FRIEDRICH-SCHILLER-UNIVERSITÄT JENA

DOCTORAL THESIS

Improving the signal-to-noise ratio in incoherent imaging

by Jan BECKER

born on 20.05.1992 in Erlangen

A thesis submitted in fulfillment of the
requirements for the degree of

Doctor rerum naturalium (Dr. rer. nat.)

in the

Physikalisch-Astronomische Fakultät

Reviewers:

Prof. Dr. Rainer HEINTZMANN
Leibniz Institute of Photonic Technologies, Jena
Institute of Physical Chemistry, Friedrich-Schiller-University Jena

Prof. Dr. Colin SHEPPARD
School of Chemistry and Molecular Biosciences, University of Wollongong
Nanophysics, Italian Institute of Technology, Genova

Prof. Dr. Jörg ENDERLEIN
Biophysics and Complex Systems
Third Institute of Physics, Georg-August-University, Göttingen

Date of disputation: 12.07.2021

Declaration of Authorship:

I hereby declare that I have prepared this thesis independently, without the unauthorized assistance of third parties and without the use of any aids and literature other than those indicated. The data and concepts taken directly or indirectly from other sources are marked with reference to the source.
In the selection and evaluation of the following material, the persons listed below have assisted me in the manner described in each case, against no payment:

1. *Dr. Katharina Reglinksi* (ZAF, Uni Jena) helped with the acquisition of the experimental data regarding the *wavelength splitting*.
2. *Dr. Takahiro Deguchi* (EMBL, Heidelberg) helped with the acquisition of the experimental data regarding the *pupil splitting*.
3. *Dr. James Manton* (MRC LMB, Cambridge) did carry out the acquisition of the experimental data regarding the *illumination splitting*.
4. *Dr. Detlef Born* (IPHT, Jena) helped me with setting up the passive submillimeter wave imager and the acquisition of the *submillimeter wave data*.

No other persons were involved in the preparation of this thesis. In particular, I have not made use of paid assistance of intermediary or consulting services (doctoral advisors or other persons) for this purpose. No one has directly or indirectly received any monetary benefits from me for work related to the content of the submitted dissertation. The work has not yet been submitted to another examination authority in the same or similar form, either in Germany or abroad.
I am familiar with the applicable doctoral regulations of the Physikalisch-Astronomische Fakultät of the Friedrich-Schiller-Universität in Jena.
Herewith I affirm that I have told the pure truth to the best of my knowledge and have not concealed anything.

Date and place: *August 22, 2021; Herzogenaurach* Signature: *Jan Becker*

"We microscopists, however, feel particularly indebted to ***Ernst Abbe****, whose tireless efforts are largely responsible for the current perfection of our instruments."*

Eduard Strassburger
Botanist, 1844 - 1912

For Liuba, my family and friends.

Abstract

Improving the signal-to-noise ratio in incoherent imaging

by Jan BECKER

Imaging plays an important role in our modern world. Applications of imaging technologies range from inspection of industrial goods to screening of patients in the healthcare sector or surveillance in public areas. Besides limitations in terms of resolution and contrast, all imaging systems show degrading image quality with increasing noise. Hence, an important aspect of any imaging application is the achievable signal-to-noise ratio (*SNR*). When neglecting typical noise sources originating from detectors or the environment, we are left with a very fundamental noise type: photon or *shot noise*. Which is due to the discrete nature of light - the photon - and the fact that imaging is associated with counting those particles. Leading to a noisy measurement process which fundamentally limits the achievable *SNR*. In this thesis that limitation will be overcome by employing an approach termed *splitting & recombination*. The basic idea is to actively influence a conventional imaging system in such a way, that different "views" from the sample are being captured, which then get computationally reconstructed to yield an effective increase in *SNR*. Four possibilities for the splitting are being examined: 1) separating the polarization of the detectable light into two orthogonal states and making use of a polarization-dependent focusing effect; 2) splitting the emission in a fluorescence microscope, as the resolution gets worse with increasing emission wavelength; 3) pupil splitting in incoherent imaging, which makes use ot the fact that high spatial frequencies can be transferred with improved *SNR*, when medium frequency components are being reduced; 4) separating the illumination in a light-sheet microscope using the Field-Synthesis concept, enabling the detection over a large field-of-view (FoV) while maintaining a axially narrow illumination sheet. A general description of the splitting & recombination idea is presented in this thesis using the polarization splitting as an example; the two main computational recombination techniques employed are: weighted averaging in Fourier space and multiview deconvolution. For each of the remaining three applications a thorough theoretical and numerical treatment is presented, as well as proof-of-principle experiments.

The second part of this thesis focuses on image reconstruction techniques of data acquired with a passive submillimeter wave imager, which has been developed for concealed threat detection in public areas. One typical challenge of such devices is to maintain a good spatio-temporal resolution at a distance larger than 10 m, while being limited to a wavelength which can penetrate through most clothes and fabrics. As the current prototype system employs a line-scanning principle and the detectors show a highly individual signal response, pre-processing is necessary to transform the data into meaningful images. These are then further processed via denoising and deconvolution algorithms to reveal small structures with improved *SNR*, so that a reliable detection of unconcealed threats becomes feasible. Another way to improve the image quality is to employ machine learning (ML) techniques. In this work denoising via a trained network is presented and compared to the more traditional reconstruction methods. The robustness of all reconstruction techniques is shown using a multitude of experimentally acquired data, exemplifying different possible real application scenarios.

Kurzfassung

Verbesserung des Signal-zu-Rausch Verhältnisses in der inkohärenten Bildgebung

von Jan Becker

Die Bildgebung spielt in unserer modernen Welt eine wichtige Rolle. Deren Anwendungen reichen von der Inspektion von Industriegütern, dem medizinischen untersuchen von Patienten im Gesundheitswesen bis zur Überwachung im öffentlichen Raum. Neben Einschränkungen in Bezug auf Auflösung und Kontrast zeigen alle bildgebenden Systeme eine abnehmende Bildqualität mit zunehmendem Rauschen. Ein wichtiger Aspekt ist daher das erreichbare Signal-zu-Rausch-Verhältnis (engl. *SNR*). Vernachlässigt man typische Rauschquellen, wie Ausleserauschen von Detektoren oder Einflüsse aus der Umgebung, bleibt eine fundamentale Art des Rauschens übrig: das Photonen - oder *Schrot*rauschen. Dies liegt an der Quantisierung des Lichts - dem Photon - und der Tatsache, dass Bildgebene Verfahren mit dem Zählen dieser diskreten Teilchen einher gehen. Demnach ist das erreichbare *SNR* in jedem bildgebenden System grundlegend begrenzt. In der vorliegenden Arbeit wird eben jene Einschränkung durch einen neuartigen Ansatz des Teilens und Rekombinierens (engl. *splitting & recombination*) des Bildgebenden Lichtes überwunden. Die Grundidee ist, ein konventionelles Bildgebungssystem aktiv so zu beeinflussen, dass verschiedene "Ansichten" der Probe aufgenommen und anschließend rechnerisch rekombiniert werden. Vier Möglichkeiten für das "splitting" wurden untersucht: 1) Aufspaltung

des detektierten Lichts in zwei orthogonale Polarisationszustände, unter Ausnutzung des Effekts das die Lichtverteilung im Fokus eines Objektivs von der Polarisation abhängt; 2) Aufspaltung der Emission in einem Fluoreszenzmikroskop, da die räumliche Auflösung mit der Emissionswellenlänge skaliert; 3) Pupillentrennung in der inkohärenten Bildgebung, welche die Tatsache ausnutzt, dass hohe Ortsfrequenzen mit verbessertem *SNR* übertragen werden können, wenn niederfrequente Komponenten blockiert werden; 4) die Aufteilung der Beleuchtung in einem Lichtblattmikroskop unter Verwendung des Field-Synthesis-Konzepts, das die Detektion über ein großes Field-of-View (FoV) bei gleichzeitiger Erhaltung einer axial schmalen Beleuchtung ermöglicht. In dieser Arbeit wird eine allgemeine Beschreibung der Idee des *splitting & recombination* am Beispiel der Polarisationsaufspaltung vorgestellt. Die beiden wichtigsten Rekombinationsalgorithmen sind: gewichtetes Mitteln im Fourier-Raum und Multiview Dekonvolution. Für jede der verbleibenden drei Anwendungen wird eine theoretische und numerische Behandlung, sowie Proof-of-Principle-Experimente zur Verifizierung vorgestellt und diskutiert. Der zweite Teil dieser Arbeit konzentriert sich auf Bildrekonstruktionsverfahren von Daten, die mit einem passiven *submillimeter wave imager* erfasst wurden, der für die Erkennung von verdeckten Bedrohungen in der Öffentlichkeit entwickelt wurde. Eine Herausforderung solcher Geräte ist es, eine gute räumlich-zeitliche Auflösung bei einer Entfernung von mehr als 10 m zu erhalten. Diese sind auf eine Wellenlänge beschränkt, welche die meisten Kleidungsstücke sowie Stoffe durchdringen kann. Um die Daten in aussagekräftige Bilder umzuwandeln ist eine gewisse Vorverarbeitung notwendig. Hierbei müssen Effekte wie die Zeilenabtastung des Detektors, sowie die sehr individuellen Detektorsensitivitäten mit berücksichtigt werden. Diese vorverarbeiteten Daten werden dann über Entrauschungs- und Entfaltungsalgorithmen weiterverarbeitet, um kleine Strukturen mit verbessertem *SNR* sichtbar zu machen. Dies ermöglicht eine zuverlässige Erkennung von verdeckten Bedrohungen, mit einer Verarbeitungsgeschwindigkeit das sich bewegende Personen nachverfolgt werden können. Eine weitere Möglichkeit zur Verbesserung der Bildqualität ist der Einsatz von Methoden des maschinellen Lernens (ML). In dieser Arbeit wird die Entrauschung mittels eines trainierten Netzwerks verwendet und mit den traditionelleren Rekonstruktionsmethoden verglichen. Die Robustheit aller Verfahren wird anhand einer Vielzahl von experimentellen Daten gezeigt, die beispielhaft für verschiedene mögliche Szenarien stehen.

Acknowledgements

I'd like to thank Prof. Dr. Rainer Heintzmann for the opportunity to pursue my PhD in his research group, his supervision and for teaching me many useful concepts ranging from Fourier space and noise to optical microscopy in general.

During my time at the IPHT I had the opportunity to share office, coffee and many discussions with different people, who greatly influenced my scientific thinking and who were always there to answer many of my questions. Just to name a few: Dr. Christian Karras, Dr. Ronny Förster, Rene Lachmann, ...

I also want to acknowledge Alexander Jügler, who together with Mathieu Ribes, helped me in progressing with the experimental realization of pupil splitting.

I want to thank Dr. Uwe Hübner for the help with designing and manufacturing the mirror device used in pupil splitting.

For the introduction into the submillimeter wave scanner and the help in setting up the system for measurements I'd like to thank Dr. Erik Heinz, Dr. Detlef Born.

Most of the experiments in this thesis were carried out in different laboratories. Which proofed itself difficult to handle as most experimental data was acquired during the Covid-19 pandemic. To all the collaborators who helped me to experimentally realize the different *splitting & recombination* ideas, I want to express my biggest gratitude: Dr. Katharina Reglinski, Dr. Takahiro Deguchi, Dr. Jonas Ries and Dr. James Manton.

In general I want to thank everybody who used to be or continused to be a member of the Bio-Nanoimaging group in Jena. The atmosphere in the group was always very positive, motivating and almost felt like a small family. Being able to share conferences, group outings and pub visits together with you guys has made my journey through the world of photons, optics and Fourier space much more enjoyable. *A big thank you, to all of you!*

Contents

1 Introduction **1**

2 Theoretical background **3**
- 2.1 Signal in incoherent imaging 3
  - 2.1.1 Image formation process 4
  - 2.1.2 Representation in Fourier space 5
  - 2.1.3 Imaging in three dimensions 7
- 2.2 Noise in imaging applications 9
  - 2.2.1 Degradation of image quality 9
  - 2.2.2 Photon or *shot* noise 11
  - 2.2.3 Out-of-focus: background noise 13
- 2.3 The signal-to-noise ratio (*SNR*) 13
  - 2.3.1 *SNR* in real space 14
  - 2.3.2 In terms of spatial frequencies 14
  - 2.3.3 Contrast and effective resolution 15
  - 2.3.4 Noise normalized OTF and detectability 18
  - 2.3.5 Sampling and pixelation effects 19
- 2.4 Methods to improve *SNR* 20
  - 2.4.1 Improving the photon collection 21
  - 2.4.2 Using computational reconstruction 23
- 2.5 Summary and discussion 26

3 Fluorescence microscopy **28**
- 3.1 Fluorescence as a contrast mechanism 28
  - 3.1.1 Visualisation using the Jablonski diagram 29
  - 3.1.2 Excitation and emission spectra 30
- 3.2 General idea: splitting & recombination 30
  - 3.2.1 Creating mutually differing sub-images 31
  - 3.2.2 Computational recombination 32
- 3.3 Splitting the fluorescence emission 40
  - 3.3.1 PSF broadening due to fluorescence 40

3.3.2 Degradation of image deconvolution 46
3.3.3 Wavelength splitting improves *SNR* 48
3.3.4 Experimental realization 50
3.4 Separation via pupil splitting 52
3.4.1 Information content in the sub-images 52
3.4.2 Evaluating the achievable *SNR* improvement 55
3.4.3 Splitting via reflection in the BFP 56
3.4.4 Experimental results 58
3.5 Illumination splitting through Field-Synthesis 61
3.5.1 Scanned light-sheet generation 62
3.5.2 Encoding axial information in 2D 66
3.5.3 Rejecting out-of-focus light using FS 67
3.5.4 Experimental setup & results 71
3.6 Summary and discussion 74

4 Submillimeter wave imaging **78**
4.1 Challenges in passive concealed threat detection 79
4.1.1 Emission & absorption of submillimeter waves 79
4.1.2 Inherent limitations in spatial resolution and *SNR* 80
4.2 The passive standoff security imager 80
4.2.1 Detection and readout scheme 80
4.2.2 Pre-processing steps 82
4.3 *SNR* enhancement via computation 85
4.3.1 Decreasing noise: denoising 86
4.3.2 Improving signal: deconvolution 88
4.3.3 Data-driven image reconstruction 92
4.4 Summary and discussion 95

5 Conclusion **97**

A Supplemental material **100**

B List of publications **139**

References **140**

List of Figures

2.1 Graphical depiction of a typical $4f$ imaging system 4
2.2 1D and 2D visualization of PSF & OTF 5
2.3 Simulation of imaging with different amount of detected light . . . 10
2.4 Poisson distribution and influence of out-of-focus blur 12
2.5 Visibility of spokes target and effective resolution limit 16
2.6 Collection efficiency, rel. *SNR* and image reconstruction 22

3.1 Jablonksi diagram and fluorescence spectrum of *DAPI* 29
3.2 Traditional imaging and *splitting & recombination* approach 31
3.3 Splitting via polarization: effects on image information, modulation and different recombination approaches 33
3.4 Weighting, noise-normalized OTF and spectral improvement factor 34
3.5 Multiview Richardson-Lucy deconvolution algorithm 37
3.6 Fluorescence emission spectra including log-normal fit 41
3.7 Comparison of peak emission vs broadened PSF 45
3.8 Broadening effect depending on emission spectrum and filter set . 46
3.9 Visibility and resolution in case of imaging a spokes target 47
3.10 Effective OTF of wavelength splitting and experimental setup . . . 49
3.11 Schematic and fluorescence emission of the Argolight sample . . . 51
3.12 PSF, OTF, axial response, weights distribution and improvement factor for pupil splitting 53
3.13 Optical setup used for pupil splitting 57
3.14 Imaging "resolution" sample using pupil splitting 58
3.15 Imaging "stair" sample using pupil splitting 60
3.16 Concept of Field-Synthesis light-sheet generation 63
3.17 Light-sheet thickness vs FoV and axial modulation introduced using Field-Synthesis to generate Bessel sheet 65
3.18 nNMF reconstruction and 3D image simulation 69
3.19 Experimental setup and results of illumination splitting 71
3.20 Experimental and deconvolved data on illumination splitting . . . 73

4.1 Concealed threat detection by submillimeter wave imaging 81

4.2 Pre-processing of submillimeter wave data 83
4.3 Image denoising of submillimeter data 87
4.4 Image deconvolution of submillimeter data 89
4.5 Deep-Learning-based denoising: principle and results 93

A.1 Effective resolution capability for two separated point sources . . . 103
A.2 Example showing the influence of localized noise on resolution . . 104
A.3 Schematic of calculating the scalar WF-OTF 105
A.4 *NCC* curves to optimize image reconstruction 112
A.5 Polarization depending focusing of light 113
A.6 Wavelength splitting & recombination result of line pattern 122
A.7 Imaging "stair" sample using wavelength splitting 124
A.8 *NCC* curves for wavelength and pupil splitting 125
A.9 Graphical autocorrelation & improvement factor for pupil splitting 126
A.10 Imaging results pupil splitting & recombination 129
A.11 The aplanatic factor modifying the pupil split 130
A.12 Schematic relating Field-Synthesis to scanned light-sheet generation 132
A.13 Rejecting out-of-focus light, through wavelenght splitting 133
A.14 Experiment results of imaging using illumination splitting 134
A.15 Mean-variance plot of submillimeter data 135
A.16 Temporal overview of recorded submillimeter wave images 137
A.17 Going from supervised to unsupervised learning-based denoising . 138

List of Tables

3.1 Visibility improvement factor for polarization splitting 39
3.2 Log-normal distribution fit of fluorescence emission spectra 41
3.3 Comparing theoretical and numerical broadening effect 45
3.4 *FWHM* of the broadened and the split & recombined PSF 49
3.5 Performance measures for pupil splitting 55
3.6 Visibility improvement factor for pupil splitting 59

A.1 Visibility improvement factor for wavelength splitting 123
A.2 Improved spatial resolution in submillimeter wave imaging 136

List of Abbreviations

SNR	Signal-to-noise ratio
WF	Widefield
BFP	Back-focal-plane
FoV	Field-of-View
PSF	Point-spread-function
NA	Numerical aperture
FWHM	Full-width-at-half-maximum
OTF	Optical-transfer-function
OS	Optical sectioning
SIM	Structured illumination microscopy
ISM	Image scanning microscopy
LSM	Laser scanning microscopy
SMLM	Single molecule localization microscopy
RL	Richardson-Lucy deconvolution
NCC	Normalized cross-correlation
ROI	Region-of-interest
DoF	Depth of focus
FS	Field-Synthesis
EDoF	Extended-depth-of-field
TES	Transition edge detector
SQUID	Superconducting-quantum-interference-device
FFT	Fast Fourier transform
DCT	Discrete cosine transform
nNMF	Non-negative matrix factorization
GPU	Graphical processing unit
CNN	Convolutional neural network

Chapter 1

Introduction

Imaging technologies have become very prominent in our everyday life. The number of mobile devices which include one or multiple cameras is steadily rising, leading to ever more digital images being taken [1]. Also in highly specialized fields such as healthcare [2], biomedical research [3, 4] or aerial observations [5], imaging plays an important role. An imaging system needs to be designed and constructed in such a way that an image of high quality forms on a detector [6]. Depending on the wavelength used, elements such as mirrors or lenses transfer information of the object under investigation (= sample) into image space [7]. There exist certain limitations in terms of spatial resolution [8] and sampling [9, 10], stemming from physical and mathematical principles. Understanding those led to advancements in imaging such as super-resolution techniques [11] or more efficient sampling strategies [12, 13]. Furthermore, a general understanding of the basic image formation principles has resulted in the invention of more powerful image restoration algorithms [14, 15, 16]. However, the performance of imaging systems in general (including the subsequent processing) is strongly dependent on the data quality [17, 18]. Uncertainties (= noise) in the measurement process lead to deviations of the acquired data from the expected (unknown) noisefree image. A simple measure of image quality is the signal-to-noise ratio (*SNR*) [19]:

$$SNR = \frac{Signal}{Noise} \tag{1.1}$$

It expresses a ratio of how much the measurement is corrupted due to unwanted noise. A minimum criterion to obtain useful information is given as: $SNR \geq 1$. *Noise* in imaging typically consists of various possible contributions. Many applications show detector noise [20], hence technological advances have been made to minimize those [21, 22]. However, shot noise as a fundamental part of measuring light, cannot be removed that easily. It originates from counting distinct particles (photons) and is inherit to light itself. In fact the shot noise limit represents a major barrier, which once overcome, can lead to breakthroughs in measurement sensitivity (e.g. gravitational wave detection [23]). Nevertheless, most techniques which have achieved this are often based on interferometry [24] or specialized quantum states of light ("squeezed light" [25]) and do not directly fit into the traditional imaging landscape. Goal of this thesis is to show how the shot noise limit can be overcome by making modest changes to conventional imaging setups as commonly found in modern biomedical research [26]. Leading to the concept of optical splitting and computational recombination (*splitting & recombination*) and its ability to improve *SNR* without the need to capture more photons. Another aspect of this work is to correct and enhance the image quality of detectors which work at the shot noise limit (i.e. bolometric detection [27] in submillimeter wave imaging) via computational approaches. Here an enhancement in terms of *SNR* and resolution will directly translate into an improved usability of such imaging devices in their respective field of application (concealed threat detection). The remainder of this thesis is structured as follows. In chapter 2 concepts of image formation and the respective influence of noise are being introduced. The definition of *SNR* in real and Fourier space, as well as different measures to quantify the quality of signal transfer and the influence of sampling are presented. Typical ways to improve the *SNR* in conventional imaging are shown, as well as their limitations and drawbacks demonstrated. In the following (chapter 3), the idea of *splitting & recombination* is introduced and applied in the field of fluorescence microscopy. Four different splitting examples are put forward: 1) separating the polarization of detected light; 2) splitting the fluorescence emission; 3) dividing the pupil of an incoherent imaging system into multiple parts and 4) splitting the illumination in light sheet microscopy through Field-Synthesis. The last three are presented in more theoretical detail, together with proof-of-principle experiments, verifying that it is possible to improve the *SNR*, given a finite photon budget. Chapter 4 focuses on the passive detection of concealed threats using submillimeter wave radiation. First, the typical characteristics of submillimeter wave radiation and its sensitive detection are discussed. Followed by a description of the necessary pre-processing steps and the application of different techniques, such as denoising and deconvolution, to computationally enhance the *SNR*. The final part of this thesis (chapter 5) comprises of a short summary, conclusion and outlook with respect to the findings this thesis has provided. This is followed by an appendix which contains additional information and derivations which are linked to throughout the main text.

Chapter 2

Theoretical background

In this chapter the two terms of equation 1.1 are subsequently being introduced: signal and noise. Both will be evaluated in real and Fourier space respectively and together yield a more detailed description of the *SNR*. Properties such as contrast, visibility, detectability and the effective resolution limit are then explained within this context. The influence of sampling is briefly discussed and followed by methods that are typically used to improve *SNR*: collecting more photons or using computational reconstruction techniques. The chapter ends with the notion that enhancing the *SNR*, without requiring to capture more photons, will bring great benefits to many application of modern imaging.

2.1 Signal in incoherent imaging

The following theoretical analysis is made under the assumption of a conventional imaging setup, here a microscope. Figure 2.1 shows a $4f$ detection system in conjunction with a single lens depicting illumination in transmission. The three lens system (illuminating, objective and tube lens) are placed in series such that the overall distance from light source to detector equals two times the sum of all focal lengths (indicted by f) [28].

A sample S is illuminated with a light distribution I, shown here as a plane wave for bright- or fluorescence widefield (WF) imaging. The detection objective captures light coming from the sample (blue) and, in combination with the tube lens, forms an (expected) image E on the pixelated detector. Here an infinity-corrected optical setup is assumed, such that a point source gets transformed into a plane

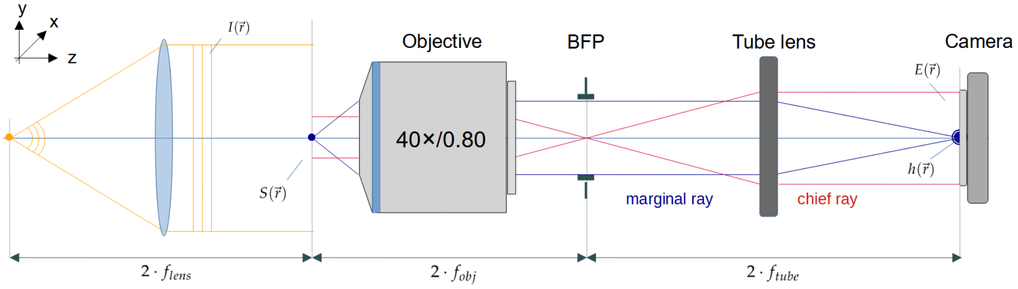


Figure 2.1: Graphical depiction of a typical $4f$ imaging system [28], such as in an infinity-corrected microscope. The illumination I of the sample S is shown in transmission, e.g. bright-field. Light carrying information about the sample is captured by an objective (blue *marginal* rays), which in combination with the appropriate tube lens results in an (expected) image E on the pixelated detector of a camera. The field-of-view (FoV) of the system is shown using red rays (*chief* rays) and indicate the magnification (here 40×). The location of the back-focal-plane (BFP) is given by the focus of a parallel plane wave (red) entering the objective.

wave in the back-focal-plane (BFP) of the detection objective [29]. The field-of-view (FoV) of the system is indicated with the red rays and shows the magnification of the image content onto the camera. A mathematical description of the image formation is given in the next section.

2.1.1 Image formation process

In this thesis the transfer of object information in a fully *incoherent* imaging system (e.g. fluorescence microscopy) is investigated. A mathematical formulation of the imaging process is given as the following *convolution* (⊗) operation [30]:

$$E(\vec{r}) = [S(\vec{r}) \cdot I(\vec{r})] \otimes h(\vec{r}) \tag{2.1}$$

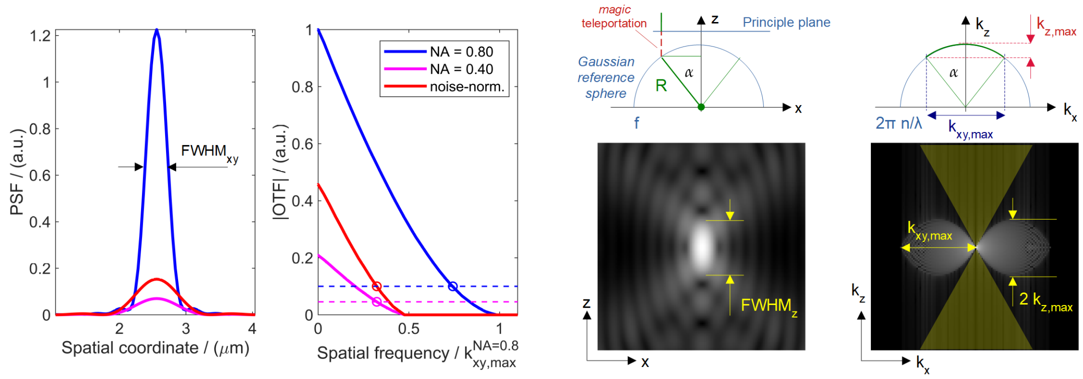
with h being the detection *point-spread function* (PSF) and $\vec{r}$ spatial coordinates. Each point in the sample S is being imaged into a broader spot (the PSF), which leads to an overall blurring of spatial information in the final image. The convolution operation itself is defined as [31]:

$$E(\vec{r}) = \int_{-\infty}^{+\infty} d\vec{r}' \;\; [S(\vec{r}') \cdot I(\vec{r}')] \cdot h(\vec{r} - \vec{r}') \tag{2.2}$$

The presented image formation theory is introduced in this thesis in a scalar description. Nevertheless, the depicted concepts can also be applied to vectorial calculations where they hold for the three vector components separately (see [32]). Using a detection objective with larger *numerical aperture NA*, yields in a reduction of the spatial blurring in the recorded image. The NA is given as the product of refractive index of the immersion medium n and the half-opening angle α of the detectable light cone:

$$NA = n \cdot \sin\alpha \tag{2.3}$$

Figure 2.2a left shows the PSF for an air immersion objective with two different NAs ($NA = 0.80$ in blue, $NA = 0.40$ in magenta; assuming equal magnification m). The distribution of h has been calculated using a vectorial model [32] with an emission wavelength $\lambda = 520$ nm. The PSF width is indicated in terms of the full-width-half-maximum ($FWHM$) of the central peak.



a b

Figure 2.2: 1D and 2D visualization of PSF & OTF using two detection objectives ($NA = 0.80$ in blue and $NA = 0.40$ in magenta), calculated using the vectorial model from [32] with $\lambda = 520$ nm. a Profile plots along the x (left) and k_x (right) axis of the PSF and $|\tilde{h}|$ respectively. Left: the lateral full-width-half-maximum is indicated $FWHM_{xy}$; reducing the NA does not only increase the width of the PSF, it also decreases the peak intensity as the number of detectable photons is smaller for the lower NA (also see Fig. 2.6a). The same can be observed in the magnitude of the OTF (right), where additionally the noise floor (100 expected photons for the $NA = 0.80$ case) is plotted (dotted lines). For a fair comparison also a *noise* normalized OTF (red) curve is depicted. Details on noise normalization is given in sec. 2.3.4. b View across the xz and k_xk_z plane, each with an adjusted gamma value of 0.25. Left: PSF along optical axis, the axial $FWHM_z$ is indicated. On top: the Gaussian reference sphere of the detection objective, the "magic teleportation" is seen as the red dotted line. Right: $|\tilde{h}|$ indicating the "missing cone" problem of normal WF detection and the finite region of support, given as the cutoff frequencies $k_{xy,max}$ and $k_{z,max}$.

A reduction of the detection NA not only leads to a broader PSF, also the number of detected photons is reduced (also see Fig. 2.6a). Leading to a much lower PSF peak value in the $NA = 0.40$ case (more to this in sec. 2.4.1).

2.1.2 Representation in Fourier space

Instead of describing image formation in real space, we can make use of Fourier analysis. For this we calculate the Fourier transform $\mathcal{F}_{dim}$ [31] of our image M:

$$\tilde{E}(\vec{k}) = \mathcal{F}_{dim}\left\{E(\vec{r})\right\} = \frac{1}{(2\pi)^{dim/2}} \int_{-\infty}^{+\infty} d\vec{r}\; E(\vec{r}) \cdot e^{i\vec{k}\cdot\vec{r}} \tag{2.4}$$

with i being the imaginary unit, $\vec{k}$ spatial frequencies, $\vec{k} \cdot \vec{r}$ a scalar product and dim the number of spatial dimensions of E (typically $dim = 2D \rightarrow \mathbb{R}^2$).
Due to the *convolution theorem* of Fourier transforms [31] we can rewrite eq. 2.1:

$$\tilde{E}(\vec{k}) = \left[\tilde{S}(\vec{k}) \otimes \tilde{I}(\vec{k})\right] \cdot \tilde{h}(\vec{k}) \tag{2.5}$$

Note that multiplication and convolution have switched in eq. 2.5 and that all involved quantities are now complex valued: $\tilde{E} \in \mathbb{C}$.
From this point on the tilde symbol indicates Fourier space and going back to real space can be achieved by using the inverse Fourier transform $\mathcal{F}^{-1}_{dim}$ [31]:

$$E(\vec{r}) = \mathcal{F}^{-1}_{dim}\left\{\tilde{E}(\vec{k})\right\} = \frac{1}{(2\pi)^{dim/2}} \int_{-\infty}^{+\infty} d\vec{k}\ \tilde{E}(\vec{k}) \cdot e^{-i\vec{k}\cdot\vec{r}} \tag{2.6}$$

Equation 2.5 introduces $\tilde{h}$ which is the *optical transfer function* (OTF) of the imaging system. It acts as a low-pass filter and limits the transfer of high spatial frequency information up to a certain cutoff. The corresponding cutoff frequency $k_{xy,max}$ is indirectly given by Abbe's diffraction limit [8], which states that there exists a periodic line structure with distance $d_{xy,min}$, that is too fine to be resolvable by an optical imaging system. For fluorescence imaging this minimal distance depends on the detection wavelength λ and the numerical aperture NA:

$$d_{xy,min} = \frac{\lambda}{2 \cdot NA} \tag{2.7}$$

Hence the cutoff frequency in lateral dimensions is given as:

$$k_{xy,max} = \frac{2\pi}{d_{xy,min}} = 2\pi \cdot \frac{2 \cdot NA}{\lambda} \tag{2.8}$$

Figure 2.2a right shows the magnitude of the OTF for the two different objectives ($NA = 0.80$ in blue & $NA = 0.40$ in magenta; for equal m and f_{obj}). It can be seen that reducing the NA narrows the region-of-support, which is defined as: $|\tilde{h}| > 0$.
Note that the $\vec{k} = \vec{0}$ (zero frequency) component of the OTFs are proportional to the amount of detectable photons from a point source, as:

$$\tilde{h}(\vec{0}) = \frac{1}{(2\pi)^{dim/2}} \int_{-\infty}^{+\infty} d\vec{r}\ h(\vec{r}) \cdot e^{i\,\vec{0}\cdot\vec{r}} = \frac{1}{(2\pi)^{dim/2}} \int_{-\infty}^{+\infty} d\vec{r}\ h(\vec{r}) \tag{2.9}$$

This is true in general: $\tilde{M}(\vec{0})$ is proportional to the number of collected photons in the image E. Figure 2.2a right shows both OTFs normalized to the $NA = 0.80$ case. The corresponding PSF is depicted on the left. Reducing the NA by a factor of two not only decreases $k_{xy,max}$ by the same amount, it additionally reduces the number of photons reaching the detector by approximately 80%, hence the $\tilde{h}(\vec{0})$ value (this differs from a reduction given by the ratio of the two NAs and is explained in sec. 2.4.1 in more detail).

2.1.3 Imaging in three dimensions

Up until now the depiction of PSF & OTF has only been as profile plots of a 2D distribution. This is a simplification as both essentially are three dimensional in structure. Figure 2.2b left shows a view of h in the xz-plane. The distribution has been calculated by propagating the electric field $\mathcal{E}$ of the corresponding intensity PSF h over varying distances Δz in the positive and negative direction along the optical axis z. The propagation itself is computed using the following Fourier transform propagator [30]:

$$\mathcal{E}(x, y, z = \Delta z) = \mathcal{F}_{2D}^{-1}\left\{\mathcal{F}_{2D}\{\mathcal{E}(x, y, z = 0)\} \cdot e^{i\, k_z(k_x, k_y) \cdot \Delta z}\right\} \tag{2.10}$$

This is an application of the *shifting* property of Fourier transforms [31] along the optical axis with (x, y, z) representing Cartesian coordinates in real space. The expression for $k_z(k_x, k_y)$ can be found by looking at the *Ewald* sphere [33], depicted in Fig. 2.2b right/top. The Ewald sphere is a compact representation of all those plane waves, which are emitted by a point source and are still detectable in the far field. However, only a finite range of those waves can be captured by an objective, which shrinks the region of detectable waves to a 3D spherical cap (green). Those plane waves are of same periodicity but travel under different angles. The maximum angle which can be collected is given by the NA. The lateral and axial extend of the cap are equal to $k_{xy,max}$ (blue) and $k_{z,max}$ (red) respectively. The latter is similarly defined as the cutoff frequency in the lateral direction and given here in terms of the NA [34]:

$$k_{z,max} = \frac{2\pi}{d_{z,min}} = \frac{2\pi}{\lambda} \cdot \left(n - \sqrt{n^2 - NA^2}\right) \tag{2.11}$$

with $d_{z,min}$ the smallest structure resolvable along the optical axis.
Making use of the dispersion relation in homogeneous media [7], which states that all plane waves reside on the shell of a sphere, gives:

$$k_z(k_x, k_y) = \sqrt{\left(2\pi \cdot \frac{n}{\lambda}\right)^2 - \left(k_x^2 + k_y^2\right)} \tag{2.12}$$

Intensity information is obtained by computing the absolute square of the electric field at $z = \Delta z$, yielding in total:

$$\left|\mathcal{E}(x, y, z = \Delta z)\right|^2 = \left|\mathcal{F}_{2D}^{-1}\left\{\mathcal{F}_{2D}\{\mathcal{E}(x, y, z = 0)\} \cdot e^{i\sqrt{\left(2\pi \cdot \frac{n}{\lambda}\right)^2 - \left(k_x^2 + k_y^2\right)} \cdot \Delta z}\right\}\right|^2 \tag{2.13}$$

The result for $\mathcal{E}$, corresponding to the electric field of h, can be seen in Fig. 2.2b left, with the *FWHM* in axial direction indicated in yellow. It is not discussed in detail at this point but conventional WF detection lacks the ability of *optical sectioning* (OS), as the laterally integrated PSF intensity does not change along z [28]. A different way to see the lack of OS is by looking at the *missing cone* problem in the 3D-OTF [30], which has been calculated as the 3D Fourier transform of the propagated PSF field and is shown in Fig. 2.2b right ($k_x k_z$-plane). The missing cone is indicated in yellow, showing that is not possible to discern any information originating from different planes perpendicular to the optical axis, as those correspond to a line along the k_z-axis in Fourier space.

It is now important to make a slight modification to the imaging equation 2.1 as we also want to incorporate the acquisition of 3D information in form of a z-stack (2D images from different axial positions in the sample). This is denoted by a subscript, e.g. E_{3D} stands for the z-stack and E_{2D} for a single image. Hence the image formation for the z-stack is written as:

$$E_{3D}(\vec{r}) = [S_{3D}(\vec{r}) \cdot I_{3D}(\vec{r})] \otimes h_{3D}(\vec{r}) \tag{2.14}$$

where $\otimes$ denotes a three dimensional convolution.

A single slice from that z-stack is given by setting $z = 0$ (the position of the detector), which yields:

$$E_{2D}(\vec{r}) = E_{3D}([x, y, z = 0]^\top) \tag{2.15}$$

with $^\top$ being the transpose of a vector or matrix.

The *projection-slice* theorem of Fourier transforms [31] connects a slice in real space to a projection along the axis perpendicular to the given slicing operation in Fourier space. Meaning that the 2D image information of three dimensional imaging process is given in Fourier space according to:

$$\tilde{E}_{2D}(k_x, k_y) = \int_{-\infty}^{+\infty} dk_z \, \left[\tilde{S}_{3D}(\vec{k}) \otimes \tilde{I}_{3D}(\vec{k})\right] \cdot \tilde{h}_{3D}(\vec{k}) \tag{2.16}$$

with $\vec{k} = [k_x, k_y, k_z]^\top$ being Cartesian coordinates in k-space. To obtain the measurement data in real space an inverse 2D Fourier transforms needs to be applied:

$$E_{2D}(\vec{r}) = \mathcal{F}_{2D}^{-1}\left\{\tilde{E}_{2D}(\vec{k})\right\} \tag{2.17}$$

In comparison to eq. 2.1, this three dimensional treatment also includes effects such as out-of-focus blur or the missing optical sectioning, which are common in widefield detection.

Note that from now on we will omit the subscript $2D$ for the sample, illumination and PSF/OTF variables, and only keep it for the image quantity E. This will help with the ease of reading, while the important information whether 2D or 3D data is being recorded is not lost.

2.2 Noise in imaging applications

As mentioned in sec. 1, noise is present in any measuring device. In imaging it mostly limits the resolution of fine structures [35], which is the primary goal of optical microscopy. Therefore it is important to understand its influence on image formation, which will later be used to define a qualitative criterion: the *SNR*.

2.2.1 Degradation of image quality

Incorporating the effects of noise into the image formation requires to make a distinction between the *noisy* measurement M_{2D} and the *expected* signal E_{2D}:

$$M_{2D}(\vec{r}) = E_{2D}(\vec{r}) + \mathcal{N}_{2D}(\vec{r}) \tag{2.18}$$

with $\mathcal{N}_{2D}$ being the noise of the respective measurement. Note that M_{2D} and $\mathcal{N}_{2D}$ represent realizations of a random process, meaning that they vary for each observation. The expectation value E_{2D} is defined according to [31] :

$$E_{2D}(\vec{r}) = \int_O dO \;\; O(\vec{r}) \cdot P\left[O(\vec{r})\right] \tag{2.19}$$

with $P[O]$ being the probability density of observation O. The spread of the actual measurements around the expectation value is given as the standard deviation $\sigma_{M_{2D}}$. Which can be calculated as the square root of the variance:

$$\sigma^2_{M_{2D}}(\vec{r}) = \int_O dO \;\; \left[O(\vec{r}) - E_{2D}(\vec{r})\right]^2 \cdot P\left[O(\vec{r})\right] \tag{2.20}$$

In noiseless imaging $\mathcal{N}_{2D} \to 0$ and hence each measurement represents the expectation value: $M_{2D} = E_{2D}$. Due to the linearity of the Fourier transform operation, equation 2.18 can readily be translated into $\vec{k}$-space:

$$\tilde{M}_{2D}(\vec{k}) = \tilde{E}_{2D}(\vec{k}) + \tilde{\mathcal{N}}_{2D}(\vec{k}) \tag{2.21}$$

Again showing the separation of image information into expectation value and variance. However, this requires the definition of expectancy and variance for complex variables, which is given in the appendix (p. 102).

Depending on the amount of noise, the measured image can deviate strongly from its expected values. This degradation of image quality can be seen in Fig. 2.3, simulated in the photon-limited regime (details to this in sec. 2.2.2).

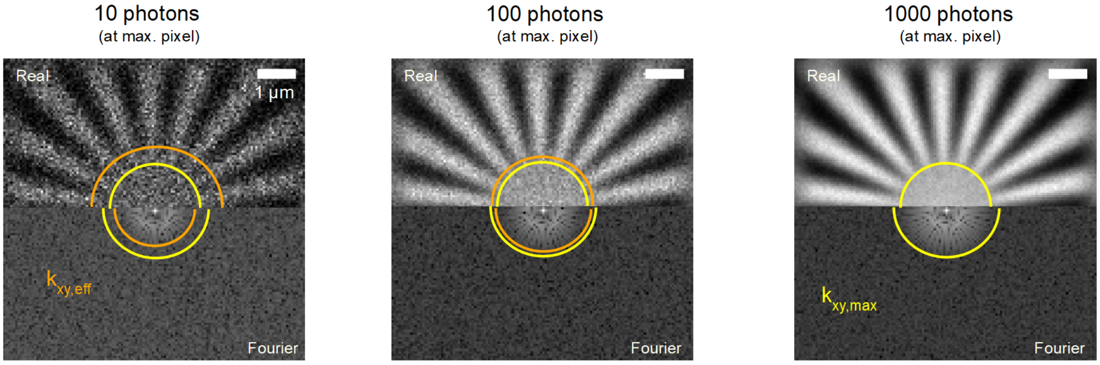


Figure 2.3: Simulation of imaging a spokes-target (18 spokes) with different amount of detected light (10, 100 or 1000 expected photons in the pixel with the maximum value) using a vectorial model [32] with $\lambda = 520$ nm and corresponding shot noise. Top: real space image (scale bar $= 1\ \mu$m); bottom: corresponding Fourier transform (displayed using γ-adjustment of $\gamma = 0.25$). The yellow semicircle indicates the fundamental cut-off frequency $k_{xy,max}$ or the corresponding radius with sufficient contrast in the real space data and is the same for all three sub-images. The effective limit $k_{xy,eff}$ (orange, introduced in sec. 2.3.3) depends on the amount of noise and can strongly deviate from the noiseless case. The loss of high spatial frequency information can be seen by the rising noise floor, reducing the effective region-of-support of the imaging system.

The top half shows the image of a spokes target (18 spokes, scale bar $= 1\ \mu$m) in real, the bottom half the corresponding information in Fourier space. Noticeably, the image quality gets better when the total number of detected photons (here given as number of photons in the brightest pixel) is increased. A more detailed explanation of this effect is given in the following section.

2.2.2 Photon or *shot* noise

As has already been discussed in sec. 1, images are fundamentally corrupted by *shot* noise. This is typically modeled using the *Poisson* distribution [36, 37]:

$$P\left[M_{2D}(\vec{r})\right] = \frac{E_{2D}(\vec{r})^{M_{2D}(\vec{r})}}{M_{2D}(\vec{r})!} \cdot e^{-E_{2D}(\vec{r})} \tag{2.22}$$

with ! denoting the factorial. Note that M_{2D} is now measured in photon counts which is a discrete quantity. Equation 2.22 gives the probability of measuring an image M_{2D}, when the underlying ground truth is given by E_{2D}. The fact that the actual measurement is corrupted by Poisson noise can be written as:

$$M_{2D}(\vec{r}) = \text{Poisson}\left\{E_{2D}(\vec{r})\right\} \tag{2.23}$$

An important property of the Poisson distribution is the equality of expectancy and variance [35], given as:

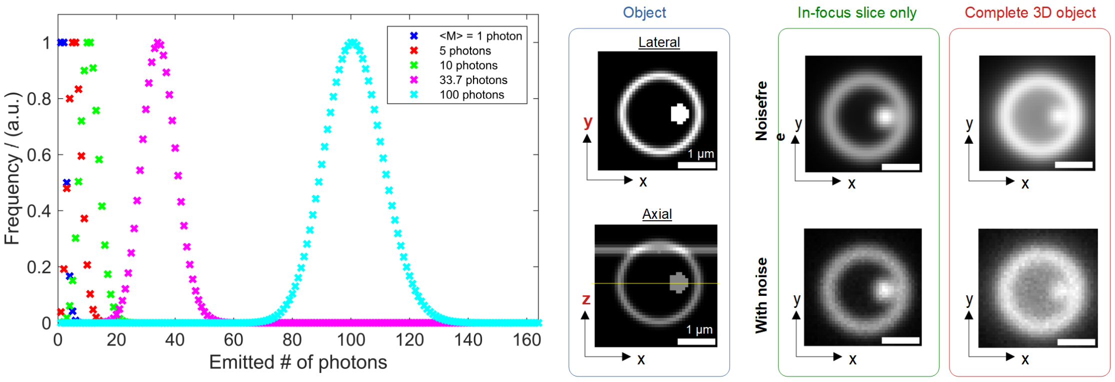
$$\sigma^2_{M_{2D}}(\vec{r}) = E_{2D}(\vec{r}) \tag{2.24}$$

Meaning that the Poisson distribution is only defined by a single parameter: the expectation value E_{2D}. Figure 2.4a shows the corresponding probability curves for five different expectation values (1, 5, 10, 33.7 and 100 expected photons in the pixel with maximum value; note that E_{2D} does not need to be integer).

The individual distributions have been normalized so that their maxima are set to 1.0. This was done so that the width of all curves could be compared more easily. It can be clearly seen that the width of the distribution increases for larger expectation values, which means that the peak value gets reduced. This is because the integral under each curve must yield unity (probability of all possible outcomes must be 100%). The typical asymmetric shape of the Poisson distribution can be observed for low values of E_{2D} (blue and red). For higher values it turns into more a Gaussian shape, which is modeled according to [37]:

$$P\left[M_{2D}(\vec{r})\right] = \frac{1}{\sqrt{2\pi \cdot E_{2D}(\vec{r})}} \cdot e^{-\frac{\left[M_{2D}(\vec{r}) - E_{2D}(\vec{r})\right]^2}{2 \cdot E_{2D}(\vec{r})}} \tag{2.25}$$

where the standard deviation of the Gaussian has been set equal to the respective expectation value. Again, the amount of noise in a specific measurement scales with the expected outcome of that very measurement.



a b

Figure 2.4: a *Poisson* distribution for five different expectation values: 1, 5, 10, 33.7 and 100 photons in the pixel with maximum value. For low photon numbers (blue and red) the typical asymmetric characteristic of the Poisson distribution can be observed. When the amount of detectable light increases, the distribution changes into a Gaussian shape, with the width proportional to the respective expectancy. Note that each maximum has been normalized to unity. b Simulated example showing the detrimental effect of out-of-focus blur. A 3D object (spherical shell, including a smaller sphere and a line parallel to x-axis) is shown on the very left (scale bar = 1 μm). Imaging (vectorial PSF [32] with $\lambda = 520$ nm and $NA = 0.80$.) has been simulated with (below) and without (top) shot noise (1000 photons in max pixel). The middle column shows the result in *thin-element* approximation, meaning that only the in-focus object slice has been used in the imaging simulation. On the right WF detection is depicted, showing a much degraded image quality, especially due to the additional contribution of out-of-focus blur.

The variance of $\tilde{M}$ (given by sum of real and imaginary variances) can be computed, in case of photon-limited imaging, according to:

$$\begin{aligned}
\sigma^2_{\tilde{M}_{2D}}(\vec{k}) &= \mathrm{Var}\left\{\tilde{M}_{2D}(\vec{r})\right\} = \mathrm{Var}\left\{\frac{1}{2\pi}\int_{-\infty}^{+\infty} d\vec{r}\ M_{2D}(\vec{r})\cdot e^{i\vec{k}\cdot\vec{r}}\right\} = \\
&= \frac{1}{2\pi}\int_{-\infty}^{+\infty} d\vec{r}\ \mathrm{Var}\left\{M_{2D}(\vec{r})\cdot e^{i\vec{k}\cdot\vec{r}}\right\} = \\
&= \frac{1}{2\pi}\int_{-\infty}^{+\infty} d\vec{r}\ \left|e^{i\vec{k}\cdot\vec{r}}\right|^2 \mathrm{Var}\left\{M_{2D}(\vec{r})\right\} = \frac{1}{2\pi}\int_{-\infty}^{+\infty} d\vec{r}\ \sigma^2_{M_{2D}}(\vec{r}) \quad (2.26)
\end{aligned}$$

with the additional assumption that M is statistically independent from pixel to pixel. Equation 2.24 states that the variance in Fourier space is proportional to number of expected photons in real space:

$$\sigma^2_{\tilde{M}_{2D}} \propto \int_{-\infty}^{+\infty} d\vec{r}\ E_{2D}(\vec{r}) = p \qquad (2.27)$$

Yielding that $\sigma^2_{\tilde{M}_{2D}}$ is proportional to the total sum of expected photons p in the image. Additionally note that the variance is independent of spatial frequencies, hence noise in Fourier space will emerge as a noise floor (in Fig. 2.2a, dotted horizontal lines). The level of this floor is given by the standard deviation $\sigma_{\tilde{M}_{2D}} \propto$

$\sqrt{p}$, hence scales with the square root of the expected photons p, while the signal is linearly depending on p. Resulting in an improved SNR, when more photons are available for imaging.

2.2.3 Out-of-focus: background noise

Shot noise is fundamental to nature and cannot be omitted in traditional imaging setups. Together with out-of-focus blur it leads to background noise which can be reduced in principle, but nevertheless plays an important role in all applications of microscopic imaging. In WF detection, unwanted light originating from different planes than the focal plane corrupts the 2D image. This effect is termed out-of-focus blur and has a negative influence on the obtainable image quality, which is exemplified in Fig. 2.4b. The imaging of a 3D object has been simulated, by directly applying equation 2.16 (vectorial PSF [32] with $\lambda = 520$ nm and $NA = 0.80$). Figure 2.4b left shows the true object, a spherical shell, with an embedded smaller sphere and a line parallel to the x-axis (scale bar = 1 μm). When simulating the imaging in a *thin-element* approximation (green, only imaging the in-focus slice of the object using eq. 2.5), the typical lateral blurring of image information (middle column; top) can be observed. Also the effect of shot noise is prominent (bottom, 1000 expected photons in the pixel with the maximum value) when it is applied.

A more realistic simulation (right, red) also incorporates out-of-focus effects, as suggested from eq. 2.16. Background blur, which originates from parts of the 3D shell outside of the focus, can be observed in the simulated 2D image. When Poisson noise is applied the outcome looks even more degraded, much worse than the noisy thin-element result. Note that this additional blurring does not originate from insufficient spatial resolution. It merely comes from the poor performance of widefield detection to reject out-of-focus light. Most popular microscope systems today, such as confocal [38, 39] or light-sheet (LS) microscopy [40, 41] show a major improvement in terms of image quality, which mainly originates from their ability to remove out-of-focus blur & noise.

2.3 The signal-to-noise ratio (*SNR*)

A criterion to determine the quality of transferring information from sample into image space, is the signal-to-noise ratio (SNR). In the following an analytical expression in real and Fourier space is derived. The conventional way to describe the *SNR* of a random variable is:

$$SNR = \frac{\text{expectation value}}{\sqrt{\text{variance}}} \tag{2.28}$$

Note that a variance $\neq 0$ has to be assumed. In a realistic scenario this is given, as optical imaging itself demands a photon number > 0, which together with shot noise always yields a variance being strictly positive.

2.3.1 *SNR* in real space

With the general definition from above, the *SNR* in a typical imaging application can be formulated in real space as:

$$SNR_{M_{2D}}(\vec{r}) = \frac{E_{2D}(\vec{r})}{\sqrt{\sigma^2_{M_{2D}}(\vec{r})}} = \frac{E_{2D}(\vec{r})}{\sqrt{E_{2D}(\vec{r})}} = \sqrt{E_{2D}(\vec{r})} \tag{2.29}$$

Meaning that *doubling* the $SNR_{M_{2D}}$ requires to *quadruple* the photon number of the expected image information, which is typically done by increasing the sample illumination or the exposure time of the detector. Note that the real space *SNR* is signal dependent: regions with no object information will not exhibit any noise contribution. Keep in mind that this means that even a perfect imaging system can only realize $SNR_{M_{2D}} = 0$ when there is no sample present [42]. A major drawback of the real space definition of the *SNR* is that a finer spatial sampling will reduce $SNR_{M_{2D}}$ although more information on the actually measured photon positions is available (more details on sampling in sec. 2.3.5).

2.3.2 In terms of spatial frequencies

A definition that is independent of the spatial sampling can be realized in Fourier space. The same procedure of calculating the *SNR* yields:

$$SNR_{\tilde{M}_{2D}}(\vec{k}) = \frac{|\tilde{E}_{2D}(\vec{k})|}{\sqrt{\sigma^2_{\tilde{M}_{2D}}(\vec{k})}} \propto \frac{|\tilde{E}_{2D}(\vec{k})|}{\sqrt{p}} \tag{2.30}$$

Note that the *SNR* is seen as a criterion determining the strength of information transfer, hence we only consider the *absolute* value of $\tilde{E}_{2D}$ in the nominator. The variance in Fourier space is proportional to the total number of photons p, but so is $|\tilde{E}_{2D}|$. This makes it more difficult to immediately see that the *SNR* improves with increasing photon number. Hence, a normalized image spectrum $\tilde{E}_{2D}^{norm.}$, independent of the detected photon number, is introduced:

$$|\tilde{E}_{2D}^{norm.}(\vec{k})| = \frac{|\tilde{E}_{2D}(\vec{k})|}{p} \tag{2.31}$$

With this the aforementioned *SNR* expression is rewritten as:

$$SNR_{\tilde{M}_{2D}}(\vec{k}) \propto \sqrt{p} \cdot |\tilde{E}_{2D}^{norm.}(\vec{k})| \tag{2.32}$$

Note that the *SNR* in Fourier spaces gives a more detailed insight how spatial frequencies are transferred in a noisy measurement. Each frequency is influenced equally by the noise floor. Reducing p by a factor of 4 will lower the noise floor for all $\vec{k}$ by 50%. However, this will affect the transfer of high spatial frequency information more strongly relative to the noise, than that of the low frequency components. Reason for this is the spatial blurring of image information, which removes high frequency content and mostly leaves low frequencies. Additionally many natural objects show a decay of their Fourier spectra that approximately follows a power law [43], meaning that high spatial frequencies are anyways underrepresented in the object under investigation. Hence, when working in the photon-limited regime, the noise floor affects high spatial frequencies more strongly, which are those that carry the most interesting information.

2.3.3 Contrast and effective resolution

The signal-to-noise ratio is a statistical measure, hence can only be used when a large number of observations have been acquired. When only a single image is to be evaluated in terms of *SNR*, often the image *contrast* is investigated. Which can be observed as a larger modulation depth or *visibility* $\mathcal{V}$ when a periodic object is being imaged. The visibility $\mathcal{V}$ is defined according to [30]:

$$\mathcal{V} = \frac{M_{max} - M_{min}}{M_{max} + M_{min}} \tag{2.33}$$

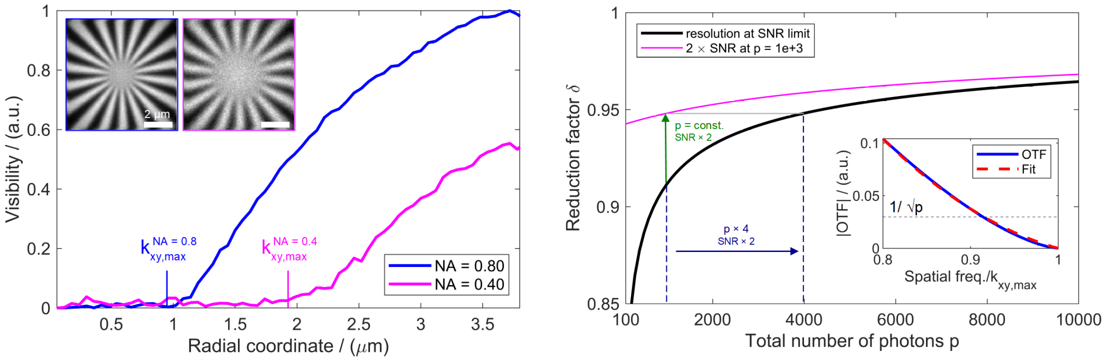
with $M_{max/min}$ corresponding to the maximum and minimum value of a fitted pattern of known periodicity in the image. Be aware that the visibility value is sensitive to an offset change, e.g. when the modulation curve is displaced by an offset ε, the following visibility $\mathcal{V}_\varepsilon$ is computed as:

$$\mathcal{V}_\varepsilon = \frac{(M_{max} + \varepsilon) - (M_{min} + \varepsilon)}{(M_{max} + \varepsilon) + (M_{min} + \varepsilon)} = \frac{M_{max} - M_{min}}{M_{max} + M_{min} + 2\varepsilon} \tag{2.34}$$

Hence, when trying to directly compare the modulation at an offset with the original curve, the following conversion needs to be applied:

$$\mathcal{V} \;=\; \mathcal{V}_\varepsilon \cdot \frac{M_{max} - M_{min}}{M_{max} - M_{min} - 2\varepsilon \cdot V_\varepsilon} \tag{2.35}$$

Figure 2.5a shows the visibility for a target with 18 spokes, imaged in simulation using two different detection objectives ($NA = 0.80$, $NA = 0.40$, same magnification m and refractive index n, 1000 photons in max. pixel of $NA = 0.8$ case).



a

b

Figure 2.5: a Visibility $\mathcal{V}$ when simulating imaging of a target with 18 spokes (scalebar = 2 μm) using two different numerical apertures ($NA = 0.80$; $NA = 0.40$, same magnification m and refractive index n) with 1000 expected photons in the pixel with the maximum value of the $NA = 0.80$ case. Decreasing the NA not only reduces the obtainable resolution limit $k_{xy,max}$, it also decreases $\mathcal{V}$. b Dependency of the reduction factor δ (defined in eq. 2.42) on the given number of expected photons p. To theoretically reach the Abbe limit a large number of photons is required, at 10.000 photons $\approx$ 96% is reached. Towards the lower photon numbers a drastic reduction of the achievable resolution is observable, as the SNR is $\propto \sqrt{p}$. Methods which can generate an SNR enhancement without requiring additional photons (such as *splitting & recombination*) can go beyond this limit (black). E.g. a method which enhances the SNR by a factor of two at $p = 1000$ photons will horizontally shift the black curve until its δ-value at $p = 4000$ photons is located at $p = 1000$. This type of SNR doubling is equivalent to capturing $4\times$ more photons, but without the drawbacks of actually having to collect all those photons. The inset shows $|\tilde{h}|$ (blue) close to the Abbe limit. Red dotted is a quadratic fit used to determine the effective resolution limit, by setting it equal to the noise floor $1/\sqrt{p}$ and solving for the abscissa (details in text).

Reducing the NA not only leads to a decrease in the attainable resolution (i.e. $k_{xy,max}$), it also results in an overall smaller SNR which presents itself as lower a visibility $\mathcal{V}$ value.

Another effect which can be used to visualize an SNR improvement is the fact that noise limits the ability to resolve fine structures. The presence of shot noise introduces an *effective* resolution limit, which in this work is defined in Fourier space as $k_{xy,eff}$. Object information can only be transferred if $|\tilde{E}_{2D}| \geq \sqrt{p}$, as then the signal does not get dominated by noise:

$$\begin{aligned} |\tilde{E}_{2D}(\vec{k})| &\geq \sqrt{p} \\ |\tilde{h}(\vec{k})| \cdot |\tilde{S}(\vec{k})| &\geq \sqrt{p} \end{aligned} \tag{2.36}$$

Indicating that the term *resolution* is always sample dependent! In the appendix (p. 102) the effective resolution limit is derived for imaging two spatially separated point sources. The expression for $k_{xy,\mathit{eff}}$ is given according to:

$$|\tilde{h}(k_{xy,\mathit{eff}})| = \tfrac{1}{\sqrt{p}} \tag{2.37}$$

Note that this is the same as looking for the intercept where the OTF reaches the noise floor defined by $1/\sqrt{p}$. To approximate this effect more quantitatively, $\tilde{h}$ is modeled using a scalar approximation, derived in the appendix (p. 105).

$$\tilde{h}(k_{xy}) \approx \frac{2}{\pi}\left[\arccos\left(\frac{k_{xy}}{k_{xy,max}}\right) - \frac{k_{xy}}{k_{xy,max}} \cdot \sqrt{1-\left(\frac{k_{xy}}{k_{xy,max}}\right)^2}\right] \tag{2.38}$$

with k_{xy} being radial spatial frequencies. From Fig. 2.2a right it can be seen that $\tilde{h}$ is slightly curved in the region $k_{xy} \geq 0.8 \cdot k_{xy,max}$. To obtain an approximation for $k_{xy,\mathit{eff}}$, we fit a quadratic curve (red dotted, Fig. 2.5b small inset) into this range and set the fitted equation equal to the noise floor, given at $1/\sqrt{p}$. The choice of $k_{xy} \geq 0.8 \cdot k_{xy,max}$ is rather arbitrary, nevertheless it corresponds to a noise floor when only 100 photons are available for imaging, which can be considered a lower practical limit.

For fitting the parabola three points are required: 1) $k_{xy,max}$ where $\tilde{h} = 0$; 2) noise floor for $p = 100$ photons and 3) an additional value in between. The parameters of the quadratic fit are derived in the appendix (p. 105) and yield:

$$\tilde{h}(k_{xy,\mathit{eff}} \geq 0.8 \cdot k_{xy,max}) \approx a \cdot \left(\tfrac{k_{xy}}{k_{xy,max}}\right)^2 + b \cdot \tfrac{k_{xy}}{k_{xy,max}} + c \tag{2.39}$$

with the three parameters:

$$a = 1.4507;\ b = -3.1314;\ c = 1.6807$$

This dependency can be used to find an effective resolution limit by setting above equation equal to $1/\sqrt{p}$ and solving for $k_{xy}/k_{xy,max}$:

$$a \cdot \left(\tfrac{k_{xy}}{k_{xy,max}}\right)^2 + b \cdot \tfrac{k_{xy}}{k_{xy,max}} + \left(c - \tfrac{1}{\sqrt{p}}\right) = 0 \tag{2.40}$$

Which yields a solution, termed the *reduction factor* $\delta(p)$.

$$\frac{k_{xy}}{k_{xy,max}} = \frac{-b-\sqrt{b^2-4a\left(c-\frac{1}{\sqrt{p}}\right)}}{2a} = \delta(p) \tag{2.41}$$

The *effective* resolution limit can now be defined by multiplying the reduction factor to the Abbe limit:

$$k_{xy,eff} = \delta(p) \cdot k_{xy,max} \tag{2.42}$$

Figure 2.5b depicts the reduction factor δ in dependency of the expected photon number p (black curve). To actually being able to reach the Abbe limit, a large number of photons needs to be captured, e.g. at 10000 photons only 96% of the Abbe limit can be observed. Towards lower number of photons a drastic drop in resolution capability is depicted, meaning that for small p the achievable resolution is dominated by noise rather than diffraction. Obtaining an enhancement in terms of $k_{xy,eff}$ requires to increase the number of detected photons or equivalently to improve the *SNR*, meaning that the performance is *SNR* limited. Any method which is able to increase the *SNR without* requiring to capture more photons (such as splitting & recombination) can achieve the same effective resolution, without the typical disadvantages of larger photon collection (such as a decrease in temporal resolution). E.g. doubling the *SNR* at a photon number of $p = 1000$ achieves the same effective resolution improvement as increasing p by a factor of four. In Fig. 2.5b this is shown as an horizontal shift of the black curve , leading to the magenta result which lies above the resolution associated to the *SNR* limit.

2.3.4 Noise normalized OTF and detectability

Directly comparing different imaging systems using OTFs might lead to false conclusions if a higher curve is simply assumed to correspond to a better imaging performance. This is because such curves do not truthfully represent the signal-to-noise behavior [44]. Lets consider the two OTFs (blue and magenta) from Fig. 2.2a right. Both the OTF curves and the noise floor scale with the total number of photon p, which is reduced in the $NA = 0.40$ case by a factor ≈ 0.21 (the deviation from the expected reduction, given by the ratio of the *NA*s, is given in sec. 2.4.1). Fig. 2.2a indicates that the imaging performance of $NA = 0.40$ is far worse. However, this is not a fair comparison as the noise floor of both imaging modalities differ. To achieve the same noise level for $\tilde{h}_{NA=0.40}$ (compared to $NA = 0.80$), the curve needs to be multiplied by a factor of $\approx 1/\sqrt{0.21} \approx 2.18$. Resulting in a noise floor at 0.1 which relates to the detection of $p = 100$ photons for the $NA = 0.80$

case. The noise-normalized result is shown in Fig. 2.2a right as the red curve and can be mathematically written as dividing the OTF by the square root of its zero frequency value $\tilde{h}(\vec{0})$, which in the $NA = 0.40$ case was given as $\tilde{h}(\vec{0}) \approx 0.21$:

$$\tilde{h_\sigma}(\vec{k}) = \frac{\tilde{h}(\vec{k})}{\sqrt{\tilde{h}(\vec{0})}} \tag{2.43}$$

with $\tilde{h}_\sigma$ being the *noise normalized* OTF corresponding to unit variance. Note that the concept of $\tilde{h}_\sigma$ has been developed by *Heintzmann*, *Wicker* and *Sheppard* in [44]. Another important quantity, which can be directly derived from the noise normalized OTF, is the *detectability* $\mathcal{D}$. It compares in-focus brightness to out-of-focus contributions and therefore describes how much more prominent the detection of small objects is compared to the background. According to [45] it is defined as:

$$\mathcal{D} = \frac{h(\vec{0})}{\sqrt{\int_{-\infty}^{+\infty} d\vec{r}\ h(\vec{r})}} \tag{2.44}$$

with $h(\vec{0})$ being the maximum of the centered PSF. A basic property of Fourier transform pairs (h and $\tilde{h}$) is that the $\vec{k} = \vec{0}$ component in one space, is given by the integral of its corresponding Fourier pair:

$$h(\vec{0}) \propto \int_{-\infty}^{+\infty} d\vec{k}\ \tilde{h}(\vec{k}) \qquad \tilde{h}(\vec{0}) \propto \int_{-\infty}^{+\infty} d\vec{r}\ h(\vec{r}) \tag{2.45}$$

Using this, the detectability D can be written in terms of Fourier components:

$$\mathcal{D} \propto \frac{\int_{-\infty}^{+\infty} d\vec{k}\ \tilde{h}(\vec{k})}{\sqrt{\tilde{h}(\vec{0})}} = \int_{-\infty}^{+\infty} d\vec{k}\ \frac{\tilde{h}(\vec{k})}{\sqrt{\tilde{h}(\vec{0})}} = \int_{-\infty}^{+\infty} d\vec{k}\ \tilde{h}_\sigma(\vec{k}) \tag{2.46}$$

Hence the integral over the noise normalized OTF is proportional to the detectability D. Looking at Fig. 2.2a left, it can be seen that the blue curve shows the highest detectability, corresponding to the largest maximum value in real space.

2.3.5 Sampling and pixelation effects

The theoretical treatment so far has not dealt with the issue of sampling and pixelation, which occurs when a camera sensor captures an image. To successfully reconstruct the original (continuous) signal, the Nyquist criterion [9, 10] needs to be fulfilled. In incoherent imaging it is given as:

$$\Delta s \leq \frac{1}{2} \cdot d_{xy,min} \tag{2.47}$$

with Δs being the sampling distance. This means that the magnification m of the optical system and the pixel size of the detector need to be chosen in such a way that they realize a sampling which is at most half the size of the resolution limit. This is because of the *aliasing* effect, which is described in the appendix (p. 106). Finer sampling is generally advantageous but interestingly reduces the *SNR* in real space, as the captured photons are distributed over more pixels. Effectively reducing the expected number of photons per pixel and hence the real space *SNR*. Nevertheless, increasing the sampling will only leave us with acquiring more information about the accurate positions of each measured photon. Making the real space *SNR* definition obsolete, especially since an *SNR* improvement can still be achieved by binning a group of pixels. The Fourier space definition of the *SNR* does not vary with changing the spatial sampling, because the noise floor is calculated as the summation of photons in the whole image. It does not matter whether the captured photons are redistributed onto a larger number of pixels, the sum of photons will stay the same and so does the Fourier space *SNR*.

Also the pixelation of the sensor has some effects on the achievable *SNR*. For a pixelated detector, the recorded intensity needs to be computed as an integral over the pixel area. This means that each pixel in the image represents a summation of spatial information, which leads to some additional blurring and decreases contrast [56]. A more detailed description on this effect is given in the appendix (p. 106). In the following this additional blur will be neglected, which does not introduce any error when the pixel size is considered to be small and Nyquist sampling is realized [46].

2.4 Methods to improve *SNR*

The previous section has shown that an improvement in terms of *SNR* is favorable in imaging, as finer spatial structures become detectable. In the photon-limited regime, the *SNR* is given by the following relationship with respect to the total number of detectable photons p:

$$SNR \propto \sqrt{p} \tag{2.48}$$

Be aware, that above equation does not encompass the spatial variation of the *SNR*. It merely motivates the two conventional strategies to improve *SNR*:

- Increasing the photon collection on the hardware side.
- Performing image reconstruction on the software level.

The former tries to improve *SNR* directly by increasing p, while latter uses mathematical models to obtain more information from the recorded data. The *SNR* improvement on the hardware level always comes with tradeoffs as described below and therefore often limits other aspects of imaging. Using computational methods requires the image formation model to correspond to reality, which is usually only partially true.

2.4.1 Improving the photon collection

As equation 2.48 suggests, the photon-limited *SNR* can be increased by collecting a larger amount of photons. Ways to achieve this is by either increasing the detection *NA*, (while maintaining the same immersion medium n and magnification m), to use a longer exposure time of the detector or to intensify the illumination of the sample. All methods come with some limitations which will be discussed in the following.

Increasing the numerical aperture:

The *NA* of the detection objective is one limiting factor in an imaging system, as only photons which are emitted within a certain solid angle Ω can be collected. This is termed the *collection efficiency* and can be regarded as being proportional to the number of detectable photons [47]. The solid angle Ω is given as:

$$\Omega \propto 1 - \cos\alpha \tag{2.49}$$

with α being the half opening angle of the collection cone. This angle α can be expressed in terms of the numerical aperture *NA*, so that the solid angle becomes:

$$\Omega \propto 1 - \sqrt{1 - \sin^2\alpha} = 1 - \sqrt{1 - \left(\tfrac{NA}{n}\right)^2} \tag{2.50}$$

As equation 2.48 states, the *SNR* is proportional to the square root of the detectable photons. The photon collection is given by the solid angle Ω, yielding:

$$SNR \propto \sqrt{1 - \sqrt{1 - \left(\tfrac{NA}{n}\right)^2}} \tag{2.51}$$

which describes the change of the relative *SNR* with respect to a varying numerical aperture. Figure 2.6a shows the corresponding curves of collection efficiency and relative *SNR* for three different immersion media (air $n = 1.00$, water $n = 1.33$, oil $n = 1.52$).

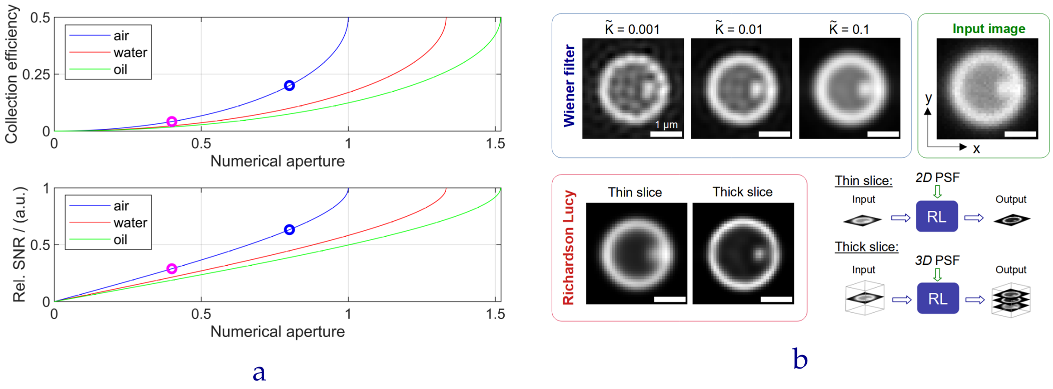


Figure 2.6: a Effects of increasing the numerical aperture *NA* on the collection efficiency and relative *SNR* in case of photon-limited imaging. Note that the maximum achievable collection efficiency of 50% and the maximum relative *SNR* is given for $NA/n = 1.0$. The blue and magenta dots represent the two OTFs shown in Fig. 2.2a. b Reconstruction results on deconvolving the noisy image from Fig. 2.4b using the Wiener filter and the RL algorithm. The RL iteration scheme enables to perform a *thick slice* reconstruction (in-focus slice shown, details in text).

The maximum achievable collection efficiency is 50%, as only the light propagating towards the detection objective can be captured. Note that this is different when two opposing objectives are being used, such as in 4Pi microscopy [48]. Changing the immersion medium n does alter the collection efficiency of the system, but it is not possible to capture more than 50% of the emitted light when going for larger n. Changing from an air to an oil immersion objective makes sense with respect to increasing the spatial resolution of the system, but does not improve the photon-limited *SNR* when the ratio NA/n is maintained. In the case of the two OTFs shown in Fig. 2.2a in blue and magenta, the reduction in photon collection is given by $\approx 21\%$ (instead of the often assumed 25%), when reducing the *NA* by a factor of two. As the relative *SNR* also decreases by a factor of two, it is crucial to perform imaging with an objective with the highest *NA* possible. Once $NA/n = 1.0$ is reached, however, an improving of *SNR* by maximizing the collection efficiency of the detection objective is not possible anymore.

Longer exposure time or higher illumination power:

Instead of capturing more photons by increasing the collection efficiency of the detection system, it is also possible to adjust external parameters to essentially achieve the same. One example would be to increase the exposure time of the detector. Doubling the exposure will double the amount of detectable photons and hence increase the *SNR* by a factor of ≈ 41%. Which is the same enhancement as averaging two consecutively acquired images with the shorter exposure time. However, both approaches come at the cost of reducing temporal resolution, which is detrimental for any recording of fast processes. Additionally they do increase the photon bleaching per frame [49]. Another approach to obtain more photons could be to increase the excitation power. When dealing with fluorescence this is not very suitable as each fluorophore only yields on average a fixed number of photons before its destruction due to photo-induced bleaching [50]. Additionally it is clear that the goal of an idealized imaging setup is to affect the sample only minimally, due to the impinging radiation. To warrant comparability we assume that the sample always emits a given number of photons.

2.4.2 Using computational reconstruction

Instead of achieving an improved *SNR* directly in the measurement process, it is also possible to use computational reconstruction algorithms to improve *SNR* in a post-processing step. Those methods require assumptions on image formation and noise characteristics, which when not met in reality, will lead to reconstruction artifacts. The used *a priori* knowledge enables algorithms to infer more information from the acquired data than initially observable. However, some caution needs to be exercised when applying those methods as the unwanted artifacts are not always identified as such easily.

In this chapter *deconvolution* is the type of image reconstruction which is introduced, a technique that aims to invert the imaging eq. 2.5. With this an estimate of the underlying sample structure S_{est} can be found. Due to the band-limit imposed onto the measured data (originating in the diffraction limit of light) and the presence of noise, this inversion cannot be done by simply computing:

$$\tilde{S}_{est}(\vec{k}) = \frac{\tilde{M}(\vec{k})}{\tilde{h}(\vec{k})} \tag{2.52}$$

as this would lead to noise amplification and render the reconstruction useless. To circumvent this problem two different strategies are introduced here: Wiener filtering [51] and iterative maximum likelihood reconstruction using the Richardson Lucy (RL) optimization strategy [52, 53].

Wiener filter based reconstruction:

This approach uses a simple filtering process to improve image quality. The filtered result S_{est} minimizes the mean-square error norm ε according to:

$$\varepsilon(\vec{r}) = \langle |S(\vec{r}) - S_{est}(\vec{r})|^2 \rangle \tag{2.53}$$

The choice of ε originates directly from the Gaussian noise model with constant variance which is implicitly assumed (see p. 107 in the appendix). Note that in this derivation the data in principle can be of arbitrarily high dimensionality.
The filtering process is performed in Fourier space as a multiplication with the Wiener filter $\tilde{W}$, so that the estimate in real space is given as:

$$S_{est}(\vec{r}) = \mathcal{F}_{dim}^{-1}\left\{ \tilde{W}(\vec{k}) \cdot \tilde{M}(\vec{k}) \right\} \tag{2.54}$$

The appendix (p. 107) shows the derivation of the Wiener filter which yields:

$$\tilde{W}(\vec{k}) = \frac{\tilde{h}^*(\vec{k})}{|\tilde{h}(\vec{k})|^2 + \tilde{K}(\vec{k})} \tag{2.55}$$

with * being the complex conjugate and $\tilde{K}$ a regularization parameter, given as:

$$\tilde{K}(\vec{k}) = \frac{\langle |\tilde{\mathcal{N}}(\vec{k})|^2 \rangle}{|\tilde{S}(\vec{k})|^2} \tag{2.56}$$

In practice the regularization parameter $\tilde{K}$ is often set to be a constant value, which is typically found empirically. Figure 2.6b (blue frame) shows the reconstruction results for three regularization values ($\tilde{K} = 0.001,\ 0.01,\ 0.1$) when the noisy image from Fig. 2.4b (bottom right, scale bar = 1 μm) is used as the input. When $\tilde{K}$ is set too low, the reconstruction will mainly result in amplified noise artifacts. Increasing $\tilde{K}$ reduces those, but comes at the cost of degrading the reconstruction quality. Meaning that the sample estimate remains blurred.

Richardson Lucy deconvolution:

A more advanced deconvolution method is based on the work of Richardson [52] and Lucy [53], which is an iterative algorithm based on Poissonian noise. A general derivation can be found in the appendix (p. 109). Note that the RL-algorithm guarantees a *positivity* constraint on the sample estimate, which exploits prior knowledge and leads typically to good reconstructions.
The algorithm begins with the initialization of the unknown sample structure

est_1. This is typically done by using the average of the measurement data M. Step 1 of the iteration is to convolve the current estimate with the given PSF $h(\vec{r})$:

Step 1:
$$\text{convEst}_l(\vec{r}) = \text{est}_l(\vec{r}) \otimes h(\vec{r}) \tag{2.57}$$

Hence convEst_l represents the corresponding image when the current sample estimate est_l is used as the unknown object. This is compared with the measured data by computing the ratio:

Step 2:
$$\text{ratio}_l(\vec{r}) = \frac{M(\vec{r})}{\text{convEst}_l(\vec{r})} \tag{2.58}$$

In case convEst_l is close to being the real unknown sample, the ratio will approach unity. Otherwise ratio_l will slightly deviate from 1, and this deviation is back-projected into sample space by convolving (again) with a point-mirrored PSF:

Step 3:
$$\text{convRatio}_l(\vec{r}) = \text{ratio}_l(\vec{r}) \otimes h(-\vec{r}) \tag{2.59}$$

The final step is to multiply convRatio_l with the current estimate:

Step 4:
$$\text{est}_{l+1}(\vec{r}) = \text{est}_l(\vec{r}) \cdot \text{convRatio}_l(\vec{r}) \tag{2.60}$$

Yielding the estimate of the following iteration, which in summary yields:

$$\text{est}_{l+1}(\vec{r}) = \text{est}_l(\vec{r}) \cdot \left[\frac{M(\vec{r})}{\text{est}_l(\vec{r}) \otimes h(\vec{r})} \otimes h(-\vec{r}) \right] \tag{2.61}$$

The challenge is to choose an appropriate number of iterations, as too many iterations will cause noise amplification. In a simulation this can be done by analyzing the normalized cross-correlation (*NCC*) and stopping the algorithm when a maximum is reached (more details in the appendix p. 111). The Richardson-Lucy update scheme typically requires many iterations to yield a good result. An acceleration technique [54] and a damping mechanism [55] (to reduce noise artifacts) will be introduced in sec. 4.3.2, where they play a crucial role.

In Fig. 2.6b the reconstruction results of processing the noisy image from Fig. 2.4b (bottom right) using RL deconvolution are shown in the red frame. When only a 2D image & PSF is used (*thin slice*, 5 accelerated iterations), the reconstruction is able to remove some of the out-of-focus blur, but fails at improving the sharpness of the in-focus structures. However, when the 3D PSF is available a three-dimensional reconstruction from a single 2D image is possible, termed a *thick slice* deconvolution. For this the measurement M is embedded into a 3D stack and set to zero everywhere except the middle slice (see Fig. 2.6b bottom). The RL reconstruction method is able to compute a 3D estimate of the underlying sample structure, which gives the algorithm the possibility to assign some of the blur to planes which are out-of-focus. This helps to improve the image quality as can be seen in Fig. 2.6b (20 accelerated iterations).

2.5 Summary and discussion

In this chapter the theory on how object information is transmitted in an incoherent imaging system has been introduced. The imaging device acts as a lowpass filter in Fourier space and limits high spatial frequencies to be transferred. Besides that, widefield detection is corrupted by out-of-focus blur, which further decreasing the image quality. Shot noise limits the imaging performance and occurs in Fourier space as a noise floor, which is proportional to $\sqrt{p}$ (with p being the total number of expected photons). As the signal scales linearly with p, the *SNR* can be improved by capturing more photons. The *SNR* in Fourier space is monotonically decreasing in terms of spatial frequencies. Meaning that small and fine details in the image will mainly be affected by noise, resulting in an effective resolution limit. This poses a problem as the general goal of imaging is to capture information with high spatio-temporal resolution, which in itself is limited to a small number of detectable photons [56]. An *SNR* enhancement automatically comes with improved contrast and detectability in single image acquisitions. Therefore it is quite crucial in any form of imaging application.

Realizing an *SNR* enhancement is typically achieved by: 1) collecting more photons; 2) using computational reconstruction methods. One limiting factor to capture light in a microscope is the detection *NA* of the objective. When $NA/n = 1.0$ is reached the collection efficiency reaches its maximum. Capturing more photons by increasing the exposure time of the detector is another way to enhance *SNR*, but comes at the cost of reduced temporal resolution and enhanced photobleaching. Just as intensifying the illumination creates a better *SNR*, but might be detrimental to the sample, especially when fluorescence is employed (photobleaching). In Fig. 2.5b it has been shown that improving *SNR* without requiring additional photons can lead to a much higher effective resolution, especially in the low photon count region. Hence improving the *SNR* on a hardware basis considering a finite number of photons, without sacrificing temporal resolution, is the motivation for the first part of this thesis. Additionally computational image reconstruction shows some strong capability to enhance *SNR* and can also be applied to already captured data. However, for it to work properly, assumptions of the underlying image formation and noise models need to match reality. Otherwise unwanted artifacts will be introduced, which might even be hard to discern from normal image content. However, a goal of this thesis is to combine both, a *SNR* improved detection scheme and appropriate computational methods, to further improve the obtainable image quality.

Chapter 3

Fluorescence microscopy

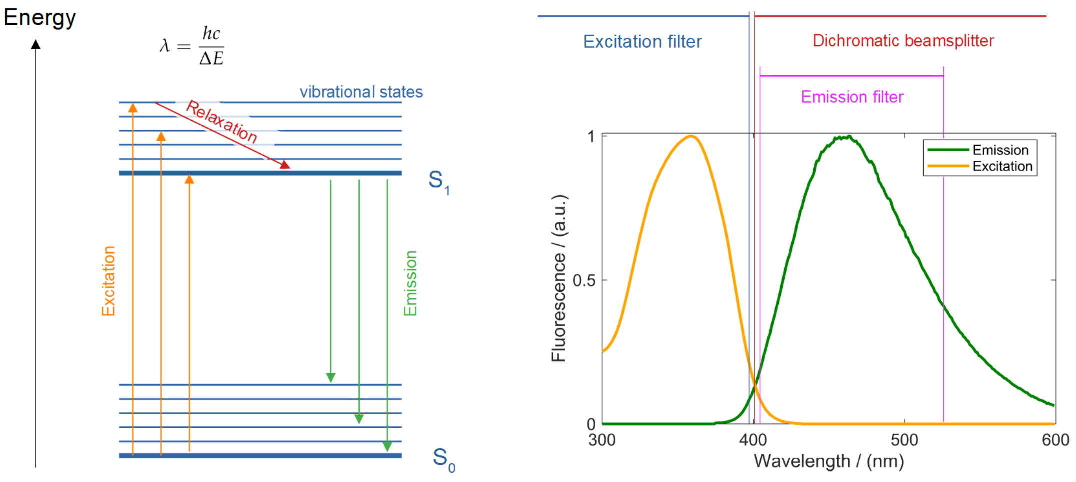
This chapter introduces the concept of *splitting & recombination* as a tool to improve the *SNR* in imaging, without requiring additional photons. This is done in the context of fluorescence microscopy, a form of incoherent imaging often employed in biomedical research. Four different splitting mechanisms are being investigated in detail: polarization, wavelength, pupil and illumination splitting. Each of these approaches will yield a respective *SNR* improvement, which is analyzed theoretically and verified experimentally.
In the beginning of this chapter, the basics on fluorescence as the underlying contrast mechanism in modern microscopy systems are introduced. This is followed by a general explanation of the *splitting & recombination* concept, using the polarization splitting as an example. Two different computational recombination techniques are presented: weighted averaging in Fourier space and multiview deconvolution. The three remaining chapters introduce the other splitting mechanisms (splitting the fluorescence emission, the detection pupil and the illumination) in more detail. In each a theoretical and experimental analysis is presented and discussed. The chapter ends with a short summary and overall discussion.

3.1 Fluorescence as a contrast mechanism

Modern light microscopes rely on fluorescence as their main source of contrast, as it enables researchers to label individual parts of their biological sample under investigation with high specificity [57, 58]. Which is not reachable when employing absorption or scattering alone. Fluorescence yields good contrast as only those regions which have been labeled with fluorophores can actually emit photons. However, the obtainable number of photons is limited by the emission characteristics of each fluorescent marker (e.g. by photobleaching [35]). The limited number of detectable photons classifies this type of imaging into the photon-limited regime. Hence improving the *SNR* will have a positive impact on the usability of many fluorescence imaging methods and is further pursued.

3.1.1 Visualisation using the Jablonski diagram

Fluorescence is best understood by looking at a *Jablonki* diagram [59], which shows a simplified representation of the distinct energy levels in a fluorescent molecule. Figure 3.1a gives an example.



a

b

Figure 3.1: a *Jablonki* diagram showing different energy levels and transitions typical in fluorescence microscopy [59]. Laser light with the correct wavelength, fitting the energy gap ΔE between two states, excites the fluorophore (orange) [60]. After some internal relaxation (red), these molecules fall back to the vibrational ground state of S_1 [61], from which fluorescence emission occurs. The emitted radiation has a slightly different wavelength which is termed the *Stokes* shift [62] of the fluorophore. Note that the emission itself is broad, as there exist many possible transitions (green) for falling back to the ground state of S_0. This can also be seen in the emission spectra in b (*DAPI*; taken from [63]). For a successful use of fluorescence the microscope has to be equipped with spectral filters: excitation (blue); dichromatic (red) + emission filter. The latter further limits the already restricted number of detectable photons p, hence the *SNR*.

The two bold horizontal lines represent two distinct energy levels in the fluorescent molecule: the ground state S_0 and the electronically excited state S_1. Both are made up of several vibrational states, which allows for a multitude of different transitions between the S_0 and S_1 state. Molecules can get excited by light with a wavelength that correpsonds to the energy gap ΔE (orange) between both transition states, given by [60]:

$$\lambda = \frac{hc}{\Delta E} \tag{3.1}$$

with h being Planck's constant and c the speed of light in vacuum.

The excited molecules lose energy through relaxation (red) and end up in the vibrational ground state of the electronically excited state S_1 [61], from which fluorescence gets emitted (green) at a different wavelength (termed *Stokes* shift [62]). The emitted light consists of a broad wavelength range, as there are many energetically possible transitions available (lower limit given for $\Delta E = S_1 - S_0$).

3.1.2 Excitation and emission spectra

Overall this leads to excitation and emission spectra similar to those shown in Fig. 3.1b, which are taken from the database *FPbase* ([63], `www.fpbase.org`). Be aware that the database does not provide in what units these spectra are given, hence in the following we will assume that they are represented as a photon density (photon number per wavelength). A certain broadness and asymmetry is noticeable in most fluorescence spectra. The area under the emission curve is proportional to the total number of emitted photons p. However, the use of fluorescence in an imaging setup requires a strong separation between the excitation and emission light. This is done using spectral filters, whose spectral ranges are indicated in Fig. 3.1b top. Note that the use of an emission filter further reduces the already limited number of measurable photons p in fluorescence microscopy, hence additionally restricting the achievable *SNR*.

3.2 General idea: splitting & recombination

As has been shown in sec. 2, improving the *SNR* will yield in an enhanced image quality. Experimental methods to achieve this usually require to capture more photons, hence inevitably come with drawbacks such as reduced temporal resolution and increased photobleaching. Note that in principle it does not matter whether the photons are collected using an increased exposure time, or whether they are split into multiple (sub-) images which are part of a time series. Each of those represent a realization of the noise process, whose expectation value E of the measurement data M is given for widefield illumination by:

$$E(\vec{r}) = S(\vec{r}) \otimes h(\vec{r}) \tag{3.2}$$

This means that one possible way to improve the *SNR* is to acquire different realizations of the noisy M and recombine them by averaging. Of course the only way to create those realizations is by repeating the measurement, e.g. by recording a time series. In this thesis the idea is to create different sub-images not by sequentially capturing images but by directly splitting image information during a single acquisition process. This will differ compared to the previous approach in the sense that each sub-image will have a different expectation value. However, using more advanced computational methods it is possible to recombine the sub-images in a way, which improves the *SNR* performance. Overall this yields an *SNR* enhancement which is achieved without the need to capture more photons. Meaning that instead of acquiring a single image with a finite photon budget in the traditional way (no splitting), it is better to split & recombine the photon budget. A comparison of both approaches is depicted in Fig. 3.2.

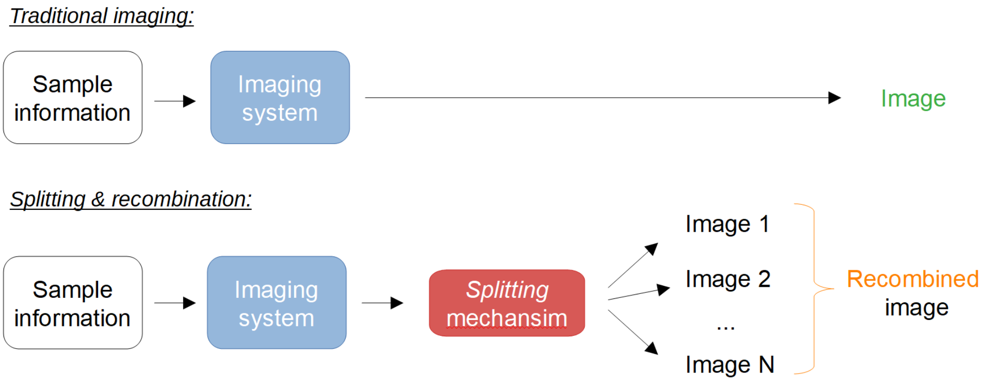


Figure 3.2: Schematic diagram indicating the difference between traditional imaging (top) and the *splitting & recombination* approach (bottom). Top: in traditional imaging information about an unknown sample is obtained via an optical system, forming a single image on a detector. The general idea is to split the image information to simultaneously obtain several (sub-) images. Different ways how to realize the splitting are described in the rest of chapter 3. All multiple images show mutually and spatially differing *SNR*, which are then subsequently computationally recombined into a final image with an improved *SNR*.

Traditional imaging (top) employs an optical system (e.g. microscope) to transfer sample information into image space (green). Instead of transferring the information directly onto the detector, a splitting into multiple simultaneously acquired sub-images (bottom) is proposed. Each of those will carry *mutually differing SNR*. Subsequent computational recombination (orange) yields a final image with improved *SNR* that does not require to capture more photons, and is equivalent in temporal resolution to the traditional approach. Different ways how to realize such a splitting will be discussed in the following.

3.2.1 Creating mutually differing sub-images

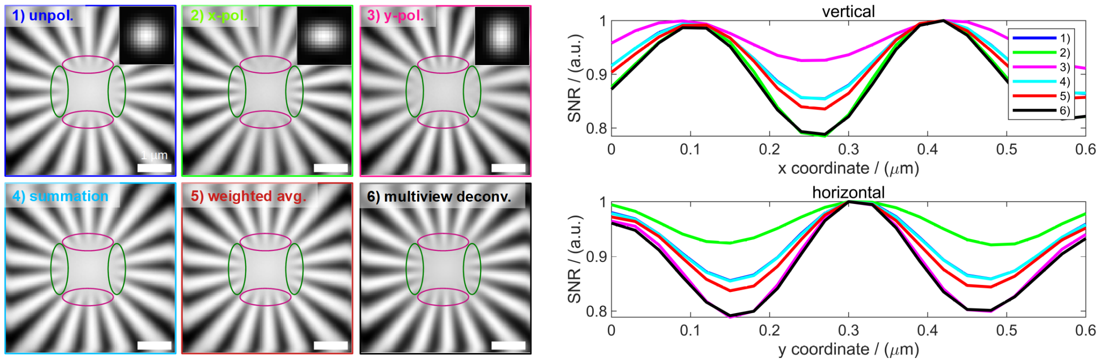
For the splitting & recombination approach to work, it is important to create sub-images with *mutually differing SNR*. For example, a 50/50%-beamsplitter in the infinity path of a microscope will fail to do so, as the two resulting sub-images will be identical. Therefore, splitting & recombination is explained using a more suitable example: the polarization dependency of the PSF.

In laser manufacturing [64] or laser scanning microscopy (LSM) [65] it is well known, that the shape of a focused laser spot can be controlled via the polarization of light in the BFP of the focusing objective. This is due to lights vectorial nature and a geometric effect introduced due to the required tilt of the light rays towards the focus, which generates differences in their interference capability. The shape of the PSF depends on this interference and therefore can be indirectly manipulated using the polarization state of the laser in the BFP. A detailed explanation on this effect can be found in the appendix (p. 113).
Note that this is also true when light is detected from a point source, instead of being focused. The *reciprocity theorem* [33] states that whenever light encounters diffraction (e.g. when it is focused), the role of light source and point of observation can be interchanged. Meaning that light emitted by a point *source* and analyzed in the infinite path of the microscope, is equivalent to a collimated laser (source) being focused into a small point. Detecting different polarization states of light will yield a mutually differing *SNR* in both sub-images, compared to the unpolarized (reference) case. For example: linearly x-polarized light leads to a narrowing of the PSF in the y and a broadening in the x direction with respect to unpolarized light (see inset in Fig 3.3 1-3; vice versa for y-polarization). Note that the idea of splitting & recombining light with respect to its polarization state was conceived by *Rainer Heintzmann* and further investigated by myself.

The method used for simulating the vectorial PSF is from [32], with $\lambda = 520$ nm, $NA = 1.4$ and $p = 10000$ (unpol.). The emitted light is split in the BFP of a microscope objective into two orthogonal linear polarization states (e.g. using a polarizing beamsplitter) and then convolved with a spokes target to create the two sub-images. With these, the *SNR* is obtained according to eq. 2.29 using 100 realizations of each measurement. The sub-images will differ in information content, e.g. the x-polarized image will show an improved modulation of the vertical oriented spokes pattern and a worsened modulation along the horizontal direction (vice versa for the y-polarization). Recombining both by a simple summation will yield the result shown in Fig. 3.3a 4), which is identical to the reference case (unpolarized; 1). Therefore, this type of recombination is not suitable, as it does not incorporate the mutually differing *SNR* into the recombined result.

3.2.2 Computational recombination

To make full use of the available information more sophisticated recombination techniques need to be applied. In this section two such methods will be discussed: *weighted averaging* in Fourier space and *multiview* deconvolution.



a b

Figure 3.3: a Polarization splitting as a first example of the *splitting & recombination* approach. Displayed is the *SNR* calculated from 100 measurements using eq. 2.29. The mechanism used is a polarization dependence of the (vectorial) PSF. 1) Imaging of the spokes target with unpolarized light (reference), we can see an isotropic blurring of spatial information, when comparing the horizontal (green) and vertical (purple) marked regions. Splitting into x- and y-polarized light (2 - 3) leads to a narrowing of image information with respect to the reference case (see PSF shape in inset, simulated using [32] with $\lambda = 520$ nm, $NA = 1.4$, $p = 10000$). Simply summing the two sub-images 4), will give a result identical to the reference case (assuming same photon budget). However, more sophisticated recombination techniques like *weighted averaging* 5) in Fourier space or *multiview* deconvolution (6; 3 accelerated iterations) lead to a significant improvement in terms of *SNR*. b Profile plot through the horizontal (green) and vertical (magenta) region indicate an improvement of modulation, when comparing the recombined (red, black) to the reference data (blue). Note that the deconvolution was limited to yield the same modulation as the split data (green & magenta). More explanation regarding weighted averaging and multiview deconvolution is given in sec. 3.2.2 and sec. 3.2.2.

Weighted averaging in Fourier space

The general idea behind *weighted averaging* in Fourier space is to add all sub-images according to their spatial frequency distribution, in such a way that the *SNR* is maximized. This method has been extensively used for image reconstruction of SIM data by *Kai Wicker* [66] and *Rainer Heintzmann* [67].

It is clear from the previous example that the x-polarized image can contribute more useful information regarding the vertical (green), but much less for the horizontal (magenta) patterns. And vice versa for the y-polarization. In Fourier space the two sub-images $\tilde{M}_{1,2}$ are given as:

$$\begin{aligned}\tilde{M}_1(\vec{k}) &= \tilde{S}(\vec{k}) \cdot \tilde{h}_1(\vec{k}) + \tilde{\mathcal{N}}_1(\vec{k}) \\ \tilde{M}_2(\vec{k}) &= \tilde{S}(\vec{k}) \cdot \tilde{h}_2(\vec{k}) + \tilde{\mathcal{N}}_2(\vec{k})\end{aligned} \tag{3.3}$$

where the subscript 1 denotes the x- and 2 the y- polarized image respectively. The total information is split into two sub-images, so is the number of detected photons per image. The noise floor in Fourier space varies with the total number of photons and is essentially characterized as the standard deviation $\sigma_{\tilde{M}_{1,2}}$. The sum of all sub-image variances needs to add up to unity (no photons are blocked):

$$\sigma^2_{\tilde{M}_1} + \sigma^2_{\tilde{M}_2} = 1 \tag{3.4}$$

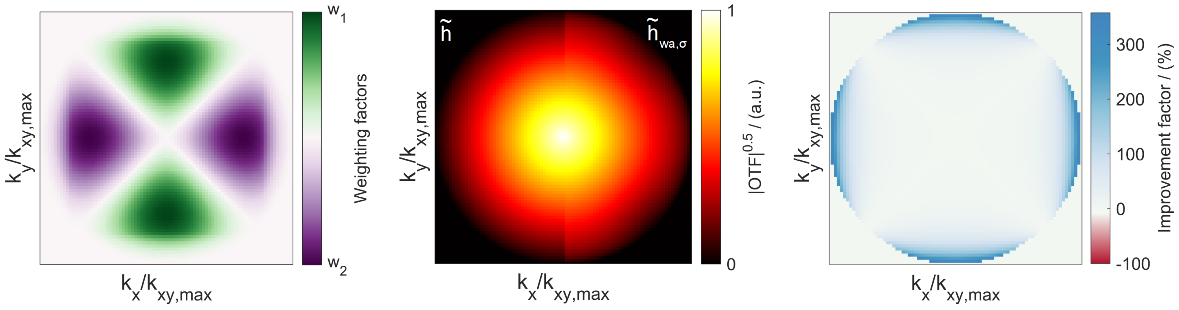
with the variances being normalized with respect to the unpolarized case.
The idea is now that for each spatial frequency, the recombined image M_{wa} is a linear mixture of information from both sub-images $M_{1,2}$. Mathematically this is written as the following weighted sum [66, 67]:

$$\tilde{M}_{wa}(\vec{k}) = \tilde{w}_1(\vec{k}) \cdot \tilde{M}_1(\vec{k}) + \tilde{w}_2(\vec{k}) \cdot \tilde{M}_2(\vec{k}) \tag{3.5}$$

with $\tilde{w}_{1,2}$ being the corresponding weights in Fourier space. Those determine how much information each sub-image will contribute to the final image M_{wa}, for each spatial frequency within the bandlimit of the detection system. The weights are chosen in such a way, that the *SNR* of the reconstructed result is maximized (see derivation on p. 114 in the appendix) and are given according to:

$$\tilde{w}_i(\vec{k}) = \frac{\tilde{h}_i^*(\vec{k})}{\sigma^2_{\tilde{M}_i}} \qquad i \in \{1, 2\} \tag{3.6}$$

The weights are the individual sub-OTFs (complex-conjugated) and scaled by the respective noise variance in Fourier space. Figure 3.4a shows a visualization of the distribution of the weighting for the polarization splitting example.



a b c

Figure 3.4: a Weighting for the two images from Fig. 3.3 2) & 3) in Fourier space (displayed is $|\tilde{w}_1| - |\tilde{w}_2|$). Horizontal information is only taken from M_1, whereas vertical information comes only from M_2. With this, *weighted averaging* is able to use the detected photons more efficiently. b Left: $|\tilde{h}|$ corresponding to the reference case (unpolarized, no splitting); right: noise-normalized OTF $|\tilde{h}_{wa,\sigma}|$ after weighted averaging recombination. The latter shows an improved transfer strength towards the high spatial frequencies. c Improvement factor (eq. 3.12) indicating that polarization dependent splitting leads to an overall enhanced information transfer, especially towards the horizontal and vertical spatial frequencies close to the cutoff $k_{xy,max}$.

For the visualization, the difference $|\tilde{w}_1| - |\tilde{w}_2|$ has been calculated, so that the green/purple color in Fig. 3.4a indicates regions where the weighting is dominated by $\tilde{M}_{1,2}$ respectively. Note that vertical structures are reconstructed mainly from image M_1 and horizontal information from M_2. This symmetry can also be seen in Fig. 3.3a, indicating the attempt to get the best out of both worlds. However, the weighted averaged result $\tilde{M}_{wa}$ does not have a constant noise floor, as its standard deviation is given according to:

$$\sigma_{\tilde{M}_{wa}}(\vec{k}) = \sqrt{\tilde{w}_1(\vec{k})\tilde{w}_1^*(\vec{k}) \cdot \sigma^2_{\tilde{M}_1} + \tilde{w}_2(\vec{k})\tilde{w}_2^*(\vec{k}) \cdot \sigma^2_{\tilde{M}_2}} \tag{3.7}$$

A noise normalization can be computed when $\tilde{M}_{wa}$ is divided by its standard deviation, which yields:

$$\tilde{M}_{wa,\sigma}(\vec{k}) = \frac{\tilde{M}_{wa}(\vec{k})}{\sigma_{\tilde{M}_{wa}}(\vec{k})} = \frac{\tilde{w}_1(\vec{k}) \cdot \tilde{M}_1(\vec{k}) + \tilde{w}_2(\vec{k}) \cdot \tilde{M}_2(\vec{k})}{\sqrt{\tilde{w}_1(\vec{k})\tilde{w}_1^*(\vec{k}) \cdot \sigma^2_{\tilde{M}_1} + \tilde{w}_2(\vec{k})\tilde{w}_2^*(\vec{k}) \cdot \sigma^2_{\tilde{M}_2}}} \tag{3.8}$$

Giving us the noise-normalized weights according to:

$$\tilde{w}_{i,\sigma}(\vec{k}) = \frac{\tilde{w}_i(\vec{k})}{\sqrt{\tilde{w}_1(\vec{k})\tilde{w}_1^*(\vec{k}) \cdot \sigma^2_{\tilde{M}_1} + \tilde{w}_2(\vec{k})\tilde{w}_2^*(\vec{k}) \cdot \sigma^2_{\tilde{M}_2}}} \tag{3.9}$$

With this equation 3.5 can be written as:

$$\tilde{M}_{wa,\sigma}(\vec{k}) = \tilde{w}_{1,\sigma}(\vec{k}) \cdot \tilde{M}_1(\vec{k}) + \tilde{w}_{2,\sigma}(\vec{k}) \cdot \tilde{M}_2(\vec{k}) \tag{3.10}$$

An effective noise-normalized OTF is defined, according to the definition of E, as:

$$\tilde{E}_{wa,\sigma}(\vec{k}) = \sqrt{\frac{|\tilde{h}_1(\vec{k})|^2}{\sigma^2_{\tilde{M}_1}} + \frac{|\tilde{h}_2(\vec{k})|^2}{\sigma^2_{\tilde{M}_2}}} \cdot \tilde{S}(\vec{k}) = \tilde{h}_{wa,\sigma}(\vec{k}) \cdot \tilde{S}(\vec{k}) \tag{3.11}$$

Figure 3.4b shows $|\tilde{h}|$, the OTF of the reference case (no splitting) and the polarization split and weighted averaged result $|\tilde{h}_{wa,\sigma}|$ side by side. They are depicted with a γ-adjustment of $\gamma = 0.5$ and show an improved transfer strength of $|\tilde{h}_{wa,\sigma}|$ towards the high spatial frequencies. For theoretical investigations on the performance of the different splitting mechanism proposed in this thesis, a measure comparing them to the reference case (no splitting) needs to be defined. This is done by calculating the *spectral* improvement factor *IF*, given as:

$$IF(\vec{k}) = \frac{|\tilde{h}_{wa,\sigma}(\vec{k})|}{|\tilde{h}(\vec{k})|} - 1 \tag{3.12}$$

IF is a number deviating from 0, depending on improvement (> 0) or degradation

(< 0) with respect to the reference case ($IF = -100\%$ being the maximum achievable degradation as then $\tilde{h}_{wa,\sigma} = 0$). The distribution of IF for the application of polarization splitting is shown in Fig. 3.4c. Significant improvement can be seen for frequencies corresponding to the horizontal and vertical direction, close to the cutoff $k_{xy,max}$. The enhancement reaches about 300% compared to the reference case (corresponding to a SNR quadrupling for those particular frequencies), indicating that the method of polarization splitting might be advantageous in the detection of very fine structures, where a high NA is required. Additionally, no spatial frequency experiences a negative IF, meaning that there is no loss of image information compared to the reference case. Note that the improvement factor is directly related to the previously introduced detectability $\mathcal{D}$ through:

$$\mathcal{D}_{wa} - \mathcal{D} \propto \int_{-\infty}^{+\infty} d\vec{k}\ IF(\vec{k}) \cdot \tilde{h}(\vec{k}) \tag{3.13}$$

Hence IF expresses a difference in detectability, but maintains information on individual spatial frequencies which are otherwise lost in the definition of $\mathcal{D}$.

Similarly, also a *spatial* improvement factor $IF_{\mathcal{V}}$ in terms of the visibility $\mathcal{V}$ can be defined, according to:

$$IF_{\mathcal{V}}(\vec{r}) = \frac{\mathcal{V}_{wa,\sigma}(\vec{r})}{\mathcal{V}(\vec{r})} - 1 \tag{3.14}$$

Which is useful when images containing periodic structures are being analyzed. A result of the attainable SNR of the weighted average reconstruction method is shown in Fig. 3.3a 5). An enhancement in terms of modulation depth becomes more clear when looking at a line profile through the green & purple marked regions, which is described in more detail in sec. 3.2.2.

Multiview deconvolution for image fusion

Another approach to recombine the split data is to perform image fusion via *deconvolution*, which tries to find an estimate of the true underlying sample structure S. Both sub-images transfer information about S (see eq. 3.3), so that in principle an iterative RL-deconvolution can be used for both images in parallel. This will result in two correction factors (convRatio$_i$), which need to be summed [68], to yield the sample estimate $\hat{S}$ for the next iteration. The algorithm is depicted in Fig. 3.5 and starts with a first estimate est$^{(1)}$, which is the average value of all measured sub-images (i.e. a constant).

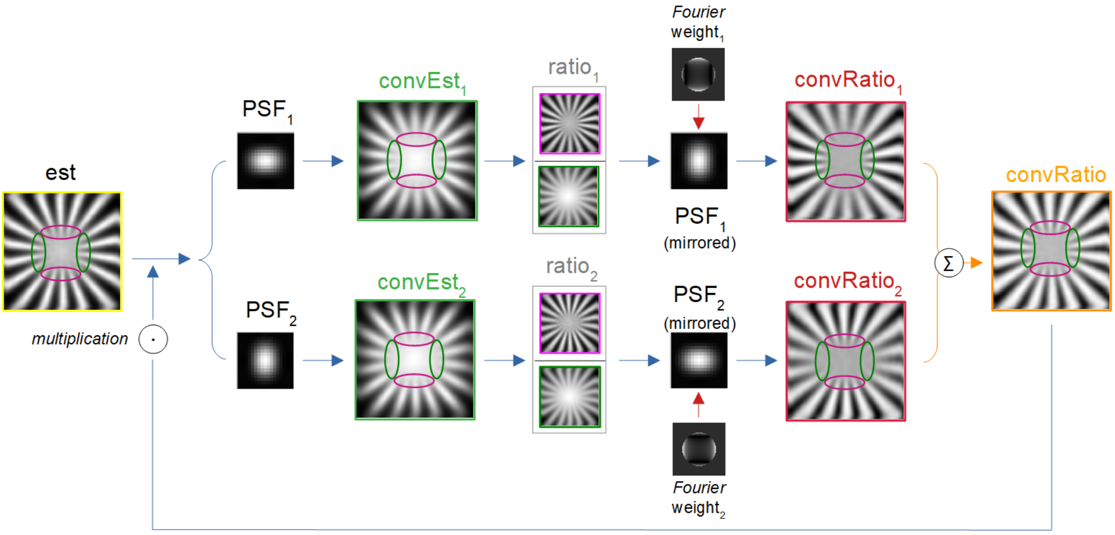


Figure 3.5: Graphical depiction of the multiview RL-deconvolution algorithm used for the computational recombination of the split image data. Each iteration starts with an estimate (est) that gets blurred using the different sub-PSFs $h_{1,2}$. With this, two ratios (gray) with respect to the measured image data $M_{1,2}$ are calculated. Both get blurred again by the point-mirrored $h_{1,2}$, yielding convRatio$_{1,2}$, which get fused using weighted averaging in Fourier space (details in text). This procedure can be described as a modification to the back-propagator of the RL-deconvolution, similar to the work in [69]. A single correction factor convRatio (orange) is obtained, which is then multiplied to the actual estimate and the next iteration begins. Note that the modification of the back-propagator increases the convergence speed of the algorithm but in principle removes the convergence guarentee of RL deconvolution.

From this sample estimate, the ideal images corresponding to the two sub-PSFs $h_{1,2}$ are being computed according to:

Step 1:
$$\text{convEst}^{(l)}_{1,2}(\vec{r}) = \text{est}^{(l)}(\vec{r}) \otimes h_{1,2}(\vec{r}) \tag{3.15}$$

with the superscript in round brackets denoting the current iteration number and the subscript the sub-images. The convolved estimates are shown in green color in Fig. 3.5. Following the main algorithm from sec. 2.4.2, two correction ratios (gray) are given as:

Step 2:
$$\text{ratio}^{(l)}_{1,2}(\vec{r}) = \frac{M_{1,2}(\vec{r})}{\text{convEst}^{(l)}_{1,2}(\vec{r})} \tag{3.16}$$

Each ratio is convolved by the respective point-mirrored PSF, yielding:

Step 3a:
$$\text{convRatio}^{(l)}_{1,2}(\vec{r}) = \text{ratio}^{(l)}_{1,2}(\vec{r}) \otimes h_{1,2}(-\vec{r}) \tag{3.17}$$

Note in Fig 3.5 how the two different correction factors clearly show the "horizontal/vertical" properties from the previous polarization splitting example.

The last step in the single view RL-deconvolution, is the multiplication of the old sample estimate (est$^{(l)}$) with the current correction factor (convRatio$^{(l)}$). Before this can be done in the multiview version, a summation of both convRatio$_{1,2}$ needs to be performed. Instead of computing an ordinary sum, a weighted version is introduced here. With the basic idea very similar to the weighted averaging approach from before: when convRatio$_{1,2}$ are being fused, distinct spatial frequencies need to be weighted differently to make better use of the obtainable information. Note that this procedure is similar to the work of [69] where the backpropagator $h(-\vec{r})$ of the RL algorithm has been modified by multiplication with a "Wiener-Butterworth filter". The general goal of the propagator modification is to accelerate the convergence of the maximum likelihood reconstruction, but comes at the cost of losing the convergence guarantee of the ordinary Richardson-Lucy deconvolution. However, in all reconstructions obtained throughout this thesis the proposed algorithm gave reasonable results, hence is regarded to be stable. Mathematically the weighted summation in Fourier space is witten as:

$$\underline{\textit{Step 3b:}} \qquad \text{convRatio}^{(l)}(\vec{r}) = \sum_{i=1}^{N=2} \left(\mathcal{F}_{dim}^{-1} \left\{ \tilde{w}_i(\vec{k}) \cdot \mathcal{F}_{dim} \{ \text{convRatio}_i^{(l)}(\vec{r}) \} \right\} \right) \tag{3.18}$$

Note that steps *3a* and *3b* can be summarized by introducing the real space equivalent of the Fourier weighting factors $w_i = \mathcal{F}_{dim}^{-1}\{\tilde{w}_i\}$:

$$\underline{\textit{Step 3:}} \qquad \text{convRatio}^{(l)}(\vec{r}) = \sum_{i=1}^{N=2} \left\{ \text{ratio}_i^{(l)} \otimes [w_i(\vec{r}) \otimes h_i(-\vec{r})] \right\} \tag{3.19}$$

The last step is to multiply convRatio$^{(l)}$ to the current sample estimate:

$$\underline{\textit{Step 4:}} \qquad \text{est}^{(l+1)}(\vec{r}) = \text{est}^{(l)}(\vec{r}) \cdot \text{convRatio}^{(l)}(\vec{r}) \tag{3.20}$$

The final multiview deconvolution update equation is given according to:

$$\text{est}^{(l+1)}(\vec{r}) = \text{est}^{(l)}(\vec{r}) \cdot \sum_{i=1}^{N=2} \left\{ \frac{M_i(\vec{r})}{\text{est}^{(l)}(\vec{r}) \otimes h_i(\vec{r})} \otimes [w_i(\vec{r}) \otimes h_i(-\vec{r})] \right\} \tag{3.21}$$

Note that in the proposed algorithm the modification of the back-propagator only takes place when multiple sub-images are used and that the corresponding weights have some physical meaning behind them, as they are given by the (scaled) sub-OTFs. This is different to the work of [69], where the modification of the back-propagator is also applied when reconstructing single images and the

parameters of the Wiener-Buttterworth filter need to be tuned by hand.
The result of the obtainable *SNR* in case of imaging the spokes target is shown in Fig. 3.3a 6). To make a fair comparison the number of (accelerated) iterations was set to 3, so that the observable modulation does not get bigger than the one corresponding to the "recorded" sub-images. Note that this chapter is more about introducing the concept of *splitting & recombination* and not about a complete evaluation of the true potential of polarization splitting.

Comparing visibility and detectability:

Figure 3.3b shows two line profiles of the green & magenta marked regions in Fig. 3.3a. Note that the split image data (green & magenta) always shows a strong enhancement in one particular direction and a very low modulation in the other. A normal summation (cyan) will yield the same result as the reference case (unpolarized, no splitting). With weighted averaging (red) it is possible to go below the conventional curve (blue) and maintain better modulation in all directions. A further improvement can be reached when the multiview deconvolution approach (black) is used, which is basically able to fuse both sub-images in such a way that the maximum attainable modulation is preserved in both directions. Values of visibility $\mathcal{V}$ and spatial improvement factor $IF_{\mathcal{V}}$ are given in Tab. 3.1:

	Unpol.	x-pol.	y-pol.	Summation	Weighted Avg.	Multiview deconv.
$\mathcal{V}_{vert.}$	7.92	12.08	3.94	7.92	9.02	11.86
$IF_{\mathcal{V},vert.}$	0.00	52.56	-50.23	0.00	13.97	49.87
$\mathcal{V}_{horiz.}$	7.75	3.93	11.76	7.79	8.83	11.61
$IF_{\mathcal{V},horiz.}$	0.00	-49.22	51.76	0.52	13.92	49.83
$\mathcal{D}$ / (a.u.)	17.15	12.13	12.13	17.15	18.07	-

Table 3.1: Visibility $\mathcal{V}$ and spatial improvement factor $IF_{\mathcal{V}}$ in the vertical and horizontal direction and the detectability $\mathcal{D}$ for the unpolarized, x- and y-polarized, as well as the recombined data (summing, weighted averaging and multiview deconvolution). Values are given in *percent*.

The values are given in percent and show that a maximum improvement is reached at $\approx 50\%$ using the multiview deconvolution approach. Also weighted averaging is able to enhance the visibility but only by $\approx 14\%$, while normal summing does not yield an improvement at all. Additionally the detectability $\mathcal{D}$ has been computed for each imaging modality and also here an improvement can only be observed when splitting & recombination is employed.
Altogether it has been shown that it makes sense to include *splitting & recombination* into the imaging process. The challenge is now to find appropriate splitting

scenarios which can be implemented in a real experimental setup and to investigate how much improvement they are capable of. This is done in this thesis with the remaining three splitting mechanism: spectral splitting of fluorescence emission, spatial splitting of the detection pupil and splitting of the illumination.

3.3 Splitting the fluorescence emission

As has been shown in chapter 3.1, the emitted fluorescence signal occupies a spectral region of about 50 - 100 nm in width. A measure of spatial resolution is the *FWHM* of the PSF, which depends on the emission wavelength of the fluorophore (see eq. 2.8). Hence, the observable PSF should be an incoherent superposition of many single wavelength PSFs, scaled by the emission spectrum and the influence of the spectral filters in the detection beam path. It will be shown that the asymmetric shape of emission spectra will lead to spatial broadening of the measurable PSF *FWHM*, compared to a PSF obtained for the peak emission wavelength. Such a broadening effect already indicates, that it might be useful to employ wavelength splitting, to be able to recombine the split data in a more meaningful way. Note that the idea of wavelength splitting was conceived and further investigated by myself, under the guidance and help of *Rainer Heintzmann*.
In this section first the effects of fluorescence emission on the width of the detection PSF is investigated, followed by a simulation showing a degradation of deconvolution results when spatial broadening is not being accounted for. A *SNR* improvement can be achieved when wavelength splitting is employed, which is theoretically and experimentally studied thereafter and even yields and improvement when a perfectly symmetrical emission spectrum is used.

3.3.1 PSF broadening due to fluorescence

Fluorescence emission typically occupies a wavelength region of about 50 - 100 nm in width and exhibits an asymmetric shape towards longer wavelengths. This has already been discussed as a result of the underlying mechanism of fluorescence in chapter 3.1 (e.g. *Kasha's* rule [61]). Fig. 3.6a shows three examples of spectra (Alexa-488, DAPI and mPlum; from [63]). It is assumed that such spectra are normalized to show the emitted photons per wavelength (= photon density) on the y-axis. To describe such spectra $\varepsilon(\lambda)$ mathematically, a *log-normal* distribution, similar to the work in [70, 71, 72], is used:

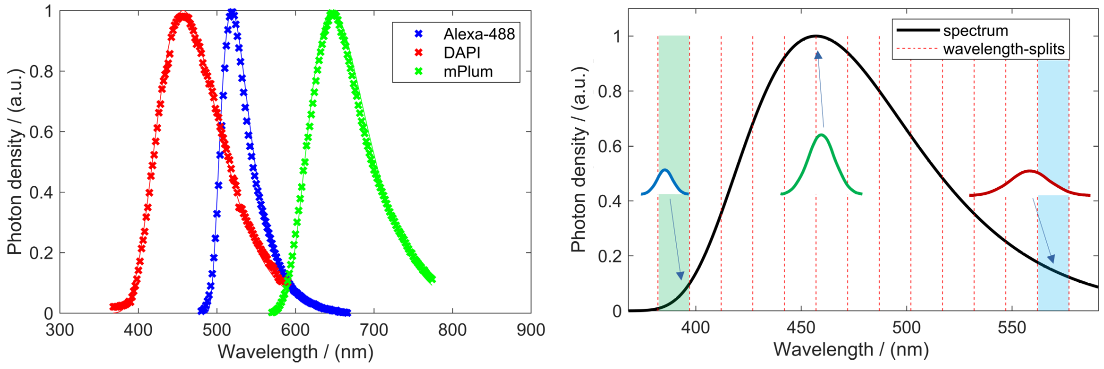
$$\varepsilon(\lambda) = \exp\left(-\frac{\ln 2}{s^2}\left[\ln\left\{1 + 2\frac{s}{w}(\lambda - \lambda_{max})\right\}\right]^2\right) \tag{3.22}$$

With λ_{max} being the peak emission of the fluorophore, w the spectral *FWHM* and s the shape parameter, indicating the skewness of the distribution. Note that for $s \to 0$ the distribution becomes a symmetric Gaussian (see p. 118 in the appendix). The area under the curve $A = \int d\lambda\ \varepsilon(\lambda)$ is not normalized to unity and will appear later as a parameter to determine the spatial broadening effect.
The results of a least-square fit of each spectrum are shown in the following table.

	Alexa-488	DAPI	mPlum
s	0.56	0.42	0.37
λ_{max} / (nm)	519	457	647
w / (nm)	43.61	93.87	87.64

Table 3.2: Results of least-squares fitting the log-normal distribution to the emission spectra $\varepsilon(\lambda)$.

Figure 3.6b depicts the fitted curve of DAPI in more detail. The resolution of an imaging system is proportional to the width of the corresponding PSF, which scales linearly with the respective emission wavelength. This is schematically depicted as the three small Gaussians which get broader towards longer wavelengths, but also scale their respective area according to the emission spectrum.



a

b

Figure 3.6: a Fluorescence emission spectra (marker) for three different fluorophores (Alexa-488, DAPI and mPlum; taken from [63]) and the corresponding least-squares fitted log-normal distribution (lines). The spectral curves are assumed to be proportional to the photon number per wavelength (= photon density). Note that fluorescence is accompanied by a broad and asymmetric emission, as described in sec. 3.1. b Fitted spectrum for DAPI including three Gaussians, indicating the wavelength scaling of the detection PSF width. The vertical lines represent the different wavelength bands, which are later used for the wavelength splitting in sec. 3.3.3.

As typical imaging sensors only detect intensities, all the spectral information is captured at once, which means that the effective PSF $h(r)$ of such a system can be described as an incoherent sum of individual "single wavelength" PSFs $h(r; \lambda)$, weighted with the corresponding emission spectrum (with $r = |\vec{r}|$):

$$h(r) = \int_{\lambda^-}^{\lambda^+} d\lambda\ \varepsilon(\lambda) \cdot h(r; \lambda) \tag{3.23}$$

with $\lambda^{\pm}$ being the limits of the recorded spectrum or an emission filter.
Equation 3.23 describes a broadening of the PSF $h(r)$ compared to $h(r; \lambda_{max})$, which is the PSF corresponding to the peak emission of the fluorophore. This is due to the fact that there are always more longer than shorter wavelengths present in the spectrum, which contribute broader PSFs than $h(r; \lambda_{max})$ to the summation. Nevertheless, $h(r; \lambda_{max})$ is often taken to be the theoretical PSF in an imaging experiment when no experimental data is available, neglecting the influence of the broad emission spectrum.
We now estimate the order of magnitude of broadening that is to be expected from this effect. To do this equation 3.22 is rewritten into:

$$\varepsilon(\lambda) = \exp\left(-a\left[\ln\left\{1 + b(\lambda - \lambda_{max})\right\}\right]^2\right) \tag{3.24}$$

with the two constants a and b given according to:

$$a = \ln 2 / s^2 \tag{3.25}$$

$$b = 2s/w \tag{3.26}$$

Next $h(r; \lambda)$ is assumed to be a simple Gaussian. The best representation in a least squares sense is given according to [73] as:

$$h(r; \lambda) = \exp\left(-\frac{r^2}{2\sigma_r^2}\right) \tag{3.27}$$

with σ_r modeling the wavelength dependent scaling of the width of the Gaussian:

$$\sigma_r = 0.21 \frac{\lambda}{NA} \tag{3.28}$$

Which is related to the *FWHM* value Δr according to:

$$\Delta r = 2\sqrt{2 \ln 2} \cdot \sigma_r \tag{3.29}$$

Lets introduce another constant $C = NA^2/(2 \cdot 0.21^2)$ and rewrite the single wavelength PSF model into:

$$h(r; \lambda) = \exp\left(-C\frac{r^2}{\lambda^2}\right) \tag{3.30}$$

The goal is now to solve the integral in eq. 3.23, using the Gaussan PSF model. To do this we express $h(r;\lambda)$ as the following Taylor series:

$$h(r;\lambda) \approx h(r;\lambda_{max}) \left[1 + 2\frac{Cr^2}{\lambda_{max}^3}(\lambda - \lambda_{max})\right] \tag{3.31}$$

And introduce some constants $c(r)$ and $d(r)$, which depend on r:

$$\begin{aligned} c(r) &= h(r;\lambda_{max}) && (3.32) \\ d(r) &= 2C \cdot r^2/\lambda_{max}^3 && (3.33) \end{aligned}$$

So that eq. 3.31 can be simplified to:

$$h(r;\lambda) \approx c(r)\left[1 + d(r) \cdot (\lambda - \lambda_{max})\right] \tag{3.34}$$

With this, the aforementioned integral can now be solved, which yields:

$$h(r) \approx c(r)\left[A + d(r) \cdot (\lambda_{CoM} - A \cdot \lambda_{max})\right] \tag{3.35}$$

with the two variables A and λ_{CoM} which are the area under the $\varepsilon(\lambda)$-curve (proportional to the total number of detectable photons p) and the "center-of-mass" wavelength respectively. Both parameters are defined as:

$$\begin{aligned} A &= \int_{\lambda^-}^{\lambda^+} d\lambda\ \exp\left(-a\left[\ln\left(1 + b(\lambda - \lambda_{max})\right)\right]^2\right) && (3.36) \\ \lambda_{CoM} &= \int_{\lambda^-}^{\lambda^+} d\lambda\ \exp\left(-a\left[\ln\left(1 + b(\lambda - \lambda_{max})\right)\right]^2\right) \cdot \lambda && (3.37) \end{aligned}$$

The analytical solution to both integrals is given in the appendix (p. 118) and leads to the following approximation of the full spectrum PSF $h(r)$:

$$h(r) \approx h(r, \lambda_{max}) \left[A + \frac{NA^2}{0.21^2} \cdot \frac{r^2}{\lambda_{max}^3}\left(\lambda_{CoM} - A \cdot \lambda_{max}\right)\right] \tag{3.38}$$

From the structure of eq. 3.38 it is noticeable that $h(r)$ can be approximated as a conventional (scaled) Gaussian PSF $A \cdot h(r;\lambda_{max})$ plus a broadening factor. The latter can be expressed as the summation of the two functions f and g:

$$h(r) \approx A \cdot \left[f(r) + g(r)\right] \tag{3.39}$$

which are given according to:

$$\begin{aligned} f(r) &= h(r;\lambda_{max}) && (3.40)\\ g(r) &= 2CD \cdot \frac{r^2}{\lambda_{max}^3} \exp\left(-C \cdot \frac{r^2}{\lambda_{max}^2}\right) && (3.41) \end{aligned}$$

To approximate how this spatial broadening affects the *FWHM* of the measurable PSF $h(r)$ a linear interpolation scheme is used, schematically shown in Fig. 3.7a. The blue curve represents $f(r)$ and $r_{1/2}$ the location of its *FWHM*. To characterize a linear relationship around this *FWHM* region, an additional point r_{max} is needed. Which is set to be the maximum of $g(r)$, depicted in green in the small inset. The corresponding intensity values of $h(r)$ are given as $f(r_{1/2}) + g(r_{1/2})$ and $f(r_{max}) + g(r_{max})$ at these particular locations. A linear model (red line) between those two points can be used to find the *FWHM* position $r_{1/2;\mathcal{B}}$ of $h(r)$ (see p. 119 in the appendix for more details). Comparing this to the old *FWHM* value Δr yields a broadening $\mathcal{B}$ of:

$$\begin{aligned} \mathcal{B} &= \Delta r_B - \Delta r = \\ &= 2\sqrt{\ln 2}\frac{\lambda_{max}}{\sqrt{C}}\left[\frac{m - \sqrt{\ln 2}\cdot\sqrt{C}D/\lambda_{max}^2}{m} - 1\right] && (3.42) \end{aligned}$$

with m being the slope of the linear expression. The broadening is depicted in Fig. 3.7b, where the peak emission model $h(r;\lambda_{max})$ (black) is compared to the broadened case; the simulated $h(r)$ (blue) and to the theoretical approximation $h_{est}(r)$ (red dotted). The latter describes the *FWHM* of the broadened result well, but is not able to reproduce the side lobes of a typical PSF.

Having established a theoretical expression of the broadening in form of eq. 3.42, the three different fluorophores shown in Fig. 3.6a can be compared. This is done in Table 3.3, for two different imaging optics ($NA = 0.80$ air and $NA = 1.40$ oil immersion) which were simulated using [32]. The theoretical prediction is compared to the numerical result which was obtained by fitting a Gaussian to the simulated PSF data, yielding a standard deviation which was converted into the corresponding *FWHM*. Note the good agreement with the numerical results that shows a spatial broadening on the order of 10 nm. When comparing the theoretically predicted broadening to the *FWHM* Δr of the peak emission-only PSF $h(r;\lambda_{max})$, it can be seen that $\mathcal{B}$ does not depend on the detection optics (NA and n). The broadening is on the order of 4% with respect to Δr for the evaluated emission spectra, hence will be difficult to detect in most imaging applications.

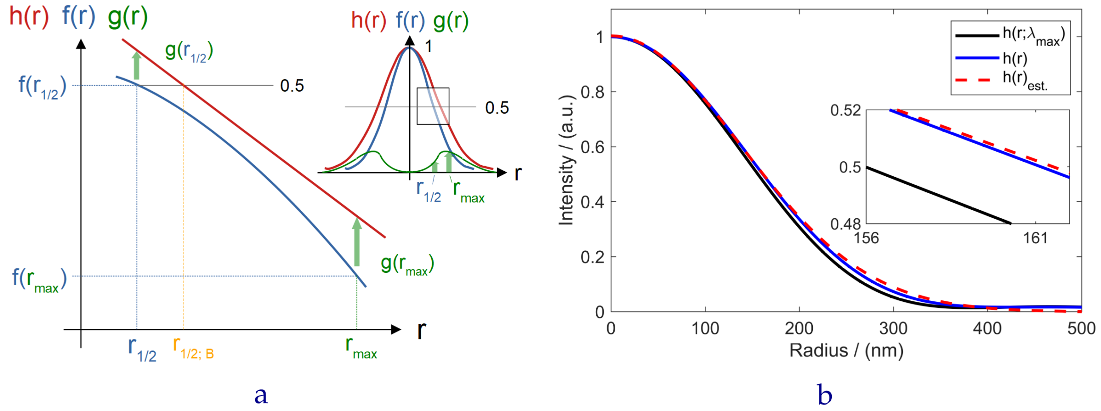


Figure 3.7: a Schematic on how to find the interpolated value of the broadened *FWHM*. The *FWHM* region of $f(r) = h(r; \lambda_{max})$ (blue) is depicted in small on the top right. $r_{1/2;B}$ corresponds to the *FWHM* value of the broadened PSF (red) and can be found using a linear fit between two points ($r_{1/2}$ and r_{max}). b Comparison of different PSFs: peak emission $h(r; \lambda_{max})$ (black), broadened simulated $h(r)$ (blue) and theoretical approximation $h_{est}(r)$ (red dotted). The inset shows a zoomed view of the *FWHM* region and indicates the spatial broadening effect.

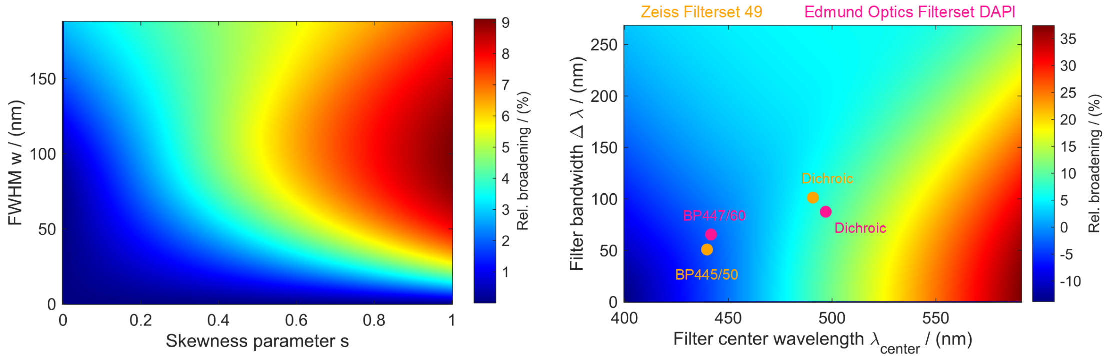
	NA/n	Δr / (nm)	$\mathcal{B}_{num.}$ / (nm)	$\mathcal{B}_{theor.}$ / (nm)	$\mathcal{B}_{theor.}/\Delta r$
Alexa-488	0.80/1.00	320.81	10.79	11.04	0.034
	1.40/1.52	183.32	6.41	6.31	0.034
DAPI	0.80/1.00	282.49	12.46	13.34	0.047
	1.40/1.52	161.42	7.40	7.62	0.047
mPlum	0.80/1.00	399.94	11.22	11.19	0.028
	1.40/1.52	228.56	6.62	6.39	0.028

Table 3.3: Comparing the broadening of the theoretical prediction $\mathcal{B}_{theor.}$ with numerical results $\mathcal{B}_{num.}$, for the three different fluorescent structures and two optical detection configurations.

Instead of looking at three specific fluorophore examples, the general dependency on the *FWHM* w and the skewness s for an emission spectrum (modeled as a log-normal curve) is investigated. Figure 3.8a shows the relative broadening $\mathcal{B}_{theor.}/\Delta r$ for a variety of values of w and s (with $\lambda_{max} = 457$ nm).

The general trend shows that especially the asymmetry of the spectrum has a strong influence on the spatial broadening of the detection PSF. Hence, choosing a wide but completely symmetric emission spectrum is more beneficial than looking for the narrowest $\varepsilon(\lambda)$-curve. Nevertheless, perfectly symmetric spectra do not exist in reality, as the molecule bandstructure forbids this.

Typical fluorescence imaging setups are often equipped with a bandpass (BP) filter in the emission path. Those are characterized by a center wavelength λ_{center} and a bandwidth $\Delta\lambda$, e.g. BP450/50 is equal to a bandwidth of $\Delta\lambda = 50$ nm around $\lambda_{center} = 450$ nm. Fig. 3.8b indicates the dependency of the broadening when such emission filters are used. Two different filter sets (excitation,



a b

Figure 3.8: a Dependency of the relative broadening $\mathcal{B}_{theor.}/\Delta r$ on the shape of the emission spectrum, given as the *FWHM* w and the skewness parameter s. Especially the latter shows a strong influence on the observable spatial broadening $\mathcal{B}$, reaching values of up to 9%. b In typical fluorescence imaging setups an emission bandpass filter is used (centered at λ_{center} with a bandwidth $\Delta\lambda$). The dependency of the relative broadening $\mathcal{B}_{theor.}/\Delta r$ shows that using such an emission filter reduces the spatial broadening, with respect to the peak emission case $h(r, \lambda_{max})$, and can even revert it (= narrowing). This effect is depicted using two commercially available filter sets (Zeiss/Edmund Optics marked orange/magenta). However, the use of an emission filter further reduces the number of detectable photons, hence the *SNR*.

dichroic & emission) for imaging DAPI have been chosen (Zeiss/Edmund Optics marked orange/magenta) and are compared with and without the use of the respective emission BP. Without the bandpass a relative broadening $\mathcal{B}_{theor.}/\Delta r$ of about $6\% - 7\%$ is obtained. However, including the BP filter in the detection will lead to a "negative" broadening, i.e. narrowing with respect to the peak emission only PSF $h(r, \lambda_{max})$. This is because the emission filter have a center wavelength $\lambda_{center} < \lambda_{max}$, meaning that longer wavelengths get blocked more. Instead of broadening the effective PSF $h(r)$, a narrowing is observed, as this effect shifts the asymmetry in the incoherent superposition towards the narrower point-spread-functions. However, the use of an emission filter reduces the number of photons which can reach the detector and therefore further limit the achievable *SNR*, which is already restricted due to photobleaching [49]. Including a BP filter in the imaging setup will reduce the broadening but also inevitably decrease the attainable *SNR*. Image reconstruction techniques can increase the *SNR* but will only yield optimal results when the correct PSF $h(r)$, instead of $h(r; \lambda_{max})$ is used.

3.3.2 Degradation of image deconvolution

As has been shown in chapter 2.4.2, deconvolution is able to enhance the *SNR*. However, it only unfolds its full potential when the correct PSF is known. In case an experimental PSF is not available, a theoretical PSF is often calculated using a model such as [32]. Which typically requires information on the optical system (NA & refractive index n) and the emission wavelength of the fluorophore used. Often the peak emission λ_{max} is taken, a parameter that is easily available for many fluorescent markers. Of course the experimentally acquired image data will be affected by the complete emission spectrum and spectral filters, not just by λ_{max} alone. When the wrong PSF model $h(r;\lambda_{max})$ is used for image deconvolution (instead of $h(r)$), this will yield in non-optimal reconstruction results. To test this, the imaging of a spokes target with a $NA = 0.8$ objective and $p = 10000$ photons has been simulated for DAPI. For each wavelength of the emission spectrum the corresponding PSF was computed (using [32]) and summed to yield $h(r)$, according to eq. 3.23. With this the spokes target was convolved and Poisson noise applied, yielding the (single) image M. In the next step the RL deconvolution algorithm was used either with $h(r;\lambda_{max})$ (peak emission-only) or $h(r)$ as the corresponding PSF model, the iteration number for an optimal reconstruction was again found by computing the normalized cross-correlation NCC (see p. 111 in the appendix). The visibility of the image reconstruction for both models (black: 85; blue: 72 accelerated iterations) are shown in Fig. 3.9a.

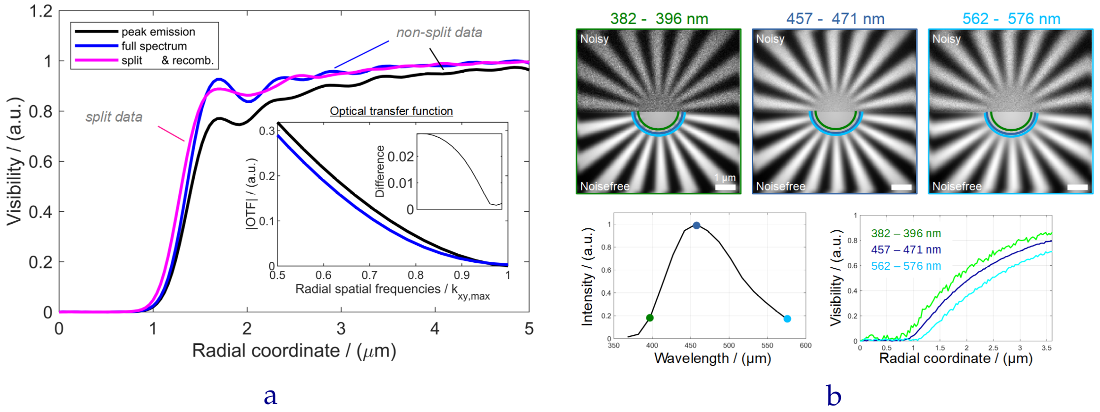


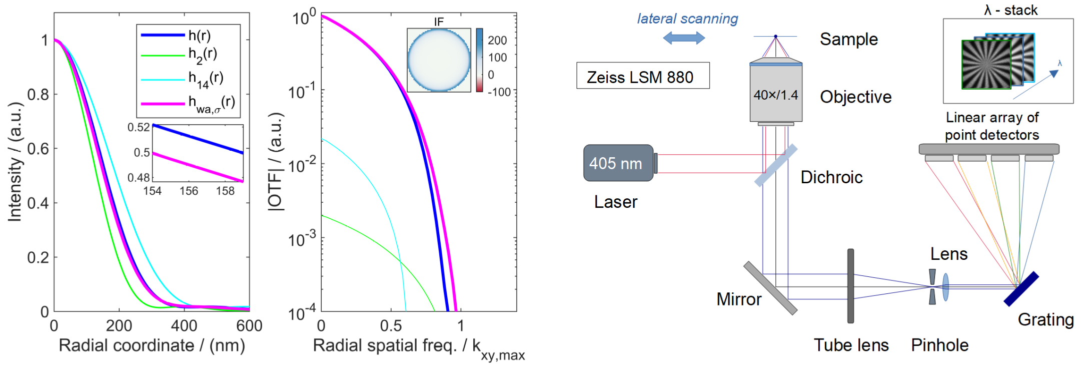
Figure 3.9: a Visibility obtained after a noisy spokes target is deconvolved using either the peak emission-only PSF $h(r;\lambda_{max})$ (black, 85 iterations) or the broadened full spectrum PSF $h(r)$ (blue, 72 iterations). The latter achieves a stronger contrast at the medium to high spatial frequencies, which can be attributed to the fact that $h(r;\lambda_{max})$ overestimates the image information content for those frequencies (see inset). b Creating mutually differing sub-images by wavelength splitting of simulated spokes target ($p = 10000$). Top: three images (with and without noise) corresponding to small, peak and large wavelength region. Note how the resolution gets worse with increasing λ. Bottom: sum of the number of photons per each wavelength band indicates the asymmetric characteristics of fluorescence emission (left); change in resolution and noise can also be seen looking at the visibility curve (right).

Comparing both curves indicates that especially structures with medium to high frequencies (towards the center of the spokes target), can be recovered with better visibility when the correct PSF model $h(r)$ is used for reconstruction. The wrong model $h(r, \lambda_{max})$ (peak emission-only) overestimates the image information content at those frequencies, as depicted with the two corresponding OTFs in the inset. Hence for any deconvolution algorithm to work correctly it is either required to know the experimental PSF, perform a blind-deconvolution [74] (much more computationally expensive) or to include the PSF broadening effect into the theoretical model and use $h(r)$ instead of $h(r; \lambda_{max})$.

3.3.3 Wavelength splitting improves *SNR*

As has been shown, fluorescence is linked to a spatial broadening effect, which can be overcome when appropriate emission filters are used. However, this comes at the cost of further reducing the number of detectable photons p and therefore limits the achievable *SNR*. The goal is to capture all emitted photons but process them in such a way that the spatial broadening can be reduced, hence *SNR* is improved. To achieve this, the emission spectrum is split into multiple wavelength bands, each corresponding to a single image. Fig. 3.9b shows three such images at peak emission and smaller/longer wavelengths, with (top) and without (bottom) noise. The different semicircle indicate the smallest resolvable structure $d_{xy,min}$ of the spokes target. An improvement in terms of resolution can be seen towards smaller wavelengths, albeit worse *SNR* due to much lower emission in this spectral region (see spectrum at bottom left). The same effect on resolution is verified when the visibility for these three images are plotted together (bottom right). In this simulation, the spectrum has been split into 15 wavelength bands. The idea is to recombine them such, that the best trade-off between resolution and *SNR* is obtained. Weighted averaging in Fourier space is one approach, the corresponding effective OTF $\tilde{h}_{wa,\sigma}$ is plotted in Fig. 3.10a (magenta curve). It is compared to the broadened case $h(r)$ (blue) and to the OTFs describing the wavelength split images $\tilde{h}_{1,14}$. The corresponding PSFs are depicted on the left, indicating the reduced width ($\approx$ 7 nm with respect to $h(r)$) for the splitting & recombination approach.

When looking at the OTFs in Fourier space it can be seen that the weighted averaged result shows an improved transfer strength for all frequencies. This enhancement gets stronger the closer k gets to the cutoff $k_{xy,max}$, as indicated by the spectral improvement factor *IF*, which reaches values up to 200%. Table 3.4 compares the *FWHM* values of the conventional $\Delta r_{\mathcal{B}}$ (no splitting, hence broadened) and of the splitting & weighted averaged recombination approach $\Delta r_{wa,\sigma}$.



a b

Figure 3.10: a PSFs and OTFs comparing when the full emission spectrum is used, with (green & cyan) or without wavelength splitting (blue), and after weighted averaged recombination (magenta). The latter results in an improved information transfer of all spatial frequencies. The spectral improvement factor *IF* is isotropic and reaches values of ≈ 200 %, compared to the broadened PSF. The PSFs have been normalized to unity, so that the change in *FWHM* is more easy to see. b Optical setup of a laser scanning microscope with the ability to simultaneously capture spectral information (such as the Zeiss LSM 880). A laser point illumination and the corresponding point detection are scanned throughout the required FoV. For each scan position a grating disperses the information such that different wavelength bands can be detected in parallel using an array of point detectors. After scanning over the complete FoV, a λ-stack with spatial and spectral information has been acquired. The system essentially acts as a confocal microscope. The pinhole in front of the dispersive element removes out-of-focus light. For this specific experiment the pinhole was fully opened, to be able to observe the effect of reducing the out-of-focus blur. Experiments were performed at the ZAF in Jena with the help of *Katharina Reglinski*.

	NA/n	Δr / (nm)	$\Delta r_{\mathcal{B}}$ / (nm)	$\Delta r_{wa,\sigma}$ / (nm)	$\frac{\Delta r_{\mathcal{B}}-\Delta r_{wa,\sigma}}{\mathcal{B}_{theor.}}$ / (%)
Alexa-488	0.80/1.00	320.81	331.85	328.31	32.07
	1.40/1.52	183.32	189.63	188.10	24.25
DAPI	0.80/1.00	282.49	295.83	288.48	55.10
	1.40/1.52	161.42	169.04	164.95	53.54
mPlum	0.80/1.00	399.94	411.13	403.06	72.12
	1.40/1.52	228.56	234.95	231.93	47.26

Table 3.4: Comparing the *FWHM* of the broadened PSF $\Delta r_{\mathcal{B}}$ with the (effective) recombined result of the wavelength split data $\Delta r_{wa,\sigma}$, for the three different fluorescent structures. The theoretical predicted broadening $\mathcal{B}_{theor.}$ can be canceled by up to ≈ 70%.

The corresponding point-spread-functions are shown in Fig. 3.10a. It can be seen that the theoretical predicted broadening $\mathcal{B}_{theor.}$ can be canceled by up to ≈ 70%, in the case of *mPlum*. Note that an experimental verification of this resolution change, which is on the order of ≈ 2% of the broadened *FWHM* $\Delta r_{\mathcal{B}}$, will be difficult. However, a more sophisticated image reconstruction method, such as multiview deconvolution will show a stronger enhancement. A result in terms of visibility for the multiview deconvolution (35 accel. iterations, *NCC* curve in p. 111) is shown as the magenta curve in Fig. 3.9a for DAPI. An improved contrast at the very high spatial frequencies, which is an indicator for the improved *SNR*, can be seen. Note that in the aforementioned evaluation only considers the imaging performance in the lateral direction. A 3D evaluation is omitted here. The wavelength-scaling of the spatial resolution is also true along the axial direction, which will lead to less out-of-focus contribution.

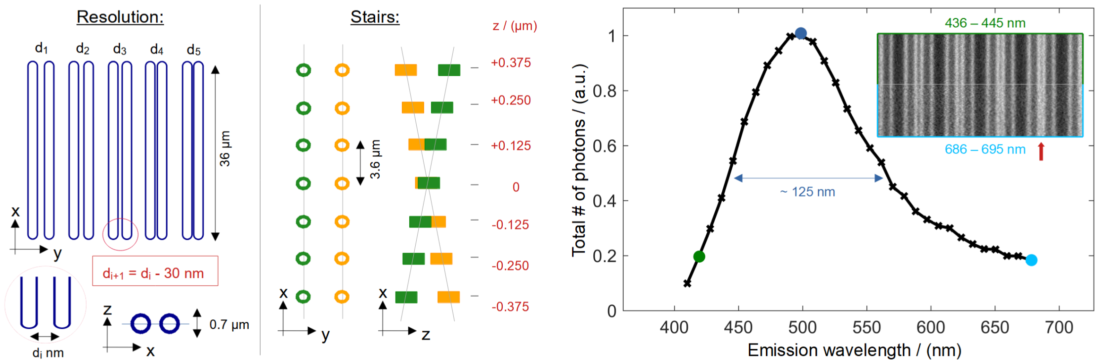
3.3.4 Experimental realization

To experimentally realize the wavelength splitting approach a more elaborate optical imaging setup is required, as spatial and spectral information have to be collected simultaneously. There exist techniques such as snapshot spectral imaging [75], which use multi-lens arrays to obtain spatially sampled spectral data. However, for the purpose of implementing a proof-of-principle experiment, a commercially available alternative was used: the laser scanning microscope LSM 880 from Carl Zeiss AG [76] with the spectral detection unit QUASAR [77]. A schematic diagram of its working principle is shown in Fig. 3.10b. The sample is excited using a focused laser spot, which is scanned over the complete FoV to yield the recorded image. The emitted signal at each point is captured ($NA = 1.4$, $n = 1.52$), de-scanned, focused through the detection pinhole and collimated with an additional lens. A grating is placed in the collimated beampath and disperses the light onto a linear array of point detectors. Each of those is therefore able to detect the image information at a different wavelength band and together yields a λ-stack of spatial and spectral information at a very high splitting efficiency (almost no loss compared to the non-split case). This system is intended to be used as a confocal microscope. Hence, the effective PSF, without any additional smearing along λ, can be written as (see p. 121 in the appendix):

$$h_{Conf.}(r;\lambda) = \Big[h_{Det.}(r;\lambda) \otimes \mathrm{p}(r)\Big] \cdot h_{Illu.}(r) \tag{3.43}$$

with $h_{Det.,Illu.}$ being the detection & illumination (widefield) PSF and p the pinhole. The purpose of the pinhole is to reject out-of-focus light, unfortunately by closing it the number of detectable photons and hence the *SNR* will be reduced. To verify the ability to improve imaging performance (in terms of *SNR*) a sample with known spatial structure is needed. In this thesis a fluorescent calibration target from Argolight [78] (*Argo-SIM*, Argolight SA, France) has been used. Pattern *E* consists of meander shaped lines with changing separation (= distance between adjacent lines) ranging from 0 to 390 nm (see schematic in Fig. 3.11a).

The axial performance of the wavelength splitting approach, is tested using pattern *I* of the Argolight calibration target. It consists of two crossing stairs, with



a b

Figure 3.11: a Lateral and axial schematic of the Argolight target. Left: line pairs (pattern *E*) with decreasing separation, from left to right, in 30 nm steps; right: two crossing stairs (pattern *I*) with a step size of 0.125 μm. b Fluorescence emission characteristics of the Argolight sample. The acquired λ-stack, shows the resolution target (pattern *E*). A visual comparison between the small (green) and large (cyan) wavelength region is shown in the small inset, indicating the wavelength dependent resolution scaling (compare line pairs at red arrow).

steps made out of open cylinders. A schematic drawing of lateral and axial view is shown in Fig. 3.11a right. This pattern enables to verify the 3D imaging performance of a system, as the same object (stair step) is located at different depths (stair step size $= 0.125$ μm), hence get more and more out-of-focus. Achieving a larger axial resolution would mean that only stair steps close to the focus plane of the detection system are visible in the recorded 2D image, while the rest is dimmed or not visible at all.

For wavelength splitting to work, an asymmetric emission spectrum of the sample under investigation is required. When calculating the total sum of the collected fluorescent signal per λ-channel (in a single time frame of pattern *E*), an asymmetric emission curve (see Fig. 3.11b) can be observed. Which resembles the shape of the fluorescence emission curves shown in Fig. 3.6a. When looking at the two images in the inset of Fig. 3.11b, corresponding to a small (green) and large (cyan) wavelength band, the change of resolution capability with λ can be seen (especially at the line pair marked with a red arrow). The *FWHM* of the captured fluorescence emission is at ≈ 125 nm and the overall shape of the emission curve indicates some asymmetry.

Experimental data for the wavelength splitting has been acquired and analyzed. Unfortunately the pinhole of the confocal system was opened, which reduces the λ-dependency of $h_{Conf.}$. Hence, the results are not fully conclusive, as the required modulation of image information with λ is strongly reduced. Nevertheless, the experimental findings are presented on p. 122 in the appendix.

3.4 Separation via pupil splitting

Fluorescence microscopy is a form of incoherent imaging, meaning that its performance is characterized by an OTF which itself can be related to the pupil of the detection system. Another approach to realize splitting & recombination is to spatially split the pupil of an incoherent imaging device, as the pupil somewhat relates to spatial frequency information of the object structure under investigation. The modification of the pupil will affect how different frequencies are transferred in each sub-image, creating the required mutually differing *SNR*. In this thesis we investigate *radial* splitting, corresponding to separating the full pupil (= widefield imaging) in an inner *disk* and outer *ring* pupil. In the following the influence of splitting on the transferable information content in the different sub-images is being studied. Note that the original idea of pupil splitting was conceived by *Rainer Heintzmann* and first investigated by *Polina Feldmann*. Whereas my contribution has been a further detailed analysis and experimental realization.

3.4.1 Information content in the sub-images

The information transfer of an incoherent imaging system is characterized using the PSF in real or OTF in Fourier space. The latter can be obtained directly from the pupil $\mathcal{P}$ as the *autocorrelation* operation $\mathcal{A}$:

$$\tilde{h}(\vec{k}) \;=\; \mathcal{A}\left\{\mathcal{P}(\vec{k})\right\} = \int_{-\infty}^{+\infty} d\vec{k}'\; \mathcal{P}^*(\vec{k}' - \vec{k}) \cdot \mathcal{P}(\vec{k}') \tag{3.44}$$

In conventional widefield imaging, the 3D (McCutchen) pupil $\mathcal{P}$ is given as a spherical cap with radius R_{max} [33], corresponding to the detectable cone of light of the detection system (see Fig. 2.2b). Hence R_{max} is given according to:

$$R_{max} = f_{obj} \cdot NA \tag{3.45}$$

The corresponding OTF (and PSF) are shown in Fig. 3.12a as the blue curve.

For low NA this spherical cap can be approximated by a 2D disk, corresponding to a WF detection pupil $\mathcal{P}_{WF}$. When a radial split is used, the full pupil is separated into two components: inner disk $\mathcal{P}_{Disk}$ and outer ring $\mathcal{P}_{Ring}$ pupil.

$$\tilde{h}_{WF}(\vec{k}) = \mathcal{A}\left\{\mathcal{P}_{WF}(\vec{k})\right\} = \mathcal{A}\left\{\mathcal{P}_{Disk}(\vec{k}) + \mathcal{P}_{Ring}(\vec{k})\right\} \tag{3.46}$$

Note that this equation also holds for high NA imaging, nevertheless we will use

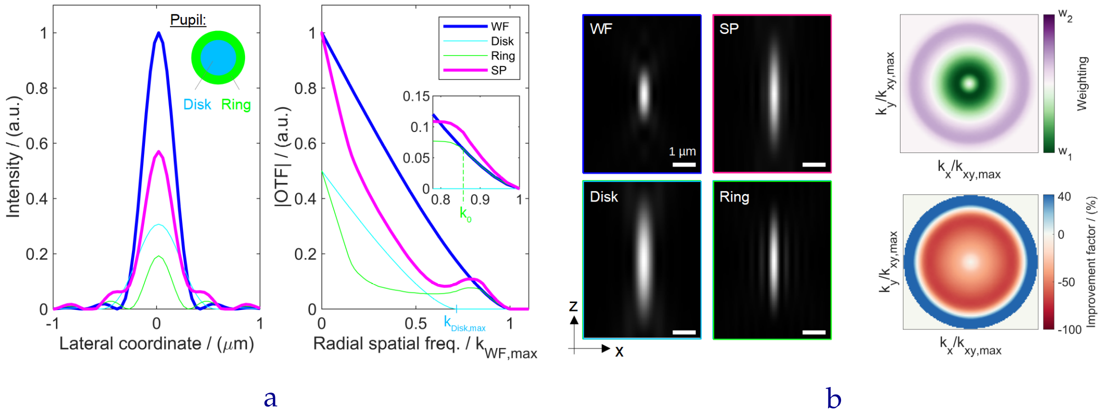


Figure 3.12: a PSF (left) and OTF (right) in case of pupil splitting. The inner disk and outer ring curves are shown in cyan and green respectively. Note that high spatial frequency information can be transferred with the same signal strength as in the WF case (blue), using the ring pupil. However with reduced noise, as the number of detected photons has been decreased to 50% for a splitting radius $R = \sqrt{0.5} \cdot R_{max}$ (neglecting the additional factor due to the changing collection efficiency as described in 2.4.1). Weighted averaging (magenta) of the two sub-images leads to an enhancement of the *SNR* towards the highest spatial frequencies. The corresponding PSF show that $h_{wa,\sigma}$ has a reduced peak and FWHM value. b Left: splitting the pupil additionally leads to a change in the depth-of-focus (DoF) of the detectable information as both sub-PSFs (inner disk and the outer ring), are enlarged along the axial direction. After recombination this yields in an extended DoF (more details on p. 127 in the appendix). Right: a visualization of the weights (top) shows that high frequency information solely originates from the outer ring sub-image, when being recombined using weighted averaging in Fourier space. The improvement factor indicates that the enhancement is isotropically distributed and reaches values of ≈ 40% in k-space.

the terms $\mathcal{P}_{Disk}$ and $\mathcal{P}_{Ring}$ to describe the split sub-pupils. Although in 3D the sub-pupils are not really "disk" or "ring"-like shaped.

When the autocorrelation integral is solved, the contributions of both sub-pupils to the widefield OTF can be expressed as:

$$\tilde{h}_{WF}(\vec{k}) = \mathcal{A}\left\{\mathcal{P}_{Disk}(\vec{k})\right\} + \mathcal{A}\left\{\mathcal{P}_{Ring}(\vec{k})\right\} + 2\left[\mathcal{P}^{*}_{Disk}(\vec{k}) \otimes \mathcal{P}_{Ring}(\vec{k})\right] \tag{3.47}$$

Which of course can be written in terms of the sub-OTFs $\tilde{h}_{Disk,Ring}$, according to:

$$\tilde{h}_{WF}(\vec{k}) = \tilde{h}_{Disk}(\vec{k}) + \tilde{h}_{Ring}(\vec{k}) + 2\left[\mathcal{P}^{*}_{Disk}(\vec{k}) \otimes \mathcal{P}_{Ring}(\vec{k})\right] \tag{3.48}$$

Meaning that the non-split OTF $\tilde{h}_{WF}$ is not just the sum of the two sub-OTFs but contains an additional term directly related to the two sub-pupils.

Knowing this, $\tilde{h}_{Ring}$ can be calculated by rearranging previous equation into:

$$\tilde{h}_{Ring}(\vec{k}) = \tilde{h}_{WF}(\vec{k}) - \tilde{h}_{Disk}(\vec{k}) - 2\left[\mathcal{P}^{*}_{Disk}(\vec{k}) \otimes \mathcal{P}_{Ring}(\vec{k})\right] \tag{3.49}$$

A different *graphical* representation of how to calculate $\tilde{h}_{Ring}$ is presented in the appendix (p. 125). Because of the splitting, the region-of-support of $\tilde{h}_{Disk}$ is reduced to $|\vec{k}| \leq k_{Disk,max}$, given by the pupil splitting radius R:

$$k_{Disk,max} = \frac{R}{R_{max}} \cdot k_{WF,max} \tag{3.50}$$

Which is depicted in Fig. 3.12a right where the cyan curve is ≥ 0. The split shown is computed for $R = \sqrt{0.5} \cdot R_{max}$, corresponding to a split with equal area of both sub-pupils. Note that for a value of $|\vec{k}| \geq k_0 = 1/2 \cdot (k_{Disk,max} + k_{WF,max})$, the additional convolution term in eq. 3.49 vanishes. Hence the OTF corresponding to the ring pupil for $|\vec{k}| \geq k_0$ is given as :

$$\tilde{h}_{Ring}\left(|\vec{k}| \geq k_0\right) = \tilde{h}_{WF}\left(|\vec{k}| \geq k_0\right) \tag{3.51}$$

Meaning that the transfer of very high spatial frequency information is solely achieved by the ring pupil itself. This effect is depicted in Fig. 3.12a right in green. Indicating that both sub-images carry mutually differing information, which can be used to improve the *SNR* via subsequent computational recombination. Note that the additional effect of reducing the collection efficiency, as described in sec. 2.4.1, has been neglected here.

The weighted averaged (effective) OTF is shown in magenta in Fig. 3.12a, indicating a strong improvement towards the cut-off frequency. This is because for $|\vec{k}| \geq k_0$, the ring pupil transfers the same information as $\tilde{h}_{WF}$ and only requires half of the available photons budget, reducing shot noise. The corresponding weighting is shown in Fig. 3.12b at the top right. Low to mid-frequency information is mostly carried by $\tilde{h}_{Disk}$, high frequency content only by $\tilde{h}_{Ring}$. When looking in real space, the corresponding effective PSF (magenta) shows a reduced peak intensity and *FWHM*. Also both sub-PSFs differ in peak and *FWHM* value, which indicates the capability of pupil splitting to create mutually differing image information. Splitting the pupil does not only affect the lateral performance of each sub-image, it also changes the depth-of-field (DoF), see Fig. 3.12b left. It is known that an annular pupil creates a "Bessel-like" beam, an elongated light distribution with strong sidelobes [79]. The inner disk pupil also creates an elongated PSF (cyan), as it corresponds to imaging with an *NA* reduced by the factor R/R_{max}. When both sub-images are recombined, the effective PSF (magenta) will also show an extended DoF. The elongation varies with R/R_{max} and is described in more detail in the appendix (p. 127). In case of the splitting shown in Fig. 3.12b ($R/R_{max} = \sqrt{0.5}$), the axial extend is roughly twice that of the widefield case.

3.4.2 Evaluating the achievable *SNR* improvement

As the magenta curve in Fig. 3.12a suggests, it is possible to recover high spatial frequency information with better *SNR* when pupil splitting is employed. However, this comes at the cost of a reduced *SNR* at the mid-frequencies. When looking at the spectral improvement factor *IF* in Fig. 3.12b, a maximum improvement IF_{max} of $\approx 42\%$ is reachable for the equal area split that was simulated. The "negative improvement" reaches a value of $IF_{min} \approx -68\%$, making it seem that more is lost than gained with pupil splitting. Nevertheless, the important point is that $IF_{min} \neq -100\%$, meaning that the transferable object information is not lost, only reduced in terms of transfer strength. This can be improved upon by image reconstruction in post-processing, whereas lost information is inevitably gone.

In general, the performance of pupil splitting varies with the splitting ratio given by R/R_{max}. Four different scenarios have been simulated and their respective spectral improvement factors are shown in the appendix (p. 126) and in Tab. 3.5.

R/R_{max}	0.25	0.50	0.75	0.95	$\sqrt{0.5} \approx 0.707$
IF_{max}	3.20	15.26	51.10	201.17	41.65
IF_{min}	-21.05	-57.32	-68.95	-68.72	-68.48
$IF > 0$	63.16	50.08	31.27	8.18	35.13
$\mathcal{D}_{wa,\sigma}/\mathcal{D}$	90.95	68.09	58.25	85.94	57.04

Table 3.5: Different performance measures for changing the pupil splitting, indicated by R/R_{max}. All values stated are given in *percent*. $IF_{max,min}$: maximum positive/negative improvement; $IF > 0$: positive improvement region; $\mathcal{D}_{wa,\sigma}/\mathcal{D}$: ratio of detectability, 100% represents the WF-case.

It can be seen that absolute value of IF_{max} and IF_{min} increases with R/R_{max}. Indicating that a strong enhancement can be achieved with decreasing width of the ring pupil. Note also that a large *IF* is only achievable at high spatial frequencies and over a limited region in *k*-space, stated by $IF > 0$. At $R/R_{max} = \sqrt{0.5}$ only 35% of all spatial frequencies within the bandlimit can be improved, compared to the WF case. Nevertheless this splitting ratio was accepted as it yields an acceptable tradeoff with respect to IF_{max}. A large improvement can only be achieved for a spatial frequencies close to the cutoff $k_{xy,max}$. Interestingly the "negative improvement" does not get much worse for values $R/R_{max} > \sqrt{0.5}$, however also relates to further limiting the measurable photons in the sub-image corresponding to the ring pupil (much narrower ring). Suggesting that for large enough photon numbers, a splitting $R/R_{max} > \sqrt{0.5}$ might yield stronger enhancements. Note that the ratio of detectability is always $< 100\%$ ($\mathcal{D}_{wa}/\mathcal{D} = 100\%$ represents the WF case) and $\mathcal{D}_{wa}/\mathcal{D}$ decreases towards larger splitting radii.

Figure 3.13 top shows the simulated results of imaging a line pattern, with increasing spatial frequencies towards the bottom, using the four different pupil settings: widefield (blue), split pupil with weighted average recombination (magenta) and the two sub-images corresponding to disk & ring pupil (cyan & green). The individual PSFs where simulated for $NA = 0.8$, $\lambda = 520$ nm and shot noise with 1000 photons in the brightest pixel of the WF image. A more detailed description on how h_{Ring} was calculated (using [32]) is given in the appendix (p. 128). It is clear that the sub-images are about half as bright as the WF case, as each only carries half of the detectable photons. Note how the image corresponding to $\mathcal{P}_{Ring}$ shows a strong modulation of the patterns corresponding to high spatial frequencies (yellow frame). When recombining the two sub-images using weighted averaging, this translates into an improved performance at those frequencies. Because the recombination makes use of all detected photons, the resulting image shows the same brightness compared to the widefield case.

3.4.3 Splitting via reflection in the BFP

To experimentally realize the pupil splitting approach, access to the BFP of an detection objective in a widefield microscope is needed. As the BFP often is located within the objective, it must be re-imaged by placing an additional lens a focal length away from the tube lens of the detection system. This will generate a (de-) magnified version of the BFP in the Fourier plane of the added lens. In the laboratory of *Jonas Ries* at the EMBL (European Molcecular Biology Laboratory, Heidelberg), such an optical system ($NA = 1.2$) has been built for enhancing the axial localization precision in single molecule localization microscopy (SMLM) [80, 81] experiments. In their work they split super- from undercritical angle fluorescence, to get additional information on the z-position of blinking fluorescent markers, by looking at ratiometric changes of the photon emission [82]. The splitting itself is achieved by placing a small mirror, mounted on top of a glass substrate with antireflection coating in the re-imaged BFP position (see Fig. 3.13, top right). Discerning light in the BFP which is either reflected by the mirror or transmitted through the glass substrate.

Note that the mirror device was manufactured at the Leibniz IPHT (Jena) and consists of a 65 nm Al_2O_3 protective on top a 200 nm thick silver layer. Four different mirror geometries (elliptical shape due to a tilt in the BFP) were put on a single 5 mm thick glass substrate. The mirror itself is tilted ($\approx 10^\circ$) so that it will reflect the inner part of the light distribution (splitting ratio $R/R_{max} \approx 68\%$), corresponding to the *disk* pupil, which gets imaged to one half of a camera (pco.edge

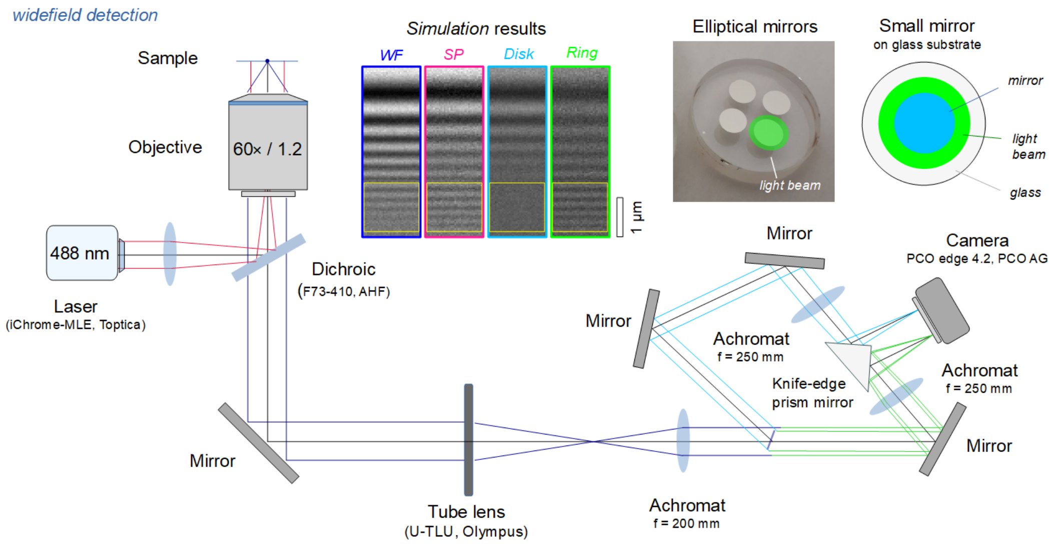
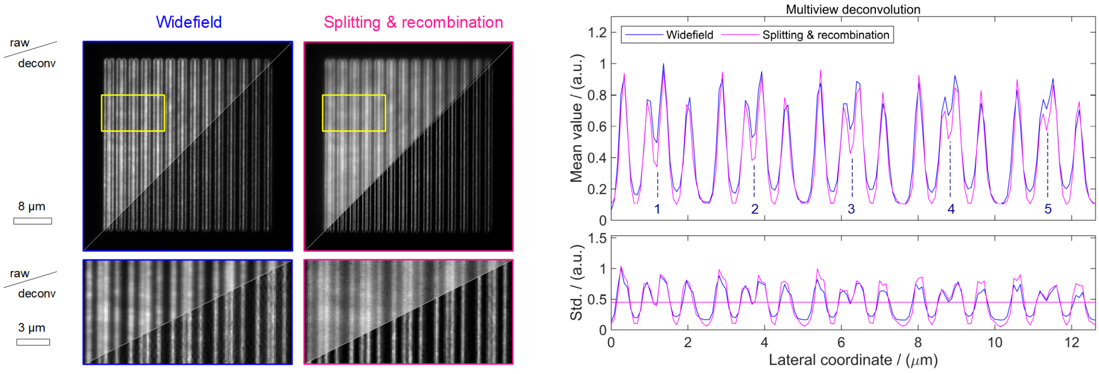

widefield detection
Sample
Objective
60× / 1.2
Simulation results
WF
SP
Disk
Ring
1 µm
Elliptical mirrors
light beam
Small mirror
on glass substrate
mirror
light beam
glass
488 nm
Laser
(iChrome-MLE, Toptica)
Dichroic
(F73-410, AHF)
Mirror
Camera
PCO edge 4.2, PCO AG
Mirror
Achromat
f = 250 mm
Knife-edge prism mirror
Achromat
f = 250 mm
Mirror
Mirror
Tube lens
(U-TLU, Olympus)
Achromat
f = 200 mm

Figure 3.13: Optical setup used for pupil splitting. The BFP of a widefield objective ($NA = 1.2$) is made accessible by placing a lens one focal length away from the tube lens of the imaging system. The pupil is split using a small mirror which is mounted on a glass substrate (top right) and placed in the position of the re-imaged BFP. It reflects the inner disk (cyan) of the light beam so that the corresponding image can be captured on a camera, using an additional lens. The information related to the ring-shaped pupil (green, transmitted) is simultaneously imaged onto the same camera, so that both sub-images are being accessible for computational recombination. A simulation (inset) compares the imaging ($NA = 0.8$) of the same line pattern with the four different pupil settings: widefield (WF), split pupil (SP) and the two sub-images (disk & ring). The yellow frame indicates high spatial frequency information which is transferred better by pupil splitting & recombination. Experiments were performed at the EMBL in Heidelberg in collaboration with *Takahiro Deguchi* and *Jonas Ries*, see [80, 81] for more details on the optical setup.

4.2, PCO AG, Germany) using an achromatic lens ($f = 250$ mm). Light which does not impinge on the mirror is transmitted through the glass (*ring* pupil) and imaged on the second half of the camera (using an achromat with $f = 250$ mm). In this way it is possible to capture two images simultaneously, while maintaining a large number of detectable photons due to the high reflection efficiency ($> 94\%$ in the visible range [80]). The necessary registration of both sub-images has been achieved by employing a cross-correlation based algorithm [83].
Again, the Argolight sample [78] (same as in sec. 3.3.4) was used to test the pupil splitting & recombination technique. The illumination is realized using a laser at an emission wavelength of $\lambda = 488$ nm (iChrome-MLE, Toptica, Germany) and the fluorescence is detected behind a quad-band beamsplitter (F73-410, AHF, Germany). A fair comparison of pupil splitting and widefield imaging (reference case) was achieved by setting the exposure time of the detector and illumination intensity to a fixed value. While the only change made to the system is whether the mirror device was placed in the BFP, or not.

3.4.4 Experimental results

To experimentally test the imaging performance of pupil splitting, the resolution target (pattern *E*) of the Argolight calibration slide (see sec. 3.3.4 for more details) were imaged, and the results are shown in Fig. 3.14a.



a b

Figure 3.14: a Comparing the images of the resolution target (pattern *E*) of the Argolight sample [78], using conventional widefield detection (blue) and pupil splitting & recombination (magenta). Depicted is the average of 100 frames before (top) and after (bottom) deconvolution. The individual number of iterations to reach the optimum reconstruction is given in the appendix (p. 124). The yellow frame indicates a region which is shown enlarged at the bottom. b Line profiles of both noise-normalized (see standard deviation plot) image results. A stronger modulation or visibility $\mathcal{V}$, for the mean value of the 100 image frames, can be seen for the pupil splitting approach, which is also depicted in Table 3.6. The weighted averaged result is given in Fig. A.10b.

For both datasets, with (magenta) and without mirror (blue), a time series of 100 images was acquired. Figure 3.14a shows the average over all frames for both imaging methods, before (top) and after (bottom) deconvolution. The necessary number required to reach the optimum reconstruction results are found by computing the *NCC* values for each iteration in a simulation (see p. 124 in the appendix). For the widefield images (blue) 24 (accelerated) iterations and for the split pupil data 18 (accelerated) iterations were performed. The PSF needed for the reconstruction was computed according to the scheme described in the appendix (p. 127), for the given experimental parameters. The yellow frame indicates a region which is enlarged on the bottom, also before and after deconvolution. The weighted averaged recombined result (Fig. 3.14a magenta top) shows some blurring, while line pairs with a very small gap can be resolved better than in the widefield case. Exemplifying the shape of the magenta curve in Fig. 3.12a, medium spatial frequencies are transferred worse (= blurring), high spatial frequencies benefit from pupil splitting (= resolving smaller structures).

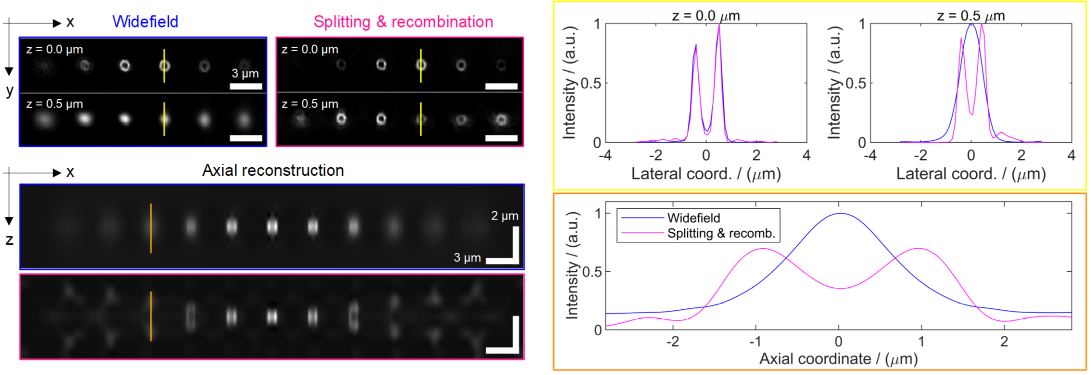
As described in the appendix (for wavelength splitting p. 122), the imaging performance of both methods (widefield & split pupil) is evaluated by calculating the pixel-wise average (Fig. 3.14b top) and standard deviation (bottom) of the

100 image frames, to ensure that the mean standard deviation (vertical line) of both imaging modalities are equal. A line profile through the center of the yellow frame indicates the modulation depth of the widefield (blue) and split & recombined (magenta) results. The curves for recombining the split data using multiview deconvolution is shown in Fig. 3.14b, while the weighted averaged approach is given in the appendix (p. 128). In order to evaluate the imaging performance in more detail, the visibility $\mathcal{V}$ and the spatial improvement factor $IF_{\mathcal{V}}$ have been analyzed and are given in Tab. 3.6.

	Line pair #	1	2	3	4	5
	Spacing / (nm)	390	360	330	300	270
	$k_{xy}/k_{xy,max}$	0.56	0.60	0.66	0.72	0.80
WA	Widefield	0.0595	0.0420	0.0269	0.0000	0.0000
	Split pupil	0.0424	0.0372	0.0308	0.0192	0.0139
	Improvement / (%)	*-28.69*	*-11.41*	*14.36*	∞	∞
MV	Widefield	0.2829	0.2383	0.1687	0.1225	0.0832
	Split pupil	0.4314	0.3713	0.2980	0.2381	0.1458
	Improvement / (%)	*32.40*	*37.00*	*53.28*	*74.15*	*52.21*

Table 3.6: The visibility $\mathcal{V}$ for the five different line pairs marked in Fig. 3.14b. Reference data is obtained by removing the mirror in the re-imaged BFP. The recombination was either done by weighted averaging (WA) or multiview deconvolution (MV). The spatial improvement factor $IF_{\mathcal{V}}$ is defined in eq. 3.14. The ratio $k_{xy}/k_{xy,max}$ is calculated for $NA = 1.2$ and $\lambda_{max} = 520$ nm.

When looking at the visibility improvement $IF_{\mathcal{V}}$ after weighted averaging, it can be seen that two of the five line structures are imaged with worse modulation depth. These two correspond to a relative spatial frequency $k_{xy}/k_{xy,max} < k_0 = 0.68$, which fall into the region of negative improvement (IF shown in Fig. 3.12b). Interestingly, the same is true for the line pair with spacing of 330 nm, nevertheless it exhibits a visibility improvement of $\approx$ 14%. A reason for this might be the aplanatic factor and its modification of the effective splitting ratio, hence modifying k_0 to ≈ 0.61 (see p. 129 in the appendix for more details). The even more closely spaced line pairs cannot be resolved in the widefield detection at all while the weighted averaged result still shows a small modulation. A "∞"-symbol was used in Tab. 3.6 to indicate the increase of effective resolution, achievable through SNR enhancement. When the deconvolved results are compared, the modulation depth of all line pairs is enhanced, and in general the improvement increases towards finer structure. The maximum enhancement of 74%, is reached for a line pair separation of 300 nm, instead of 270 nm. This might be attributed to misalignment when both sub-images are being registered to each other. Overall, this indicates that the more elaborate post-processing using multiview deconvolution, enables pupil splitting & recombination to transfer sample information more robustly than widefield imaging. The axial performance is similarly tested as in sec. 3.3.4, using the "crossing stairs" sample (pattern *I*, step size of 0.25 μm) of the Argolight calibration target. Both datasets (split & non-split) have been processed using a thick-slice reconstruction (described in sec. 2.4.2) using 18 (accelerated) iterations for both. The theoretical PSF was computed with 11 axial planes and the reconstructed volumes were upsampled by a factor of 12 for visualization. Due to the slight extended-depth-of-field (EDoF) effect of pupil splitting, the stairs of the test target stay in focus over a larger axial distance. While the normal widefield data is out-of-focus immediately, depicted in Fig. 3.15a top.



a b

Figure 3.15: a 3D reconstruction of the imaged stair sample (pattern *I*, step size 0.25 μm). Top: in-focus and out-of-focus ($z = 0.5$ μm) xy-slices for conventional widefield detection and the pupil splitting approach. Bottom: Axial view of the thick slice reconstruction from a single focal plane exposure, showing that the stair-like shape of the sample can be reconstructed using the split pupil data (11 slices upsampled by a factor of 12 for visualization). b Line profiles through images in a. Top: comparison of resolving the cylindrical structure at the in-focus and at an out-of-focus plane, which is only possible using the split pupil. Bottom: showing the axial reconstruction indicating the ability to reconstruct the stair-like shape of the pattern from 2D images only.

A line profile through one of the stair steps at an in-focus ($z = 0$ μm) and out-of-focus ($z = 0.5$ μm) slice of the reconstruction is shown in Fig. 3.15b top (yellow). In-focus (left), both detection schemes are able to visualize the cylindrical shape of the individual stair steps. When going out-of-focus (right) only pupil splitting maintains the central dip in the curve. More interestingly, pupil splitting is able to recover the stair-like nature of the test sample which is shown in Fig. 3.15a bottom, in contrast to conventional widefield detection. The two arms of the crossing stairs sample (shown in orange and green in Fig. 3.11a) can be reconstructed using pupil splitting, only a mirror image appears (above and below the focus plane). Also see Fig 3.15b, which shows a line profile through one of the out-of-focus stair steps.

3.5 Illumination splitting through Field-Synthesis

So far, the splitting has only been realized in the *detection* of a widefield microscope. Where it is crucial to capture as many photons as possible, to obtain sufficient *SNR*. When splitting a detected fluorescence signal, the additional redistribution of photons into multiple sub-images becomes problematic. This is because at low photon numbers it is possible that image information in the sub-images falls below the noise floor (in Fourier space), making it inevitably lost for further processing. Instead of splitting the detection of an imaging system, it might be beneficial to split the illumination. Where many photons are available so that the redistribution does not necessarily come at any cost in photon loss. Of course, when a larger number of photons is available the impact of shot noise on the image quality is marginal (large photon numbers = good *SNR*). Nevertheless, the problem of out-of-focus blur in widefield detection still needs to be addressed.

Hence, in this chapter the idea is to split the illumination of a light-sheet system, capture the corresponding sub-images and recombine those to generate an *SNR* improvement in thick and dense samples, by removing out-of-focus blur. Note that this idea is not new, e.g. in structured illumination microscopy (SIM) [84, 85, 86], the illumination of a widefield microscope is split into multiple shifted sinusoidal patterns and the recorded sub-images are recombined to yield a reconstruction with improved spatial resolution. Hence illumination splitting seems to be promising and is applied here in a different way to light-sheet microscopy.

Widefield detection is not able to remove out-of-focus contributions, which can strongly deteriorate the imaging quality. This led to the (re-) invention of light-sheet microscopy [41, 87], where the sample is illuminated with a thin sheet of light from the side, overlapping with the focal plane of the detection objective. Figure 3.16, bottom right shows that this type of optical arrangement not only automatically removes out-of-focus information, as the fluorescence signal can only be emitted from a plane defined by the illumination sheet. It also distributes the excitation light very effectively, only in a small volume where the detection has its optimal focus. Hence light-sheet microscopy is regarded a very gentle method with respect to phototoxicity and photobleaching [88]. Different type of light-sheet geometries have been investigated: Gaussian [89], Bessel [90], Airy [91] and lattice (square and hexagonal) [92]. The overall goal in "light-sheet engineering" is to create a long and thin illumination structure, so that excitation only takes place in the focal plane of the detection objective, while a large FoV can be investigated. Generating arbitrary light-sheets can be achieved by scanning a focused laser spot in one (lateral) direction [93]. A mask in the BFP of the illumination

objective determines what type of light-sheet is being generated, e.g. an annular mask creates a Bessel focus and therefore yields a time averaged "Bessel-sheet" (see Fig. 3.14, top row). However, recently a new method for scanned light-sheet generation has been found: *Field-Synthesis* (FS) [94, 95].
In this chapter we want to use FS to realize illumination splitting, by generating a set of images with modulated axial image information. Subsequent computational recombination of this data will yield a suppression of out-of-focus information, compared to light-sheet imaging without illumination splitting.

3.5.1 Scanned light-sheet generation

The conventional way to create a light-sheet is by scanning a laser focus along the x-axis (see Fig. 3.16 top), termed *beam scanning* (BS). Note that the z/y coordinate is defined to be the optical axis of the detection/illumination objective, respectively. In beam scanning, an electric field distribution in the BFP $\tilde{\mathcal{E}}$ is focused using an objective, yielding the respective electric field $\mathcal{E}$ in sample space. The corresponding *intensity* is given as $|\mathcal{E}|^2$, which after horizontal scanning (along the x-axis) yields an effective light-sheet I_{BS}. Mathematically this is written as a convolution of $|\mathcal{E}|^2$ with a line along the x-axis, modeled as $\delta(y,z)$:

$$I_{BS}(x,y,z) = |\mathcal{E}(x,y,z)|^2 \otimes \delta(y,z) \tag{3.52}$$

Note that the aforementioned convolution basically only acts in one dimension:

$$\begin{aligned} I_{BS}(x,y,z) &= \iiint_{-\infty}^{+\infty} dx'dy'dz' \; |\mathcal{E}(x',y',z')|^2 \cdot \delta(y-y',z-z') = \\ &= \int_{-\infty}^{+\infty} dx' \; |\mathcal{E}(x',y,z)|^2 \end{aligned} \tag{3.53}$$

With the integral showing a smearing effect of $|\mathcal{E}|^2$ along the x-coordinate.
The general idea of *Field-Synthesis* is that the scanned illumination I_{BS} can also be obtained by performing a line scanning in the BFP of the excitation objective. First, lets express $\mathcal{E}(x,y,z)$ in Fourier space using the propagator in eq. 2.13 as:

$$\mathcal{E}(x,y,z) = \mathcal{F}^{-1}_{k_xk_z}\left\{\mathcal{F}_{xz}\left\{\mathcal{E}(x,y=0,z)\right\} \cdot e^{i\sqrt{\left(2\pi\cdot\frac{n}{\lambda}\right)^2 - \left(k_x^2+k_z^2\right)}\cdot y}\right\} \tag{3.54}$$

with $\mathcal{F}_{xz}/\mathcal{F}^{-1}_{k_xk_z}$ indicating a 2D Fourier transform in the xz/k_xk_z-plane.

A laser, focused into a line, is scanned along k_x over the complete mask (e.g. annular) in the BFP (see Fig. 3.16 bottom). The corresponding field in the BFP is

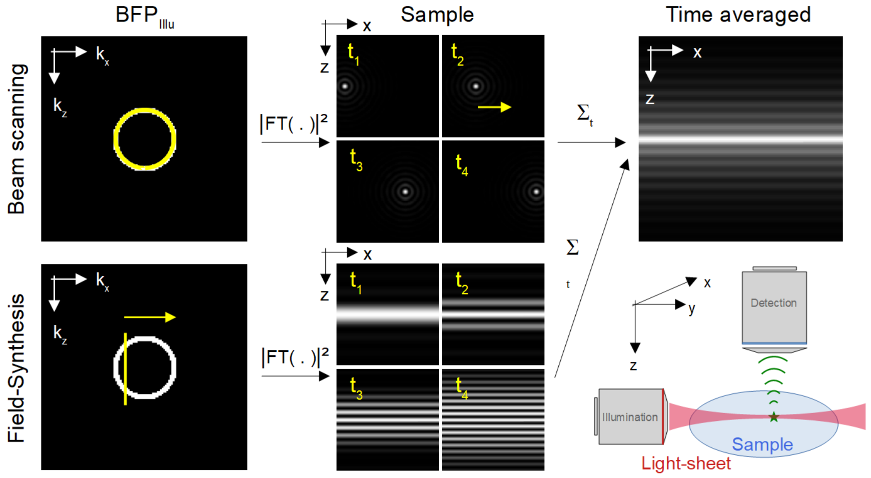


Figure 3.16: Concept of Field-Synthesis: in conventional beam scanning light-sheet generation (top) the mask in the BFP of the illumination objective is fully illuminated and creates a laser focus at the nominal focal point. This spot is laterally scanned (along x), which creates a temporally averaged light-sheet. In *Field-Synthesis* (FS) the same is achieved by scanning a illumination line (along k_x) in the BFP of the excitation objective (bottom). This will result in time-varying sub-illuminations, which when added incoherently will give the same overall illumination distribution as the beam scanning method. A detailed explanation of the Field-Synthesis concept is also shown in the appendix (p. 132). Bottom right: typical optical setup of light-sheet microscopy, with two orthogonally placed illumination & detection objectives.

parameterized as a multiplication with a shifted version of the line illumination:

$$\mathcal{E}_t(x,y,z) = \mathcal{F}^{-1}_{k_x k_z}\left\{\mathcal{F}_{xz}\left\{\mathcal{E}(x,y=0,z)\right\}\cdot e^{i\sqrt{\left(2\pi\cdot\frac{n}{\lambda}\right)^2-\left(k_x^2+k_z^2\right)}\cdot y}\cdot\delta(k_x-t)\right\} \tag{3.55}$$

with t denoting the different scan positions in the BFP. $\mathcal{E}_t$ can be expressed as a convolution $\otimes_{xz}$ along the xz-coordinates:

$$\mathcal{E}_t(x,y,z) = \mathcal{E}(x,y,z)\otimes_{xz}\left[e^{-i\cdot tx}\cdot\delta(z)\right] \tag{3.56}$$

with $\delta(z)$ being the Fourier transform of $\delta(k_x)$. The convolution operation in its integral form is given as:

$$\begin{aligned}\mathcal{E}_t(x,y,z) &= \iint_{-\infty}^{+\infty} dx'dz'\ \mathcal{E}(x',y,z')\cdot e^{-i\cdot t(x-x')}\cdot\delta(z-z') = \\ &= \int_{-\infty}^{+\infty} dx'\ \mathcal{E}(x',y,z)\cdot e^{-i\cdot t(x-x')} = \\ &= e^{-i\cdot tx}\int_{-\infty}^{+\infty} dx'\ \mathcal{E}(x',y,z)\cdot e^{i\cdot t\cdot x'}\end{aligned} \tag{3.57}$$

Note that this is a Fourier transform along the x' coordinate.

$$\mathcal{E}_t(x,y,z) = e^{-i\cdot tx} \cdot \mathcal{F}_{x'}\Big\{\mathcal{E}(x',y,z)\Big\} \tag{3.58}$$

The intensity I_t, corresponding to the respective line-scan in the BFP, is given as:

$$I_t(x,y,z) = |\mathcal{E}_t(x,y,z)|^2 = \left|\mathcal{F}_{x'}\Big\{\mathcal{E}(x',y,z)\Big\}\right|^2 \tag{3.59}$$

Within a single exposure, the detector adds up all those intensities, yielding I_{FS}:

$$I_{FS}(x,y,z) \quad = \quad \int_{-\infty}^{+\infty} dt\ I_t(x,y,z) = \int_{-\infty}^{+\infty} dt\ \left|\mathcal{F}_{x'}\Big\{\mathcal{E}(x',y,z)\Big\}\right|^2 \tag{3.60}$$

Remember that x' and t are Fourier coordinates with respect to each other. Using *Plancherel's* theorem [31], the Fourier transformation can be removed:

$$I_{FS}(x,y,z) \quad = \quad \int_{-\infty}^{+\infty} dx'\ |\mathcal{E}(x',y,z)|^2 \tag{3.61}$$

With this we see that generating the light-sheet using beam scanning or by scanning an illuminating line in the BFP, as suggested in Field-Synthesis, is essentially the same. A more graphical description of the connection between *Field Synthesis* and conventional beam scanning is given in the appendix (p. 132).

An advantage of Field-Synthesis is, that it distributes the overall illumination dose that a specific position in the sample receives, over the whole duration of the scanning process. In contrast to beam scanning, where a specific point in the sample gets all the excitation in a short time period (i.e. when illumination and sample point coincide). FS drastically reduces the maximum illumination that a sample has to endure (see Fig. A.12b) and therefore decreases photodamage [50]. A fundamental tradeoff in light-sheet imaging is that of optimizing FoV and light-sheet thickness, which governs the ability to reject out-of-focus light. The latter is directly connected to the *SNR* for thick & dense samples (see sec. 2.2.3). When a collimated laser beam (top hat) is focused by an objective, the resulting light distribution (along y) first converges, reaches a minimum value (the waist w_0) and then diverges again (see Fig 3.17a). The FoV is given as the distance along the y-coordinate, in which the width of the intensity distribution stays approximately constant at w_0. Depending on the *NA* of the objective, w_0 can be reduced. But so is the FoV, meaning that the ability to remove out-of-focus light and the size of the FoV are directly linked to each other (see $d_{xy,min}$ and $d_{z,min}$ in sec. 2). In Gaussian optics this is quantified via an equation connecting the *Rayleigh length* z_R (distance from waist until beam cross-section has doubled) with the beam waist w_0 [29]:

$$z_R = \frac{\pi \cdot w_0^2}{\lambda} \tag{3.62}$$

Hence it is not possible to achieve a large FoV and a very thin light-sheet, by focusing light through an objective lens.

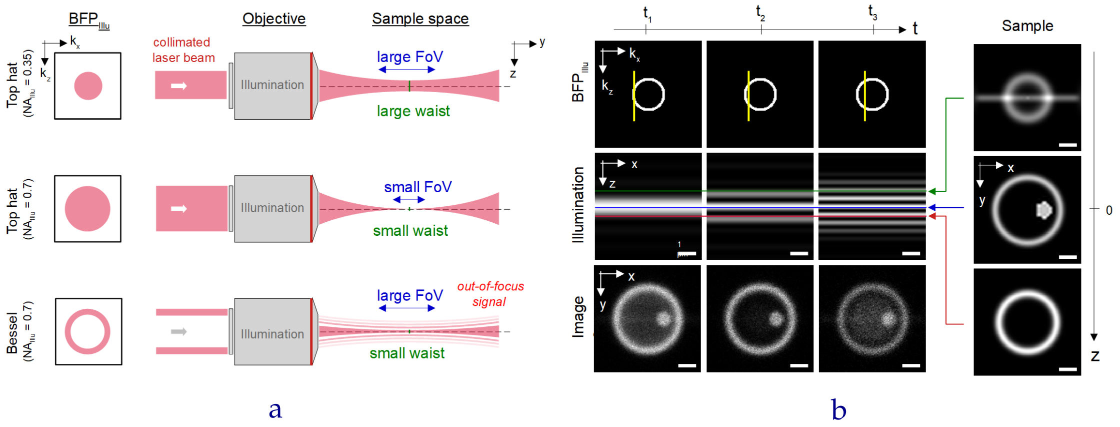


Figure 3.17: a Creating a light-sheet by focusing a collimated laser beam (top hat in BFP) using an objective will result in a specific detection FoV and light-sheet thickness. The ability to remove out-of-focus light depends on latter and can only be improved at the cost of a reduced FoV. When an annular mask is inserted into the BFP, the resulting illumination pattern consists of a thin & long central peak, surrounded by strong sidelobes. The latter excites fluorescence in planes outside of the detection focal plane, hence leads to a strong degradation of *SNR* for thick & dense samples (due to out-of-focus blur). b Creating time-varying sub-illuminations using Field-Synthesis, leads to a modulation of axial image information. To visualize this effect the resulting images of a simulated 3D spherical object (right) are depicted for three different scan positions in the illumination BFP (bottom). Note how only the plane in-focus (blue) of the detection objective is illuminated in all cases, while the other axial information (green & red lines) gets modulated.

There are different ways how to tackle this problem: one would be to scan a very thin light-sheet in the y-direction and to stitch the different in-focus regions together. This can be achieved by synchronizing the line-readout of a CCD camera with the scan position in the sample, as introduced by Dean et al. [96, 97]. Another idea is to use propagation-invariant beams, like a Bessel or Airy beam, for generating the light-sheet [91]. Those maintain a narrow center over a large FoV, however, always are accompanied by strong side lobes (see Fig. 3.17a). Which result in excitation of sample structure outside of the plane of focus of the detection objective, leading to additional blur and further degradation of the attainable *SNR*. Ways, how to remove this drawback, is to either use a non-linear effect like two-photon absorption, decreasing the sidelobes by effectively squaring the excitation distribution [98]. Or by computational post-processing as it is done in Bessel-beam plane microscopy [90, 92], where structured illumination is used to effectively cancel the side lobes.

3.5.2 Encoding axial information in 2D

As can be seen from Fig. 3.17b, Field-Synthesis enables to capture an full-FoV image for each scan position in the BFP of the illumination objective. This is not possible in beam scanning, where the full FoV is never illuminated at once. Additionally notice that the individual sub-illuminations (in Field-Synthesis) are approximately constant over the detection FoV, but show a distinct modulation along the z-axis (see Fig. 3.16). Only the excitation maximum remains in focus of the detection objective (blue in Fig. 3.17b), whereas out-of-focus information is modulated. This means that while line scanning in the BFP, it is possible to capture images where the axial response changes (due to the illumination modulation), but the in-focus information maintains maximal. Note that the general idea to use FS to improve the imaging performance of a light-sheet system has been simultaneously conceived by myself and *James Manton*, with whom I then collaborated on this project. The aforementioned effect of modulating axial image content has been simulated by imaging the 3D spherical object from sec. 2.2.2, where conventional widefield imaging ($NA_{Det.} = 1.1$) suffers from strong out-of-focus blur. Sample emission at three different z-positions are shown in Fig. 3.17b on the right. As *imaging* is inherently three dimensional, information from a multitude of planes will end up in the 2D image. However, due to the varying illumination introduced through Field-Synthesis, each sub-image will consist of differing axial information. E.g. the image at t_1 shows some considerable intensity in the interior of the spherical object (see the line-like feature from the line in the object being slightly out-of-focus), while at t_2 this information is largely removed. When relating this to the respective sub-illuminations (middle row), we can conclude that the modulated image content cannot originate from the detection focus plane. Based on this principle more information on the axial position of the different object layers (red and green) can be obtained. Interestingly, it is advantageous to use a Bessel light-sheet here, due to its stronger ability to modulate axial information. E.g. due to the annular mask in the BFP, the illumination at t_3 is a high frequency pattern along the z-axis, which corresponds to a strong modulation of axial information. An alternative approach, which is not followed in this thesis, is to generate the light-sheet excitation by beam scanning and recording the fluorescence in an ISM-like detection scheme [99, 100]. Such acquired data might give similar or even better access to axial information, enabling imaging with high *SNR* in thick samples. In the following, Field-Synthesis is used to acquire multiple sub-images that correspond to the time-varying Bessel sub-illuminations shown in Fig. 3.17b. Hence, each sub-image will vary in information content, as the sample excitation is differently axially modulated for each scan-position in the BFP of the illumination objective. The sub-images are then being computational recombined and compared to Field-Synthesis as acquired by recording the total intensity over a complete scan only (equivalent to the beam scanning method).

3.5.3 Rejecting out-of-focus light using FS

As discussed before, the goal is to generate a Bessel light-sheet using the Field-Synthesis concept to remove out-of-focus information by computationally recombining multiple sub-images, while maintaining a large detection FoV. Note that we want to process 2D images only, and leave the opportunity to acquire a z-stack aside. The idea is to remove out-of-focus blur from a 2D image, without requiring to axially scan through the sample. When using FS to generate the excitation, the resulting change of image information is structured along the axial dimensions, with no effect in the xy-plane. Again, weighted averaging in Fourier space is one possible approach to recombine the recorded data (comparable to the reconstruction of sub-images with varying illumination in SIM). In this thesis we will follow a different approach, namely to employ non-negative matrix factorization (nNMF) to find a representation of the acquired data that corresponds to an in- and out-of-focus contribution. Similar to the wavelength & pupil splitting, multiview deconvolution can also be used to reconstruct the underlying (3D) object structure from the illumination split data.

Non-negative matrix factorization:

Non-negative matrix factorization (nNMF) is a technique which tries to find a non-orthogonal basis for a given dataset (e.g. an image), while requiring the factorization to be strictly positive. As typical for matrix factorization schemes (like principal-component analysis PCA [101] and independent-component-analysis ICA [102]), it can be used for dimensionality reduction [103, 104]. Where an underlying, unknown and lower dimensional structure is to be inferred from higher dimensionality data. Using nNMF for image fusion has already been employed in photography [105], especially for recombining multi-focus image data [106]. In the following nNMF is introduced in a mathematical context and then related to the problem of rejecting out-of-focus light by employing the FS concept.
The measured 2D sub-images are denoted as M_i, with i indicating the i-th scan position in the illumination BFP. Each image will be vectorized into $\vec{M}_i$, so that

all measurements can be summarized into a single measurement matrix **M**:

$$\mathbf{M} = \left[\vec{M}_1;\ \vec{M}_2;\ ...\ ;\ \vec{M}_N\right]^\top \tag{3.63}$$

with $^\top$ representing the matrix transpose and N the number of sub-images.
The goal of non-negative matrix factorization is to express **M** as a matrix multiplication of two (non-negative) matrices **V** and **H** of lower dimensionality:

$$\mathbf{M} \approx \mathbf{V}\,\mathbf{H} \tag{3.64}$$

V corresponds to a set of *basis* vectors, that describe the underlying data structure. **H** are the respective *load* vectors, indicating how the different basis vectors superimpose to form the individual measurements $\vec{M}_i$. The measured 2D images M_i consist of P number of pixels, meaning that **M** is a $P \times N$ matrix. The corresponding two matrices **V** and **H** are then characterized by a reduced dimensionality R, which is a parameter to be set by the user of the nNMF algorithm:

$$\mathbf{V} \in \mathbb{R}^{P\times R}; \qquad \mathbf{V} \geq 0 \tag{3.65}$$

$$\mathbf{H} \in \mathbb{R}^{R\times N}; \qquad \mathbf{H} \geq 0 \tag{3.66}$$

An estimation of **V** and **H** is found by minimizing the Frobenius norm F [31]:

$$\begin{aligned} \operatorname{argmin}_{\mathbf{V},\mathbf{H}} \quad & F(\mathbf{V},\mathbf{H}) = \\ = \operatorname{argmin}_{\mathbf{V},\mathbf{H}} \quad & ||\mathbf{M} - \mathbf{VH}||_2^2 = \\ = \operatorname{argmin}_{\mathbf{V},\mathbf{H}} \quad & \sqrt{\textstyle\sum_{i,j} |\mathbf{M} - \mathbf{VH}|_{i,j}^2} \end{aligned} \tag{3.67}$$

which is typically chosen for Gaussian noise, and is assumed to be correct in the application presented here (large enough photon numbers; see eq. 2.25). There exist different iterative schemes [103, 104], which perform the minimization in eq. 3.67 and find a solution to the aforementioned matrix factorization problem. However, the typical challenge in applying nNMF is to choose the reduced dimensionality R, as it strongly determines how **M** gets represented by **V** and **H**. A natural choice of R should originate from some logical and physical motivation. Note that in illumination splitting all excitation patterns exhibit their maxima in the focal plane of the detection objective (see Fig. 3.17b), while out-of-focus contributions are being modulated. Hence it makes sense to factorize **M** into an in- and out-of-focus part, as all sub-images contain similar in- but varying out-of-focus information. This is aimed for with $R = 2$ and is shown in Fig. 3.18a.

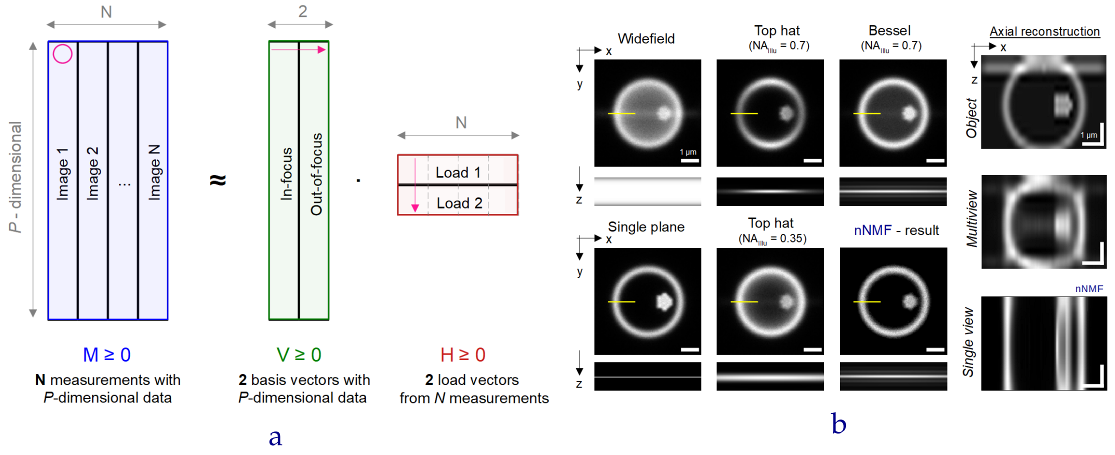

N
P - dimensional
Image 1
Image 2
...
Image N
≈
2
In-focus
Out-of-focus
·
N
Load 1
Load 2
M ≥ 0
N measurements with P-dimensional data
V ≥ 0
2 basis vectors with P-dimensional data
H ≥ 0
2 load vectors from N measurements
a
Widefield
Top hat (NA$_{Illu}$ = 0.7)
Bessel (NA$_{Illu}$ = 0.7)
Axial reconstruction
x
y
z
1 µm
Single plane
Top hat (NA$_{Illu}$ = 0.35)
nNMF - result
Object
Multiview
Single view
nNMF
b

Figure 3.18: a nNMF reconstruction of N measured images into an in- and out-of-focus contribution. The in-focus information corresponds to a weighted sum of all sub-images, with weights given according to $\mathbf{H}^{-1}$ (details in text). b Comparison of different detection modalities when imaging the 3D spherical object from Fig. 3.17b (1000 photons in the maximum pixel of the WF data, $NA_{Det.} = 1.1$). Widefield detection exhibits a lot of out-of-focus blur, light-sheet (top hat, $NA_{Illu.} = 0.7$) removes out-of-focus light but only provides a small FoV. The Bessel sheet yields a larger FoV but also generates considerable side lobes, creating unwanted background. Using Field-Synthesis enables to record multiple sub-images corresponding to axially varying sub-illuminations. Recombining those was done using the nNMF approach or by multiview deconvolution (10 accelerated iterations, thick slice; right) to reconstruct a 3D object estimate. Note that the multiview deconvolution enables to reconstruct a much more accurate three-dimensional representation of the ground truth object, compared to the nNMF processed data.

From $i = 1 \dots N$ number of images M_i (blue), two basic components (green) and two load vectors (red) are being extracted. The in-focus representation of the dataset can be found as a weighted sum, with weights given by $\mathbf{H}^{-1}$:

$$\mathbf{V}_{1,j} = \sum_{i=1}^{N} \mathbf{M}_{i,j} \cdot \mathbf{H}^{-1}_{1,i} \tag{3.68}$$

assuming that the first column of $\mathbf{V}$ corresponds to the in-focus information. In practice this might not necessarily be given, as the Frobenius norm (shown in eq. 3.67) can exhibit local minima [107]. Which means that repeating the factorization can yield different outcomes (especially changing the column order of in- and out-of-focus representation). To overcome this, most algorithms typically are run multiple times and use randomly chosen starting values of $\mathbf{V}$ and $\mathbf{H}$. Their final output is given as the particular factorization that yields the smallest value of $F(\mathbf{V}, \mathbf{H})$. In this thesis the *nnmf*-function of Matlab [108] was used (with 5 random initialization of $\mathbf{V}$ and $\mathbf{H}$), which is based on an algorithm described in Berry et al. [109]. To identify the in-focus contribution in $\mathbf{V}$, the two column vectors of $\mathbf{V}$ are first reshaped to yield two images of size $\sqrt{P} \times \sqrt{P}$ (assuming square image format), which are then subsequently Fourier transformed. The out-of-focus image will contain more noise, hence its spectrum (magnitude) will be closer to the noise floor than that of the in-focus representation. By thresholding both spectra, using the *isodata*-algorithm [110, 111] (single threshold), it is possible to discern which of the two column vectors of **V** corresponds to in- and out-of-focus light. The effect of the nNMF processing in case of imaging the 3D spherical object with different light-sheet geometries, is shown in Fig. 3.18b (1000 photons in the maximum pixel of the WF data). Conventional widefield detection will lead to strong ouf-of-focus blur and a reduced *SNR*. Top hat light-sheet excitation drastically reduces the unwanted background. However, this can only be achieved over a limited FoV (top hat, $NA_{Illu.} = 0.7$). To obtain the full FoV $NA_{Illu.}$ needs to be decreased, which inevitable will lead to more out-of-focus contribution as the excitation sheet gets thicker ($w_0^2 \propto z_R$). When illuminating with a Bessel light-sheet a large FoV can be recorded, while out-of-focus light from the strong excitation sidelobes corrupts the image. Creating the Bessel light-sheet with Field-Synthesis (23 sub-illuminations), the nNMF algorithms is able to remove most of the out-of-focus information, while maintaining the larger FoV. This can also be seen in Fig. A.13a, which compares different line profiles marked in Fig. 3.18b. Note that there is no guarantee that the factorization achieved with nNMF will correspond to a splitting into in- and out-of-focus information. Meaning that the nNMF processing of artificial objects like points, lines and planes in 3D might yield unreasonable results. Nevertheless, for the depicted simulation and the experimental data in sec. 3.5.4, the nNMF-algorithm did factorize into in- and out-of-focus representations, making it seem that for more natural objects this processing is justifiable. Interestingly, the nNMF-based reconstruction does not require any information on the optical system, it solely relies on the recorded data. Which makes it an easier to use alternative to multiview deconvolution.

Multiview deconvolution:

As for the previous *splitting & recombination* ideas, multiview deconvolution can also be applied in the proposed illumination splitting approach. However, the algorithm, as described in section 3.2.2, needs to be modified to yield:

$$\mathrm{est}^{(l+1)}(\vec{r}) = \mathrm{est}^{(l)}(\vec{r}) \cdot \sum_{i=1}^{N} \left\{ \frac{M_i(\vec{r})}{\left[\mathrm{est}^{(l)}(\vec{r}) \cdot I_i(\vec{r})\right] \otimes h(\vec{r})} \otimes h(-\vec{r}) \right\} \tag{3.69}$$

with the sub-images M_i being modulated through the axially varying sub-illuminations distribution I_i, as given by:

$$M_i(\vec{r}) = [S(\vec{r}) \cdot I_i(\vec{r})] \otimes h(\vec{r}) + \mathcal{N}(\vec{r}) \tag{3.70}$$

The results of performing a thin slice deconvolution (see sec. 2.4.2) on the data shown in Fig. 3.18b are depicted in the appendix (p. 133). A 3D deconvolution (thick slice) enables to obtain information about the three-dimensional structure of the sample under investigation. Results of the (3D) multiview deconvolution and the nNMF preprocessed (3D) deconvolved data is shown in Fig. 3.18b right. Note that recombining the split data in a multiview deconvolution yields a much better reconstruction of the underlying 3D object.

3.5.4 Experimental setup & results

To experimentally realize the illumination splitting, the custom-built light-sheet setup of *James Manton* at the Laboratory of Molecular Biology (LMB) in Cambridge was used. Its optical layout (see Fig. 3.19a) was matched to implement the idea of light-sheet generation via Field-Synthesis: a laser beam (LBX-488-100-CSB-PPA, Oxxius) is first focused in one dimension (cylindrical lens, LJ- 1695RM-A, Thorlabs) to create a line which is subsequently scanned over the pupil of the illumination objective (54-10-7, Special Optics; $NA_{Illu.} = 0.7$) by rotating a galvo-mirror (GVS001, Thorlabs) which is placed conjugate to the sample plane.

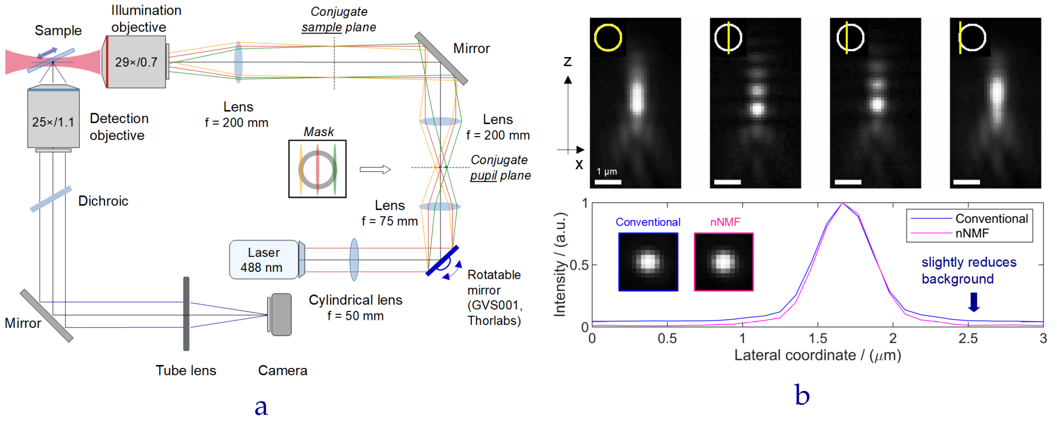
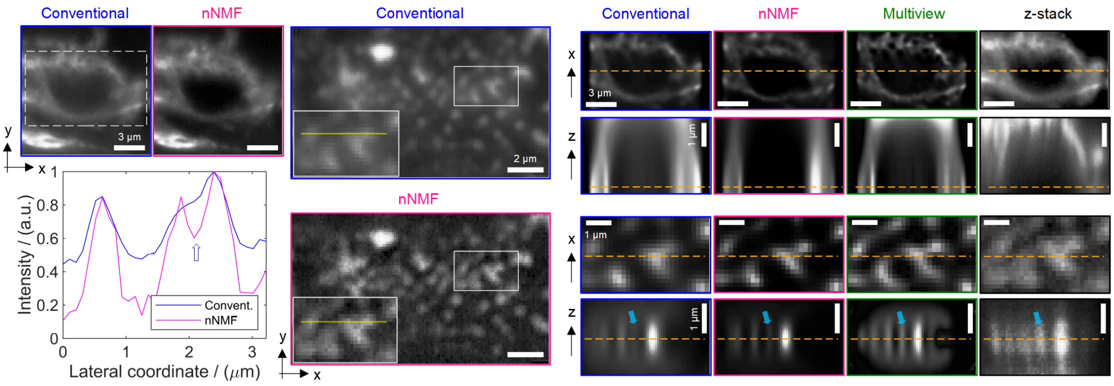

Sample
Illumination objective
29×/0.7
Conjugate sample plane
Mirror
25×/1.1
Detection objective
Lens f = 200 mm
Mask
Lens f = 200 mm
Conjugate pupil plane
Dichroic
Lens f = 75 mm
Laser 488 nm
Rotatable mirror (GVS001, Thorlabs)
Cylindrical lens f = 50 mm
Mirror
Tube lens
Camera
a
z
x
1 µm
Conventional
nNMF
Conventional
nNMF
slightly reduces background
Intensity / (a.u.)
1
0.5
0
0
0.5
1
1.5
2
2.5
3
Lateral coordinate / (µm)
b

Figure 3.19: a Experimental setup enabling illumination splitting in a light-sheet setup. A cylindrical lens transforms a rotational-symmetric laser beam (e.g. 488 nm) into a line, which is scanned over the illumination pupil by rotating a mirror in a conjugate sample plane. When an annular mask is put into the pupil, the scanning yields a Bessel-sheet illumination in the sample. The emitted fluorescence is captured by an objective ($NA_{Det.} = 1.1$) and imaged onto a camera. The experimental setup is designed and maintained by *James Manton* at the LMB in Cambridge. b Top: experimental data showing the axial view (from a z-stack) of an isolated bead (Tetraspeck, diameter = 100 nm), by illuminating it using the Field-Synthesis concept (11 sub-illuminations). Note how the emission of the bead varies with the scan position in the illumination BFP (similar to Fig. 3.17b). Bottom: when the corresponding in-focus slices (xy-plane) are processed using the nNMF-method, a slight reduction of background can be achieved (see line profile).

The pupil of the illumination objective is re-imaged so that it is freely accessible and an annular mask was introduced. The detection system is a conventional widefield setup with $NA_{Det.} = 1.1$ (CFI75 Apo 25XC), imaged onto a sCMOS camera (pco.edge 4.2, PCO AG, Germany). To test the alignment and overall performance of the system, fluorescent beads (diameter = 100 nm, Tetraspeck, Thermofisher Scientific) on a coverslip, have been imaged by sequentially acquiring a volume (z-stack; see Fig. A.14b), corresponding to each scan-position in the illumination BFP (11 sub-illuminations). The three dimensional information was used to observe the emission of a single bead, while employing the Field-Synthesis scanning. The results are depicted in Fig. 3.19b top, for three scanning positions and for the *conventional* light-sheet generation (left, full ring pupil), which was realized by summing the recorded sub-images (corresponding to beam scanning light-sheet generation). The effects of spherical aberrations are observable, nevertheless, the overall shape is that of a typical WF-PSF. The beads emission is modulated according to the illumination profile, generated by the respective scan position in the illumination BFP. The corresponding in-focus (lateral) sub-images are processed using the nNMF method described before, which yields in a slight background reduction (see line profile in Fig. 3.19b), compared to a pixel-wise summing of the data (conventional). The effect is not very strong, as a point sample does not represent an ideal object to be imaged by the proposed illumination splitting technique. A better option would be a structure which emits a signal at an out-of-focus region, which then corrupts the image in the detection plane (e.g. the spherical object used in the simulation shown in Fig. 3.17b). To achieve a similar effect experimentally, a sample of Vimentin, stained with Atto647N (Atto-Tec, Germany), has been imaged. Note that the nucleus of the cells were not labeled, hence did not emit any signal. As the Vimentin wraps around the whole nucleus, it does emit some light at out-of-focus regions (e.g. behind the nucleus), which can be seen in the 2D in-focus image as background haze (when focusing close to the nucleus membrane). The image results (in-focus slice only) are shown in Fig. 3.20a with a γ-adjustment of $\gamma = 1.5$. Note how the background haze typically associated with the Bessel illumination is removed (unwanted excitation of out-of-focus structures). Interestingly, this effect can also be observed at rather flat objects, such as randomly distributed fluorescent beads (same Tetraspeck as before) on a coverslip. Figure 3.20a right shows the two in-focus images corresponding to summing (conventional) and the nNMF processing of the acquired sub-images. The latter is able to remove some of the background light, which is most likely to be caused by overlapping beads (e.g. clustering). Also note that the general shape of the in-focus structures is not altered by nNMF processing. A line profile (yellow line) indicates an improved modulation depth, making some previously hidden structure visible (arrow).



a b

Figure 3.20: All images have been γ-adjusted with $\gamma = 1.5$. a Experimental data showing the nucleus (black) in a Vimentin probe (top left, stained with Atto674N) and beads immobilized on a coverslip (right), imaged using Bessel-sheet illumination by employing the Field-Synthesis concept. Due ot the sidelobes of the excitation, out-of-focus blur is generated which can be seen at the nucleus (3D object) and the gap between the distributed beads. However, when the modulated in-focus slices (shown on p. 134 in the appendix) are being processed using the nNMF method, the unwanted out-of-focus information can be removed. Particularly the beads sample shows an increase in contrast (inset) and even an improvement of modulation (line profile, yellow). The conventional method refers to the sub-images being pixel-wise added. b Thick slice reconstruction results for the Vimentin and the beads sample (10 accelerated iterations). Top always shows the in-focus slice and bottom the axial view through the slice in the xy-plane marked with an orange line. As already shown on simulated data in Fig. 3.18b, using the sub-images for multiview deconvolution helps to obtain a better reconstruction result along the z-axis. A recorded z-stack is shown on the very right. For the beads sample two arrow markers (cyan) have been added to highlight features which are observable in the z-stack and the multiview reconstruction, while being absent or underrepresented in the reconstructions of the conventional and nNMF processed data. E.g. the deconvolved nNMF result shows good reconstruction in the lateral view, but underrepresented the structure at the cyan arrow.

Figure 3.20b shows thick slice image reconstruction results for the Vimentin and beads sample in Fig. 3.20a (also with $\gamma = 1.5$), which are compared to a z-stack recording depicted on the very right. The top row indicates the in-focus slice of the reconstruction, marked with the orange line in the bottom row which represents an axial view. Similar to the simulation results in Fig. 3.18b right, processing the sub-images directly in a multiview deconvolution enables to obtain a more accurate representation of the underlying 3D object. With the bead sample it is noticeable that some of the structures, marked with cyan arrows, are either not visible or poorly represented in the single view reconstruction (conventional & nNMF). Due to the complex structure of the experimental data it was not possible to theoretically compute *NCC*-curves and optimize the iteration to reach an optimum reconstruction result. The presented data was obtained with 10 (accelerated) iterations for all reconstructions, varying this number did not change the overall impression that the multiview deconvolution is superior with respect to reconstruct 3D information. However, when looking at the in-focus slice of the beads sample, it seems that the deconvolved nNMF result is providing a more clear view on the clustered beads. Nevertheless, it is highly under representing the structure marked with the cyan arrow in the axial view. A representation of the recorded 3D information of the sample, indicating the modulation of axial image information introduced in the 2D images due to Field-Synthesis, is given in the appendix (p. 134). Due to limited access and the Covid-19 pandemic, it was not possible to optimize the experimental imaging or get data with improved quality. In the authors view there still needs to be more experimental work done to show the benefits and limitations of the proposed illumination splitting idea.

3.6 Summary and discussion

In this chapter the concept of *splitting & recombination* was introduced and put into context of fluorescence microscopy. The overall goal is to create sub-images with mutually differing *SNR*, which are then computationally recombined to yield an overall improved *SNR*. The general idea was first presented using the polarization splitting example, where a polarization-dependent focusing effect was exploited. Two computational reconstruction methods have been illustrated: weighted averaging in Fourier space and multiview deconvolution. The first has the advantage of being linear, hence gives the ability to predict the recombined outcome in form of a noise-normalized effective OTF $\tilde{h}_{wa,\sigma}$. For the multiview deconvolution, the Richardson-Lucy update scheme based on [68] was used. However, the projection of the different views into a single estimate was modified into a weighted summation. Note that this increases the convergence speed of the RL iterations, while the convergence guarantee in principle is lost. Additionally, deconvolution allows to reconstruct 3D information from 2D images, which can only give good results when multiple views (or images) are available. Three distinct splitting mechanisms were presented, each including a detailed theoretical analysis and a proof-of-principle experiment.

Wavelength splitting:

Fluorescence emission is typically characterized by a spectrum which is broad and asymmetric with respect to the emission wavelength λ (tail towards larger λ). This asymmetry effectively creates a spatial broadening which further degrades the resolution achievable by an imaging system. This broadening increases when the emission spectrum becomes more asymmetric (towards larger λ) and can be reduced when an appropriate emission filter is used (cutting away more longer wavelengths). However, this will further reduce the number of detectable photons and therefore also limit the achievable SNR. By splitting the fluorescence emission into multiple wavelength bands, computational recombination ensures to get more information out of the captured photons, leading to an improved SNR. This was verified using pattern E (line pairs) and I (crossing stairs) from an Argolight calibration target (introduced in sec. 3.3.4), which emits broad fluorescence emission in the visible spectral region. It was imaged using a Zeiss LSM 880 with spectral detection unit, enabling to simultaneously capture spatial and spectral information. Weighted averaging showed a slight enhancement of visibility of $IF_{\mathcal{V}} \approx 5\%$ at the smallest resolvable line pair. The multiview deconvolution makes much better use of the available information and reaches a maximum of improvement $IF_{\mathcal{V}} \approx 79\%$, compared to the non-split data. The three dimensional image reconstruction of the "crossing stairs" sample yields a reduction in axial blur on the order of 10%. Depending on the shape of the $\varepsilon(\lambda)$-curve it might also make sense to obtain irregularly sampled spectral data, so that the amount of light per spectral channel stays constant. Such that the object information corresponding to the emission tail at smaller wavelengths is not lost in noise. The microscope system was located at a different campus (ZAF, Universität Jena) and could not be aligned perfectly due to time restrictions and the Covid-19 pandemic. In general the options to optimize the imaging settings were limited, leaving it open whether better data could have been produced.

Pupil splitting:

The next idea for splitting & recombination is to manipulate the spatial frequency information in the pupil of the optical system. This is achieved by radially splitting the pupil of an incoherent imaging system into two sub-pupils. Using weighted averaging in Fourier space to recombine the split data shows an SNR enhancement towards higher spatial frequencies, while some degradation is visible at the mid-frequencies. The latter is not limiting in the sense that the signal can still be recovered using post-processing techniques, as $SNR \gg 1$. The splitting additionally creates an extended-depth-of-field (EDoF) effect, as both sub-PSFs are elongated along the z-axis. The experimental verification of the advances of pupil splitting have been investigated using an optical setup, build and operated by the group of *Jonas Ries* at the EMBL in Heidelberg. In their setup the BFP of a microscope objective is accessible, such that a small mirror device can split the light distribution in the pupil into an inner *disk* and outer *ring* part. Both contributions were redirected onto two neighboring halves of an sCMOS camera. Again the Argolight sample with the resolution (pattern E) and the crossing stairs (pattern

I) were imaged and after weighted average recombination showed the aforementioned performance at mid- and high spatial frequencies. The advantage of pupil splitting becomes apparent as some very fine structures can only be resolved using the splitting & recombination approach. Multiview deconvolution is even able to undo the reduced information transfer at the mid-frequencies, while generating a visibility improvement of $> 30\%$. The axial reconstruction of the two crossing stairs shows the aforementioned EDoF effect and the ability to reconstruct the stair-like shape of the object, while only measuring 2D images. Note that the reconstruction is symmetric with respect to the optical axis, which is due to the numerically computed sub-PSFs which are symmetric as well. An idea to overcome this would be to include PSF aberrations, present in most experimental data, such as spherical aberration or astigmatism [6]. The idea of using more than just a single radial split is obvious but was not followed in much detail as every additional split will automatically decrease the *SNR* per sub-image. Which might result in some information being lost in the respective noise floor. When implementing the pupil splitting, it is important to ensure that both channels are in-focus on the camera to obtain the best recombination of lateral structures. This was achieved by moving the imaging lenses in front of the camera by hand and looking for the sharpest image. Some small misalignment along the optical axis could be beneficial as long as both sub-PSFs still overlap. This would result in an artificially enlarged depth-of-field, while information above and below the focus could be attributed to the differently shifted sub-PSFs. It would also solve the described symmetry problem along the z-axis.

Illumination splitting:

Performing the splitting in the illumination, as compared to the detection (e.g. in wavelength & pupil splitting), has the advantage that the splitting itself does not necessarily limit the amount of photons per sub-image. Shifting the focus from photon-limited imaging (i.e. shot noise), to additional factors that can degrade *SNR* such as out-of-focus light. The presented idea is to generate a light-sheet illumination, using the concept of Field-Synthesis [94]. In light-sheet imaging the ability to remove out-of-focus contributions is given by the thickness of the excitation structure. However, typically a small thickness is linked to a reduced detection FoV, making it difficult to achieve good out-of-focus rejection over a large FoV. One possibility to get around this would be to create the light-sheet by using a Bessel beam (annular mask in illumination BFP), consisting of a thin and long central peak and strong sidelobes. Unfortunately, the latter excite out-of-focus structures which will corrupt the 2D widefield image with additional unwanted background. Field-Synthesis can be used to create the (temporally averaged) Bessel sheet, resulting in a series of axially modulated sub-illuminations. The main advantage of Field-Synthesis given in the literature is its capability to reduce the peak irradiation at the sample, which minimizes phototoxicity and photobleaching. Another benefit that it provides is the opportunity to capture sub-images, corresponding to each of the sub-illuminations. Because the maximum excitation for each sub-illumination always coincides with the focal plane of the detection objective and the out-of-focus information is modulated in the sub-images (2D only), it is possible to remove the unwanted out-of-focus information. The computational recombination is either done using an algorithm based on non-negative matrix factorization (nNMF) or by multiview deconvolution. The former tries to separate the measurement data into two contributions: in- and out-of-focus information. Multiview deconvolution can be used to obtain a 3D estimate of the true underlying object (thick slice), as it incorporates different 2D views corresponding to the axially modulated excitation patterns. The experimental verification of the presented ideas were done using a custom-made light-sheet system, build by *James Manton* at the LMB (Cambridge). A Vimentin and a bead sample were imaged and showed the expected axial modulation of fluorescence emission. Using only the 2D in-focus slices from the recorded z-stack, and processing them with the proposed nNMF algorithm confirmed the idea to remove out-of-focus light via illumination splitting. The axial reconstruction using the thick slice method showed that the most accurate estimate of the underlying 3D images structure can be reached using a multiview deconvolution. Due to Covid-19 it was not possible for the author to actually visit the experimental setup at the LMB. Hence the presented data can only be regarded as first proof-of-principle results and a further investigation with an optimized image acquisition needs to be realized in the future.

In general it should be noted that when image reconstruction methods (e.g. deconvolution) are used, this should be accompanied with some information on whether a sensible or optimal solution has been reached. This is partially done in this thesis by computing *NCC* curves from simulated data. Whenever two datasets were processed with the same number of iterations, the development of their individual reconstruction quality was monitored.

Chapter 4

Submillimeter wave imaging

In the previous chapter a hybrid method, combining experimental and computational steps, was introduced and applied to fluorescence microscopy. However, often it is also possible to achieve an *SNR* enhancement with computational post-processing alone. In the following this will be exemplified in the field of *submillimeter wave imaging*. This wavelength regime is also known as terahertz radiation and comprises of wavelengths from 10 μm to 1 mm [112]. In contrast to detecting light in the visible, submillimeter wave detection is not as widely distributed in a commercial sense. The corresponding frequencies lie in a range where electronic detection is not fast enough and optical detectors aren't efficient, termed the "THz gap" [113]. However, submillimeter waves have interesting properties, such as the ability to penetrate through fabrics while being low energetic (non-harmful because non-ionizing). One of their possible applications is *passive concealed threat detection*, which has the goal to identify possible threats, worn at the human body but hidden beneath clothes from a safe distance [114, 115]. The detection can be done in such a way, that the suspect is unaware of the surveillance. Enabling to plan further police measures, while having proof of their eligibility in case of legal actions in front of a court.

In the following the challenges and limitations using submillimeter waves for passive concealed threat detection will be introduced and discussed. Next, a brief description of the prototype of the security imager and the subsequent pre-processing steps are given. Since the amount of submillimeter wave signal emitted from the human body is small and the detection is shot noise limited, the image quality is governed by a low *SNR*. An improvement in *SNR* is generated by means of computation, employing denoising and deconvolution algorithms. The last section will be a short summary and discussion.

4.1 Challenges in passive concealed threat detection

The idea behind passive concealed threat detection is to identify suspects that are carrying unidentified (potentially dangerous) objects under their clothes, without the use of an active radiation source. So that the screening of such a person can be carried out without their knowledge, minimizing the risk of creating a dangerous situation when trying to detain the suspect. This means that a possible security imager must be able to detect signals from a distance > 10 m, while maintaining a FoV of $\approx 1 \times 2$ m^2, encompassing an upright standing person. The system must be fast enough to track moving individuals, while also allowing for a high enough spatial resolution (≈ 1 cm) and *SNR*, to be able to successfully identify small objects, i.e. the size of a handgun. Making sure that a smartphone is not falsely marked as a potential threat. Another important aspect of this type of investigation has to do with data privacy [116, 117]. It is required by German law that the identity of an individual can not be directly inferred from the submillimeter wave data alone. A passive imaging modality suits perfectly to this as it only reveals an outline of the person. However, the most important property that needs to be met by the passive security imager is that of efficiently detecting radiation which penetrates through clothes.

4.1.1 Emission & absorption of submillimeter waves

As stated before, the wavelength range of submillimeter waves is given between 10 μm and 1 mm. The (spectral) detection window of the security imager needs to be set such, that the suitable radiation shows a high transmissivity through typical clothes. For most fabrics the transmission decreases for wavelengths $\lambda < 1$ mm, e.g. see Fig. 1 in [118] or [119]. This coincides with the *atmospheric window*, meaning that for wavelengths much smaller than 1 mm the humidity in the air will block most of the respective radiation. Making electromagnetic waves with $\lambda \geq 1$ mm more suitable for concealed threat detection. However, the emission of electromagnetic radiation in this respective wavelength regime is low. According to *Wien's* displacement law [120], the maximum emission of a black body, at $T = 20°C = 293$ K, is given at $\lambda_{max} \approx 10$ μm. The black body emission curve is asymmetric and exhibits a tail towards longer wavelengths (similar to fluorescence emission spectra). Meaning that wavelengths at around 1 mm correspond to a spectral region where only a very small number of photons is being emitted. Just from the emission characteristics alone it would make sense to go to wavelengths $\lambda < 1$ mm. However, then the transmission through fabrics gets worse, which is problematic in the application of concealed threat detection. Therefore it is important to employ a very sensitive detection scheme.

4.1.2 Inherent limitations in spatial resolution and *SNR*

Besides the ability to detect a signal that carries information about concealed threats, it is also important to achieve high enough spatial resolution and *SNR*. The ability to detect smaller structures scales with the wavelength and is indirectly proportional to the detection *NA* (see eq. 2.8). Suggesting that a choice of $\lambda \geq 1$ mm, comes at the cost of reduced spatial resolution. Also note that the detection *NA* gets smaller when imaging further away from the detector. Overall this means that the ability to detect concealed structures in the submillimeter wave regime comes with a tradeoff: either a good spatial resolution is achieved ($\lambda \leq 1$ mm) but the transmission efficiency through clothes is reduced, or radiation is transmitted well ($\lambda \geq 1$ mm) but the spatial resolution is worsened. As the main criterion for the success of the security imager lies in the ability to detect concealed threats, this tradeoff is approached by setting $\lambda \lesssim 1$ mm. Hence, the limiting factor is now given by the achievable signal-to-noise ratio. Which, among others, depends on the efficiency of the detection technology used.

4.2 The passive standoff security imager

In the previous section some criteria for the successful use of the security imager were defined. During the last 10 - 15 years, research at the Leibniz Institute of Photonic Technology (IPHT) in Jena has resulted in several prototypes, which have gradually come closer to meet those requirements, e.g. in [118, 121, 122].

4.2.1 Detection and readout scheme

The latest prototype is based on a Cassegrain telescope design (see Fig. 4.1) and consist of a main mirror, which has a diameter of 1 m. Together with an imaging wavelength of $\lambda = 0.75$ mm -1 mm, this results in a spatial resolution of ≈ 1 cm -2 cm. The imaging concept is that of a line-scanner, meaning that a full FoV is generated by scanning the image over an array of detectors. To do this, the small mirror of the Cassegrain is tiltable, achieving a FoV of ≈ 1 m $\times 2$ m at a distance of 10 m -15 m. The detector array is located at the center of the optical axis, behind an opening in the main mirror and consists of 8 modules, each with 2 rows of 8 detectors (see blue dots in Fig. 4.1). Note that both rows are slightly shifted

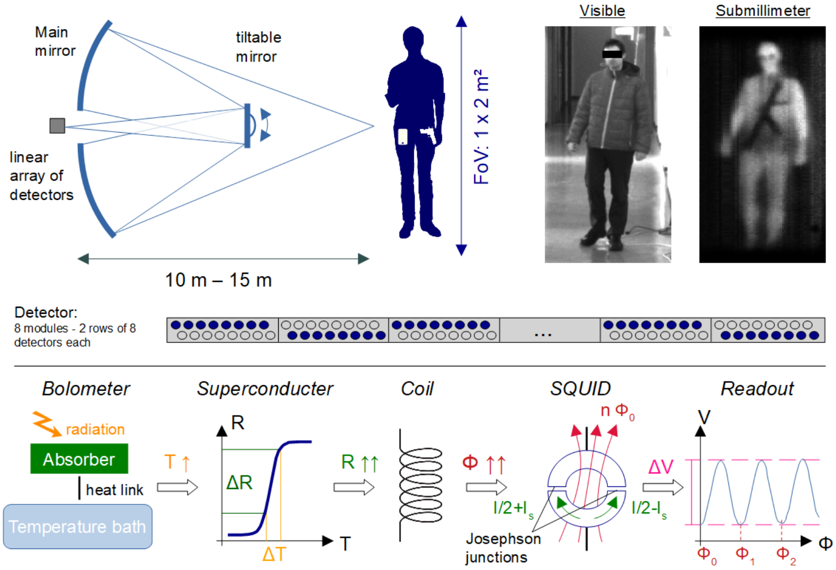

Main mirror
tiltable mirror
linear array of detectors
10 m – 15 m
FoV: 1 x 2 m²
Visible
Submillimeter
Detector:
8 modules - 2 rows of 8 detectors each
...
Bolometer
Superconducter
Coil
SQUID
Readout
radiation
Absorber
heat link
Temperature bath
T ↑
R
ΔR
ΔT
T
R ↑↑
Φ ↑↑
n Φ₀
I/2+Iₛ
I/2-Iₛ
Josephson junctions
ΔV
V
Φ₀
Φ₁
Φ₂
Φ

Figure 4.1: The passive submillimeter wave imager used in this thesis consists of a large main mirror (diameter ≈ 1 m) and a small, tiltable, mirror which reflects the captured radiation onto a linear array of detectors. The scanning provides a FoV of about 1 m $\times 2$ m at a distance of 10 m -15 m. Feed-horn antennas direct the acquired signal onto detectors, which are tuned for the required wavelength region of 0.75 mm -1 mm. The detectors are arranged in 8 modules, which are made up of to two rows with 8 detectors each. In the prototype, only one row per module (blue dots) was read out. For the detection of submillimeter waves a bolometric response is used. In short: radiation gets absorbed by a superconducting transition-edge sensor (TES) and the resulting temperature change T is transferred into a change of magnetic flux Φ, which is amplified using a super-conducting-quantum-interference device (SQUID). Relating the voltage measurable across the SQUID to the absorbed photon energy. Top right: comparing image results in the visible and the submillimeter range.

with respect to each other, the reason being that otherwise Nyquist sampling is not fulfilled. Each detector and the corresponding feed horn antenna, which couples the radiation into the sensor, are too big to be placed next to each other, at Nyquist sampling distance. In total 128 detectors are available, however, only 64 of them (in alternating rows) have been readout in the data presented. Due to a lack of multiplexed readout capability in the current prototype.

The actual detector technology is borrowed from the field of particle detection. So-called microbolometers are used, as they are tunable in their detection range, meaning they can be specifically designed to work in the required wavelength regime of $\lambda = 0.75$ mm -1 mm. A short graphical representation of the whole detection process is shown in Fig. 4.1 bottom. Their working principle is as follows: submillimeter wave radiation gets absorbed and heats up a transition-edge-sensor (TES) [123], which shows a strong response in electric resistance R. As the TES is run at cryogenic temperatures (< 1 K) it exhibits superconducting properties, resulting in a very sharp dependency of R with respect to T. A connected circuit transforms this, via a coil arrangement, into a change of magnetic flux Φ.

The challenge is now to amplify this fairly weak signal, without introducing additional noise. This is done using superconducting quantum interference devices (SQUIDs) [124, 125]. Which consists of two Josephson junctions [126], separating a superconducting loop into two halves. The Josephson effect describes, similarly to quantum tunneling, the flow of a supercurrent (here Cooper pairs [127]) through a potential barrier. When a current I is applied to the SQUID and no external magnetic flux Φ is present, I splits into the two arms, yielding $I_{1,2} = I/2$ for each. A weak external magnetic field will now induce a screening current I_s, which will flow in the opposing direction in one of the arms ($I_1 = I/2 - I_s$) and concurrent in the other arm ($I_2 = I/2 + I_s$), as it tries to cancel the external field (corresponding to the absorption event). When the current in the latter exceeds a critical value I_c, a voltage can be measured across the barrier. However, the magnetic flux Φ going through the SQUID can only be an integer of the magnetic flux quantum Φ_0. This means that as soon as the external field gets larger than $n \cdot \Phi_0/2$ (with $n \in \mathbb{N}$), the induced screening current I_s will change its direction, leading to a modulation of detectable voltage change ΔV. With this a measurable electric signal can be related to the incoming submillimeter wave radiation. However, notice that this response is not unambiguous as the voltage can be attributed to multiple integers of the magnetic flux quantum Φ_0. Additionally, the image acquisition needs to be fast enough to track walking individuals. The available prototype captured data at up to 25 frames per second [118].

4.2.2 Pre-processing steps

The imager prototype used for the concealed threat detection comes with some caveats. Pre-processing is required to reconstruct an image from the acquired raw data and to correct different imaging artifacts. The following pre-processing steps work on a subset of the recorded data (e.g. 10 frames), where the required correction parameters are obtained automatically and are recomputed for each new set of frames. Note that this type of content-based processing was deliberately aimed for in the project that funded parts of this PhD work. The reason being that for an easy-to-use system, which could later be turned into a commercial product, the amount of required calibration steps should be minimized. All processing steps introduced in the following are designed and tested, so that they work robustly with the raw data acquired from the current prototype. As the security imager records data according to a line-scan in the vertical direction, every second frame needs to be flipped vertically. Raw data for three different datasets are shown in Fig. 4.2 (*20180905-03* refers to the third dataset recorded on 5.09.2018).

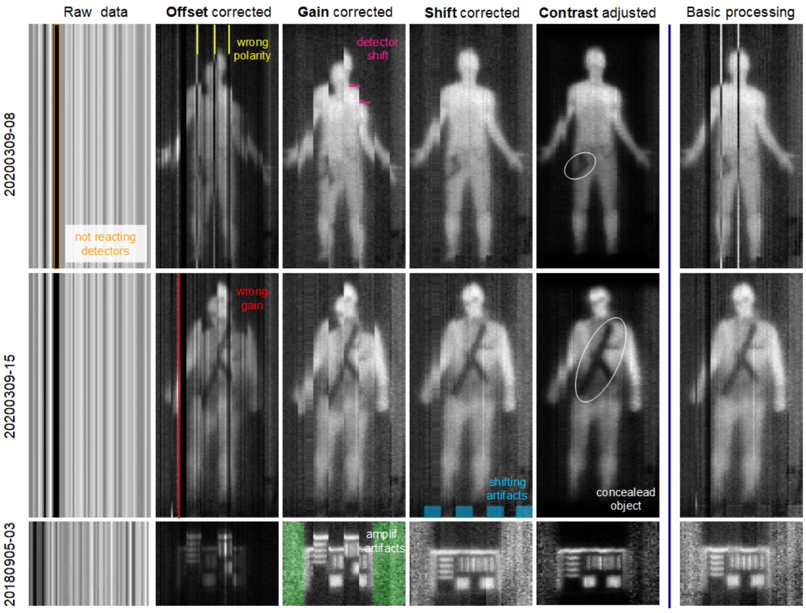


Figure 4.2: Overview of multiple pre-processing steps for three different datasets (vertical). From left to right: raw data, vertical lines as the detector array is vertically scanned over the scene; offset corrected: removing individual detector offsets by subtracting dark images; gain corrected: getting uniform sensitivity of individual detectors; shift corrected: remove axial shift of detector modules; contrast adjusted: applying scaling to ensure maximum contrast from back- to foreground. Some artifacts to point out: depending on cool-down procedure some detectors might not work (orange) or show a limited sensitivity (red); due to the wiring process of the detector modules to the read-out, a polarity error (yellow, measured value gets smaller with increasing radiation) occurs; to ensure uniformity of the measured signal, weaker sensor responses are amplified, leading to strong noise effects (green) especially when the detectors have not seen any signal within the evaluates time series; due to the specific arrangement of the detector modules a vertical shift of image information can be observed (pink), de-shifting leads to wrap-around artifacts (blue). Two objects ("handgun" made out of aluminum and wooden "AK-47") were hidden in trouser pockets (gun) or underneath jacket (rifle). Last column shows the basic processing, consisting only of background subtraction and a de-shifting operation with fixed shift values.

The vertical lines correspond to the individual sensor pixels, scanned vertically. A direct image of the scene is not observable due to strongly deviating offset values of each detector. For more details regarding this and the varying gain of each detector see p. 135 in the appendix. Note that those detector characteristics drift over time and more importantly, change whenever the whole system is cooled down to its operating temperature of < 1 K (compare different dates of acquisition in Fig. 4.2). This requires the recording of appropriate dark images (no submillimeter signal in scene) anytime when the security imager is set up to be used for concealed threat detection. Offset correction is achieved by subtraction of the dark images, resulting in the recorded scenes shown in the second column of Fig. 4.2. However, the image data still comprises of a number of artifacts. Some of the detectors are not working (orange), hence their value need to be interpolated from adjacent sensors. Others show some inverted polarity (red): the measured signal becomes weaker when more radiation is absorbed by the TES.

The reason for this is the incorrect wiring of detectors to the readout electronics. Apart from the offset of each individual detector, also its sensitivity differs (see slope in Fig. A.15a). Uniformity is enforced by calculating a *max*-projection over all frames and along the vertical scanning direction. With this a reference value is obtained for each detector, comparing it with respect to the full FoV maximum. These individual values are written into the diagonal of a 64×64 matrix $\mathbf{Q}$. The matrix formulation is useful when the effects of detector crosstalk (not present in the shown data) is to be included, as this can be achieved by setting the values in the first minor diagonal of $\mathbf{Q}$. The goal is now to find a gain vector $\vec{G}$ which, when multiplied with $\mathbf{Q}$ yields the identity vector $\vec{1}$ (indicating uniformity):

$$\vec{1} = \mathbf{Q} \cdot \vec{G} \tag{4.1}$$

with $\vec{1}$ and $\vec{G}$ being vectors of size 64×1. By inverting this equation, a least-squares optimized estimate of the correct gain values $\vec{G}_{est.}$ can be computed:

$$\vec{G}_{est.} = \mathbf{Q}^{-1} \cdot \vec{1} \tag{4.2}$$

Each element of $\vec{G}_{est.}$ corresponds to one of the 64 detectors and scales their individual values to yield an approximately uniform FoV. However, this will also lead to pure noise amplification (green), when some of the detectors haven't seen any signal (besides background) within the evaluated image sequence. In the aforementioned project this was aimed to be corrected for by obtaining the outline of the person under investigation with the help of cameras working in the visible wavelength range. Results on this are not presented in this thesis. Apart from differences in individual detector responses, also the line-scanning itself results in artifacts. For example: the specific meander-shaped geometrical layout of the detector (see Fig. 4.1) results in a vertical shift of every second detector module (pink). This is corrected for by averaging the recorded signal for each module (8 detectors) and using a correlation-based algorithm [83] to estimate the shift with respect to each adjacent module. Once the shift in pixels has been computed, a de-shifted version of the image data is obtained by applying an appropriate phase ramp in Fourier space (*shifting* theorem). Naturally, this leads to wrap-around artifacts at the vertical edges of the image (blue) (originating from the discrete Fourier transform). The visibility of those are damped by multiplying a cosine shaped windowing function after the de-shifting operation. Not shown in Fig. 4.2 is the shift correction between subsequent frames. When operating the security imager prototype, the scanning results in some oscillation of the whole detection unit. This leads to a considerable vertical shift between consecutive frames and some minor stretches along the vertical dimension. When neglecting the latter and assuming that the recorded scene does not change too much in between, the shift can be estimated and corrected for similarly by applying an appropriate phase ramp. The last step of the pre-processing is a contrast adjustment. To do this the series of frames to be processed is split into two halves $M_{up,down}$, each corresponding to the up- and down-scanning of the tilt-mirror. These two temporal stacks are concatenated along the fourth dimension and an pixel-wise average $\bar{M}$ is computed. This value represents the mean value after one complete vertical scan (up and down). A correction factor $C_{up,down}$ for each of the two (temporal) stacks is calculated, according to:

$$C_{up,down}(\vec{r}) = \frac{\bar{M}(\vec{r})}{|M_{up,down}(\vec{r})|^2 + 0.1} \tag{4.3}$$

These correction factors are multiplied to the respective temporal series $M_{up,down}$ and then rearranged into a single data stack, comprising of both up- and down-scanned images. What this does is that the background in the recorded scene is damped, while the person/object under investigation is amplified in terms of its signal display. After pre-processing, objects which were hidden under clothes (encircled in white) could be identified more easily: 1) mock-up gun made out of aluminium, put into pocket; 2) wooden "AK-47" underneath a jacket. The resolution target in the bottom row of Fig. 4.2 has a spacing of 1.5 cm and indicates a loss of horizontal resolution. However, this is still enough to visually identify the handgun, making it clear that the successful detection of hidden threats is indeed possible. The last column of Fig. 4.2 shows a basic processing of the submillimeter wave data, which only employs background correction and a fixed de-shifting of the meander-shaped detector pattern. The difference in terms of contrast and the presence of imaging artifacts, clearly highlights the importance of employing pre-processing. Especially when dealing with small objects like guns or knives, where the image formation in the existing prototype comes to its technical and physical limits. Another important aspect of the aforementioned pre-processing is the capability to self-calibrate through the acquired image data. No additional user input is needed, the pre-processing obtains the necessary correction parameters from a short time series of 10 or more frames.

4.3 *SNR* enhancement via computation

To further improve the threat-detection capability of the security imager, additional *SNR* improvement seems to be advantageous. This is because the detected signal in the submillimeter wave regime is not restricted to dangerous objects, e.g. also smartphones are visible. To achieve a successful identification of harmful objects (= threats), the best possible *SNR* is required. Improvements on the hardware level will not be possible, due to the existing strong technical limitations. Therefore, only a *SNR* enhancement on the computational level is feasible. The three following methods will be looked at: decreasing noise, improving signal and data-driven image reconstruction.

4.3.1 Decreasing noise: denoising

The first idea is to try to remove additive noise from the acquired data, without compromising the inherit signal:

$$Image = Signal + Noise$$

The goal is to subtract noise from the image data, such that a good estimate of the true underlying signal is obtained. Making a clear distinction between signal and noise is difficult, as each image contains an individual realization of a statistical noise process, making it impossible to predict the exact noise contribution. In the appendix (p. 135) it is shown that the noise present in the raw data corresponds to the shot noise regime. Making it clear that processing these raw frames by some sort of averaging seems to be a suitable choice, as shot noise is bias-free. In Fig. 4.3, different denoising techniques are depicted. On the very left the fully pre-processed images from Fig. 4.2 are shown, while next to it, the results of applying a uniform filter are displayed. Mathematically this filtering is given as:

$$M_{unif.}(\vec{r}) = M(\vec{r}) \otimes \tfrac{1}{9}\left[\begin{smallmatrix}1&1&1\\1&1&1\\1&1&1\end{smallmatrix}\right] \tag{4.4}$$

with $1/9 \cdot \left[\begin{smallmatrix}1&1&1\\1&1&1\\1&1&1\end{smallmatrix}\right]$ being a 3×3 normalized filter kernel.

This filter replaces each pixel in the image with the arithmetic average of its surrounding neighbors. It is clear that the averaging over adjacent pixels fundamentally does reduce the spatial resolution. However, it is easy to implement and can be used on any type of image data. In Fig. 4.3, a reduction of noise can be observed, while some spatial resolution is lost (bottom). To overcome the problem with the reduced resolution, an adaptive filtering approach can be used. Changing the parameters of the filter kernel according to the local image content [128].

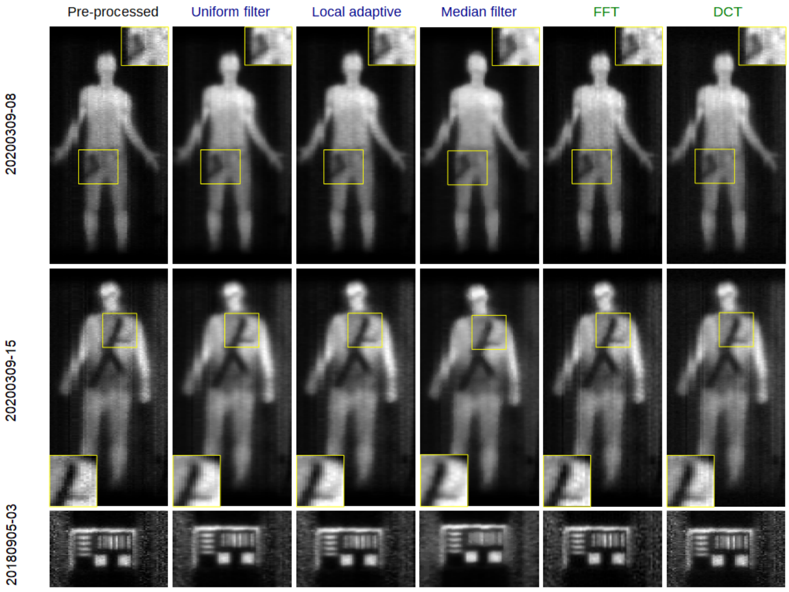


Figure 4.3: Overview of different *denoising* strategies. From left to right: pre-processed data (shown in Fig. 4.2); applying a uniform (averaging) filter will remove noise but also smoothen image details; this can be improved upon by applying a locally adaptive filter routine [128]; using a median filter does remove noise but also leads to some additional spatial blur; representing the data in Fourier space (FFT) helps to get rid of noise, as high frequency components, mainly corresponding to noise, are simply set to zero. The shape of the hidden objects is not altered by the denoising, nevertheless grainy patches become visible; a better alternative to use is the discrete cosine transform (DCT) as it provides a more dense signal representation [129], making it easier to remove noises by setting most of the resulting DCT coefficients to zero.

The idea is to calculate the mean μ and variance σ^2 of the image by employing:

$$\mu(\vec{r}) = M(\vec{r}) \otimes \frac{1}{9}\begin{bmatrix}1&1&1\\1&1&1\\1&1&1\end{bmatrix} \tag{4.5}$$

$$\sigma^2(\vec{r}) = \left(M^2(\vec{r}) \otimes \frac{1}{9}\begin{bmatrix}1&1&1\\1&1&1\\1&1&1\end{bmatrix}\right) - \mu^2(\vec{r}) \tag{4.6}$$

with M the measured image and a uniform filter kernel (filter size $= 3 \times 3$). The filtered image $M_{filt.}$ is obtained by applying the following computation:

$$M_{filt.}(\vec{r}) = \mu(\vec{r}) + \frac{\max\{\sigma^2(\vec{r}) - \nu^2, 0\}}{\sigma^2(\vec{r})} \cdot \Big[M(\vec{r}) - \mu(\vec{r})\Big] \tag{4.7}$$

Each pixel i in the original image gets bias corrected and then normalized by its relative difference between the observable variance and the global noise variance ν (with $\sigma^2 - \nu^2 \leq 0$). The result looks very similar to what was obtained using the uniform filter, only a very small increase in obtainable spatial resolution might be detected. Another filter operation that is typically used to denoise images is the median filter. It replaces the pixel at the center of a neighborhood (here 3×3) with the median value of all adjacent pixels. The median operation is more robust with

respect to outliers, which are typically introduced due to noise. However, similar to the uniform filter it can reduce the resulting spatial resolution.
A different approach for noise removal, is to represent the measured image in another basis. Such that noise and signal can be split according to some predefined characteristics. One possibility would be to use Fourier space. Where additive (pixel-independent) noise occurs as a constant noise floor. The measured signal in an imaging system is typically bandlimited, meaning that there are some pixels in the Fourier transform of the image data, which only encode noise (in a region beyond the bandlimit). We can set those to zero and apply an inverse Fourier transform, the result is shown in Fig. 3.3a. Noise has been reduced, while fine image details get only slightly altered. However, more grainy patches become visible. Instead of using the full Fourier series, the data can also be represented by a discrete cosine transform (DCT). It typically encodes image information with much fewer coefficients than the FFT, meaning that it concentrates most of the signal at the lowest frequency components [129]. Removing higher frequencies therefore will separate noise from signal contribution more efficiently. Figure 4.3 shows the corresponding result, indicating an improvement in terms of image quality with respect to the raw data.

4.3.2 Improving signal: deconvolution

As mentioned before, it might be crucial to further improve the spatial resolution and *SNR* of the security imager. So that the shape of small objects can more easily be detected and identified, improving the classification into normal and harmful objects. Therefore we will look at *deconvolution* as a tool to perform image reconstruction, assuming the following image formation model:

$$Image = Object \otimes PSF + Noise$$

with $\otimes$ denoting the convolution operation, *PSF* the point-spread function of the security imager and *Object* the unknown object, a potential threat. The PSF is theoretically modeled as an Airy pattern [7], whose width was adjusted by looking at the spatial blur of the resolution target in Fig. 4.2.
In deconvolution the goal is to reverse above equation, to obtain a good estimate of the unknown object. This is non-trivial as noise gets amplified in the inversion process, leading to an unsatisfactory reconstruction of object information.

The straightforward inversion is typically replaced by a Wiener filter (see sec. 2.4.2), where the regularization parameter $\tilde{K}$ needs to be adjusted depending on

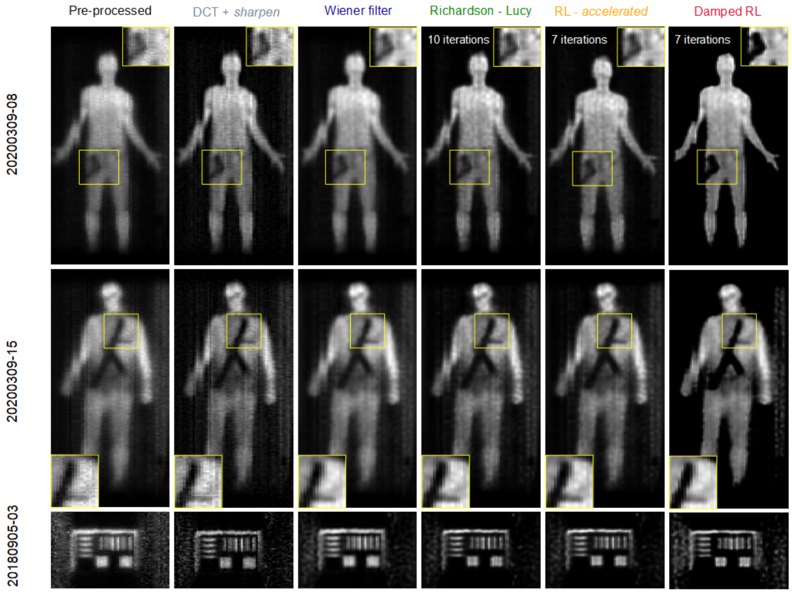


Figure 4.4: Overview of different *deconvolution* strategies. From left to right: pre-processed data; applying DCT-denoising + a sharpening filter to emulate improved spatial resolution while using a denoising method; using a Wiener filter ($\tilde{K} = 0.05$) generates a smoother representation but does not yield a strong improvement in terms of resolution (see "20180905-03"); this can be overcome by the Richardson-Lucy (RL) algorithm, which exhibits stronger noise artifacts and additionally requires a longer computation time (10 iterations); acceleration of the iterations is possible via work of [54] (7 accel. iterations), but cannot remove the noise artifacts; a damped version [55] ($\gamma = 0.01$) of the accelerated RL method is able to strongly reduce those, yielding a strong *SNR*.

the noise level (and image content). Figure 4.4 shows that the Wiener filtered result ($\tilde{K} = 0.05$) exhibits less noise and is not affected by pre-processing artifacts, which can be seen when the DCT-denoised image from Fig. 4.3 is sharpened (by substracting a Laplace-filtered version [130]). However, in terms of spatial resolution, the Wiener filter does not yield in a strong improvement. A better approach would be to use an iterative algorithm such as Richardson-Lucy (RL) deconvolution, which has been introduced in sec. 2.4.2 as:

$$\mathrm{est}_{l+1}(\vec{r}) = \mathrm{est}_{l}(\vec{r}) \cdot \left[\frac{M(\vec{r})}{\mathrm{est}_{l}(\vec{r}) \otimes h(\vec{r})} \otimes h(-\vec{r}) \right] \tag{4.8}$$

with M being the measured image data, h the corresponding PSF and est_l the algorithms estimate of the underlying object structure at iteration l. An enhancement in terms of resolution is observable (see "20180905-03"). But comes at the cost of amplified noise artifacts, which are quite common in RL deconvolution. Another drawback is the increase in computation time compared to direct filtering such as the Wiener filter. Note that the computational speed is of interest for the particular application of concealed threat detection. Hence iterative reconstruction methods need to be optimized with respect to their runtime.

Accelerating Richardson-Lucy deconvolution:

Different methods of accelerating the RL algorithm have been proposed in the literature (e.g. [54, 131, 132, 133]). In this thesis the work of Biggs et al. [54] is followed, who proposed a modified line-search method. To implement their work, the estimate (est_l) needs to be replaced by a current estimate ($\mathrm{est}_l^{(\mathrm{curr.})}$), so that the aforementioned RL iteration equation becomes:

$$\mathrm{est}_{l+1}(\vec{r}) = \mathrm{est}_l^{(\mathrm{curr.})}(\vec{r}) \cdot \left[\frac{M(\vec{r})}{\mathrm{est}_l^{(\mathrm{curr.})}(\vec{r}) \otimes h(\vec{r})} \otimes h(-\vec{r}) \right] \tag{4.9}$$

The current estimate depends on the history of the previous reconstructions. After each iteration it gets modified by a multiplication with a difference term involving the estimate of the previous iteration:

$$\mathrm{est}_l^{(\mathrm{curr.})}(\vec{r}) = \mathrm{est}_l(\vec{r}) + \lambda_l \cdot \left[\mathrm{est}_l(\vec{r}) - \mathrm{est}_{l-1}(\vec{r}) \right] \tag{4.10}$$

To control the influence of the previous reconstructions on the current estimate the parameter λ_l is introduced, acting as a weighting factor which itself changes per iteration. Its value is given by comparing the residuals of the current and the previous iteration according to:

$$\lambda_l = \frac{\sum_{\vec{r}} \mathrm{res}_l(\vec{r}) \cdot \mathrm{res}_{l-1}(\vec{r})}{\sum_{\vec{r}} \mathrm{res}_{l-1}(\vec{r}) \cdot \mathrm{res}_{l-1}(\vec{r})} \tag{4.11}$$

with the residuals $\mathrm{res}_{l,l-1}$ given as:

$$\begin{aligned} \mathrm{res}_l(\vec{r}) &= \mathrm{est}_l^{(\mathrm{curr.})}(\vec{r}) - \mathrm{est}_l(\vec{r}) \\ \mathrm{res}_{l-1}(\vec{r}) &= \mathrm{est}_{l-1}^{(\mathrm{curr.})}(\vec{r}) - \mathrm{est}_{l-1}(\vec{r}) \end{aligned} \tag{4.12}$$

To ensure stability $0 \leq \lambda_l \leq 1$ is enforced. Using this acceleration approach a similar reconstruction quality compared to the non-accelerated version with 10 iterations can be achieved already after 7 iterations. Yielding a reduction in processing time on the order of 30%.

Damping the noise amplification:

However, still the typical noise amplification of RL deconvolution remains. A solution to this would be to find an automatic stopping criterion of the algorithm, preventing it from further amplifying noise. This comes with two problems: 1) it limits the achievable reconstruction as the iteration might end too early; 2) different regions in the image might require varying stopping criteria. Instead, a damping mechanism is introduced, which makes sure that noise amplification gets suppressed. In [55] this is achieved by modifying $\text{ratio}_l = M/(\text{est}_l \otimes h)$ into:

$$\text{ratio}_l(\vec{r}) \;\; \rightarrow \;\; 1 + G_l(\vec{r}) \cdot \left[\, \text{ratio}_l(\vec{r}) - 1 \right] \tag{4.13}$$

including a damping factor G_l. This parameter is ought to change ratio_l in such a way, that in case of convEst being close to M the value of the effective ratio_l is set to unity. This is accomplished by choosing G_l according to [55] as:

$$G_l(\vec{r}) = g_l(\vec{r})^{g_m - 1} \cdot [g_m - (g_m - 1) \cdot g_l(\vec{r})] \tag{4.14}$$

with g_m being a constant which is typically set to $g_m = 10$ [55].
The parameter g_l is given according to:

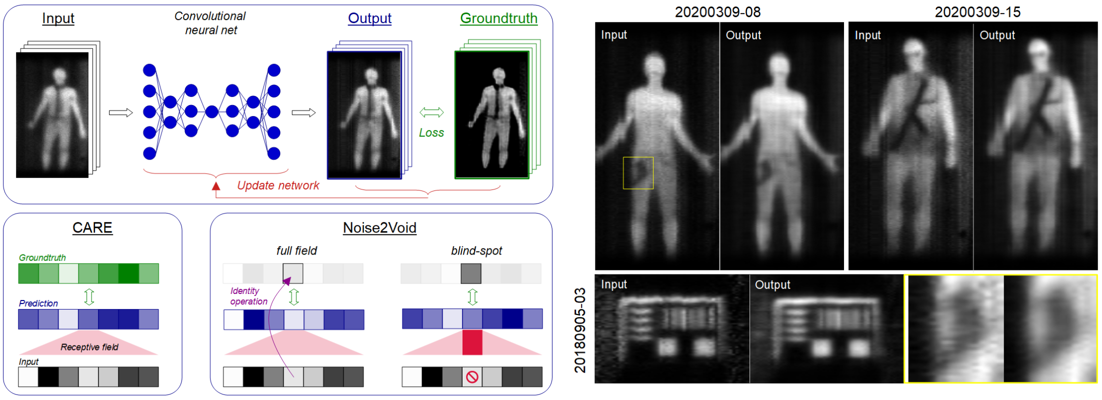
$$g_l(\vec{r}) = \gamma^{-1} \cdot \Big(M(\vec{r}) \cdot \ln[\text{ratio}_l(\vec{r})] + \text{convEst}_l(\vec{r}) - M(\vec{r}) \Big) \tag{4.15}$$

which is a function measuring the reconstruction error in relation to some threshold value γ. In case of a good reconstruction both, g_l and G_l, will approach zero, resulting in the effective ratio_l becoming 1. Which prevents noise amplification at larger iteration numbers. Is the reconstruction bad, then g_l will increase which yields a much larger value of G_l. Hence the effective ratio_l will be $\gg 1$ and the algorithm tries to improve the reconstruction in subsequent iterations. Note that this approach is spatially dependent, meaning that different regions in the image will be damped individually. Which enables an algorithm that yields better reconstructions of fine details, and at the same time homogeneous areas without strong noise amplification. Figure 4.4 shows the results of this algorithm for $\gamma = 0.01$. Note the homogeneous surface of the upper body of both persons, indicating the reduced noise amplification. Additionally the contrast of the concealed threats is much enhanced. It might seem that spatial resolution is lost, as the handgun looks slightly bigger in the reconstruction. However, when looking at the resolution target (bottom) a clear improvement in resolution becomes observable. Note that the damping strategy also works with the acceleration scheme of [54], as shown here. Enabling fast reconstruction of high *SNR* data from the security imager. The different image reconstruction methods are validated against the resolution line pattern (*20180905*), by looking at the visibility $\mathcal{V}$ and the obtainable improvement factor $IF_{\mathcal{V}}$ which is depicted in the appendix (p. 135). A maximum enhancement of 120% is obtainable for the accelerated and damped Richardson-Lucy algorithm (7 iterations). On page 136 in the appendix, two datasets (*20200309-08* and *20200309-15*) are shown as a sequence of five frames side by side. Indicating the sufficient temporal resolution to successfully observe a walking (top) and a turning (bottom) person, without missing some of the movements. However, the shape of the concealed threat can only be acquired from the data, when the object is facing the submillimeter wave imager. E.g. a flat object at the side of the person will not result in any detectable contrast, meaning that also image reconstruction techniques will not be able to improve the image quality (except with additional information from external sensors).

4.3.3 Data-driven image reconstruction

Using post-processing as a tool to improve *SNR* is limited in the sense that all reconstruction methods require some sort of mathematical model describing signal and noise. There are typically many more effects in real experiments, which cannot all be described in a single theoretical framework. Especially, since some of them might be too complicated to yield a straightforward description. A possibility to overcome this, is to obtain all information about the underlying reality of the measurement process from the recorded data itself. This is possible when a large dataset is presented to a computer algorithm, such that it can learn the underlying structure by trying to predict those measurements. A realization of this paradigm appeared in recent years due to the advance of *Deep-Learning* [134, 135], as the use of more powerful Graphical Processing Units (GPUs) [136] became feasible. Especially *convolutional neural networks* (CNNs) have lead to impressive results in image analysis (e.g. segmentation [137, 138]), when compared to more traditional methods. Figure 4.5a shows a graphical sketch of such a neural net. It consists of different layers of neurons (blue circles) and connections between them. The processing of a single neuron is shown in the appendix (p. 136). Each input signal from the previous layer is weighted with some weighting factors w_i, summed and processed by a non-linear function (activation), yielding the output of that single neuron. The corresponding weights of all neural connections are found by training on pairs of recorded images and groundtruth data.

In case of image denoising this is achieved by acquiring data corresponding to low *SNR* (image M) and high *SNR* as an groundtruth estimate (image E; green in Fig. 4.5a), as done in *Content Aware Image Reconstruction* (CARE) [140]. In each training iteration, the neural network tries to remove noise such that the loss of the network prediction compared to the groundtruth data is minimized. The obtained error is backpropagated through the neural network and effectively


Input
Convolutional neural net
Output
Groundtruth
Loss
Update network
CARE
Groundtruth
Prediction
Receptive field
Input
Noise2Void
full field
blind-spot
Identity operation
20200309-08
20200309-15
Input
Output
Input
Output
20180905-03
Input
Output

a b

Figure 4.5: Deep-Learning-based denoising: principle and results. a A convolutional neural net (blue) is used to remove noise from submillimeter wave images. This typically requires pairs of low and high *SNR* data, the latter act as an estimate of the underlying groundtruth (green). Training on low *SNR* images only is possible if a masking strategy is used, yielding in a so-called blind-spot network [139]. Removing the information of the pixel ought to be estimated from the receptive field, makes sure that the net is not learning the identity operation. b Results after Deep-Learning-based image denoising. Note that only *20200309-15* was part of the training batch. For all datasets the image quality gets improved, however spatial resolutions still gets compromised (see handgun inset or resolution line pattern).

changes the weight parameters w in such a way that the predicted result in future iterations is improved [141]. Mathematically this is written as:

$$\underset{\theta}{\operatorname{argmin}} \sum_{(i)} \sum_{\vec{r}} \mathcal{L}\left\{ \underbrace{f_\theta[M^{(i)}(\vec{r})];}_{prediction} \ \underbrace{E^{(i)}(\vec{r})}_{groundtruth} \right\} \tag{4.16}$$

with f_θ being the mapping realized by the neural network, depending on the model parameters θ (e.g. weights w, ...) and the i^{th} pair of training data. $\sum_{\vec{r}} \mathcal{L}$ represents a loss function, which is chosen to be the mean-square error loss:

$$\mathcal{L}\left\{ f_\theta[M^{(i)}(\vec{r})]; E^{(i)}(\vec{r}) \right\} = \left| f_\theta[M^{(i)}(\vec{r})] - E^{(i)}(\vec{r}) \right|^2 \tag{4.17}$$

A major disadvantage of this approach is the requirement to collect groundtruth data. This is not only cumbersome but might even be impossible in certain scenarios. Hence a method that can train such a network on low *SNR* images alone, hence only on M, is required. Lehtinen et al. [142] introduced this in form of their *Noise2Noise*-network, which requires two images with varying noise (denoted by i and j) but same underlying signal structure to train the network.

$$\begin{aligned} M^{(i)}(\vec{r}) &= E^{(i)}(\vec{r}) + \mathcal{N}^{(i)}(\vec{r}) \\ M^{(j)}(\vec{r}) &= E^{(i)}(\vec{r}) + \mathcal{N}^{(j)}(\vec{r}) \end{aligned} \tag{4.18}$$

Both having the same expectation value, enabling to make an appropriate adjustment of the minimization process:

$$\underset{\theta}{\operatorname{argmin}} \sum_{(i)} \sum_{\vec{r}} \mathcal{L}\Big\{ \underbrace{f_\theta[M^{(i)}(\vec{r})];}_{prediction} \; \underbrace{M^{(j)}(\vec{r})}_{\text{"}groundtruth\text{"}} \Big\} \tag{4.19}$$

where the target distribution $E^{(i)}$ was replaced with the noisy measurement $M^{(j)}$. Because the noise varies in each image ($M^{(i)}$ and $M^{(j)}$), the network is not able to predict an underlying noise process, enabling it to solely focus on $E^{(i)}$. Meaning that it learns the denoising due to the lack of noise predictability. A more detailed explanation of this is given in the appendix (p. 136).

Noise2Void [139, 143] achieves the same, without requiring a pair of noisy images. The only assumption made is that the noise is pixel-wise independent, meaning that it is impossible to predict the noise from surrounding neighbors. Contrary to the signal, where adjacent pixels carry some information making it possible to infer the central pixels signal value from the surrounding neighborhood. However, to prevent the network form learning the identity operation (i.e. just forwarding the input pixel), a masking scheme is required, leading to a so-called *blind-spot* network. Figure 4.5a bottom shows a comparison between the CARE and the Noise2Void strategy. In CARE the prediction of a single pixel is achieved by accumulating information from within a certain neighborhood, the *receptive field*. In Noise2Void the receptive field must be masked so that no information of the pixel which is supposed to be estimated is used in the training process. This prevents the network form learning the identity operation and enables to learn image denoising from a single image acquisition. Re-writing the minimization into:

$$\underset{\theta}{\operatorname{argmin}} \sum_{(i)} \sum_{\vec{r}} \mathcal{L}\Big\{ \underbrace{f_\theta[M_{BS}^{(i)}(\vec{r})];}_{prediction} \; \underbrace{M_{BS}^{(i)}(\vec{r})}_{\text{"}groundtruth\text{"}} \Big\} \tag{4.20}$$

with M_{BS} being the masked (or blind-spot) version of the noisy measurements, which is used for both: input and reference data. Enabling to train denoising on a single measurement, without requiring to capture groundtruth data or multiple noisy measurements. To train a Deep-Learning-based denoiser on the submillimeter wave data, the Noise2Void plugin for Fiji [144] has been used. It was trained on the image sequence *20200309-15* for approximately 10 hours on a GPU (Nvidia Geforce RTX 2070 Super). The results of applying the network to different images is shown in Fig. 4.5b, which also incorporates data that was never presented to the network during that training period. For all datasets the image quality does improve, as mainly noise is removed. Nevertheless, also this type of denoising is not able to perfectly maintain signal structure, as can be seen especially at the line pairs of the resolution target. An advantage of Deep-Learning-based denoising is that once it is trained, its runtime is low. In this sense, denoising through neural networks seems to be ideal in the use of the envisioned security application. Denoising alone might not be enough, due to physically (and technologically) limitations further resolution enhancement techniques could be required. Common image reconstruction methods based on CNNs have been adapted to also predict a deconvolved result [145, 146]. As these method are still in their development phase, an application to the passive submillimeter wave data has not yet been pursued.

4.4 Summary and discussion

In this chapter the image quality of a passive submillimeter wave imaging system has been improved, as it lacks *SNR* and spatio-temporal resolution. The goal of the presented imager is to detect concealed threats, e.g. a handgun or knife hidden underneath clothes. This can only be realized by imaging at a very distinct wavelength region $\lambda = 0.75$ mm -1 mm, as the corresponding radiation is required to be able to penetrate through fabrics. This choice of λ also sets the attainable optical resolution to 1 cm -2 cm at a distance of 10 m -15 m, which is barely enough for the proposed surveillance application. Naturally, improving spatial resolution and the *SNR* will positively affect the ability to identify possible threats. Those enhancements have been achieved in post-processing.

Pre-processing:

Each detector exhibits different offset and gain values which are corrected by subtracting previously acquired dark images and by adjusting the detector response through least-squares fitting. Due to the scanning motion several different geometric artifacts can be found in the data, which are corrected by automatically estimating the required shift using a correlation-based algorithm. Wrap-around artifacts have been removed by damping using a windowing function, while proper image normalization has been performed to generate the actual images.

Denoising:

To further improve the image quality, the present noise needs to be reduced. Two types of very basic denoising strategies have been investigated: 1) spatial filtering; 2) clipping in a transformed space. Former indicated limited use as the separation of noise & signal in real space can only be partially achieved. When first transforming the image into Fourier space, the nature of band-limited detection can be used to remove high-frequency noise-only components. Using the discrete cosine transform (DCT) shows even better denoising results, as DCT has a stronger ability to concentrate sample information in a smaller region in *k*-space.

Deconvolution:

If a further improvement in image quality is required, more elaborate techniques such as deconvolution can be used. Simple denoising & subsequent sharpening of the image results in strong artifacts. Wiener filtering is a fast reconstruction method, however, it also generates considerable noise artifacts. An iterative approach such as Richardson-Lucy deconvolution provides more flexibility in the reconstruction to yield good results, but also requires a longer runtime. The iteration process has been accelerated by the method of Biggs [54], reducing the computation time by 30%. To further overcome the problem of noise amplification, the damping strategy from [55] has been implemented.

Deep-Learning:

A novel method for image reconstruction is based on Deep-Learning. The main idea is to obtain information on the underlying sample structure of an object by investigating a large number of representative example images. A network can be trained to remove noise and artifacts not only when groundtruth data is available but also from single images by apply a masking strategy [139]. When applied to the recorded submillimeter data an enhancement of image quality was observable, compared to traditional methods. Deep-Learning-based deconvolution techniques are currently being developed and might bring an unprecedented enhancement in image restoration. An advantage of Deep-Learning-based methods is that after training their application is computationally inexpensive and therefore perfectly suited to the live application of concealed threat detection.

Chapter 5

Conclusion

The performance of any imaging system does not only depend on its ability to transfer *sample* information, but also on the amount of *noise* present in the respective measurement process. Meaning that imaging itself is not only characterized by the attainable spatio-temporal resolution alone, but rather depends on the achievable signal-to-noise ratio (*SNR*).

Many noise sources can be minimized or completely eliminated, e.g. by improving detector technology. However, the detection of photons poses a fundamental noise limit (shot noise), which cannot be reduced through technological advancements. Any noise type is essentially characterized by the variance it imposes on the measurement. In case of shot noise the variance is equal to the expected number of photons, meaning that the shot noise limited *SNR* depends on the amount of detectable photons. Often this number is limited (e.g. finite photon budget in fluorescence microscopy), which inherently restricts the achievable *SNR*. A bad *SNR* is able to drastically reduce image quality, even for advanced methods such as super-resolution techniques. Hence achieving maximum *SNR* at low photon numbers is of strong interest, as this would benefit many imaging applications. In principle the *SNR* can be improved using two different approaches: 1) collect more photons on a hardware basis; 2) enhance *SNR* computationally by software. The latter requires mathematical models describing the actual measurement process correctly. For example: in deconvolution an image formation and noise model is specifically assumed, from which the corresponding reconstruction algorithm is derived. In case the underlying true image formation or noise characteristics differ from that assumption, this will lead to poor reconstruction results and thus restrict their use. However, an advantage of image reconstruction is that it also works in post-processing, enabling to improve already recorded datasets. In this thesis, that advantage has been used with image data of a passive submillimeter wave imager, where an optimized pre-processing strategy was developed to later obtain a strong *SNR* improvement using image reconstruction methods. A better reconstruction result often requires the respective algorithm to either have more information on the image acquisition available (e.g. PSF of optical system, sparsity constraint on sample, ...), allow for a longer computation time (e.g. iterative methods) or both. For the application of concealed threat detection it is quite important that the image enhancement methods work robustly and fast enough so that an potential suspect can be tracked efficiently. An accelerated & damped Richardson-Lucy algorithm has been used to improve *SNR* and resolution of the acquired data and has proved to fulfill the robustness and speed requirements needed. Nevertheless, in the future such an iterative algorithms might be replaced by Deep-Learning methods, where the effort on adjusting reconstruction parameters is shifted towards the training period. Meaning that the actual reconstruction step can be executed at a very low runtime. However, the training step requires the use of graphical processing units (GPUs), as it is very computationally expensive. In the current work only Deep-Learning-based *denoising* has been used, first ideas how to realize an equivalent *deconvolution* technique are still under development. In general, the use of computational reconstruction always comes with the burden of obtaining artifacts which usually cannot be discerned from actual signal structures. This is especially true if noisy data is used, hence improving the *SNR* already at the acquisition level also has a positive effect on the subsequent image reconstruction.

Signal-to-noise enhancement by developing increasingly sophisticated imaging devices has been one of the major directions of research in fields related to imaging for the past 10 - 20 years. When looking at photon detection in the visible range, the current technology is far advanced so that any further *SNR* improvements requires to build even more specialized imaging setups which often come with tradeoffs. For example: whenever shot noise limits image quality it is necessary to increase the number of detectable photons, which is often achieved by increasing the exposure time of the detector, at the cost of a reduced temporal resolution. Meaning that many modern imaging technologies do get their improvement of image quality from some sort of tradeoff. E.g. localization-based microscopy requires the collection of a large number of photons to obtain a high improvement in terms of spatial resolution, whenever the samples moves to quickly this technique cannot be applied as motion blur will occur (due to the low temporal resolution); STED or confocal imaging rely on strong illumination intensities over a short amount of time which is disadvantageous in terms of photobleaching and phototoxicity. In this thesis an approach has been suggested, which tries to obtain an improved *SNR* without demanding any increase in photon number, termed *splitting & recombination*. This is realized by modifying the experimental setup of a conventional imaging system, such that a splitting of image information occurs. Subsequent computational recombination is then able to make more use of the split data, achieving an improvement in terms of *SNR*. The presented work exemplifies four different splitting mechanisms: 1) polarization; 2) fluorescence emission; 3) pupil and 4) illumination splitting. The individual implementation of the respective splitting mechanism require minor adjustments in the imaging setup (e.g. re-imaging of detection objective BFP, ...) but essentially enable the recording of multiple *views*, often even simultaneously. In principle this idea is already being followed with Image Scanning Microscopy (ISM) [99, 100], where it was noticed that replacing the detection pinhole in a confocal microscope with a fast spatially resolved detector, yields superior *SNR* while an improvement in terms of resolution and optical sectioning is achieved via post-processing. In the presented splitting & recombination approach, the computational reconstruction is either done by multiview deconvolution or weighted averaging in Fourier space. The latter gives the opportunity to theoretically describe the potential *SNR* enhancement when looking at a noise-normalized effective optical transfer function. A further improvement in terms of reconstruction capability might be achievable when using Deep-Learning methods, as it would be possible for a neural network to make use of multiple sub-images as an input with the aim to reconstruct the underlying (most probable) sample distribution. Note that splitting the data can come at the cost of a reduced field-of-view in case a single camera is used, or requires to run two cameras simultaneously. However, as the size of camera detectors will increase in the future, this seems to be a small price to pay. In case of the fluorescence splitting the acquisition of the individual wavelength bands can be achieved in different ways, but always come with some drawback: either reduced temporal resolution (time sequential recording) or spatial resolution (snap shot imagers, such as in [75]). Note that in this thesis additional noise sources (such as detector readout noise, ...) were neglected. In imaging scenarios where those largely determine the unwanted noise contribution splitting & recombination will not be able to achieve a strong *SNR* enhancement. Hence it is quite crucial that all other noise sources are minimized as much as possible.

A fundamental challenge in optical imaging applications is obtaining as much information about the sample as possible with the minimum available amount of photons. The potential impact of this thesis lies in principles and their implementations of how to make better use of photons in an imaging application - a crucial step for future advances in the photon limited imaging regime.

Appendix A

Supplemental material

Contents:

Theoretical background: page

1. Expectation value and variance of a complex quantity 102
2. Effective resolution measure for two separated point sources 102
3. Scalar widefield OTF and quadratic fit at large frequencies 105
4. Sampling, Nyquist limit and the pixel form factor 106
5. Image reconstruction using a Wiener filter 107
6. Maximum likelihood reconstruction using RL method 109
7. Optimize reconstruction using normalized cross-correlation *NCC* 111

Fluorescence microscopy:

1. Polarization effects in vectorial PSFs 113
2. Weighted averaging in Fourier space 114
3. Log-normal distribution for vanishing skewness 118
4. Solution to the integrals representing A and λ_{CoM} 118
5. Estimating the spatial broadening $\mathcal{B}$ 119
6. The scanning confocal point spread function 121
7. Image results of wavelength splitting & recombination 122
8. Estimating the optimum iteration number using *NCC* 124

9. Calculating the OTF via autocorrelation of the pupil 125
10. Improvement factor for varying pupil splitting radius 126
11. Elongating the depth-of-field in pupil splitting 127
12. Computing the sub-pupils using the Richards & Wolf model 128
13. Image results of pupil splitting & recombination 128
14. The aplanatic factor modifying the pupil splitting 129
15. Field-Synthesis reduces phototoxicity 132
16. Reducing out-of-focus light through illumination splitting 133
17. Image results of illumination splitting & recombination 134

Submillimeter wave imaging:

1. Mean-variance relationship of submillimeter detectors 135
2. Improved spatial resolution in submillimeter wave imaging 135
3. Temporal representation of the recorded video data 136
4. Supervised and unsupervised learining based denoising 136

Theoretical background:

1 Expectation value and variance of a complex quantity

When dealing with the complex-valued image data in Fourier space, expectancy and variance need to be quantified. General definitions of those are [37]:

$$\begin{aligned} \tilde{E}_{2D}(\vec{k}) &= \tilde{E}_{2D,Re}(\vec{k}) + i \cdot \tilde{E}_{2D,Im}(\vec{k}) && \text{(A.1)} \\ \sigma^2_{\tilde{M}_{2D}}(\vec{k}) &= \sigma^2_{\tilde{M}_{2D,Re}}(\vec{k}) + \sigma^2_{\tilde{M}_{2D,Im}}(\vec{k}) && \text{(A.2)} \end{aligned}$$

With the subscript Re / Im indicating real and imaginary part respectively. It is important to see that expectancy and variance are defined over two different domains:

$$\begin{aligned} \tilde{E}_{2D}(\vec{k}) &\in \mathbb{C} && \text{(A.3)} \\ \sigma^2_{\tilde{M}_{2D}}(\vec{k}) &\in \mathbb{R} && \text{(A.4)} \end{aligned}$$

If $\tilde{M}$ gets multiplied by a constant, the change in variance is given according to [31] as:

$$\sigma^2_{c\cdot\tilde{M}_{2D}}(\vec{k}) = cc^* \cdot \sigma^2_{\tilde{M}_{2D}}(\vec{k}) \tag{A.5}$$

with cc^* being equal to the absolute square of c.

2 Effective resolution measure for two separated point sources

In sec. 2.3.3 it has been shown that an effective resolution can be defined using the basic criterion $|\tilde{E}_{2D}| \geq \sqrt{p}$, which leads to:

$$|\tilde{h}(\vec{k})| \cdot |\tilde{S}(\vec{k})| \geq \sqrt{p} \tag{A.6}$$

Indicating that the effective resolution limit depends on the sample structure under investigation. Conventionally two separated point-sources are chosen and the resolution of the optical system is given as the smallest distance at which both points can be visually separated from each other. A criterion which is often stated for this particular example is the *Rayleigh criterion* [147]. For two point objects, separated by $d_{xy,min}$, the spectrum of the sample structure can be written as:

$$\tilde{S}(k_{xy}) = p \cdot \cos\left(\tfrac{1}{2} d_{xy,min} \cdot k_{xy}\right) \tag{A.7}$$

with p indicating the total number of expected photons and k_{xy} lateral spatial frequencies. The absolute value of this is given according to:

$$|\tilde{S}(k_{xy})| = \sqrt{p^2 \cdot \cos^2\left(\tfrac{1}{2} d_{xy,min} \cdot k_{xy}\right)} \tag{A.8}$$

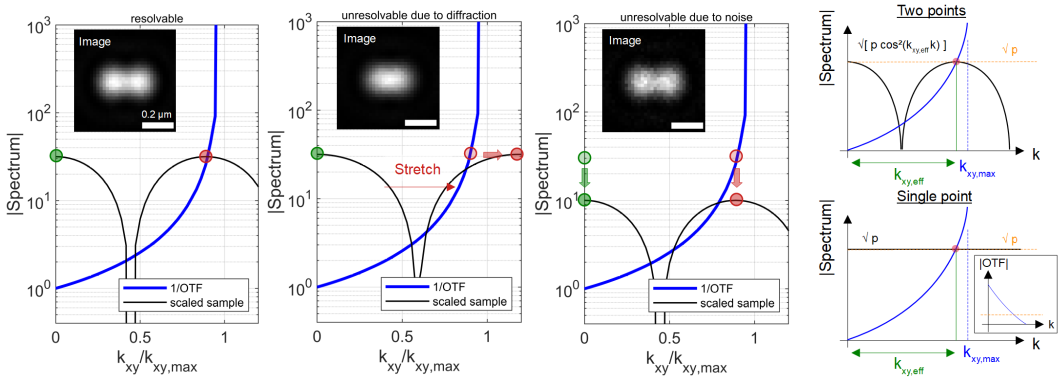
Rewriting the basic criterion from eq. A.6, yields:

$$\frac{|\tilde{S}(k_{xy})|}{\sqrt{p}} \geq \frac{1}{|\tilde{h}(k_{xy})|} \tag{A.9}$$

Which means that the scaled sample spectrum needs to be bigger than the inverse of the |OTF|. Using the aforementioned spectrum of the two point objects leads to:

$$\sqrt{p \cdot \cos^2\left(\tfrac{1}{2} d_{xy,min} \cdot k_{xy}\right)} \geq \frac{1}{|\tilde{h}(k_{xy})|} \tag{A.10}$$

This expression is graphically depicted in Fig. A.1, with blue being the inverse OTF and black the scaled sample spectrum.



a b c d

Figure A.1: Resolution capability for two separated point sources, including shot noise. a shows the inverse OTF (blue, see text) and the scaled sample spectrum (black) in log-scale on the |Spectrum|-axis. The criterion for resolving the two separated point objects is that the maximum object information (red dot) lies above the OTF curve. This is fulfilled in the first example and the small inset shows that both point objects are still resolvable. Moving those closer together will yield the black curve to be stretched along the $k_{xy}/k_{xy,max}$-axis, resulting in the red dot leaving the region where $|\tilde{E}_{2D}| \geq \sqrt{p}$. Hence the imaged points are note separable any more. Another possibility to achieve this is when the number of detectable photons p is small, resulting a vertical shift of the black curve. Effectively making the two (real space) points not separable although the separation is still within Abbe's resolution limit $d_{xy,min}$. Be aware that also your personal perception can play a role in what you label "resolvable" or not. Schematic showing that the effective resolution limit for two separated point sources is equal to imaging a single point and solving for the intercept of noise floor.

The red dot marks the maximum object frequency which needs lie above the blue curve, so that both object points can still be successfully separated. When the distance between the two object points $d_{xy,min}$ is reduced, the black curve in Fig. A.1 scales inversely along the horizontal axis (red arrow), resulting in the red dot leaving the $|\tilde{E}_{2D}| \geq \sqrt{p}$ region. As the simulation in the inset suggests, both points cannot be separated anymore. Another possibility to achieve this is to decrease the total number of photons p. This yields in a vertical shift of the black curve, also moving the red dotted outside the $|\tilde{E}_{2D}| \geq \sqrt{p}$ region. The image in real space depicts the same two point sources as in Fig. A.1a, however with much stronger shot noise making it more difficult to isolate both. Note that also your perception plays a role and that your brain might expect to see two isolated points in above figure.

The position of the red dot in terms of spatial frequencies is given when $\tilde{S}$ has reached one period of the $\cos^2$ at:

$$\tfrac{1}{2} d_{xy,min} \cdot k_{xy} = \pi \tag{A.11}$$

The position of this point on the |Spectrum|-axis of this graph is then given according to:

$$\sqrt{p \cdot \cos^2 \pi} = \sqrt{p} \tag{A.12}$$

Overall this yields the following effective resolution criterion, when imaging two separated point sources:

$$|\tilde{h}(k_{xy})| \geq \tfrac{1}{\sqrt{p}} \tag{A.13}$$

Note that this is exactly the same as when searching for the intercept of $|\tilde{h}|$ with the noise floor, located at $1/\sqrt{p}$. As for a single point the sample spectrum is given as:

$$\tilde{S}(k_{xy}) = p \rightarrow |\tilde{S}(k_{xy})| = p \tag{A.14}$$

According to the criterion $|\tilde{E}_{2D}| \geq \sqrt{p}$, this yields in:

$$\frac{p}{\sqrt{p}} = \sqrt{p} \geq \frac{1}{|\tilde{h}(k_{xy})|} \rightarrow |\tilde{h}(k_{xy})| \geq \tfrac{1}{\sqrt{p}} \tag{A.15}$$

Which again is the same as the "two point" effective resolution limit. Figure A.1d shows this in same form as the two point resolution criterion in Fig. A.1a-A.1c.

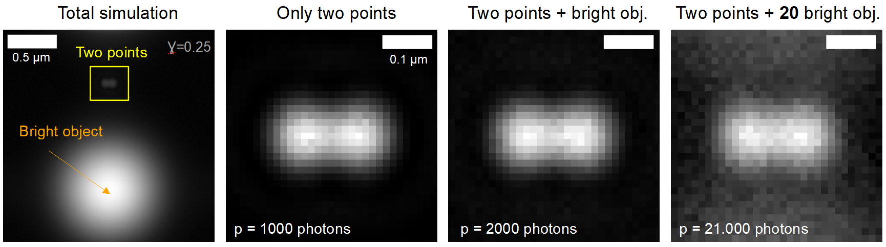


Figure A.2: Example showing the influence of localized noise on spatial resolution. In this simulation two point sources are put offset from the center of the FoV. The distance between both objects is large enough so that they can be resolve with a total number of photons $p = 1000$. A bright object is added at the bottom of the FoV and its brightness is scaled to match the brightness of both imaged points. However, when the brightness of the big spot at the bottom is drastically increased, the light corrupts the region around the two distinct point sources. Yielding in both points not being able to be separated, although the number of available photons is large.

Another problem arising when trying to define a reliable resolution criterion is the fact, that shot noise is localized in real space. Which makes it difficult to find a resolution criterion based in Fourier space, as the Fourier transform generally represents the whole image in terms of spatial frequency. An example: for a certain number of photons p and separation $d_{xy,min}$, two point objects can be resolved. Now, if a very bright object is added at a distance from these two points, it is possible that the light of the bright object "pollutes" the region of the two points sources. This can go so far that both points cannot be resolved anymore, as the bright object basically dominates the field-of-view. Such an example is shown in Fig. A.2.

Note that by adding the bright object, the total number of photons has been increased drastically ($p = 1000 \rightarrow 21000$). We should now be able to easily resolve the two point objects, according to the previously introduced effective resolution limit. However, this is not the case, as the criterion is based in Fourier space and cannot deal with situations as described above. Because of this the effective resolution limit used in this thesis is limited to isolated, single objects of interest.

3 Scalar widefield OTF and quadratic fit at large frequencies

The OTF of an imaging system can be calculate from the pupil function, which is a spherical cap on the *Ewald* sphere [33] (see Fig. 2.2b right top). For low *NA* detection the pupil can be approximated by a disk, as it is possible to neglect the k_z-direction. The OTF can then be computed from the pupil $\mathcal{P}_{2D}$ as an *autocorrelation* $\mathcal{A}$ [30]:

$$\mathcal{A}\left\{\mathcal{P}_{2D}(\vec{k})\right\} = \int_{-\infty}^{+\infty} d\vec{k}' \; \mathcal{P}_{2D}(\vec{k}') \cdot \mathcal{P}^{*}_{2D}(\vec{k}' - \vec{k}) \tag{A.16}$$

In an ideal system (aberration free), the pupil is strictly real-valued [31]. Then the autocorrelation can be visualized as shown in Fig. A.3 (based on [30]).

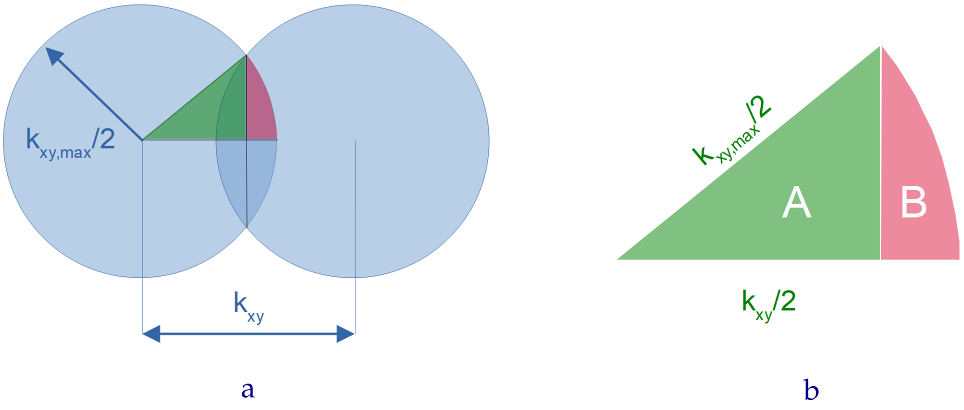


Figure A.3: a Graphical representation of the autocorrelation of the 2D-pupil (based on [30]). Take two copies of the pupil (= disk), separate them by k_{xy} and calculate the overlapping area. Do this for all possible shifts and plot the obtained overlap. This curve is equal to the widefield OTF in scalar approximation. b To obtain the overlap for a specific separation we need to calculate the area B and multiply it by four. This is done by first calculating a circular sector which has the area $A + B$ and then substracting the area of the triangle with area A.

Take two copies of the pupil, with the diameter equal to $k_{xy,max}$. Separate them by shifting one of them by k_{xy} into one direction, as shown in Fig. A.1a. Calculate the overlapping area for all possible k_{xy} and plot them in a graph. The result will be the falling curve shown in Fig. 2.2a right in blue, as for increasing the separation the overlap gets smaller. To obtain the overlap area, we need to calculate the area of B in Fig. A.1b and multiply it by four. B is given as a circular sector, which area can be calculated as [31]:

$$A + B = \tfrac{\theta}{2\pi} \cdot \left(\tfrac{k_{xy,max}}{2}\right)^2 \pi \tag{A.17}$$

with θ being the angle to the left, in the green triangle. Its area (A) is given by:

$$A = \tfrac{1}{2}\tfrac{k_{xy}}{2} \cdot \sqrt{\left(\tfrac{k_{xy,max}}{2}\right)^2 - \left(\tfrac{k_{xy}}{2}\right)^2} \tag{A.18}$$

Replacing θ with an arccos-expression and then calculating $4 \cdot B$, gives the OTF $\tilde{h}$ as:

$$\begin{aligned} \tilde{h}(k_{xy}) &= 4 \cdot [(A+B) - A] = & \text{(A.19)} \\ &= \left(\frac{k_{xy,max}}{2}\right)^2 \cdot \left[\arccos\left(\frac{k_{xy}}{k_{xy,max}}\right) - \frac{k_{xy}}{k_{xy,max}} \cdot \sqrt{1 - \left(\frac{k_{xy}}{k_{xy,max}}\right)^2}\right] & \text{(A.20)} \end{aligned}$$

Finally, the expression is normalized with respect to the full area of the pupil:

$$\tilde{h}(k_{xy}) \approx \frac{2}{\pi}\left[\arccos\left(\frac{k_{xy}}{k_{xy,max}}\right) - \frac{k_{xy}}{k_{xy,max}} \cdot \sqrt{1 - \left(\frac{k_{xy}}{k_{xy,max}}\right)^2}\right] \tag{A.21}$$

This is the theoretically calculated OTF for a low *NA* detection system. To find the effective resolution limit, $\tilde{h}$ needs to be approximated in a region close to the cutoff frequency $k_{xy,max}$. This is done by fitting a quadratic equation using the following three points:

$$\begin{aligned} \frac{k_{xy}}{k_{xy,max}} &= 1.0 = k_1 &\rightarrow\quad \tilde{h}\left(\frac{k_{xy}}{k_{xy,max}}\right) &= 0 = \tilde{h}_1 \\ \frac{k_{xy}}{k_{xy,max}} &= 0.8 = k_2 &\rightarrow\quad \tilde{h}\left(\frac{k_{xy}}{k_{xy,max}}\right) &= 0.1 = \tilde{h}_2 \\ \frac{k_{xy}}{k_{xy,max}} &= 0.9 = k_3 &\rightarrow\quad \tilde{h}\left(\frac{k_{xy}}{k_{xy,max}}\right) &= 0.0375 = \tilde{h}_3 \end{aligned}$$

So that the three points are given as $(k_1, \tilde{h}_1)$, $(k_2, \tilde{h}_2)$ and $(k_3, \tilde{h}_3)$.
The fitted equation is given as:

$$\tilde{h}(k_{xy} \geq 0.8 \cdot k_{xy,max}) \approx a \cdot \left(\frac{k_{xy}}{k_{xy,max}}\right)^2 + b \cdot \frac{k_{xy}}{k_{xy,max}} + c \tag{A.22}$$

The linear system involved can be solve for the three parameters a, b and c, yielding:

$$\begin{aligned} a &= \frac{k_1 \cdot (\tilde{h}_3 - \tilde{h}_2) + k_2 \cdot (\tilde{h}_1 - \tilde{h}_3) + k_3 \cdot (\tilde{h}_2 - \tilde{h}_1)}{(k_2 - k_1) \cdot (k_3 - k_1) \cdot (k_2 - k_3)} = 1.4507 \\ b &= \frac{k_1^2 \cdot (\tilde{h}_3 - \tilde{h}_2) + k_2^2 \cdot (\tilde{h}_1 - \tilde{h}_3) + k_3^2 \cdot (\tilde{h}_2 - \tilde{h}_1)}{(k_1 - k_2) \cdot (k_1 - k_3) \cdot (k_3 - k_2)} = -3.1314 \\ c &= \frac{k_1^2 \cdot (k_2\tilde{h}_3 - k_3\tilde{h}_2) + k_2^2 \cdot (k_3\tilde{h}_1 - k_1\tilde{h}_3) - k_3^2 \cdot (k_1\tilde{h}_2 - k_2\tilde{h}_1)}{(k_2 - k_1) \cdot (k_3 - k_1) \cdot (k_2 - k_3)} = 1.6807 \end{aligned}$$

With these parameters the OTF can now be approximated according to equation A.22.

4 Sampling, Nyquist limit and the pixel form factor

Any measuring device approximates a continuous function $f_{contin.}$ with a sampled version $f_{sampl.}$. This can be mathematically expressed as a multiplication with a δ-comb (here only in shown in 1D for simplification) [35, 46]:

$$f_{sampl.}(x) = \left[f_{contin.}(x) \otimes rect(x)\right] \cdot \sum_m \delta(x - m\Delta s) \tag{A.23}$$

with Δs being the sampling interval in real space.
The convolution with the *rect*-function models the photosensitive area of each pixel in a

camera sensor. In Fourier space this equation can be rewritten as:

$$\tilde{f}_{sampl.}(k) = \left[\tilde{f}_{contin.}(k) \cdot \mathcal{F}_{1D}\{rect(x)\}\right] \otimes \sum_m \delta\left(k - m\tfrac{2\pi}{\Delta s}\right) \tag{A.24}$$

with $\mathcal{F}_{1D}\{rect(x)\}$ being the so-called *pixel form factor*.
Aforementioned equation states that in frequency space the sampled version $\tilde{f}_{sampl.}$ is first multiplied by the pixel form factor, which can be thought of as a step that effectively reduces $\tilde{h}$ at high spatial frequencies (decreasing contrast in real space [56]). Throughout this thesis we assume that the pixel size is reasonably small, and we can neglect the influence of the pixel form factor, so that:

$$\tilde{f}_{sampl.(k)} = \tilde{f}_{contin.}(k) \otimes \sum_m \delta\left(k - m\tfrac{2\pi}{\Delta s}\right) \tag{A.25}$$

Effectively this means that the Fourier transform of the sampled function consists of multiple copies at a spacing interval of $2\pi/\Delta s$. The smallest structure that can be imaged with an (linear) imaging system is the point-spread-function h, which is given in Fourier space as the OTF $\tilde{h}$. It has a finite width, the band limit given as $2 \cdot k_{xy,max}$ (as diameter). *Nyquist* sampling states that the individual copies should not overlap [9, 10], otherwise the effect of *aliasing* appears [34]. To fulfill Nyquist sampling the following criterion is given:

$$\frac{2\pi}{\Delta s} \geq 2 \cdot k_{xy,max} = 2 \cdot \frac{2\pi}{d_{xy,min}} \tag{A.26}$$

Or in real space given as a minimum sampling distance:

$$\Delta s \leq \tfrac{1}{2} \cdot d_{xy,min} \tag{A.27}$$

Otherwise *aliasing* occurs and the original function $\tilde{f}_{contin.}$ cannot be recovered successfully. Throughout this thesis it is assumed that correct sampling is used and *SNR* degradation due to the pixelation effect is neglected.

5 Image reconstruction using a Wiener filter

Lets define the difference between true S and estimated sample S_{est} using a Gaussian distribution, given as:

$$P[S(\vec{r}), S_{est}(\vec{r})] \propto e^{\frac{-|S(\vec{r}) - S_{est}(\vec{r})|^2}{2 \cdot \sigma^2}} \tag{A.28}$$

Note that we have assumed that the standard deviation σ is independent from $\vec{r}$.
The log-likelihood of S_{est} being the true sample distribution is given as:

$$\mathcal{L}(\vec{r}) = \int_O dO \;\ln P[S(\vec{r}), S_{est}(\vec{r})] = \left\langle \ln\left(P[S, S_{est}](\vec{r})\right) \right\rangle \tag{A.29}$$

with the integration being over all possible outcomes O. The logarithm applied to the Gaussian function leaves the log-likelihood according to:

$$\mathcal{L}(\vec{r}) \propto \langle |S(\vec{r}) - S_{est}(\vec{r})|^2 \rangle = \varepsilon(\vec{r}) \tag{A.30}$$

Which is defined here as the error metric ε and represents the mean-square error. Using *Plancherel's* theorem [31] this can be expressed in Fourier space as:

$$\tilde{\varepsilon}(\vec{k}) = \langle|\tilde{S}(\vec{k}) - \tilde{S}_{est}(\vec{k})|^2\rangle \tag{A.31}$$

The estimated sample is given, after filtering with the Wiener filter $\tilde{W}$, according to:

$$\tilde{S}_{est}(\vec{k}) = \tilde{W}(\vec{k}) \cdot \tilde{M}(\vec{k}) \tag{A.32}$$

The measured data is given as the sum of expectation value $\tilde{E}$ and noise term $\tilde{\mathcal{N}}$:

$$\begin{aligned} \tilde{M}(\vec{k}) &= \tilde{E}(\vec{k}) + \tilde{\mathcal{N}}(\vec{k}) = \\ &= \tilde{S}(\vec{k}) \cdot \tilde{h}(\vec{k}) + \tilde{\mathcal{N}}(\vec{k}) \end{aligned} \tag{A.33}$$

With this the error metric $\tilde{\varepsilon}$ can be written as:

$$\tilde{\varepsilon}(\vec{k}) = \langle\left|\tilde{S}(\vec{k}) \cdot \left[1 - \tilde{W}(\vec{k})\tilde{h}(\vec{k})\right] - \tilde{W}(\vec{k})\tilde{\mathcal{N}}(\vec{k})\right|^2\rangle \tag{A.34}$$

Computing the $|.|^2$ operation leads to:

$$\begin{aligned} \tilde{\varepsilon}(\vec{k}) &= \langle|\tilde{S}(\vec{k})|^2 \cdot \left|\left[1 - \tilde{W}(\vec{k})\tilde{h}(\vec{k})\right]\right|^2 + |\tilde{W}(\vec{k})|^2|\tilde{\mathcal{N}}(\vec{k})|^2 \\ &- 2\,\mathrm{Re}\left\{\tilde{S}(\vec{k}) \cdot \left[1 - \tilde{W}(\vec{k})\tilde{h}(\vec{k})\right] \cdot \tilde{W}^*(\vec{k})\tilde{\mathcal{N}}^*(\vec{k})\right\}\rangle \end{aligned} \tag{A.35}$$

The expectation value is linear, hence the previous equation can be simplified to:

$$\begin{aligned} \tilde{\varepsilon}(\vec{k}) &= |\tilde{S}(\vec{k})|^2 \cdot \left|\left[1 - \tilde{W}(\vec{k})\tilde{h}(\vec{k})\right]\right|^2 + |\tilde{W}(\vec{k})|^2\langle|\tilde{\mathcal{N}}(\vec{k})|^2\rangle \\ &- 2\,\mathrm{Re}\left\{\tilde{S}(\vec{k}) \cdot \left[1 - \tilde{W}(\vec{k})\tilde{h}(\vec{k})\right] \cdot \tilde{W}^*(\vec{k})\langle\tilde{\mathcal{N}}(\vec{k})^*\rangle\right\} \end{aligned} \tag{A.36}$$

Note that the expression inside of Re is essentially multiplying $\tilde{S} \cdot \langle\tilde{\mathcal{N}}^*\rangle$. As the true sample and noise are assumed to be uncorrelated, this multiplication yields zero. Which further simplifies the error term $\tilde{\varepsilon}$ into:

$$\begin{aligned} \tilde{\varepsilon}(\vec{k}) &= |\tilde{S}(\vec{k})|^2 \cdot \left[1 - \tilde{W}(\vec{k})\tilde{h}(\vec{k})\right]\left[1 - \tilde{W}^*(\vec{k})\tilde{h}^*(\vec{k})\right] \\ &+ \tilde{W}(\vec{k})\tilde{W}^*(\vec{k}) \cdot \langle|\tilde{\mathcal{N}}(\vec{k})|^2\rangle \end{aligned} \tag{A.37}$$

To find the minimum of the mean-square error, its first derivative needs to be computed and set equal to zero. Be aware that complex calculus needs to be applied here, i.e. the *Wirtinger* derivatives [31]. Which results in:

$$\frac{\partial\tilde{\varepsilon}(\vec{k})}{\partial\tilde{W}(\vec{k})} = |\tilde{S}(\vec{k})|^2 \cdot \left[\tilde{W}^*(\vec{k})\tilde{h}(\vec{k})\tilde{h}^*(\vec{k}) - \tilde{h}(\vec{k})\right] + \tilde{W}^*(\vec{k}) \cdot \langle|\tilde{\mathcal{N}}(\vec{k})|^2\rangle \tag{A.38}$$

The advantage of Wirtinger derivatives is that those can be easily represented in term of real and imaginary part:

$$\frac{\partial\tilde{\varepsilon}(\vec{k})}{\partial\tilde{W}(\vec{k})} = \frac{\partial\tilde{\varepsilon}(\vec{k})}{\partial\tilde{W}_{Re}(\vec{k})} - i\frac{\partial\tilde{\varepsilon}(\vec{k})}{\partial\tilde{W}_{Im}(\vec{k})} \tag{A.39}$$

with $\tilde{W} = \tilde{W}_{Re} + i \cdot \tilde{W}_{Im}$. Therefore $\partial\tilde{\varepsilon}/\partial\tilde{W}$ needs to be separated into Re & Im part:

$$\frac{\partial\tilde{\varepsilon}(\vec{k})}{\partial\tilde{W}_{Re}(\vec{k})} = |\tilde{S}(\vec{k})|^2 \cdot \left[\tilde{W}_{Re}(\vec{k})|\tilde{h}(\vec{k})|^2 - \tilde{h}_{Re}(\vec{k})\right] + \tilde{W}_{Re}(\vec{k})\langle|\tilde{\mathcal{N}}(\vec{k})|^2\rangle \tag{A.40}$$

$$i \cdot \frac{\partial\tilde{\varepsilon}(\vec{k})}{\partial\tilde{W}_{Im}(\vec{k})} = |\tilde{S}(\vec{k})|^2 \cdot \left[\tilde{W}_{Im}(\vec{k})|\tilde{h}(\vec{k})|^2 + \tilde{h}_{Im}(\vec{k})\right] + \tilde{W}_{Im}(\vec{k})\langle|\tilde{\mathcal{N}}(\vec{k})|^2\rangle \tag{A.41}$$

Setting both, real and imaginary part, equal to zero gives us:

$$\tilde{W}_{Re}(\vec{k})\left[|\tilde{S}(\vec{k})|^2 \cdot |\tilde{h}(\vec{k})|^2 + \langle|\tilde{\mathcal{N}}(\vec{k})|^2\rangle\right] = |\tilde{S}(\vec{k})|^2\tilde{h}_{Re}(\vec{k}) \tag{A.42}$$

$$\tilde{W}_{Im}(\vec{k})\left[|\tilde{S}(\vec{k})|^2 \cdot |\tilde{h}(\vec{k})|^2 + \langle|\tilde{\mathcal{N}}(\vec{k})|^2\rangle\right] = -|\tilde{S}(\vec{k})|^2\tilde{h}_{Im}(\vec{k}) \tag{A.43}$$

This can now be fully separated to find an expression for the Wiener filter [51]:

$$\tilde{W}_{Re}(\vec{k}) = \frac{|\tilde{S}(\vec{k})|^2 \cdot \tilde{h}_{Re}(\vec{k})}{|\tilde{S}(\vec{k})|^2 \cdot |\tilde{h}(\vec{k})|^2 + \langle|\tilde{\mathcal{N}}(\vec{k})|^2\rangle} = \frac{\tilde{h}_{Re}(\vec{k})}{\tilde{h}(\vec{k})\tilde{h}^*(\vec{k}) + \frac{\langle|\tilde{\mathcal{N}}(\vec{k})|^2\rangle}{|\tilde{S}(\vec{k})|^2}} \tag{A.44}$$

$$\tilde{W}_{Im}(\vec{k}) = \frac{-|\tilde{S}(\vec{k})|^2 \cdot \tilde{h}_{Im}(\vec{k})}{|\tilde{S}(\vec{k})|^2 \cdot |\tilde{h}(\vec{k})|^2 + \langle|\tilde{\mathcal{N}}(\vec{k})|^2\rangle} = \frac{-\tilde{h}_{Im}(\vec{k})}{\tilde{h}(\vec{k})\tilde{h}^*(\vec{k}) + \frac{\langle|\tilde{\mathcal{N}}(\vec{k})|^2\rangle}{|\tilde{S}(\vec{k})|^2}} \tag{A.45}$$

After further simplification this ends with the final Wiener filter equation:

$$\tilde{W}(\vec{k}) = \tilde{W}_{Re} + i \cdot \tilde{W}_{Im} = \frac{\tilde{h}^*(\vec{k})}{\tilde{h}(\vec{k})\tilde{h}^*(\vec{k}) + K(\vec{k})} \tag{A.46}$$

with the regularization parameter $\tilde{K}$:

$$\tilde{K}(\vec{k}) = \frac{\langle|\tilde{\mathcal{N}}(\vec{k})|^2\rangle}{|\tilde{S}(\vec{k})|^2} \tag{A.47}$$

6 Maximum likelihood reconstruction using RL method

Richardson-Lucy deconvolution is a statistical method [52, 53], hence tries to maximize the probability $P(S)$ of S being the true sample. Using *Bayes* theorem [31] this can be rewritten in terms of the probability $P(M|S)$, describing how likely it is to detect the image M under the assumption that S is the underlying sample.

$$P[S(\vec{r})] = \frac{P[M(\vec{r})|S(\vec{r})] \cdot P[S(\vec{r})]}{P[M(\vec{r})]} \tag{A.48}$$

With $P[M]$ being a sample independent factor, which can be neglected for the maximization process. $P[S]$ is regarded as *a priori* knowledge about the sample and is typically used as a regularization option. We omit this option here and only try to maximize $P[M|S]$, which is given as the Poisson distribution (with expectancy E):

$$P[M(\vec{r})|S(\vec{r})] = \frac{E(\vec{r})^{M(\vec{r})}}{M(\vec{r})!} \cdot e^{-E(\vec{r})} \tag{A.49}$$

The overall likelihood $\mathcal{L}$ to measure the given image, is the product of all the individual pixel probabilities:

$$\mathcal{L} = \prod_{\vec{r}} P[M(\vec{r})|S(\vec{r})] \tag{A.50}$$

Instead of maximizing the likelihood, the negative *log*-likelihood is minimized:

$$-\ln \mathcal{L} = \sum_{\vec{r}} P[M(\vec{r})|S(\vec{r})] = \sum_{\vec{r}} \left[E(\vec{r}) + M(\vec{r})! - M(\vec{r}) \cdot \ln E(\vec{r})\right] \tag{A.51}$$

The idea is now to follow the gradient of $\mathcal{L}$ to find the minimum. An iterative update equation can be defined, according to (with l being the current iteration):

$$S_{est}^{(l+1)}(\vec{r}) = S_{est}^{(l)}(\vec{r}) - \lambda \cdot \frac{\partial(-\ln \mathcal{L})}{\partial S(\vec{r})} \tag{A.52}$$

with $\partial\mathcal{L}/\partial S$ being the gradient which is to be descending and λ the step size which can be adjusted. Note that the likelihood $\mathcal{L}$ is a function of the estimated sample S, because:

$$E(\vec{r}) = S(\vec{r}) \otimes h(\vec{r}) = \int_{-\infty}^{+\infty} d\vec{r}'\ S(\vec{r}') \cdot h(\vec{r} - \vec{r}') \tag{A.53}$$

The derivative yields:

$$\begin{aligned} \frac{\partial(-\ln \mathcal{L})}{\partial S(\vec{r})} &= \sum_{\vec{r}} \left[\frac{\partial E(\vec{r})}{\partial S(\vec{r})} - M(\vec{r}) \cdot \frac{1}{E(\vec{r})} \cdot \frac{\partial E(\vec{r})}{\partial S(\vec{r})}\right] = \\ &= \sum_{\vec{r}} \frac{\partial E(\vec{r})}{\partial S(\vec{r})} \left(1 - \frac{M(\vec{r})}{E(\vec{r})}\right) \end{aligned} \tag{A.54}$$

To solve this the derivative of the expectation value of our measurement with respect to the sample distribution is needed:

$$\frac{\partial E(\vec{r})}{\partial S(\vec{r}'')} = \frac{\partial \int_{-\infty}^{+\infty} d\vec{r}'\ S(\vec{r}') \cdot h(\vec{r} - \vec{r}')}{\partial S(\vec{r}'')} \tag{A.55}$$

Note that $\vec{r}''$ has been introduced to make a distinction between the spatial coordinate of the overall sum operation in eq. A.54 and the coordinate in the denominator of the derivation. There is only a single contribution to this integral, when $S(\vec{r}') = S(\vec{r}'')$:

$$\frac{\partial E(\vec{r})}{\partial S(\vec{r}'')} = \frac{\partial}{\partial S(\vec{r}'')} \Big(S(\vec{r}'') \cdot h(\vec{r} - \vec{r}'')\Big) = h(\vec{r} - \vec{r}'') \tag{A.56}$$

With this, the gradient of $-\ln \mathcal{L}$ becomes:

$$\frac{\partial(-\ln \mathcal{L})}{\partial S(\vec{r})} = \sum_{\vec{r}} \left[h(\vec{r} - \vec{r}'') \cdot \left(1 - \frac{M(\vec{r})}{E(\vec{r})}\right)\right] \tag{A.57}$$

Which is a convolution operation, abbreviated according to:

$$\frac{\partial(-\ln \mathcal{L})}{\partial S(\vec{r})} = \left(1 - \frac{M(\vec{r})}{E(\vec{r})}\right) \otimes h(-\vec{r}) \tag{A.58}$$

Using this result the previous update equation for the iterative reconstruction can be written as:

$$S_{est}^{(l+1)}(\vec{r}) = S_{est}^{(l)}(\vec{r}) + \lambda \cdot \left[\left(\frac{M(\vec{r})}{E(\vec{r})} - 1 \right) \otimes h(-\vec{r}) \right] \tag{A.59}$$

The gradient consists of two contributions:

$$S_{est}^{(l+1)}(\vec{r}) = S_{est}^{(l)}(\vec{r}) + \lambda \cdot \left(\frac{M(\vec{r})}{E(\vec{r})} \otimes h(-\vec{r}) \right) - \lambda \cdot \left(1 \otimes h(-\vec{r}) \right) \tag{A.60}$$

From this expression it becomes apparent that the typical additive iteration equation can be changed into a multiplicative one by setting:

$$\lambda = \frac{S_{est}^{(l)}(\vec{r})}{1 \otimes h(-\vec{r})} \tag{A.61}$$

Note that the denominator (at first λ) effectively yields a constant value (e.g. 1) as the convolution goes over an infinitely extended plane:

$$1 \otimes h(-\vec{r}) = \int_{-\infty}^{+\infty} d\vec{r}'\ h(\vec{r} - \vec{r}') \propto 1 \tag{A.62}$$

Altogether this will yield the final Richardson-Lucy iteration:

$$S_{est}^{(l+1)}(\vec{r}) = S_{est}^{(l)}(\vec{r}) \cdot \left[\frac{M(\vec{r})}{E(\vec{r})} \otimes h(-\vec{r}) \right] \tag{A.63}$$

Expressing the expectation value E as the (forward) convolution of $S_{est}^{(l)}$ with h, yields:

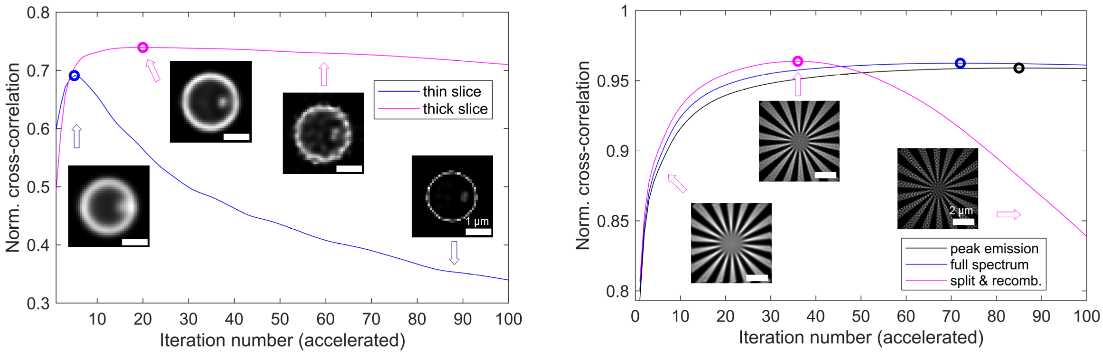
$$S_{est}^{(l+1)}(\vec{r}) = S_{est}^{(l)}(\vec{r}) \cdot \left[\frac{M(\vec{r})}{S_{est}^{(l)}(\vec{r}) \otimes h(\vec{r})} \otimes h(-\vec{r}) \right] \tag{A.64}$$

7 Optimize reconstruction using normalized cross-correlation

When using the iterative Richardson Lucy update scheme to perform a maximum likelihood reconstruction, the difficulty is to know when the algorithm has reached its optimum reconstruction result. With experimental data this often ends with the user reevaluating the results after a set of increasing iterations and visually determining for which "the best reconstruction" was achieved. Of course this all is very subjective and not really based on good scientific principles. When the problem of reconstructing an object estimate from **simulated** data is approached, there is a way to determine the optimal iteration number to stop the reconstruction. Simply the current estimate (est) needs to be compared to the known groundtruth object (S) from the simulation. This comparison is done by computing the *normalized cross-correlation* (NCC), which is given according to:

$$NCC_l = \frac{\frac{1}{N}\sum_{\vec{r}} \left[S(\vec{r}) - \bar{S}(\vec{r}) \right] \cdot \left[\mathrm{est}_l(\vec{r}) - \bar{\mathrm{est}}_l\,(\vec{r}) \right]}{\sigma_S \cdot \sigma_{\mathrm{est}_l}} \tag{A.65}$$

with $\bar{\cdot}$ denoting an average value, N the number of pixels in the reconstruction and $\sigma_{S,\mathrm{est}_l}$ indicating the standard deviation calculated for S and the l-th estimate est_l respectively. In case of the reconstruction results, which are shown in Fig. 2.6b bottom, the two NCC curves are shown in Fig. A.4a.



a b

Figure A.4: a *NCC* curves for the thin (blue) & thick (magenta) slice reconstruction shown in Fig. 2.6b. The optimum (accelerated) iteration for the thin slice method is reached earlier (5 accelerated iterations) but does not yield the same reconstruction quality as the thick slice method (20 accelerated iterations). Note that the gradual drop of both curves is much stronger for the thin slice method, meaning that the thick slice reconstruction might give good results even when the optimal iteration point is already passed. b *NCC* curves for deconvolving the simulated image data corresponding to the wavelength split in sec. 3.3.2. Blue and black indicating reconstructing the non-split data (single view), while magenta employs splitting & recombination (multiview).

Note that for the *thick slice* reconstruction the NCC-value has been calculated by comparing only the in-focus slice. It can be seen that both curves increase strongly after the first couple of iterations, in case of the thin slice method (blue) the maximum is reached after 5 accelerated iterations from which one the reconstruction quality quickly reduces (small inset at 88 accelerated iterations). The thick slice method (magenta) takes a bit longer to reach its maximum (20 accelerated iterations) which also yields a better reconstruction of S as $NCC_{thick} > NCC_{thin}$. The gradual drop of the NCC curve is not as strong as for the thin slice method, meaning that also beyond the optimum iteration the tick slice method might yield acceptable reconstruction results. Figure A.4b shows the *NCC* curves for the thin slice reconstructions corresponding to the simulated wavelength splitting in sec. 3.3.2. Here a spokes target was imaged using DAPI, meaning that the broad emission spectrum of DAPI requires $h(r)$ from eq. 3.23 to be the corresponding PSF model in the deconvolution. Interestingly, deconvolving the image (single view) with $h(r)$ (blue) or $h(r; \lambda_{max})$ (black) yields similar *NCC* curves, but deviate more strongly in the obtainable visibility, as shown in Fig. 3.9a. When performing a multiview deconvolution (magenta) the reconstruction is improved and already reached an optimum after a lower number of iterations. The point of maximum *NCC* values are: peak-emission-only at 85 iterations; full spectrum at 72 iterations and split & recombined at 36 iterations.

Note that in general the shape of the *NCC* curve cannot really be predicted before starting the reconstruction. It all depends on the available *SNR* and the object itself. This means that in principle every image reconstruction result needs to be accompanied by an *NCC* curve, making sure that the actual optimum iteration is reached. However, in reality this is often not the case as the quality of image reconstructions is often guided by visual perception rather than scientific numbers. In the case of only looking at some biological structures, without a comparison between different imaging modalities, this seems to be okay. But when comparing two methods it is important to work in the optimal regime of both approaches to yield a fair comparison.

Fluorescence microscopy:

1 Polarization effects in vectorial PSFs

When light is focused under high angles (large *NA*), the polarization of the light becomes important. In the BFP of the objective the electric field can be written as two orthogonal polarization states, x- and y-polarized:

$$\vec{E}_{BFP}(\vec{r}\,') = \begin{pmatrix} E^x_{BFP}(\vec{r}\,') \\ E^y_{BFP}(\vec{r}\,') \end{pmatrix} \tag{A.66}$$

with the superscript indicating x- or y-polarized light and $\vec{r}\,'$ being coordinates in the BFP. This is graphically presented in Fig. A.5 in green (x-pol.) and magenta (y-pol.). Focusing means to change the direction of propagation so that the light meets at a single point, the focus. Which requires a rotation of the propagation direction of the light rays.

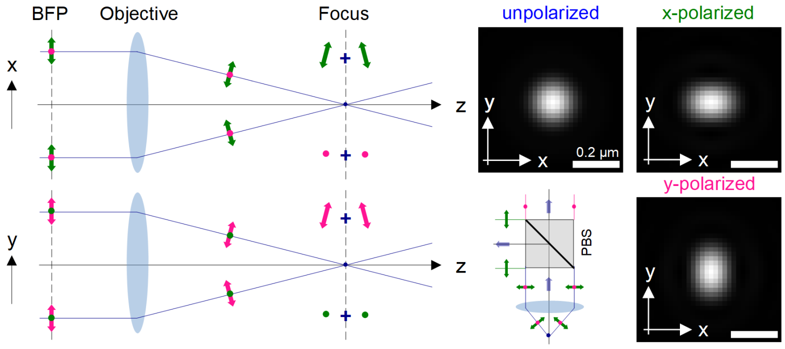


Figure A.5: Left: high *NA* focusing of light shown in the xz- and yz-plane. The electric field in the BFP is decomposed into two orthogonal linear polarization states (x & y polarized), depicted in green and magenta. When the light is focused these polarization states have to be tilted according to their BFP position, so that the propagation direction aligns with the focal point. This leads to imperfect interference for the x and y polarization, in different directions. Right: the quality of interference gives the width of the overall intensity distribution in the respective direction. Hence, filtering x and y polarized light in the BFP, using a polarization sensitive beamsplitter (PBS), will give an enhanced information transfer in one particular direction.

Note that this rotation depends on the position $\vec{r}\,'$ in the BFP and leads to the appearance of a z-polarized component in the focal region:

$$\vec{E}_{focus}(\vec{r}) = \begin{pmatrix} E^x_{focus}(\vec{r}) \\ E^y_{focus}(\vec{r}) \\ E^z_{focus}(\vec{r}) \end{pmatrix} \tag{A.67}$$

The corresponding intensity distribution is given as the *incoherent* superposition of the three individual polarization states, according to:

$$I(\vec{r}) = |\vec{E}_{focus}(\vec{r})|^2 = |E^x_{focus}(\vec{r})|^2 + |E^y_{focus}(\vec{r})|^2 + |E^z_{focus}(\vec{r})|^2 \tag{A.68}$$

When a lens focuses light, it does it in such a way the light can interfere perfectly to yield an intensity distribution which is also known as the PSF. In case the interference capability is somehow reduced (e.g. by an imperfect lens or additional aberrations), then the shape of the resulting PSF will change, e.g. the width will increase. Hence, the imaging

performance depends on how well light can interfere at the focus. The focusing of the two linearly polarized contributions $E_{BFP}^{x,y}$ leads to a change in interference capability. For example: when looking in the xz-plane the y-polarized light can interfere such as in an ideal case, as the rotation did not affect this polarization component. However, the x-polarized part experiences a slight reduction of interference. Note that this is effectively the same in the yz-plane, only that the polarization states have changed. Altogether this will yield an isotropic intensity distribution with certain width.
When only considering one of the two polarizations suddenly the interference capability varies with the spatial direction in the focal plane. E.g. the x-pol. contribution interferes worse along the x-direction than in the y-direction, because the focusing has distributed the electric field associated with E_{BFP}^{x} into two new components $E_{focus}^{x,z}$. Effectively decreasing the interference capability and resulting in a narrowed/widened PSF in the y/x direction. The same is true for the y-polarized light, only that x and y coordinate are being swapped. With the help of a polarization sensitive beamsplitter (PBS), the emitted light of a fluorphore (that emits isotropically) can be split into two orthogonal components, so that polarization splitting & recombination can be employed.

2 Weighted averaging in Fourier space

As introduced in section 3.2.2 two sub-images need to be recombined into a weighted averaged result M_{wa} (the generalization to N images is straightforward). In Fourier space this yields:

$$\tilde{M}_{wa}(\vec{k}) = \tilde{w}_1(\vec{k}) \cdot \tilde{M}_1(\vec{k}) + \tilde{w}_2(\vec{k}) \cdot \tilde{M}_2(\vec{k}) \tag{A.69}$$

Note that the weights are spatial frequency dependent and still might be complex-valued, which means that the variance of the averaged result $\tilde{M}_{wa}$ is given as:

$$\sigma^2_{\tilde{M}_{wa}}(\vec{k}) = |\tilde{w}_1(\vec{k})|^2 \cdot \sigma^2_{\tilde{M}_1} + |\tilde{w}_2(\vec{k})|^2 \cdot \sigma^2_{\tilde{M}_2} \tag{A.70}$$

The expectation value of $\tilde{M}_{wa}$ (termed $\tilde{E}_{wa}$), is given as:

$$\tilde{E}_{wa}(\vec{k}) = \tilde{w}_1(\vec{k}) \cdot \tilde{E}_1(\vec{k}) + \tilde{w}_2(\vec{k}) \cdot \tilde{E}_2(\vec{k}) \tag{A.71}$$

The goal is to maximize the SNR in Fourier space,:

$$SNR(\vec{k}) = \frac{|\tilde{E}_{wa}(\vec{k})|}{\sqrt{\sigma^2_{\tilde{M}_{wa}}(\vec{k})}} \tag{A.72}$$

An expression for $|\tilde{E}_{wa}|$ is given as:

$$\begin{aligned} |\tilde{E}_{wa}(\vec{k})|^2 &= |\tilde{w}_1(\vec{k})|^2 \cdot |\tilde{E}_1(\vec{k})|^2 + |\tilde{w}_2(\vec{k})|^2 \cdot |\tilde{E}_2(\vec{k})|^2 \\ &+ 2\,\mathrm{Re}\left\{\tilde{w}_1(\vec{k})\tilde{w}_2^*(\vec{k}) \cdot \tilde{E}_1(\vec{k})\tilde{E}_2^*(\vec{k})\right\} \end{aligned} \tag{A.73}$$

Weights and image information can be written in terms of real and imaginary parts:

$$\tilde{w}_1(\vec{k}) = \tilde{w}_{1,Re}(\vec{k}) + i \cdot \tilde{w}_{1,Im}(\vec{k}) \tag{A.74}$$

$$\tilde{w}_2^*(\vec{k}) = \tilde{w}_{2,Re}(\vec{k}) - i \cdot \tilde{w}_{2,Im}(\vec{k}) \tag{A.75}$$

$$\tilde{E}_1(\vec{k}) = \left(\tilde{h}_{1,Re}(\vec{k}) + i \cdot \tilde{h}_{1,Im}(\vec{k})\right) \cdot \tilde{S}(\vec{k}) \tag{A.76}$$

$$\tilde{E}_2^*(\vec{k}) = \left(\tilde{h}_{2,Re}(\vec{k}) - i \cdot \tilde{h}_{2,Im}(\vec{k})\right) \cdot \tilde{S}^*(\vec{k}) \tag{A.77}$$

Multiplying the weights yields:

$$
\begin{aligned}
\tilde{w}_1(\vec{k})\tilde{w}_2^*(\vec{k}) &= \underbrace{\Big(\tilde{w}_{1,Re}(\vec{k})\tilde{w}_{2,Re}(\vec{k}) + \tilde{w}_{1,Im}(\vec{k})\tilde{w}_{2,Im}(\vec{k})\Big)}_{\tilde{w}_{Re}(\vec{k})} \\
&+ i \cdot \underbrace{\Big(\tilde{w}_{1,Im}(\vec{k})\tilde{w}_{2,Re}(\vec{k}) - \tilde{w}_{1,Re}(\vec{k})\tilde{w}_{2,Im}(\vec{k})\Big)}_{\tilde{w}_{Im}(\vec{k})} \\
&= \tilde{w}_{Re}(\vec{k}) + i \cdot \tilde{w}_{Im}(\vec{k})
\end{aligned} \tag{A.78}
$$

Similarly, multiplying the image information gives:

$$
\tilde{E}_1(\vec{k})\tilde{E}_2^*(\vec{k}) = \Big[\tilde{h}_{Re}(\vec{k}) + i \cdot \tilde{h}_{Im}(\vec{k})\Big] \cdot |\tilde{S}(\vec{k})|^2 \tag{A.79}
$$

with the real and imaginary part of $\tilde{h}$ given as:

$$
\tilde{h}_{Re}(\vec{k}) = \tilde{h}_{1,Re}(\vec{k})\tilde{h}_{2,Re}(\vec{k}) + \tilde{h}_{1,Im}(\vec{k})\tilde{h}_{2,Im}(\vec{k}) \tag{A.80}
$$

$$
\tilde{h}_{Im}(\vec{k}) = \tilde{h}_{1,Im}(\vec{k})\tilde{h}_{2,Re}(\vec{k}) - \tilde{h}_{1,Re}(\vec{k})\tilde{h}_{2,Im}(\vec{k}) \tag{A.81}
$$

Computing the cross-term of $\tilde{M}_{wa}$ gives:

$$
\frac{1}{|\tilde{S}(\vec{k})|^2} \cdot \mathrm{Re}\Big\{\tilde{w}_1(\vec{k})\tilde{w}_2^*(\vec{k}) \cdot \tilde{E}_1(\vec{k})\tilde{E}_2^*(\vec{k})\Big\} = \tilde{w}_{Re}(\vec{k})\tilde{h}_{Re}(\vec{k}) - \tilde{w}_{Im}(\vec{k})\tilde{h}_{Im}(\vec{k}) \tag{A.82}
$$

Simplifying this expression leads to:

$$
\frac{\mathrm{Re}\Big\{\tilde{w}_1(\vec{k})\tilde{w}_2^*(\vec{k}) \cdot \tilde{E}_1(\vec{k})\tilde{E}_2^*(\vec{k})\Big\}}{|\tilde{S}(\vec{k})|^2} = \tilde{w}_{1,Re}(\vec{k}) \cdot \tilde{A}(\vec{k}) + \tilde{w}_{1,Im}(\vec{k}) \cdot \tilde{B}(\vec{k}) \tag{A.83}
$$

with the two new variables $\tilde{A}$ and $\tilde{B}$ given as:

$$
\tilde{A}(\vec{k}) = \tilde{w}_{2,Re}(\vec{k}) \cdot \tilde{h}_{Re}(\vec{k}) + \tilde{w}_{2,Im}(\vec{k}) \cdot \tilde{h}_{Im}(\vec{k}) \tag{A.84}
$$

$$
\tilde{B}(\vec{k}) = \tilde{w}_{2,Im}(\vec{k}) \cdot \tilde{h}_{Re}(\vec{k}) - \tilde{w}_{2,Re}(\vec{k}) \cdot \tilde{h}_{Im}(\vec{k}) \tag{A.85}
$$

A further simplification of eq. A.73 yields :

$$
\begin{aligned}
\frac{|\tilde{E}_{wa}(\vec{k})|^2}{|\tilde{S}(\vec{k})|^2} &= |\tilde{w}_1(\vec{k})|^2 \cdot |\tilde{h}_1(\vec{k})|^2 + |\tilde{w}_2(\vec{k})|^2 \cdot |\tilde{h}_2(\vec{k})|^2 \\
&+ 2 \cdot \Big[\tilde{w}_{1,Re}(\vec{k}) \cdot \tilde{A}(\vec{k}) + \tilde{w}_{1,Im}(\vec{k}) \cdot \tilde{B}(\vec{k})\Big]
\end{aligned} \tag{A.86}
$$

With this the expression for the *SNR* can be written as:

$$
\frac{SNR(\vec{k})}{|\tilde{S}(\vec{k})|} = \sqrt{\frac{|\tilde{w}_1(\vec{k})|^2 \cdot |\tilde{h}_1(\vec{k})|^2 + |\tilde{w}_2(\vec{k})|^2 \cdot |\tilde{h}_2(\vec{k})|^2 + 2 \cdot \Big[\tilde{w}_{1,Re}(\vec{k}) \cdot \tilde{A}(\vec{k}) + \tilde{W}_{1,Im}(\vec{k}) \cdot \tilde{B}(\vec{k})\Big]}{|\tilde{w}_1(\vec{k})|^2 \cdot \sigma^2_{\tilde{M}_1} + |\tilde{w}_2(\vec{k})|^2 \cdot \sigma^2_{\tilde{M}_2}}} \tag{A.87}
$$

Note that:

$$
|\tilde{w}_1(\vec{k})|^2 = \tilde{w}^2_{1,Re}(\vec{k}) + \tilde{w}^2_{1,Im}(\vec{k}) \tag{A.88}
$$

$$
|\tilde{w}_2(\vec{k})|^2 = \tilde{w}^2_{2,Re}(\vec{k}) + \tilde{w}^2_{2,Im}(\vec{k}) \tag{A.89}
$$

Yielding the following expression for the scaled *SNR*:

$$\frac{SNR(\vec{k})}{|\tilde{S}(\vec{k})|} = \sqrt{\frac{\tilde{Z}(\vec{k})}{\tilde{N}(\vec{k})}} \tag{A.90}$$

with:

$$\begin{aligned}
\tilde{Z}(\vec{k}) &= \left[\tilde{w}^2_{1,Re}(\vec{k}) + \tilde{w}^2_{1,Im}(\vec{k})\right] \cdot |\tilde{h}_1(\vec{k})|^2 + \left[\tilde{w}^2_{2,Re}(\vec{k}) + \tilde{w}^2_{2,Im}(\vec{k})\right] \cdot |\tilde{h}_2(\vec{k})|^2 \\
&+ 2 \cdot \left[\tilde{w}_{1,Re}(\vec{k}) \cdot \tilde{A}(\vec{k}) + \tilde{w}_{1,Im}(\vec{k}) \cdot \tilde{B}(\vec{k})\right] \qquad \text{(A.91)} \\
\tilde{N}(\vec{k}) &= \left[\tilde{w}^2_{1,Re}(\vec{k}) + \tilde{w}^2_{1,Im}(\vec{k})\right] \cdot \sigma^2_{\tilde{M}_1} + \left[\tilde{w}^2_{2,Re}(\vec{k}) + \tilde{w}^2_{2,Im}(\vec{k})\right] \cdot \sigma^2_{\tilde{M}_2} \qquad \text{(A.92)}
\end{aligned}$$

The goal is to find an expression for the weights $\tilde{w}_{1,2}$ which maximizes the *SNR*, given the set of images $\tilde{M}_{1,2}$ which have been measured. As the weights might be complex valued, the *Wirtinger* derivatives need to be used. For the real part of $\tilde{W}$ this gives:

$$\begin{aligned}
\frac{1}{|\tilde{S}(\vec{k})|} \cdot \frac{\partial SNR(\vec{k})}{\partial \tilde{w}_{1,Re}(\vec{k})} &= \frac{1/SNR(\vec{k})}{\tilde{N}^2(\vec{k})} \qquad \text{(A.93)} \\
&\cdot \left(\left[\tilde{w}_{1,Re}(\vec{k}) \cdot |\tilde{h}_1(\vec{k})|^2 + \tilde{A}(\vec{k})\right] \cdot \tilde{N}(\vec{k}) - \tilde{w}_{1,Re}(\vec{k}) \cdot \sigma^2_{\tilde{M}_1} \tilde{Z}(\vec{k})\right)
\end{aligned}$$

And analogously for the imaginary part:

$$\begin{aligned}
\frac{1}{|\tilde{S}(\vec{k})|} \cdot \frac{\partial SNR(\vec{k})}{\partial \tilde{w}_{1,Im}(\vec{k})} &= \frac{1/SNR(\vec{k})}{\tilde{N}^2(\vec{k})} \qquad \text{(A.94)} \\
&\cdot \left(\left[\tilde{w}_{1,Im}(\vec{k}) \cdot |\tilde{h}_1(\vec{k})|^2 + \tilde{B}(\vec{k})\right] \cdot \tilde{N}(\vec{k}) - \tilde{w}_{1,Im}(\vec{k}) \cdot \sigma^2_{\tilde{M}_1} \tilde{Z}(\vec{k})\right)
\end{aligned}$$

Now both equations are set equal to zero to find a criterion that maximizes the *SNR*:

$$\begin{aligned}
\left[\tilde{w}_{1,Re}(\vec{k}) \cdot |\tilde{h}_1(\vec{k})|^2 + \tilde{A}(\vec{k})\right] \cdot \tilde{N}(\vec{k}) &= \tilde{w}_{1,Re}(\vec{k}) \cdot \sigma^2_{\tilde{M}_1} \cdot \tilde{Z}(\vec{k}) \qquad \text{(A.95)} \\
\left[\tilde{w}_{1,Im}(\vec{k}) \cdot |\tilde{h}_1(\vec{k})|^2 + \tilde{B}(\vec{k})\right] \cdot \tilde{N}(\vec{k}) &= \tilde{w}_{1,Im}(\vec{k}) \cdot \sigma^2_{\tilde{M}_1} \cdot \tilde{Z}(\vec{k}) \qquad \text{(A.96)}
\end{aligned}$$

It is not trivial to see a possible solution immediately. But note that the equations show a certain symmetry, which becomes more obvious if the OTFs and weights are restricted to being **real-valued** only. With this in mind, only equation A.95 needs to be fulfilled, where $\tilde{A}$ and $\tilde{Z}$ take a much simpler form:

$$\begin{aligned}
\tilde{A}(\vec{k}) &= \tilde{w}_{2,Re}(\vec{k}) \cdot \tilde{h}_{1,Re}(\vec{k})\tilde{h}_{2,Re}(\vec{k}) \qquad \text{(A.97)} \\
\tilde{Z}(\vec{k}) &= \tilde{w}^2_{1,Re}(\vec{k}) \cdot \tilde{h}^2_{1,Re}(\vec{k}) + \tilde{w}^2_{2,Re}(\vec{k}) \cdot \tilde{h}^2_{2,Re}(\vec{k}) \\
&+ 2 \cdot \tilde{w}_{1,Re}(\vec{k})\tilde{h}_{1,Re}(\vec{k}) \cdot \tilde{w}_{2,Re}(\vec{k})\tilde{h}_{2,Re}(\vec{k}) \qquad \text{(A.98)}
\end{aligned}$$

Note that $\tilde{Z}$ now represents a binomial equation which can be written as:

$$\tilde{Z}(\vec{k}) = \left[\tilde{w}_{1,Re}(\vec{k})\tilde{h}_{1,Re}(\vec{k}) + \tilde{w}_{2,Re}(\vec{k})\tilde{h}_{2,Re}(\vec{k})\right]^2 \tag{A.99}$$

The left side of equation A.95 is now given as:

$$\begin{aligned} & \tilde{h}_{1,Re}(\vec{k}) \cdot \left[\tilde{w}_{1,Re}(\vec{k}) \cdot \tilde{h}_{1,Re}(\vec{k}) + \tilde{w}_{2,Re}(\vec{k}) \cdot \tilde{h}_{2,Re}(\vec{k})\right] \cdot \tilde{N}(\vec{k}) \\ = \quad & \tilde{w}_{1,Re}(\vec{k}) \cdot \sigma^2_{\tilde{M}_1} \cdot \left[\tilde{w}_{1,Re}(\vec{k})\tilde{h}_{1,Re}(\vec{k}) + \tilde{w}_{2,Re}(\vec{k})\tilde{h}_{2,Re}(\vec{k})\right]^2 \end{aligned} \tag{A.100}$$

with an simplified expression for N according to:

$$\tilde{N}(\vec{k}) = \tilde{w}_{1,Re}(\vec{k})^2 \cdot \sigma^2_{\tilde{M}_1} + \tilde{w}_{2,Re}(\vec{k})^2 \cdot \sigma^2_{\tilde{M}_2} \tag{A.101}$$

After canceling terms, the following criterion must be fulfilled:

$$\begin{aligned} & \tilde{h}_{1,Re}(\vec{k}) \cdot \left[\tilde{w}_{1,Re}(\vec{k})^2 \cdot \sigma^2_{\tilde{M}_1} + \tilde{w}_{2,Re}(\vec{k})^2 \cdot \sigma^2_{\tilde{M}_2}\right] \\ = \quad & \tilde{w}_{1,Re}(\vec{k}) \cdot \sigma^2_{\tilde{M}_1} \cdot \left[\tilde{w}_{1,Re}(\vec{k})\tilde{h}_{1,Re}(\vec{k}) + \tilde{w}_{2,Re}(\vec{k})\tilde{h}_{2,Re}(\vec{k})\right] \end{aligned} \tag{A.102}$$

The trick is now to make the terms in front of both brackets equal, by setting:

$$\tilde{w}_{1,Re}(\vec{k}) = \frac{\tilde{h}_{1,Re}(\vec{k})}{\sigma^2_{\tilde{M}_1}} \tag{A.103}$$

Interestingly, this choice of the weights fulfills the above criterion:

$$\begin{aligned} & \tilde{h}_{1,Re}(\vec{k}) \cdot \left[\frac{\tilde{h}^2_{1,Re}(\vec{k})}{\sigma^2_{\tilde{M}_1}} + \frac{\tilde{h}^2_{2,Re}(\vec{k})}{\sigma^2_{\tilde{M}_2}}\right] \\ = \quad & \tilde{h}_{1,Re}(\vec{k}) \cdot \left[\frac{\tilde{h}^2_{1,Re}(\vec{k})}{\sigma^2_{\tilde{M}_1}} + \frac{\tilde{h}^2_{2,Re}(\vec{k})}{\sigma^2_{\tilde{M}_2}}\right] \qquad \square \end{aligned} \tag{A.104}$$

Hence, the weights needed for the *weighted averaging* in Fourier space approach are:

$$\tilde{w}_{1,2}(\vec{k}) = \frac{\mathrm{Re}\{\tilde{h}_{1,2}(\vec{k})\}}{\sigma^2_{\tilde{M}_{1,2}}} \tag{A.105}$$

Meaning that each sub-OTF (real-valued only) needs to be rescaled by the amount of noise variance present in Fourier space. From experience in simulations, it has been seen that replacing the Re$\{.\}$-operation by the complex-conjugate $.^*$, does not reduce the achievable *SNR*. Note that due to complexity the latter version has not been directly derived from the criteria eq. A.96.

Finally, the following weights are used to optimize the *SNR* in weighted averaging:

$$\tilde{w}_{1,2}(\vec{k}) = \frac{\tilde{h}^*_{1,2}(\vec{k})}{\sigma^2_{\tilde{M}_{1,2}}} \tag{A.106}$$

3 Log-normal distribution for vanishing skewness

The exponent of the emission spectrum model is given in eq. 3.22 as:

$$-\ln 2\left[\frac{\ln\left(1+2\frac{s}{w}(\lambda-\lambda_{max})\right)}{s}\right]^2 \tag{A.107}$$

For $s \rightarrow 0$ the term inside the bracket becomes:

$$\lim_{s\rightarrow 0}\frac{\ln\left(1+2\frac{s}{w}(\lambda-\lambda_{max})\right)}{s}=\frac{0}{0} \tag{A.108}$$

Hence the rule of *l'Hospital* [31] needs to be applied, which yields:

$$\lim_{s\rightarrow 0}\frac{\frac{2}{w}(\lambda-\lambda_{max})}{1+2\frac{s}{w}(\lambda-\lambda_{max})}=\frac{2}{w}(\lambda-\lambda_{max}) \tag{A.109}$$

With this the exponent of $\varepsilon(\lambda)$ can be written as:

$$-\ln 2\left[\frac{2}{w}(\lambda-\lambda_{max})\right]^2 \tag{A.110}$$

Relating the *FWHM* w to the standard deviation σ_ε of a Gaussian function via:

$$w=2\sqrt{2\ln 2}\cdot\sigma_\varepsilon \tag{A.111}$$

Which gives the typical Gaussian exponent:

$$-\frac{(\lambda-\lambda_{max})^2}{2\sigma_\varepsilon^2} \tag{A.112}$$

Showing that the emission spectrum $\varepsilon(\lambda)$ for $s \rightarrow 0$ is given as a Gaussian function, hence becomes perfectly symmetric with respect to λ_{max}.

4 Solution to the integrals representing A and λ_{CoM}

The integral defining A, the area under the $\varepsilon(\lambda)$-curve, is given as:

$$A=\int_{\lambda^-}^{\lambda^+} d\lambda\;\exp\left(-a\left[\ln\left(1+b(\lambda-\lambda_{max})\right)\right]^2\right) \tag{A.113}$$

First the following substitution is introduced:

$$\begin{aligned} u &= 1+b(\lambda-\lambda_{max}) \\ du/d\lambda &= b \end{aligned} \tag{A.114}$$

Yielding the following integral:

$$A=\frac{1}{b}\int_{1+b(\lambda^- -\lambda_{max})}^{1+b(\lambda^+ -\lambda_{max})} du\;\exp\left(-a\left[\ln u\right]^2\right) \tag{A.115}$$

Another substitution is used to further simplify the expression:

$$\begin{aligned} v &= \ln\left[u\right] \\ dv/du &= 1/u \end{aligned} \tag{A.116}$$

Resulting in an integral which can be solved using the Gaussian error function erf [31]:

$$\begin{aligned} A &= \frac{1}{b}\int_{\ln[1+b(\lambda^{-}-\lambda_{max})]}^{\ln[1+b(\lambda^{+}-\lambda_{max})]} dv \;\; \exp\left(-av^2+v\right) = && \text{(A.117)} \\ &= \frac{1}{b}\frac{\sqrt{\pi}\,\sqrt[4a]{e}}{2\sqrt{a}}\left[erf\left(\frac{2av-1}{2\sqrt{a}}\right)\right]_{v=\ln[1+b(\lambda^{-}-\lambda_{max})]}^{v=\ln[1+b(\lambda^{+}-\lambda_{max})]} && \text{(A.118)} \end{aligned}$$

Note that A does not depend on r or λ.
A similar approach can be followed for λ_{CoM}, the "center-of-mass" wavelength. Using the same substitutions (u) the following expression can be found:

$$\begin{aligned} \lambda_{CoM} &= \int_{1+b(\lambda^{-}-\lambda_{max})}^{1+b(\lambda^{+}-\lambda_{max})} du \;\frac{1}{b}\left(\lambda_{max}+\frac{u-1}{b}\right)\exp\left(-a\left[\ln u\right]^2\right) && \text{(A.119)} \\ &= \int_{1+b(\lambda^{-}-\lambda_{max})}^{1+b(\lambda^{+}-\lambda_{max})} du \;\left(\frac{\lambda_{max}}{b}+\frac{u-1}{b^2}\right)\exp\left(-a\left[\ln u\right]^2\right) && \text{(A.120)} \end{aligned}$$

Note that this integral can be split into two parts:

$$\lambda_{CoM} = A\cdot\lambda_{max}+\frac{1}{b}\left(I-A\right) \tag{A.121}$$

with I another integral to solve.

$$I = \frac{1}{b}\int_{1+b(\lambda^{-}-\lambda_{max})}^{1+b(\lambda^{+}-\lambda_{max})} du \;\, u\cdot\exp\left(-a\left[\ln u\right]^2\right) \tag{A.122}$$

This can be achieved by using the same substitutions as previously, which yields:

$$\begin{aligned} I &= \frac{1}{b}\int_{\ln[1+b(\lambda^{-}-\lambda_{max})]}^{\ln[1+b(\lambda^{+}-\lambda_{max})]} dv \;\; \exp\left(-av^2+2v\right) dv = \\ &= \frac{1}{b}\frac{\sqrt{\pi}\sqrt[a]{e}}{2\sqrt{a}}\left[erf\left(\frac{av-1}{\sqrt{a}}\right)\right]_{v=\ln[1+b(\lambda^{-}-\lambda_{max})]}^{v=\ln[1+b(\lambda^{+}-\lambda_{max})]} && \text{(A.123)} \end{aligned}$$

5 Estimating the spatial broadening $\mathcal{B}$

In the following the spatial broadening effect, in terms of an increased *FWHM*, is theoretically investigated. To do this lets recall the two functions $f(r)$ and $g(r)$ that make up the broadened PSF:

$$\begin{aligned} f(r) &= h(r;\lambda_{max}) && \text{(A.124)} \\ g(r) &= 2CD\cdot\frac{r^2}{\lambda_{max}^3}\exp\left(-C\cdot\frac{r^2}{\lambda_{max}^2}\right) && \text{(A.125)} \end{aligned}$$

with $D = 1/b\cdot(I/A-1)$ and I being an integral solved on page 118. With this, the broadened PSF $h(r)$ can be expressed as:

$$h(r) \approx A\left[f(r)+g(r)\right] \tag{A.126}$$

An expression for the new *FWHM* value Δr_B is given according to:

$$\Delta r_B = 2\cdot r_{1/2;B} \tag{A.127}$$

with $r_{1/2;B}$ being the radial coordinate where $h(r)$ has reached $1/2 \cdot \max\{h(r)\}$.
Figure 3.7a shows how $r_{1/2;B}$ is linearly interpolated. To do this, two points (marked with green arrows) are defined. One of them corresponds to the *FWHM* position of the peak only PSF $h(r; \lambda_{max})$, the other one corresponds to the maximum of $g(r)$.

$$\left(r_{1/2}\, ,\, f(r_{1/2}) \quad = \quad \tfrac{1}{2} \right) \tag{A.128}$$

$$\left(r_{max}\, ,\, f(r_{max}) \quad + \quad g(r_{max}) \right) \tag{A.129}$$

The latter can be found by calculating the first derivative of $g(r)$:

$$\frac{\partial g(r)}{\partial r} = 2CD \cdot \frac{2r}{\lambda_{max}^5} \exp\left(-C \frac{r^2}{\lambda_{max}^2} \right) \left(\lambda_{max}^2 - Cr^2 \right) \tag{A.130}$$

Setting this expression to zero and solving for r gives:

$$r_{max} = \pm \frac{\lambda_{max}}{\sqrt{C}} \tag{A.131}$$

Knowing this the values of $f(r_{max})$ and $g(r_{max})$ can be found.

$$g_{max} \quad = \quad g(r_{max}) = \frac{2D}{\lambda_{max} \cdot e} \tag{A.132}$$

$$f_{max} \quad = \quad f(r_{max}) = 1/e \tag{A.133}$$

The linear relationship (red line in Fig. 3.7a) can be expressed in the form of:

$$y = m \cdot r + t \tag{A.134}$$

with m being the slope and t the vertical offset.
According to basic math, the slope is given as:

$$m = \frac{f_{max} + g_{max} - f_{1/2} - g_{1/2}}{r_{max} - r_{1/2}} \tag{A.135}$$

For this, the values of $r_{1/2}$ and $g_{1/2}$ need to be known:

$$r_{1/2} \quad = \quad \sqrt{\ln 2} \cdot \frac{\lambda_{max}}{\sqrt{C}} = \sqrt{\ln 2} \cdot r_{max} \tag{A.136}$$

$$g_{1/2} \quad = \quad g(r_{1/2}) = \ln 2 \cdot \frac{D}{\lambda_{max}} \tag{A.137}$$

With this the slope m can be computed, yielding the following expression:

$$m = \sqrt{C} \cdot \frac{1/e \cdot (1 + 2 \cdot D/\lambda_{max}) - 1/2 - \ln 2 \cdot D/\lambda_{max}}{\left(1 - \sqrt{\ln 2}\right) \lambda_{max}} \tag{A.138}$$

The offset t can now be found by using the two points and the linear relationship:

$$t = 1/2 + \ln 2 \cdot D/\lambda_{max} - m \cdot r_{1/2} \tag{A.139}$$

Finally, to solve for $r_{1/2;B}$, $r = (1/2 - t)/m$ needs to be calculated:

$$r_{1/2;B} = \frac{\sqrt{\ln 2} \cdot \lambda_{max}/\sqrt{C} \cdot m - \ln 2 \cdot D/\lambda_{max}}{m} \tag{A.140}$$

This can be converted into a *FWHM* value by multiplying by two:

$$\Delta r_B = 2\sqrt{\ln 2} \cdot \frac{\lambda_{max}/\sqrt{C} \cdot m - \sqrt{\ln 2} \cdot D/\lambda_{max}}{m} \tag{A.141}$$

Now lets introduce the broadening $\mathcal{B}$, the difference between both *FWHM* values:

$$\begin{aligned} \mathcal{B} &= \Delta r_B - \Delta r = \\ &= 2\sqrt{\ln 2}\frac{\lambda_{max}}{\sqrt{C}} \left[\frac{m - \sqrt{\ln 2} \cdot \sqrt{C} D/\lambda_{max}^2}{m} - 1 \right] \end{aligned} \tag{A.142}$$

6 The scanning confocal point spread function

In a fluorescence confocal microscope the sample is illuminated by a focused laser spot and recorded using a point detector through a small pinhole. For a particular scan position $\vec{s}$ the following intensity response is found [7]:

$$M(\vec{r};\vec{s},\vec{r}_{\mathrm{p}}) = \int_{-\infty}^{+\infty} d\vec{s}\ S(\vec{r} - \vec{s}) \cdot h_{Illu.}(\vec{s}) \cdot h_{Det.}(\vec{r}_{\mathrm{p}} - \vec{s}) \tag{A.143}$$

with $\vec{r}_{\mathrm{p}}$ being the coordinates of the point detector. Note that the sample has been shifted by $\vec{s}$ while the illumination is given as the point-spread function $h_{Illu.}$, the smallest attainable laser focus, centered around the current scan position. The acquired intensity signal is given as an integration over the detector coordinates:

$$M(\vec{r}) = \int_{-\infty}^{+\infty} d\vec{r}_{\mathrm{p}}\ \mathrm{p}(\vec{r}_{\mathrm{p}}) \cdot M(\vec{r};\vec{s},\vec{r}_{\mathrm{p}}) \tag{A.144}$$

with p being the pinhole. Writing the convolution in integral form yields:

$$M(\vec{r}) = \int_{-\infty}^{+\infty} d\vec{r}_{\mathrm{p}}\ \mathrm{p}(\vec{r}_{\mathrm{p}}) \int_{-\infty}^{+\infty} d\vec{s}\ S(\vec{r} - \vec{s}) \cdot h_{Illu.}(\vec{s}) \cdot h_{Det.}(\vec{r}_{\mathrm{p}} - \vec{s}) \tag{A.145}$$

Exchanging both integrals and rearranging the quantities accordingly gives:

$$M(\vec{r}) = \int_{-\infty}^{+\infty} d\vec{s}\ S(\vec{r} - \vec{s}) \cdot h_{Illu.}(\vec{s}) \int_{-\infty}^{+\infty} d\vec{r}_{\mathrm{p}}\ \mathrm{p}(\vec{r}_{\mathrm{p}})\ h_{Det.}(\vec{s} - \vec{r}_{\mathrm{p}}) \tag{A.146}$$

In the last equation a symmetric detection PSF was assumed so that $h_{Det.}(\vec{r}_{\mathrm{p}} - \vec{s}) = h_{Det.}(\vec{s} - \vec{r}_{\mathrm{p}})$. Note that the integral with respect to $d\vec{r}_{\mathrm{p}}$ is given as a convolution in terms of $\vec{s}$, hence:

$$M(\vec{r}) = \int_{-\infty}^{+\infty} d\vec{s}\ S(\vec{r} - \vec{s}) \cdot h_{Illu.}(\vec{s}) \cdot \Big(\mathrm{p}(\vec{s}) \otimes h_{Det.}(\vec{s})\Big) \tag{A.147}$$

With this the image is given as a convolution with an effective PSF:

$$M(\vec{r}) = S(\vec{r}) \otimes \Big[h_{Illu.}(\vec{r}) \cdot \Big(\mathrm{p}(\vec{r}) \otimes h_{Det.}(\vec{r})\Big)\Big] = S(\vec{r}) \otimes h_{Conf.}(\vec{r}) \tag{A.148}$$

7 Image results of wavelength splitting & recombination

In the experimental realization of wavelength splitting, a fluorescent calibration target from Argolight [78] (*Argo-SIM*, Argolight SA, France) has been imaged, using the LSM 880 (see Fig. 3.10b). A λ-stack was recorded with 32 spectral channels, which was repeated 10 times to effectively obtain multiple realizations of the noisy measurement. This is required to compute the mean (= signal) and standard deviation (= noise) values per pixel, so that a possible *SNR* enhancement can be evaluated. The reference data, termed splitting & *sum*, was calculated by summing over all wavelength channels. This corresponds to the non-split case, as all photons are simply added as stated in eq. 3.23. The non-split data was not acquired additionally with the LSM 880, as the system uses a different detector when the spectral detection (QUASAR module) is not used. A fair comparison between non-split and split data requires to use the same detector, as any difference in detection performance would strongly alter the ability to compare both datasets. The non-split reference data is shown in Fig. A.6a blue after averaging the 10 time frames: the top half depicts the averaged raw data, the bottom half the average of the deconvolved time frames, processed using the RL method with 30 (accelerated) iterations (see p. 124) and a theoretically derived PSF $h(r)$ (broadened according to the spectrum shown in Fig. 3.11b). A region in the FoV is marked with a yellow box and depicted enlarged at the bottom, comparing the average of raw and deconvolved data.

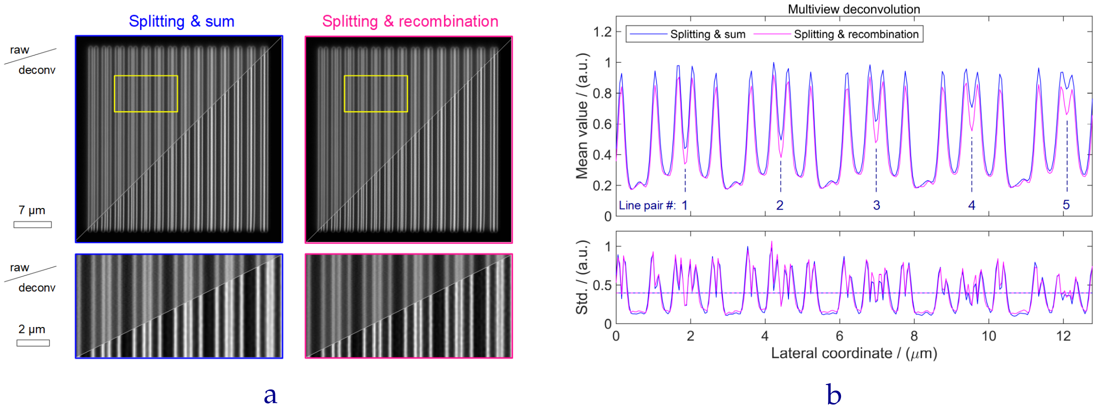


Figure A.6: a Comparison between summing or recombining (weighted average, multiview deconvolution) the experimentally acquired λ-stack. The data was recorded as a time series with 10 frames, the depicted images represents the average of those 10 images. The top half either shows the raw data (blue) or the weighted averaged result (magenta). On the bottom the deconvolved results of both are depicted. A zoomed region of the FoV is shown on the bottom, again comparing non- with deconvolved data. b Horizontal line profile, through the center of the yellow frame in a, showing the mean and standard deviation for both imaging strategies after deconvolution. Note that both signals are noise-normalized to the same average standard deviation. Meaning that an enhanced modulation of the line profile indicates an improvement in *SNR*. Five different line pairs with decreasing gaps are marked and evaluated in more detail in Tab. A.1.

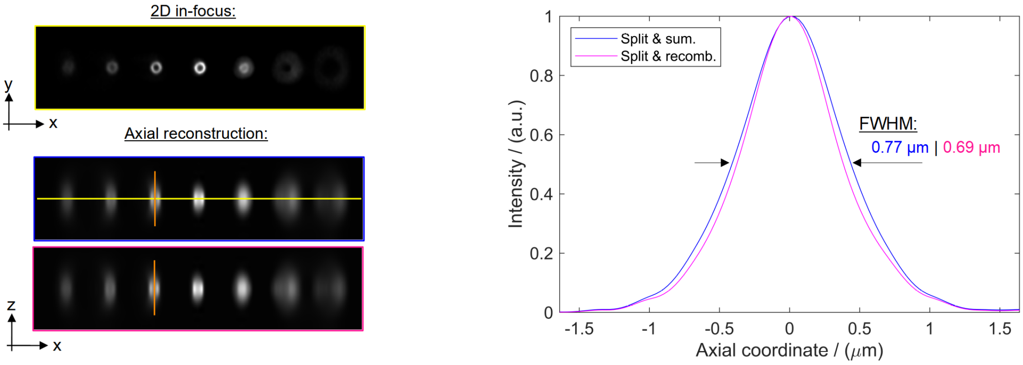
To apply the wavelength splitting idea, the λ-stack in each time frame was sequentially processed via weighted averaging in Fourier space (with noise normalization). The same dataset was also processed using the multiview deconvolution with 39 (accelerated) iterations (see Fig. A.8a) and the corresponding single wavelength PSFs $h(r;\lambda)$, scaled according to the spectrum shown in Fig. 3.11b. The average of the 10 reconstructed time frames is depicted in the bottom half, for the full FoV and the enlarged view (yellow frame). The improvement in terms of the *SNR* is analyzed in more detail using a horizontal line profile through the center of the yellow frame in Fig. A.6a. The modulation of the deconvolved line patterns are shown in Fig. A.6b top. To make sure that the datasets (non-split and split) are properly noise normalized, the standard deviation over the 10 time frames was computed and depicted in at the bottom of Fig. A.6b. The line indicates the average value of the pixel-wise computed standard deviation and was used as a reference to normalize the split data accordingly. When this is done, both datasets show the same amount of noise and the modulation of the mean value (top) is a direct indicator for the *SNR* performance of the respective system. Five different line pairs are marked in Fig. A.6b, the respective gap for each line pair is given in Table A.1. Which also shows the visibility and the spatial improvement factor $IF_{\mathcal{V}}$ (eq. 3.14).

	Line pair #	1	2	3	4	5
	Spacing / (nm)	330	300	270	240	210
	$k_{xy}/k_{xy,max}$	0.54	0.60	0.66	0.74	0.85
WA	Summing	0.1403	0.1131	0.0704	0.0426	0.0124
	Recombination	0.1443	0.1169	0.0727	0.0437	0.0131
	Improvement / (%)	*2.82*	*3.30*	*3.31*	*2.76*	*5.66*
MV	Summing	0.3801	0.3291	0.2222	0.1454	0.0576
	Recombination	0.3907	0.3567	0.2713	0.1938	0.1032
	Improvement / (%)	*6.46*	*8.39*	*22.11*	*33.26*	*79.23*

Table A.1: The visibility $\mathcal{V}$ for the five different line pairs shown in Fig. A.6b after averaging 10 time frames. Reference data is obtained by summing the λ-stack, while the recombination was done either by weighted averaging (WA) or multiview deconvolution (MV). The *improvement* is given as $IF_{\mathcal{V}}$ (defined in eq. 3.14), the relative difference between summing and the respective recombination method. The ratio $k_{xy}/k_{xy,max}$ is calculated for $NA = 1.4$ and $\lambda_{max} = 500$ nm.

In the case of weighted averaging (WA) the improvement factor stays approximately constant at a value of 3%. Only the line pair with the smallest separation shows a larger $IF_{\mathcal{V}}$ value of $\approx 6\%$. A much stronger enhancement can be achieved when the multiview (MV) deconvolution is used. The maximum improvement factor again is given at the smallest line spacing, this time reaching $\approx 80\%$. In contrast to the weighted averaged result, the obtained improvement increases more strongly from coarse to finer structures. Indicating that the attainable *SNR* improvement due to wavelength splitting & recombination mostly shows its benefits when small structures are being reconstructed.

The axial performance of the wavelength splitting approach, is tested using pattern *I* of the Argolight calibration target ("stairs" in Fig. 3.11a). This time only a single λ stack was acquired and the reference case (non-split) was obtained by summing over all wavelengths. A thick slice deconvolution (see sec. 2.4.2) was performed with both datasets, either as a single view (non-split) or multiview (split data) version, using the 3D-PSFs computed as already described previously for the deconvolution of the resolution target (pattern *E*). Note that no evaluation of optimal iteration using an *NCC* curve was done here, meaning that both datasets where deconvolved using 10 (accelerated) iterations. The primary goal of the presented analysis is to give a first impression on 3D deconvolution aspects of wavelength split data, while more work on optimization still needs to be done. Figure A.7a right shows an in-focus slice (yellow) of the reconstruction of the non-split data, quantifying the axial resolution as only two stair steps are nicely visible. The three dimensional reconstructions for the non-split (blue) and split (magenta) data are presented to the right (with 11 slices spaced by 0.5 μm, upscaled by factor of 12 for visualization). Note how the reconstruction indicates the axial position of the different cylinders with stronger blurring. However, it was not possible to infer the crossing stair type of geometry from the data. In Fig. A.7b a line profile (orange) through the stair at $z = -0.125\ \mu$m is shown. The improved axial resolution is indicated as a reduction of *FWHM* by ≈ 10 %.



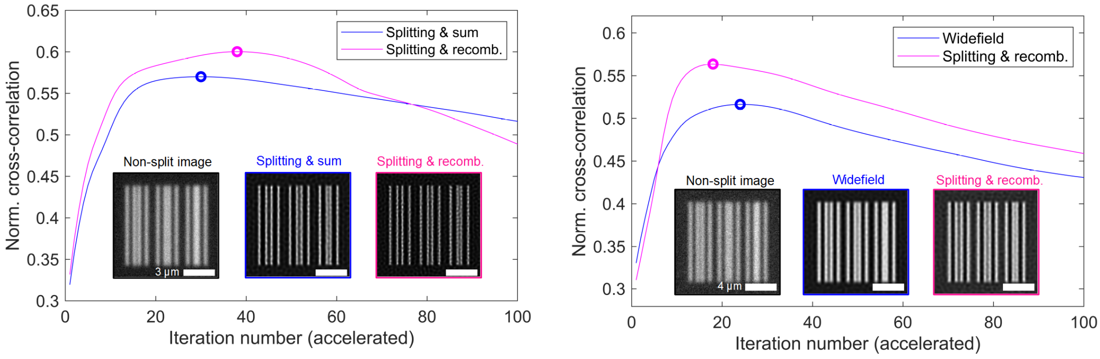
a b

Figure A.7: a Schematics and thick-slice deconvolution results (see sec. 2.4.2) of imaging pattern *I* of the Argolight sample, used to evaluate the 3D performance of the imaging system. Left: lateral and axial scheme of the crossing stairs sample with a step size of 0.125 μm; right: axial views of the thick-slice reconstruction using 10 (accelerated) iterations for the split (magenta) and non-split (blue) data (see text for more detail). b Line profile through the 3D reconstructed stair step at $z = -0.125\ \mu$m (orange line), indicating a slightly narrower axial response of $\approx$ 10%.

As mentioned in section 3.3.4, the problem with the acquired data lies in the fact that the detection pinhole was opened. Therefore $h_{Conf.}$ in equation 3.43 is dominated by p and only weakly depends on the emission wavelength λ. Meaning that the necessary wavelength-scaling of the *FWHM* of $h_{Conf.}$ is reduced. Hence, the experimentally found improvement in terms of visibility $\mathcal{V}$ and axial *FWHM* cannot be fully attributed to realized wavelength splitting.

8 Estimating the optimum iteration number using *NCC*

When the multiview deconvolution method is used for recombining the split data, it is important to compare the reference (non-split) and splitting & recombined data in a *fair* way. This can only be achieved when the image reconstruction for each method is optimized, in the sense that both have reached their optimal reconstruction result. Which most probably will occur at different number of iterations, for each dataset (non-split & split) individually. Finding the optimum number of iterations with experimental data is difficult and often leads to unfair comparisons when unintentionally one method is iterated closer to the optimum iteration number than the other. Here we try to circumvent this by simulating the imaging of a line pattern, similar to the resolution pattern of the Argolight sample used for the wavelength (sec. 3.3) and pupil (sec. 3.4) splitting. Poisson noise was applied and matched to the observed photon numbers in the experimental data. To do this the average and standard deviation along the vertical direction have been computed for: 1) a line structure; 2) a region in between lines where no signal was measured. In the simulation the number of photons in the brightest pixel of the expected image and some global background have been varied by hand, such that the same ratio of average to standard deviation values are obtained. Each dataset was then deconvolved using the RL algorithm, while for each (accelerated) iteration the *NCC* value was computed. Fig. A.8 shows the curves for wavelength (A.8a) and pupil (A.8b) splitting, the reference case (non-split) in blue and the split & recombined result in magenta.



a b

Figure A.8: *NCC* curves for wavelength a and pupil splitting b. The underlying image data was simulated and matched to fit the experimentally observations as close as possible. These curves are used to estimate the point of optimum iteration (marked with circle) in the RL deconvolution scheme, for the reference (blue) and split & recombined (magenta) data. Note that the order of which reconstruction reaches its maximum first is not fixed. In both cases (wavelength & pupil splitting) the split & recombined reconstruction using multiview deconvolution yields a better result in terms of *NCC* value.

In both splitting cases it can be seen that the optimal point of reconstruction is reached at different iterations. Interestingly, the order of which dataset (non-split or split) reaches its maximum earlier is not the same, an effect which can also be seen in Fig. A.4. In both splitting scenarios (wavelength & pupil) the split & recombined approach results in a larger maximum *NCC* value, meaning that the split data enables better image reconstructions, compared to the non-split counterpart. For the *wavelength splitting* the maximum *NCC* values are: splitting & sum at 30, splitting & recombination at 39 iterations. Note how the magenta curve drops rapidly for large iteration numbers, similar to what is depicted in Fig. A.4b. In the *pupil splitting* approach the maximum *NCC* values are reached at: widefield (non-split) at 24, splitting & recombination at 18 iterations. This time the *NCC* curves look similar, only with an *NCC* offset.

9 Calculating the OTF via autocorrelation of the pupil

The optical transfer function of an imaging system can be computed from the pupil function $\mathcal{P}$ using an *autocorrelation* operation:

$$\tilde{h}(\vec{k}) \;=\; \mathcal{A}\left\{\mathcal{P}(\vec{k})\right\} = \int_{-\infty}^{+\infty} d\vec{k}'\; \mathcal{P}(\vec{k}')\cdot\mathcal{P}^{*}(\vec{k}'-\vec{k}) \tag{A.149}$$

Which is graphically presented in Fig. A.9a for a disk (top) and annular pupil (bottom).

In ideal imaging (constant phase in the pupil), the autocorrelation can be represented as the following geometric operations:

1. Take two copies of the pupil (e.g. ring shape) and overlap their center.
2. Shift them apart by $k_{xy}/k_{xy,max}$ where $k_{xy,max}$ is the diameter of the full pupil.
3. The area of overlap (dark green in Fig. A.9a) with respect to the area of the full pupil ($= R_{max}^2 \cdot \pi$) represents the OTF value at the relative frequency $k_{xy}/k_{xy,max}$.

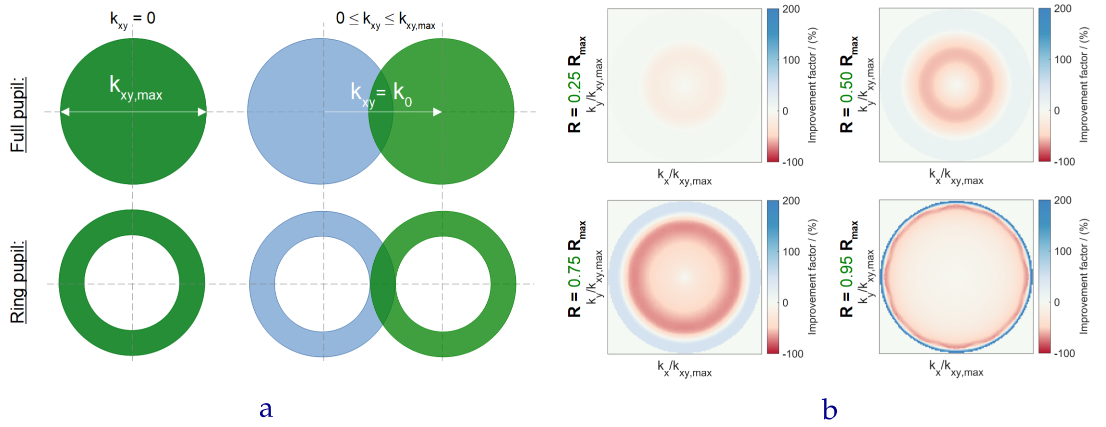


Figure A.9: a Graphical representation of obtaining the transfer strength (= OTF) for two different pupil types: conventional (full pupil, top) and annular (ring pupil, bottom). Each pupil has a diameter which corresponds to the cut-off limit of the incoherent imaging system $k_{xy,max}$. The transfer strength for a particular k_{xy} can be obtained by separating the two copies such that the distance between them corresponds to k_{xy}. The ratio of the overlap area divided by the area of the full pupil, corresponds to the OTF value at this particular spatial frequency k_{xy}. Note that for $|\vec{k}| > k_0$ the overlap for the full and ring pupil is exactly the same. Indicating that both OTFs share the same value (also see Fig.A.7a). b Spectral improvement factors *IF* for different splitting radii. The narrower the annular pupil gets, the higher the maximally achievable improvement factor. However, that stronger enhancement is limited to a smaller region in Fourier space. In case the object under investigation mainly consists of very high spatial frequencies (like a grid) it makes sense to use a narrow annular pupil for detection.

In case of the annular pupil the overlap is much reduced for most separations, resulting in a low OTF value for most spatial frequencies. However, there exists a separation k_0 for which the spatial overlap in case of the full and ring pupil are equal (see Fig. A.9a right). Which is the main effect that is used in pupil splitting: an imaging system with an annular pupil is capable of transferring high spatial frequencies with the same signal strength as a conventional WF system. However, it only requires half of the photons (when $R = \sqrt{0.5} \cdot R_{max}$), meaning that splitting & recombination can be employed to improve upon *SNR*.

10 Improvement factor for varying pupil splitting radius

The spectral improvement factor *IF* is used to characterize the performance of the pupil splitting compared to the conventional WF detection. Figure A.9b shows the *IF* for four different splitting radii, characterized by R/R_{max}. While the annular pupil is changing from a thick into a thin ring, the region of enhancement (blue) does the same. To improve a larger fraction of spatial frequencies within the bandlimit, it is required to make the ring wide. However, the maximum improvement which can be obtained using the pupil splitting gets stronger when the width of the annular pupil is decreased. Hence, there is a tradeoff between obtaining a strong improvement for only a small fraction of spatial frequencies, or enhancing a larger region in *k*-space but being limited to a lower maximum *IF*. Table 3.5 shows different performance measures in dependency of R/R_{max}.

11 Elongating the depth-of-field in pupil splitting

Pupil splitting does also affect the axial performance of the detection system. This will lead to an elongation of the depth-of-field (DoF), as both sub-PSF will be able to transfer more axial information content into their respective 2D image.

The axial extent of a WF detection PSF can be estimated to be given by $d_{z,min}$:

$$d_{z,min} = \frac{\lambda}{n \cdot [1 - \cos\alpha]} \tag{A.150}$$

with λ the wavelength of light, n the refractive index of the immersion medium and α the half-opening angle of the detection cone. Expressing $d_{z,min}$ in terms of the NA yields:

$$d_{z,min} = \frac{\lambda}{n} \left[1 - \sqrt{1 - \left(\tfrac{NA}{n}\right)^2} \right]^{-1} \tag{A.151}$$

Here the definition $NA = n \cdot \sin\alpha$ and the identity $\sin^2\alpha + \cos^2\alpha = 1$ have been used. Meaning that the axial extend of h_{WF} and h_{Disk} are given according to:

$$\begin{aligned} d_{z,WF} &= \frac{\lambda}{n} \left[1 - \sqrt{1 - \left(\tfrac{NA}{n}\right)^2} \right]^{-1} \\ d_{z,Disk} &= \frac{\lambda}{n} \left[1 - \sqrt{1 - \left(\tfrac{R}{R_{max}}\right)^2 \cdot \left(\tfrac{NA}{n}\right)^2} \right]^{-1} \end{aligned} \tag{A.152}$$

For h_{Ring} it can be used that the sub-pupils add up to yield the (full) widefield pupil:

$$k_{z,WF} = k_{z,Disk} + k_{z,Ring} \tag{A.153}$$

Because the autocorrelation operation & Fourier transforms are linear operators, the axial extend of h_{Ring} is given as:

$$d_{z,Ring} = \frac{d_{z,WF} \cdot d_{z,Disk}}{d_{z,Disk} - d_{z,WF}} \tag{A.154}$$

Computing the axial extent for h_{WF} and the two sub-PSFs h_{Disk} and h_{Ring} for $NA = 1.2$, $n = 1.333$, $\lambda = 500$ nm and $R/R_{max} = \sqrt{0.5}$ yields:

$$d_{z,WF} = 664 \text{ nm} \tag{A.155}$$

$$d_{z,Disk} = 1640 \text{ nm} \tag{A.156}$$

$$d_{z,Ring} = 1117 \text{ nm} \tag{A.157}$$

Meaning that both sub-PSF show an axial extend enlarged by roughly a factor of two. Note that weighted averaging in Fourier space is not a linear operation in real space, meaning that we cannot predict the axial extent $d_{z,wa}$ of the recombined results directly from $d_{z,Disk}$ and $d_{z,Ring}$. Nevertheless, it can be said that for the given example of splitting with $R/R_{max} = \sqrt{0.5}$ the axial extent of $h_{wa,\sigma}$ must be roughly elongated by a factor of two. This agrees with the simulated results shown in Fig. 3.12b.

12 Computing the sub-pupils using the Richards & Wolf model

Instead of calculating the individual OTFs directly from the pupil function (using the autocorrelation operation), it is also possible to obtain it via Fourier transforming a PSF:

$$\tilde{h}(\vec{k}) = \mathcal{F}_{dim}\{h(\vec{r})\} \tag{A.158}$$

This is helpful because there already exist many accurate models to compute h in real space, e.g. the Richards & Wolf method (RW) introduced in [32]. These models encompass effects such as focusing under high angles, the aplanatic factor and the polarization of light. The Richards & Wolf method requires input parameters such as: numerical aperture, refractive index of the immersion medium and the polarization state. The RW method (denoted by f_{RW}) is able to calculate the accurate shape of the PSF for a widefield system in terms of its electric field, the amplitude-spread-function a:

$$a_{WF}(\vec{r}) = f_{RW}(NA, n, \text{Pol.}) \tag{A.159}$$

$$a_{Disk}(\vec{r}) = f_{RW}(NA, n, \text{Pol.}) \tag{A.160}$$

The corresponding intensity PSFs are given according to:

$$h_{WF}(\vec{r}) = \left|a_{WF}(\vec{r})\right|^2 \tag{A.161}$$

$$h_{Disk}(\vec{r}) = \left|a_{Disk}(\vec{r})\right|^2 \tag{A.162}$$

To calculate the PSF corresponding to the ring pupil we use the fact that $\tilde{\mathcal{P}}_{Ring}$ can be expressed in terms of $\tilde{\mathcal{P}}_{WF}$ and $\tilde{\mathcal{P}}_{Disk}$:

$$\tilde{\mathcal{P}}_{Ring}(\vec{k}) = \tilde{\mathcal{P}}_{WF}(\vec{k}) - \tilde{\mathcal{P}}_{Disk}(\vec{k}) \tag{A.163}$$

In general, a is given as the inverse Fourier transform $\mathcal{F}_{dim}^{-1}$ of the pupil:

$$a(\vec{r}) = \mathcal{F}_{dim}^{-1}\{\tilde{\mathcal{P}}(\vec{k})\} \tag{A.164}$$

Which means that a_{Ring} is given as the following subtraction:

$$a_{Ring}(\vec{r}) = a_{WF}(\vec{r}) - a_{Disk}(\vec{r}) \tag{A.165}$$

Hence, the corresponding intensity point-spread-function h_{Ring} can be directly computed from a_{WF} and a_{Disk}, according to:

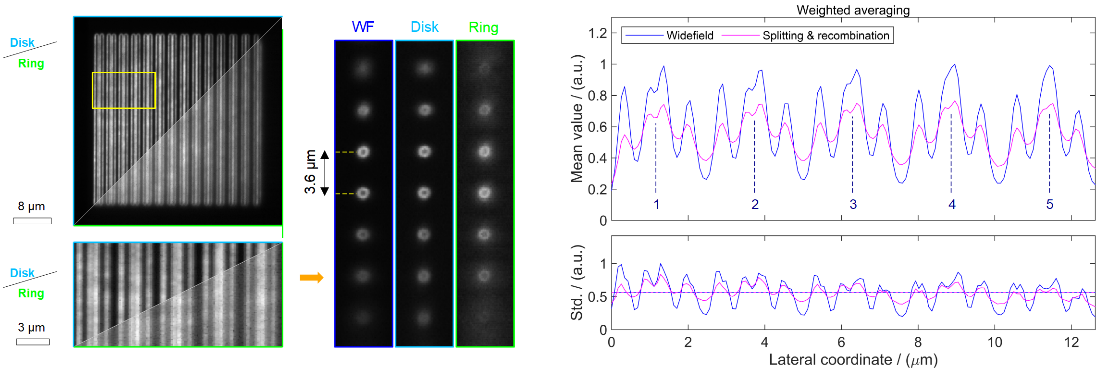
$$h_{Ring}(\vec{r}) = |a_{Ring}(\vec{r})|^2 = \left|a_{WF}(\vec{r}) - a_{Disk}(\vec{r})\right|^2 \tag{A.166}$$

So that the RW model f_{RW} can be used to calculate h_{WF} and h_{Disk} directly and h_{Ring} indirectly by using the previous equation.

13 Image results of pupil splitting & recombination

The acquired data of the pupil splitting experiment is shown in Fig. A.10a. Left the resolution target (pattern E, average of 100 frames) and right the "crossing stairs" (pattern I, single frame) of the Argolight calibration target [78]. When comparing the two images, corresponding to *disk* (cyan) and *ring* pupil (green), it is noticed that the latter looks more blurred. Yet it still recovers the very fine spatial structures. The blur is due to the much worse transfer of medium spatial frequencies through the ring pupil, see green OTF in

Fig. 3.12a. High frequency information maintains its transfer strength, but becomes more clearly visible as relative shot noise is reduced through the splitting.



a

b

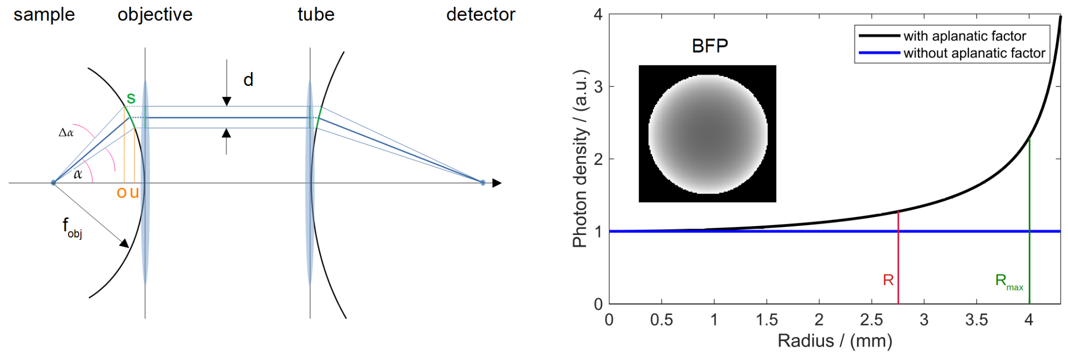
Figure A.10: a Imaging results pupil splitting & recombination of the resolution (left) and stair (right) sample of the Argolight calibration target. Marked with cyan color are the sub-images corresponding to the inner disk pupil, cyan those related to the ring pupil. Latter shows some strong blurring, while still maintaining information about the very fine structures. Pupil splitting generates a slight extended-depth-of-field effect, which is indicated at the orange arrow, in the imaged "stair" sample. b Line profile through resolution target (center of yellow box) for the weighted averaged recombination. The change of modulation depth for the different line pairs corresponds to the theoretical prediction in Fig. 3.12b.

The extended-depth-of-field (EdoF) effect obtained in pupil splitting can be seen in both sub-images of the "stair" target marked by the orange arrow. The inner disk pupil (cyan) corresponds to imaging with a lower *NA*, hence exhibits a longer axial range of detection. Detection through the annular pupil (cyan) also leads to EDoF, as the corresponding PSF corresponds to a Bessel [79] type of intensity distribution (more details in sec. 127).
Figure A.10b shows the line profile through the resolution target for widefield and weighted average recombination. The pixel-wise mean and standard deviation has been computed, so that the data is shown noise-normalized (bottom). Any improvements in terms of *SNR* are observable as an enhancement in terms of visibility $\mathcal{V}$. As suggested from the spectral improvement factor *IF* in Fig. 3.12b, medium spatial frequencies are transferred worse when pupil splitting is employed. However, the very small structures (corresponding to high frequencies) are transferred better, enabling to detect a modulation in the split data, which previously has been impossible using conventional widefield detection.

14 The aplanatic factor modifying the pupil splitting

When a point source emits light, the corresponding intensity distribution in the BFP of an objective is not uniformly distributed. For higher emission angles α the light emitted into a specific solid angle (see Fig. A.11a) needs to be projected onto a smaller area. This effect is termed the *aplanatic factor* and results in a measurable intensity distribution in the BFP, which increases sharply towards the edge of the pupil (see Fig. A.11b, inset).

To quantify this effect, a comparison between the circumference s of the spherical cap and the diameter d of the disk in the parallel beampath (see Fig. A.11a), needs to be performed. Note that this essentially is a 1D problem, as the "compression" of the light only takes place in a single direction.



a b

Figure A.11: a A point source emits light under an angle α into a light cone with opening angle $\Delta\alpha$. The light is then projected from the Gaussian reference sphere onto the pupil, which results in a "squeezing" of intensity towards larger emission angle α. This effect is typically named the *aplanatic factor* γ. Reason for this behaviour is that an area element s on the spherical cap is converted into a smaller segment d in the parallel beam path. Hence the photon density is increased, leading to higher intensity values towards the edge of the BFP. b Photon density in the BFP (= photons per unit area) in dependency of the radial position in blue (no aplanatic effect in blue). The inset shows the BFP of an microscope objective including the aplanatic factor for emission, indicating how more photons are "squeezed" towards the rim of the pupil.

Therefore, the aplanatic factor γ is given as the ratio:

$$\gamma(\alpha, \Delta\alpha) = \frac{s(\Delta\alpha)}{d(\alpha)} \tag{A.167}$$

with α being the angle under which the light ray hits the Gaussian reference sphere and $\Delta\alpha$ the angular width of the resulting spherical cap. The circumference s is calculated according to:

$$s(\Delta\alpha) = 2\pi f_{obj} \cdot \tfrac{\Delta\alpha}{2\pi} = f_{obj} \cdot \Delta\alpha \tag{A.168}$$

with f_{obj} being the focal length of the objective which also determines the obtainable FoV. The diameter d of the disk is given as the difference between the two triangle sides o & u:

$$\begin{aligned} o &= f_{obj} \cdot \sin(\alpha + \Delta\alpha) && \text{(A.169)} \\ u &= f_{obj} \cdot \sin(\alpha - \Delta\alpha) && \text{(A.170)} \end{aligned}$$

Making use of the addition theorem of sine functions [31]:

$$\sin(x \pm y) = \sin x \cos y \pm \cos x \sin y \tag{A.171}$$

With this it is possible to express o and u as:

$$\begin{aligned} o &= f_{obj} \cdot (\sin\alpha \cos\Delta\alpha + \cos\alpha \sin\Delta\alpha) && \text{(A.172)} \\ u &= f_{obj} \cdot (\sin\alpha \cos\Delta\alpha - \cos\alpha \sin\Delta\alpha) && \text{(A.173)} \end{aligned}$$

Which simplifies to the following expression for the diameter d:

$$d(\alpha) = o - u = 2f_{obj} \cdot \cos\alpha \sin\Delta\alpha \tag{A.174}$$

The overall aplanatic factor is now given according to:

$$\gamma(\alpha, \Delta\alpha) = \frac{1}{\cos\alpha} \cdot \frac{\Delta\alpha}{2\sin\Delta\alpha} \tag{A.175}$$

Lets consider that expression for small $\Delta\alpha$, hence:

$$\lim_{\Delta\alpha\to 0} \gamma(\alpha, \Delta\alpha) = \frac{1}{\cos\alpha} \cdot \frac{0}{0} \tag{A.176}$$

Applying the rule of *l'Hospital* leads to:

$$\lim_{\Delta\alpha\to 0} \gamma(\alpha, \Delta\alpha) = \frac{1}{\cos\alpha} \cdot \frac{1}{2\cos\Delta\alpha \cdot \frac{1}{2}} = \frac{1}{\cos\alpha} \tag{A.177}$$

Hence the aplanatic factor (in detection) is given as:

$$\gamma(\alpha) = \frac{1}{\cos\alpha} \tag{A.178}$$

Instead of expressing γ in terms of the emission angle α, it can also be rewritten by substituting the following:

$$\alpha = \arcsin\left(\frac{r}{n \cdot f_{obj}}\right) \tag{A.179}$$

So that the dependency on the radial position r in the BFP becomes more clear. The aplanatic factor is then given as:

$$\gamma(r) = \frac{1}{\cos\left[\arcsin\left(\frac{r}{n \cdot f_{obj}}\right)\right]} \tag{A.180}$$

The corresponding curve can be interpreted as a photon density in the BFP and is plotted as the black curve in Fig. A.11b, for the experimental parameters used in sec. 3.4.4.
Note that when the aplanatic effect is not considered, that the photon density would be a constant. Hence, splitting the detection pupil with radius R_{max} into a smaller disk (with radius R) and a ring, will only lead to an equal area split for $R/R_{max} = \sqrt{0.5}$ without accounting for the aplanatic factor. This can be seen when calculating the integral:

$$A(r_{min}, r_{max}) = \int_0^{2\pi} d\phi \int_{r_{min}}^{r_{max}} dr\ r = \pi \cdot \left[r_{max}^2 - r_{min}^2\right] \tag{A.181}$$

with $r_{min,max}$ being the radii integrated over in the BFP.
Obtaining the ratio for the disk pupil $A(0, R)$ and the widefield pupil $A(R, R_{max})$ gives:

$$\frac{A(0,R)}{A(R,R_{max})} = \frac{R^2}{R_{max}^2 - R^2} = \frac{1}{(R_{max}/R)^2 - 1} \tag{A.182}$$

which is equal to 0.5 for $R/R_{max} = \sqrt{0.5}$.
When the aplanatic factor is included, the integral changes to:

$$A(r_{min}, r_{max}) = 2\pi \cdot \int_{r_{min}}^{r_{max}} dr\ r \cdot \gamma(r) = 2\pi \cdot \left(f_{obj} n\right)^2 \cdot \left[1 - \sqrt{1 - \left(\frac{r}{f_{obj} n}\right)^2}\right]_{r_{max}}^{r_{min}} \tag{A.183}$$

Obviously this changes the ratio $A(0,R)/A(R,R_{max})$ and therefore also the effective splitting ratio. E.g. for the experimental parameters in sec. 3.4.4, the values of $A(0,R)$ and $A(R,R_{max})$ are given as:

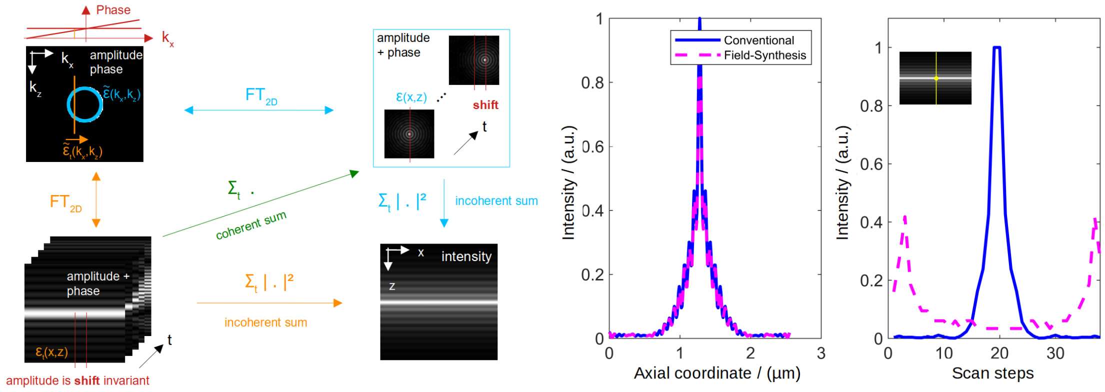
$$A(0,R) \propto 0.2740 \tag{A.184}$$

$$A(R,R_{max}) \propto 0.7260 \tag{A.185}$$

with $f_{obj} = f_{tube}/M = 3$, $n = 1.333$, $R_{max} = 4.0$ mm, $R/R_{max} = 0.68$ and $R = 2.75$ mm. Yielding a ratio of $A(0,R)/A(R,R_{max}) = 0.3774$, which is equivalent to a splitting ratio of $R/R_{max} = \sqrt{0.3774} = 0.61$. Meaning that the pupil splitting realized in the experiment does not yield an equal area split, and effectively changes the region of improvement ($IF > 0$) by setting $k_0 = 0.5 \cdot (1 + 0.61)/k_{xy,max} = 0.805/k_{xy,max}$.

15 Field-Synthesis reduces phototoxicity

The conventional way of creating a scanned light-sheet is by beam scanning (BS). The BFP of the illumination objective is homogeneously illuminated with light which is focused into the sample. In Fig. A.12a this is exemplified for an annular mask, meaning that the objective lens converts the BFP electric field $\tilde{\mathcal{E}}$ into a focused Bessel spot $\mathcal{E}$ (top right).


Phase
k_x
k_x
k_z
amplitude
phase
Ẽ(k_x,k_z)
Ẽ_t(k_x,k_z)
FT_2D
FT_2D
amplitude + phase
ε(x,z)
shift
t
Σ_t ·
coherent sum
Σ_t | · |²
incoherent sum
Σ_t | · |²
incoherent sum
amplitude + phase
ε_t(x,z)
t
amplitude is shift invariant
x
z
intensity
Conventional
Field-Synthesis
Intensity / (a.u.)
Axial coordinate / (µm)
0
1
2
3
0
0.2
0.4
0.6
0.8
1
Intensity / (a.u.)
Scan steps
0
10
20
30

a b

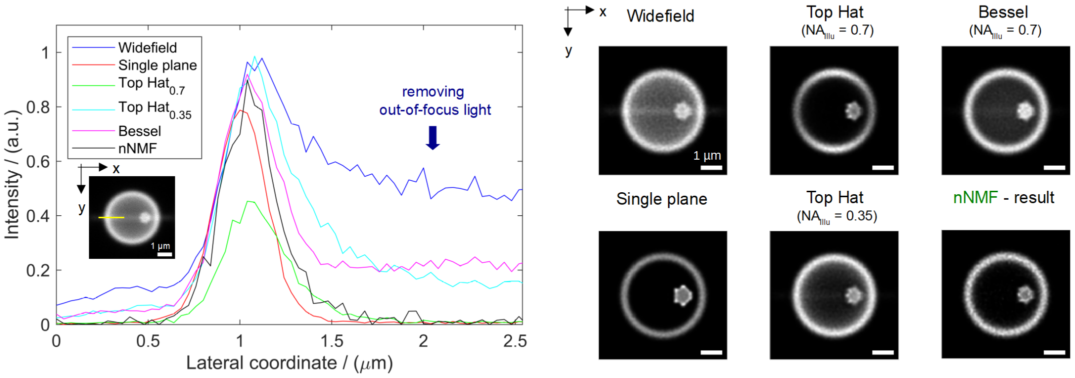
Figure A.12: a Schematic describing the Field-Synthesis theorem: the goal is to create a time-averaged light-sheet illumination (bottom right). This can either be done by scanning a focused laser spot along the x-direction (top right, beam scanning), or by line scanning in the BFP (top left, Field-Synthesis). The connections between the respective electric fields indicated are explained in more details in the text. b Left: comparing the illumination profile of conventional beam scanning (blue) with Field-Synthesis (magenta). Right: Latter reduces the maximum illumination dose (here by a factor of 2.5), as the overall illumination gets distributed over the whole scanning window . Which benefits the sample in terms of phototoxicity and photobleaching [94, 95].

In BS this spot gets scanned along the x-axis, to create a time-averaged light-sheet (cyan path in A.12a). The scanning is achieved by applying a phase ramp $e^{i \cdot t k_z}$ in the BFP, where t corresponds to the different time steps. The Field-Synthesis theorem states that the same intensity distribution can be achieved when a line parallel to k_z is scanned in k_x-direction in the BFP of the illumination objective. For each scan position a distinct electric field $\tilde{\mathcal{E}}_t$ is obtained, which summed *coherently* yield $\tilde{\mathcal{E}} = \int dt\ \tilde{\mathcal{E}}_t$.

All line scans together are equivalent with illuminating the full pupil (here annular mask) at once. Due to the linearity of the Fourier transform, we can conclude that the Bessel spot $\mathcal{E}$ in real space can be generated by coherently summing the real space analog of $\mathcal{E} = \int dt\ \mathcal{E}_t$. The resulting electric field will coincide with $(x,z) = (0,0)$, the same as in the BS approach when no phase ramp is applied. In case a phase ramp is applied in the BFP (to move $\mathcal{E}$ along the x-axis), this will not affect the individual $\tilde{\mathcal{E}}_t$. This is because the line scan is perpendicular to the phase ramp direction, meaning that each line scan will not experience any phase change along k_x. However, the phase between subsequent scan positions will change and this results in the shifting of $\mathcal{E}$ along the x-axis. Of course, the incoherent summation of FS anyways neglects this varying phase information between different pupil scan positions. Meaning that all shifted illumination foci $\mathcal{E}$ are already encoded in the incoherent summation of $\mathcal{E}_t$. So that the incoherent superposition of $\mathcal{E}_t$ effectively corresponds to when all possible phase ramps are applied simultaneously, creating the scanned light-sheet. A.12b left shows a line plot along the z-axis (yellow line) and compares the temporally averaged light-sheet created using beam scanning (blue) or Field-Synthesis (magenta). Note how both curves are identical, nevertheless FS brings some advantage. Field-Synthesis distributes the same illumination dose over a larger amount of time. In beam scanning, a point in the sample (e.g. yellow point in Fig. A.12b) gets all its excitation within a very short period of time (blue curve). In Field-Synthesis the illumination is distributed more evenly over the whole excitation volume (see Fig. A.12b, magenta). Reducing the maximum illumination dose drastically (here by a factor of 2.5), making Field-Synthesis much more gentle in terms of phototoxicity [94, 95].

16 Reducing out-of-focus light through illumination splitting

Generating a Bessel illumination sheet using the Field-Synthesis concept, enables to record multiple sub-images that correspond to axially varying sub-illuminations. Figure 3.18b depicts simulation results for imaging a 3D spherical object, with different imaging modalities. To visualize the ability to remove out-of-focus light when the sub-images are recombined, a line profile (yellow) is plotted in Fig. A.13a.



a b

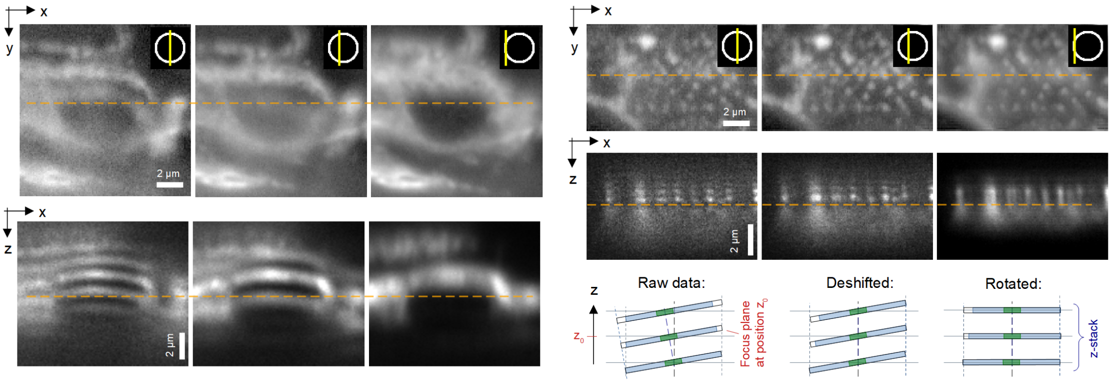
Figure A.13: a Line profiles through the images shown in Fig. 3.18b . It can be clearly seen how light-sheet illumination (cyan) removes some of the out-of-focus light, present in the widefield data (blue). Creating the Bessel sheet (magenta) will yield in unwanted background contributions, due to the strong sidelobes exciting out-of-focus regions. The nNMF processing (black) enables to remove this additional out-of-focus light and maintains background-free imaging over a large FoV. b Thin slice deconvolved results of images shown in Fig. 3.18b (20 accelerated iterations). Note that the nNMF result corresponds to data which has been pre-processed using nNMF.

The region inside of the imaged spherical object indicates that post-processing the data using the nNMF-algorithm (described in sec. 3.5.3) enables to remove out-of-focus contributions. While not introducing any further blurring (see walls of spherical object and compare to Top hat $NA_{Illu.} = 0.35$ result) and maintaining a larger FoV.

Figure A.13b shows the deconvolved (thin slice) results of the images shown in Fig. 3.18b after 20 accelerated iterations. Note that these results should just give an exemplary impression on what is possible when deconvolution is performed. Because the nNMF processed data has most of its out-of-focus light already removed, the image deconvolution has it easier to obtain good reconstruction results.

17 Image results of illumination splitting & recombination

Figure A.14 shows the lateral in-focus slice (top) of the recorded z-stack, corresponding to the axial modulation of the excitation pattern (bottom). Note how the lateral information varies when the scan-position in the illumination BFP is changed. This effect is used in illumination splitting, as the in-focus structure is present in all sub-images. For the bead sample this effect is smaller, which is probably due to the smaller axial extend. Note that the optical configuration shown in Fig. 3.19a, shows a light-sheet system where both objective (illumination & detection) are on the same side of the coverslip. Hence, the recorded "z-stack" is tilted and laterally shifted with respect to the optical axis of the detection objective. This is undone by pre-processing the data with a shift correction and optional rotation of the acquired volume information. Together with the data, *James Manton* (LMB) provided a Fiji-script [144] which performed the necessary corrections.



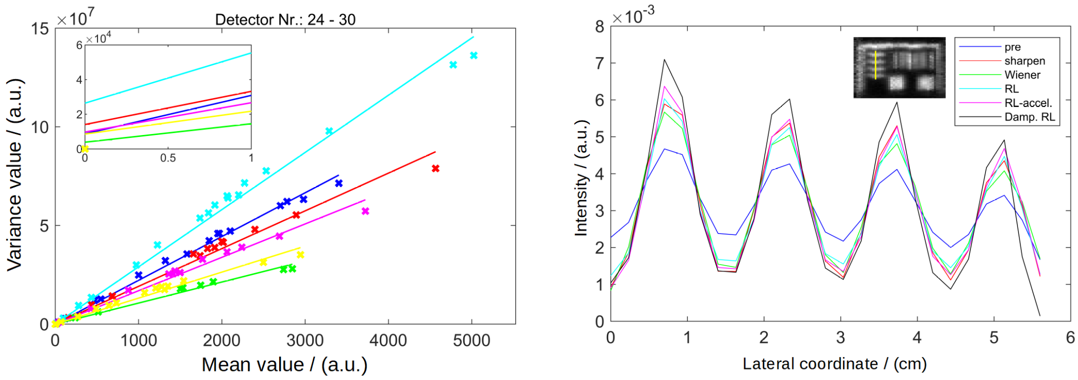
a b

Figure A.14: Experimental results of imaging the Vimentin probe (a) and fluorescent beads (b) using illumination splitting through Field-Synthesis (all data is shown with a γ-adjustment of $\gamma = 1.5$). Top: individual in-focus images for three different scan positions in the illumination BFP. Bottom: axial view acquired from a z-stack indicating the axial modulation of the illumination. a Note the change of image information in the lateral view (top), which is directly related to the axial modulation in the side view. Meaning that when imaging a 3D object, any axially modulated illumination will have an influence on the 2D image information that is being captured. b Top: Compared to the nucleus of the Vimentin sample, the distribution of beads can be regarded as relatively flat. Nevertheless, the illumination splitting alters some of the image content in the lateral view as well. Note that the captured data suffers from spherical aberrations. Bottom: the data was acquired with a light-sheet configuration with both objectives (illumination & detection) on the same side of the coverslip (see Fig. 3.19a). Hence the acquired "z-stack" is tilted and shifted with respect to the z-axis. This is corrected by performing a shift correction and rotation of the captured volume using a Fiji [144]-script provided by *James Manton* (LMB).

Submillimeter wave imaging:

1 Mean-variance relationship of submillimeter detectors

In every imaging application, it is necessary to calibrate the sensor data. One important aspect is the ratio of *mean* to *variance* value for each detector, as it quantitatively describes the effects of noise on the image data. For example: shot noise is characterized as a straight line in the *mean-variance* plot, as it is governed by a Poisson distribution which exhibits: expectancy = variance. To be able to observe this relationship it is necessary to analyze a sequence of image data where all possible signal levels are present, e.g. a person walking through the complete FoV. For each detector in this data stack, the measured signal range has been split into N bands and for each of those the mean and the variance value has been calculated. The result is shown in Fig. A.15a.



a

b

Figure A.15: a *Mean-variance* plot of a scene with a person walking through the camera FoV. The different colors indicate the detectors from module Nr. 4, the lines are fitted using a least-squares approach. Each detector for itself is working in the shot noise regime (expectancy = variance), however, they all differ in sensitivity (= slope) and offset. Later is depicted in the small inset and shows a maximum difference of $\approx 2.75 \cdot 10^4$, explaining the "streaky" appearance of the raw data in Fig. 4.2. b Modulation pattern obtained via a line profile (marked yellow in resolution target). Blue corresponds to the data after pre-processing, without any image enhancement. The ability to improve upon spatial resolution is shown as a visibility $\mathcal{V}$ increase, especially for the accelerated and dampened version of the Richardson-Lucy deconvolution, in Tab. A.2.

The *mean-variance* curves for the 8 detectors of the 4th detector sub-module are shown in different colors. They all represent a linear relationship, which has been confirmed by fitting a line using a least-squares approach. However, each detector shows an individual sensitivity (= slope) and offset behavior, which need to be corrected for when the image data is to be generated. Especially the offset characteristics (see small inset) show a maximum difference of the different detectors on the order of $\approx 2.75 \cdot 10^4$, i.e. four orders of magnitude. Note that these measured dependencies of the individual pixels drift over time and change completely whenever the system has to be setup again. Due to thermal fluctuations the individual detector response changes, which leads to the requirement of consistently obtaining calibration data after each system setup.

2 Improved spatial resolution in submillimeter wave imaging

As seen from sec. 4.1, submillimeter wave imaging is limited in terms of *SNR* and therefore also in spatial resolution. Post processing techniques, such as image deconvolution, try to overcome this. In chapter 4.3.2 different methods have been shown: sharpening, Wiener filter and Richardson Lucy deconvolution. For latter an acceleration [54] and dampening step [55] was introduced. The result of imaging the resolution line pattern is shown as a profile plot in Fig. A.15b. For all post-processing an enhancement in modulation depth or visibility $\mathcal{V}$ can be observed. The exact values of the spatial improvement facgtor $IF_{\mathcal{V}}$ are depicted in Table A.2.

	Pre	Sharpen	Wiener	RL	RL-accel.	Damp. RL
M_{max} / (a.u.)	0.0041	0.0052	0.0049	0.0052	0.0055	0.0060
M_{min} / (a.u.)	0.0022	0.0012	0.0014	0.0015	0.0013	0.0011
Visibility / (%)	30.94	62.13	56.65	54.18	60.40	68.74
Improvement / (%)	0.0	100.8	83.1	75.1	95.2	122.2

Table A.2: Table showing the visibility $\mathcal{V}$ and its spatial improvement factor $IF_{\mathcal{V}}$ in *percent*. An enhancement of $IF_{\mathcal{V}} = 100\%$ corresponds to a doubling of modulation. RL = 10 (accelerated) iterations; Accel. + Damp. RL = 7 (accelerated) iterations.

The technique of denoising (DCT) and sharpening (subtracting Laplace filtered version) yields in a doubling of visibility. However, when looking at Fig. 4.4 this results in strong artifacts such as noise amplification. Wiener filtering gives a slight decrease in enhancement but is not as strongly affected by restoration artifacts. Conventional Richardson-Lucy yields in 75% improvement which can be strongly improved (120%) by when the dampened version [55] is used.

3 Temporal representation of the recorded video data

Besides providing a large enough *SNR* and spatial resolution, the passive submillimeter wave security imager also needs to be able to acquire the data fast enough. So that the movement of a potential suspect can be followed. The prototype at the IPHT did acquire the image data at ≈ 12 frames per second (fps). Processing the data was done using the accelerated & damped Richardson-Lucy at > 4 fps on a Laptop (Intel i7 8565U; 1.8 GHz, 4 cores) but not optimized (e.g. parallelized). Figure A.16 shows a series of frames, each captured in an time interval of $t \approx 0.5$ seconds.

In both cases the movement of both person is nicely resolved in time. Note that those scenes indicate that a concealed threat can only be detected when it directly faces the security imager. This is not surprising as the underlying contrast mechanism is to measure a temperature difference between object and surrounding media (body).

4 Supervised and unsupervised learning-based denoising

In Deep-Learning-based image reconstruction a neural network must be trained to recover the underlying sample structure from low *SNR* images. Those networks consists of multiple layers of neurons (blue circle in Fig. A.17a) which are mutually connected. A single neuron processes information from its adjacent neighbors x_i by computing a weighting sum (weights w_i), which is input to a non-linear function (activation function) yielding the output of the neuron. The interaction between all neurons in the different layers is controlled by setting each individual weight w_i, which represents a vast number of adjustable parameters of the network [141]. Their respective values are optimized during a step called network training, which typically is very time consuming.
When pairs of noisy image and groundtruth data are available, this is called *supervised* training and is shown in Fig. A.17b as the CARE network [140]. It uses a convolutional

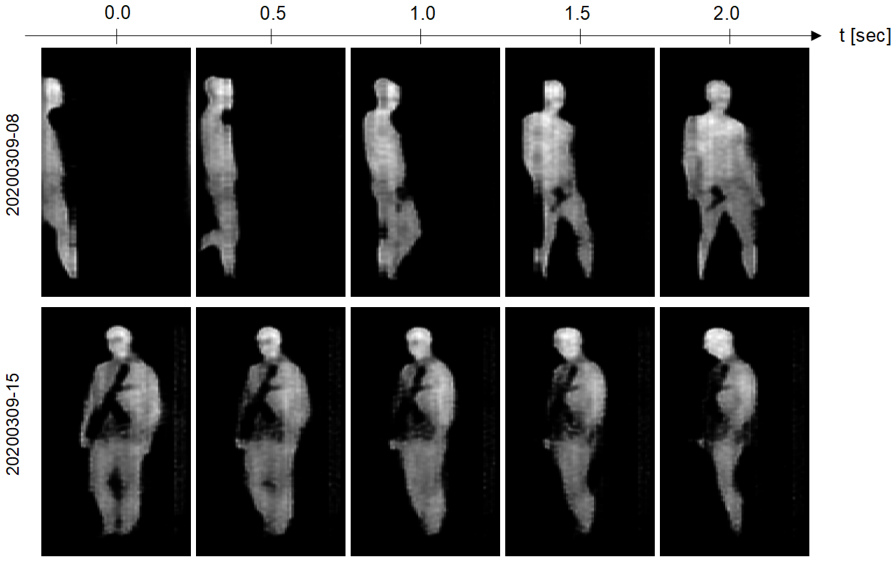
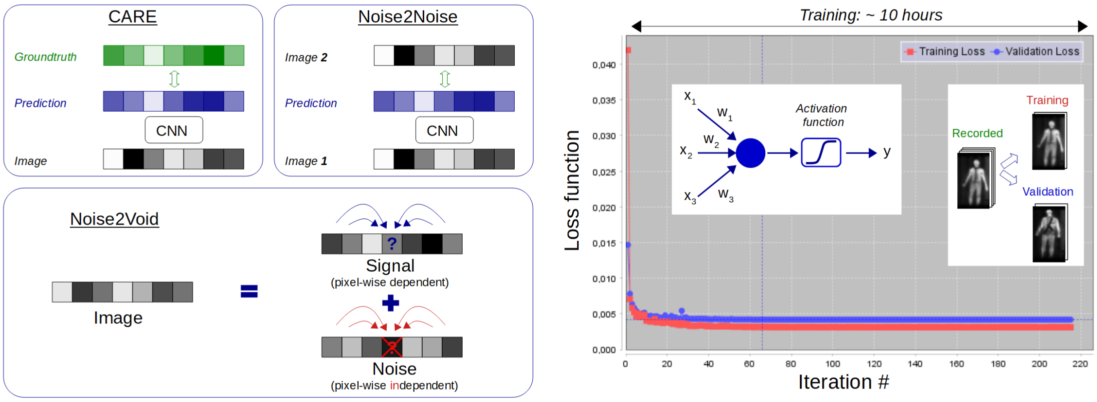


Figure A.16: Temporal overview of recorded submillimeter wave imaging. Top: person walking from the left to the middle of the FoV. Note how the hidden handgun is not detectable while the person is facing in walking direction. Bottom: person with hidden rifle turning around. Again, when the concealed threat faces away from the security imager, no weapon can be detected.

neural network (CNN) [138] to predict denoised images. It is also possible to train the CNN by using multiple low *SNR* images only (without groundtruth knowledge). The idea behind this is that both sub-images carry the same signal information but a varying noise contribution [142]. As the network is not able to predict the noise of the reference image, it will learn find the underlying sample structure of both sub-images. Noise2Void [139, 143] goes even further and enables to train on single noisy images alone. However, the noise is required to be pixel-wise independent. Which is the case for the most common noise contributions, such as shot noise. The basic idea behind Noise2Void is that a network will not be able to predict the noise in a pixel from its surrounding, nevertheless can still find some information on the sample structure in a close neighborhood. Making it possible to successfully train the network according. However, this only works when a masking strategy is used, preventing the measured information at the pixel which is ought to be predicted to be part of the receptive field of the network (see Fig. 4.5a).

The denoising network for the submillimeter wave data was applied using a Fiji-plugin [144] that was made available by the Jug lab [139]. The plugin first separates a provided dataset into a training (red) and a validation (blue) batch. Latter is unseen by the network and can be used to verify the accuracy of the obtained reconstruction result. The two curves depict the loss function for training and validation and are shown in Fig. A.17a. For the first 40 iterations a quick drop is observable, from which one the improvement is much more gradually. Note that the training error is always slightly below the validation loss, as the validation data is unseen by the network and not directly trained for.



a b

Figure A.17: Supervised and unsupervised learning based denoising. a Top left: Content aware image reconstruction (CARE) [140] enables to denoise an image using a convolutional neural network (CNN). It requires the knowledge of groundtruth data (green) which is compared in training to the networks prediction (blue). Top right: Lehtinen et al. [142] have shown that the groundtruth data can be replaced by a second recorded image, consisting of the same underlying signal but varying noise. Bottom: *Noise2Void* [139, 143] shows that it is also possible to train the CNN using a single image. The required assumption is that the underlying signal can be reconstructed from a neighborhood, while the noise is pixel-wise independent and cannot be predicted from adjacent neighbors. b Loss curves for training (red) and validation (blue) after ≈ 10 hours; a recorded dataset is randomly separated into a training & validation data. Both curves quickly decrease, while the training loss is always slightly reduced, indicating that a sensible solution has been found. The processing of a single neuron depends on the input values x_i, the weights w_i and a non-linear activation function.

Appendix B

List of publications

Paper:

- Dasgupta A, Deschamps J, Matti U, Hübner U,Becker J, Strauss S, Jungmann R, Heintzmann R, Ries J. Direct supercritical angle localization microscopy for nanometer 3d superresolution. Nature Communications. **2021** Feb 19;12(1):1-9.
- Heintzmann R, Becker J; Measuring by Darkness? Let there be light!, (Version v0.1), Zenodo, http://doi.org/10.5281/zenodo.3629784, (**2020**)
- Becker J, Heintzmann R. PSF broadening due to fluorescence emission. bioRxiv. **2019** Jan 1.
- Jügler A, Becker J, Förster R, Heintzmann R. Adaptive holographic region of interest illumination with oblique angles for use in single molecule localization microscopy. InEuropean Conference on Biomedical Optics **2019** Jun 23 (p. 11076-1). Optical Society of America.
- Jügler A, Becker J, Then P, Heintzmann R. Holographic Region-of-Interest with Oblique Illumination. InFrontiers in Optics **2018** Sep 16 (pp. FM4E-5). Optical Society of America.
- Becker J, Förster R, Heintzmann R. Better than a lens-A novel concept to break the *SNR*-limit, given by Fermat's principle. arXiv preprint arXiv:1811.
08267. **2018** Nov 20.

Talks:

- PSF broadening due to fluorescence emission: principle and implications on imaging, Quantitative Bioimaging Conference (QBI, **2020**), Oxford, Session: Point spread function analysis
- Better than a lens, Focus on Microscopy (FoM, **2019**), London, Session: Optical, Theory III: Image Formation, Modelling I

References

[1] *Smartphones Cause Photography Boom*. URL: `https://www.statista.com/chart/10913/number-of-photos-taken-worldwide/`.

[2] Hedvig Hricak. "2016 New Horizons Lecture: Beyond Imaging—Radiology of Tomorrow". In: *Radiology* 286.3 (2018). PMID: 29346031, pp. 764–775. DOI: `10.1148/radiol.2017171503`. eprint: `https://doi.org/10.1148/radiol.2017171503`. URL: `https://doi.org/10.1148/radiol.2017171503`.

[3] Rongguang Liang. *Biomedical optical imaging technologies: design and applications*. Springer Science & Business Media, 2012.

[4] Reiner Salzer. *Biomedical Imaging: Principles and Applications*. John Wiley & Sons, 2012.

[5] Mary B Stuart, Andrew JS McGonigle, and Jon R Willmott. "Hyperspectral imaging in environmental monitoring: a review of recent developments and technological advances in compact field deployable systems". In: *Sensors* 19.14 (2019), p. 3071.

[6] Herbert Gross, Fritz Blechinger, and Bertram Achtner. *Handbook of optical systems*. Vol. 1. Wiley Online Library, 2005.

[7] Jerome Mertz. *Introduction to optical microscopy*. Cambridge University Press, 2019.

[8] Ernst Abbe. "Beitraege zur Theorie des Mikroskops und der mikroskopischen Wahrnehmung". In: *Archiv fuer mikroskopische Anatomie* 9.1 (1873), pp. 413–418.

[9] Claude Elwood Shannon. "Communication in the presence of noise". In: *Proceedings of the IRE* 37.1 (1949), pp. 10–21.

[10] Harry Nyquist. "Certain topics in telegraph transmission theory". In: *Transactions of the American Institute of Electrical Engineers* 47.2 (1928), pp. 617–644.

[11] Stefan W Hell et al. "The 2015 super-resolution microscopy roadmap". In: *Journal of Physics D: Applied Physics* 48.44 (2015), p. 443001.

[12] Joel A Tropp et al. "Beyond Nyquist: Efficient sampling of sparse bandlimited signals". In: *IEEE transactions on information theory* 56.1 (2009), pp. 520–544.

[13] Andrea Massa, Paolo Rocca, and Giacomo Oliveri. "Compressive sensing in electromagnetics:a review". In: *IEEE Antennas and Propagation Magazine* 57.1 (2015), pp. 224–238.

[14] GMP Van Kempen et al. "A quantitative comparison of image restoration methods for confocal microscopy". In: *Journal of Microscopy* 185.3 (1997), pp. 354–365.

[15] Peter J Verveer and Thomas M Jovin. "Efficient superresolution restoration algorithms using maximum a posteriori estimations with application to fluorescence microscopy". In: *JOSA A* 14.8 (1997), pp. 1696–1706.

[16] Peter J Verveer, Mark J Gemkow, and Thomas M Jovin. "A comparison of image restoration approaches applied to three-dimensional confocal and wide-field fluorescence microscopy". In: *Journal of microscopy* 193.1 (1999), pp. 50–61.

[17] WA Edelstein et al. "The intrinsic signal-to-noise ratio in NMR imaging". In: *Magnetic resonance in medicine* 3.4 (1986), pp. 604–618.

[18] Johan Philip and Kjell Carlsson. "Theoretical investigation of the signal-to-noise ratio in fluorescence lifetime imaging". In: *JOSA A* 20.2 (2003), pp. 368–379.

[19] TE Gureyev et al. "Complementary aspects of spatial resolution and signal-to-noise ratio in computational imaging". In: *Physical Review A* 97.5 (2018), p. 053819.

[20] Y Reibel et al. "CCD or CMOS camera noise characterisation". In: *The European Physical Journal Applied Physics* 21.1 (2003), pp. 75–80.

[21] J-L Gach et al. "A New Digital CCD Readout Technique for Ultra–Low-Noise CCDs". In: *Publications of the Astronomical Society of the Pacific* 115.811 (2003), p. 1068.

[22] Yue Chen et al. "Column-parallel digital correlated multiple sampling for low-noise CMOS image sensors". In: *IEEE Sensors Journal* 12.4 (2011), pp. 793–799.

[23] J Abadie et al. "A gravitational wave observatory operating beyond the quantum shot-noise limit". In: *Nature Physics* 7.12 (2011), p. 962.

[24] William N Plick, Jonathan P Dowling, and Girish S Agarwal. "Coherent-light-boosted, sub-shot noise, quantum interferometry". In: *New Journal of Physics* 12.8 (2010), p. 083014.

[25] Benjamin J Lawrie et al. "Quantum sensing with squeezed light". In: *ACS Photonics* 6.6 (2019), pp. 1307–1318.

[26] Gregor PC Drummen. *Fluorescent probes and fluorescence (microscopy) techniques illuminating biological and biomedical research*. 2012.

[27] Boris S Karasik and Robin Cantor. "Demonstration of high optical sensitivity in far-infrared hot-electron bolometer". In: *Applied Physics Letters* 98.19 (2011), p. 193503.

[28] James Pawley. *Handbook of biological confocal microscopy*. Vol. 236. Springer Science & Business Media, 2006.

[29] Bahaa EA Saleh and Malvin Carl Teich. *Fundamentals of photonics*. john Wiley & sons, 2019.

[30] Joseph W Goodman. *Introduction to Fourier optics*. Roberts and Company Publishers, 2005.

[31] Ilja N Bronstein et al. *Taschenbuch der mathematik*. Vol. 1. Springer-Verlag, 2012.

[32] B Richards and E Wolf. "Electromagnetic diffraction in optical systems, II. Structure of the image field in an aplanatic system". In: *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* 253.1274 (1959), pp. 358–379.

[33] Max Born and Emil Wolf. *Principles of optics: electromagnetic theory of propagation, interference and diffraction of light*. Elsevier, 2013.

[34] Rainer Heintzmann. "Band limit and appropriate sampling in microscopy". In: *Cell biology*. Elsevier, 2006, pp. 29–36.

[35] Colin JR Sheppard, Min Gu, and Maitreyee Roy. "Signal-to-noise ratio in confocal microscope systems". In: *Journal of Microscopy* 168.3 (1992), pp. 209–218.

[36] Lucas J van Vliet, Damir Sudar, and Ian T Young. "Digital fluorescence imaging using cooled CCD array cameras invisible". In: *Cell Biol* 3 (1998), pp. 109–120.

[37] Robert L Lucke. "Fourier-space properties of photon-limited noise in focal plane array data, calculated with the discrete Fourier transform". In: *JOSA A* 18.4 (2001), pp. 777–790.

[38] Tony Wilson. "Resolution and optical sectioning in the confocal microscope". In: *Journal of microscopy* 244.2 (2011), pp. 113–121.

[39] CJR Sheppard and T Wilson. "The theory of the direct-view confocal microscope". In: *Journal of microscopy* 124.2 (1981), pp. 107–117.

[40] Jim Swoger et al. "Multi-view image fusion improves resolution in three dimensional microscopy". In: *Optics express* 15.13 (2007), pp. 8029–8042.

[41] Ernst HK Stelzer and Steffen Lindek. "Fundamental reduction of the observation volume in far-field light microscopy by detection orthogonal to the illumination axis: confocal theta microscopy". In: *Optics Communications* 111.5-6 (1994), pp. 536–547.

[42] Marin van Heel and Michael Schatz. "Information: to Harvest, to Have and to Hold". In: *arXiv preprint arXiv:2009.03223* (2020).

[43] WH Hsiao and RP Millane. "Effects of occlusion, edges, and scaling on the power spectra of natural images". In: *JOSA A* 22.9 (2005), pp. 1789–1797.

[44] Heintzmann Rainer; Wicker Kai and Colin Sheppard. "Standardization of the Optical Transfer Function". 2020; *manuscript under review*.

[45] XS Gan and CJR Sheppard. "Detectability: a new criterion for evaluation of the confocal microscope". In: *Scanning* 15.4 (1993), pp. 187–192.

[46] E.A. INGERMAN et al. "Signal, noise and resolution in linear and nonlinear structured illumination microscopy". In: *Journal of Microscopy* 273.1 (2019), pp. 3–25. DOI: `https://doi.org/10.1111/jmi.12753`. eprint: `https://onlinelibrary.wiley.com/doi/pdf/10.1111/jmi.12753`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1111/jmi.12753`.

[47] Qiaole Zhao, Ian T Young, and Jan Geert Sander De Jong. "Photon budget analysis for fluorescence lifetime imaging microscopy". In: *Journal of Biomedical Optics* 16.8 (2011), p. 086007.

[48] Stefan Hell and Ernst HK Stelzer. "Properties of a 4Pi confocal fluorescence microscope". In: *JOSA A* 9.12 (1992), pp. 2159–2166.

[49] Alberto Diaspro et al. "Photobleaching". In: *Handbook of biological confocal microscopy*. Springer, 2006, pp. 690–702.

[50] RA Hoebe et al. "Controlled light-exposure microscopy reduces photobleaching and phototoxicity in fluorescence live-cell imaging". In: *Nature biotechnology* 25.2 (2007), pp. 249–253.

[51] Norbert Wiener. *Extrapolation, interpolation, and smoothing of stationary time series: with engineering applications*. MIT Press, 1950.

[52] William Hadley Richardson. "Bayesian-based iterative method of image restoration". In: *JoSA* 62.1 (1972), pp. 55–59.

[53] Leon B Lucy. "An iterative technique for the rectification of observed distributions". In: *The astronomical journal* 79 (1974), p. 745.

[54] David SC Biggs and Mark Andrews. "Acceleration of iterative image restoration algorithms". In: *Applied optics* 36.8 (1997), pp. 1766–1775.

[55] RJ Hanisch, RL White, et al. "Image restoration using the damped richardson-lucy method". In: *The Restoration of HST Images and Spectra II*. Space Telescope Science Institute, 1994, pp. 104–110.

[56] Ernst HK Stelzer. "Contrast, resolution, pixelation, dynamic range and signal-to-noise ratio: fundamental limits to resolution in fluorescence light microscopy". In: *Journal of Microscopy* 189.1 (1998), pp. 15–24.

[57] Cédric Vonesch et al. "The colored revolution of bioimaging". In: *IEEE signal processing magazine* 23.ARTICLE (2006), pp. 20–31.

[58] Andreas Ettinger and Torsten Wittmann. "Fluorescence live cell imaging". In: *Methods in cell biology*. Vol. 123. Elsevier, 2014, pp. 77–94.

[59] Aleksander Jablonski. "Efficiency of anti-Stokes fluorescence in dyes". In: *Nature* 131.3319 (1933), pp. 839–840.

[60] James Franck and EG Dymond. "Elementary processes of photochemical reactions". In: *Transactions of the Faraday Society* 21.February (1926), pp. 536–542.

[61] Michael Kasha. "Characterization of electronic transitions in complex molecules". In: *Discussions of the Faraday society* 9 (1950), pp. 14–19.

[62] George G Guilbault. *Practical fluorescence*. Vol. 3. CRC Press, 1990.

[63] Talley J Lambert. "FPbase: A community-editable fluorescent protein database". In: *Nature methods* 16.4 (2019), p. 277.

[64] Qinggele Li, Isabelle Ledoux-Rak, and Ngoc Diep Lai. "Influence of incident beam polarization on intensity and polarization distributions of tight focusing spot". In: *Advanced Device Materials* 1.1 (2015), pp. 4–10.

[65] Ralf Dorn, Susanne Quabis, and Gerd Leuchs. "The focus of light—linear polarization breaks the rotational symmetry of the focal spot". In: *Journal of modern optics* 50.12 (2003), pp. 1917–1926.

[66] Kai Wicker. "Increasing resolution and light efficiency in fluorescence microscopy". PhD thesis. Kings College London, 2010.

[67] Rainer Heintzmann et al. "Resolution enhancement by subtraction of confocal signals taken at different pinhole sizes". In: *Micron* 34.6-7 (2003), pp. 293–300.

[68] Maria Ingaramo et al. “Richardson–Lucy deconvolution as a general tool for combining images with complementary strengths”. In: *ChemPhysChem* 15.4 (2014), pp. 794–800.

[69] Min Guo et al. “Accelerating iterative deconvolution and multiview fusion by orders of magnitude”. In: *BioRxiv* (2019), p. 647370.

[70] Donald B. Siano and David E. Metzler. “Band Shapes of the Electronic Spectra of Complex Molecules”. In: *The Journal of Chemical Physics* 51.5 (1969), pp. 1856–1861. DOI: `10.1063/1.1672270`. eprint: `https://doi.org/10.1063/1.1672270`. URL: `https://doi.org/10.1063/1.1672270`.

[71] E.K. Rooney and A.G. Lee. “Fitting fluorescence emission spectra of probes bound to biological membranes”. In: *Journal of Biochemical and Biophysical Methods* 12.3 (1986), pp. 175 –189. ISSN: 0165-022X. DOI: `https://doi.org/10.1016/0165-022X(86)90031-X`. URL: `http://www.sciencedirect.com/science/article/pii/0165022X8690031X`.

[72] C. E. Tyner, W. D. Drotning, and H. G. Drickamer. “Asymmetric line shapes of localized optical excitations in condensed systems”. In: *Journal of Applied Physics* 47.3 (1976), pp. 1044–1047. DOI: `10.1063/1.322742`. eprint: `https://doi.org/10.1063/1.322742`. URL: `https://doi.org/10.1063/1.322742`.

[73] Bo Zhang, Josiane Zerubia, and Jean-Christophe Olivo-Marin. “Gaussian approximations of fluorescence microscope point-spread function models”. In: *Appl. Opt.* 46.10 (2007), pp. 1819–1829. DOI: `10.1364/AO.46.001819`. URL: `http://ao.osa.org/abstract.cfm?URI=ao-46-10-1819`.

[74] Aurélie Jost et al. “Optical sectioning and high resolution in single-slice structured illumination microscopy by thick slice blind-SIM reconstruction”. In: *PloS one* 10.7 (2015), e0132174.

[75] Nathan A Hagen and Michael W Kudenov. “Review of snapshot spectral imaging technologies”. In: *Optical Engineering* 52.9 (2013), p. 090901.

[76] Joseph Huff. “The Fast mode for ZEISS LSM 880 with Airyscan: high-speed confocal imaging with super-resolution and improved signal-to-noise ratio”. In: *Nature Methods* 13.11 (2016), p. 958.

[77] *Tadja, Dragoo and Michael W. Davidson. 34-Channel QUASAR Detection Unit. Carl Zeiss AG.* URL: `https://www.zeiss.com/microscopy/us/solutions/reference/all-tutorials/spectral-imaging/34channel-quasar-detection-unit.html`.

[78] Arnaud Royon and Noël Converset. “Quality Control of Fluorescence Imaging Systems: A new tool for performance assessment and monitoring”. In: *Optik & Photonik* 12.2 (2017), pp. 22–25.

[79] David McGloin and Kishan Dholakia. “Bessel beams: diffraction in a new light”. In: *Contemporary Physics* 46.1 (2005), pp. 15–28.

[80] Anindita Dasgupta et al. “Direct Supercritical Angle Localization Microscopy for Nanometer 3D Superresolution”. In: *bioRxiv* (2020).

[81] Anindita Dasgupta et al. “Direct supercritical angle localization microscopy for nanometer 3d superresolution”. In: *Nature Communications* 12.1 (2021), pp. 1–9.

[82] Joran Deschamps, Markus Mund, and Jonas Ries. “3D superresolution microscopy by supercritical angle detection”. In: *Optics Express* 22.23 (2014), pp. 29081–29091.

[83] Manuel Guizar-Sicairos, Samuel T Thurman, and James R Fienup. “Efficient subpixel image registration algorithms”. In: *Optics letters* 33.2 (2008), pp. 156–158.

[84] Rainer Heintzmann and Christoph G Cremer. “Laterally modulated excitation microscopy: improvement of resolution by using a diffraction grating”. In: *Optical Biopsies and Microscopic Techniques III*. Vol. 3568. International Society for Optics and Photonics. 1999, pp. 185–196.

[85] Mats GL Gustafsson. “Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy”. In: *Journal of microscopy* 198.2 (2000), pp. 82–87.

[86] Mats GL Gustafsson et al. “Three-dimensional resolution doubling in wide-field fluorescence microscopy by structured illumination”. In: *Biophysical journal* 94.12 (2008), pp. 4957–4970.

[87] Ernst HK Stelzer. “Light-sheet fluorescence microscopy for quantitative biology”. In: *Nature methods* 12.1 (2015), pp. 23–26.

[88] Emmanuel G Reynaud et al. “Guide to light-sheet microscopy for adventurous biologists”. In: *Nature methods* 12.1 (2014), p. 30.

[89] Bo-Jui Chang, Kevin M Dean, and Reto Fiolka. “Systematic and quantitative comparison of lattice and Gaussian light-sheets”. In: *Optics Express* 28.18 (2020), p. 27052.

[90] Thomas A Planchon et al. “Rapid three-dimensional isotropic imaging of living cells using Bessel beam plane illumination”. In: *Nature methods* 8.5 (2011), p. 417.

[91] Tom Vettenburg et al. “Light-sheet microscopy using an Airy beam”. In: *Nature methods* 11.5 (2014), pp. 541–544.

[92] Bi-Chang Chen et al. “Lattice light-sheet microscopy: imaging molecules to embryos at high spatiotemporal resolution”. In: *Science* 346.6208 (2014).

[93] Philipp J Keller et al. “Reconstruction of zebrafish early embryonic development by scanned light sheet microscopy”. In: *science* 322.5904 (2008), pp. 1065–1069.

[94] Bo-Jui Chang et al. “Universal light-sheet generation with field synthesis”. In: *Nature methods* 16.3 (2019), pp. 235–238.

[95] Bo-Jui Chang and Reto Fiolka. “Light-sheet engineering using the Field Synthesis theorem”. In: *Journal of Physics: Photonics* 2.1 (2019), p. 014001.

[96] Kevin M Dean et al. “Deconvolution-free subcellular imaging with axially swept light sheet microscopy”. In: *Biophysical journal* 108.12 (2015), pp. 2807–2815.

[97] Tonmoy Chakraborty et al. “Light-sheet microscopy of cleared tissues with isotropic, subcellular resolution”. In: *Nature methods* 16.11 (2019), pp. 1109–1113.

[98] Andres Flores Valle and Johannes D Seelig. “Two-photon Bessel beam tomography for fast volume imaging”. In: *Optics express* 27.9 (2019), pp. 12147–12162.

[99] Claus B Müller and Jörg Enderlein. “Image scanning microscopy”. In: *Physical review letters* 104.19 (2010), p. 198101.

[100] CJ R SHEPPARD. “Super-resolution in confocal imaging”. In: *Optik (Stuttgart)* 80.2 (1988), pp. 53–54.

[101] Nandakishore Kambhatla and Todd K Leen. “Dimension reduction by local principal component analysis”. In: *Neural computation* 9.7 (1997), pp. 1493–1516.

[102] Christopher J James and Christian W Hesse. “Independent component analysis for biomedical signals”. In: *Physiological measurement* 26.1 (2004), R15.

[103] Nicolas Gillis. *The Why and How of Nonnegative Matrix Factorization*. 2014. arXiv: `1401.5226 [stat.ML]`.

[104] Daniel Lee and H Sebastian Seung. “Algorithms for non-negative matrix factorization”. In: *Advances in neural information processing systems* 13 (2000).

[105] Junying Zhang et al. “Image fusion based on nonnegative matrix factorization”. In: *2004 International Conference on Image Processing, 2004. ICIP’04.* Vol. 2. IEEE. 2004, pp. 973–976.

[106] Yongxin Zhang et al. “Multi-focus image fusion based on non-negative matrix factorization and difference images”. In: *Signal processing* 105 (2014), pp. 84–97.

[107] Nicolas Gillis and François Glineur. “Accelerated multiplicative updates and hierarchical ALS algorithms for nonnegative matrix factorization”. In: *Neural computation* 24.4 (2012), pp. 1085–1105.

[108] MATLAB. *9.7.0.1190202 (R2019b)*. Natick, Massachusetts: The MathWorks Inc., 2018.

[109] Michael W Berry et al. “Algorithms and applications for approximate nonnegative matrix factorization”. In: *Computational statistics & data analysis* 52.1 (2007), pp. 155–173.

[110] Flavio R Velasco. *Thresholding using the ISODATA clustering algorithm*. Tech. rep. MARYLAND UNIV COLLEGE PARK COMPUTER SCIENCE CENTER, 1979.

[111] TW Ridler, S Calvard, et al. “Picture thresholding using an iterative selection method”. In: *IEEE trans syst Man Cybern* 8.8 (1978), pp. 630–632.

[112] SS Dhillon et al. “The 2017 terahertz science and technology roadmap”. In: *Journal of Physics D: Applied Physics* 50.4 (2017), p. 043001.

[113] Gwyn P Williams. “Filling the THz gap—high power sources and applications”. In: *Reports on Progress in Physics* 69.2 (2005), p. 301.

[114] Michael C Kemp. “A review of millimetre-wave and terahertz technology for detection of concealed threats”. In: *2008 33rd International Conference on Infrared, Millimeter and Terahertz Waves*. IEEE. 2008, pp. 1–2.

[115] A. Agurto et al. “A Review of Concealed Weapon Detection and Research in Perspective”. In: *2007 IEEE International Conference on Networking, Sensing and Control*. 2007, pp. 443–448. DOI: `10.1109/ICNSC.2007.372819`.

[116] H Bello-Salau, AF Salami, and M Hussaini. “Ethical analysis of the full-body scanner (FBS) for airport security”. In: *Advances in Natural and Applied Sciences* 6.5 (2012), pp. 664–672.

[117] Andreas Traut et al. “Körperscanner–Sicherheiten und Unsicherheiten”. In: *forum kriminalprävention*. Vol. 1. 2010, pp. 14–20.

[118] E Heinz et al. “Passive 350 GHz video imaging systems for security applications”. In: *Journal of Infrared, Millimeter, and Terahertz Waves* 36.10 (2015), pp. 879–895.

[119] Richard Knipper et al. “THz absorption in fabric and its impact on body scanning for security application”. In: *IEEE Transactions on Terahertz Science and Technology* 5.6 (2015), pp. 999–1004.

[120] Eugene Hecht. *Optik*. Walter de Gruyter GmbH & Co KG, 2018.

[121] Erik Heinz et al. “Passive submillimeter-wave stand-off video camera for security applications”. In: *Journal of Infrared, Millimeter, and Terahertz Waves* 31.11 (2010), pp. 1355–1369.

[122] Erik Heinz et al. “Toward high-sensitivity and high-resolution submillimeter-wave video imaging”. In: *Optical Engineering* 50.11 (2011), p. 113204.

[123] Kent D Irwin and Gene C Hilton. “Transition-edge sensors”. In: *Cryogenic particle detection*. Springer, 2005, pp. 63–150.

[124] RL Fagaly. “Superconducting quantum interference device instruments and applications”. In: *Review of scientific instruments* 77.10 (2006), p. 101101.

[125] S Ariyoshi et al. “Terahertz imaging with a direct detector based on superconducting tunnel junctions”. In: *Applied physics letters* 88.20 (2006), p. 203503.

[126] Brian Josephson. “Possible new effect in superconducting tunneling”. In: *Phys. Lett.* 1 (1962), pp. 251–253.

[127] John Bardeen, Leon N Cooper, and John Robert Schrieffer. “Theory of superconductivity”. In: *Physical review* 108.5 (1957), p. 1175.

[128] Jae S Lim. “Two-dimensional signal and image processing”. In: *ph* (1990).

[129] Guoshen Yu and Guillermo Sapiro. “DCT image denoising: a simple and effective image denoising algorithm”. In: *Image Processing On Line* 1 (2011), pp. 292–296.

[130] Rafael C Gonzalez, Richard Eugene Woods, and Steven L Eddins. *Digital image processing using MATLAB*. Pearson Education India, 2004.

[131] Timothy J Holmes and Yi-Hwa Liu. “Acceleration of maximum-likelihood image restoration for fluorescence microscopy and other noncoherent imagery”. In: *JOSA A* 8.6 (1991), pp. 893–907.

[132] HM Adorf et al. “Accelerating the Richardson-Lucy restoration algorithm”. In: *European Southern Observatory Conference and Workshop Proceedings*. Vol. 41. 1992.

[133] Edward S Meinel. “Origins of linear and nonlinear recursive restoration algorithms”. In: *JOSA A* 3.6 (1986), pp. 787–799.

[134] Ian Goodfellow et al. *Deep learning*. Vol. 1. 2. MIT press Cambridge, 2016.

[135] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. “Deep learning”. In: *nature* 521.7553 (2015), pp. 436–444.

[136] James Bergstra et al. “Theano: Deep learning on gpus with python”. In: *NIPS 2011, BigLearning Workshop, Granada, Spain*. Vol. 3. Citeseer. 2011, pp. 1–48.

[137] Sergey Zagoruyko and Nikos Komodakis. “Learning to compare image patches via convolutional neural networks”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, pp. 4353–4361.

[138] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. “U-net: Convolutional networks for biomedical image segmentation”. In: *International Conference on Medical image computing and computer-assisted intervention*. Springer. 2015, pp. 234–241.

[139] Alexander Krull, Tim-Oliver Buchholz, and Florian Jug. “Noise2void-learning denoising from single noisy images”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2019, pp. 2129–2137.

[140] Martin Weigert et al. “Content-aware image restoration: pushing the limits of fluorescence microscopy”. In: *Nature methods* 15.12 (2018), pp. 1090–1097.

[141] Michael T McCann, Kyong Hwan Jin, and Michael Unser. “Convolutional neural networks for inverse problems in imaging: A review”. In: *IEEE Signal Processing Magazine* 34.6 (2017), pp. 85–95.

[142] Jaakko Lehtinen et al. “Noise2noise: Learning image restoration without clean data”. In: *arXiv preprint arXiv:1803.04189* (2018).

[143] Alexander Krull, Tomas Vicar, and Florian Jug. “Probabilistic noise2void: Unsupervised content-aware denoising”. In: *arXiv preprint arXiv:1906.00651* (2019).

[144] Johannes Schindelin et al. “Fiji: an open-source platform for biological-image analysis”. In: *Nature methods* 9.7 (2012), pp. 676–682.

[145] Anna S. Goncharova et al. *Improving Blind Spot Denoising for Microscopy*. 2020. arXiv: `2008.08414 [eess.IV]`.

[146] Hirofumi Kobayashi et al. *Image Deconvolution via Noise-Tolerant Self-Supervised Inversion*. 2020. arXiv: `2006.06156 [cs.CV]`.

[147] Lord Rayleigh. “Investigations in optics with special reference to the spectroscopy”. In: *Philos. Mag* 8 (1879), p. 261.