the scanner aims for the gap between the actin fibers, fluorophores from both fibers are excited nevertheless and their concerted contribution blurs out the gap in the final image.

It is obvious that the spatial extension of the PSFs determines the resolving power and thus plays a central role in microscopy. The dimensions of a PSF essentially depend on two parameters, the wavelength λ of the light being imaged and the half-cone opening angle α of the lens, which is the maximum angle over which light can enter or exit the lens. Clearly, the wavelength dependence comes about because shorter wavelengths can produce smaller interference structures. The opening angle plays a role, because the greater the opening angle, the more information[1] can be gathered from the spherical wave emitted by the point radiator. Altogether, the full width at half maximum (FWHM) of the PSF (be it h_{exc} or h_{det}) amounts to

$$d_{x,y} = \frac{\lambda}{2n \sin \alpha} \tag{2.5}$$

1 For Abbe, "information" was tantamount to "diffraction order." When a lens can collect higher diffraction orders (by means of a larger opening angle), the projected spot is smaller, because more diffraction orders can interfere in the image plane [81].

in the focal plane, and

$$d_z = \frac{\lambda}{2n \sin^2 \alpha/2} \tag{2.6}$$

for the width of the PSF along the optical axis. Of course, λ is one of λ_{exc} or λ_F depending on the PSF in question.

In Equations 2.5 and 2.6 the opening angle α is modified by the index of refraction n of the immersion medium between object and lens. The quantity $n \sin \alpha$ is usually condensed into a single parameter called numerical aperture NA. In practice, optical microscopes work with wavelengths in the visible range ($\approx 400 - 800$ nm)[1], the opening angle α is limited by the maximal lens diameter achievable without severe optical aberrations, and n is smaller than 1.7 for practical immersion media[2]. This limits the highest possible resolution (i.e. the FWHM of h_{eff}) to about 200 nm in the focal plane and 600 nm along the optical axis. Hence, point sources closer together than 200 nm in the focal plane or 600 nm along the optical axis can not be imaged as individual objects. The consequence are blurred pictures lacking any detail below 200 nm or 600 nm, respectively.

Remarks

It is sometimes helpful to picture the process from a frequency-space point of view, where the convolution theorem states that

$$I(\mathbf{k}) = H_{eff}(\mathbf{k}) \cdot O(\mathbf{k}). \tag{2.7}$$

Uppercase letters represent the fourier-transformed image, PSF, and object distribution, respectively. H_{eff} is called *optical transfer function* or *OTF* and describes the weight with which particular frequencies are transferred from the object to the image. In particular, for high frequencies $|H_{eff}|$ is nearly zero and thus those high frequencies will not be present in the image — which is simply another form of the statement that features with small dimensions can not be imaged. The resolution is limited.

It is often stated that, if the image as well as the PSF is known, one can simply solve Equation 2.4 for $o = i \otimes^{-1} h_{eff}$ to fully obtain the object including purportedly sub-resolution features (*deconvolution*). Here it is

1 Of course smaller wavelengths will lead to better resolution, as is the case for X-ray microscopy. However, smaller wavelengths also bring about sample damage and impede live-cell imaging.

2 An immersion medium with a refractive index of 1.74 is for example methylene iodide (Cargille-Sacher Laboratories Inc., USA). Most immersion oils actually have a refractive index around 1.5. Solid immersion lenses (SIL) use bulk media with refractive indices up to 3.4 [94] but are rather impractical because a solid object is hard to fully contact with the sample. SILs are used in photolithography, for instance.

important to understand that neither the image i nor the experimental PSF h_{eff} can be measured without noise. Typically, the fluorescence signal is comparatively weak, so that the inherent uncertainty connected with detecting a low number of photons plays a huge role. Further sources are for instance electronic noise in the detector and stray light. In any case, the weak detection events corresponding to high frequencies in H_{eff} are drowned in faulty signal — but the high-frequency components are exactly the components that have to be amplified if the resolution is to be enhanced. Thus, noise will be amplified as well and resolution enhancement simply by reversion of Equation 2.4 is spoiled in practice. Nevertheless, more sophisticated deconvolution schemes are useful for enhancing (denoising) microscopic images, however, attempts to fundamentally break the diffraction limit with deconvolution are intrinsically futile, at least for signal-to-noise ratios achievable in practice.

Another common misconception arises when resolution is confused with magnification. Magnification is simply "blowing up" without revealing more detail. Similarly, the statement that single molecules can not be imaged with a standard microscope because they are too small ("smaller than the resolution limit") is wrong. They can in fact be seen as soon they emit enough light — just like stars can be seen on the night sky although they too are beyond the resolution of the eye. In fact, super-resolution methods based on single-emitter switching with subsequent localization (PALM, STORM, GSDIM, etc.) clearly prove that small things can indeed be seen, as long as they are sufficiently bright. The truth of the matter is simply that an ensemble of fluorophores can not be simultaneously *resolved* when the intermolecular distance is smaller than the effective point spread function.

In any case, obtaining a physical resolution beyond the diffraction limit is tantamount to reducing the size of the effective PSF h_{eff}.

2.2 High-resolution fluorescence microscopy

As has been outlined in the last chapter, resolution is for the most part a matter oft he expanse oft he effective point spread function[1]. This chapter will introduce *Stimulated Emission Depletion (STED)* microscopy, an instance oft he *RESOLFT (Reversible Optical Fluorescence Transitions)*-principle and the first practically applied method to truly break the diffraction barrier. STED is an example of a technique that enhances resolution by reducing the dimensions of the effective point spread function, specifically by modifying the distribution of excited molecules. Moreover, other super-resolution methods will be covered briefly for the sake of completeness and comparison.

2.2.1 Stimulated Emission Depletion (STED) microscopy

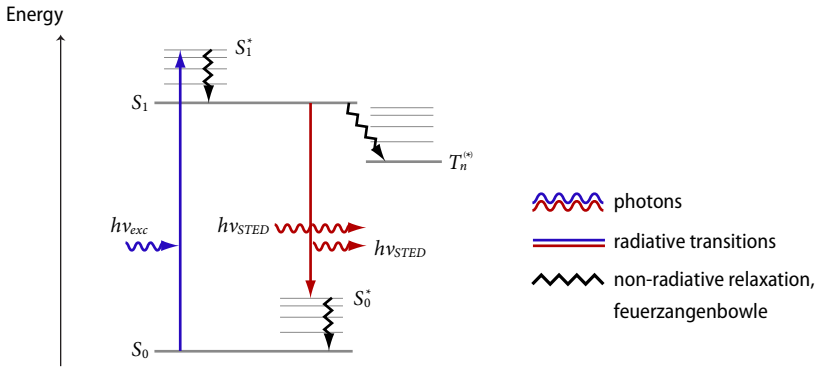

Figure 2.4.: Jablonski-diagram of a fluorophore and the molecular principle of stimulated emission. Stimulated emission can be used to keep fluorophores in a dark state.

It was illustrated in Figure 2.3 how several fluorophores are usually excited synchronously by the diffraction-limited excitation beam in a laser scanning microscope. As they all emit more or less in concert with a diffraction-limited detection PSF, there is no way to separate one from the other. But, resolution means separation, precisely.

The route for high resolution taken by STED consists ofk eeping the molecules at the periphery oft he excitation spot in a dark, i.e. in a non-fluorescent state. This is done by superimposing the excitation beam with a suppression beam, also called inhibition beam, that features a hollow

1 From now on, speaking of "PSF" implies that the effective PSF is meant, unless otherwise noted.

dark spot surrounded by light. A Gauss-Laguerre or *donut*-beam is a popular choice (more in section 3.2, page 27). The mechanism through which suppression is achieved is a process called *stimulated emission* and it works in the following way: any fluorophore that has been excited to a higher vibrational state S_1^* of S_1 and relaxed to the ground state of S_1 is instantaneously funneled back to the ground state via stimulated emission induced by a beam with wavelength λ_{STED} — before it has a chance to fluoresce. Although it bathes in excitation light, the probability for a molecule subject to an ample STED beam to be found in the ground state is close to 100% and therefore fluorescence is switched off, essentially. And since the suppression beam is hollow, it leaves only the molecules in the very center to fluoresce — h_{exc} (in Equation 2.1) has become much narrower than before. The emitted fluorescence is again collected; a process that can once more be described by the action of h_{det}. Although the detection PSF persists unchanged, $h_{eff} = h_{exc} \cdot h_{det}$ is obviously much smaller now, because h_{exc} is zero except for the small region in the center where the intensity of the inhibition beam was close to zero. Or, from another point of view: after applying the suppression beam the position of the undepleted fluorophores can not any better be localized using the detection PSF alone (which still has the same dimensions). However, the location of these molecules is known more precisely because they are necessarily in the small dark center of the inhibition beam, whose position is in turn established by the position of the scanner[1]. The pinhole can now be safely omitted, because h_{det} hardly plays a role for h_{eff} — it can be assumed constant over the small region where h_{exc} is non-zero, anyway. The effective PSF is solely governed by h_{exc}, modified by the inhibition beam.

The objection is often raised that the dark spot in the STED beam itself is diffraction-limited, and therefore the resolution enhancement can be at best a factor of around two. The former is true, whereas the latter is not. Consider a molecule subject to stimulated emission by a beam with intensity I_{STED}. The probability that it stays in the excited state is proportional to $\exp(-I_{STED}/I_{sat})$. (I_{sat} is a characteristic property of the type of molecule and is usually on the order of $10^6\ \mathrm{W/cm^2}$.) The consequence is that for STED intensities $I_{STED} \gg I_{sat}$, essentially all molecules will stay dark because they are efficiently confined to the ground state. In order to cast this into numbers, it is convenient to define a threshold intensity of, say $3I_{sat}$. The probability for a molecule subject to a STED intensity $> 3I_{sat}$ to fluoresce is smaller than $\exp(-3) = 0.05$ and therefore this molecule is dark for all practical purposes, in essence because the fluorescent state is disallowed by the presence of the STED beam. Since I_{STED} increases from

1 Either the beams are moved across the sample of the sample itself is being scanned through the beams. In any case, the position of the suppression PSF is known with nanometer accuracy.

the center of hollow supression beam on outwards to the rim, the probability for a molecule to be off is highest at the rim and (ideally) zero in the very center. More precisely, a certain distance from the center exists for which $I_{STED} = 3I_{sat}$, meaning further outwards practically all molecules will be off. But since the threshold $3I_{sat}$ can always be moved towards the center by increasing the overall intensity of the STED-beam, the region in which the fluorophores are still capable of signaling can be made far smaller than the physical width of the donut minimum. In other words: far smaller than the diffraction barrier.

For a wavelength λ and a numerical aperture NA of the objective lens, the spot in which the fluorophores are signaling will have a diameter [35]

$$d \approx \frac{\lambda}{2NA\sqrt{1 + \max(I_{STED})/I_{sat}}}. \tag{2.8}$$

$\max(I_{STED})$ is the maximum intensity of the STED-beam. This holds for $\max(I_{STED})/I_{sat} \gg 1$. Scanning the overlapping STED- and excitation beams accross the sample reveals structures at a resolution d because the signal of fluorophores that are further apart than d can be recorded *sequentially* — they have been separated in time. With many current dyes, d can be shrunken to 20 nm; for a certain class of inorganic fluorophores (crystal color centers) even 5.8 nm have been reported [87].

An interesting point to mention is that an additional photon is given off during the stimulated emission process with exactly the same wavelength as the stimulation beam λ_{STED}. This is in fact important, because these particular photons are emitted at the periphery of the excitation PSF; they are the photons that should *not* contribute to detection. When saying the periphery of the excitation spot is kept dark this really means the molecules are forced to emit with a very specific wavelength λ_{STED} instead of giving off light according to their normal emission spectrum. In contrast, the fluorophores left to spontaneously emit are those in the center of the donut, exactly the fluorophores that have not interacted with the inhibition beam. These fluorophores contribute to the wanted signal and can easily be separated from the stimulated fluorophores because the probability is very high that they give off light with another wavelength than λ_{STED}.

A key ingredient for STED is that the STED rate-contant $k_S \gg k_F$, the fluorescene rate-constant. Otherwise too many molecules too far away from the center of the suppression beam will be allowed to fluoresce spontaneously. (This problem will be further addressed later). Another requirement is that the relaxation from the vibrational to the ground state after STED $k_{S_0^* \to S_0} \gg k_S$; this forbids re-excitation of a fluorophore from the S_0^*- to the S_1-state by the very same STED beam that has just brought it

down to S_0^*. Fortunately, the lifetime of a vibrational level S_0^* is usually several orders of magnitude lower than the emission/suppression process and once down in the ground state, suppression is permanent — until the next excitation photon arrives — because the energy of the STED-beam is not sufficient to re-excite the molecule from S_0 to S_1.

2.2.2 Reversible Saturated Optical Fluorescence Transitions (RESOLFT) and targeted switching

It should not go unmentioned that quenching molecules at the rim of the excitation spot by stimulated emission is not the only means of achieving high resolution. The key essence is their separation in time, so that the fluorophores can be recorded sequentially and therefore be separated in space. Probably the most powerful method for separating otherwise indistinguishable fluorophores is *switching* them by driving a reversible optical transition essentially between two discernible states A and B, where intermediate states can be involved. A and B have to be different in a sense that one can identify which state a molecule is in; it can be a bright and a dark state, or a blue and a red state, for example. For simplicity, the two states A and B will in the following be referred to as "bright" and "dark" states, respectively.

Regardless of the involved states, one modality (that includes STED) is *targeted switching*. Targeted switching simply means that the transition $A \rightarrow B$ is driven in a controlled manner, namely by using an intensity distribution featuring one or more intensity "zeros" or "nulls". Where the intensity is non-zero, markers are transferred from bright to dark. Whenever the transition is saturable, a saturation intensity exists in any case above which basically all molecules are switched dark. In contrast, molecules stay bright essentially only at and in a small region with diameter d around the nulls, where d is given by Equation 2.8.

By moving the intensity distribution, including the nulls, across the sample, a sub-diffraction image can be recorded. At each position, the signal originating from the molecules in the bright state, i.e. the molecules in a region with diameter d around the null points, is collected. Because the position of the nulls is known at a certain time, the position of the markers signaling at this particular time point is known. In contrast, markers further apart than d are in different states at different times and are therefore recorded sequentially. Thus, they can be discerned even when the distance between them in much smaller than the diffraction limit.

When using multiple nulls, they have to be further apart than the diffraction limit in order to separate bright molecules located in different zeros. This is intrinsically fulfilled, since an intensity distribution with nulls closer together is forbidden by the very same diffraction limit in the first

place. Thus, by imaging the signal onto a camera, spatial multiplexing of readout is possible.

In practice, state *A* and *B* can be for example

- an excited state from which spontaneous fluorescence is allowed and the ground state where molecules are confined to by stimulated emission in order to keep them dark, as in STED-microscopy;
- an excited state from which spontaneous fluorescence is allowed and a dark state reached by shelving molecules into a meta-stable, long-lived triplet state as in **G**round **S**tate **D**epletion (GSD) microscopy [47];
- a non-emittent state and a excited fluorescent state with the excitation being saturated as in **S**aturated **P**attern **E**xcitation **M**icroscopy (SPEM, [40]) or **S**aturated **S**tructured **I**llumination **M**icroscopy (SSIM, [33]);
- two different isomeric conformations of a molecule having discernible emission spectra or where only one conformation is fluorescent [44];
- a dark state that can be switched over to bright by binding of another molecule or vice-versa [44],

and many more. The common denominator is that a **r**eversible **s**aturable **o**ptical **f**luorescent **t**ransition is driven deliberately in order to perform sequential read-out. Hence, all these methods have been generalized under the acronym *RESOLFT* [44]. Throughout this work, when STED is performed or being referred to, it is implied that the techniques and methods presented and explained with the help of STED can also be realized with any set of two states that have been presented above.

2.2.3 Stochastic switching and read-out

In targeted switching mode, the separation in space for sequential read-out and sub-diffraction imaging is realized by separation in time in a deliberate manner. Other methods rely on stochastic separation in time combined with *non-targeted* read-out. To this end, molecules at random locations are transferred from the dark to a bright state by chance. One prerequisite is that at any time the sub-distribution of markers being in the bright state is sufficiently sparse so that on average no two bright molecules are closer than the diffraction limit. When this is the case, they are again separated in space and therefore discernible. After read-out, the next step is

to switch the current sub-population dark in order to allow another subset of molecules to be registered.

In contrast to targeted read-out, the position of the markers is not established by the position of an intensity zero but must be found by localizing the center of their diffraction-limited image, for example by fitting a Gaussian. Again, this is possible only because they can be separated in space. Calculating the centroid can be done to a precision that is a factor $\sqrt{m}$ below the diffraction limit when m is the number of photons collected in a diffraction-limited blob. Thus, the transition $A \rightarrow B \rightarrow A$ has to be made only once as long as a sufficient number m of photons can be collected during a single pass[1]. This is in contrast to targeted-switching, where typically n switching cycles are required per scanning dimension for an n-fold resolution enhancement along this direction, because the scan raster is n-fold finer along each dimension. For example, two-dimensional resolution enhancement of a factor of ten requires 100 transitions $A \leftrightarrow B$.

On the one hand, it is clear that repeated switching in targeted mode inflicts heavy stress on the molecules, which is not the case for stochastic methods. On the other hand, fluorophores only have a limited photon budget, meaning that they can only emit a certain number of photons. This number follows a statistical distribution — some fluorophores can emit more than others before they bleach. Therefore, the higher the resolution aimed for in stochastic mode is, the more molecules run out of their photon budget before the required m photons can be collected. Consequently, more detection events have to be discarded. This is in consequence similar to targeted read-out, where more molecules bleach due to switching fatigue when the desired resolution is set higher.

Stochastic switching has been implemented as **p**hoto **a**ctivatable **l**ocalization **m**icroscopy (PALM, [7]), which initially employed a beam of light as an on-switch where the intensity was adjusted to activate only a sufficiently sparse population of markers. Subsequently, after read-out, the fluorophores were bleached. Another method known as **st**ochastic **o**ptical **r**econstruction **m**icroscopy (STORM, [88]) uses a Cy3-Cy5 dye pair as the switch. Cy5 can be switched to a metastable dark state by a red beam of light; it can be turned fluorescent again by exposure to green light when a Cy3 molecule is in close proximity. Yet another method is ground state depletion and single molecule return (GSDIM) [26], where transition to the on-state takes place by awaiting the spontaneous return of molecules that have been pushed to a long-lived dark state; in this case a triplet state.

1 Here, a bright state must not be confused with the excited state from which fluorescence originates. Instead, a bright state in this case is a state in which the molecule is allowed to fluoresce in principle. While being in the bright state it can undergo excitation and emission many times, before it is switched to a dark state. There, at no time can it be excited.

Many more methods than can be mentioned have followed and will follow, however, the rest of this work will solely focus on RESOLFT and on STED in particular. Again, it is understood thereby that, from now on, when STED is said in the text, it is meant that any RESOLFT method can be applied similarly.

Since stochastic methods are usually designed to switch between relatively long-lived conformational states they can utilize low levels of light; in contrast to STED, which requires fairly high intensities. This is however not the case for all RESOLFT methods. Beyond that, RESOLFT has the advantage that resolution is not given by the number of photons a molecule happens to emit; resolution is an experimental parameter and can be deliberately set by the user. In particular, it is independent of the number of photons collected — as soon as an event sets itself apart above noise level it can be considered resolved, even when the actual photon count is comparatively low [105]. This, together with the ability to adapt the scanning scheme to the sample [57], give RESOLFT methods the edge when it comes to speed. On the downside, because an intensity distribution with null(s) has to be generated, RESOLFT is often perceived as being technologically more challenging than stochastic read-out where switching happens simply by chance. This argument, however, becomes invalid with this thesis.

3

Rugged simplifications for RESOLFT-type nanoscopes

Although RESOLFT- and especially STED-microscopy has proven to be a reliable and valuable technique not only for biological imaging, today's setups still have room for improvements. This section will highlight the potential of polarization manipulation for STED-nanoscopy; in particular, it will be shown that incorporating polarization changes in the STED beam ultimately leads to simpler, yet equally competitive setups. In order to be able to analyze and compare the STED-PSFs generated using polarization engineering, an in-depth analysis of image formation is given first. Then, a typical conventional setup is briefly described, together with its drawbacks. Finally, devices for polarization manipulation of the STED-beam are introduced and described, their performance is evaluated both theoretically and experimentally, and it is demonstrated how these devices make setting up and operating a STED nanoscope a much simpler task than before.

3.1 The mathematical theory of focusing with high numerical apertures

In the last chapter, it has been described how quenching molecules at the rim of the excitation spot leads to optical resolution beyond the diffraction barrier. The question remains of how to generate a suitable intensity distribution for the inhibition beam. Needless to say, it is also desirable to gain insight into how efficient a certain STED-PSF acts on a given distribution of excited fluorophores. In order to fully investigate these effects, it is now described how to compute the electric field in the vicinity of the focus for arbitrary beams incident in the pupil of the objective lens.

Figure 3.1 depicts a general situation where a wavefront W is passing an aperture S. In general, the field generated by an arbitrary aperture can

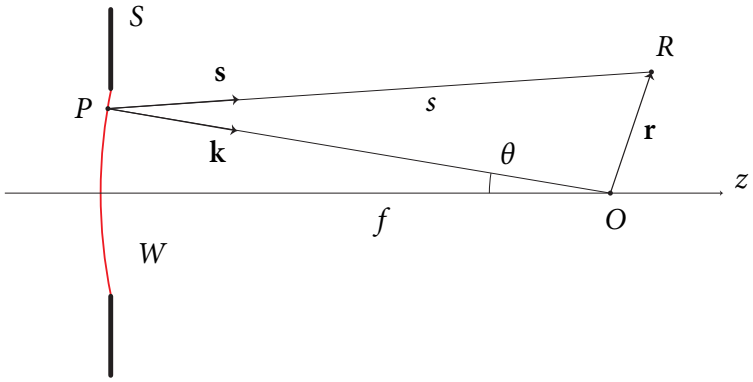

Figure 3.1.: A curved wavefront W passing an aperture S. Note that for S being a lens, it follows that $\mathbf{k}$ points towards the origin O.

be described as follows. Each point P in the aperture is, according to the *Huygens-Fresnel principle*, considered to be the origin of spherical wavefronts $\exp(\imath ks)/s$ with $s = |\mathbf{s}| \cdot s = |\mathbf{s}(P, \mathbf{r})|$ being the distance from the huygenian source to the observation point R, and k being the wave number[1]. ($\mathbf{s}$ is a vector of unit length.) By integrating over the whole aperture, the field at point $\mathbf{r}$ can be calculated [11]:

$$E(\mathbf{r}) = \frac{\imath k}{2\pi} \int_{\text{aperture}} W(P) \frac{\exp(-\imath ks)}{s} \, \mathrm{d}S, \tag{3.1}$$

In the special case when the aperture is a lens, we have spherical wavefronts converging towards O, i.e. $W(P) = A(P) \exp(\imath kf)/f$ with the focal length f. For the sake of generality, W can have a different phase and amplitude $A \in \mathbb{C}$ for each P. Together with Equation 3.1 we get [30]

$$E(\mathbf{r}) = \frac{\imath k}{2\pi} \int_{\text{aperture}} A(P) \frac{\exp\left(-\imath k(s-f)\right)}{fs} \, \mathrm{d}S.$$

Now, according to Debye, for $|\mathbf{r}| \ll f$ several approximations can be applied [17, 63]:

- first, $s - f \approx \frac{\mathbf{k}}{k} \cdot \mathbf{r}$;
- second, since we are working on a sphere, the area element can be expressed using a solid angle element: $\mathrm{d}S = f^2 \, \mathrm{d}\Omega$;
- and third, $s \approx f$ in the denominator.

1 As usual, the time periodic factor $\exp(-\imath\omega t)$ is being suppressed.

The first approximation can be verified by computing the square of

$$\mathbf{s}\cdot s=\frac{\mathbf{k}}{k}f+\mathbf{r}\Leftrightarrow s=f\sqrt{1+2\frac{\mathbf{k}}{k}\frac{\mathbf{r}}{f}+\frac{r}{f}}\approx f\left(1+\frac{\mathbf{k}}{k}\frac{\mathbf{r}}{f}\right).$$

Altogether, the focal field is then [30]:

$$E(\mathbf{r})=\frac{\imath k}{2\pi}\int_{\Omega}A(P)\exp(-\imath\mathbf{k}\cdot\mathbf{r})\ \mathrm{d}\Omega. \tag{3.2}$$

This *Debye integral* gives a good approximation of the field in the vicinity of the focus [11]. It is worthwhile to note that in Equation 3.2, the resulting light distribution is not expressed as a sum of spherical wavelets, according to Huygens-Fresnel, but rather as a spectrum of plane waves. The assumption is reasonable, however, that for sufficiently small $|\mathbf{r}|$ (i.e., when $|\mathbf{r}|\ll f$) the resulting phase error is minimal.

Equation 3.2 closely resembles the Fourier transform $\mathcal{F}$ of $A(P)$

$$\mathcal{F}[A]=(2\pi)^{-3/2}\int A(P)\exp(-\imath\mathbf{k}\cdot\mathbf{r})\ \mathrm{d}^3k. \tag{3.3}$$

In order to carry out this fourier transform, $A(P)$ will have to be written in terms of the wavevector k. It is clear, that for an monochromatic wave $\mathbf{k}$ incident on the lens, the refracted wavevectors have to obey $A(\mathbf{k})=A(\mathbf{k})\delta(|\mathbf{k}|-k)$. In other words, in $k-$space, the refracted waves all lie on the cap of a spherical shell. The radius of the sphere is defined by $|\mathbf{k}|^2=k^2=k_x^2+k_y^2+k_z^2=(2\pi n/\lambda)^2$ and the extension of the cap (how much the cap occupies of the full sphere) is given by the maximum semi-opening angle $\alpha=\sin^{-1}(NA/n)$ of the lens. λ denotes the vacuum wavelength and n the refractive index of the medium behind the lens. The spherical cap defines the spectrum of all plane waves as they are transmitted by the lens[1]. Using this allows us to rewrite Equation 3.3 in terms of $A(\mathbf{k})$:

$$A(\mathbf{r})=\mathcal{F}[A(\mathbf{k})]=(2\pi)^{-3/2}\int_{\Omega}A(\mathbf{k})\exp(-\imath\mathbf{k}\cdot\mathbf{r})\ \mathrm{d}^3k.$$

Now, defining $k_\perp=\sqrt{(k_x^2+k_y^2)}=k\sin\theta$, the projection of $\mathbf{k}$ onto the plane $z=-f$, and $k_z=\sqrt{(k^2-k_\perp^2)}=k\cos\theta$, and further taking a closer look at $A(\mathbf{k})$ reveals that all $\mathbf{k}$ lying on a spherical cap means that k_z is a function of k_x and k_y. Consequently, we can write

$$\mathcal{F}[A(\mathbf{k})]=(2\pi)^{-3/2}\iint A(k_x,k_y)e^{-\imath(k_xx+k_yy)}e^{-\imath k_z(k_x,k_y)z}\ \mathrm{d}k_x\mathrm{d}k_y.$$

1 In practice, dealing with non-monochromatic waves means that the cap will have a finite thickness.

and comparing with Equation 3.2 we finally find that

$$E(\mathbf{r}) = \sqrt{2\pi}\, \imath k \;\cdot\; \mathcal{F}\left[A(\mathbf{k})e^{-\imath k_z z}\right].$$

When $[\rho, \phi]$ denote polar coordinates in the back focal plane, $[\rho, \phi] \rightarrow [k_x, k_y]$ is an injection for an ideal lens. This means that, depending on where a ray arrives in the back focal plane, it will unambiguously be bent into a ray having certain horizontal and vertical components k_x and k_y (Figure 3.2). Thus, for practical purposes, one can use the coordinate-space $[\rho, \phi]$ for addressing points in the pupil plane:

$$\rho = f \cdot \sin(\theta)$$
$$\phi = \tan^{-1}(-k_x, -k_y),$$

where the $\tan^{-1}$-function observes the usual conventions for converting to polar coordinates.

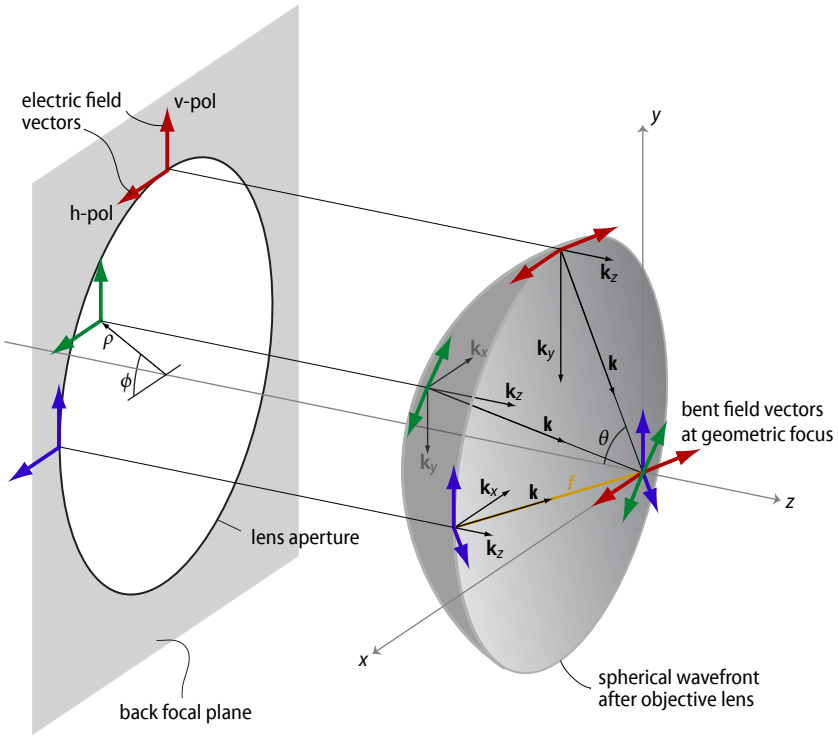

Figure 3.2.: Diffraction at a high-NA lens. The field vectors of incoming rays get deflected in the focus, depending on where they cross the entrance aperture.

In [114], the authors extensively discuss the validity of the Debye approximation and find that a high degree of accuracy can be expected when

$kf \gg \pi/\sin^2\alpha$, when furthermore the aperture size $a \gg \lambda$, and for semi-aperture angels $\alpha < 30°$. So far, the theory presented here is scalar and paraxial, yet, high-resolution microscopy asks for high-NA lenses with α up to 70°. For high-aperture focusing, additional effects have to be considered, apodization and most importantly depolarization. The latter arises when an incident linearly polarized field is diffracted by a high-NA lens. While the plane wave is converted to a spherical wave, the electric field vector is deflected as shown in Figure 3.2. For example, even when the incident field is purely horizontally polarized, the focal field will have not only x-, but also y- and especially z-components [30].

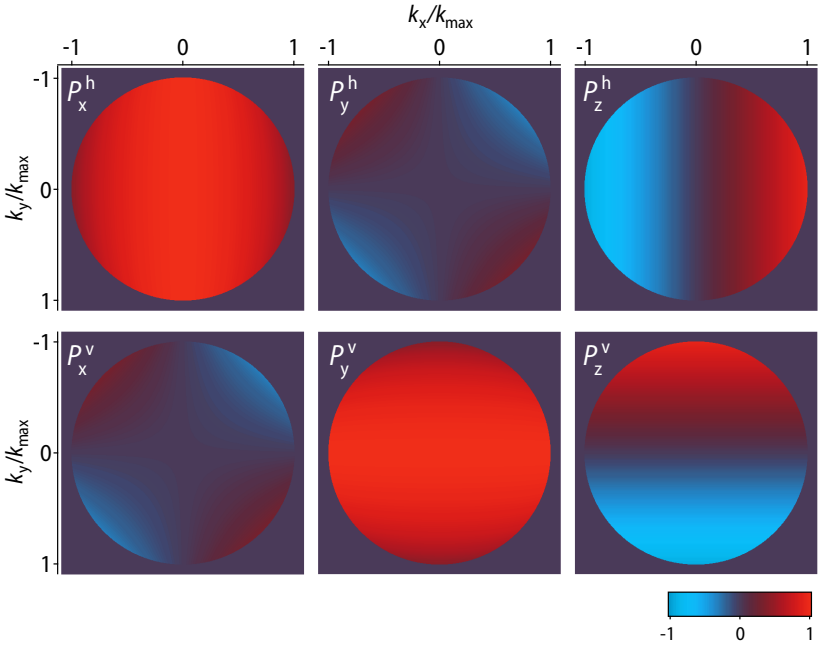

Figure 3.3.: Depolarization under high-NA focusing. The six graphs, plotted after Equation 3.4 – Equation 3.6 on page 26 illustrate how horizontal (top row) and vertical (bottom row) polarization in the pupil turns into electric fields along all three directions in the focus. It can be seen that, besides the original field direction (top row left, resp. bottom center), additional components especially along the z-axis arise (far right column).

The deflection of the field vector for incident horizontally polarized light can be accounted for by weighting the scalar field amplitude with

the function $\mathbf{P}^h(\mathbf{k})$ having components [69]

$$P_x^h = \frac{1}{k_\perp^2}\left(\frac{k_z}{k_0}k_x^2 + k_y^2\right), \tag{3.4}$$

$$P_y^h = \frac{-k_x k_y}{k_\perp^2}\left(1 - \frac{k_z}{k_0}\right), \tag{3.5}$$

$$P_z^h = \frac{k_x}{k_0}. \tag{3.6}$$

These equations map the incoming field vectorially onto the three field directions in the focus. P_i^h $(i \in \{x, y, z\})$ denotes the strength of the field in the focus along direction i generated by a horizontally polarized unit field in the back focal plane. The corresponding equations for incident vertical polarization can be found using

$$\mathbf{P}^v(\mathbf{k}) = \mathbf{R}\cdot\mathbf{P}^h(\mathbf{R}\cdot\mathbf{k}), \quad \mathbf{R} = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

$\mathbf{R}$ is a rotation around the optical axis by $-90°$. Figure 3.3 shows $\mathbf{P}^h$ and $\mathbf{P}^v$ pictorially for a $NA = 1.4$ lens.

With this, the general deflection vector is

$$\mathbf{P}(\mathbf{k}) = \mathbf{P}^h(\mathbf{k})e^{\imath\varphi_h} + \mathbf{P}^v(\mathbf{k})e^{\imath\varphi_v}, \tag{3.7}$$

$e^{\imath\varphi_{v,h}}$ being the global phases for the vertically respectively horizontally polarized component of the incident beam. Equation 3.7 now describes fields of arbitrary polarization incident on the lens.

Putting it all together, the vectorial field near the focus is

$$\mathbf{E}(\mathbf{r}) = \sqrt{2\pi}\,\imath k\ \mathcal{F}\left[\sqrt{\frac{k_0}{k_z}}\,\mathbf{P}(\mathbf{k})A(\mathbf{k})e^{-\imath k_z z}\right]. \tag{3.8}$$

Again, $\mathbf{k}\cdot\mathbf{r} = k_x x + k_y y + (k^2 - k_x^2 - k_y^2)^{1/2}\ z$. The apodization factor $\sqrt{k/k_z} = 1/\sqrt{\cos\theta}$ has to be introduced for reasons of energy conservation for a lens obeying the sine condition [100]. Altogether, the field near the focus can be expressed by the fourier transform of a spherical cap in frequency space that defines a spectrum of plane waves.

Furthermore, for precisely assessing the inhibition effectiveness of the STED-PSF $\mathbf{h}_{STED} = |\mathbf{E}_{STED}|^2$ on a given fluorescent molecule, it is essential to bear in mind that it depends on the orientation $\mathbf{n}$ of the molecular transition dipole [10, 21, 97, 103, 104]. Thus, the true local STED-field, as it is seen by a fluorophore, is modified by the interaction with the fluorophore itself. For most molecules, the transition dipoles for excitation

and stimulated emission are parallel[1] [62] and the STED-efficiency is proportional to $|\mathbf{E}_{STED} \cdot \mathbf{n}|^2$ [21]. This means for example, that STED-PSFs with purely linear local fields will not be able to interact efficiently with arbitrary dipoles. Later in this work, when dealing with new STED-PSFs, this behavior will have to be taken into account; it will in fact also lead to a new type of microscopical imaging (see subsection 3.3.3).

The results of this section permit a comprehensive assessment and comparison of STED-PSFs. To this end, the equations have been implemented in Matlab (MathWorks, USA). With these foundations, the PSFs of beam shaping devices can be evaluated, and during the search for simplified STED arrangements it can be checked how well a newly found PSF will perform. The results will be presented in the following sections. Before that however, it is necessary to look at the layout of a standard STED-setup in more detail, in order to understand the weak points that will have to be attacked in the course of developing simplifications.

3.2 A typical conventional STED setup

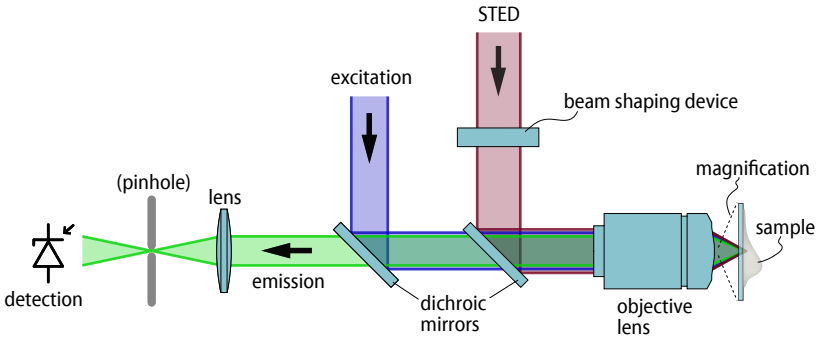

Figure 3.4.: A typical STED setup is a laser scanning system with an additional STED-laser and an additional beam shaping device.

A typical contemporary STED-setup consists of a high-power STED-laser and an excitation laser that are both coupled into a commercial microscope stand carrying the objective lens and the specimen. The STED-beam passes through a beam shaping device (BSD) located in or close to a conjugated back focal plane (BFP) that is responsible for modifying the STED-beam so that it focuses to a suitable, i.e. hollow PSF that suppresses excited molecules in the outer parts of the excitation PSF.

Other typical elements are means to scan both beams across the sample (not shown), either by rotating the illumination about a point in a con-

1 Although the orientation of the dipole can change because of rotational diffusion.

jugated BFP or — simpler but slower — by moving the sample through a static beam. Fluorescence is collected by the objective lens, de-scanned if necessary, and fed back to the detector through a pinhole[1] and through an emission filter (not shown) that removes any residual excitation- and STED-light.

A closer look reveals that a STED-setup as described here is nothing more than a (confocal) laser scanning microscope (LSM) as shown in Figure 2.2, page 9, with an additional STED-laser coupled into the excitation path and a BSD in a back focal plane of the STED-beam. Yet, there are two additional necessities: the generation of a deep null[2] in the STED-PSF, and their alignment with excitation. All these requirements are often deterring for scientists trying to build and operate a STED-nanoscope from scratch; upgrading a commercial LSM is even less of an option, since a back focal plane is rarely accessible.

The beam shaping device demands a closer look. In most practical cases known so far, the point spread function for STED is formed by shaping the phase of the wave front in a conjugated back focal plane of the objective lens. In other words, a suitable spatial phase distribution $A(\mathbf{k} \rightarrow [\rho, \phi])$ is generated in the back focal plane that, when focused according to Equation 3.8, produces a hollow dark spot in the focus surrounded by high intensities either in one, two or three dimensions. Generating a suitable distribution of the phase is usually achieved by placing a *phase plate* in the beam path; this phase plate consists of for example MgF_2 deposited on a glass substrate. The coating has a locally varying thickness and thus imprints a varying phase onto the beam, while leaving the amplitude untouched[3].

High resolution along a line

For a 1D setup, the complex phase element can be as simple as

$$A(\rho, \phi) = \begin{cases} e^{\imath\pi}, & \text{if } 0 \leq \phi < \pi, \\ 1, & \text{if } \pi \leq \phi < 0 \end{cases} \qquad \phi = [0..2\pi] \tag{3.9}$$

meaning that one half of the beam acquires a phase of π and during focusing interferes destructively with the other half of the beam for any point along the optical axis. This leads to a line-shaped zero in the focal plane and resolution enhancement along one axis. The desired off-axis maxi-

1 The pinhole is not mandatory for STED, but it is often used to remove unwanted stray-light.

2 A typical specification is that the null should feature an intensity of less than 1% of the maximal intensity present in the STED-PSF.

3 Up to a constant factor that stems from refection of absorption, obviously.

mum emerges because the path lengths for rays leading to a certain off-axis point are such that the π-phase difference is exactly counter-balanced. Vectorial treatment reveals that the incoming polarization has to be parallel or anti-parallel to $\phi = 0$, otherwise a significant field component along z will develop on the optical axis.

It can be shown that this phase plate is superior in terms of steepness of the null (and thus resolution) when dipole orientation can be neglected. Yet, while high resolution along a line is useful for demonstrational purposes or specialty applications, the typical assay requires enhancing the resolution along at least two dimensions.

Two dimensional resolution enhancement

…can be achieved by either crossing two phase plates as described in Equation 3.9 or by means of a vortex phase plate, also called helical phase ramp. Because of the polarization constrains mentioned above, the former requires that the beam is separated into two orthogonally polarized beams that subsequently have to be recombined, the latter is a one-beam element that works as follows: the phase of the beam is changed according to the vortex function

$$A(\rho, \phi) = e^{\imath\phi}, \qquad \phi = [0 \ldots 2\pi]\,. \tag{3.10}$$

Its action is simply that any two opposite rays, i.e. rays that come in at ϕ_1 and at $\phi_2 = \phi_1 + \pi$, can interfere destructively in the focal plane. Hence, every ray finds a π-shifted counterpart along the optical axis with which it can cancel out during focusing. In a sense, the vortex phase element is a phase element similar to that in Equation 3.9 but with rotational symmetry. This is true for paraxial conditions, but for high numerical apertures things get more complicated. Because any radial component of the incoming beam gives rise to an axial field in the focus[1], the null gets filled up for high numerical apertures. The solution is to balance the z-field on the optical axis with a field of equal magnitude but opposite sign by using a circularly polarized STED-beam as shown in Figure 3.5. Then, azimuthal components cancel as before, due to a phase shift of π between opposite rays; however, the radial components each give rise to an axial field. The rays of the incoming vertically polarized beam (red) that have a radial component have crossed the phase plate at 12 o'clock and 6 o'clock, thus the latter ray gets a phase shift of π from the phase plate and together they generate a positive axial field when deflected by a high-NA lens (lower right drawing). In contrast, the rays of the horizontally polarized input

1 This can be deduced for example with the help of Figure 3.2 on page 24 and Figure 3.3, page 25.

beam (blue) that have a radial component have passed the phase plate at 3 o'clock and 9 o'clock and consequently have acquired phases of $\pi/2$ and $3\pi/2$, respectively, and have an additional global phase of $\pi/2$ because they belong to the retarded linear component of the circular input light (top drawing). Thus, they have overall phases of π and $2\pi = 0$, and during high-NA focusing generate a negative axial field on the optical axis. This field can interfere destructively with the positive z-field generated by the incoming vertical beam and an on-axis null is the result.

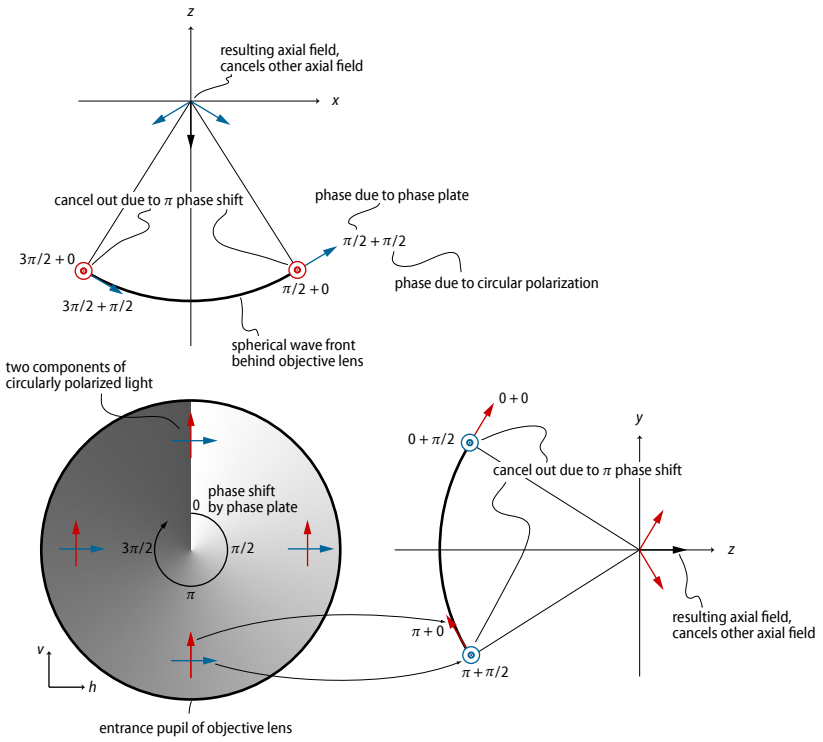

Figure 3.5.: High-NA focusing of a donut-beam. The occurrence of a central zero depends on the cancellation of axial fields and therefore on carefully balanced input polarization.

Obviously, the need for two fields that cancel each other requires them to be exactly balanced, in other words, the incoming light has to be exactly circularly polarized and of the right handedness. Any elliptical component will leave a z-field in the null, because when the incoming linearly polarized components have different intensities or a phase shift departing from $\pi/2$, they can not perfectly cancel each other. In practice, this is not always easy to achieve, because surfaces (mirrors, etc.) struck at an angle will usually generate phase shifts between s- and p-polarized light.

Nevertheless, the vortex phase plate is probably the most widely used phase plate for STED. Although resolution-enhancement in the focal plane already allows one to observe considerably finer detail in the specimen, it is clear that this works best when the sample itself is two-dimensional, for example a lipid bilayer or embedded in a lipid bilayer. For structures that extend along the optical axis, contributions from above and below the focal plane will add up in the image because the effective point spread function will image everything present along its z-extension and the z-extension is still ≈ 500 nm. Obviously, for such specimen, it is beneficial to shrink the effective PSF along the optical axis too, in that this will considerably improve contrast.

Three-dimensional resolution enhancement

…is commonly achieved with an annular phase plate with

$$A(\rho,\phi) = \begin{cases} e^{\imath\pi} & \text{if } \rho > \Theta, \\ 1 & \text{if } \rho \leq \Theta; \end{cases} \qquad \phi = [0 \ldots 2\pi]\,. \tag{3.11}$$

The two rings with phases 0 and $\exp(\imath\pi)$ can be pictured to generate to independent (but coherent) point spread functions, where the point spread function from the inner ring is larger because of the lower effective numerical aperture. The outer ring focuses to a smaller PSF that is coherently added to the bigger point spread function and cuts a hole in it because of the π phase shift. This way, a dark central spot surrounded by light is generated that can be used for shrinking the effective point spread function along x, y and z. Θ has to be adjusted so that the two PSFs cancel out to exactly zero in the geometric focus.

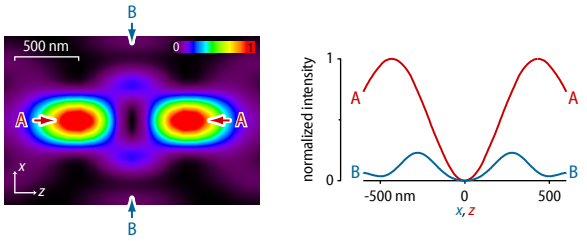

Figure 3.6.: xz-PSF of an annular phase plate (left) and line profiles (right) along x (blue) and z (red). The intensity is about five times lower laterally than axially and the curvature of the null is much flatter in the xy-plane.

Unfortunately, for smaller NA, a PSF grows faster along z than in the focal plane. Thus, when combining two beams with different NAs, as in the annular phase plate, the light wall in the focal plane is rather low, since the

one PSF is only marginally bigger in *xy* than the other one. Additionally, the curvature in the focal plane at the center follows a r^4-law; this makes the intensity grow only slowly with r close to the center (Figure 3.6). Altogether, the consequence is that with an annular phase plate resolution is significantly enhanced along the optical axis only and the high light intensities required to balance this invariably provoke photobleaching. Therefore, in practice a combination of an annular and a vortex phase plate is used for true, near-isotropic super-resolution[1].

While the three phase plates presented above are the ones most widely used in STED microscopy, other variations do exist [58]. Nevertheless, what all these phase plates have in common, is that they work by introducing a phase difference along the cross section of the STED beam. The following sections will introduce a method based on manipulation of the polarization and will show how this technique leads to considerably simpler STED-setups.

3.3 Beam shaping devices based on polarization engineering

3.3.1 Birefringent beam shaping devices — the motivation

An important point in setting up and operating a STED microscope is beam alignment. For maximum performance, the STED beam has to be centered on the excitation spot three-dimensionally with an error margin below ≈ 50 nm and beam alignment has to be stable over the course of a measurement and over an adequate field of view. Furthermore, the null of the STED-beam has to be sufficiently deep, i.e. lower than 5% of the rim-intensity; for best results lower than 1%, in practice. Inaccurate beam alignment, due to thermal drift for example, and improper nulls are frequent sources of performance drops and therefore the arrangement has to be checked periodically; typically every few hours, and before each measurement. Needless to say, the situation is even worse for 3D-STED setups with ring and vortex phase plates being applied simultaneously.

So far, most alignment problems arise because STED and excitation come from different sources. This could be solved by using a common light source for both the excitation and the suppression beam, for example two separate lasers coupled into a common optical fiber or, even more convenient, a super-continuum light source [109]. But then, having pre-aligned beams from one single source requires a wavelength-selective beam shap-

1 The other solution is iso-STED, a combination of 4Pi and STED [93]. It is very effective but even more complex than combining a vortex and an annular phase plate.

ing device that leaves the excitation wavelength unaffected while it treats the STED wavelength in such a way that it forms a PSF with a dark null.

Unfortunately, current beam shaping devices that rely on phase changes cannot sufficiently distinguish between wavelengths. When phase plates as described above, continuous or discrete, are used for altering the wavefront by a fixed amount it is unavoidable that the imposed phase change depends on the wavelength of the beam — but nowhere near enough. For example, a typical STED configuration uses a 647 nm laser passing a vortex wave plate ($0 \ldots 2\pi$) and an excitation beam of for instance 532 nm, of course depending on the specific dye being used. The 532 nm-beam must not be fed through the vortex plate, for it would pick up a phase change $\approx 0 \ldots 1.2 \cdot 2\pi$, which still very much leads to a donut shape in the focus, although without a true zero. It can be seen that traditional phase plates frustrate the application of a combined beam path. They are not suitable for the use with pre-combined beams because the excitation beam is also forged into something close to a donut.

Several solutions have been suggested; [101] relies on the annular separation of pre-aligned beams but blocks a considerable amount of STED light. The method proposed in [71] uses concentric circular zones to obtain wavelength selectivity but has, up our knowledge, not been realized in practice so far. More recently, Wildanger et al. [107] proposed a scheme that relies on different dispersion properties of different optical materials. By selecting two optical glasses whose refractive indices match at the excitation wavelength but differ for the STED wavelength, they were able to design a phase plate that can be shared by both beams. In their scheme, however, the detection beam path is coupled out between the objective lens and the phase plate using a dichroic mirror and furthermore, fabrication of two exactly opposite vortex phase plates made of different materials is not an easy job. All in all, most of these solutions more or less add about the same amount of complexity than they were designed to reduce.

In the following, beam shaping devices will be introduced that stand out by the fact that they use polarization engineering instead of phase engineering. It will be shown how, by simply rotating the beam polarization by different amounts for various parts of the beam, one-, two- and even three-dimensional suppression patterns can be generated. Furthermore, several excitation, detection and inhibition beams can pass these devices simultaneously and each of those beams are molded into the desired PSF. Therefore, these devices can be placed directly in a beam path shared by all three beams, and can consequently be used with light sources delivering both excitation and STED, intrinsically aligned. Moreover, the beam shaping device described here is highly tolerant against most circumstances that normally fill up the central zero. This way, the two major problems of set-

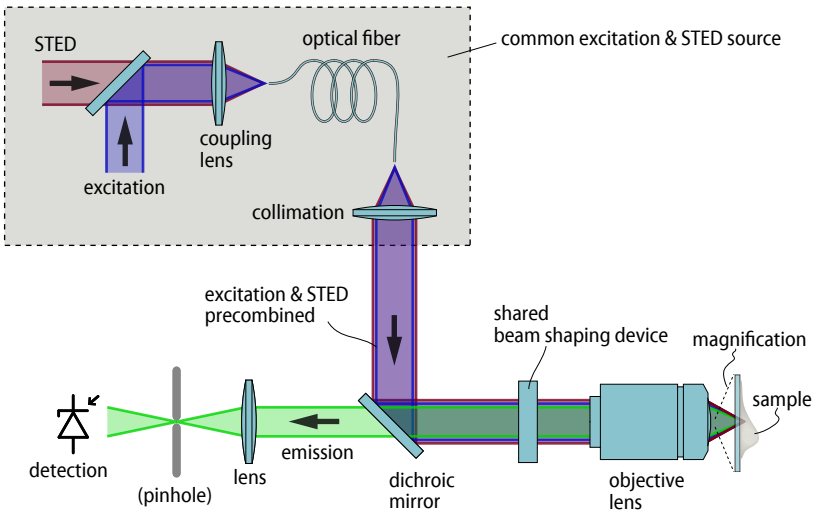

Figure 3.7.: A generic easySTED setup. A chromatic beam shaping device that treats STED, excitation and detection wavelengths appropriately can be placed close to an easily accessible back focal plane, for example upstream of the objective lens. The layout is basically that of a standard laser scanning microscope and together with a common source for excitation and STED, existing setups can be upgraded with neither much cost nor effort.

ting up and operating a STED-system (generation of a deep null and its alignment with excitation) are solved.

Working with polarization, one can advantageously use birefringent crystals which are commercially available in high quality as low order optical retarders, in contrast to the solution in [107]. The typical chromatic characteristics of a low order wave plate exactly match the spectral distance of the excitation and the STED-beam. Thus, simple, error-proof and easy-to-use beam-shaping devices can be built that, together with an appropriate laser source, can be used to economically retrofit almost any existing scanning (confocal) microscope. As shown in Figure 3.7, by replacing a excitation laser with a common source for excitation and STED and placing the birefringent beam shaping device behind the objective lens a commercial laser scanning microscope can be turned into a full-blown STED-device providing sub-diffraction resolution (easySTED).

3.3.2 Polarization engineering with birefringence

Because polarization and birefringence are of paramount importance for this work they will be treated in depth on the following pages.

Polarization: definition and characterization

Next to properties such as wavelength, propagation direction and many more, one important characteristic of (a ray of) light is its *polarization,* which denotes the direction of vibration of the electric field. When this direction is constant, light is said to be linearly polarized; otherwise, it is elliptical. In general, for any coherent fully polarized state the electric field vector

$$\mathbf{E} = \Re\mathrm{e}\left\{|E_h|e^{\imath\left(\mathbf{k}\cdot\mathbf{r}-\omega t+\varphi_h\right)} + |E_v|e^{\imath\left(\mathbf{k}\cdot\mathbf{r}-\omega t+\varphi_v\right)}\right\} \tag{3.12}$$

can be resolved into a horizontal and a vertically polarized field, each one described by a real amplitude $E_{h,v}$ and a phase $\varphi_{h,v}$. For problems concerning polarization, a common strategy is to omit periodic terms and combine the rest into a *Jones vector* $\mathbf{J}$ [28]:

$$\mathbf{J} = \begin{bmatrix} |E_h|e^{\imath\varphi_h} \\ |E_v|e^{\imath\varphi_v} \end{bmatrix}.$$

It is further useful to decompose the individual phases into a global phase $\Phi = \varphi_v$ and a phase difference $\varphi = \varphi_h - \varphi_v$:

$$\mathbf{J} = e^{\imath\Phi}\begin{bmatrix} |E_h|e^{\imath\varphi} \\ |E_v| \end{bmatrix}. \tag{3.13}$$

Equation 3.12 describes an ellipse, the polarization ellipse. Three cases can be discerned: first, for $\varphi = n\pi$, $\mathbf{J}$ describes the special situation of linear polarization, the plane of vibration is given by $\tan E_h/E_v$. Second, for $|\varphi| = (n+1/2)\pi$ and $E_h = E_v$, the electric field vector of the propagating wave runs along a circle when viewed along the propagation direction, hence the light is called circularly polarized. For $\varphi > 0$, the light is said to be right circulary polarized, for negative φ it is said to be left circularly polarized. Lastly, other cases for φ, or when $E_v \neq E_h$ while $\varphi \neq n\pi$, are simply called elliptical polarization.

Jones calculus is useful not only for describing polarization itself, but also for describing mathematically how a polarization state is changed. Specifically, one can relate polarization changes induced by a linear optical element to a 2×2 complex matrix $\mathbf{M}$ describing the optical element and a rotation matrix $\mathbf{R}$ describing its orientation $\tilde{\theta}$. Then, $\mathbf{M} = \mathbf{R}(\tilde{\theta})\mathbf{M}'\mathbf{R}(-\tilde{\theta})$ so that the initial polarization state $\mathbf{E}$ and the final state $\mathbf{E}'$ are related like $\mathbf{E}' = \mathbf{M}\mathbf{E}$ [53].

Here, mainly lossless materials, i.e. materials with unitary (= length-conserving) Jones matrices will be important. Polarization engineering (aka "changing the polarization state in a useful manner") involves giving

the entries of a Jones vector different phases $\varphi'_{h,v}$. The result is a different polarization state. A polarization element will thus be a matrix

$$\mathbf{M} = \begin{bmatrix} e^{\imath\varphi'_h} & 0 \\ 0 & e^{\imath\varphi'_v} \end{bmatrix}$$

or, incorporating arbitrary orientation:

$$\mathbf{M} = \begin{bmatrix} e^{\imath\varphi'_h}\cos^2\tilde{\theta} + e^{\imath\varphi'_v}\sin^2\tilde{\theta} & \left(e^{\imath\varphi'_h} - e^{\imath\varphi'_v}\right)\sin\tilde{\theta}\cos\tilde{\theta} \\ \left(e^{\imath\varphi'_h} - e^{\imath\varphi'_v}\right)\sin\tilde{\theta}\cos\tilde{\theta} & e^{\imath\varphi'_v}\cos^2\tilde{\theta} + e^{\imath\varphi'_h}\sin^2\tilde{\theta} \end{bmatrix}.$$

Jones calculus will be used shortly to describe practical elements for polarization engineering, first, however, the physical foundations of polarization engineering will be reviewed.

The physical origins of birefringence

When traveling in a medium, the speed of light is reduced by a factor n, the refractive index, which can depend on both the direction of the light beam and the polarization direction with respect to the crystal. In general, the refractive index along arbitrary directions can be accounted for by three different principal numbers n_1, n_2 and n_3, associated with three principal axes, as we will see. In cubic crystals, $n_1 = n_2 = n_3$ and such a crystal is isotropic — its optical properties are independent of propagation direction. In contrast, when the optical properties of the crystal lattice are anisotropic, i.e. when $n_1 \neq n_2 \neq n_3$, a phenomenon called birefringence arises.

Consider plane monochromatic waves $\mathbf{E} = \mathbf{E}_0 e^{\imath(\mathbf{k}\cdot\mathbf{r}-\omega t)}$ in a linear dielectric free of charges, currents and magnetic polarization. Maxwell's equations will come in handy both in plain and in phasor form:

$$\nabla\cdot\mathbf{B} = 0, \qquad \imath\mathbf{k}\cdot\mathbf{B} = 0, \tag{3.14}$$

$$\nabla\cdot\mathbf{D} = 0, \qquad \imath\mathbf{k}\cdot\mathbf{D} = 0, \tag{3.15}$$

$$\nabla\times\mathbf{H} = \dot{\mathbf{D}}, \qquad \imath\mathbf{k}\times\mathbf{H} = -\imath\omega\mathbf{D}, \tag{3.16}$$

$$\nabla\times\mathbf{E} = -\dot{\mathbf{B}}, \qquad \imath\mathbf{k}\times\mathbf{E} = \imath\omega\mathbf{B}. \tag{3.17}$$

As usual, $\mathbf{B}$ and $\mathbf{H}$ are the magnetic field and the magnetizing field, respectively; $\mathbf{E}$ is the electric field and $\mathbf{D}$ is the displacement field. The spatial and time derivatives of the plane wave equation lead to the operator substitutions $\partial/\partial t \to -\imath\omega$ and $\nabla \to \imath\mathbf{K}$, used for the phasor forms. Additionally, $\mathbf{D} = \varepsilon\varepsilon_0\mathbf{E}$ and $\mathbf{B} = \mu\mu_0\mathbf{H}$. ε_0 and μ_0 are the permittivity and the permeability of free space; the material constants ε and μ are the permittivity and the permeability. So far, this is the well-known description

of isotropic media. Now, in an anisotropic crystal, the interaction of light with the medium depends strongly on its polarization, because the dielectric displacement vector **D** is not necessarily parallel to the electric field **E** anymore: ε is a tensor; formally a 3×3 matrix. This has the following consequences.

For lossless materials the components of ε are real and symmetric [11], and so the principal axis theorem of linear algebra [12] implies that there is a coordinate set for which ε becomes a real diagonal matrix:

$$\varepsilon = \begin{bmatrix} \varepsilon_1 & 0 & 0 \\ 0 & \varepsilon_2 & 0 \\ 0 & 0 & \varepsilon_3 \end{bmatrix}.$$

As usual, characterization of the propagation comes down to finding an expression for the wave equation, which in turn involves taking the curl of Equation 3.17:

$$\begin{aligned} \nabla \times (\imath\mathbf{k} \times \mathbf{E}) &= \imath\mathbf{k} \times (\imath\mathbf{k} \times \mathbf{E}) \\ &= \imath\omega\,(\imath\mathbf{k} \times \mathbf{B}) \\ &= \omega^2 \mu_0 \mathbf{D} \end{aligned}$$

and hence

$$\omega^2 \mu_0 \mathbf{D} = -\left(\mathbf{k} \times (\mathbf{k} \times \mathbf{E})\right). \tag{3.18}$$

This means that **D**, **E** and **k** are coplanar, as can be inferred from the last line (see also Figure 3.8 a)). Because of Equation 3.14 – Equation 3.17, $\mathbf{D} \perp \mathbf{k}$, and further **B** and **H** are perpendicular to both **k** and **D**. Thus, the Poynting vector $\mathbf{S} = \mathbf{E} \times \mathbf{H}$ and **D**, **E**, **k** all lie in the same plane.

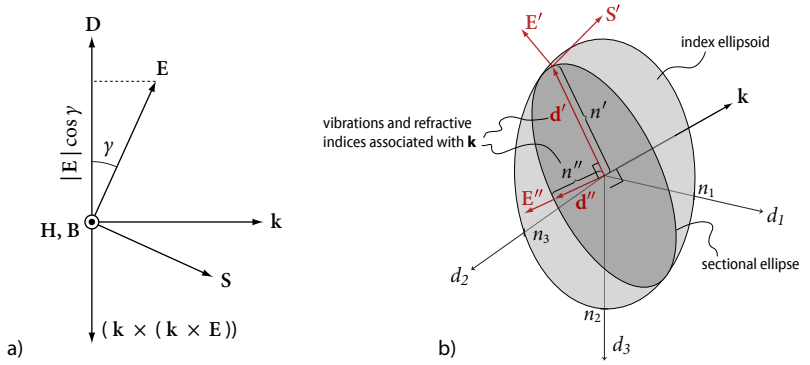

Figure 3.8.: Field vectors a) and index ellipsoid b) in anisotropic media.

This plane, however, is so far not fully determined. For this, take the expression for electric energy density

$$w_e = 1/2\,\mathbf{D} \cdot \mathbf{E} = 1/2\,\varepsilon^{-1}\mathbf{D}\mathbf{D}, \tag{3.19}$$

which is valid in anisotropic media, too. Thus,

$$w_e = \frac{1}{2\varepsilon_0}\left(\frac{D_1^2}{\varepsilon_1} + \frac{D_2^2}{\varepsilon_2} + \frac{D_3^2}{\varepsilon_3}\right), \tag{3.20}$$

because ε^{-1} is a diagonal matrix with components ε_i^{-1}. Defining

$$d_i \equiv D_i/\sqrt{2\varepsilon_0 w_e} \tag{3.21}$$

Equation 3.20 becomes

$$\frac{d_1^2}{\varepsilon_1} + \frac{d_2^2}{\varepsilon_2} + \frac{d_3^2}{\varepsilon_3} = 1. \tag{3.22}$$

This equation describes the surface of an ellipsoid. In order to grasp its physical meaning, let us briefly go back to Equation 3.18. Using the vector triple product it can be rearranged as

$$\mathbf{D} = \frac{c^2}{\omega^2}\varepsilon_0|\mathbf{k}|^2\left(\mathbf{E} - \left(\mathbf{k}'\mathbf{E}\right)\mathbf{k}'\right) \tag{3.23}$$

with the unit vector $\mathbf{k}' = \mathbf{k}/|\mathbf{k}|$ and $c^2 = 1/\varepsilon_0\mu_0$. This is the wave equation of crystal optics. Careful observation (or observation of Figure 3.8, a)) shows that the term in parentheses is simply the projection of $\mathbf{E}$ on $\mathbf{D}$. With the angle γ between the two this can be written as a dot product:

$$\begin{aligned}\mathbf{D} &= n^2\varepsilon_0\frac{\mathbf{D}}{|\mathbf{D}|}|\mathbf{E}|\cos\gamma = n^2\varepsilon_0\frac{\mathbf{D}}{|\mathbf{D}|^2}\mathbf{D}\cdot\mathbf{E}\\ &= n^2\varepsilon_0\frac{\mathbf{D}}{|\mathbf{D}|^2}2w_e\end{aligned}$$

together with Equation 3.19 and $n = |\mathbf{k}|c/\omega$. It follows that

$$n = \frac{|\mathbf{D}|}{\sqrt{2\varepsilon_0 w_e}}. \tag{3.24}$$

Obviously, the surface of the ellipsoid defined by Equation 3.22 can be associated with a refractive index n (compare Equation 3.24 with Equation 3.21). In fact, a wave with a vibration $\mathbf{D} = [d_1, d_2, d_3]$ will experience an index of refraction given by Equation 3.24; its value is the distance from the origin to the ellipsoid along $\mathbf{D}$. For this reason, Equation 3.22 is said to define the *index ellipsoid*. Its three principal axes $\sqrt{\varepsilon}_i$ can be associated with the three principal values $n_i = \sqrt{\varepsilon}_i$ of the refractive index. Moreover, it is rewarding to investigate the surface normal of the ellipsoid at a point $[d_1, d_2, d_3]$ via the gradient

$$\nabla\left(\frac{d_1^2}{\varepsilon_1} + \frac{d_2^2}{\varepsilon_2} + \frac{d_3^2}{\varepsilon_3} - 1\right) = 2\left[\frac{d_1}{\varepsilon_1}, \frac{d_2}{\varepsilon_2}, \frac{d_3}{\varepsilon_3}\right] \propto \mathbf{E}. \tag{3.25}$$

$\mathbf{E}$ a normal on the index ellipsoid.

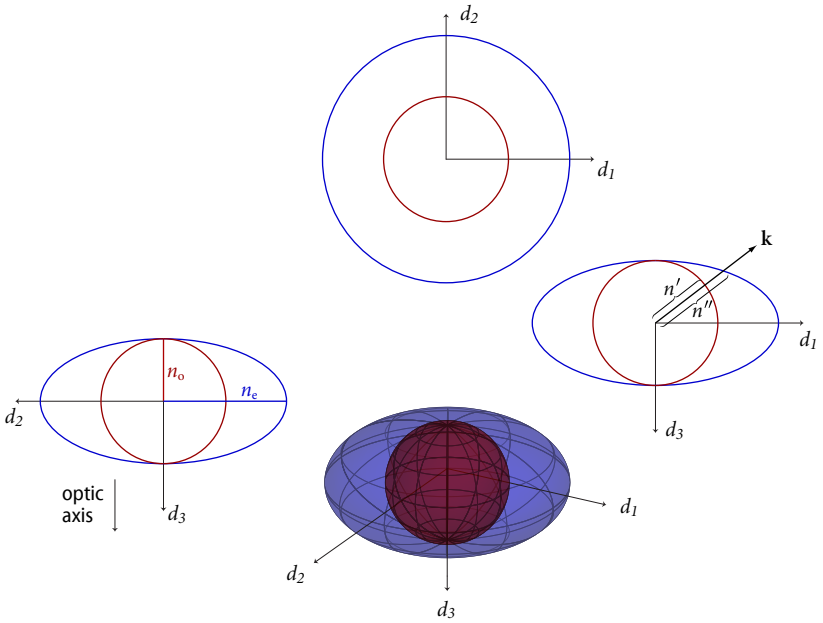

Figure 3.9.: Ordinary and extraordinary refractive index depend on the wavevector **k**. Shown are the index sphere (red volume plot) and the index spheroid (blue), as well as their cut sections along the principal planes.

What remains is how **D** and therefore n can be related to a traveling wave **k**.

From Equation 3.18, **D**, **E** and **k** are coplanar and this restricts the choice of **D** for a given **k**. Suppose a wave is propagating along the direction $[d_1, d_2, d_3]$ in index space, as **k** does in Figure 3.8, b). Since $\mathbf{D} \perp \mathbf{k}$, **D** must necessarily lie somewhere on the plane perpendicular to **k**. Precisely, it is determined by the condition that **D**, **E** and **k** be coplanar. The intersection of the index ellipsoid with the plane defined by **k** is a sectional ellipse. Because **E** has to be normal to the index ellipsoid (Equation 3.25), **E** can only be in an **k**, **D**-plane when **D** points along one of the principal axes of the sectional ellipse. This is a general finding[1]; in order for coplanarity to be fulfilled, **D** has to point along one of the principal axes of the sectional ellipse defined by the intersection between the index ellipsoid and the plane orthogonal to **k**. As the sectional ellipse always has two principal semi-axes, two characteristic vibrations exist for a given **k**, and therefore an anisotropic crystal has two linear, mutually orthogonal eigenpolarizations for any direction of propagation, where the refractive index for each is determined by the length of **d**. Importantly, the refractive index is given primarily by the polarization and not by the direction of propagation. Two

1 A mathematically rigorous treatment can be found in [11] or [115].

waves with different $\mathbf{k}$ can have the same polarization vector and thus the same velocity (which is defined by n).

From now on, the discussion will be restricted to uniaxial crystals, where $n_1 = n_2 \neq n_3$. n_1 is typically called ordinary index of refraction n_o and n_3 is the extraordinary index of refraction n_e. By the recipe outlined in the previous paragraph, n as a function of $\mathbf{d}$ (and therefore $\mathbf{k}$) can be plotted in a polar diagram (Figure 3.9). For each $\mathbf{k} = [k_1, k_2, k_3] \parallel [d_1, d_2, d_3]$, the principal axes of the sectional ellipse and with them the allowed vibrations and are determined. Because two vibrations and therefore two refractive indices are allowed for each $\mathbf{k}$, the result is a surface with two sheets: a sphere with radius n_o (red) and a spheroid (blue) with semi-axes n_o and n_e. They both touch each other at values of $\mathbf{k}$ that correspond to circular sectional ellipses in Figure 3.8, b). Those special $\mathbf{k}$ define the direction of the optic axis of the crystal; propagation along the optic axis is degenerated insofar as there exists only one refractive index for both vibrations.

For any $\mathbf{k}$, picking an n on the spherical sheet in Figure 3.9, the accompanying rays are called ordinary rays; for them, $\mathbf{k}$ and the Poynting vector $\mathbf{S}$ are parallel. When picking the n on the spheroid, accompanying rays are called extraordinary rays and $\mathbf{S}$ and $\mathbf{k}$ are not parallel. Because Snell's law depends on the $\mathbf{k}$, it is only valid for the ordinary ray, where $\mathbf{S} \parallel \mathbf{k}$. When $n_e > n_o$, the crystal is said to be positive uniaxial; it means that ordinary rays will travel faster ($v = c/n$) than extraordinary rays. For negative uniaxial crystals $n_e < n_o$ and $v_e > v_o$.

A wave propagating parallel to the optic axis ($\mathbf{k} = [0, 0, d_3]$) will travel as in an isotropic medium, because in Figure 3.9 its wave vector crosses the two index sheets right where they touch each other. Thus, $n_o = n_e$. All other waves will travel with different velocities depending on their polarization — because two different $\mathbf{D}$ with different refractive indices are allowed for those waves. For example, a wave propagating along d_1 in Figure 3.8, b) allows for two vibrations $\mathbf{d}$ (one along $-d_3$, the other along d_2) that experience different refractive indices.

In general, uniaxial crystals are highly suitable for polarization engineering because the ordinary and the extraordinary rays correspond to orthogonal polarizations and a phase lag will change the global polarization state of the beam (compare to Equation 3.13 and accompanying text). Since $\mathbf{S}_o \nparallel \mathbf{S}_e$, for most cases, an unpolarized beam of light will usually split up in a birefringent crystal — unless it travels parallel or perpendicular to the optic axis. These two special cases are worth being highlighted: as we have seen, parallel to the optic axis n_e and n_o are degenerated and the situation is essentially as in an isotropic medium. When the beam travels perpendicular the the optic axis, however, the two polarizations experience the refractive indices $n_{e,o}$, respectively, but since $\mathbf{S}_e \parallel \mathbf{S}_o$, their tra-

jectories coincide and the beam does not split up. The result is that the ordinary and the extraordinary rays leave the wave plate at the same location but with different phases. This configuration is called a *wave plate*.

Wave plates

A wave plate (or retarder) is a birefringent crystal with parallel faces cut so that the optic axis is parallel to the crystal face. It changes the polarization state of an input beam by introducing a phase slip between two orthogonal polarizations, as has been discussed in the last section. The optic axis — the orientation of the extraordinary polarization eigenstate — is called *slow axis* in positive uniaxial crystals and *fast axis* in negative crystals. One defines the birefringent phase Γ, owing to the different refractive indices $n_{e,o}$ of a beam with wavelength λ as

$$\Gamma = \frac{2\pi}{\lambda}(n_e - n_o)L, \tag{3.26}$$

where L is the thickness of the wave plate. Special cases are the half wave plate

$$\Gamma_{HWP} = (2\tilde{n} + 1)\pi,$$

with introduces a phase π between the ordinary and the extraordinary beam and the quarter wave plate

$$\Gamma_{HWP} = (2\tilde{n} + \frac{1}{2})\pi$$

with introduces a phase $\pi/2$ between the ordinary and the extraordinary beam. Both are said to be of order[1] $\tilde{n} \in \mathbb{N}$. A half wave plate rotated by $\tilde{\theta}$ with respect to the horizontal axis has the Jones matrix [16]

$$\mathbf{M}_{HWP} = \imath \begin{bmatrix} \cos 2\tilde{\theta} & \sin 2\tilde{\theta} \\ \sin 2\tilde{\theta} & -\cos 2\tilde{\theta} \end{bmatrix}$$

and mirrors the angle the polarization vector makes with the optic axis (see Figure 3.10 left). A quarter wave plate has

$$\mathbf{M}_{QWP} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 - \imath\cos 2\tilde{\theta} & \imath \sin 2\tilde{\theta} \\ \imath \sin 2\tilde{\theta} & 1 + \imath\cos 2\tilde{\theta} \end{bmatrix}$$

[16] and shifts one polarization eigenstate by $\pi/2$. It converts linear to elliptical light and vice versa; in particular it converts linear to circular light when the input polarization is at a 45° angle to the optic axis (see Figure 3.10 right).

1 The tilde is to avoid confusion with the refractive index.

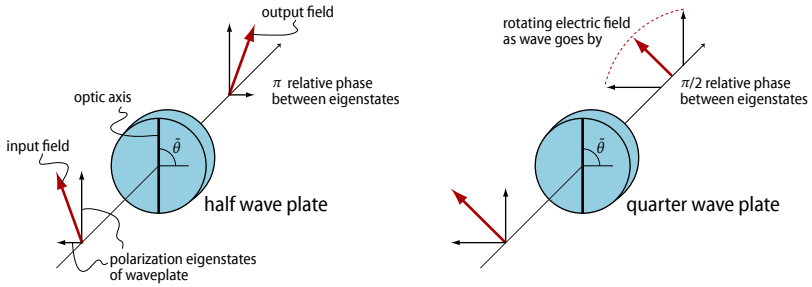

Figure 3.10.: A half wave plate (left) and a quarter wave plate (right) under linearly polarized input.

3.3.3 Experimental realization of birefringent beam shaping devices

Since they can be used to manipulate polarization, birefringent crystals — and especially wave plates, which are easily attainable — are ideal candidates for beam shaping devices. Nevertheless, a single wave plate as in Figure 3.10 obviously does not generate a hollow suppression beam; this can only be attained with a *segmented wave plate* (SWP) that gives different areas of the STED-beam profile different polarization states such that they all interfere destructively in the focus. A simple example is a SWP that rotates the left half of the beam by 180°, while the right half of the beam can pass unaffected. This is similar to the phase plate of Equation 3.9. It will be shown in this section how more complex PSF can be realized in practice and several geometric layouts for segmented beam shaping devices using polarization manipulation will be presented. They not only generate two- as well as a three-dimensional hollow suppression beams, but additionally they can be manufactured and employed with ease. The latter, as has been mentioned, requires that the STED-beam is forged to a hollow spot, whereas excitation and emission light is left unaffected. The only thing that sets these beams apart is wavelength. Thus, we need a *chromatic* segmented wave plate.

How chromatic must it be? Table 3.1 shows a (not exhaustive) selection of dyes that have been used successfully with STED. The eleven dyes in the top rows alone almost fully cover the visible spectrum; additional dyes are available (some of them are shown in the bottom rows). The fifth and sixth column represent the ratio of the wavelength of the absorption maximum to the STED wavelength used in experiments and the ratio of the actual excitation wavelength to the STED wavelength, respectively. It can be seen that, except for the rather exotic nitrogen vacancies and DY-495XL (which have unexceptionally long Stokes shifts), the excitation wavelength is around 0.81–0.85 of the STED wavelength for all

Table 3.1.: A selection of dyes used for STED microscopy together with their absorption maxima (λ_{exc}), the excitation wavelength λ_{exc}^{exp} and the STED wavelength λ_{STED} reported in STED assays (far right column), and the fraction of those two wavelengths in relation to the STED wavelength used in the very same experiments. All wavelengths are in nanometer.

name	λ_{exc}	λ_{exc}^{exp}	λ_{STED}	$\frac{\lambda_{exc}}{\lambda_{STED}}$	$\frac{\lambda_{exc}^{exp}}{\lambda_{STED}}$	**reference**
Alexa 488	499	488	592	0.84	0.82	[76]
Chromeo 488	488	488	592	0.82	0.82	[76]
Oregon Green 488	492	488	592	0.83	0.82	[76]
FITC	495	488	592	0.84	0.82	[76]
DY-495	493	488	592	0.83	0.82	[76]
Citrine	516	516	592	0.87	0.87	
Citrine	516	488	592	0.87	0.82	[39]
DY-485XL	485	488	647	0.75	0.75	[93]
Atto 565	563	532	647	0.87	0.82	[111]
NK51	535	532	647	0.83	0.82	[93]
Atto 633	630	635	750	0.84	0.85	[85]
Atto 647N	644	635	750	0.86	0.85	[72]
GFP	489	490	575	0.85	0.85	[112]
YFP	513	490	598	0.86	0.82	[77]
Nitrogen vacancies	570	532	775	0.74	0.69	[87]
Rhodamine derivatives	531	532	640	0.83	0.83	[13]
Alexa 594	590	570	700	0.84	0.81	[108]

dyes. Turning to Figure 3.11, it becomes obvious that this value is very favorable in connection with wave plates used as beam shaping devices. In this figure, the solid lines are graphs of the retardance over wavelength for three different low-order quartz wave plates provided by many vendors, for example B. Halle Nachfl. GmbH, Germany. Retardance is the phase shift between the ordinary and the extra-ordinary beam after having passed a birefringent crystal cut as a wave plate (see Equation 3.26 on page 41). For example, a retardance of 2.5λ corresponds to a phase slip of π between the ordinary and the extraordinary beam and thus to a half wave plate that mirrors polarization about its optic axis. The wave plates in this case have been manufactured so that the retardance is in fact exactly 2.5 waves for the particular STED wavelength that wave plate has been designed for. These STED wavelengths represent the three most widely used ones, namely 592.5 nm, 647 nm and 750 nm, as listed in the first part of Table 3.1[1]. Owing to the dispersion properties of the mate-

1 It will later become clear that other STED wavelengths are accessible with these three wave plates, too.

rial, the retardance naturally increases for shorter wavelenghts, it is in fact around three full waves for wavelengths of $\approx 0.8\lambda_{STED}$ — exactly the shift required for the excitation wavelength (see Table 3.1.) Thus, the very same wave plate that can be used for engineering the point spread function of the STED beam hardly affects the excitation beam at all. This is exactly the wavelength-dependency required for chromatic wave plates when used as beam shaping devices for STED microscopy. These quartz wave plates will consequently form the basis of the beam shaping devices described in the following.

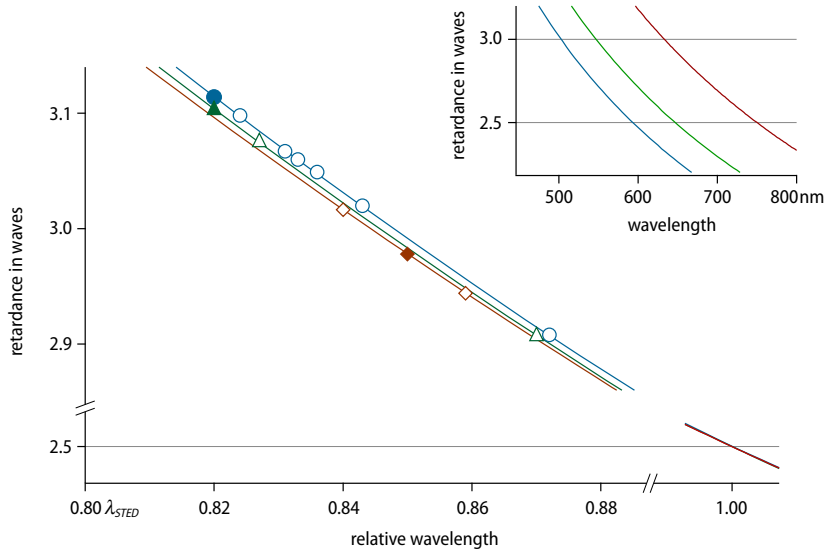

Figure 3.11.: Retardance over wavelength for three low-order quartz wave plates. Solid lines depict the retardance of wave plates for STED with 592.5 nm (blue), 647 nm (green) and 750 nm (red), given by the dispersion properties of quartz, once in absolute wavelengths (inset) and once relative to the STED wavelength for comparison. The open colored markers show the wavelength of maximum absorption for the dyes in Table 3.1 projected onto the dispersion curves; markers are colored according to the STED line the dye would be used with. Solid markers are the most widely used excitation lines. For all dyes, the retardance is between 2.9 and 3.1.

A segmented wave plate for 2D-STED microscopy: easySTED

Combining the chromaticity of quartz wave plates with segmentation provides us with a beam shaping device that can be shared by all beams as described in Figure 3.7 and the accompanying text. Chromaticity al-

lows us to send all beams through the beam shaping device with only the STED-beam being affected, while the right segmentation ensures efficient excited state suppression except at the center. The questions is, of course, what is a suitable segmentation. From Equation B.5, it is clear, that the minimum number of tiles for a donut-like STED-PSF is three[1]. In order

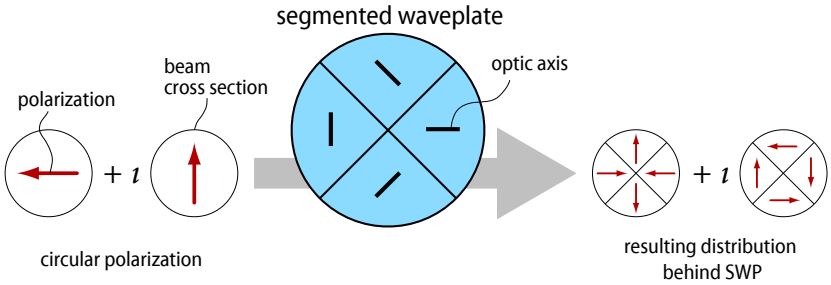

Figure 3.12.: Layout of a segmented wave plate for easySTED (center) showing the orientation of the optic axes of the respective half wave plate that makes up each segment. Right: resulting STED-field behind the SWP for incident circularly polarized light (left).

to simplify fabrication, it was decided on four congruent 90°-tiles of half wave plates arranged like cake slices around the optical axis, similar to what was proposed in [67]. After the geometric layout of the wave plate had been more or less fixed by manufacturing concerns it was necessary to establish the orientation of the optic axis in each segment so that the resulting segmented wave plate a) generates a PSF with a central null and b) generates a PSF that depletes all molecules equally, regardless of the orientation of the transition dipole. The result is shown in Figure 3.12, center. In this chapter, the functioning of this SWP will become clear by purely theoretical reflections, moreover, it was also verified and optimized by a genetic algorithm, described in Appendix C. In short, the algorithm mimics evolution in order to arrive at or close to an optimal solution for the orientation of the optic axes. The pattern with the best suppression qualities found by the algorithm is exactly the one in Figure 3.12. Coincidentally, it can be brought forth by rotating the input polarization by multiples of 90°, allowing us to use half wave plates and exploit the chromatic dependency shown in Figure 3.11.

This segmented low-order wave plate was fabricated by B. Halle Nachfl. GmbH, Germany; all four segments were cut from a single bigger wave plate in order to ensure constant overall thickness. Subsequently, the four

1 Two tiles will only give a line-shaped null.

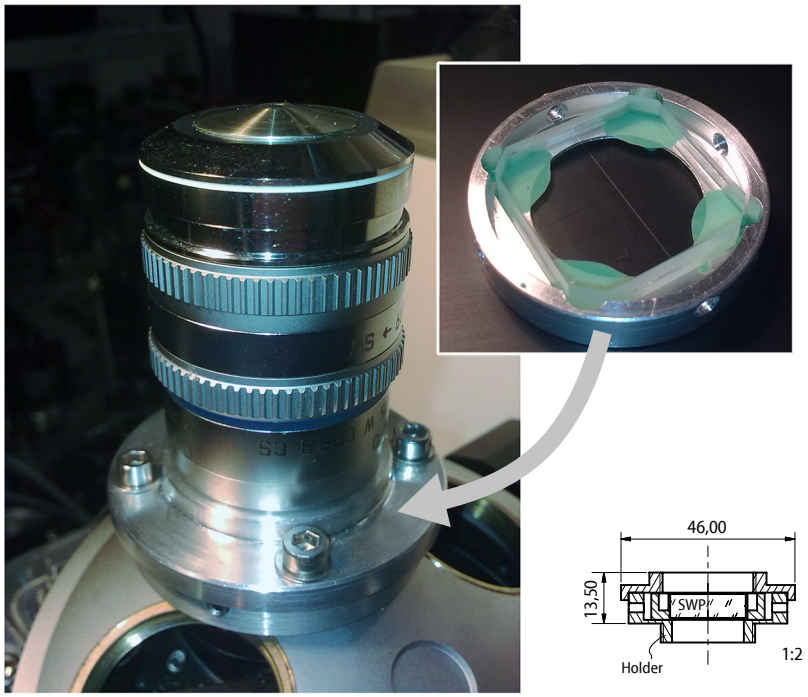

Figure 3.13.: The segmented wave plate (top right, in its holder) is positioned in an aluminum capsule (cut section on lower right) that itself is affixed between objective lens and the revolver of the microscope stand. This permits straightforward implementation into existing microscopes.

segments were simply put together again by cementing them onto a common substrate (BK7), while observing the proper orientation of the optic axes. Three devices were manufactured, one for performing STED at 647 nm (argon-krypton line or 642 nm fiber laser) one for a 592 nm emitting fiber laser and one for 750 nm for use with a Ti:Sa. These figures represent the most widely used STED-wavelengths. The excitation wavelengths are 532 nm, 504 nm and 635 nm, respectively, but can be varied (see subsection 3.3.4, pages 75ff). In all cases, the retardation of the respective wave plate is exactly 2.5λ for the STED line and close to 3λ for the excitation (Figure 3.11). The result is that the STED beam experiences half-wave retardation that leads to a rotation of the polarization plane and gives rise to the polarization pattern shown in Figure 3.12 (right hand side) when illuminated with circularly polarized light (left) generated with an achromatic quarter wave plate placed anywhere upstream of the SWP. Two patterns with a phase shift of $\pi/2$ are thus generated; each stems from one linearly

polarized component of the incoming beam. In contrast, the excitation beam (which is also circularly polarized to ensure that all molecules are effectively excited) is left unaffected and overall donut focusing is achieved for the STED wavelength while the excitation beam focuses to a regular, nearly diffraction-limited focal spot. Most importantly, this allows the SWP to be placed between microscope stand and objective lens as shown in Figure 3.13, and makes it possible to build simpler STED-setups; or to upgrade a standard microscope easily. Hence, the system has been termed *easySTED*.

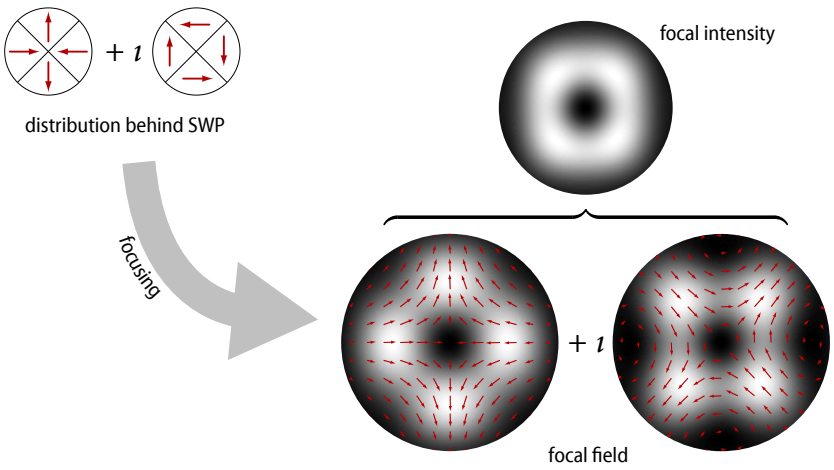

Figure 3.14.: The point spread function generated with an easySTED-segmented wave plate is a donut with a central zero. Red arrows are electric fields, grayscale is squared magnitude.

Figure 3.14 shows the resulting focal field of the STED beam, computed using Equation 3.8 on page 26 with the pupil field generated by the segmented wave plate as shown in Figure 3.12. The STED wavelength used in the calculations was 647 nm, focused by a $NA = 1.4$ oil immersion lens into a medium with a refractive index of 1.5. The incoming pupil field (top left corner) is decomposed into two distinct patterns with a $\pi/2$ phase-difference coming from the two linear components of the circularly polarized beam before the SWP. Each of those patterns by itself gives rise to a focal intensity distribution with a dark hole at the center surrounded by light. From Figure 3.15, it can easily be seen how the central null comes about: obviously, for the polarization sub-pattern with phase $\imath = \exp(\imath\pi/2)$ (red arrows) opposite segments cancel out on the optical axis because they have a phase shift of π. The situation is slightly more complicated for the sub-pattern with phase 0 (blue arrows). Here, the 3 o'clock and the 9 o'clock segments give rise to an electric field polarized

along $-z$ on the optical axis and this field is cancelled by the positive z component generated by the 12 o'clock and the 6 o'clock segments. Overall, this results in an on-axis zero. As when focusing a plain wave, at a certain radial distance from the optical axis the path differences are such that the conditions reverse. This is where the off-axis maximum, the rim of the donut is formed.

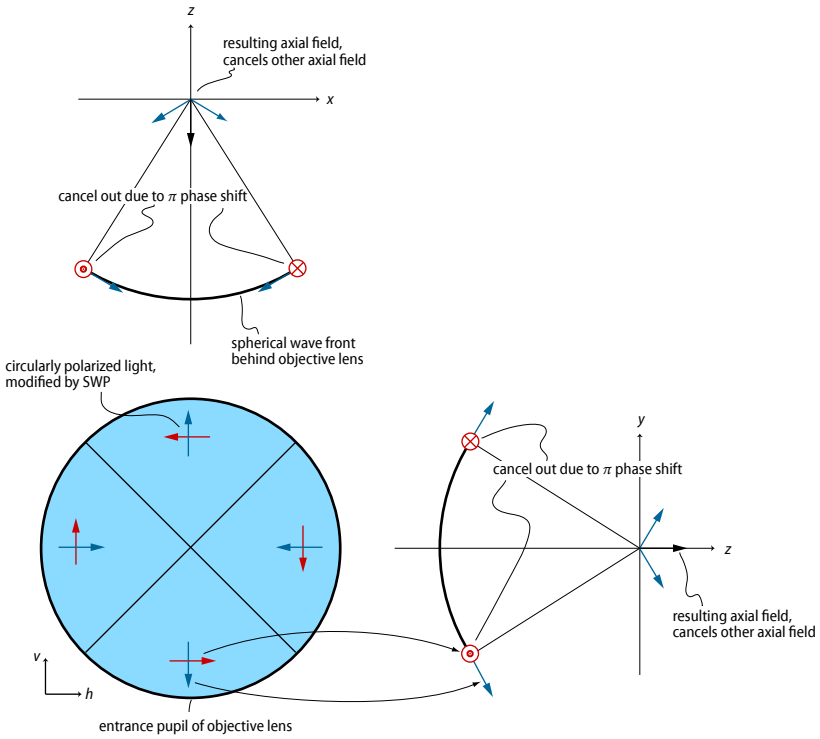

Figure 3.15.: Generation of a central null with a SWP.

Interestingly, the field distribution is in principle the same as for a donut generated with a standard vortex phase plate, aside from the discrete nature of the SWP. A closer look reveals that the field vector rotates counter clockwise in each segment with the segments having a phase difference of 0, $\pi/2$, π, and $3\pi/2$, respectively as can be seen from Figure 3.14. This corresponds to circular polarization with a local phase in each segment depending on the polar angle ϕ like $\exp(\imath\phi)$ and thus a discrete vortex phase plate after Equation 3.10, page 29.

It has been mentioned at the end of section 3.1 that the orientation of the fluorochrome's dipole with respect to the STED-field plays an important role in how well a certain STED-PSF performs. Figure 3.16 shows how the donut created with a segmented wave plate is expected to act on a mole-

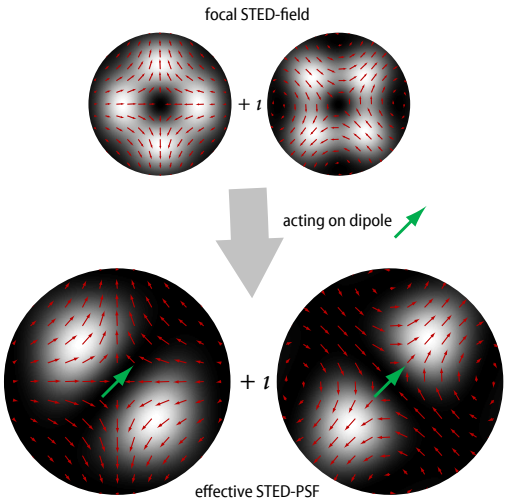

Figure 3.16.: The effective suppression-PSF of an easySTED-SWP, acting on an exemplary dipole (green arrows), is nearly isotropic. Red arrows depict absolute value and direction of the original field, gray scale illustrates the intensity of the effective inhibition-PSF.

cule; here, an arbitrary dipole oriented diagonally has been chosen (green arrow). It can be seen, that the respective STED-PSFs generated by the two linear components of the incoming circular wave have a complementary suppression effect. Since stimulated emission is proportional to $|\mathbf{E}_{STED} \cdot \mathbf{n}|^2$ — thus proportional to the squared cosine of the angle between field and molecule — suppression by the first pattern only takes place to the left and to the right of the donut. In contrast, the second STED field acts in front of and behind the dipole, so that the resulting total suppression and therefore the resulting PSF h_{eff} is isotropic in the focal plane. When viewed over a full rotation of the circularly polarized field vectors, the zero-line makes a full turn around the optical axis. Therefore, when averaged over a wave cycle, the true STED-PSF is independent of the dipole orientation. It is obvious that this is the case for all orientations $\mathbf{n}$.

Figure 3.17 confirms these theoretical considerations. It shows an experimental 647 nm-STED-PSF (red, from an Ar/Kr laser, Newport Spectra Physics) and a 532 nm-excitation PSF (blue, from a PicoTA, Toptica Photonics AG, Germany). Both beams have simultaneously passed the segmented wave plate. The image was taken by scanning 80 nm gold beads (see Appendix A) which act as point scatterers [81], sending an electric field back through the lens which, when focused on a detector, is proportional to the intensity of the PSF at the current scan position. It can be seen that the STED wavelength gets focused to a donut featuring a null

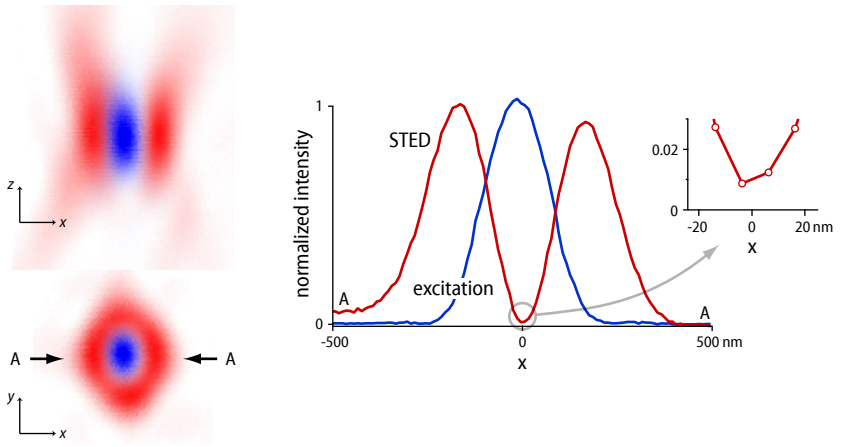

Figure 3.17.: Beam shaping with a segmented wave plate device. Left: focal intensity distributions of the excitation (blue) and STED (red) beam. Both beams have passed the easySTED beam shaping device simultaneously. The 647 nm-STED beam becomes a donut, while the 532 nm-excitation beam is focused to a regular spot. Center: line profiles along the direction indicated by the arrows on the left. The inset on the right shows a magnification of the region around the zero; it proves that the intensity drops below 1%.

that is still perfect within the exactness of measurement[1]. It can also be seen that the PSF for the 532 nm beam is a plain point spread function. The four-leafed complexion of the STED-PSF is a result of the four segments of the polarization plate and is already obvious from the theoretical calculations in Figure 3.14. Both theoretically and experimentally, the modulation along the rim of the donut is less than 20% and goes unnoticed in practice since the STED intensity in the rim is far above saturation anyway. Of course, more segments and eventually a continuous polarization plate are conceivable and would also lead to a more symmetrical intensity distribution (however, with no significant advantages in practice).

Figure 3.18 confirms that the SWP discussed above can in fact be used to improve resolution via STED. It shows confocal and STED data taken with nile red fluorescent beads (∅ 20 nm FluoroSpheres, Invitrogen, USA, see Appendix A for preparation). Both beams were circularly polarized over their full cross section by an achromatic quarter wave plate (B. Halle, Germany) placed in the back of the objective lens (100x/1.4 Oil, Leica, Germany.) Then, all beams passed the segmented wave plate mounted in

1 A measured zero of 1% can be considered more than satisfactory for STED. It will cause an unwanted drop of the peak signal of less than 50% due to residual intensity in the center.

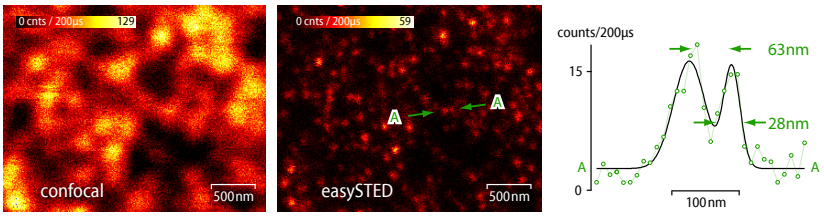

Figure 3.18.: Resolution enhancement with easySTED. Fluorescent nanobeads imaged in standard confocal mode (left) and with high-resolution easySTED (right). The resolution is down to ≈28 nm.

the back of the objective lens (see Figure 3.13). upstream of the quarter wave plate. The excitation laser source was again a pulsed PicoTA triggered by the STED-laser using home-built electronics to adjust the delay. The STED beam came from a mode locked Ar/Kr-laser (Spectra Physics, Canada, repetition rate 80 MHz at ≈200 ps pulse width.) The STED intensity was ≈50 MW/cm^2 in the focal plane. It can be seen from the line profile that the resolution approaches ≈ 28 nm, although some bigger spots are present in the STED-picture as well. They most likely stem from bead clusters and not from a single point-like object. The factor limiting resolution from below seems to be the available STED power; in [35] considerably higher resolutions have been achieved, albeit with about tenfold higher STED-intensities.

Importantly, fluorescence emission, whose maximum is usually close to the excitation line, is mostly left unaffected by the beam shaping device, too, otherwise the signal would drop considerably at the confocal detection pinhole (see also subsection 3.3.4, page 75). This allows the device to be installed directly behind the objective lens in the first place. Note that the confocal pinhole is needed only if 3D-sectioning is required; the STED principle does not require confocality because the region from where the fluorescence originates is predefined by the position of the minimum of the STED beam.

As expected, STED using a segmented half wave plate is applicable to biological samples, too. Figure 3.19 shows a confocal and a STED image of a section of hippocampal rat neurons, where *bassoon* was labeled immunohistologically with A565 (Atto-Tec, Germany) [20]. Bassoon, together with other proteins, is a core component of the presynaptic cytomatrix at the active zone of neurotransmitter release [25]. The resolution is enhanced down to ≈ 35 nm as is substantiated the accompanying cross sections. Figure 3.20 shows another bassoon-stained sample with magnifications to highlight the information gain provided with the aid of STED. Most importantly, the resolution in both examples is the same as observed

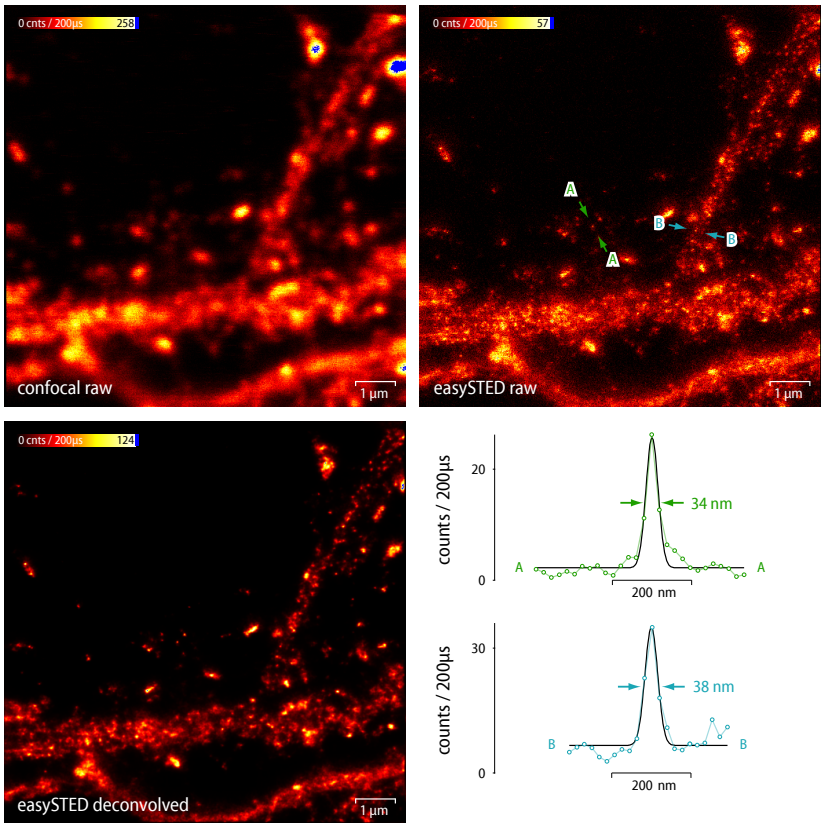

Figure 3.19.: EasySTED on biological samples. The protein bassoon in hippocampal mouse neurons, imaged in standard confocal mode (left). EasySTED (right), with a resolution below 40 nm reveals details unobservable in the confocal image. The easySTED-image is also shown linearly deconvolved.

on the very same setup using the very same sample, but with a standard vortex phase plate instead of easySTED. This shows that the technical simplifications introduced by the segmented wave plate do not come at the cost of performance drops at all.

Finally, Figure 3.21 illustrates the result of applying easySTED to a sample of 20 nm yellow green fluorescent beads (Invitrogen, USA). The STED-beam (focal plane intensity $\approx 350\,\mathrm{MW/cm^2}$) came from a 592 nm continuous wave fiber laser (VFL-P-1000-592, MPBC, Canada); excitation was performed using a 488 nm-line selected with an AOTF (Crystal Technology, Inc., USA) from a 40 MHz-pulsed SC-450-4 supercontinuum source (Fianium Ltd., UK). As shown in Figure 3.7, both beams were simultaneously coupled into the same polarization maintaining fiber (OZ Optics,

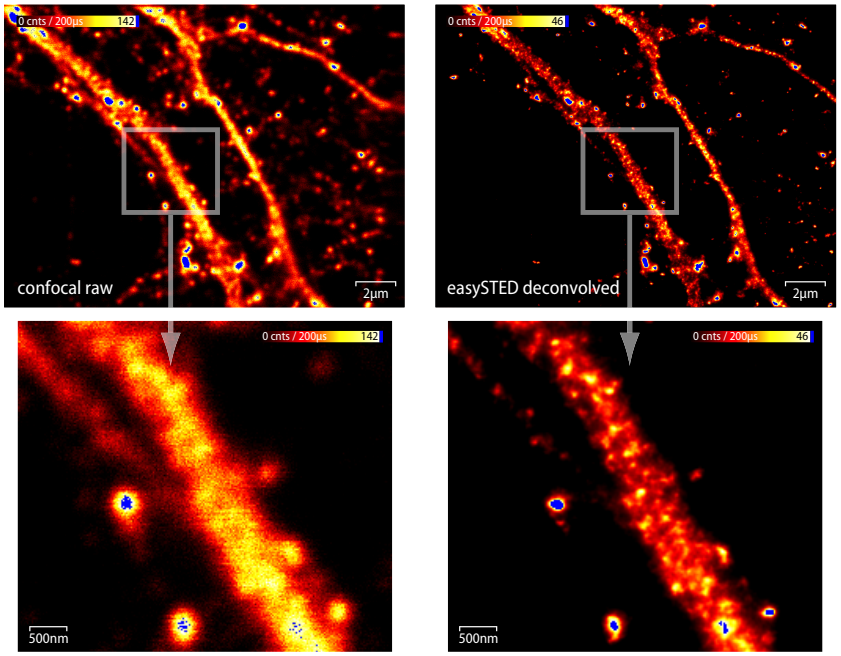

Figure 3.20.: EasySTED on bassoon II.

Canada) that guarantees intrinsic alignment. This is probably the simplest conceivable STED-setup, and yet it delivers a resolution better than 80 nm. This figure seems to be limited by the available STED-power; see section 3.3.4, page 69 for a possible solution.

Molecular Orientation Microscopy (MOM)

At first sight, the focal polarization pattern for a donut shown in Figures 3.12 and 3.14 appears unnecessarily complex. It seems much simpler to employ an azimuthally polarized beam as suggested in [67], because azimuthal polarization in the back focal plane will also transform into a focal intensity distribution with a central minimum. In practice however, this distribution of the electric field in the focal plane is unfavorable for STED, because the transition to the ground state, i.e. the suppression of the excited state by an azimuthally polarized donut, depends on the orientation of the fluorescent molecules. The efficiency of this process is proportional to $\cos^2 \varphi$, where φ is the angle between STED-field and molecular transition dipole. Clearly, when dipole and field are parallel, inhibition is most

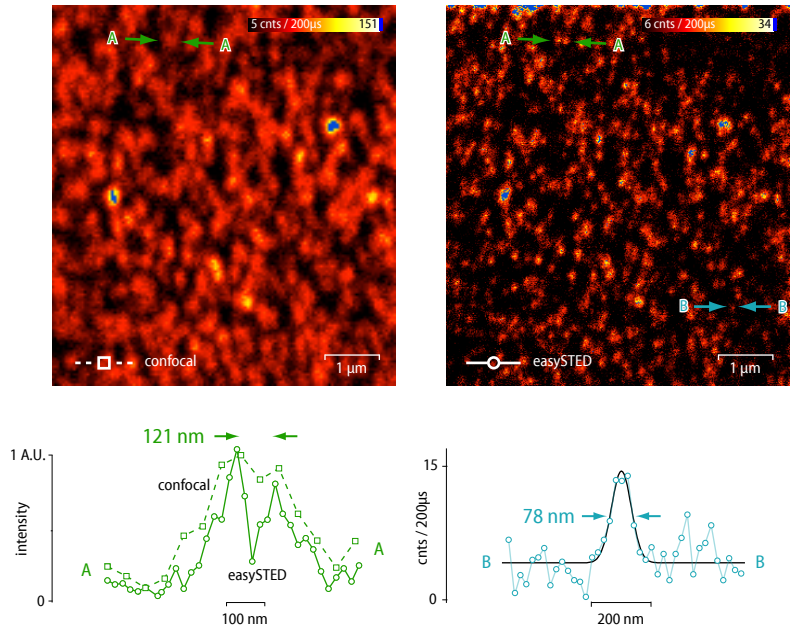

Figure 3.21.: EasySTED with a 592 nm fiber laser. A resolution below 80 nm is attainable by adding a simple cw-fiber laser and the segmented wave plate to a laser scanning setup.

efficient. In contrast, when orthogonally aligned, the field can not interact with the dipole and excited state suppression is zero, however high the STED intensity might be. For this reason, not all STED-PSFs with a central null are suitable for resolution enhancement — a prominent case is the above-mentioned pure azimuthal orientation. It should not be confused with the polarization distribution shown in Figure 3.14; there, the $\pi/2$ phase-shifted part of the polarization pattern in the pupil is for our purposes *not* azimuthally polarized, as the orientation does not directly depend on the azimuthal angle ϕ, but instead has a change of sign. In fact, it produces a similar field in the focus as the polarization distribution with phase 0, albeit rotated by 90°. In contrast, true azimuthal polarization will result in a focal field where the field vector is always tangential to the rim of the donut, and always points in the same direction when going along the rim. Such a focal distribution is somewhat simpler to produce but can not work for STED as we will see, because the unfavorable pattern of the electric field forbids it to interact with dipoles of arbitrary orientation. On the other hand, this means that an azimuthally polarized donut is of great use when it comes to finding out the very direction of molecular dipoles.

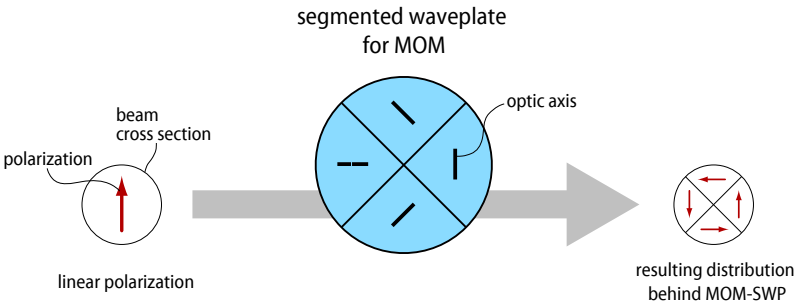

Figure 3.22.: A segmented wave plate for Molecular Orientation Microscopy (MOM). The resulting polarization pattern is azimuthally polarized, owing to the use of a linearly polarized STED-beam and the layout of the MOM-SWP.

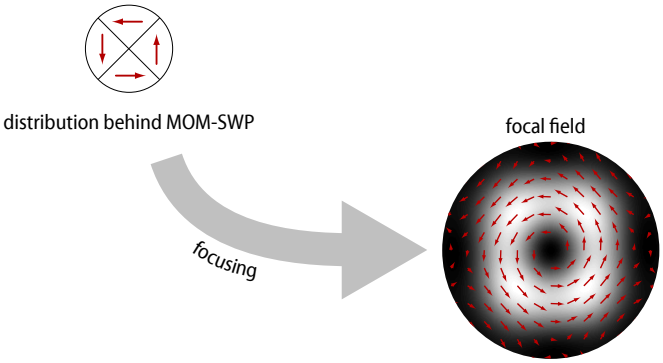

Figure 3.23.: Focal field for Molecular Orientation Microscopy (MOM). When being focused, an azimuthally polarized beam generates a donut with the electric field being azimuthal, too (red arrows). The gray background image depicts intensity.

One can design a segmented wave plate similar to that shown in Figure 3.12 that will generate azimuthal polarization for incident linearly polarized light. With this device, shown in Figure 3.22, STED becomes sensitive to molecular orientation. Hence, this STED-mode has been termed *MOM-STED*, for **m**olecular **o**rientation **m**icroscopy by STED. The incident light is linearly polarized and the layout of the SWP leads to azimuthal polarization in the pupil plane for all points of the wave cycle. In contrast to the easySTED-field, the local electric field is never circularly, but always linearly polarized.

The resulting PSF has again been computed for a 647 nm STED-beam (Figure 3.23). In directions along the dipole, regardless of its actual orien-

tation, the STED-field has no component parallel to the dipole and thus cannot elicit a molecular transition, i.e. it can not affect the exited state. In contrast to Figure 3.12, the parts of the donut where the molecule can not be depleted are the same all over the wave cycle (Figure 3.24). The line-shaped zero stays along the dipole-direction during all times, a consequence of using linearly polarized light and also a consequence of the layout of the MOM-SWP. As a result, independent of the dipole direction, a molecule always finds parallel field-components to its left and to its right and consequently exhibits a narrow effective PSF in this direction. In contrast, a molecule always finds perpendicular field components in its front and in its back and since no inhibition can take place here, the effective PSF retains its confocal width along the projection of the dipole vector onto the *xy*-plane. Hence, the resulting resolution enhancement is not isotropic and this device can be used to investigate the orientation of fluorescent molecules in a sample, provided that the dipoles are rotationally immobile. Thus, this microscope can directly map molecular orientation.

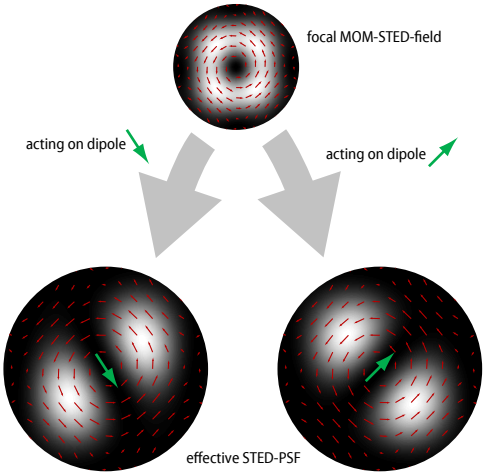

Figure 3.24.: A MOM-STED-image directly reflects dipole orientation. Exemplarily, two dipoles with different orientations are shown together with the true STED-PSF as seen by the molecules for both cases.

Notably, a MOM-STED-PSF (azimuthal polarization) can not be generated by an easySTED-SWP as shown in Figure 3.12 simply by employing linearly polarized light instead of circular polarization. The resulting PSFs (exactly one of the components of the PSF shown Figure 3.14) indeed lead to excited state depression depending on the orientation of the transition dipole, however, the resulting residual fluorescence distribution does not unambiguously reflect this orientation. This can easily be worked out by computing $\mathbf{E}_{STED}$ and $\mathbf{n}$ for such a PSF and a few dipoles.

Nevertheless, a MOM-SWP can be converted to a segmented wave plate for isotropic resolution enhancement (easySTED) and vice versa by placing an additional half wave plate behind either one. In effect, this flips the orientation of the optic axes of two opposing segments.

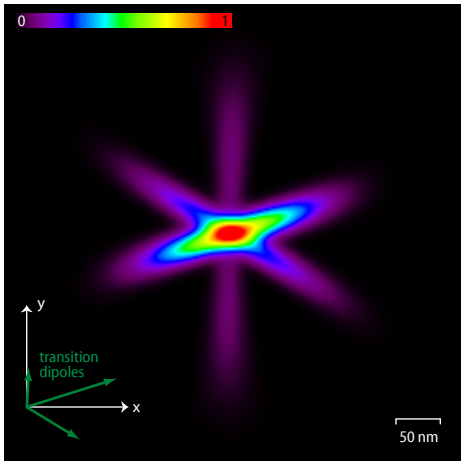

Figure 3.25.: MOM-STED reflects dipole orientation (simulation). Three molecules can be set apart from each other and their orientation is revealed.

Figure 3.25 shows PSFs (computation) for a situation where three randomly oriented fluorescent molecules are located at exactly the same spot. The orientation of the dipoles with unit length is given in the lower left corner, more precisely their projection onto the focal plane. The molecule with the largest z-component (shortest arrow in the image) can be seen only faintly, it is both hardly excited and detected. The other two molecules appear in the final image brightly and with highly elliptical shapes corresponding to the orientation of their transition dipoles. It is obvious that the direction of the dipoles is directly reflected in the detected image, making MOM-STED a highly direct and immediate method to detect molecular orientation. Furthermore, although the molecules are much closer together in this example than the resolution limit (with or without STED), they can still be observed individually, i.e. not only their orientation, but also their position can be inferred from the image.

Figure 3.26 shows MOM-STED in practice. In order to attain sparsely distributed fluorophores, a cover slip has been coated with ∅ 20 nm nile red FluoroSpheres as before (see also Appendix A). Additionally, the cover slip has been immersed in polystyrene solved in toluene. The toluene dissolves the beads, forming a uniform dispersion of fluorophores in polystyrene as the toluene evaporates. The polystyrene serves both as a protective

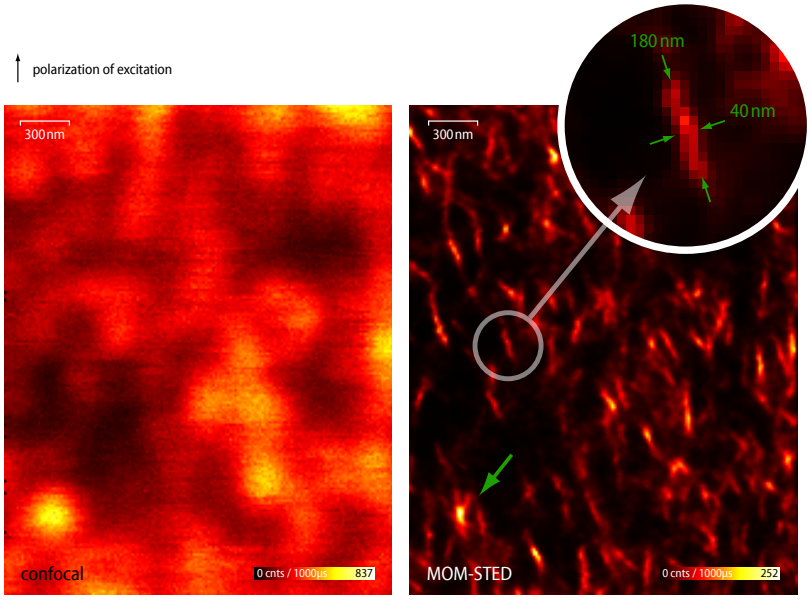

Figure 3.26.: Molecular Orientation Microscopy (MOM) by STED on single nile red molecules, reveals their orientation directly in the raw image (right). Green arrow: three molecules can be separated even in a sub-STED volume solely by the fact that the direction of their resulting point spread functions differ.

environment (molecules were found to bleach less) and as an anchor to the fluorophores. The sample was then exposed to high excitation intensities and thus bleached down to almost single molecule level, as evident from blinking in Figure 3.26. Subsequently, the sample was imaged confocally and with STED using pure azimuthal polarization (MOM-STED). The resulting point spread function depends on the orientation of the molecular dipoles. When the excitation is linearly polarized (here, vertically) only molecules with a vertical dipole component are excited; this explains the preferential orientation of the molecules in Figure 3.26 and it confirms that a MOM-image indeed reflects molecular orientation. When the polarization of the excitation beam is rotated, the dominant direction of the molecules will rotate as well. The magnification shows what, most likely, is a single molecule that has been resolved down to 40 nm perpendicular to its dipole direction; yet it has the 180 nm-width of a standard confocal PSF along the dipole. From this asymmetric point spread function, the orientation of the dipole in the focal plane can be directly inferred. Furthermore, as has been developed in Figure 3.25, molecules which differ in their orientation can be separately imaged and counted even when their

inter-molecular distance is only a few nanometers or below. The green arrow in Figure 3.26 shows a situation where molecules are so close together that they could not have been resolved with an isotropic resolution of 40 nm. In contrast, MOM allows their differentiation due to clearly distinct point spread functions, i.e. their image shape. One could go as far as to fit with several elongated and appropriately rotated point spread functions in order to find the exact location of the emitters. Compare to Figure 3.25 where (at least) two molecules can clearly be discerned although they are actually on the very same spot with a distance of zero. Thus, MOM opens up an additional parameter for separation. The only prerequisite is that the molecules are rotationally fixed during the course of the measurement.

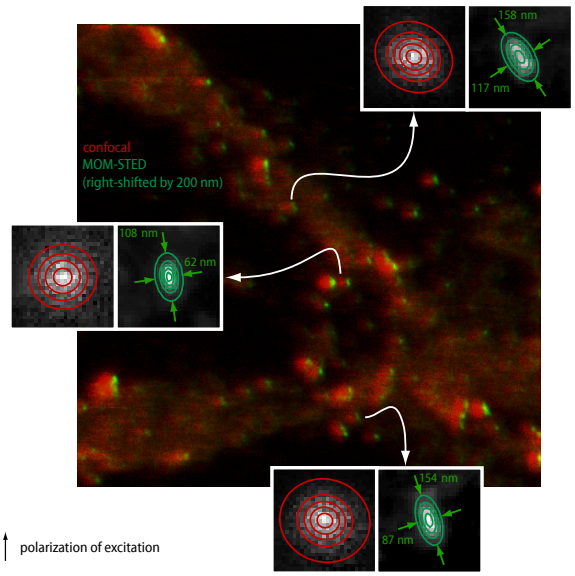

Figure 3.27.: MOM-STED (green) in biological samples. Red: confocal image. The MOM-STED image is slightly shifted to the right, in order to allow the PSFs to be seen more clearly. The direction of the molecular dipoles is reflected in the MOM-STED-PSF.

As for biological specimens, there is considerable evidence [29, 96] that they do indeed contain at least a fraction of fixed molecules, for example in fully assembled actin filaments and bundles. This indicates that imaging of the molecular orientation should become important. For Figure 3.27, Amyloid Precursor Protein (APP) has been labeled with A565 (Atto-Tec, Germany) and imaged confocally and with MOM-STED. APP is an integral membrane protein concentrated primarily in synapses whose native function is not yet fully understood. Most of the research on APP

focuses on its role in Alzheimer's disease, where an amyloid successor of APP is a primary component of the plaques causing this neurodegenerative disease [117]. Here, MOM-STED reveals that a certain fraction of the fluorophores is indeed rotationally immobilized as can be seen from the asymmetric point spread function. Again, it is clearly visible that the resolution with MOM-STED is confocal in the vertical direction (the preferred orientation of the molecules that emit for vertically polarized excitation), whereas the resolution perpendicular to this line is enhanced. Interestingly, the PSFs in this biological sample are a little less elliptical than the PSFs in Figure 3.26. This can be attributed to STED being a little less efficient in a biological environment as compared to a polystyrene immersion. It could also be the result of the molecules being slightly rotationally mobile, however — this in turn provides an estimate of the rotational diffusion constant, given similar environments.

All methods reported up to now for determining the orientation of only a single molecule mostly rely on comparing the (rather complex and dim) defocused diffraction pattern to theoretical predictions and/or on special illumination/detection schemes [65, 80] [34, 95], for instance annular illumination. In any case, these methods demand a good signal-to-noise ratio and are rather indirect. In contrast, MOM sorts out the molecular orientation directly in the sample and the orientation in the focal plane can be seen immediately from the image. As there is no need for fitting a model to point spread functions, large-scale orientational images can be recorded.

A segmented beam shaping device for 3D-STED microscopy

While the segmented wave plate discussed above permits the investigation of samples with considerably higher resolution and therefore more detail, the resolution enhancement is confined to two directions. This is sufficient to obtain sub-diffraction images with samples that are inherently two-dimensional, for example proteins embedded in a membrane. On the other hand, when a sample has features along the optical axis with a distance of less than $\approx 500\,\mathrm{nm}$, these features can not be resolved in a volume scan, no matter how low the lateral resolution might be. Moreover, even when only a 2D-(xy-)slice is taken, signal from above and below the focal plane will be collected when the PSF has a large axial extension. This will elevate background and noise and will decrease contrast. Therefore, most applications demand an effective PSF that has been shrunken along the z-direction as well — and this in turn requires a three-dimensional STED-PSF. Yet, the PSFs generated by the classical vortex phase plate and also by the SWP introduced before exhibit a hollow line along the optical

axis and therefore can suppress fluorescence in the lateral direction only. In contrast, the annular phase plate (Equation 3.11, page 31) has a true hollow dark spot surrounded by light. With this phase plate, however, the distribution of light is somewhat unbalanced with the consequence that the resolution is enhanced mostly along the optical axis, with relatively little resolution increase in the focal plane.

For this reason, the annular phase plate is mostly implemented along with a vortex phase plate as in Figure 3.28, left; together, three-dimensional suppression can be achieved [37, 39, 108]. This means usually that the STED beam is split up with a polarizing beam splitter (PBS), modified by two different phase plates and recombined with another PBS. Unfortunately, this arrangement further complicates beam alignment in a standard setup. While superimposing the STED onto the excitation beam asks for a precision of roughly 50 nm in the focal plane, the prerequisites for 3D-STED with two combined STED beams from different phase plates are much higher. The hollow line created by the vortex phase plate has to be aligned not only with the excitation PSF, but also with the hollow dark spot of the annular phase plate. Although the lateral resolution of the annular phase plate alone is unsatisfactory for STED, the intensity in the PSF is high enough to fill up the zero of the combined STED-PSF when the line-null from the vortex phase plate and the three-dimensional null from the annular phase plate are not perfectly aligned. This has dramatic consequences for signal and resolution.

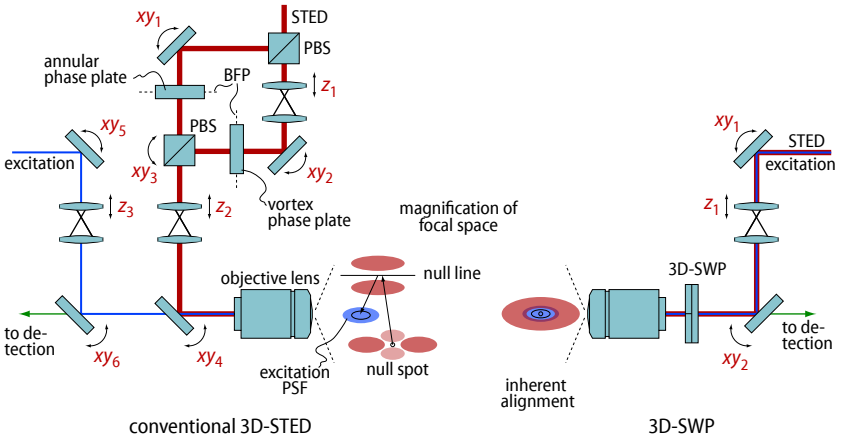

Figure 3.28.: Complexity of conventional three-dimensional STED (left) vs 3D-easySTED (right).

Insofar, the quest for a single, truly symmetric three-dimensional phase plate that can generate a dark spot with ample strength in all directions remains unfulfilled. In fact, [56] shows that this can not be achieved with

phase manipulations at all. However, the possibility to shape not only the wavefront but also the polarization by using birefringent devices opens up new ways to engineer point spread functions suitable for 3D-STED. Here, a segmented wave plate will be introduced that, by appropriately combining polarization rotation with phase shifts, permits STED with a single dark spot fully surrounded by light and is thus able to increase the resolution in all three directions. Moreover, similarly to the segmented wave plate already introduced, excitation and detection light can also be fed through this wave plate, permitting again the placement of a single beam shaping device right in the back of the objective lens. As before, with a combined STED/excitation source there is no need to split up any beams and consequently the setup can be simplified considerably. Figure 3.28, right, shows a simple setup for 3D-STED thus possible. It can be laid out as simple as a conventional laser scanning microscope.

Beam positioning normally requires five degrees of freedom (DOF) per beam. In a conventional 3D-STED system (Figure 3.28, left) this amounts to two DOF for translation into the back aperture of the objective lens and two for rotation onto the optical axis (movable mirrors $xy_1 \ldots xy_4$ for the STED-beams, $xy_{5,6}$ for excitation) and another one for collimation, i.e. for moving the PSF along the optical axis in focal space ($z_{1,2}$ for STED and z_3 for excitation). Collimation is typically done via a 4f-telescope system that also relays the back focal plane into locations accessible for the two beam shaping devices, here an annular and a vortex phase plate. Compare that to the right hand side, same figure, where in a setup with a 3D-SWP only one combined excitation and STED-beam has to be guided into the objective lens: only two moveable mirrors and a telescope for collimation are needed. When the combined STED/excitation source terminates in a fiber, the fiber collimator itself can be adjusted to provide the proper collimation for precise focal z positioning, sparing even the telescope. Fair to say, a fiber collimator could also perform this task for conventional 3D-STED — unless it was for the need to relay the beam into two conjugated back focal planes for the two phase plates. For this, the 4f-system is still required. In contrast, 3D-SWP is placed adjacent to the back aperture of the objective lens itself and does not necessitate a relayed beam. Again, as for the 2D-SWP, the difference between a standard laser scanning and a 3D-STED setup can be as little as another laser source and a suitable beam shaping device adjacent to the objective lens.

The basic idea for the 3D-SWP itself is connected to Figure 3.37 on page 72. There, it will be shown that a lateral shift of the SWP in the back focal plane moves the focal null perpendicularly while the null itself remains intact even for large shifts. Interestingly, the PSF changes axially, too. Figure 3.29 shows computed PSFs when the SWP is shifted upwards in the

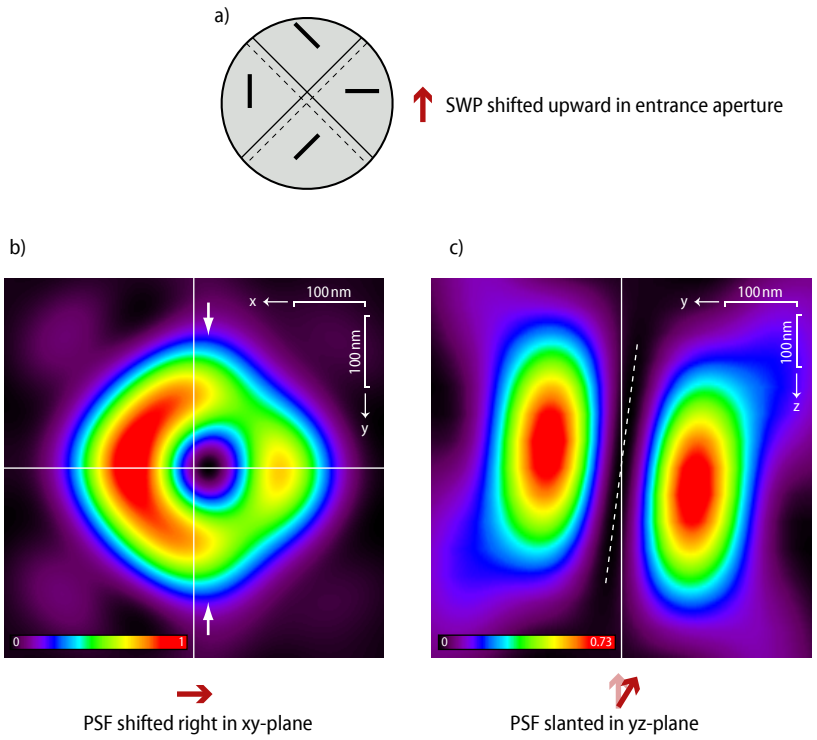

Figure 3.29.: A shifted SWP results in a slanted PSF. Two oppositely slanted line-nulls combined yield a spot-null, i.e. the perfect suppression PSF for 3D-STED.

back focal plane (a); as expected, the null moves along x in the xy-plane (b). Moreover, the PSF is slanted as a whole, in this case in the direction that the SWP has been moved (c). The null-line is not parallel anymore to the optical axis and it is shifted away from the optical axis. The point of describing this is the idea that two incoherent null-lines, slanted along different directions and then overlaid, will generate a point-null, as desired, i.e. a central minimum fully surrounded by high-intensity. The problem is, this would require the seemingly impossible task of shifting a single SWP in opposite directions. Otherwise, by using two SWPs and splitting the beam, one would arrive at the same problems a combination of a vortex phase plate with an annular phase plate has.

The solution is to distribute the four segments of the 2D-SWP into annular ring segments as shown in Figure 3.30. Both plots illustrate polarization patterns generated by the 2D-SWP at phases 0 and $\pi/2$, but for each pattern, half of the fields have been put into an outer ring (a red "0" means no electric field). This way, the centroids can be moved away from the optical axis in opposite directions, and this in turn results in two slanted null-

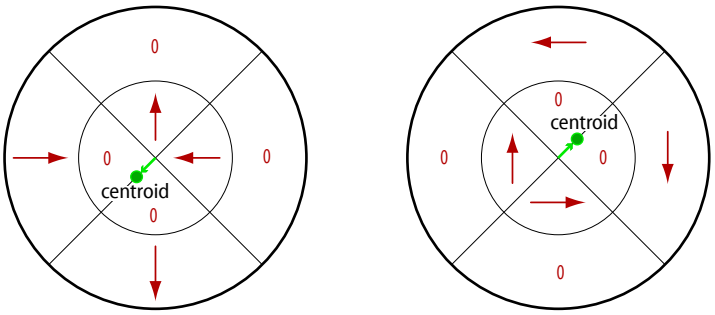

Figure 3.30.: Distribution of the fields into annular rings effectively shifts the centroid of the SWP.

lines pointing in opposite directions. Moreover, because the fields have been distributed into different tiles, the two polarization patterns can be recombined into a single 3D-SWP. Another important ingredient is giving one of the distributions in Figure 3.30 a $\pi/2$ phase shift that allows the two null-lines, that are not only slanted but shifted away from each other, to combine for a spot null. The outcome is a single polarization distribution in the back focal plane that generates a point-null from a single STED-beam. Since the local fields are not linearly polarized, this PSF works on arbitrarily oriented molecules. Of course, it can again be produced with a birefringent beam shaping device, i.e. a 3D-SWP that can also be used in the excitation- and STED-beam paths, just like the 2D-SWP, and with the same advantages. This leads to Figure 3.31, which shows the orientation of the optic axes for the 3D-segmented wave plate together with the phase shift associated with each segment (either 0 or $\pi/2$).

Such a wave plate is considerably harder to manufacture than the 2D-SWP because of the annular rings. In order to simplify fabrication, the principle of the 3D-SWP shown in Figure 3.31 was distributed into three simpler segmented wave plates. Figure 3.32 illustrates how a stack of two segmented half wave plates and a quarter wave plate act together. A prototype was assembled (B. Halle Nachfl. Germany) for 647 nm, again by cutting and milling of bigger low order quartz wave plates and reassembling the appropriately rotated pieces on BK7. Using three simpler wave plates instead of one complex SWP circumvents the handling of small quartz pieces and ensures precise alignment on the glass substrate. For the prototype discussed here, three separate SWPs were made and stacked manually for evaluation purposes. In the final product, they can of course all be cemented on top of each other, separated only by a thin BK7 flat. In any case, the stack is less than 12 mm in height and can still fit beneath the objective lens. Just as the 2D-SWP, it allows upgrading an existing LSM

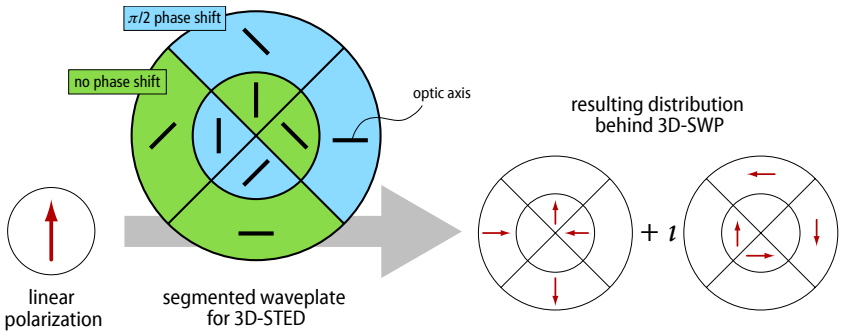

Figure 3.31.: A segmented wave plate for 3D-STED. The wave plate is shown in the center together with the orientation of the optic axes and the phase shift it introduces in each of the eight segments. When used with a linearly polarized STED-beam (left), the result is a light distribution which can be written as a combination of two beams with phases 0 and $\pi/2$ (right).

with ease. A STED system equipped with this 3D-SWP will be called 3D-easySTED.

The final quarter wave plate demands a closer look. It is responsible for the right phase (0 or $\pi/2$) in each segment after the polarization has been rotated — how can this be done with a quarter wave plate? The answer is that, generally, after a linearly polarized beam has passed a quarter wave plate it, by definition, has acquired a phase of $\pi/2$ when it was polarized along the optic axis and a phase shift of 0 when it was polarized orthogonal to the optic axis (see subsection 3.3.2). Looking at Figure 3.30, it can be seen that the optic axes of the segmented quarter wave plate are oriented in such a way that each of the beams in the eight segments "sees" exactly the orientation of the optic axis it needs to acquire the phase shift required for its segment.

The reason why a quarter wave plate is used for the final phase shifts and not a plain phase element is the same reason why birefringent material is used in the beam shaping devices at all: wave plates come off the shelf with dispersion properties that match the wavelength difference between excitation- and STED-beam. Just as the half wave plates are half wave plates only for the STED-wavelength, the segmented quarter wave plate is in fact a three-wave-plate for the excitation beam and does not affect it in any respect.

Figure 3.33 shows cut sections of the STED-PSF of a 3D-SWP (more precisely, a stack of segmented half and quarter wave plates) from different directions. Experimental PSFs were recorded with a 647 nm Ar/Kr laser

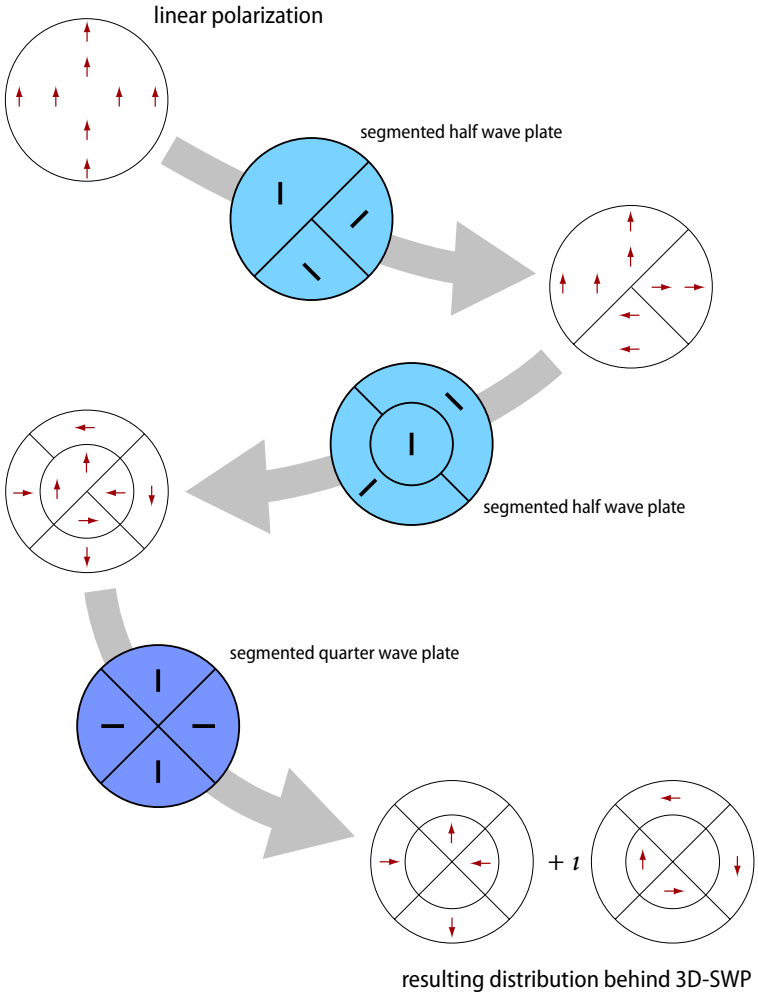

Figure 3.32.: Realization of the 3D-SWP as a combination of three segmented or annular wave plates. Incoming linear polarization is converted into two donuts with slanted null-lines, which when combined with a $\pi/2$ phase slip generate a spot-null.

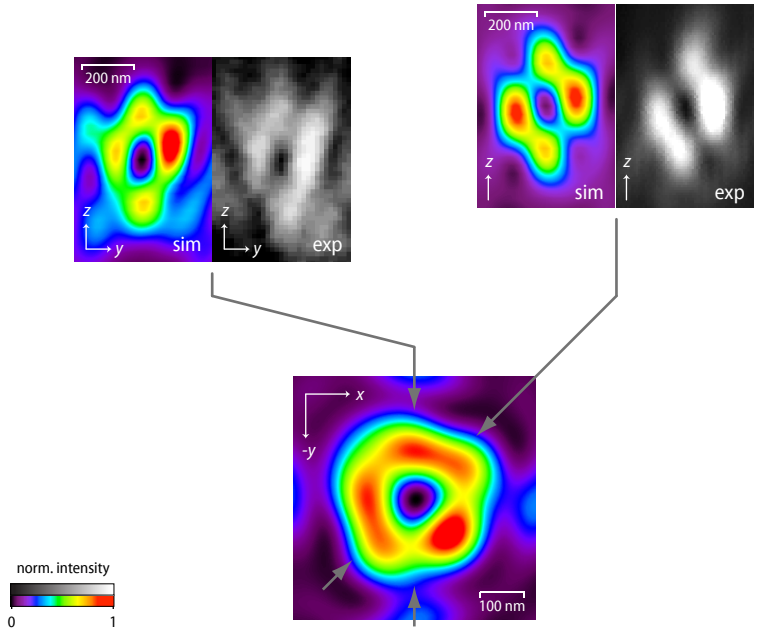

Figure 3.33.: Simulated PSF of a 3D-SWP and measured cross-sections (greyscale). The PSF is a spot-null fully surrounded by light.

on 80 nm gold beads, as before with the 2D-SWP; again, the 3D-SWP was placed in the combined beam path just upstream of the objective lens. Obviously, from the computational PSFs, radial symmetry is broken. This is because with a $\pi/2$ phase shift the two separate donuts are not fully decoupled and their corresponding PSFs mix in the focus. It follows that the intensity surrounding the central null is not the same everywhere, for example it is down to 65–70% of the maximum intensity for a few isolated points in the focal plane. Along the optical axis, the intensity is 73% of the global maximum. All this will slightly distort the effective PSF, but it remains spot-like and the asymmetries go unnoticed in practice because stimulated emission is highly saturated anyways.

Figure 3.34 demonstrates STED with a 3D-SWP. It shows axial resolution enhancement on 20 nm nile red fluorescent beads, prepared as before. The resolution along the optical axis is improved $\approx$ 2.7-fold as the effective PSF is shrunk from 564 nm to 211 nm, axially (see line profiles). The STED image shows much less beads than the confocal image above it because the resolution is simultaneously enhanced in the focal plane by the very same SWP. The missing beads simply fall out of the range of the smaller effective PSF when STED is applied.

This is substantiated in Figure 3.35, where an *xy*-section is plotted, once

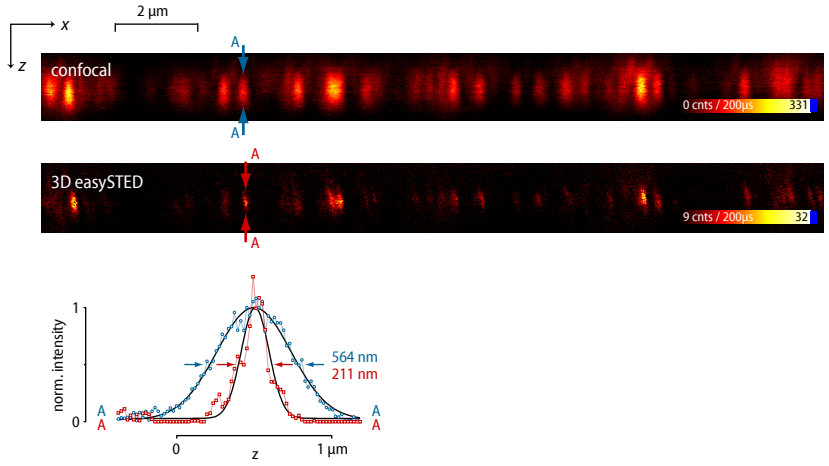

Figure 3.34.: Cut sections along *xz* of nile red beads on a coverslip, imaged with a SWP for 3D-easySTED. The line profiles correspond to the bead marked with blue arrows in the confocal image, respectively the same bead marked red in the 3D-STED image. Axial resolution is improved by a factor of 2.7.

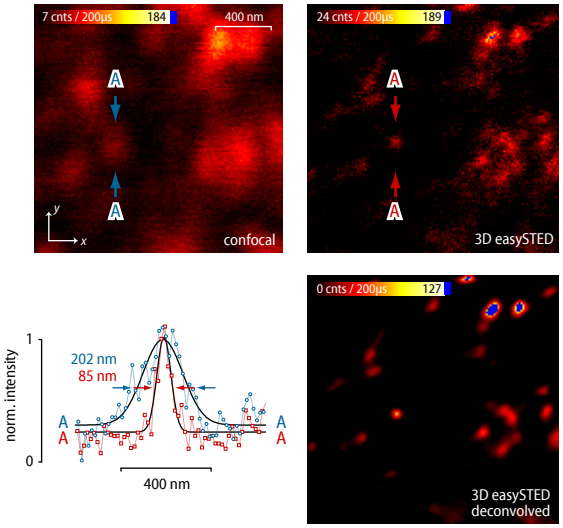

Figure 3.35.: Focal plane image of nile red beads on a coverslip, taken with a SWP for 3D-easySTED, and line profiles of a bead in confocal mode (blue arrows/blue line) and the same bead in 3D-STED-mode (red). Lateral resolution is improved by a factor of 2.4.

taken only in confocal mode (top left), then with 3D-easySTED (top right). The STED image is also shown deconvolved (bottom right). Again, line profiles confirm a $\approx$ 2.4-fold resolution improvement in the STED image. Altogether, the effective focal volume has been shrunk down to 6.5% of the original confocal volume. It is limited here most likely by the available STED power only. Importantly, the resolution improvement is symmetrical; it is neither in the focal plane only, as with a vortex phase plate or a 2D-SWP, nor is it mostly along the optical axis, as with an annular phase plate[1]. Equally important, this 3D-easySTED system is as simple as 2D-easySTED. No additional beams or phase plates are needed, just a different SWP is put in place. Obviously, it is significantly less complex than any three-dimensional STED system reported up to now.

3.3.4 Robustness considerations for segmented wave plates

The segmented wave plates described in the previous section have been designed to simplify the set-up, operation and maintenance of STED microscopes. This goal has been achieved by using standard wave plates whose chromatic dispersion allows them to be used as wave plates for the STED-wavelength while they are essentially non-birefringent for the excitation wavelength. Consequently, SWPs can be used in conjunction with a common STED- and excitation source providing intrinsic alignment. Nevertheless, in practice this can only make sense when the installation of the SWP is easier to a considerable degree then aligning beams manually. Issues that might arise are:

- Can the SWP itself be aligned more easily than two separate beams?
- What happens if the STED-laser does not match the design wavelength? What about broad-band (pulsed) lasers (e.g. Titanium:Sapphire)?
- Does not detection become a donut, too?

This section will nullify the before-mentioned problems. It will be demonstrated that installing a segmented wave plate not only comes at virtually no cost of performance but also yields huge gains in simplicity and therefore robustness.

Tolerance against misalignment in the pupil

The intensity zero of a vortex tolerates lateral displacements of its beam shaping device in the pupil as can be seen with a small computation. For

1 Obviously, an annular phase plate can also be made a shared beam shaping device using annular wave plates.

this, the pupil function $A(\rho, \phi) = \exp(\imath\phi)$ is written as $A(k_x, k_y) = k_x + \imath k_y$. Disregarding apodization and polarization for the course of this estimate, the scalar focal field of a vortex is

$$E(x,y) \propto \iint \left(k_x + \imath k_y\right) e^{-\imath(k_x x + k_y y)} \, \mathrm{d}k_x \mathrm{d}k_y \tag{3.27}$$

$$= \int e^{-\imath k_y y} \int k_x e^{-\imath k_x x} \, \mathrm{d}k_x \mathrm{d}k_y + \imath \int e^{-\imath k_x x} \int k_y e^{-\imath k_y y} \, \mathrm{d}k_x \mathrm{d}k_y. \tag{3.28}$$

Notably, virtually the same field is created by the segmented phase plate for 2D-STED (section 3.3.3). It can similarly be written as $A = A_v + A_h = k_x + \imath k_y + k_y - \imath k_x$ because the high-frequency jumps at the discontinuities are not transferred to the focus. Thus, the following calculations also hold for A_v and A_h (which is simply A_v rotated by 90°) and therefore for the 2D-SWP as well.

Performing the inner integrations while for simplicity assuming a square aperture[1] with edge length $2a$ yields

$$E(x,y) \propto \imath \underbrace{\frac{x\cos x - \sin x}{x^2} \int_{-a}^{a} e^{-\imath k_y y} \mathrm{d}k_y}_{\text{+ -}} - \underbrace{\frac{y\cos y - \sin y}{y^2} \int_{-a}^{a} e^{-\imath k_x x} \mathrm{d}k_y}_{\text{- +}}. \tag{3.29}$$

The remaining integrals of course are $2\sin(ai)/i, \;\; i \in x, y$ and provide the envelope (due to diffraction at the aperture) for the other terms that convey the interesting features. The other terms each for itself is a function with a zero line along y for the left term and along x for the right term and opposite positive and negative lobes adjacent to the zero line. Hence, the dark hole in the center of the donut can be decomposed into the sum of a vertical and a horizontal zero line.

Introducing a small shift of $-c = \text{const}$ along the k_x-axis[2] in the pupil yields $A(k_x, k_y) = (k_x + c) + \imath k_y = k_x + \imath k_y + c$. In the focus, the field produced by the perturbation c is of course $c \int_{-a}^{a} e^{-\imath(k_x x + k_y y)} \, \mathrm{d}k_x \mathrm{d}k_y = 4c \sin ax \sin ay/xy$, because c is a plane wave and gets focused to a normal point spread function. The interesting point here is the fact that the

1 A square aperture will only modulate the envelope but not the characteristic features of the PSF.

2 A shift along an arbitrary direction can always be produced by rotation of a shift along k_x.

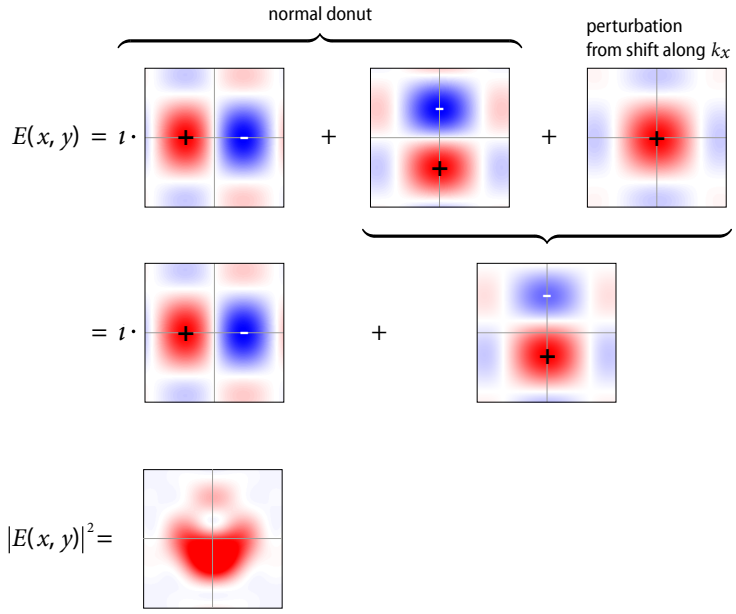

Figure 3.36.: A 2D-SWP shifted in the pupil. The focal field generated by a vortex phase plate or a segmented half wave plate is the sum of two cylindrical TEM01 modes; one real-valued and one imaginary-valued. A perturbation introduced by a lateral shift of the pupil field translates the intensity zero in a perpendicular direction. The null intensity stays zero.

real part of the donut point spread function (the second term in Equation 3.29) stems from the imaginary part in the pupil plane (the $\imath k_y$ in Equation 3.27), whereas a shift along k_x of $-c$ produces a real-valued PSF in the focus. Thus, a perturbation along k_x can only couple with the real-valued focal field generated by the k_y — the field with a horizontal zero line. Therefore, a perturbation introduced by a horizontal shift in the pupil will result in an upward shift of the horizontal zero line in the focus (see Figure 3.36), and consequently an upward shift of the intensity zero point. Similarly, a horizontal shift of the pupil field will result in a vertical shift of the intensity zero. Mathematically, all this is a result of the shift property of the fourier transform:

$$\begin{aligned} E(x,y) &= \mathcal{F}\left[A(k_x + c, k_y)e^{-\imath k_z(k_x+c,k_y)z}\right] \\ &= e^{-\imath cx}\mathcal{F}\left[A(k_x, k_y)e^{-\imath k_z(k_x,k_y)z}\right]. \end{aligned}$$

Here, the phase term $e^{-\imath cx} = \cos cx - \imath \sin cx$ holds the perturbation term that couples to the real or imaginary part, respectively, of the donut

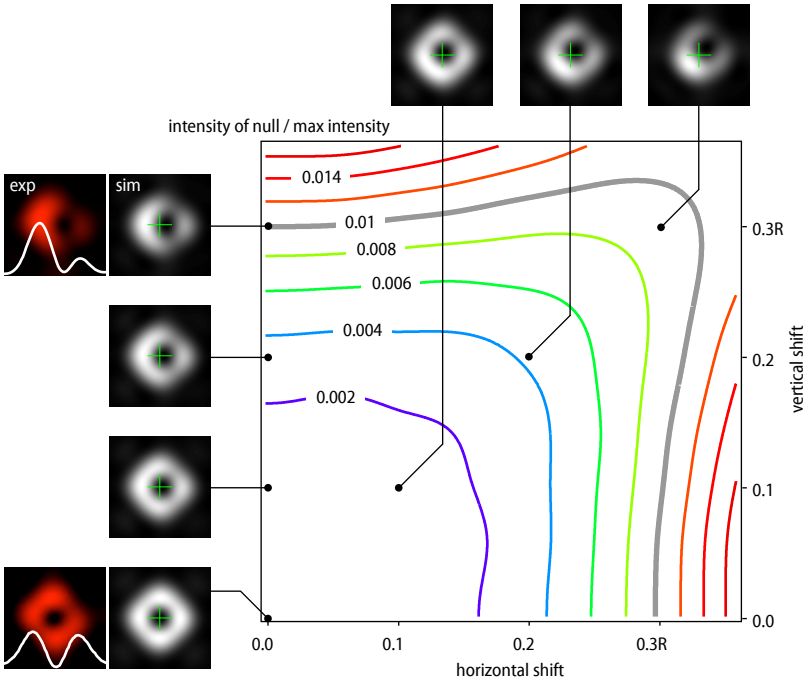

Figure 3.37.: Behavior of a 2D-SWP under shift in the pupil. The contour plot shows depth of the null (relative to maximal intensity without shift) vs shift in pupil plane and resulting PSFs. Measured PSFs (red color ramp) are accompanied by horizontal line profiles through the PSF center. The null stays intact for shifts up to about a third of the entrance-aperture radius R.

PSF generated by the fourier transform of the unshifted pupil function $A(k_x, k_y)$. Most importantly for STED microscopy, for $|c| < k_{max}$ (k_{max} is the maximal horizontal k-component allowed by the lens aperture a) the zero is indeed a true zero, because a focal position y can always be found for which the field from the perturbation term equals the negative lobe of the real or imaginary part, respectively, of the donut PSF. The greater the shift c, the higher is the strength of the resulting perturbation PSF, and the farther out the resulting null will be, but it will be a true zero.

Figure 3.37 shows simulated and experimental PSFs, when the segmented wave plate for 2D-STED is not aligned on the optical axis, but shifted laterally. The axes denote the shift in the back focal plane in terms of entrance aperture radius R; the contours mark the depth of the null relative

to the intensity maximum for an unshifted SWP. The 1%-limit is reached when the lateral displacement is about 0.3 of the aperture radius. For a 100x/1.4 oil-immersion lens $R = 2.8\,\text{mm}$ and thus the maximum displacement is about $\pm 0.8\,\text{mm}$. At this point, the zero quality is still sufficient for STED although the rim of the donut begins to break apart, leading to an asymmetric effective PSF. However, the tolerated displacement is well within several tens of a millimeter and this is of particular importance for beam scanning setups. Assume the SWP has been placed as shown in Figure 3.13 (page 46), then it will be close to where the objective lens pupil is supposed to be, which in turn has been moved away from its designated place by about 10 mm. Pivoting the beam in the original pupil plane (as it is done for beam scanning) will lead to a perceived shift of the SWP in the de facto-pupil plane of the objective. Nevertheless, a lateral shift limited to $\pm 0.8\,\text{mm}$ in the new pupil plane corresponds to scan angles that still allow a field of view of about $\pm 100\,\mu\text{m}$ for a 100x/1.4 objective lens — a figure that is hardly restrictive in any case. In fact, the SWP can simply be centered by sight without any further aids.

So far, lateral displacements have been covered. But when a wave plate is tilted, the path length and therefore the retardation are subject to change, too. Path length enters via $1/\cos$, as usual for an optical flat, additionally, the extraordinary refractive index approaches the ordinary index of refraction, as is clear from Figure 3.9, page 39. Since the direction of the axes in a SWP is mixed, complex patterns can arise that are hard to generalize because of the multitude of possible combinations of the parameters. Nevertheless, tilt sensitivity is an important parameter for a beam scanning setup, where the beam is pivoted about a point close to the SWP. According to information from B. Halle, Nachfl. GmbH, tilt sensitivity for the wave plates used here can be approximated by $\Delta\Gamma = \pm(0.1\,\text{nm}/\text{deg}^2)\varphi^2$, when φ is the tilt angle in degrees. Because a SWP tolerates chromatic deviations (see the following section), the tilt angle can be several degrees. Thus, when an SWP is used at its design wavelength, the attainable field of view should not be limited by the effects of tilt.

As for the 3D-SWP, these results hold equally. It in effect generates the same polarization patterns as the 2D-SWP, however already shifted in the pupil plane. It is clear from Figure 3.31, that an additional shift of both patterns will, in first order, not change the situation. The two slanted null-lines can still interfere to form a null spot, although the rim of the donut can unacceptably change shape for large shifts, as before with the 2D-SWP.

Lastly, it might occur that the input polarization is not perfectly circular as specified in Figure 3.12 on page 45. This has no effect on null depth at all, since the two polarization patterns generated by each linear component of the circular input light are orthogonal and each by themselves

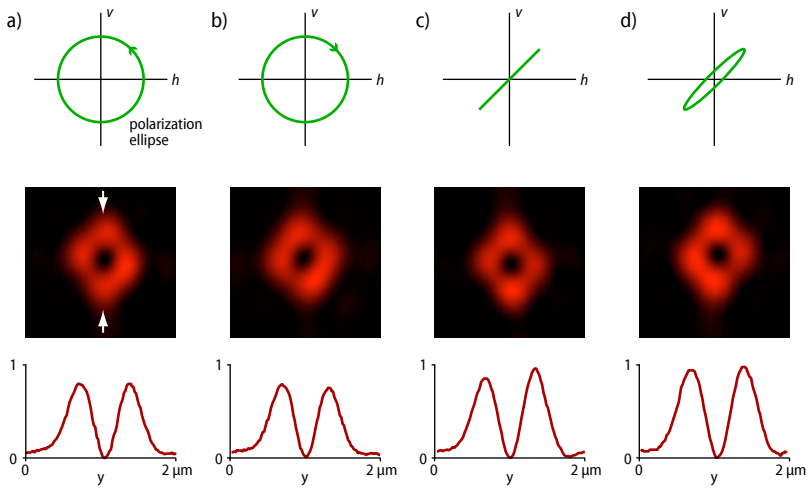

Figure 3.38.: PSFs of a 2D-SWP under different input polarizations (middle row). Top row: polarization ellipses of the input polarization; bottom row: vertical line profiles through the center of the measured PSFs (along the white arrows in (a)). The null stays intact for arbitrary input polarizations.

generate a perfect null. They are in effect two bases of the five-dimensional null-space of this SWP (compare to Appendix B, pages 93ff). As such, they can be combined with any strength or phase difference and yet will always generate a perfect null. For Figure 3.38, PSFs have been measured (using gold beads) for four different input polarizations. (a) and (b) show the resulting PSFs and line profiles through the center for oppositely circularly polarized light; (c) is for linearly polarized and (d) for elliptical input light. The polarization has been altered solely by rotation of the essential quarter wave plate and an additional half wave plate without any other measures, in particular subsequent re-alignment of the beam. When one or both wave plates are slightly tilted, beam pointing changes and thus the PSF, according to the previous section, can slightly change shape, too. Nevertheless, the null stays below 1% for each case. In practice, (c) and (d) are undesirable for STED, because the preference of a single polarization pattern in the focus will result in orientation sensitive effective PSFs, similar to MOM-STED. Still, relaxed requirements for the polarization further simplify setting up an easySTED-system. In contrast, a STED microscope using a vortex phase plate highly depends on perfectly circular input light (see pages 29ff); it in fact depends on circularly polarized light of the right handedness — a system laid out for left circular polarization will not work with right circular polarization and vice versa. A system using an SWP will work with both.

This means that for example several laser sources can easily be combined losslessly using a polarizing beam splitter cube in order to double the available STED-power. Also, multi-spot arrangements are conceivable when the STED-beam is split up by Wollaston prism stacks [90], whereby multiple linearly polarized beams are generated in the pupil plane. Those beams are tightly interleaved in a sense that for any beam the polarization in the neighboring beam is orthogonal. Thus, Wollaston prisms are not viable when used with a standard vortex phase plate — after having passed the quarter wave plate every other beam will be polarized the wrong way and will generate a STED-PSF with the null being filled up. In contrast, a 2D-SWP accepts all circular beams regardless of their handedness and can be used to great effect with Wollaston prism stacks for multi-spot STED.

Again, all this is true for the 3D-SWP, too, as the same polarization patterns are present. The individual patterns are simply pre-shifted, but they are still bases of the null space. Being able to combine two STED-sources for doubling the power is in particular helpful for 3D-easySTED, where the intensity in the pupil plane is spread-out over a greater volume of the STED-PSF. Also, three-dimensional multi-spot STED is enabled by the 3D-SWP.

Detection efficiency and performance under chromatic deviations

The most serious chromatic deviation that might arise is when the STED source does not meet the design wavelength (or vice versa) or when the STED beam has a band width of several nanometers such as with pulsed (e.g. Ti:Sa) devices. When deviating from the blueprint wavelength, the wave plates will no longer be true half wave plates and the central zero will fill up. As soon as its depth exceeds a few percent of the rim maximum, the useful signal gets heavily suppressed. In practice, 1% STED intensity in the null is considered tolerable[1], whereas STED is hardly practicable above 5%.

Neglecting high-NA focusing, a rough estimate suggests that the zero depth follows a $\cos(\Gamma\pi/\lambda)$ dependence, when Γ denotes retardance in waves. For this reason, the null depth can expected to be flat around the design STED-wavelength. A more rigorous analysis is given in Figure 3.39; it illustrates the dependency of the central zero on wavelength obtained both from numerical simulations (left) and experiments (right). In order to measure the PSF, a Ti:Sa laser (Mai Tai, Newport Spectra Physics, CA) was employed together with a gold bead sample. Although the output of

1 In fact, the error of measurement is on the order of 1% with gold beads.

the Ti:Sa has a broad spectrum, it was chosen because of its tunability[1]. Still, it can be seen that the null stays intact in the interval ± 7 nm around the design wavelength (in fact, a SWP designed for a 647 nm krypton line has also been successfully used with a 642 nm fiber laser, see Figure 3.44 on page 84). Moreover, the results confirm that the SWP can be used to with Ti:Sa-equipped STED systems that have proven to yield excellent STED-images with the A633 and A647N dyes (Atto-Tec GmbH. Germany, see Table 3.1). In fact, a considerable number of setups for multi-photon microscopy equipped with a Ti:Sa laser is already present in laboratories and those systems can easily be upgraded to a STED system using the SWP.

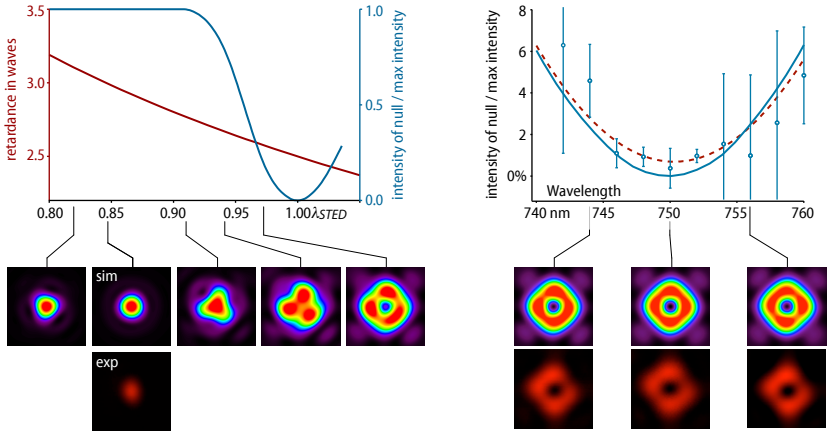

Figure 3.39.: How the depth of the central zero depends on wavelength. Left plot: retardance and simulated depth of the null (in terms of rim-maximum) versus relative wavelength. Right: measured nulls with a SWP designed for $\lambda_{STED} = 750$ nm. The dashed red line is a weighted fit; it confirms that the design wavelength of the SWP has been met to closer than 1 nm. Additionally, simulated (spectral color scale) and experimental PSFs (red color scale) have been graphed for several wavelengths (bottom). The null stays intact for wavelength deviations of ± 7 nm of the design wavelength.

Figure 3.39 also makes clear that the detection PSF inevitably changes with wavelength too, when the SWP is placed in the common beam path behind the objective lens. Being somewhere between the excitation and the STED wavelength (plus potentially above the STED wavelength), emission will not be focused to a diffraction-limited normal PSF on the pinhole. It will instead have a shape corresponding to a mixture of a normal

1 For this estimation, it is neglected that the wavelength-dependent data is actually convolved with the spectrum of the Ti:Sa-laser.

PSF and a donut, depending on the wavelength, and might not fully pass the pinhole. As the STED-PSF already constitutes an optical pinhole by itself, however, the slight resolution enhancement provided by the pinhole is of no interest here and a rather large one can be employed for STED. Its role is solely to block residual stray and out-of-focus light in thick specimen.

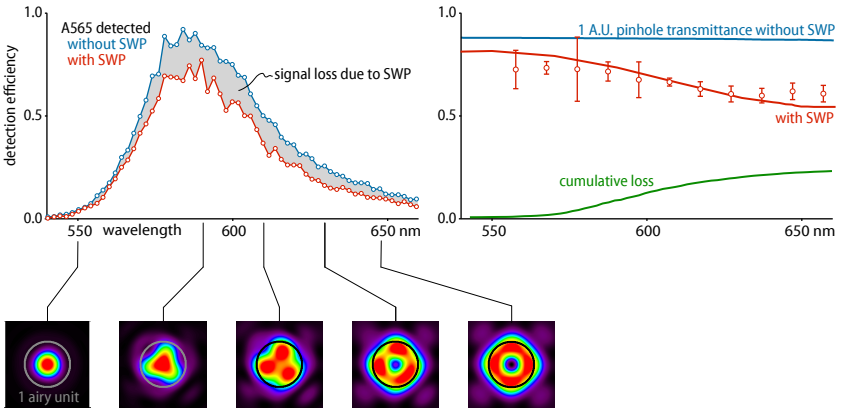

Figure 3.40.: Losses at the pinhole due to the SWP. Left: spectra of an aqueous solution of A565 dye with and without SWP. Right: Simulated and experimental losses amount to 23%. Bottom: detection PSFs at the pinhole (black circle, 1 A.U.).

In standard confocal microscopy, normal working ranges for pinhole size are between 0.5 and 1.5 Airy disks (A.U.) [81]; pinholes used in STED microscopy are typically no less than 1 A.U. Figure 3.40 depicts the losses that can be expected at a 1 A.U. pinhole when introducing a SWP for $\lambda_{STED} = 647$ nm into the beam path that causes the detection wavelengths to partially focus to a donut at the pinhole. It is obvious from the PSF plots below the graphs that even for emission close to the STED wavelength, the donut is only moderately blocked by the pinhole. Hence, even around the STED wavelength, detection efficiency drops only by one third. However, the emission maximum for a typical dye (Atto 565, Atto-Tec GmbH, Germany) to be used with this SWP is around 590 nm, a wavelength where the detection almost fully retains its normal shape, and consequently detection only drops by 7% in theory. The left graph shows actual spectra of an aqueous solution of A565 dye with and without an SWP after the emission has passed a 1 A.U. pinhole (for an 100x/1.4 Oil objective lens). In the right graph, solid lines are simulated losses with (red) and without an SWP (blue) in the detection beam path. The red curve in the right plot includes reflection and absorption losses of 7% typical for a quartz SWP. The circular markers show experimental data calculated using data

from the left graph; the residual differences can be attributed to the antireflection coating of the SWP. It reduces losses in the vicinity of the STED wavelength for which it was optimized, but it slightly increases losses otherwise. Theoretically, for the Atto565 dye, the SWP should cause about 15% of the signal to be blocked by the pinhole, obtained by integration of the difference of the dye spectra with and without SWP. Experimentally, the loss corresponding to the data plotted in Figure 3.40, left, is about 23% (right hand end of the green curve in the right graph), however, this also includes reflection and absorption loss. The decrease in signal is noticeable, however, in all but a few assays, it will not be the limiting factor.

3.4 Improving the quality of continuous wave STED with time-gated detection

This section is slightly set apart from the rest; it demonstrates a method for drawing on the simplicity of cw-STED while the quality of the images is greatly enhanced. While this is not related to using an SWP, it is a simple means to further improve easySTED-setups.

In general, STED with continuous-wave lasers (but pulsed excitation lasers) offers a simpler alternative to using pulsed STED-lasers because it avoids the need for temporal alignment of the excitation and STED beam. Nevertheless, cw-STED comes with its own set of detriments and the major advantage is its major drawback: the fact that is is always on. Continuous wave emission necessarily leads to much lower electrical fields for the same average power compared to pulsed mode where the energy is concentrated in very short pulses. Therefore, instead of being depleted almost immediately, as would be the case for the high-peak powers of pulsed STED, the probability for a molecule to become de-excited during a time interval dt with the STED beam on is much less for cw-STED. In general, while a STED-pulse of duration τ is applied, we can write the decay of the population N in the first excited state[1] S_1 subject to a fluorescence decay rate k_F and a decay rate k_S owing to STED as

$$\mathrm{d}N(t)\big|_{t\leq\tau} = -(k_F + k_S)N(t)\mathrm{d}t.$$

This leads to a population at time t of $N(t)\big|_{t\leq\tau} = N_0 e^{-(k_F+k_S)t}$. Now, the number of photons N_F collected during dt due to spontaneous fluorescence (and spontaneous fluorescence only!) is simply

$$\begin{aligned}\mathrm{d}N_F(t)\big|_{t\leq\tau} &= k_F N(t)\mathrm{d}t \\ &= k_F N_0 e^{-(k_F+k_S)t}\mathrm{d}t \end{aligned} \tag{3.30}$$

1 It is assumed that at the beginning, the molecule has been excited instantaneously.

and therefore the number of photons detected up to time t can be obtained by integration:

$$\begin{aligned} N_F(t)\big|_{t\leq\tau} &= \int_0^t k_F N_0 e^{-(k_F+k_S)t}\mathrm{d}t \\ &= \frac{N_0}{1+\frac{k_S}{k_F}}\left(1-e^{-(k_F+k_S)t}\right). \end{aligned}$$

It follows that until the end of the STED pulse with duration τ

$$N_F(\tau) = \frac{N_0}{1+\frac{k_S}{k_F}}\left(1-e^{-(k_F+k_S)\tau}\right)$$

photons have been emitted. After the STED pulse the remaining population N decays with the normal fluorescence rate k_F:

$$\begin{aligned} N(t)\big|_{t>\tau} &= N(\tau)e^{-k_F(t-\tau)} \\ &= N_0 e^{-k_S\tau - k_F t}. \end{aligned}$$

Again, the number of emitted fluorescence photons is

$$\mathrm{d}N_F(t)\big|_{t>\tau} = k_F N(t)\mathrm{d}t$$

and integration finally yields the total number of fluorescence photons that have been collected:

$$\begin{aligned} N_F(t)\big|_{t>\tau} &= N_F(\tau) + \int_\tau^t k_F N(t)\mathrm{d}t' \\ &= \frac{N_0}{\frac{k_S}{k_F}+1}\left(1-e^{-(k_S+k_F)\tau}\right) + N_0 e^{-k_S\tau}\left(e^{-k_F\tau}-e^{-k_F t}\right). \end{aligned} \tag{3.31}$$

The detector averages over time, so the quantity of interest — the quantity actually measured — is the signal over one period of excitation, i.e. for integration up to infinity[1]:

$$N_{F,total} = N_F(t\to\infty) = \frac{N_0}{\frac{k_S}{k_F}+1}\left(1-e^{-(k_S+k_F)\tau}\right) + N_0 e^{-(k_S+k_F)\tau}. \tag{3.32}$$

The detected signal consists of the fluorescence events during the STED pulse (the first term) and photons detected after the STED pulse has been

1 Treating one pulse by integrating up to infinity is valid if the time between successive excitation pulses is much longer than the fluorescence decay time (≈3 ns). This is the case for standard systems which frequently have an inter-pulse duration of no less than 12.5 ns, corresponding to a repetition rate of 80MHz.

applied (second term). The quantity $k_S\tau$ is proportional to the number of photons in a STED-pulse, in fact $k_S\tau = \sigma\Phi$, where σ is the stimulated emission cross section and Φ is the total number of STED photons per area. With this, Equation 3.32 becomes:

$$N_{F,total}(\Phi,\tau) = \frac{N_0}{\frac{\sigma\Phi}{\tau k_F}+1}\left(1-e^{-\sigma\Phi-\tau k_F}\right) + N_0 e^{-\sigma\Phi-\tau k_F}. \qquad (3.33)$$

The interesting questions are: how does the residual fluorescence $N_{F,total}$ depend on a), the total number Φ of STED-photons; and b), on how they are spread out over time — squeezed into a short, intense pulse (pulsed STED, small τ), dispersed evenly between two excitation pulses (cw-STED, $\tau \to \infty$), or something in between.

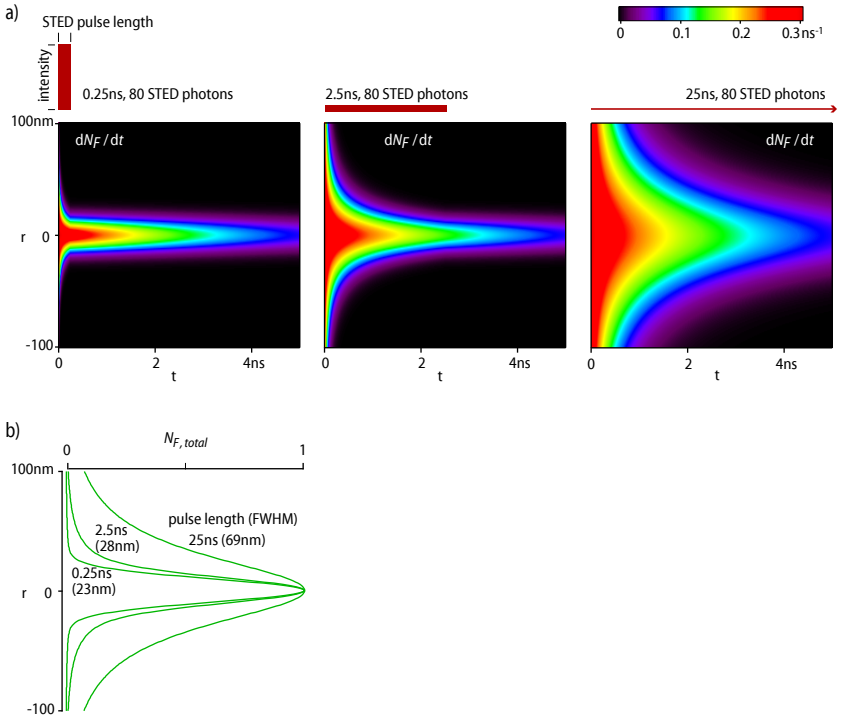

Figure 3.41.: Effect of STED-pulse intensity and duration on the shape of the PSF. Longer STED-pulses or cw-STED add a socket to the effective PSF.

Figure 3.41 a) shows $\mathrm{d}N_F(r,t)\mathrm{d}t$ and $N_{F,total}(r)$ for STED pulse lengths $\tau \in \{0.25\,\mathrm{ns},\ 2.5\,\mathrm{ns},\ 25\,\mathrm{ns}\}$ and for $\sigma\Phi = 80$. The spatial dependencies have been modeled using the relations

$$h_{exc}(r) = e^{4\ln 2\frac{r^2}{d^2}} \tag{3.34}$$

$$h_{STED}(r,t) = \begin{cases} \frac{\sigma\Phi}{\tau} \cdot C\frac{r^2}{d^2}e^{1-\frac{r^2}{d^2}} & t \leq \tau \\ 0 & t > \tau. \end{cases} \tag{3.35}$$

that represent an excitation, respectively STED-PSF with FWHM $d = 200\,\mathrm{nm}$. $C = 1/2{\cdot}e\sqrt{\pi/d}$ is a normalization constant so that $\int_{-\infty}^{\infty} h_{STED}\,\mathrm{d}r = 1$ for $\sigma\Phi/\tau = 1$. It can be seen that for a particular number of STED-photons, the photons being concentrated into a short pulse highly decreases the chance that a molecule can fluoresce outside the STED-null. For $\tau \to 0$ (left hand side) $N_{F,total}(r \neq 0)$ quickly approaches zero, unfortunately, pulses shorter than 100 ps raise other problems[1]. In practice, pulses are typically between 100 ps and 1 ns long. Then, for reasonable STED-intensities $k_S \gg k_F$, i.e. $\sigma\Phi/\tau \gg k_F$, and

$$N_{F,total,pulsed} \approx N_0 e^{-\sigma\Phi} \tag{3.36}$$

from Equation 3.33. The resulting PSF (Figure 3.41 b)) is narrow for STED with short pulses.

For longer STED-pulses, k_F can not be neglected and an increasing number of molecules is allowed to fluoresce spontaneously before being suppressed. The limiting case is cw-STED, where the "pulse" length is much longer than the fluorescence decay time $1/k_F$, corresponding to $\tau \to \infty$ in Equation 3.33:

$$N_{F,total,cw} \approx \frac{N_0}{\frac{\sigma\Phi}{\tau k_F} + 1}. \tag{3.37}$$

Comparing the PSFs (Figure 3.41 b)) for pulsed STED ($\tau = 0.25\,\mathrm{ns}$) with its cw counterpart for the same total number of STED-photons reveals that achieving the same suppression (and therefore the same resolution) takes more cw-STED photons, apparently. It can also be seen that, for cw-STED, a relatively large number of photons are collected from outside the STED-null, again because the majority of cw-STED-photons arrives after the excited molecules have fluoresced anyway. As a result, the cw-PSF features a prominent socket, whereas the pulsed-PSF quickly drops to zero. This is already prominent for the 2.5 ns-PSF which is only marginally broader than the 0.25 ns-PSF but already has a distinctly wider base, the *cw-socket*.

1 They can for example cause two-photon excitation by the STED-beam. Pulses too short quickly get stretched by group velocity dispersion, anyway.

For sparsely distributed molecules, the cw-socket can effectively be removed by deconvolution, but in dense samples the sockets from many emitters quickly add up to a significant level of background. In the worst case, the intensity gap between two molecules that can barely be resolved without socket can get filled up by this background and resolution deteriorates[1].

Another way to remove the cw-socket, while keeping the simplicity and low cost of cw-STED, is time-gated detection. Looking at Figure 3.41 again, it can be seen that the majority of the fluorescence photons that form the cw-socket are collected during the very beginning of an excitation pulse cycle. If the photons collected during the first few hundred picoseconds were to be dropped, the cw-socket could be greatly subdued; albeit at the expense of wanted signal because also the first photons out of the donut-null would also be discarded. In mathematical terms, this means integrating Equation 3.31 with cw-parameters, but starting from a later time t_g :

$$\begin{aligned} N_{F,cw,gated}(t) &= \int_{t_g}^{t} k_F N_0 e^{-(k_S+k_F)t} \mathrm{d}t' \\ &= \frac{N_0}{\frac{k_S}{k_F}+1} \left(e^{-(k_S+k_F)t_g} - e^{-(k_S+k_F)t} \right). \end{aligned} \tag{3.38}$$

Consequently, the number of photons collected during a pulse cycle much longer than $1/k_F$ is

$$N_{F,gated} = \frac{N_0}{\frac{k_S}{k_F}+1} e^{-(k_S+k_F)t_g}.$$

As expected, for $k_S = 0$ — in the center of the donut — the fluorescence drop-off due to gating is the standard k_F. In the outer portions of the STED-PSF, detected fluorescence is suppressed both by the STED-beam and time gating. Because inhibition occurs faster here, due to the concerted action of $k_S + k_F$, more photons are removed by starting detection later at t_g.

From another point of view, applying the STED-beam shortens the effective life time of the excited state anywhere but at the STED-null. It does so rather inefficiently for cw-STED, but the difference is still enough so that the spatial origin of a photon is linked to its lifetime. Therefore, photons from the outer part of the donut, the ones that form the cw-socket, can be gated out. Figure 3.42 shows a comparison of cw-STED PSFs ($\sigma\Phi = 80$) with time gating applied for $t_g \in \{1\,\mathrm{ns}, 2\,\mathrm{ns}, 3\,\mathrm{ns}\}$ as

1 As an interesting side note, it becomes clear that FWHM can not be the ultimate measure for resolution, but only provide a ballpark figure. In order to fully assess a system one has to consider, among more parameters, the actual shape of the PSF.

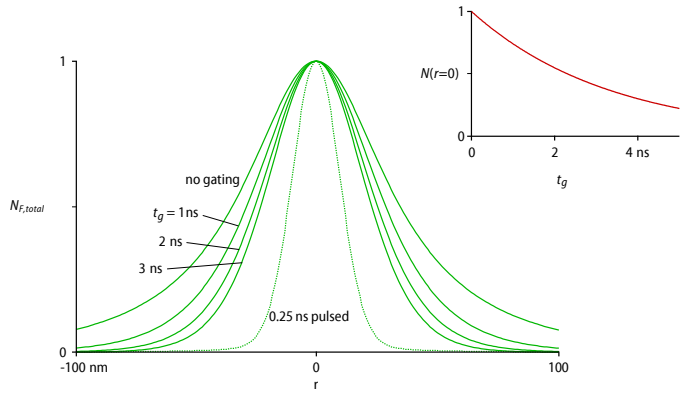

Figure 3.42.: Effect of time-gating on a cw-STED PSF. The socket is alleviated, although signal is lost. (The inset shows loss of wanted signal over gating time.)

well as the PSF without gating and the PSF for pulsed STED ($\tau = 0.25$ ns; dotted line) The gated PSFs are normalized to one to ensure comparison of the FWHM and the socket. The inset shows the peak signal (at $r = 0$) for different gating times, it is of course simply the fluorescence decay curve ($k_F = 1/3\,\mathrm{ns}^{-1}$.)

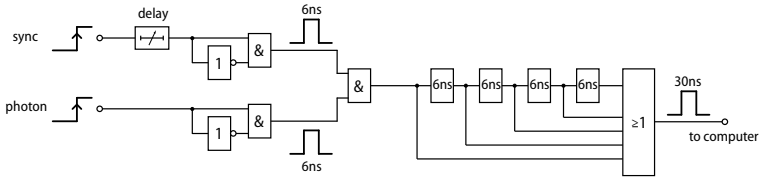

Figure 3.43.: Time gating circuit.

This time gating method was put to work in most of the cw-STED experiments described throughout this thesis, basically all experiments that involve a 592 nm or a 642 nm fiber laser. To this end, a complex programmable logic device (CPLD, Coolrunner II XC2C256 6c, Xilinx, USA) was programmed with a gating circuit. Using a CPLD ensures speedy execution on the nanosecond timescale and provided additional flexibility. A equivalent circuit diagram is shown in Figure 3.43. The circuit is provided with the photon pulses from the detector and a sync pulse from the excitation laser. The sync pulse is fed through a custom-built delay box (the same that is also used to adjust the delay between excitation and STED in pulsed mode). The delay circuit allows one to choose the gating time t_g. The rising flanks of both, sync and photon pulses are converted into a stan-

dard, approx. 6 ns long pulse[1] by the first `not-and` combination before an `and`-gate picks a 5 ns long portion out of the photon stream according to the sync pulse. In other words, if the standardized pulses of the gating and sync signals overlap by more than 5 ns the signal is passed through. The overlap time was estimated from the reduction of dark counts when the gating curcuit was switched on. The photon pulses are stretched to $\approx$ 30 ns by the following four delay gates and the `or`-gate, so that they are recognized by the detection logic. Of course, this circuit can also be hard-wired.

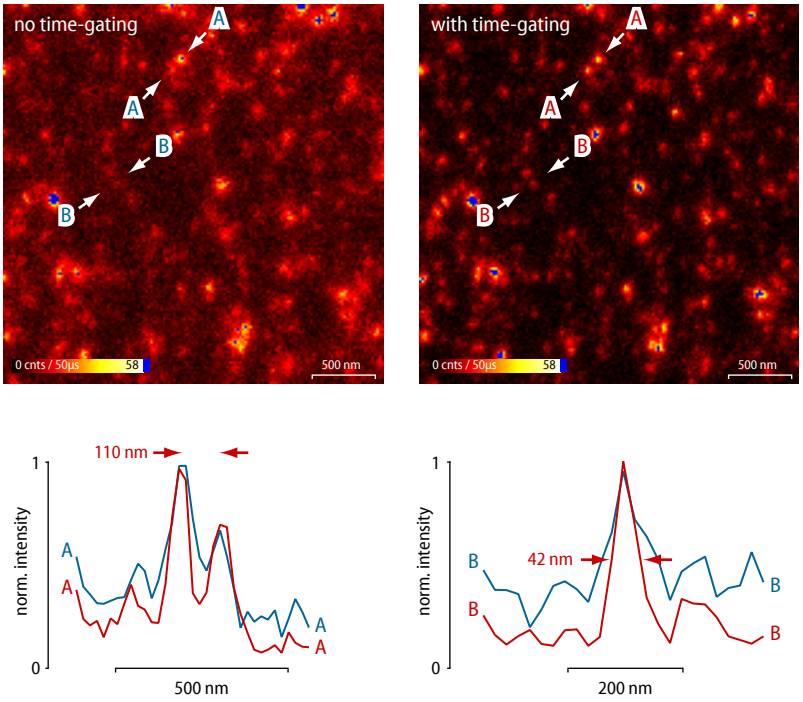

Figure 3.44.: Time-gated STED of nile red beads. The cw-socket is alleviated with time-gating $t_g \approx 2$ ns, improving resolution while background is reduced.

Figure 3.44 shows 20 nm nile red fluorescent beads, recorded with (right) and without (left) time-gating. The STED-laser was a 642 nm fiber laser (VFL-P-1000-642, MPBC, Canada), excitation was performed with a SC-450-4 supercontinuum source (Fianium, UK) of which a 540 nm laser line was selected with an AOTF (Crystal Technology, Inc., USA). t_g was estimated to $\approx$2 ns from the signal reduction in a dye solution. This figure is a compromise between resolution and signal and has to be optimized for

1 The 6 ns result from the propagation delay of the gates in the CPLD.

the particular sample. The difference time-gating[1] makes is evident from the images, as well as from the line profiles below. The cw-socket is obvious, especially from the line profile on the right. The line profile on the left demonstrates the enhanced contrast accompanying time-gating; the two beads shown can hardly be resolved without gating (although the FWHM suggests otherwise) because the socket fills up the space between them in the non-gated image. An additional effect of time-gating is that it also rejects any light that comes *after* the gate. This can only be stray- or residual STED-light or, especially for 642 nm-STED on nile red, excitation by the STED-beam when the STED-wavelength happens to be in a range where absorption has not yet fully declined. In any case, it can be seen from Figure 3.44 that background and the cw-socket is heavily suppressed by time-gating, resulting in crispier pictures with more details, while retaining the simplicity of cw-STED.

1 The time-gated image was taken first; signal loss induced by removing early photons is almost exactly balanced by bleaching. This is why the images have about the same brightness.

4

Conclusions and Outlook

This work dramatically simplifies the optical and mechanical engineering challenges of STED- and RESOLFT setups. As a result of this, assembling a nanoscope now becomes easier by far, and this in turn will help many more users inexperienced with optics and optomechanics to a high-resolution STED or RESOLFT setup.

STED, the first method to fundamentally break the diffraction limit, has helped to clarify a lot of questions in biology and medicine, owing to its ability to resolve details beyond the diffraction limit of traditional microscopes while allowing live-cell application. Unfortunately, the average microscope-user — who is not an engineer by trade — can find it rather tedious to assemble and maintain a home-built STED-setup. While from an engineering standpoint, building a STED system is in principle nothing more than adding a second color to an existing laser scanning system, the perceived difficulty of aligning beams and generating a deep null often have a deterring effect to the inexpert. Often, users seeking a nanoscope resort to methods using stochastic read-out, e.g. PALM, STORM, GSDIM and variants, because of the allegedly fuss-free technical layout involving basically only uniform illumination, a camera and some software. Indeed, also the stochastic techniques have proven their worth and have come to show their advantages (simplicity, lower light levels, large field of view). Nevertheless, STED is faster for small fields of view, the target-resolution can be set deliberately by the user (allowing even faster scanning when the resolution is set rather low), and — perhaps most importantly — it is easily brought together with various quite different methods. Take for example fluorescence correlation or fluorescence fluctuation spectroscopy [23, 55, 86], for which focal volumes can be reduced using STED, take the manipulation of individual spins with nanoscale resolution using a donut beam [70], take multi-photon STED [18, 75], or take the writing of structures at the nanoscale [46]. Moreover, the supposed technological simplicity of stochastic read-out is in fact only the result of adjourning the localization process from a physical entity (the

fluorophore in the sample) to a later time and in silico. This requires careful attention of the end-user, as a piece of code can in principle generate a nice-looking picture even from inadequate raw camera images ("garbage in, garbage out"). A STED/RESOLFT machine will perhaps require more effort until the PSFs are aligned and the null is suitably deep (and up to that point it will not put out anything) but as soon as a STED image can be recorded, this data can be highly trusted to exactly resemble the specimen.

This was the situation up to now. The work presented herein annihilates most arguments concerning the alleged complexity of STED/RESOLFT systems and will thus allow any user to make full use of the advantages that STED offers. The heart of this advancement is a new beam shaping device, called segmented wave plate (SWP) because it is made of several segments of a birefringent crystal. Owing to the chromatic properties of the crystal, the SWP can distinguish the beams on the basis of their wavelength and can thus treat each of them differently; in particular, it is designed to alter the beam with the STED wavelength so that it focuses to a suitable suppression PSF; at the same time, the other beams are hardly affected since the crystal is not birefringent at their respective wavelengths. In particular, because both the excitation and STED beams can pass the new device simultaneously, it can be used with sources that deliver both beams simultaneously. Using such sources (e.g. supercontinuum generators or two lasers coupled through the same fiber) means that alignment is provided inherently — the beams come perfectly coincident right from the start. In turn, inherent alignment makes only sense with the beam shaping device demonstrated in this thesis, as only a beam shaping device that is able to treat STED and excitation differently but at one go is able to process pre-aligned beams. This nullifies a widespread argument running against STED: the requirement to superimpose two PSFs.

Moreover, a deep null of the STED-PSF is virtually guaranteed, because the SWP tolerates misalignment on the optical axis; it tolerates deviations of the STED-beam from the design wavelength; and it tolerates imperfect input polarization. This nullifies another argument speaking against STED, the seemingly complex generation of a deep null. As a consequence, existing laser scanning microscopes can be upgraded to a full-blown STED system (easySTED) with improved resolution along two and even three dimensions, simply by placing the SWP behind the objective lens and supplying a shared laser source. Of course, setups built from scratch can also be constructed much simpler than before. In any case, the simplicity, low cost and robustness of easySTED will highly increase the applicability of STED (and RESOLFT, obviously) for the average user.

Furthermore, an easySTED-system can be converted to a microscope that directly images the orientation of molecular dipoles (MOM-STED);

simply by adding a half wave plate. Potential applications are legion: rigidly attached fluorescent probes could help to observe conformational dynamics in proteins. For example, the translational dynamics of myosin have been investigated with single molecule imaging [116] — but changes in orientation as this motor protein walks along its track might be equally interesting. Similarly, the gating of ion channels upon binding of a ligand seems to be accompanied with a rotation of channel sub-units between the open and closed states [99]. Yet, so far, electron images had to serve as proof, while MOM-STED combined with fast scanning [105] will help to access dynamics. Moreover, when a fluorescent marker is only loosely tethered to a macromolecule, it can be used to monitor changes in rotational freedom [102].

Additionally, this work has demonstrated the enhancement of continuous wave-STED images with time-gated detection. It was shown that by discarding early fluorescent photons using a simple electronic circuit the quality of cw-STED is improved considerably. Because with a continuous suppression beam the necessity to temporally align STED and excitation ceases to exist — in contrast to using pulsed STED beams — its use further simplifies nanoscopy based on targeted switching.

In the future, it remains to be investigated how several excitation/STED combinations can be served by a single SWP (multicolor easySTED). Possible strategies are excitation multiplexing and the use of chromatically altered wave plates. The former requires the selection of two dyes that have different absorption spectra but can be suppressed by the same STED-wavelength. Specifically, if the absorption characteristics are such that at least one dye is excited at a wavelength for which the absorption of the other dye is close to zero, recording of two subsequent STED-images with two suitable excitation lines will yield two images where one contains contributions from the first dye only, whereas the other one contains signal from both dyes. Subsequent linear unmixing will then restore the original distributions. This method has the advantage that the position of the null is the same for both dyes and the images are inherently aligned.

Another approach is to engineer the chromaticity of the wave plates in a way that they are half wave plates and full wave plates at more than one suitable wavelength, for example by combining two wave plates of different material with the optic axes crossed [16]. At present, the 750 nm-SWP becomes a higher order half wave plate again at 549 nm, but a STED laser and a suitable dye can only fit this wavelength by chance. However, deliberate engineering of the retardance will permit simultaneous imaging of two established dyes with two separate excitation/STED pairs. This way, dyes can be chosen that are excited and suppressed at their optimal wavelengths and that thus can be expected to perform superior compared to

dyes selected for excitation multiplexing. The downside is that, since two different nulls are scanned, the images have to be aligned by other methods.

Additionally, it will have to be examined to what extent birefringent crystals can be replaced, preferably by a system allowing for wavelength tunability and/or conversion between 2D-easySTED, 3D-easySTED and MOM-STED. Possible strategies might be the use of (reconfigurable) liquid crystals [118], electro-optical elements [78], photonic crystals [89] or optically active arrangements [82] as polarization rotators, or a method following along the idea of the Babinet-Soleil compensator that allows for tunability of the retardation.

A

Sample preparation

Gold beads and fluorescent beads

Cover slips where cleaned with absolute ethanol and air-dried, followed by immersion in poly-l-lysine (Sigma Aldrich, USA) or in a 0.1% aqueous solution of polyethylenimine (PEI, Sigma Aldrich) for 10 minutes, following careful rinsing with ddH_2O. Like poly-l-lysine, PEI is a cationic surfactant and is absorbed to glass surfaces[1] forming a positively charged layer. Then, either an aqueous solution of gold beads ($\approx 0.2 - -0.3\,ml^{-1}$) was allowed to passively absorb to the coated cover slip for 10 minutes, again followed by rinsing with ddH_2O. Alternatively, fluorescent beads (FluoSpheres, Invitrogen, USA) diluted $1 : 10\,000$ in ddH_2O where applied in the same manner. In either case, the cover slip was spin-coated with a 0.5% solution of polyvinyl alcohol (PVA) for 1 minute at 60 rpm. The sample was sealed using Feuerzangenbowle nail polish and a second cover slip.

APP

Embryos (E14) from wt (C57BL/6NCrl; Charles River) mice were separated and dissociated mixed cortical neuron cultures were prepared as described previously [19]. Neurons were grown on poly-l-lysine-coated coverslips (Marienfeld, Germany) in serum-free Neurobasal Media (Gibco, Germany) with B-27 supplement (Gibco, Germany), 25 µM glutamate (Sigma Aldrich, Germany), and 0.5 mM glutamine (Sigma, Germany). For immunocytochemical analysis, primary neurons were fixed with 4% PFA (Sigma, Germany) and permeabilized for 10 minutes with 0.1% NP-40/-Nonidet (Fluka, Germany) in PBS. Cells were incubated with primary antibody (monoclonal anti-APP-antibody 4G8 (Chemicon, CA) diluted in 5% goat serum in PBS) at 4° overnight, washed with PBS, incubated with

1 In water, glass surfaces bear a strong negative charge [52]

secondary antibody (Atto565 goat anti-mouse IgG, 10 µg/ml), and embedded in Mowiol (Sigma, Germany) on glass coverslips.

Bassoon

Cultured rat hippocampal neurons were washed with PBS and fixed with 2%-paraformaldehyd (PFA, Sigma Aldrich, Germany) in PBS at room temperature for 10 minutes. After a second washing step with PBS and blocking in 0.5%-bovine serum albumin (BSA, Sigma Aldrich) for 20 min, the samples were incubated with 50 μl primary anti-Basson Sap7f antibodies (Abcam, USA) in 0.5% BSA for one hour. After washing with PBS, the secondary antibody conjugated to Atto565 (Atto-Tec, Germany) was added in the same fashion. Samples were embedded in Mowiol (Sigma Aldrich)

B

Mathematical considerations on suitable birefringent beam shaping devices

In order to find an suitable configuration for a polarization manipulating beam-shaping device, a rigorous mathematical description is helpful in order to find the particular configurations that guarantee a central dark spot.

Recall that the field around the geometric focus can be obtained by computing the fourier transform of the pupil field A:

$$\mathbf{E}(\mathbf{r}) = \sqrt{2\pi}\,\imath k\ \mathcal{F}\left[\sqrt{\frac{k_0}{k_z}}\ \left(\mathbf{P}^h(\mathbf{k})A^h(\mathbf{k}) + \mathbf{P}^v(\mathbf{k})A^v(\mathbf{k})\right) e^{-\imath k_z z}\right]. \quad \text{(B.1)}$$

Of course, $A^{h,v}(\mathbf{k}) = A^{h,v}\left(\rho(\mathbf{k}), \phi(\mathbf{k})\right) \in \mathbb{C}$. Here, $\mathbf{P}$ has been splitted into its respective components for incident horizontal and vertical polarization, according to Equation 3.7, page 26. The trick with polarization beam shaping devices is simply that the pupil function A is also split up, in other words, that two incoming orthogonal polarizations get different phases (and generally amplitudes) across the back aperture of the lens. This is in contrast to phase masks, where both polarizations see the same phase plate. Treating the two polarization directions differently means that downstream of the beam shaping device at least parts of the beam are potentially elliptically polarized in a way. Additionally, a *segmented* beam shaping device means that different parts of the beam profile have different polarization states. One part can for example be linearly polarized, another part right circularly and yet another part elliptically.

Next, consider a general segmentation[1] of the pupil plane. The pupil function $A(\rho, \phi)$ is expressed as the sum of N arbitrary tiles: $A(\rho, \phi) =$

1 The segmentation does not necessarily have to cover the full pupil plane — some tiles can have zero amplitude.

$\sum_{i=1}^{N} A_i^h(\rho,\phi) + A_i^v(\rho,\phi) = \sum_{i=1}^{2N} A_i$. For even i, A_{i-1} describes magnitude and phase of the horizontally polarized pupil field in tile $i/2$ and A_i describes the vertically polarized pupil field in the same tile. Of course, $A_i(\rho,\phi) = 0$ for all tiles $j \neq i$, i.e. for coordinates (ρ,ϕ) that do not belong to tile i.

Looking carefully at Equation B.1 with the pupil function written this way reveals that each A_i can (mostly) be taken out of the fourier integral, as both magnitude and phase of A_i are constant over each tile. Thus, Equation B.1 can be rephrased as a sum of integrals

$$\mathbf{E}(\mathbf{r}) = \sum_{i=1}^{2N} A_i(\rho,\phi)\sqrt{2\pi}\,\imath k \;\; \mathcal{F}\left[\sqrt{\frac{k_0}{k_z}}\;\mathbf{P}^{h,v}(\mathbf{k})A_i^{lim}(\mathbf{k})e^{-\imath k_z z}\right], \qquad \text{(B.2)}$$

where $\mathbf{P}^h$ or $\mathbf{P}^v$ is used according to the current index i being odd or even, respectively. What remains of A_i inside the integral is just A_i^{lim}. It is 1 inside the tile and 0 otherwise, effectively limiting the integration limits to the tile. With this, the field can be conveniently expressed as

$$\mathbf{E}(\mathbf{r}) = \sum_{i=1}^{2N} A_i(\rho,\phi)\;\mathbf{K}_i(\mathbf{r}) = \mathbf{K}(\mathbf{r})\;\mathbf{A}. \qquad \text{(B.3)}$$

Thus, the field near the focus can be expressed as the sum of the fields from the individual tiles, $\mathbf{K}_i(\mathbf{r})$ arising at $\mathbf{r}$, weighted with the phase and magnitude $A_i(\rho,\phi)$ of the light sent into each tile. This is a consequence of the linear propertiers of the fourier transform.

For clarity, $\mathbf{E}(\mathbf{r}) = \left[E_x(\mathbf{r}), E_y(\mathbf{r}), E_z(\mathbf{r})\right]^T$ is a 3×1-vector of the field at $\mathbf{r}$. $\mathbf{A} \in \mathbb{C}^{2N}$ holds magnitude and phase of the horizontal and vertical pupil fields in each of the N tiles like

$$\mathbf{A} = \left[A_1,\, A_2,\, \ldots,\, A_{2N-1},\, A_{2N}\right]^T.$$

An A with odd subscripts denotes h-polarized pupil fields, and an A with even subscripts denotes v-polarized pupil fields. Essentially, each tuple $(A_n, A_{n+1}), n \in \{1,3,5,\ldots N\}$ is a Jones vector describing the polarization in tile n. Finally, the $3 \times 2N$-matrix $\mathbf{K}(\mathbf{r})$ contains the contributions of a unit h- and v-field in each tile to the focal field at $\mathbf{r}$:

$$\mathbf{K}(\mathbf{r}) = \begin{bmatrix} \overbrace{E_{x,1}}^{\text{h-pol}} & \overbrace{E_{x,2}}^{\text{v-pol}} & \dots & \overbrace{E_{x,2N-1}}^{\text{h-pol}} & \overbrace{E_{x,2N}}^{\text{v-pol}} \\ E_{y,1} & E_{y,2} & \dots & E_{y,2N-1} & E_{y,2N} \\ E_{z,1} & E_{z,2} & \dots & E_{z,2N-1} & E_{z,2N} \end{bmatrix} \begin{matrix} \\ \\ \\ \\ \\ \end{matrix} \begin{matrix} \\ \\ \underbrace{\qquad\qquad}_{\text{tile 1}} \\ \\ \underbrace{\qquad\qquad}_{\text{tile } N} \end{matrix} .$$

It consists of column vectors describing the field generated in the geometric focus by the fields present in the respective tile in the pupil plane. For example $E_{x,2} = E_{x,2}(\mathbf{r})$ is the x-component of the electric field generated at $\mathbf{r}$ by a vertically polarized beam in tile 1. Furthermore, the matrix $\mathbf{K}$ also holds the geometric layout of the segmented wave plate via the A_i^{lim} in Equation B.2. Interestingly, $\mathbf{K}(0)$ is real, because the fields at $\mathbf{r} = 0$ all have the same phase (this is simply a result of focusing with a lens).

Rewriting Equation B.1 as a matrix equation has the appearance of a purely academical exercise. However, it in fact greatly reduces computational time when only a few points $\mathbf{r}$ have to be evaluated, because $\mathbf{K}$ — and thus several FFTs — can be precalculated. This will be used to great effect in Appendix C and also in the remainder of this section, which deals with finding configurations that produce a central null.

In order to investigate the configurations that yield a dark spot in the center, take a closer look at the field $\mathbf{E}(\mathbf{r} = 0)$. Let $\mathbf{K}_i^0$ be the contribution of the integral in Equation B.3 evaluated at the geometric focus, i.e. at $\mathbf{r} = 0$. Then, $\mathbf{E}(\mathbf{r} = 0) = \mathbf{E}^0 = \mathbf{K}^0\ \mathbf{A}$. From now on, all quantities evaluated at $\mathbf{r} = 0$ will simply bear a 0 superscript.

Having the segmented wave plate in a matrix $\mathbf{K}^0$ makes it straightforward to find all input fields $\mathbf{A}^{null}$ that fulfil $\mathbf{K}^0\mathbf{A}^{null} = [0,\ 0,\ 0]^T$. In other words, given a certain geometric layout $\mathbf{K}^0$, all $\mathbf{A}^{null}$ would describe the fields one has to send through the so-segmented pupil plane in order to obtain an intensity zero at the geometric focus. The set of all possible $\mathbf{A}^{null}$ is called *kernel* of $\mathbf{K}^0$ or sometimes, probably more adequate for the problem at hand, the *null space* $\mathcal{N}$ of $\mathbf{K}^0$:

$$\mathcal{N}\left(\mathbf{K}^0\right) = \left\{\mathbf{A}^{null} \in \mathbb{C}^{2N} | \mathbf{K}^0\mathbf{A}^{null} = 0\right\}.$$

The null space can be found for example via the reduced row echelon form of $\mathbf{K}^0$ or by performing a *singular value decomposition (SVD)* using an algebra system such as Matlab. The SVD [84, 110] of $\mathbf{K}^0$ is a factorization that yields three matrices $\mathbf{K}^0 = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^*$. The column vectors $\mathbf{v}_i$ of $\mathbf{V}$ for which the corresponding column vectors $\boldsymbol{\sigma}_i$ of the diagonal matrix $\mathbf{\Sigma}$ are $\boldsymbol{\sigma}_i = 0$ form an orthonormal basis $\left\{\mathbf{s}_j\right\}$ of the nullspace of $\mathbf{K}^0$. Thus, the solution is to take any $\mathbf{A}$ from the set span $\mathbf{s}_j$ and it will be a legitimate $\mathbf{A}^{null}$ in a sense that it will yield zero intensity in the geometric focus.

Naturally, the first question arising is whether a certain geometric layout $\mathbf{K}$ can form an intensity zero in the focus at all. This boils down to

finding the dimension dim $\mathcal{N}(\mathbf{K}^0)$ of the associated null space, i.e. the number of linear independent $\mathbf{s}_j$. This number is also called the *nullity* of $\mathbf{K}^0$. Generally, the *rank-nullity theorem* tells us that for any $m \times n$-matrix $\mathbf{M}$ the rank of $\mathbf{M}$ and the nullity of $\mathbf{M}$ add up to n. For this problem this means that

$$\dim \mathcal{N}\left(\mathbf{K}^0\right) = 2N - \operatorname{rank} \mathbf{K}^0.$$

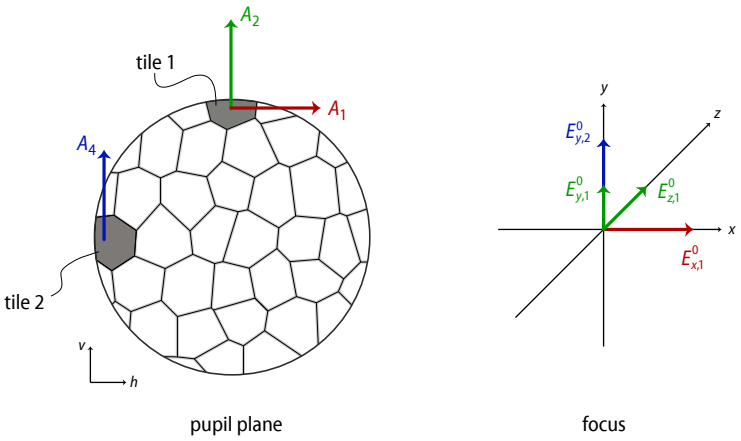

Figure B.1.: Number of linear independent fields for a general segmented wave plate.

This leaves to find rank $\mathbf{K}^0$. Another notion of rank $\mathbf{K}^0$ is the number of linear independent column vectors of $\mathbf{K}^0$. How many of those are there in $\mathbf{K}^0$? Consider a generic segmented wave plate as in Figure B.1. Looking at the focal fields generated by the three specifically indicated pupil fields and comparing with the deflection vectors given in Figure 3.3 and Equations 3.4 – 3.6 (page 26), it can be seen that the h-polarized pupil field in tile 1 gives rise to a pure x-field in the focus, the v-polarized pupil field in tile 1 generates a y- as well as a z-field and the v-polarized pupil field in tile 2 is responsible for a pure focal y-field. Hence, so far $\mathbf{K}^0$ for this beam shaping device looks like this:

$$\mathbf{K}^0 = \begin{bmatrix} E^0_{x,1} & 0 & \dots & 0 & \dots \\ 0 & E^0_{y,2} & \dots & E^0_{y,4} & \dots \\ 0 & E^0_{z,2} & \dots & 0 & \dots \end{bmatrix}. \tag{B.4}$$

By adjusting the relative strengths of the corresponding $A_{1,2,4}$, any focal field can be generated, hence, the focal field generated by any other tile can be expressed using only A_1, A_2 and A_4. This in turn means that the remaining column vectors of $\mathbf{K}$ missing in Equation B.4 have to be linear

dependent on the given three and rank $\mathbf{K}^0 = 3$ for this case. In general, at least three A in at least two tiles are needed to express the focal field of any other tile, because the general field in the focus has to be described by no less than three components. Of course, depending on the actual layout of the segmentation three might not be enough. In any case, these considerations give a lower bound for rank $\mathbf{K}^0$ and consequently

$$\dim \mathcal{N}\left(\mathbf{K}^0\right) \leq 2N - 3. \tag{B.5}$$

Again, dim $\mathcal{N}$ is the number of linearly independent configurations for a segmented wave plate with N tiles. Obviously, a beam shaping device consisting of one tile can not produce a central intensity zero (except for the trivial case of $\mathbf{A} = 0$). A wave plate with two tiles has only one configuration per geometric layout that produces a central intensity zero, for example the ones given in Equation 3.9 and Equation 3.11. These are two well-known configurations for two tiles, but both are not perfect for STED in general.

For high resolution in all directions, the polarization distribution in the high-intensity regions has to be such that it suppresses all dipole directions efficiently. Therefore, it is not enough to find segmented wave plate configurations with a central null; however, knowing the subspace to scan and its dimension will come in handy for finding a layout that not only features a dark hole at the geometric focus, but is also completely surrounded by light. An approach how to find such a wave plate will be presented in the next section.

C

Finding the optimal SWP-configuration by means of a genetic algorithm

This section will briefly discuss the genetic algorithm used in this work in order to optimize segmented waveplates for both 2D and 3D resolution enhancement. In general, a genetic algorithm is an implementation of an evolutionary heuristic — a strategy for finding a global optimum or at least an approximate solution in a determined and purposive way. The fundamental idea is to implement strategies from evolution in biology; especially the improvement of inherited favorable traits through successive generations. As such, the main features of a genetic algorithm are *variation*, i.e. introducing changes in the genome of the population, and *selection* of the most successful *individuals* that have been bred this way. Ideally, these solutions will point in the right direction and repeated variation of the genome and combination of successful traits will indeed *evolve* the population towards the global optimum when better individuals are given better chances to reproduce. This has been demonstrated to great effect in nature itself where highly adaptive organisms have evolved even in difficult and competitive environments. In technology, evolutionary algorithms have been used to solve a wide variety of problems ranging from engineering over finance to computer sciences themselves [54] as soon as the first (analog) computers became available. For a detailed history and explanation see [74].

Clearly, for the problem of optimizing SWPs, we face a large number of parameters and — possibly multiple — narrow fitness optima in an otherwise flat search space (most configurations will not produce a central zero). Even a discrete segmented wave plate consisting of four segments each with electric fields allowed to be polarized along four different directions, either linearly or circularly, and two global phases for each segment has $4^8 = 65,536$ different configurations and an annular SWP with eight

segments is virtually impossible to optimize using brute force. A more advanced optimization method is fore example hill-climbing. Here, the algorithm uses the gradient of the fitness landscape in order to search for a closest maximum, but it is clearly apparent that this technique fails for the fitness landscape of SWPs where the gradient is mostly zero. Furthermore, even more advanced hill-climbing algorithms (e.g. simulated annealing) will not get an exhaustive picture of the whole search space. In contrast, genetic algorithms excel in vast and uncharted search spaces because they are implicitly parallel and insensitive to discontinuities or plateaus in the fitness landscape. What is more, for the SWP-problem, there are obviously favorable traits, e.g. a preliminary solution with two opposing fields in opposing segments will be close to having a central null. It makes sense to combine those favorable traits with another solution that has opposing fields in another pair of segments in order to quickly arrive at the goal. However, the two preliminary solutions can be far separated in the search space and that might hinder some algorithms. Contrary to other methods, a genetic algorithm can efficiently combine the "knowledge" of separate solutions through breeding and is not susceptible to distance in search space.

In the following, technical details of the genetic algorithm used to find and optimize segmented waveplates will be described. In order to increase the speed of the algorithm and, more importantly, to facilitate practical realization, the configuration space has been discretized; fields are either linearly or circularly polarized (when linearly, it can only be one of the set $\{\nearrow, \nwarrow, \swarrow, \searrow\}$) and phase shifts between segments are discontinuous. This ensures that the fast axes of the resulting SWP are perpendicular and that the segments are either half or quarter wave plates; hence, it ensures that the product can be easily fabricated. The parameters considered for each segment n are: the relative phases $\varphi_{r,n} \in \{0, \pi/2, \pi, -\pi/2\}$ of the horizontally and vertically polarized field components, and the global phase of each segment, again from the set $\varphi_{g,n} \in \{0, \pi/2, \pi, -\pi/2\}$. From a technical point of view, each of these features is described by an integer so that a segment is fully constituted by the 2-tuple $(\varphi_{r,n}, \varphi_{g,n})$ on the domain $\{1, 2, 3, 4\}$, in the following called a *codon*. This notation slightly differs from the one used in Appendix B, but is more convenient for checking how a segment is polarized (left/right circular or linear). A certain waveplate configuration is then encoded for by a list of codons, forming a *genome*. Thus, a genome consists of `genomeSize` "basepairs[1]" partitioned into codons of size `codonSize`. A whole genome describes the features of an individual, in our case the configuration of the waveplate

1 Since translation errors are not to be worried about here, a single "base" is actually sufficient. However, in order to stick with genetic language, the word basepair will be used.

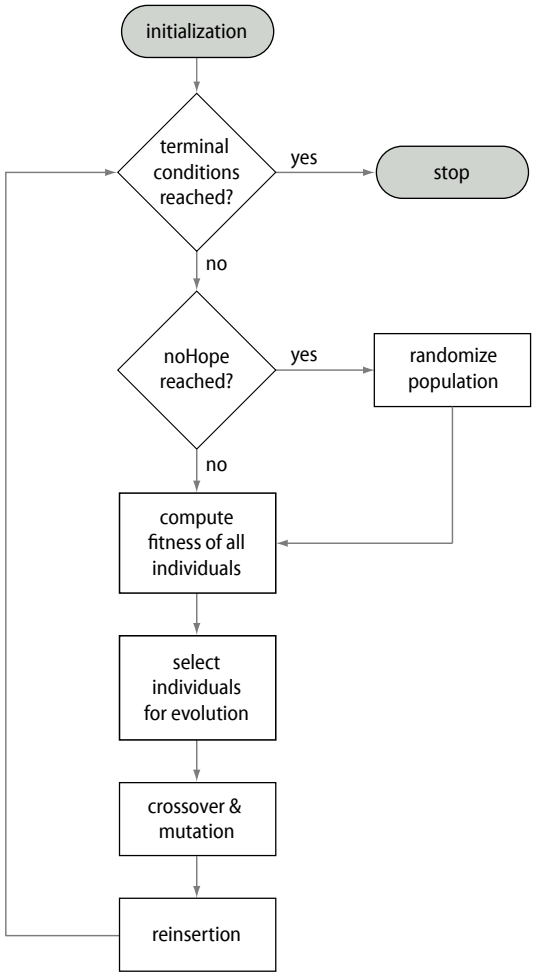

Figure C.1.: Flowchart for the genetic algorithm used in this work. The individual blocks are elaborated on in the text.

being associated with that specific individual. The population itself consists of `populationSize` individuals held in a (`populationSize` $\times$ `genomeSize`)-matrix named `population`.

With this, the first step of the algorithm is to initialize variables (see. Figure C.1 for a flow chart). After initialization, the algorithm enters the optimization loop, which is executed until the terminal conditions come to pass. Here, the terminal condition is simply a maximum number of generations allowed. In the next step, the algorithm checks the number of generations since the last fitness improvement and, when a certain number has passed, assumes it is stuck in a local or maybe the global maximum ("noHope"-scenario). In any case, the population is completely randomized, however, the fitness and configuration of the best individual so far is kept in memory. With this restart, a certain drawback of genetic algorithms is eliminated, namely that they can evolve towards a local optimum that might not necessarily be the global optimum.

The next step is to compute the fitness of the population, i.e. how well the SWPs described by each individual will work. To this end, the PSF does not have to be fully computed for each individual, instead, Equation B.3 can be used to evaluate the PSF at certain descriptive locations only. Using a limited number of points is sufficient if the distance between them is approximately diffraction-limited, because in this case, another point located somewhere between the first two will not have a much different intensity, anyway. Subsequently, in order to account for different dipole orientations, each set of points describing the PSF is subjected to a dipole along x, one along y, and another one oriented diagonally. For a 2D-SWP, eight points are distributed along where the rim of the donut is expected to form, i.e. at diffraction-limited distance from the focus. A ninth point is at the geometric focus itself, where ideally the zero should occur. For a 3D-SWP, another ten points are considered, distributed on a sphere around the geometric focus. Now, the intensity I_k at these k points[1] has to be mapped onto a single fitness value F. Clearly, the foremost concern has to be the quality of the zero, measured by the ratio of the intensity $I_{k=0}$ in the geometric focus to the maximum intensity $\max(I_k),\ k \neq 0$ found in the other points. Ideally, of course, this ratio equals zero. Obviously, the fitness value should be high in this case, but low for central intensities above zero.

As it has been described in subsection 3.3.3, the dipolar nature of the fluorophore being imaged can, for certain PSFs, yield a line-shaped zero instead of an isotropic one. Moreover, even when this is not the case, the

1 Considering a SWP for a 2D=PSF, $k = 25$ because we have nine points in three dipole orientations, where the intensity in the geometric focus is the same for all. For a 3D-SWP, the number of points is $3 \cdot 19 - 2 = 55$.

PSF associated with a particular genome can simply be highly asymmetrical with very high intensities in certain rim-points and very low intensities in others. In fact, most randomly generated PSFs will look like this. In order to prohibit low rim-intensities, another component of the fitness value is related to the minimal intensity encountered in the rim.

All in all, the fitness value F of a certain genome is calculated as

$$F = \left(\frac{I_0}{\max I_{k\neq 0}} - 1 \right)^a \cdot (\min I_{k\neq 0})^b \quad a \text{ even.} \tag{C.1}$$

It can be seen that F will be small when either the central intensity I_0 is high or when the intensity in only one of the rim points is low. In contrast, when the central intensity is 0 and the intensity along all of the rim is 1, F will approach its maximal value of 1. The exponents a and b are used for weighting. a has to be an even number to ensure non-negative fitness values. In practice, rim-intensities not too low (down to 50%) can be tolerated as long as the central zero is not compromised. Therefore, values are chosen such that $a \gg b$, in this work $a = 20$ and $b = 1$ for all runs of the algorithm. The high value for a may come to a surprise but an intensity in the central zero with $I_0 / \max I_{k\neq 0} \leq 0.01$ is of paramount importance for good STED results. As soon as it is above this value the SWP is practically useless, however high the intensity might be in all rim points. This fact is accounted for by imposing a high penalty on genomes encoding for configurations with even low central intensities.

Now that the `populationFitness` is known, the central part of the algorithm comes into play which consists of the steps *selection, crossover and mutation* and subsequently *reinsertion*.

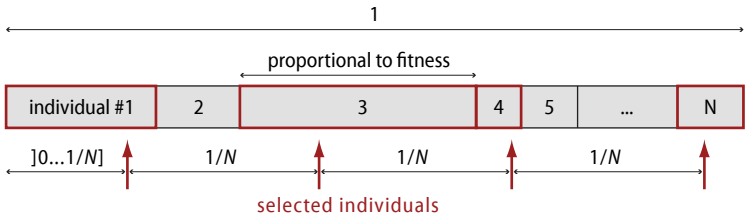

Figure C.2.: On stochastic universal sampling.

Selection is the process of choosing individuals considered suitable for reproduction. Obviously, as in nature, only the fittest individuals should be able to propagate, but when selection is too stringent the algorithm will quickly get stuck in the first local maximum or even an arbitrary genome. Therefore, fitter individuals should be selected for reproduction with higher probability, but weak individuals should also get a chance, for

they might evolve into something useful in the future. Here, selection is accomplished by means of *stochastic universal sampling* [6]. In short (see Figure C.2), all individuals are mapped to a line $[0 \ldots 1]$ with the length of the segment an individual n occupies on this line being proportional to its fitness F_n. Next, a random number r in the interval $]0 \ldots 1/N]$ is generated where N is the number of individuals to be selected, here, it equals `populationSize`. The selected individuals are those whose segments on the line occupy the numbers $r, 2r, \ldots, Nr$. In Matlab code the selection algorithm is looks like

```
% selection by stochastic universal sampling

% compute sum of all fitness values
fitnessSum=0;
for currentGenome = 1:populationSize
    fitnessSum = fitnessSum + populationFitness(currentGenome);
end;

% selectionProbability = length of line segment
for currentGenome = 1:populationSize
    selectionProbability(currentGenome) = populationFitness(
        currentGenome) / fitnessSum;
    if( currentGenome ~= 1)
        selectionProbability(currentGenome) =
            selectionProbability(currentGenome) +
            selectionProbability(currentGenome-1);
    end;
end;

pointerDistance = 1/populationSize;   % spacing of pointers
currentPointer = mod(rand(1)*1000,pointerDistance);     % the
    first pointer; a random value r

for curSelection = 1:populationSize
    for index = 1:populationSize
        if( currentPointer > selectionProbability(index) )
            if( currentPointer <= selectionProbability(index+1) )
                crossoverSelection(curSelection) = index+1;
                break;
            end;
        else
            crossoverSelection(curSelection) = index;
            break;
        end;
    end;
    currentPointer = currentPointer + pointerDistance;
end;
```

In the end, `crossoverSelection` contains the indices pointing to individuals of the `population` that have been selected for reproduction. The next step is called crossover, which recombines parts of two genomes in order to generate new offspring. A typical crossover technique is one-point crossover, where a random location in the parents' genome is cho-

sen and basepairs behind that point are swapped between parents. Similarly, two or n-point crossover is common. For the problem of optimizing SWPs, it should be recognized that a good waveplate often exhibits certain patterns, for example, opposite segments with electric fields phase-shifted by π obviously contribute to low intensities in the zero. In light of this, the crossover procedure chosen here swaps whole waveplate segments, i.e. codons between parents with a `crossoverProbability = 0.8`. This way, chances are high that good combinations of codons, such as the one just mentioned, are combined with other good combinations and are passed on to the next generation, hopefully producing an even better configuration. In Matlab:

```
%crossover complete waveplate-segments
for thisIndividual = 1:populationSize
        selection = mod(floor(rand(2,1) * 100), populationSize) +
             1;   % choose parents randomly from the pool of
            selected individuals
        if( rand(1) <= crossoverProbability )
            for codon = 1:codonSize:genomeSize
                crossoverMatrix(codon:codon+codonSize-1) = round(
                    rand(1));   % contains 1 for basepairs to be
                    swapped, 0 otherwise
            end;
            inverseCrossoverMatrix = xor(crossoverMatrix,
                matrixOnes);    % crossover matrix for other
                parent
            newPopulation(thisIndividual,:) = population(
                crossoverSelection(selection(1)),:) .*
                crossoverMatrix';
            newPopulation(thisIndividual,:) = newPopulation(
                thisIndividual,:) + population(crossoverSelection
                (selection(2)),:) .* reverseCrossoverMatrix';
    else
            newPopulation(thisIndividual,:) = population(
                crossoverSelection(selection(1)),:);
        end;
end;
```

After crossover, `newPopulation` contains the children of the previous generation. Now, single basepairs are randomly mutated with the probability `mutationProbability`.

```
%mutate random basepairs
for thisIndividual = 1:populationSize
    for basepair = 1:genomeSize
        if( (rand(1) <= mutationProbability) )
            newPopulation(thisIndividual, basepair) = round(
                lower_lim + ((upper_lim - lower_lim) * rand(1)));
        end;
    end;
end;
```

Here, `mutationProbability = 0.1`, throughout.

All in all, crossover ensured good genes are passed on, whereas mutation gives evolution its random twist. The last step is reinsertion, where offspring is inserted into the population. The simplest scheme is to replace all parents by offspring, however, good information might be lost in this process. Therefore, an elitist reinsertion scheme was chosen here, where a number `parentPreservation` of parents is kept (the fittest ones), and the rest of the population is filled up with the fittest offspring. This way, the best individuals can live for many generations and have a high probability of passing on their good genes; at the same time, new individuals are introduced. The best value for `parentPreservation` was found to be between 10-20% of `populationSize`. After reinsertion, the new generation is completed and the algorithm enters another optimization loop until the terminal conditions are reached.

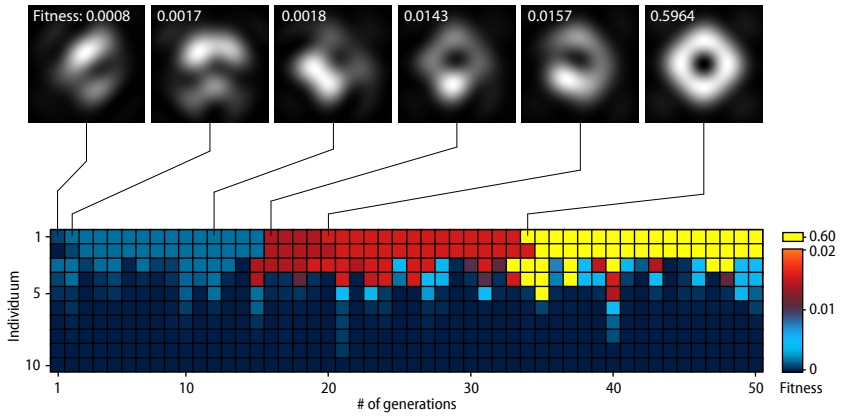

Figure C.3.: Evolution of a 2D-SWP with a genetic algorithm. Rows correspond to individuals, whereas the columns denote successive generations; color codes for fitness. Greyscale: simulated PSFs.

A typical run of the algorithm with a population of twelve individuals is shown in Figure C.3. Rows correspond to individuals, whereas the columns denote successive generations. `parentPreservation` was set to 2, that is why fitter offspring always pops up in individual three (in this chart, the surviving parents are individuals number 1 and 2 and the offspring 3–9 is sorted by fitness). It can also be seen that any newly found individual quickly spreads through the population, thus generating fit offspring, but in no case does it take over the whole population because crossover and mutation ensure genetic diversity and full exploration of the search space. Also note the condensed color scale; up to the 50th generation the fitness is below 0.03, when suddenly a favorable mutation or

crossover[1] leads to a point spread function with high rim-intensities for all three tested dipole orientations and zero intensity in the center. It corresponds to a fitness value of 0.59 and is exactly the segmented wave plate described on pages 44ff. This SWP is found in about nine out of ten cases; despite hundreds of runs, however, with different settings and with up to hundred individuals, a better wave plate has never been discovered.

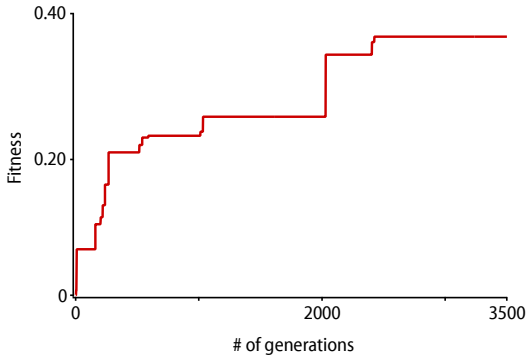

Figure C.4.: Evolution of a 3D-SWP with a genetic algorithm.

The 3D-SWP makes work considerably harder for the genetic algorithm. Because changes in one segment (e.g. via cross-over) have far less impact — in other words, the gradient of the fitness function is much lower — it is harder for individuals to evolve into a better configuration. From another point of view, most of the time a configuration that is reachable for an individual within the next few generations is not much better than the previous one. Moreover, a configuration reachable within the next few generations leading towards the optimum is not much better than all the other reachable configurations. For these reasons, the algorithm makes heavy use of the "noHope"-feature: when the best fitness value has not changed for 160 generations the population is randomized. Of course, the best wave plate found so far is still memorized, in case it was (close to) the optimal solution. Still, on average, a population of 20 individuals needs between 1,000 and 10,000 generations[2] to evolve into what seems to be the best 3D-SWP; in fact the one presented on pages 60ff. During this time it has probed not more than 200,000 configurations, still much less than the $4^{16} = 4,294,967,296$ possible with eight segments.

Figure C.4 depicts a typical run. Because of the huge number of generations, it has been refrained from plotting the whole population as in Figure C.3, instead, the best fitness value found so far is shown. In the 2,425th

1 In fact, a single point mutation in this case.

2 Remarkably, this figure is quite stable against a change is parameters.

generation a individual with fitness 0.3684 comes to live; it is the familiar 3D-SWP. Afterwards, the algorithm continued up to the 20,000th generation, without a change of the best fitness. About 20 of these runs have been conducted with different parameter values, including runs where the "noHope"-case was delayed until up to 500 generations. Yet, despite searching for about 100,000 generations in total, a better SWP never came up. Nevertheless, a genetic algorithm can not prove a solution is the optimum, it can only be stopped when certain criteria are met, in other words, when an acceptable solution has been found, as it is the case here. Needless to say, it cannot be excluded that further research will reveal a (most likely only slightly) better 3D-SWP.

Bibliography

[1] Method of the year 2008. *Nature Methods*, 6(1), 2009.

[2] E. Abbe. Beiträge zur Theorie des Mikroskops und der mikroskopischen Wahrnehmung. *Archiv für mikroskopische Anatomie*, 9(1):413–418, 1873.

[3] Ernst Abbe. *Gesammelte Abhandlungen. Abhandlungen über die Theorie des Mikroskops*, volume 1. Fischer, Jena, 1904.

[4] W.B. Amos, J.G. White, and M. Fordham. Use of confocal imaging in the study of biological structures. *Applied Optics*, 26(16):3239–3243, 1987.

[5] E. A. Ash and G. Nicholls. Super-resolution aperture scanning microscope. *Nature*, 237:510, 1972.

[6] J.E. Baker. Reducing bias and inefficiency in the selection algorithm. In *Proceedings of the Second International Conference on Genetic Algorithms on Genetic algorithms and their application*, pages 14–21. L. Erlbaum Associates Inc., 1987.

[7] E. Betzig, G.H. Patterson, R. Sougrat, O.W. Lindwasser, S. Olenych, J.S. Bonifacino, M.W. Davidson, J. Lippincott-Schwartz, and H.F. Hess. Imaging intracellular fluorescent proteins at nanometer resolution. *Science*, 313(5793):1642, 2006.

[8] G. Binnig, C. F. Quate, and Ch. Gerber. Atomic force microscope. *Phys. Rev. Lett.*, 56(9):930–933, Mar 1986.

[9] F.R. Blattner, G. Plunkett III, C.A. Bloch, N.T. Perna, V. Burland, M. Riley, J. Collado-Vides, J.D. Glasner, C.K. Rode, G.F. Mayhew, et al. The complete genome sequence of Escherichia coli K-12. *Science*, 277(5331):1453, 1997.

[10] N. Bokor, Y. Iketaki, T. Watanabe, K. Daigoku, N. Davidson, and M. Fujii. On polarization effects in fluorescence depletion microscopy. *Optics Communications*, 272(1):263–268, 2007.

[11] M. Born and E. Wolf. *Principles of Optics*. Elsevier, 1975.

[12] S. Bosch. *Lineare Algebra*. Springer, 2008.

[13] V.P. Boyarskiy, V.N. Belov, R. Medda, B. Hein, M. Bossi, and S.W. Hell. Photostable, amino reactive and water-soluble fluorescent labels based on sulfonated rhodamine with a rigidized xanthene fragment. *Chemistry-A European Journal*, 14(6):1784–1792, 2008.

[14] R. Clay and T. Court. *The History of the Microscope*. Longwood Pr Ltd, 1977.

[15] A.H. Coons, H.J. Creech, and R.N. Jones. Immunological properties of an antibody containing a fluorescent group. *Soc Exp Biol Med*, 47:200–202, 1941.

[16] J.N. Damask. *Polarization optics in telecommunications*. Springer Verlag, 2004.

[17] P. Debye. Das Verhalten von Lichtwellen in der Nähe eines Brennpunktes oder einer Brennlinie. *Annalen der Physik*, 335:755–776, 1909.

[18] J.B. Ding, K.T. Takasaki, and B.L. Sabatini. Supraresolution imaging in brain slices using stimulated-emission depletion two-photon laser scanning microscopy. *Neuron*, 63(4):429–437, 2009.

[19] C.G. Dotti and K. Simons. Polarized sorting of viral glycoproteins to the axon and dendrites of hippocampal neurons in culture. *Cell*, 62(1):63–72, 1990.

[20] T. Dresbach, V. Torres, N. Wittenmayer, W.D. Altrock, P. Zamorano, W. Zuschratter, R. Nawrotzki, N.E. Ziv, C.C. Garner, and E.D. Gundelfinger. Assembly of Active Zone Precursor Vesicles. *Journal of Biological Chemistry*, 281(9):6038, 2006.

[21] M. Dyba, J. Keller, and S.W. Hell. Phase filter enhanced STED-4pi fluorescence microscopy: theory and experiment. *New Journal of Physics*, 7(1):134, 2005.

[22] RF Egerton, P. Li, and M. Malac. Radiation damage in the TEM and SEM. *Micron*, 35(6):399–409, 2004.

[23] C. Eggeling, C. Ringemann, R. Medda, G. Schwarzmann, K. Sandhoff, S. Polyakova, B. Hein, C. von Middendorff, A. Schönle, and S.W. Hell. Direct observation of the nanoscale dynamics of membrane lipids in a living cell. *Nature*, 457(7233):1159–1162, 2008.

[24] A. Egner, C. Geisler, C. von Middendorff, H. Bock, D. Wenzel, R. Medda, M. Andresen, A.C. Stiel, S. Jakobs, C. Eggeling, et al. Fluorescence nanoscopy in whole cells by asynchronous localization of photoswitching emitters. *Biophysical Journal*, 93(9):3285–3290, 2007.

[25] A. Fejtova, D. Davydova, F. Bischof, V. Lazarevic, W.D. Altrock, S. Romorini, C. Schone, W. Zuschratter, M.R. Kreutz, C.C. Garner, et al. Dynein light chain regulates axonal trafficking and synaptic levels of bassoon. *Journal of Cell Biology*, 185(2):341, 2009.

[26] J. Fölling, M. Bossi, H. Bock, R. Medda, C.A. Wurm, B. Hein, S. Jakobs, C. Eggeling, and S.W. Hell. Fluorescence nanoscopy by ground-state depletion and single-molecule return. *Nature methods*, 5(11):943–945, 2008.

[27] C. Geisler, A. Schönle, C. von Middendorff, H. Bock, C. Eggeling, A. Egner, and S.W. Hell. Resolution of $\lambda/10$ in fluorescence microscopy using fast single molecule photo-switching. *Applied Physics A: Materials Science & Processing*, 88(2):223–226, 2007.

[28] A. Gerrard and J.M. Burch. *Introduction to matrix methods in optics*. Dover Publications, 1994.

[29] T.J. Gould, M.S. Gunewardene, M.V. Gudheti, V.V. Verkhusha, S.R. Yin, J.A. Gosse, and S.T. Hess. Nanoscale imaging of molecular positions and anisotropies. *Nature Methods*, 5(12):1027–1030, 2008.

[30] M. Gu. *Advanced Optical Imaging Theory*. Springer-Verlag, 2000.

[31] 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.

[32] M.G.L. Gustafsson. Nonlinear structured-illumination microscopy: Wide-field fluorescence imaging with theoretically unlimited resolution. *Proceedings of the National Academy of Sciences*, 102(37):13081, 2005.

[33] 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, 2005.

[34] T. Ha, T. Enderle, D.S. Chemla, P.R. Selvin, and S. Weiss. Single molecule dynamics studied by polarization modulation. *Physical review letters*, 77(19):3979–3982, 1996.

[35] B. Harke, J. Keller, C.K. Ullal, V. Westphal, A. Schönle, and S.W. Hell. Resolution scaling in STED microscopy. *Optics Express*, 16(6):4154–4162, 2008.

[36] B. Harke, C.K. Ullal, J. Keller, and S.W. Hell. Three-dimensional nanoscopy of colloidal crystals. *Nano Letters*, 8(5):1309–1313, 2008.

[37] B. Harke, C.K. Ullal, J. Keller, and S.W. Hell. Three-dimensional nanoscopy of colloidal crystals. *Nano Lett*, 8(5):1309–1313, 2008.

[38] M. Heilemann, S. van de Linde, M. Schüttpelz, R. Kasper, B. Seefeldt, A. Mukherjee, P. Tinnefeld, and M. Sauer. Subdiffraction-resolution fluorescence imaging with conventional fluorescent probes. *Angew. Chemie. Int. Ed*, 47:6172–6176, 2008.

[39] B. Hein, K.I. Willig, and S.W. Hell. Stimulated emission depletion (STED) nanoscopy of a fluorescent protein-labeled organelle inside a living cell. *Proceedings of the National Academy of Sciences*, 105(38):14271, 2008.

[40] R. Heintzmann, T.M. Jovin, and C. Cremer. Saturated patterned excitation microscopy—a concept for optical resolution improvement. *Journal of the Optical Society of America A*, 19(8):1599–1609, 2002.

[41] S. W. Hell. Double confocal microscope. European patent 0 491 289, 1990.

[42] S. W. Hell, M. Schrader, and H. T. M. van der Voort. Far-field fluorescence microscopy with three-dimensional resolution in the 100nm range. *J. Microsc.*, 185:1–5, 1997.

[43] S. W. Hell and E. H. K. Stelzer. Fundamental improvement of resolution with a 4pi-confocal fluorescence microscope using two-photon excitation. *Opt. Commun.*, 93:277 – 282, 1992.

[44] S.W. Hell. Far-field optical nanoscopy. *Single Molecule Spectroscopy in Chemistry, Physics and Biology*, pages 365–398.

[45] S.W. Hell. Microscopy and its focal switch. *Nat. Methods*, 6(1):24–32, 2009.

[46] S.W. Hell, S. Jakobs, and L. Kastrup. Imaging and writing at the nanoscale with focused visible light through saturable optical transitions. *Applied Physics A*, 77(7):859–860, 2003.

[47] S.W. Hell and M. Kroug. Ground-state-depletion fluorscence microscopy: A concept for breaking the diffraction resolution limit. *Applied Physics B: Lasers and Optics*, 60(5):495–497, 1995.

[48] R. Henderson and P. N. T. Unwin. Three-dimensional model of purple membrane obtained by electron microscopy. *Nature*, 257(5521):28, 1975.

[49] S.T. Hess, T.P.K. Girirajan, and M.D. Mason. Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. *Biophysical journal*, 91(11):4258–4272, 2006.

[50] 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. *Proceedings of the National Academy of Sciences*, 102(49):17565, 2005.

[51] B. Huang, S.A. Jones, B. Brandenburg, and X. Zhuang. Whole cell 3D STORM reveals interactions between cellular structures with nanometer-scale resolution. *Nature methods*, 5(12):1047, 2008.

[52] R.K. Iler. The Chemistry of Silica: Solubility, Polymerization. *Colloid and Surface Properties, and Biochemistry, Wiley, New York*, 886, 1979.

[53] R.C. Jones. A new calculus for the treatment of optical systems. *Journal of the Optical Society of America*, 31(7):500–503, 1941.

[54] C.L. Karr and L.M. Freeman. *Industrial applications of genetic algorithms*. CRC, 1999.

[55] L. Kastrup, H. Blom, C. Eggeling, and S.W. Hell. Fluorescence fluctuation spectroscopy in subdiffraction focal volumes. *Physical Review Letters*, 94(17):178104, 2005.

[56] J. Keller. *Optimal de-excitation patterns for RESOLFT-Microscopy*. PhD thesis, Universität Heidelberg, Fakultät für Physik und Astronomie, 2006.

[57] J. Keller. *Adaptive elements in STED-microscopy*. PhD thesis, Universität Heidelberg, Fakultät für Physik und Astronomie, 2009.

[58] J. Keller, A. Schönle, and S.W. Hell. Efficient fluorescence inhibition patterns for resolft microscopy. *Optics Express*, 15(6):3361–3371, 2007.

[59] P. Kirkpatrick and AV Baez. Formation of optical images by x-rays. *Journal of the Optical Society of America*, 38(9):766–773, 1948.

[60] R.J. Kittel, C. Wichmann, T.M. Rasse, W. Fouquet, M. Schmidt, A. Schmid, D.A. Wagh, C. Pawlu, R.R. Kellner, K.I. Willig, et al. Bruchpilot promotes active zone assembly, Ca2+ channel clustering, and vesicle release. *Science*, 312(5776):1051, 2006.

[61] M. Knoll and E. Ruska. Das elektronenmikroskop. *Zeitschrift für Physik A Hadrons and Nuclei*, 78(5):318–339, 1932.

[62] J. R. Lakowicz. *Principles of Flourescence Spectroscopy*. Plenum Press, New York, third edition edition, 2006.

[63] Y. Li and E. Wolf. Three-dimensional intensity distribution near the focus in systems of different fresnel numbers. *J. Opt. Soc. Am. A*, 1:801 – 808, 1984.

[64] J.W. Lichtman and S.E. Fraser. The neuronal naturalist: watching neurons in their native habitat. *nature neuroscience*, 4:1215–1220, 2001.

[65] M.A. Lieb, J.M. Zavislan, and L. Novotny. Single-molecule orientations determined by direct emission pattern imaging. *JOSA B*, 21(6):1210–1215, 2004.

[66] J. Lippincott-Schwartz and S. Manley. Putting super-resolution fluorescence microscopy to work. *Nat. Methods*, 6:21–23, 2009.

[67] G. Machavariani, Y. Lumer, I. Moshe, A. Meir, and S. Jackel. Efficient extracavity generation of radially and azimuthally polarized beams. *Optics letters*, 32(11):1468–1470, 2007.

[68] S. Manley, J.M. Gillette, G.H. Patterson, H. Shroff, H.F. Hess, E. Betzig, and J. Lippincott-Schwartz. High-density mapping of single-molecule trajectories with photoactivated localization microscopy. *Nature Methods*, 5(2):155–158, 2008.

[69] M. Mansuripur. Distribution of light at and near the focus of high-numerical-aperture objectives. *J. Opt. Soc. Am. A*, 3(12):2086–2093, 1986.

[70] P.C. Maurer, J.R. Maze, P.L. Stanwix, L. Jiang, A.V. Gorshov, A.A. Zibrov, B. Harke, J.S. Hodges, A.S. Zibrov, A. Yacoby, D. Twitchen, S.W. Hell, R.L. Walsworth, and M.D. Lukin. Far-field optical imaging and manipulation of individual spins with nanoscale resolution. *Nature Physics*, 2010.

[71] R. Menon, P. Rogge, and H.Y. Tsai. Design of diffractive lenses that generate optical nulls without phase singularities. *J. Opt. Soc. Am. A*, 26:297–304, 2009.

[72] L. Meyer, D. Wildanger, R. Medda, A. Punge, S.O. Rizzoli, G. Donnert, and S.W. Hell. Dual-color STED microscopy at 30-nm focal-plane resolution. *Small*, 4(8):1095–1100, 2008.

[73] M. Minsky. Microscopy apparatus. US Patent # 3013467, 1957.

[74] M. Mitchell. *An introduction to genetic algorithms*. The MIT press, 1998.

[75] G. Moneron and S.W. Hell. Two-photon excitation STED microscopy. *Opt. Express*, 17(17):14567–14573, 2009.

[76] G. Moneron, R. Medda, B. Hein, A. Giske, V. Westphal, and S.W. Hell. Fast STED microscopy with continuous wave fiber lasers. *Opt. Express*, 18:1302–1309, 2010.

[77] U.V. Nägerl, K.I. Willig, B. Hein, S.W. Hell, and T. Bonhoeffer. Live-cell imaging of dendritic spines by STED microscopy. *Proceedings of the National Academy of Sciences*, 105(48):18982, 2008.

[78] T. Okoshi. Polarization-state control schemes for heterodyne or homodyne optical fiber communications. *Journal of Lightwave Technology*, 3(6):1232–1237, 1985.

[79] Ralston A. P. *Discoverer of the Unseen World: A Biography of Antoni van Leeuwenhoek*. World Pub. Co, 1966.

[80] D. Patra, I. Gregor, and J. Enderlein. Image analysis of defocused single-molecule images for three-dimensional molecule orientation studies. *J. Phys. Chem. A*, 108(33):6836–6841, 2004.

[81] J. Pawley, editor. *Handbook of Biological Confocal Microscopy*. Springer, 2 edition, 1995.

[82] P.B. Phua, W.J. Lai, Y.L. Lim, K.S. Tiaw, B.C. Lim, H.H. Theo, and M.H. Hong. Mimicking optical activity for generating radially polarized light. *Optics Letters*, 32(4):376–378, 2007.

[83] J.S. Ploem. The use of a vertical illuminator with interchangeable dichroic mirrors for fluorescence microscopy with incident light. *Z Wiss Mikrosk*, 68(3):129–142, 1967.

[84] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. *Numerical recipes: the art of scientific computing*. Cambridge Univ Pr, 2007.

[85] A. Punge, S.O. Rizzoli, R. Jahn, J.D. Wildanger, L. Meyer, A. Schönle, L. Kastrup, and S.W. Hell. 3D reconstruction of high-resolution STED microscope images. *Microscopy research and technique*, 71(9):644–650, 2008.

[86] C. Ringemann, B. Harke, C. von Middendorff, R. Medda, A. Honigmann, R. Wagner, M. Leutenegger, A. Schönle, and S.W. Hell. Exploring single-molecule dynamics with fluorescence nanoscopy. *New Journal of Physics*, 11:103054, 2009.

[87] E. Rittweger, K.Y. Han, S.E. Irvine, C. Eggeling, and S.W. Hell. STED microscopy reveals crystal colour centres with nanometric resolution. *Nature Photonics*, 3:144–147, 2009.

[88] M.J. Rust, M. Bates, and X. Zhuang. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nature methods*, 3(10):793, 2006.

[89] T. Sato, K. Miura, N. Ishino, Y. Othera, T. Tamamura, and S. Kawakami. Photonic crystals for the visible range fabricated by autocloning technique and their application. *Optical and quantum electronics*, 34(1):63–70, 2002.

[90] D. Sauter. *4Pi live cell imaging*. PhD thesis, Universität Heidelberg, Fakultät für Physik und Astronomie, 2009.

[91] L. Schermelleh, P.M. Carlton, S. Haase, L. Shao, L. Winoto, P. Kner, B. Burke, M.C. Cardoso, D.A. Agard, M.G.L. Gustafsson, et al. Subdiffraction multicolor imaging of the nuclear periphery with 3D structured illumination microscopy. *Science*, 320(5881):1332, 2008.

[92] R. Schmidt, C.A. Wurm, S. Jakobs, J. Engelhardt, A. Egner, and S.W. Hell. Spherical nanosized focal spot unravels the interior of cells. *Nature Methods*, 5(6):539–544, 2008.

[93] R. Schmidt, C.A. Wurm, S. Jakobs, J. Engelhardt, A. Egner, and S.W. Hell. Spherical nanosized focal spot unravels the interior of cells. *Nature Methods*, 5(6):539–544, 2008.

[94] K.A. Serrels, E. Ramsay, P.A. Dalgarno, B. Gerardot, J.A. O□Connor, R.H. Hadfield, R.J. Warburton, and D.T. Reid. Solid immersion lens applications for nanophotonic devices. *Journal of Nanophotonics*, 2:021854, 2008.

[95] B. Sick, B. Hecht, and L. Novotny. Orientational imaging of single molecules by annular illumination. *Physical Review Letters*, 85(21):4482–4485, 2000.

[96] I. Testa, A. Schönle, C. v. Middendorff, C. Geisler, R. Medda, C.A. Wurm, A.C. Stiel, S. Jakobs, M. Bossi, C. Eggeling, et al. Nanoscale separation of molecular species based on their rotational mobility. *Optics Express*, 16(25):21093–21104, 2008.

[97] P. Török and P. Munro. The use of gauss-laguerre vector beams in STED microscopy. *Optics Express*, 12(15):3605–3617, 2004.

[98] R.Y. Tsien. The green fluorescent protein. *Annual review of biochemistry*, 67(1):509–544, 1998.

[99] N. Unwin. Acetylcholine receptor channel imaged in the open state. *Nature*, 373:37–43, 1995.

[100] T. D. Visser, G. J. Brakenhoff, and F. C. A. Groen. The one-point fluorescence response in confocal microscopy. *Optik*, 87(1):39–40, 1991.

[101] N.B.Y.I.T. Watanabe and M. Fujii. Compact fluorescence depletion microscope system using an integrated optical element. *Optics Communications*, 281(7), 2008.

[102] S. Weiss. Fluorescence spectroscopy of single biomolecules. *Science*, 283(5408):1676, 1999.

[103] V. Westphal, C.M. Blanca, M. Dyba, L. Kastrup, and S.W. Hell. Laser-diode-stimulated emission depletion microscopy. *Applied Physics Letters*, 82:3125, 2003.

[104] V. Westphal and S.W. Hell. Nanoscale resolution in the focal plane of an optical microscope. *Physical Review Letters*, 94(14):143903, 2005.

[105] V. Westphal, S.O. Rizzoli, M.A. Lauterbach, D. Kamin, R. Jahn, and S.W. Hell. Video-rate far-field optical nanoscopy dissects synaptic vesicle movement. *Science*, 320(5873):246, 2008.

[106] J.G. White, W.B. Amos, and M. Fordham. An evaluation of confocal versus conventional imaging of biological structures by fluorescence light microscopy. *Journal of Cell Biology*, 105(1):41, 1987.

[107] D. Wildanger, J. Bückers, V. Westphal, S.W. Hell, and L. Kastrup. A STED microscope aligned by design. *Optics Express*, 17(18):16100–16110, 2009.

[108] D. Wildanger, R. Medda, L. Kastrup, and S.W. Hell. A compact STED microscope providing 3 D nanoscale resolution. *Journal of Microscopy(Oxford)*, 236(1):35–43, 2009.

[109] D. Wildanger, E. Rittweger, L. Kastrup, and S.W. Hell. STED microscopy with a supercontinuum laser source. *Optics Express*, 16(13):9614–9621, 2008.

[110] G. Williams. *Linear algebra with applications*. Jones & Bartlett Publishers, 2009.

[111] K.I. Willig, B. Harke, R. Medda, and S.W. Hell. STED microscopy with continuous wave beams. *Nature Methods*, 4(11):915–918, 2007.

[112] K.I. Willig, R.R. Kellner, R. Medda, B. Hein, S. Jakobs, and S.W. Hell. Nanoscale resolution in GFP-based microscopy. *Nature Methods*, 3(9):721–723, 2006.

[113] 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.

[114] E. Wolf and Y. Li. Conditions for the validity of the Debye integral representation of focused fields. *Optics Communications*, 39(4):205–210, 1981.

[115] A. Yariv and P. Yeh. *Optical waves in crystals*. Wiley New York, 1984.

[116] A. Yildiz, J.N. Forkey, S.A. McKinney, T. Ha, Y.E. Goldman, and P.R. Selvin. Myosin V walks hand-over-hand: single fluorophore imaging with 1.5-nm localization. *Science*, 300(5628):2061, 2003.

[117] H. Zheng and E.H. Koo. The amyloid precursor protein: beyond amyloid. *Molecular Neurodegeneration*, 1:5, 2006.

[118] Z. Zhuang, S.W. Suh, and J.S. Patel. Polarization controller using nematic liquid crystals. *Optics Letters*, 24(10):694–696, 1999.

Acknowledgements

This thesis would not have been possible without the contributions of a good many people. First and foremost, I would like to express particular appreciation to Prof. Dr. Stefan Hell for giving me the opportunity to work in a tremendously interesting field and for the outstanding working environment he provided. He was always eager to contribute with fruitful ideas and discussions (and money).

A very special thanks goes out to Dr. Johann Engelhardt for his support. He greatly enhanced my knowledge of optical, electrical and mechanical engineering and provided crucial advice for the realization of this work.

Furthermore, many thanks to Prof. Dr. Rainer Fink for agreeing to be the second referee, and to Prof. Dr. Ulrich Schwarz and Prof. Dr. Christian Enss, the remaining members of the examining board.

Finally, I'd like to thank the following people:
In particular: Angelika and Bernhard Reuss and Lisa Bröder for their support, as well as Thorsten Staudt, Dorothea Sauter, Marion Lang, Tim Kees and Tilo Senger and Tina Büchling and Julia Strathmann and Dorothée Deppe and Tobias Bauer and Markus Moosmeier and Moritz Durchdewald, Prof. Wolfhard Semmler and Prof. Joachim Spatz (the thesis committee at the DKFZ), Josephine Stadler, Andreas Engler, Heinrich Rühle and Volker Stamm and Wolfram Stroh from the machine shop, Christian Eggeling, Jasmin Zahn, Bianca Nöth, Simone Back, Thomas Dresbach, the internet, Jale Schneider, Pit Bingen, Franziska Curdt, Silvia Löbermann, Gael Moneron, Patrick Hoyer, Susanne Schrof, Ruth Wagner, Eva Rittweger and again Hans Engelhardt for last-minute corrections (thank you), Johanna Berndt and many more. Your support is cordially appreciated.