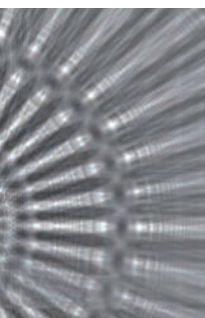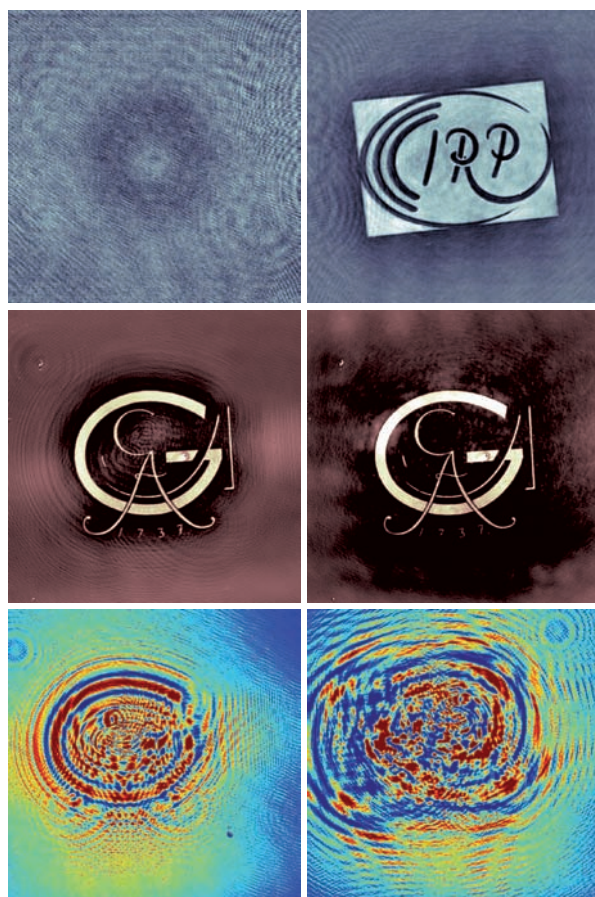Anna–Lena Robisch

Phase retrieval for object and probe in the optical near–field

Universitätsverlag Göttingen

Anna-Lena Robisch

Phase retrieval for object and probe in the optical near-field

Anna-Lena Robisch

Phase retrieval for object and probe in the optical near-field

Address of the Author
Anna-Lena Robisch
e-mail: anna-lena.robisch@phys.uni-goettingen.de

Dissertation
for the award of the degree of
"Doctor rerum naturalium"
of the Georg-August-Universität Göttingen
within the doctoral program PROPHYS
of the Georg-August University School of Science (GAUSS)

Thesis Committee:
Prof. Dr. Tim Salditt
Prof. Dr. D. Russell Luke

Preface of the series editors

The Göttingen series in x-ray physics is intended as a collection of research monographs in x-ray science, carried out at the Institute for X-ray Physics at the Georg-August-Universität in Göttingen, and in the framework of its related research networks and collaborations.

It covers topics ranging from x-ray microscopy, nano-focusing, wave propagation, image reconstruction, tomography, short x-ray pulses to applications of nanoscale x-ray imaging and biomolecular structure analysis.

In most but not all cases, the contributions are based on Ph.D. dissertations. The individual monographs should be enhanced by putting them in the context of related work, often based on a common long term research strategy, and funded by the same research networks. We hope that the series will also help to enhance the visibility of the research carried out here and help others in the field to advance similar projects.

Prof. Dr. Sarah Köster
Prof. Dr. Tim Salditt
Editors
Göttingen June 2014

Preface to the present volume

Unknown object and unknown wave field – asking for the reconstruction of both complex-valued functions from the same experimental recordings seems like asking for too much. Even ptychographic algorithms could previously not provide phase retrieval for extended illumination wave fronts.

Now, the present monograph teaches us how the simultaneous recovery of object and wave front can be accomplished by the combination of suitable transverse and longitudinal translations. An important step to further emancipate the optical scientist from the dictate of ideal wave assumptions …

Prof. Dr. Tim Salditt

Contents

1 Introduction

X-ray microscopy is a non-invasive imaging technique that provides insight into structural details on the nanometer-scale. Since the discovery of X-rays by Wilhelm Conrad Röntgen in 1895, the use of this radiation opened up the opportunity to access details undetectable with visible light. Today, apart from electron microscopy, X-ray structural analysis is the principal tool to resolve materials at atomic resolution. Until the late 20th century, this high resolution imaging technique was basically restricted to crystalline samples which due to their structure amplify the scattered signal. With the development of bright and coherent X-ray sources, nowadays, far-field diffraction patterns of non-crystalline samples can be measured. Furthermore, X-ray imaging is not limited to the optical far-field. Magnified, holograms of the sample can be recorded by using highly curved wave fronts.

An analysis of the far- or near-field intensity distributions is necessary to obtain a sharp and directly accessible image of the object of interest which reveals its fine details in the familiar way of visible light microscopy.

Efficient focusing devices such as lenses exist for light in the spectral range perceivable by the human eye. Lenses provide a magnified image of the sample which is straightforward to interpret. However, X-ray optics are much less performing and usually do not achieve a simple and direct recording of a sharp and highly resolved image. Instead here, the task of lenses is delegated to the software side: Algorithmic tools are necessary to invert the scattered intensity and to reconstruct a complex valued, focused image of the sample.

A common approach is to use iterative optimization tools that minimize the difference between the recorded intensity distribution and the current iterate of the reconstructed structure. The purpose of these algorithms is to refine the reconstruction further and further such that it finally mimics the sample.

However, highly resolved images of the object of interest can only be obtained, if there is an appropriate strategy at hand to separate the illumination from the information of the sample itself. It is intuitively clear that if there is no possibility to distinguish between features of the object and features of the illumination, no reliable image can be reconstructed.

There exist different tools to correct for artifacts resulting from the illumination. They range from straightforward but erroneous division of the recorded hologram by the measured, pure intensity distribution of the X-ray microscope (flat-field correction) to more elaborated techniques which independently or simultaneously (ptychography) reconstruct object and illumination from multiple far-field diffraction patterns of a sample illuminated at different lateral positions.
So far, in the optical near-field the simultaneous reconstruction of object and (unperturbed) illumination has not yet been achieved.

The purpose of this thesis is to establish a method similar to the technique of far-field ptychography which allows simultaneously retrieving object and probe in the optical near-field without the conventional flat-field correction. Primary experimental inspiration is taken from full-field holographic X-ray imaging techniques rather than from methods using far-field diffraction patterns recorded with a confined illumination. Yet, primary inspiration for the algorithm of simultaneous phase retrieval of object and illumination is taken from far-field ptychography. Seen from a near-field imaging point of view, it is clear that lateral scanning of the sample (as used in far-field ptychography) does not provide enough diversity to the data. In contrast, here variations of the distance between sample and detector cause holograms which differ much more than holograms which are only distinguishable by lateral displacement. Therefore, a full data set for near-field ptychography which allows reconstructing object and probe needs to consist of holograms obtained from displacement of the sample in all three dimensions of space. In contrast to far-field ptychography, it is not sufficient to translate the sample only perpendicular to the illumination at a fixed defocus position.

The thesis is organized in four main parts. In chapter two, topics of Fourier-optics are addressed which concern the physical background of X-ray microscopy, coherence, propagation of electromagnetic waves in free space and in matter, as well as resolution and the phase problem in coherent imaging.

Chapter three presents an introduction to the algorithmic tools of near-field propagation and optimization procedures. Selected mathematical aspects of iterative projection algorithms are presented without the claim of completeness. In the remaining of chapter three, the connection to phase retrieval is established. Properties of phase retrieval projection operators are explored. The Gerchberg-Saxton and the error reduction algorithm are briefly introduced. They are both based on the method of alternating projections. Next, more difficult techniques like the (extended) ptychographical engine and the difference map based ptychography are reviewed. With this in mind, the importance of lateral diversity in far-field imaging is explained.

Chapter four expands in depth on the background and development of near-field ptychography with lateral and longitudinal diversity. A detailed explanation of the implementation in parallel and cone beam geometry, as well as a first study based on simulations is presented.

Chapter five shows experimental realizations of near-field ptychography. Experiments with coherent, visible light and hard X-rays of synchrotron sources are presented. Parallel beam geometry as well as a magnifying cone beam microscopy setting was used. A comparison of near-field ptychography with holographic image reconstruction based on the contrast transfer function is provided.

Near-field ptychography is not only a high resolution imaging technique with low noise, but also a method to characterize extended wave fronts.

In chapter six, the thesis is concluded by a summary and an outlook.

2 X-ray microscopy and the phase problem

This chapter is dedicated to give a historic and a conceptual overview of X-ray optics. Without the claim of completeness, the development of coherent X-ray imaging techniques is briefly reviewed. Necessarily, these techniques are strongly influenced by and intertwined with methods designed for visible light optics and electron microscopy. After a historic excursion, selected aspects of coherent optics are addressed. Topics concerning technical and conceptual ingredients of lensless X-ray microscopy such as coherence, propagation in free space and in matter, different imaging regimes, maximum achievable resolution, as well as the phase problem for non-crystalline samples are discussed.

2.1 From Gabor holography to lensless X-ray microscopy

Driven by the motivation to see atoms with electron microscopy, D. Gabor faced the problem that in 1948 the best producible electron lenses could "*be compared in optical perfection to a raindrop rather than to a microscope objective*" [1]. He came up with the ingenious idea to completely leave out the objective lens of the electron microscope and instead to record the (near-field) diffraction pattern of the electron beam [2] on a photographic plate. In his Nobel Lecture (1971), he details:

"After pondering this problem for a long time, a solution suddenly dawned on me, one fine day at Easter 1947 [...]. Why not take a bad electron picture, but one which contains the whole information, and correct it by optical means? It was clear to me for some time that this could be done, if at all, only with coherent electron beams, with electron waves which have a definite phase. But an ordinary photograph loses the phase completely, it records only the intensities. No wonder we lose the phase, if there is nothing to compare it with! Let us see what happens if we add a standard to it, a "coherent background" " [1].

When using a coherent illumination, this blurred image on the photographic plate, which D. Gabor called hologram, is an interference pattern of the primary wave transmitting the sample and the secondary waves scattered by the sample. Hence, it encodes the phase information of the sample. Regions of maximum intensity on the photographic plate correspond to regions where the transmitting and emitting waves had the same phase. An inverse mask of the photographic plate can now be created. If this mask is illuminated by a beam with the same properties of the previously used illumination of the sample, the photons will be diffracted such that they resemble the wave field which once was recorded on the photographic plate:

"The interference of the object wave and of the coherent background or "reference wave" will then produce interference fringes. There will be maxima wherever the phases of the two waves were identical. Let us make a hard positive record, so that it transmits only at the maxima, and illuminate it with the reference source alone. Now the phases are of course right for the reference source A [see Fig. 2.1 taken from [1]], but as at the slits the phases are identical, they must be right also for B

[see Fig. 2.1]; therefore the wave of B must also appear, reconstructed" [1].
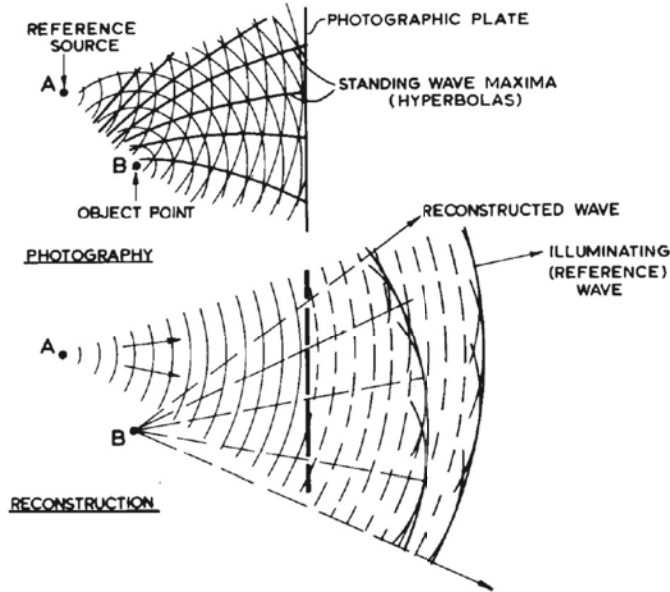
Figure 2.1: Principle of holography. Figure taken from [1].

Since in 1948 lasers as a source of coherent light had not yet been available, D. Gabor used a high pressure mercury lamp combined with a small pinhole in order to have enough intensity and coherence for recording sufficiently structured holograms. As the mercury lamp had a small coherence length of around 0.1 mm, D. Gabor and his assistant I. Williams arranged the whole optical setup on one axis (inline holography); the necessary exposure times were in the order of minutes [1]. Because of these cumbersome experimental conditions, holography found broad application in science only after the development of lasers in 1963 which provide a long coherence length and high flux.

In early electron microscopy appropriate lenses were not disposable. This was the motivation for D. Gabor to invent a lensless microscopy technique. Similar challenges are encountered in modern X-ray optics. The basic idea of D. Gabor works for X-ray microscopy as well: Leave out the lenses in a microscope setup and instead use image processing techniques. Indeed, D. Gabor's work was inspired by W. L. Bragg, who recorded the diffraction patterns of crystals with X-rays and afterwards drilled holes in the photographic plates at the positions of the diffraction spots. An image of the crystal structure can be observed when placing such a mask in an appropriate optical setup [3, 4].

A very prominent and early application of X-rays was in the field of crystallography, where the intensity scattered by crystalline samples is recorded. In 1980,

D. Sayre proposed to adapt soft X-ray diffraction techniques for non-crystalline materials such as single biological cells and their organelles [5, 6]. Hence, resolution would not be limited by the available physical lenses, but the largest angle of diffraction measurable by the detector [7]. The idea of X-ray coherent diffractive imaging (CDI) illustrated in Fig. 2.2(a) was born: An X-ray beam of high flux and coherence hits a non-crystalline sample. The scattered beam is recorded in the optical far-field while the still very bright primary beam is blocked by a beam stop to protect the detector. Since highly coherent X-rays and algorithmic tools to interpret these diffraction data were necessary, it still lasted twenty more years until D. Sayre's hypothesis could be successfully demonstrated and analyzed in experiment by J. Miao. The latter combined X-ray diffraction from a non-crystalline sample with an iterative scheme of data analysis originally developed for electron microscopy [8].
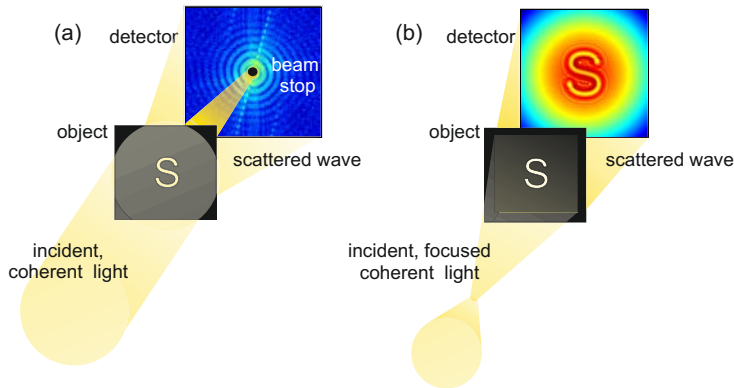
Figure 2.2: (a) Setup for coherent diffractive imaging. The sample S is illuminated by a coherent, parallel X-ray beam. The scattered intensity is recorded in the optical far-field. In order to protect the detector from radiation damage, a beam stop has to be used to block the primary beam. (b) Setup for Fresnel near-field imaging. A coherent, focused X-ray beam produces curved wave fronts impinging on a sample S. The interference between transmitted and scattered beam is recorded in the Fresnel near-field regime. No beam stop is needed.

Despite this success, CDI has two main disadvantages essentially resulting from the same observation [9]. First, due to the low probability for a non-crystalline sample to scatter X-rays under considerably wide angles, the sample has to be exposed to a sufficiently large number of incident photons. This leads to radiation damage of the object of interest. Second, a beam stop is needed to protect the X-ray detector, because most of the photons transmit the sample without being scattered at all. Yet, the regions covered by the beam stop contain important information such as the small angle scattering signal, which encodes among others the overall shape of the sample. This information is lost and has to be acquired by other means, like additional images recorded with a lower number of incident

photons and without a beam stop or by other microscopy techniques (see [8]). In addition, it is possible to construct an estimate of the small angle scattering signal by iterative phase retrieval techniques which are left unconstrained in the regions of missing data [9]. Another drawback of CDI is its insensitivity to translations of the sample in the plane perpendicular to the optical axis. Finally, a high degree of spatial coherence covering at least the dimensions of the sample is required in order to successfully interpret the measured data [9].

While CDI is a technique to record far-field diffraction patterns, different imaging methods for the optical near-field were developed as well. They are much more related to the original idea of holography proposed by D. Gabor. Important examples are the holographic reconstruction algorithms designed by P. Cloetens *et al.* at the end of the 20th century. These algorithms are based on numerical back-propagation of full-field holograms and on the inversion of the contrast transfer function [10,11]. High resolution can be obtained by illuminating the sample with a divergent cone beam (i.e. a beam with considerable phase curvature) [10,12,13]. Thus, it is possible to record a magnified Fresnel hologram (see Fig. 2.2(b)).

In the year 2003, K. A. Nugent *et al.* also proposed to illuminate a non-crystalline specimen by a highly divergent beam [14]. The difference from the just described holographic near-field techniques is that the decay of the illumination is recorded as well. Whereas in the center of the diffraction pattern, a magnified Fresnel hologram can be seen, higher angle scattering of the specimen is imaged at the outer regions of the field of view. This method became known as Fresnel CDI. In the year 2006, it was demonstrated in experiment by G. J. Williams *et al.* [15]. Full field holographic imaging has several advantages compared to CDI. Due to a highly divergent illumination, the whole detector is more or less homogeneously illuminated. No wide angle scattering signal needs to be collected, a fact that considerably reduces the required dose. As a consequence, radiation damage of sample and detector (no beam stop is necessary) is reduced. Furthermore the demand on spatial coherence of the illumination is less stringent [9,16].

2.2   Coherence - a short overview

In the last section it was emphasized that an essential ingredient of coherent imaging techniques is the use of a coherent illumination. Coherence describes the ability of two wave fields to interact (interfere) and thereby – depending on their relative phases – either enhance or diminish their summed intensity. These intensity modulations are denoted as fringes.

A different point of view is to regard coherence as a property of the statistical nature of an optical field $\Psi(\boldsymbol{r}, t)$ – namely as correlations between any two space-time points of $\Psi(\boldsymbol{r}, t)$. This will lead to the definition of the complex degree of coherence $\gamma(\boldsymbol{r}_1, \boldsymbol{r}_2, \tau)$, which is a measure for the strength of the mentioned space-time correlations, and to a generalized law of interference.

Correlations in space can be considered independently of correlations in time. The loss of spatial coherence with increasing separation of two points $\boldsymbol{r}_1 = (x_1, y_1)$ and $\boldsymbol{r}_2 = (x_2, y_2)$ is given by the transversal coherence length. Similarly, the loss of

temporal correlations due to the lack of monochromaticity is given by the coherence time and the longitudinal coherence length.

Because of thermal and quantum-mechanical effects, an X-ray wave field $\Psi(\boldsymbol{r}, t)$ undergoes random fluctuations, both in space and time. Consequently, X-ray fields are stochastic processes. Correlations between any two space-time points $(\boldsymbol{r}_1, t_1)$ and $(\boldsymbol{r}_2, t_2)$ of such a field are described by the mutual coherence function $\Gamma(\boldsymbol{r}_1, \boldsymbol{r}_2, \tau)$ [17]

$$\Gamma(\boldsymbol{r}_1, \boldsymbol{r}_2, \tau) = \langle \Psi(\boldsymbol{r}_1, t+\tau) \Psi^*(\boldsymbol{r}_2, t) \rangle \ , \tag{2.1}$$

where τ is the time difference $t_2 - t_1$. The angular brackets denote the ensemble average. It can be replaced by the time average when considering ergodic systems [9, 17]:

$$\Gamma(\boldsymbol{r}_1, \boldsymbol{r}_2, \tau) = \lim_{T \to \infty} \frac{1}{T} \int_0^T \Psi(\boldsymbol{r}_1, t+\tau) \Psi^*(\boldsymbol{r}_2, t) \, dt \ . \tag{2.2}$$

Similarly, the self-correlation of $\Psi(\boldsymbol{r}, t)$ is defined as the time average over all instances of its squared modulus $|\Psi(\boldsymbol{r}, t)|^2 = \Psi(\boldsymbol{r}, t)\Psi^*(\boldsymbol{r}, t)$

$$\Gamma(\boldsymbol{r}_1, \boldsymbol{r}_1, \tau = 0) = \langle \Psi(\boldsymbol{r}_1, t)\Psi^*(\boldsymbol{r}_1, t) \rangle \ , \tag{2.3}$$
$$\Gamma(\boldsymbol{r}_2, \boldsymbol{r}_2, \tau = 0) = \langle \Psi(\boldsymbol{r}_2, t)\Psi^*(\boldsymbol{r}_2, t) \rangle \ . \tag{2.4}$$

The self-correlations described by equations 2.3 and 2.4 are called intensity or irradiance $I(\boldsymbol{r})$ of the field $\Psi(\boldsymbol{r}, t)$ [9]. Normalization of the mutual coherence function by the self-correlations $\Gamma(\boldsymbol{r}_1, \boldsymbol{r}_1, \tau = 0)$ and $\Gamma(\boldsymbol{r}_2, \boldsymbol{r}_2, \tau = 0)$ results in the definition of the complex degree of coherence [17]

$$\gamma(\boldsymbol{r}_1, \boldsymbol{r}_2, \tau) = \frac{\Gamma(\boldsymbol{r}_1, \boldsymbol{r}_2, \tau)}{\sqrt{\Gamma(\boldsymbol{r}_1, \boldsymbol{r}_1, \tau = 0)\Gamma(\boldsymbol{r}_2, \boldsymbol{r}_2, \tau = 0)}} \ . \tag{2.5}$$

To illustrate the meaning of the complex degree of coherence and how it reflects statistical properties of $\Psi(\boldsymbol{r}, t)$, the intensity of two superimposed wave fields at a point $\boldsymbol{r}_P$ emanating from two small holes of a pierced screen at positions $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ at some distance with respect to $\boldsymbol{r}_P$ shall be calculated. The situation is sketched in Fig. 2.3. The distance between $\boldsymbol{r}_i$ and $\boldsymbol{r}_P$ is denoted as $\Delta_{i,P}$ with $i = 1, 2$. Following the argumentation of the textbooks by D. Paganin [17] and E. Hecht [18], the wave field at $\boldsymbol{r}_P$ at time t can be described as

$$\Psi(\boldsymbol{r}_P, t) = K_1 \Psi \left(\boldsymbol{r}_1, t - \frac{\Delta_{1,P}}{c} \right) + K_2 \Psi \left(\boldsymbol{r}_2, t - \frac{\Delta_{2,P}}{c} \right) \ . \tag{2.6}$$

Due to the finite speed of light, the fields arriving at $\boldsymbol{r}_P$ at time t are retarded, i.e. they result from fields at $\boldsymbol{r}_i$ which had been emitted at a time $t' = t - \frac{\Delta_{i,P}}{c}$ earlier than t. The parameters K_1 and K_2 are propagators, which depend on the size of the pinholes, their relative location with respect to $\boldsymbol{r}_P$ and the phase difference of the secondary waves emanating at $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ with respect to the wave incident on the pierced screen. They are purely imaginary [18] and of equal
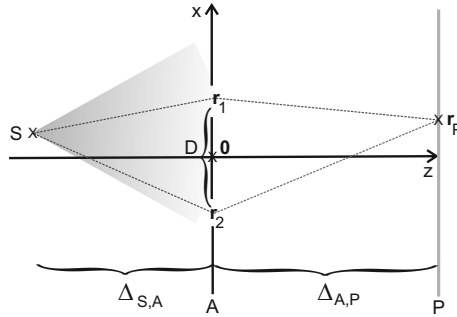
Figure 2.3: Light emitted by a source S is illuminating a pierced screen placed at distance $\Delta_{S,A}$ with respect to S. Behind the screen the interference of the two waves emitting the pinholes at positions $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ is recorded at a distance $\Delta_{A,P}$ with respect to the screen.

sign, such that the product $K_1 K_2^* = K_1^* K_2 = |K_1||K_2|$ is positive and real [18]. Inserting equation 2.2 in the self-correlation function given in equation 2.3 results in the intensity of $\Psi(\boldsymbol{r}_P, t)$

$$
\begin{aligned}
I(\boldsymbol{r}_p) &= \langle \Psi(\boldsymbol{r}_P, t)\, \Psi^*(\boldsymbol{r}_P, t) \rangle \\
&= I_1(\boldsymbol{r}_P) + I_2(\boldsymbol{r}_P) + \\
&\quad 2|K_1||K_2|\mathrm{Re}\left[\left\langle \Psi\left(\boldsymbol{r}_1, t - \frac{\Delta_{1,P}}{c}\right) \Psi^*\left(\boldsymbol{r}_2, t - \frac{\Delta_{2,P}}{c}\right)\right\rangle\right] ,
\end{aligned} \quad (2.7)
$$

where $I_i(\boldsymbol{r}_P)$ are the intensities at $\boldsymbol{r}_P$ resulting from $K_i \Psi(\boldsymbol{r}_i, t - \frac{\Delta_{i,P}}{c})$ in the absence of the pinhole at $\boldsymbol{r}_j$ with $j \neq i$. Shifting the origin of time[1] by

$$
t \to t + \frac{\Delta_{2,P}}{c} \quad (2.8)
$$

and defining the time difference τ

$$
\tau = \frac{\Delta_{2,P}}{c} - \frac{\Delta_{1,P}}{c} , \quad (2.9)
$$

results in

$$
I(\boldsymbol{r}_p) = I_1(\boldsymbol{r}_P) + I_2(\boldsymbol{r}_P) + 2|K_1 K_2^*|\, \mathrm{Re}\left[\Gamma(\boldsymbol{r}_1, \boldsymbol{r}_2, \tau)\right] . \quad (2.10)
$$

With

$$
\begin{aligned}
I_1(\boldsymbol{r}_p) &= |K_1|^2 \left\langle \Psi\left(\boldsymbol{r}_1, t - \frac{\Delta_{1,P}}{c}\right) \Psi^*\left(\boldsymbol{r}_1, t - \frac{\Delta_{1,P}}{c}\right)\right\rangle \\
&= |K_1|^2 \langle \Psi(\boldsymbol{r}_1, t)\, \Psi^*(\boldsymbol{r}_1, t) \rangle \\
&= |K_1|^2\, \Gamma(\boldsymbol{r}_1, \boldsymbol{r}_1, \tau = 0) ,
\end{aligned} \quad (2.11)
$$

---

[1] This is justified, as Ψ is assumed to be statistically stationary [17].

and with

$$I_2(\boldsymbol{r}_p) = |K_2|^2 \, \Gamma(\boldsymbol{r}_2, \boldsymbol{r}_2, \tau = 0) \; , \tag{2.12}$$

we find that

$$|K_1|\,|K_2| = \frac{\sqrt{I_1(\boldsymbol{r}_P)I_2(\boldsymbol{r}_p)}}{\sqrt{\Gamma(\boldsymbol{r}_1, \boldsymbol{r}_1, 0)\Gamma(\boldsymbol{r}_2, \boldsymbol{r}_2, 0)}} \; . \tag{2.13}$$

Hence, with the definition of the complex degree of coherence given in equation 2.5, equation 2.10 becomes

$$I(\boldsymbol{r}_p) = I_1(\boldsymbol{r}_P) + I_2(\boldsymbol{r}_P) + \underbrace{2\sqrt{I_1(\boldsymbol{r}_P)I_2(\boldsymbol{r}_P)}\, \mathrm{Re}\left[\gamma(\boldsymbol{r}_1, \boldsymbol{r}_2, \tau)\right]}_{\text{interference term}} \; . \tag{2.14}$$

This equation is known as the generalized interference law. In the limiting case of a completely incoherent field, correlations between any two space-time points do not exist. Consequently, $\gamma(\boldsymbol{r}_1, \boldsymbol{r}_2, \tau) = 0$ and the intensity at $\boldsymbol{r}_P$ given by equation 2.14 is just the sum of the intensities $I_1(\boldsymbol{r}_P)$ and $I_2(\boldsymbol{r}_P)$. In case that the fields $\Psi(\boldsymbol{r}_i, t)$ emitting the pinholes are perfectly monochromatic with frequency ω, i.e.

$$\Psi(\boldsymbol{r}_i, t) = \sqrt{I(\boldsymbol{r}_i)} \exp\left[i(\varphi(\boldsymbol{r}_i) - \omega t)\right], \tag{2.15}$$

the complex degree of coherence is

$$\gamma(\boldsymbol{r}_1, \boldsymbol{r}_2, \tau) = \exp\left[i(\varphi(\boldsymbol{r}_1) - \varphi(\boldsymbol{r}_2) - \omega\tau)\right]. \tag{2.16}$$

It is of unit modulus and hence the interference term can adopt the maximum and minimum values of $\pm 2\sqrt{I_1(\boldsymbol{r}_P)I_2(\boldsymbol{r}_P)}$. In practice, a completely monochromatic field is very unlikely. Instead of considering strict monochromaticity, it is more realistic to deal with quasi monochromatic fields [17]. These fields have a limited spectral range $\overline{\omega} - \frac{1}{2}\Delta\omega \leq \omega \leq \overline{\omega} + \frac{1}{2}\Delta\omega$ around a mean frequency $\overline{\omega}$. In analogy to the argumentation given later in section 2.3.1, a quasi monochromatic field can be decomposed into its monochromatic components

$$\Psi(\boldsymbol{r}, t) = \frac{1}{\sqrt{2\pi}} \int_{\overline{\omega} - \frac{1}{2}\Delta\omega}^{\overline{\omega} + \frac{1}{2}\Delta\omega} \psi_\omega(\boldsymbol{r}) \exp(-i\omega t)\, d\omega. \tag{2.17}$$

Replacement of ω by $\omega = \overline{\omega} + \delta\omega$ where $\delta\omega$ is the difference of ω with respect to the mean frequency $\overline{\omega}$ leads to

$$\Psi(\boldsymbol{r}, t) = \exp(-i\overline{\omega}t) \underbrace{\frac{1}{\sqrt{2\pi}} \int_{-\frac{1}{2}\Delta\omega}^{+\frac{1}{2}\Delta\omega} \psi_{\overline{\omega}+\delta\omega}(\boldsymbol{r}) \exp(-i(\delta\omega)t)\, d(\delta\omega)}_{:=\mathcal{A}(\boldsymbol{r},t)} \; . \tag{2.18}$$

Equation 2.18 describes a monochromatic field of frequency $\overline{\omega}$ modulated by a complex envelope $\mathcal{A}(\boldsymbol{r}, t)$ that varies with frequencies $\delta\omega$. For the reason that $\delta\omega$ is much smaller than $\overline{\omega}$, the variations caused by the complex envelope are much

slower than those caused by the monochromatic field $\exp(-i\overline{\omega}t)$. The complex degree of coherence is then

$$\gamma(\boldsymbol{r}_1, \boldsymbol{r}_2, \tau) = \exp(-i\overline{\omega}\tau) \frac{\langle \mathcal{A}(\boldsymbol{r}_1, t+\tau)\, \mathcal{A}^*(\boldsymbol{r}_2, t)\rangle}{\sqrt{\langle |\mathcal{A}(\boldsymbol{r}_1, t)|^2\rangle \langle |\mathcal{A}(\boldsymbol{r}_2, t)|^2\rangle}} \ . \tag{2.19}$$

It can be shown that for a quasi monochromatic field the absolute value of the degree of coherence is between zero and one (see for example [17])

$$0 \le |\gamma(\boldsymbol{r}_1, \boldsymbol{r}_2, \tau)| \le 1 \ , \tag{2.20}$$

where the limit of zero correlations corresponds to a completely incoherent field and the limit of full correlations describes a fully coherent field. Intermediate states are attributed to partial coherence.

2.2.1 Temporal coherence

The mutual intensity $\Gamma(\boldsymbol{r}_1, \boldsymbol{r}_2, \tau)$ is an eight dimensional function (six spatial and two temporal coordinates). However, it is easier to consider spatial and temporal correlations independently from each other.

Temporal coherence describes the ability of a beam of light to interfere with a delayed version of itself which has not been shifted in space [19]. In this case the mutual intensity $\Gamma(\boldsymbol{r}_1, \boldsymbol{r}_2, \tau)$ reduces to the self-correlation of the field Ψ:

$$\Gamma(\boldsymbol{r}, \boldsymbol{r}, \tau) = \lim_{T\to\infty} \frac{1}{T} \int_0^T \Psi(\boldsymbol{r}, t+\tau)\Psi^*(\boldsymbol{r}, t)\, dt \ . \tag{2.21}$$

Monochromatic light is temporally coherent. Consider two monochromatic waves where one of them is delayed by the time τ with respect to the other one. These waves will keep a fixed phase relation governed by τ for infinite time. If Ψ is not strictly monochromatic, i.e. it contains frequencies $\overline{\nu} - \frac{1}{2}\Delta\nu \le \nu \le \overline{\nu} + \frac{1}{2}\Delta\nu$ with $\nu = \omega/(2\pi)$ and it is described by equation 2.18, the delayed copy $\Psi(\boldsymbol{r}, t+\tau)$ will loose its correlation with $\Psi(\boldsymbol{r}, t)$ over time. As the complex amplitude $\mathcal{A}(\boldsymbol{r}, t)$ of such a field changes on time scales $\frac{1}{\Delta\nu}$, waves delayed by $\tau_c \approx \frac{1}{\Delta\nu}$ loose their correlation with respect to the non-delayed wave [19]. The time difference τ_c is called coherence time.

The coherence time τ_c is related to the temporal or longitudinal coherence length, which is defined as the (spatial) length after which two waves with slightly different wavelength are in phase opposition [20]. If the first wave at wavelength λ has oscillated N times, the second wave must have oscillated $N - 1/2$ times at wavelength $\lambda + \Delta\lambda$ in order to be in phase opposition with the first wave (see Fig. 2.4(a)). Therefore one finds

$$N\lambda = \left(N - \frac{1}{2}\right)(\lambda + \Delta\lambda) \tag{2.22}$$

$$\Rightarrow \quad N \approx \frac{1}{2}\frac{\lambda}{\Delta\lambda} \ . \tag{2.23}$$
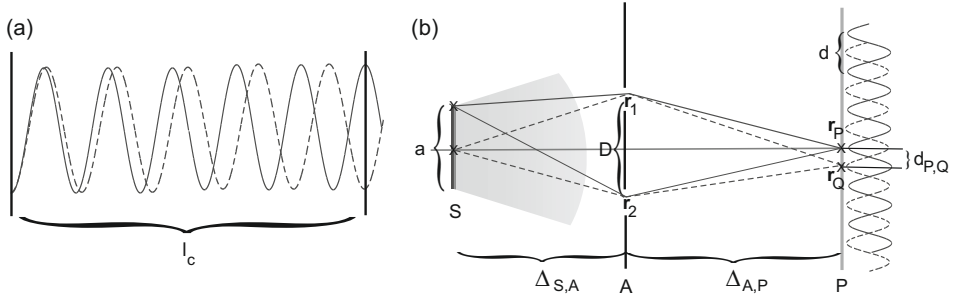
Figure 2.4: (a) Illustration of the temporal coherence length l_c as the the length at which two propagating waves of slightly different frequency are in anti-phase [20]. (b) Illustration of the spatial coherence length as the distance D of two pinholes, for which overlapping fringes resulting from interference patterns caused by the central element and one border element of the source extinguish each other [20]

.

Multiplication of equation 2.23 by the wavelength λ results in the longitudinal or temporal coherence length l_c:

$$l_c \approx \frac{1}{2} \frac{\lambda^2}{\Delta\lambda} \ . \tag{2.24}$$

The exact prefactor of l_c depends on the power spectral density of the source [19]. The longitudinal coherence length can also be deduced by multiplication of τ_c with the wavelength [21] and by using the approximation

$$\frac{\Delta\lambda}{\lambda} \approx \frac{\Delta\nu}{\nu} \ . \tag{2.25}$$

The expression $\frac{\Delta\lambda}{\lambda}$ is called bandwidth. For third generation synchrotrons using the (111) reflection of a Si-crystal monochromator, $\frac{\Delta\lambda}{\lambda} \approx 1.3 \cdot 10^{-4}$ [20]. For $\lambda \approx 0.1$ nm this results in a longitudinal coherence length of

$$l_c \approx 0.4 \ \mu m \ . \tag{2.26}$$

The Wiener-Khintchine theorem
The Wiener-Khintchine theorem connects the spectral composition of a source of light with the temporal coherence. It can be shown that the Fourier-transform of the self-correlation function (r, r, τ) given in equation 2.21 is proportional to the power spectrum of the optical field which monitors to the spectral energy distribution of the light [22]. Hence, a measure of the temporal coherence length can be obtained by the $1/e$-decay of the power spectrum of the source.

2.2.2 Spatial coherence

In contrast to temporal coherence, spatial coherence describes the ability of a wave field to interfere with a spatially shifted, but not delayed copy of itself [19].

To quantify spatial correlations, two quasi monochromatic waves with wavelength λ and time delay $\tau = 0$ are considered. Similar to the longitudinal coherence length l_c one can define a transverse coherence length $l_{x,y}$. It is the distance D between two apertures at which interference fringes from the central element of the source are extinguished by interference fringes resulting from a border element of the source [20]. The situation is sketched in Fig. 2.4(b). Interference fringes of maximal intensity produced by the central part of the source appear at angles $m\frac{\lambda}{D}$ with $m \in \mathbb{Z}$. They are spaced by the distance $d \approx \sin\left(\frac{\lambda}{D}\right)\Delta_{A,P} \approx \frac{\lambda}{D}\Delta_{A,P}$ on the screen at distance $\Delta_{A,P}$. Interference maxima caused by a border element of the source appear at shifted angles $m\frac{\lambda}{D} + \frac{a}{2\Delta_{S,A}}$. Hence, the central maxima of the first and the second interference fringes are spaced by the distance $d_{P,Q} = \frac{a\Delta_{A,P}}{2\Delta_{S,A}}$. If the maxima of the first interference pattern should coincide with minima of the second interference pattern, then

$$d_{P,Q} = \frac{1}{2}d = \frac{1}{2}\frac{\lambda}{D}\Delta_{A,P} = \frac{a\Delta_{A,P}}{2\Delta_{S,A}} \ . \tag{2.27}$$

Solving for the slit separation D results in

$$D = \frac{\lambda\Delta_{S,A}}{a} \ . \tag{2.28}$$

The distance D defined by equation 2.28 is called transversal coherence length l_x. For a two dimensional source, it is

$$l_x = \frac{\lambda\Delta_{S,A}}{a_x} \quad \text{and} \quad l_y = \frac{\lambda\Delta_{S,A}}{a_y} \ , \tag{2.29}$$

where a_x is the extension of the source in the horizontal direction and a_y the extension in the vertical direction. Typically, synchrotron beamlines optimized for coherent imaging have source sizes of 20(h)×200(v) µm^2 [23]. Assuming a source-sample distance of 20 m and a wavelength of 0.1 nm, the coherently illuminated area is

$$l_x \times l_y = 100 \times 10 \quad \text{µm}^2 \ . \tag{2.30}$$

The Van-Citter-Zernike theorem
The Van-Citter-Zernike theorem is the complement of the Wiener-Khintchine theorem. It states that the complex degree of spatial coherence is proportional to the Fourier transform of the intensity distribution of the source of light [22]. Hence, a measure of the transversal coherence length can equivalently be obtained by determining the $1/e$ decay of the Frauenhofer diffraction pattern of the source.

Finally, a rule of thumb for CDI-techniques is that the lateral extent of the sample must fit in the coherently illuminated area and the maximum path length difference of two rays diffracted by the sample must be less than the longitudinal coherence length [20, 23].

2.3 Physics of light propagation

A difference between the two lensless X-ray microscopy techniques (CDI and Fresnel CDI) introduced in the previous section is the optical regime where the scattered radiation is recorded. It is determined by the propagation distance of light. In this section, aspects of the physics of propagation of electromagnetic fields are presented. Among several ways to derive propagation of light, here the angular spectrum method is used. The argumentation basically follows the textbook of D. Paganin [17]. After that, different imaging regimes are explored which describe Fresnel and Frauenhofer diffraction. The Fresnel scaling theorem used for propagation of paraxial, spherical waves is introduced.

2.3.1 Free space propagation

Maxwell's equations for free space give insight into the behavior and properties of the electromagnetic field in vacuum. These equations can be combined to the wave equations for the electric and magnetic fields [17]:

$$\left(\frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \nabla^2\right)\boldsymbol{E}(x,y,z,t) = \boldsymbol{0} \ , \tag{2.31}$$

$$\left(\frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \nabla^2\right)\boldsymbol{B}(x,y,z,t) = \boldsymbol{0} \ , \tag{2.32}$$

where c is the speed of light, $\boldsymbol{E}$ and $\boldsymbol{B}$ are the electric and magnetic field, x, y, z are spatial coordinates and t is time. For hard X-rays, it is justified to use a scalar representation of the vector equations 2.31 and 2.32:

$$\left(\frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \nabla^2\right)\Psi(x,y,z,t) = 0 \ , \tag{2.33}$$

where $\Psi(x,y,z,t)$ represents one component of E_x, E_y, E_z, B_x, B_y, or B_z. In general, $\Psi(x,y,z,t)$ is polychromatic, i.e. it is a superposition of different monochromatic ψ_ω with frequency $\omega = 2\pi\nu$ and with $\nu = \frac{c}{\lambda}$. Each ν represents a single color of the spectrum of Ψ. The field Ψ can be decomposed into monochromatic components $\psi_\omega(x,y,z)$ with the time dependent, oscillating parts $\exp(-i\omega t)$ separated off:

$$\Psi(x,y,z,t) = \frac{1}{\sqrt{2\pi}}\int_0^\infty \underbrace{\psi_\omega(x,y,z)}_{\text{monochromatic components}} \exp(-i\omega t)\ d\omega \ . \tag{2.34}$$

Inserting equation 2.34 into 2.33 leads to the Helmholtz equation describing the evolution of the monochromatic and time independent ψ_ω:

$$\left(\nabla^2 + k^2\right)\psi_\omega(x,y,z) = 0 \ , \tag{2.35}$$

where $k = \frac{\omega}{c} = \frac{2\pi}{\lambda}$. Particularly simple solutions of the Helmholtz equation are plane waves:

$$\psi_\omega^{\text{PW}}(x,y,z) = \exp(i\boldsymbol{k}\boldsymbol{r}) \ , \tag{2.36}$$

with $\boldsymbol{k} = (q_x, q_y, q_z)$ and $|\boldsymbol{k}| = k = \sqrt{q_x^2 + q_y^2 + q_z^2} = \frac{2\pi}{\lambda}$ is the wave vector pointing in the direction of propagation and $\boldsymbol{r} = (x, y, z)$ the vector pointing at (x, y, z). Equation 2.36 can be reformulated into a plane wave at $z = 0$ and an additional complex exponential function denoted as free space propagator:

$$\psi_\omega^{\mathrm{PW}}(x, y, z = \Delta) = \underbrace{\exp\left[i(q_x x + q_y y)\right]}_{=\psi_\omega^{\mathrm{PW}}(x,y,0)} \cdot \underbrace{\exp\left[i\sqrt{k^2 - q_x^2 - q_y^2}\,\Delta\right]}_{\text{free space propagator}}. \qquad (2.37)$$

This equation describes the propagation of a plane wave in free space from $z = 0$ to $z = \Delta$ by multiplication of the plane wave at $z = 0$ with the free space propagator $\exp\left[i\sqrt{k^2 - q_x^2 - q_y^2}\,\Delta\right]$ that includes the propagation distance $z = \Delta$. To propagate arbitrary wave fields, it is necessary to define the Fourier transform $\mathcal{F}$ and its inverse $\mathcal{F}^{-1}$:

$$
\begin{aligned}
\tilde{f}(q_x, q_y) &= \mathcal{F}\left[f(x, y)\right] = \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \exp\left[-i(q_x x + q_y y)\right] dx\, dy\,, \quad (2.38) \\
f(x, y) &= \mathcal{F}^{-1}\left[\tilde{f}(q_x, q_y)\right] = \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \tilde{f}(q_x, q_y) \exp\left[i(q_x x + q_y y)\right] dq_x\, dq_y\,. \quad (2.39)
\end{aligned}
$$

An arbitrary, monochromatic wave field $\psi_\omega(x, y, z = 0)$ in real space is related to the wave field $\tilde{\psi}_\omega(q_x, q_y, q_z = 0)$ in reciprocal space by the inverse Fourier transform $\mathcal{F}^{-1}$:

$$\psi_\omega(x, y, z = 0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \tilde{\psi}_\omega(q_x, q_y, z = 0) \exp\left[i(q_x x + q_y y)\right] dq_x\, dq_y\,. \tag{2.40}$$

This expression states that a wave field $\psi_\omega(x, y, 0)$ can be written as a superposition of plane waves $\exp\left[i(q_x x + q_y y)\right]$ with complex amplitudes $\tilde{\psi}_\omega(q_x, q_y, z = 0)$ which are propagating under different angles with respect to the optical axis. The amplitudes $\tilde{\psi}_\omega(q_x, q_y, z = 0)$ are denoted as angular spectrum. Each of these single plane waves can be propagated by application of the free space propagator

$$
\begin{aligned}
\psi_\omega(x, y, z = \Delta) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \tilde{\psi}_\omega(q_x, q_y, z = 0) \exp\left[i\sqrt{k^2 - q_x^2 - q_y^2}\,\Delta\right] \\
&\quad \exp\left[i(q_x x + q_y y)\right] dq_x\, dq_y\,.
\end{aligned}
\tag{2.41}
$$

Equation 2.41 can be expressed in a compact way by the Fourier transform operator notation

$$\psi_\omega(x, y, z = \Delta) = \mathcal{F}^{-1}\left[\mathcal{F}\left[\psi_\omega(x, y, 0)\right] \cdot \exp\left[i\sqrt{k^2 - q_x^2 - q_y^2}\,\Delta\right]\right]\,. \tag{2.42}$$

In the following, the free space propagator is approximated for waves propagating under small angles with respect to the optical axis (paraxial approximation or Fresnel diffraction). For this purpose, the square root in the exponent of the propagator is expanded into a binomial series up to first order in $\delta = (q_x^2 + q_y^2)$ [17]

$$\sqrt{k^2 - \delta} \approx k - \frac{\delta}{2k} = k - \frac{q_x^2 + q_y^2}{2k} \ . \tag{2.43}$$

Hence, equation 2.42 reads in paraxial approximation as

$$\psi_\omega(x, y, z = \Delta) \approx \exp(ik\Delta)$$
$$\mathcal{F}^{-1}\left[\mathcal{F}\left[\psi_\omega(x, y, z = 0)\right] \exp\left\{\frac{-i\Delta(q_x^2 + q_y^2)}{2k}\right\}\right] \ . \tag{2.44}$$

After separating the constant phase offset $\exp(ik\Delta)$, this is the free space propagation $\mathcal{D}_\Delta$ for Fresnel diffraction or near-field propagation of ψ_ω used throughout this thesis

$$\mathcal{D}_\Delta[\psi_\omega] := \mathcal{F}^{-1}\left[\mathcal{F}[\psi_\omega] \cdot \exp\left\{-\frac{i\Delta}{2k} \cdot (q_x^2 + q_y^2)\right\}\right] \ . \tag{2.45}$$

2.3.2 Fresnel and Frauenhofer regime

The exponential chirp function $\exp\left[\frac{ik}{2\Delta}(x'^2 + y'^2)\right]$ defined in the starting plane of propagation with coordinates (x', y') determines the imaging regime. In case that this chirp can be neglected, the imaging regime is called far-field or Frauenhofer regime, otherwise it is called short distance- or Fresnel near-field regime. To illustrate how this exponential chirp affects propagation, equation 2.44 needs be reformulated into a real space convolution. Following the convolution theorem [17], a real space convolution of two functions $f(x, y)$ and $g(x, y)$ is related to a multiplication in Fourier space[2]:

$$\begin{aligned} f(x, y) \otimes g(x, y) &:= \int_{-\infty}^{\infty} f(\xi) f(x - \xi)\, d\xi \\ &= 2\pi\, \mathcal{F}^{-1}\left[\mathcal{F}\left[f(x, y)\right] \cdot \mathcal{F}\left[g(x, y)\right]\right] \ . \end{aligned} \tag{2.46}$$

We reformulate equation 2.44

$$\begin{aligned} \psi_\omega(x, y, z = \Delta) &\approx \exp(ik\Delta)\, \mathcal{F}^{-1}\left[\mathcal{F}\left[\psi_\omega(x, y, z = 0)\right] \exp\left\{\frac{-i\Delta(q_x^2 + q_y^2)}{2k}\right\}\right] \\ &= 2\pi \mathcal{F}^{-1}\left[\mathcal{F}\left[\psi_\omega(x, y, z = 0)\right] \mathcal{F}\left[h(x, y, \Delta)\right]\right], \end{aligned} \tag{2.47}$$

---

[2] A proof of the convolution theorem can be found in [24].

where the impulse response $h_\Delta(x,y)$ is defined as

$$
\begin{aligned}
h_\Delta(x,y) &:= \mathcal{F}^{-1}\left[H_\Delta(q_x,q_y)\right] \text{ and} \\
H_\Delta(q_x,q_y) &:= \frac{\exp\left(ik\Delta\right)}{2\pi}\exp\left\{\frac{-i\Delta(q_x^2+q_y^2)}{2k}\right\} .
\end{aligned}
\tag{2.48}
$$

The function $H_\Delta(q_x,q_y)$ is called transfer function. With the transfer function, the impulse response $h_\Delta(x,y)$ can be calculated analytically (for an explicit derivation see for example [17]):

$$
h_\Delta(x,y) = -\frac{ik}{2\pi\Delta}\exp\left(ik\Delta\right)\exp\left[\frac{ik(x^2+y^2)}{2\Delta}\right] .
\tag{2.49}
$$

Application of the convolution theorem given in equation 2.46 transforms equation 2.47 into

$$
\begin{aligned}
\psi_\omega(x,y,\Delta) &= \psi_\omega(x,y,0)\otimes h_\Delta(x,y) \\
&= -\frac{ik}{2\pi\Delta}\exp\left(ik\Delta\right)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\psi_\omega(x',y',0)\cdot \\
&\quad \exp\left[\frac{ik[(x-x')^2+(y-y')^2]}{2\Delta}\right]\,dx'dy' \\
&= \frac{k}{i2\pi\Delta}\exp\left(ik\Delta\right)\exp\left[\frac{ik}{2\Delta}(x^2+y^2)\right]\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\psi_\omega(x',y',0)\cdot \\
&\quad \exp\left[\frac{ik}{2\Delta}(x'^2+y'^2)\right]\exp\left[\frac{-ik}{\Delta}(xx'+yy')\right]\,dx'dy'.
\end{aligned}
\tag{2.50}
$$

This is the Fresnel-Kirchhoff diffraction integral. The original field $\psi_\omega(x',y',0)$ is decorated with the chirp $\exp\left[\frac{ik}{2\Delta}(x'^2+y'^2)\right]$, which contains the propagation distance Δ. After that, this decorated field is Fourier transformed. The result is multiplied by a complex prefactor, that contains an exponential function with a quadratic phase and depends on the propagation distance.
It is useful to define the Fresnel number F

$$
F := \frac{a^2}{\lambda\Delta} ,
\tag{2.51}
$$

where a is the diameter of the region within which $\psi_\omega(x',y',0)$ is not negligible. For $F \ll 1$, i.e. propagation distances much larger than the wavelength λ, such that $\lambda\cdot\Delta \gg a^2$, the Fresnel-Kirchhoff integral (equation 2.50) can be approximated by

$$
\begin{aligned}
\psi_\omega(x,y,\Delta) &= \frac{k}{i2\pi\Delta}\exp\left(ik\Delta\right)\exp\left[\frac{ik}{2\Delta}(x^2+y^2)\right] \\
&\quad \cdot \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\psi_\omega(x',y',0)\exp\left[\frac{-ik}{\Delta}(xx'+yy\prime)\right]\,dx'dy' \\
&= \frac{k}{i2\pi\Delta}\exp\left(ik\Delta\right)\exp\left[\frac{ik}{2\Delta}(x^2+y^2)\right]\mathcal{F}\left[\psi_\omega(x',y',0)\right].
\end{aligned}
\tag{2.52}
$$

Equation 2.52 is the Frauenhofer diffraction integral for far-field propagation. For large distances the optical field in the destination plane is the Fourier transform of the field in the starting plane multiplied by a complex amplitude and a paraxial, spherical wave $\exp{(ik\Delta)}\exp{\left[\frac{ik}{2\Delta}(x^2+y^2)\right]}$.

2.3.3 The Fresnel scaling theorem
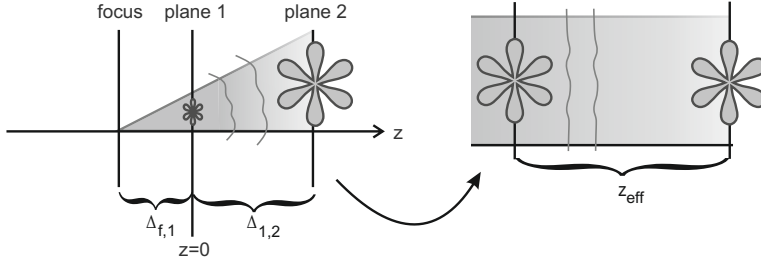
Figure 2.5: The Fresnel scaling theorem allows propagating a cone beam like a plane wave when considering an effective coordinate system.

As introduced the section 2.3.1, the Fresnel-Kirchhoff integral describes free space propagation of an arbitrary wave field $\psi(x',y',0)$ over a distance $\Delta_{1,2}$, resulting in the wave field $\psi(x,y,\Delta_{1,2})$ (for the notation of distances see Fig. 2.5). In order to keep the notation simple, the index ω used to indicate monochromatic waves is skipped, but monochromaticity is assumed in the following as well. The complex prefactor of the Fresnel-Kirchhoff integral without the quadratic phase factor is

$$\mathcal{A}(\Delta_{1,2}) = \frac{k}{i2\pi\Delta_{1,2}} \exp{(ik\Delta_{1,2})}. \tag{2.53}$$

Recall the Fresnel-Kirchhoff integral:

$$
\begin{aligned}
\psi(x,y,\Delta_{1,2}) \;=\;\; & \mathcal{A}(\Delta_{1,2})\exp\left[\frac{ik}{2\Delta_{1,2}}(x^2+y^2)\right] \\
& \int_{\infty}^{\infty}\int_{\infty}^{\infty} \psi(x',y',0)\cdot\exp\left[\frac{ik}{2\Delta_{1,2}}\cdot\ x'^2+y'^2)\right] \\
& \cdot\ \exp\left[\frac{ik}{\Delta_{1,2}}\cdot(x'x+y'y)\right]dx'dy'.
\end{aligned}
\tag{2.54}
$$

If $\psi(x',y',\Delta_{f,1})$ is a plane wave at distance $\Delta_{f,1}$ and with $\boldsymbol{k}$ parallel to the optical axis, it can be expressed as

$$\psi^{\mathrm{PW}}(x',y',\Delta_{f,1}) = \exp{(ik\Delta_{f,1})}, \tag{2.55}$$

whereas a spherical wave $\psi^{\mathrm{SW}} = \exp(ikr)$ in paraxial approximation

$$r = \sqrt{\Delta_{f,1}^2 + x^2 + y^2} \approx \Delta_{f,1} + \frac{1}{2\Delta_{f,1}}(x^2+y^2) \tag{2.56}$$

with defocus distance $\Delta_{f,1}$ is

$$
\begin{aligned}
\psi^{\mathrm{SW}}(x', y', 0) &\approx \exp\left(ik\sqrt{\Delta_{f,1}^2 + x'^2 + y'^2}\right) \\
&\approx \underbrace{\exp\left(ik\Delta_{f,1}\right)}_{=\psi^{\mathrm{PW}}} \exp\left[\frac{ik}{2\Delta_{f,1}}\left(x'^2 + y'^2\right)\right] \; .
\end{aligned}
\tag{2.57}
$$

Inserting the point source illumination into the Fresnel-Kirchhoff-integral leads to

$$
\begin{aligned}
\psi^{\mathrm{SW}}(x, y, \Delta_{1,2}) &= \mathcal{A}(\Delta_{1,2}) \exp\left[\frac{ik}{2\Delta_{1,2}}(x^2 + y^2)\right] \\
&\cdot \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \psi^{\mathrm{PW}}(x', y', 0)\exp\left[\frac{ik}{2\Delta_{f,1}}\left(x'^2 + y'^2\right)\right] \\
&\cdot \exp\left[\frac{ik}{2\Delta_{1,2}} \cdot \left(x'^2 + y'^2\right)\right] \exp\left[\frac{-ik}{\Delta_{1,2}} \cdot (x'x + y'y)\right] \; dx'dy'.
\end{aligned}
\tag{2.58}
$$

With the help of Fig. 2.5 the geometrical magnification of the image of an object with respect to its original size can be determined to:

$$
M = \frac{\Delta_{f,1} + \Delta_{1,2}}{\Delta_{f,1}} \; .
\tag{2.59}
$$

With equation (2.59), the amplitude of $\psi^{\mathrm{SW}}(x, y, \Delta_{1,2})$ is

$$
\begin{aligned}
\left|\psi^{\mathrm{SW}}(x, y, \Delta_{1,2})\right| &= \frac{|\mathcal{A}(z_{\mathrm{eff}})|}{M} \Big| \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \psi^{\mathrm{PW}}(x', y', 0)\exp\left[\frac{ik}{2z_{\mathrm{eff}}} \cdot \left(x'^2 + y'^2\right)\right] \\
&\cdot \exp\left[\frac{-ik}{z_{\mathrm{eff}}} \cdot \left(x' \cdot \frac{x}{M} + y' \cdot \frac{y}{M}\right)\right] \; dx'dy' \Big|,
\end{aligned}
\tag{2.60}
$$

where $z_{\mathrm{eff}} = \frac{\Delta_{1,2}}{M}$. For comparision, the amplitude of a plane wave propagated over a distance z_{eff} and calculated at coordinates $(x/M, y/M)$ is

$$
\begin{aligned}
\left|\psi^{\mathrm{PW}}(x/M, y/M, z_{\mathrm{eff}})\right| &= |\mathcal{A}(z_{\mathrm{eff}})| \Big| \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \psi^{\mathrm{PW}}(x', y', 0)\exp\left[\frac{ik}{2z_{\mathrm{eff}}} \cdot \left(x'^2 + y'^2\right)\right] \\
&\cdot \exp\left[\frac{-ik}{z_{\mathrm{eff}}} \cdot \left(x' \cdot \frac{x}{M} + y' \cdot \frac{y}{M}\right)\right] \; dx'dy' \Big|,
\end{aligned}
\tag{2.61}
$$

Hence, one obtains [17]

$$
\left|\psi^{\mathrm{SW}}(x, y, \Delta_{1,2})\right| = \frac{1}{M} \left|\psi^{\mathrm{PW}}(x/M, y/M, z_{\mathrm{eff}})\right| \; .
\tag{2.62}
$$

Reconstruction and propagation techniques used for a parallel beam geometry can directly be applied to intensities recorded in a magnifing cone beam setting. But it is necessary to introduce an effective coordinate system.

2.3.4 Interaction between light and matter: The projection approximation

Propagation of electromagnetic waves in linear, isotropic, static, non magnetic materials with no current and no charge densities and an electron density that is slowly varying compared to the wavelength of the electromagnetic radiation is described by the scalar wave equation in materials [17]:

$$\left[\epsilon(x,y,z)\mu_0\frac{\partial^2}{\partial t^2} - \nabla^2\right]\Psi(x,y,z,t) = 0 \; , \tag{2.63}$$

where ϵ is the electric permittivity and μ_0 is the magnetic permeability in vacuum. Since the assumptions concerning the underlying materials are essential to derive equation 2.63, they shall be explained in the following.

Linear materials respond in a linear way to electromagnetic fields, i.e. the electric displacement is $\boldsymbol{D} = \epsilon\boldsymbol{E}$ with the electric permittivity ϵ, similarly the magnetic induction is $\boldsymbol{B} = \mu\boldsymbol{H}$ with the magnetic permeability μ. Assuming linear response to electromagnetic fields excludes all ferroelectric and ferromagnetic materials, because they depend on the history of the material. In addition, all effects of non-linear optics are excluded. Isotropic materials respond in the same way to electromagnetic fields no matter from which direction they are exposed to the fields; ϵ and μ are scalars. The assumption of isotropy excludes materials such as non-cubic crystals. For static materials, the electric permittivity and the magnetic permeability are independent of time. Non magnetic materials have zero magnetic permeability $\mu = 0$. Materials with no current and charge densities are neutral and without internal electric currents. Slowly varying materials are described by a permeability and a permittivity which are constant over length scales of the wavelength.

Equation 2.63 has a similar structure as equation 2.33 for electromagnetic waves in free space. Both equations differ in the prefactor of the second order time derivative that contains the speed of light which is either $c = \frac{1}{\sqrt{\epsilon_0\mu_0}}$ in vacuum with permeability μ_0 and vacuum permittivity ϵ_0 or $c' = \frac{1}{\sqrt{\epsilon\mu}}$ in material.

Similar to the derivation of free space propagation, the next step is to decompose $\Psi(x,y,z,t)$ into its monochromatic, time independent components $\psi_\omega(x,y,z)$ (see equation 2.34) and insert this decomposition into equation 2.63. Using the definition of the frequency dependent refractive index n_ω

$$n_\omega = c\sqrt{\epsilon_\omega(x,y,z)\mu_0} = \sqrt{\frac{\epsilon_\omega(x,y,z)}{\epsilon_0}} \; , \tag{2.64}$$

with frequency dependent permittivity $\epsilon_\omega(x,y,z)$, this results in the inhomogeneous Helmholtz equation [17]

$$\left(\nabla^2 + n_\omega^2(x,y,z)k^2\right)\psi_\omega(x,y,z) = 0 \; . \tag{2.65}$$

Equation 2.65 can be further simplified for a beam like wave, i.e. a wave that propagates essentially in the direction of z. The main variations of beam like

waves are perpendicular to the optical axis, whereas variations in the direction of propagation can be neglected. While transmitting a material, this wave is only scattered under small angles with respect to the direction z and hence can be regarded as a perturbed plane wave with envelope $\psi'_\omega(x,y,z)$ [17]

$$\psi_\omega(x,y,z) = \psi'_\omega(x,y,z)\exp(ikz) \ . \tag{2.66}$$

Inserting equation 2.66 into 2.65 and neglecting second order derivatives with respect to z leads to a wave equation for $\psi'_\omega(x,y,z)$

$$\left(2ik\frac{\partial}{\partial z} + \nabla^2_\perp + k^2(n^2_\omega(x,y,z)-1)\right)\psi'_\omega(x,y,z) = 0 \ , \tag{2.67}$$

where

$$\nabla^2_\perp := \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \ . \tag{2.68}$$

Equation 2.67 is known as the inhomogeneous, paraxial or parabolic wave equation. Furthermore, it is assumed, that the perturbation $\psi'_\omega(x,y,z)$ is weak enough that light scattered along neighboring trajectories does not interact. Therefore, the second order derivative $\nabla^2_\perp = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$ of $\psi'_\omega(x,y,z)$ can be set to zero as well. The inhomogeneous, paraxial equation becomes

$$\frac{\partial}{\partial z}\psi'_\omega(x,y,z) \approx \frac{ik}{2}(1-n^2_\omega(x,y,z))\psi'_\omega(x,y,z) \ , \tag{2.69}$$

with solutions

$$\psi'_\omega(x,y,z_1) = \exp\left[\frac{k}{2i}\int_{z_0}^{z_1}(n^2_\omega(x,y,z)-1)\,dz\right]\psi'_\omega(x,y,z_0) \ . \tag{2.70}$$

In the X-ray regime, the refractive index n_ω is close to unity and often written as [17]

$$n_\omega(x,y,z) = 1 - \delta_\omega(x,y,z) + i\beta_\omega(x,y,z) \ . \tag{2.71}$$

The parameter δ_ω is proportional to the number density of electrons $\rho(x,y,z) \approx 1/\text{Å}^3$, the scattering amplitude per electron $r_0 = 2.82\cdot10^{-5}$ Å and the quadratic wavelength λ. Thus, it describes the scattering properties of the medium and is [25]

$$\delta_\omega = \frac{\lambda^2}{2\pi}r_0\rho \ . \tag{2.72}$$

With a wavelength of $\lambda = 1$Å, δ_ω is of the order of 10^{-6} to 10^{-5}. The parameter β_ω describes the absorption properties of the medium and in general is about one order of magnitude lower than δ_ω. Hence, both δ_ω and β_ω are small compared to one, such that in the expression $1-n^2_\omega(x,y,z)$ second order components in δ_ω and β_ω can be set to zero

$$1-n^2_\omega(x,y,z) \approx 2(\delta_\omega - i\beta_\omega) \ . \tag{2.73}$$

Inserting this approximation in equation 2.70 leads to

$$\psi'_\omega(x,y,z_1) \approx \exp\left[-ik \int_{z_0}^{z_1} \delta_\omega(x,y,z)dz\right] \exp\left[-k \int_{z_0}^{z_1} \beta_\omega(x,y,z)dz\right] \psi'_\omega(x,y,z_0) .$$
(2.74)

The first exponential function is the accumulated phase shift along the streamlines of propagation through a material. For a homogeneous material δ_ω does not depend on z, nor does it depend on x or y, such that the overall phase shift $\Delta\varphi$ of a material with thickness $\Delta z = z_1 - z_0$ is

$$\Delta\varphi = -k\delta_\omega\Delta z .$$
(2.75)

The second exponential function with $\beta_\omega > 0$ describes the absorption accumulated along the beam path. Both parts together are denoted as the object's complex valued transmission function

$$o(x,y) = \exp\left[-ik \int_{z_0}^{z_1} \left(\delta_\omega(x,y,z) - i\beta_\omega(x,y,z)\right)\, dz\right] ,$$
(2.76)

that modulates the incoming wave $\psi'_\omega(x,y,z_0)$.
Equation 2.74 states that the wave field $\psi'_\omega(x,y,z_1)$ can be approximated by a product of the object's transmission function and the probe $\psi'_\omega(x,y,z_0)$:

$$\psi'_\omega(x,y,z_1) = o(x,y) \cdot \psi'_\omega(x,y,z_0) .$$
(2.77)

This approximation is called projection approximation. The perturbed wave field directly behind the object is denoted as exit surface wave or exit wave. The projection approximation can be used for an object of thickness

$$\Delta z < \frac{2(\delta x^2 + \delta y^2)}{\lambda} ,$$
(2.78)

where $\sqrt{\delta x^2 + \delta y^2}$ is the smallest detail of the sample that should be resolved (see e.g. [26]) and Δz is the thickness of the object along the optical axis. It illustrates that within an interval Δz neighboring rays of light do not interact.

2.4 Resolution

The purpose of microscopy is to provide insight into small structures otherwise invisible for the human eye. The smaller these visualized features are, the better is the resolution of the microscope. Here, concepts of resolution are explained. The Rayleigh criterion is introduced. The influence of the source size on the achievable resolution will be discussed. After this theoretical introduction, two practical methods to determine the resolution of an image are presented.

2.4.1 Rayleigh resolution, resolution of a microscope and resolution in CDI

There are two ways arrive at Frauenhofer diffraction [22]. (1) It is the Fourier transform relationship between two field distributions in planes separated by a long distance Δ. (2) It is the diffraction in the plane where an optical system forms the image of the source of light which has passed through its components. Since these components are lenses with apertures of finite extent, the width of the beam transmitting the system is restricted. This causes a spreading of the image in the observation plane.
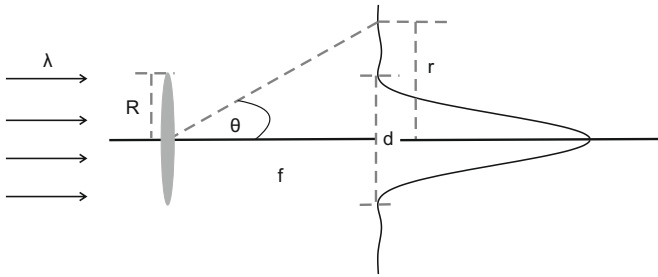
Figure 2.6: A lens focusses light to spot which is spread by diffraction caused by the finite aperture of the lens.

In case that the restricting aperture is a lens of circular geometry with radius R and focal length f, an image of the source is observed in the focal plane. It is described by a squared Bessel function J_1 [27] (Airy pattern) which depends on the angular coordinate θ and is symmetrical about the axis of symmetry of the aperture (see Fig. 2.6)

$$I(\theta) = I_0 \cdot \left(2 \frac{J_1 \left(\frac{2\pi R}{\lambda} \sin(\theta) \right)}{\frac{2\pi R}{\lambda} \sin(\theta)} \right)^2 . \tag{2.79}$$

With the approximation that for small angles θ

$$\sin(\theta) \approx \frac{r}{f} , \tag{2.80}$$

where r is a radial coordinate in the focal plane, one finds for the width d of the central lobe [24]

$$d = 1.22 \frac{\lambda f}{R} . \tag{2.81}$$

Light originating from a point source and passing through a lens is not focused to a sharp point again, but to a spread pattern of radial symmetry (see Fig. 2.6). Hence, Frauenhofer diffraction describes the spreading of light around an image (here the image of the source). It can be used to define a measure for the resolving power of

an optical system. By convenience, two points are denoted as just resolvable, when the center of the Airy pattern surrounding the image of the first point source falls exactly on the first zero of the Airy pattern surrounding the image of the second point source [24]. This criterion is known as Rayleigh resolution[3]. The minimum resolvable distance is

$$\delta = 0.5 \cdot d = 0.61 \frac{\lambda f}{R} \ . \tag{2.82}$$

The resolving power of a microscope is also limited by diffraction. In an ideal system described by ray optics, light that has been emitted/diffracted by an object of size G placed at a distance g with respect to a lens L and has passed through this lens forms an image at a distance b behind L. This image is not disturbed by aberrations and is a scaled, inverted copy with size B of the original object. In a real imaging system, diffraction leads to a convolution of the ideal magnified or demagnified image with the Frauenhofer diffraction pattern of the lens pupil [24]. The result is a smoothed image showing less fine details than actually inherent in the object.
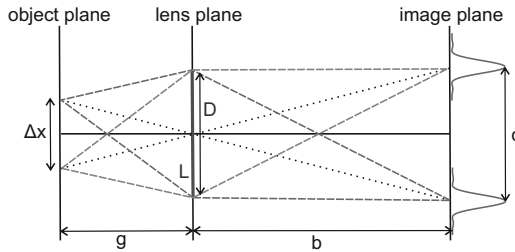
Figure 2.7: Sketch to determine the minimum resolvable distance of a microscope, figure adapted from [27].

With this in mind, one can see from Fig. 2.7 that the distance d between two image points which are spread by diffraction has to be

$$d \geq 1.22 \frac{\lambda b}{D} \ , \tag{2.83}$$

where D is the diameter of the lens. With a lens, a (de-)magnified image of size B of an object of size G is formed following the relation

$$\left| \frac{B}{G} \right| = \frac{b}{g} \ , \tag{2.84}$$

where g is the distance between object and lens and b is the distance between image and lens. Therefore, the minimum resolvable distance Δx_{min} between two object points is

$$\Delta x_{\text{min}} = d_{\text{min}} \cdot \frac{g}{b} = 1.22 \frac{\lambda b}{D} \cdot \frac{g}{b} = 1.22 \frac{\lambda g}{D} \ . \tag{2.85}$$

---

[3] Note that strictly speaking the Rayleigh criterion is only true for incoherent light in paraxial approximation. For coherent light, interference between the diffracted light hast to be taken into account as well.

When the object is placed in the focal plane of the lens, i.e. $g \approx f$, equation 2.85 results in

$$\Delta x_{\min} = 1.22 \frac{\lambda f}{D} \ . \tag{2.86}$$

The fraction $\frac{D}{f}$ is twice the sine of the the angular semi-aperture $\theta_{\max}$ under which light can be 'collected' by the lens [27]. Together with the refractive index of the medium between object and lens, one can define the numerical aperture NA of the system (see Fig. 2.8(a))

$$\mathrm{NA} = n \sin(\theta_{\max}) \ . \tag{2.87}$$

It is a dimensionless number that characterizes the the angular acceptance of the setup, i.e. its 'light gathering power' [28]. With this definition we find for the resolution of a microscope

$$\Delta x_{\min} = 0.61 \frac{\lambda}{\mathrm{NA}} \ . \tag{2.88}$$

This states that the used lenses determine the numerical aperture and hence together with the wavelength the maximum possible resolution.
Resolution can be increased by decreasing the wavelength and/or by increasing the numerical aperture.
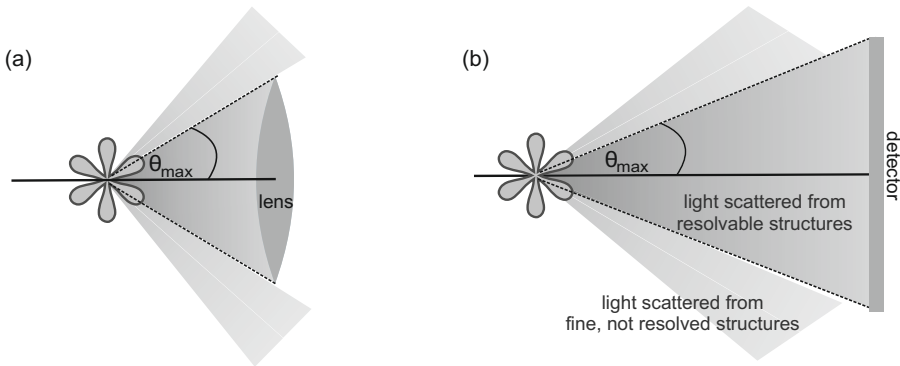
Figure 2.8: (a) Numerical aperture $\sin(\theta_{\max})$ for lens-based systems. (b) Numerical aperture $\sin(\theta_{\max})$ for (Fresnel-) CDI. Figure adapted from [7].

In coherent diffractive imaging – lacking lenses – the maximum detectable angle under which radiation is scattered, is defined by the lateral extent of the detecting device [7]. Therefore, the numerical aperture is determined by the field of view of the detector (see Fig. 2.8(b)). To arrive at a high numerical aperture, either the diameter of the lenses (for visible-light optics) or the diameter of the detector (for CDI) has to be increased.
X-ray wavelengths are between 0.1 nm and 10 nm, whereas visible light is found at wavelengths between 380 nm and 640 nm. With a detector covering a field of

view of approximately 10 cm in diameter and placed 5 m apart from the source of light, using X-rays, structural details of

$$\delta \approx 0.61 \frac{1 \cdot 10^{-10} \text{ m} \cdot 5 \text{ m}}{5 \cdot 10^{-2} \text{ m}} \approx 6 \cdot 10^{-9} \text{ m} \qquad (2.89)$$

can be resolved - a distance which is close to atomic distances of $1 \cdot 10^{-10}$ m.

2.4.2 Resolution in Fresnel CDI

The resolution in Fresnel coherent diffractive imaging is mainly limited by the numerical aperture and the pixel size of the detector, as well as by the source size and the wavelength.

In section 2.3.3, it was shown that the propagation of a cone beam can be performed in the same way as the propagation of a parallel beam by introducing an effective coordinate system. An object placed at distance $\Delta_{f,1}$ in the defocus of a point source causes a magnified hologram at distance $\Delta_{1,2}$. The corresponding magnification is

$$M = \frac{\Delta_{f,1} + \Delta_{1,2}}{\Delta_{f,1}} \ , \qquad (2.90)$$

which depends on both, the focus-to-sample distance $\Delta_{f,1}$, as well as the sample-to-detector distance $\Delta_{1,2}$. The closer the sample is moved towards the focus, the larger is the magnification. In contrast, shifting the sample towards the detector reduces M (see Fig. 2.9(a)).
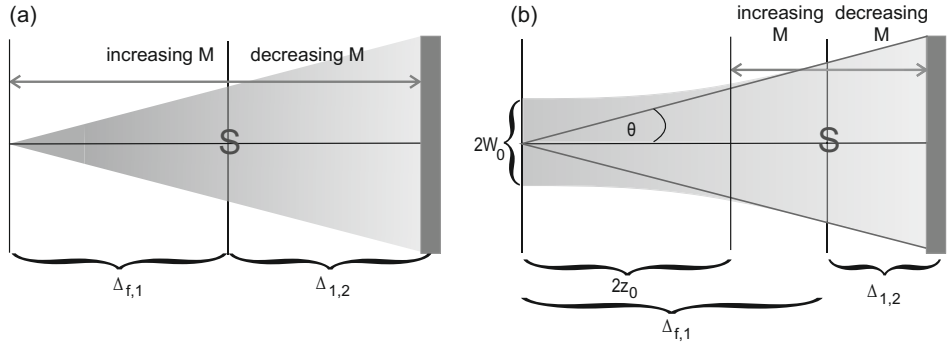
Figure 2.9: (a) Cone beam holography allows to record magnified holograms compared to parallel beam holography. The magnification depends on the relative positions between focus and sample and between sample and detector. Shifting the sample towards the source increases the magnification of the hologram recorded at a distance $\Delta_{f,1} + \Delta_{1,2}$ with respect to the source. (b) An extended source size limits the maximum achievable magnification, because the sample has to be positioned at distances $\Delta_{f,1}$ longer than twice the Rayleigh length z_0.

Once the sample is sufficiently close to the detector, i.e. the magnification M is

approximately one, the resolution essentially depends on the pixel size of the detector. Hence, when shifting the sample closer and closer towards the focus, the resolution is basically limited by the wavelength and the extent of the detector. Yet, this is only true for beams focused down to very small spot sizes. A converging/diverging, paraxial beam can be modeled more realistically by a Gaussian beam with focal spot size W_0 and wavelength λ. The complex amplitude of a Gaussian beam can be described by [21]

$$\psi^{\text{GB}}(x, y, z) = A_0 \frac{W_0}{W(z)} \exp\left[-\frac{x^2 + y^2}{W^2(z)}\right] \exp\left[-ikz - ik\frac{x^2 + y^2}{2R(z)} + i\zeta(z)\right] ,$$

(2.91)

where A_0 is a complex constant, $W(z)$ describes the width of the Gaussian beam. The beam width $W(z)$ depends on the focal spot of diameter $2W_0$

$$W(z) = W_0\sqrt{1 + \left(\frac{z}{z_0}\right)^2} .$$

(2.92)

The curvature of the Gaussian beam is related to the propagation distance z and is described by $R(z)$

$$R(z) = z\left[1 + \left(\frac{z_0}{z}\right)^2\right] ,$$

(2.93)

where z_0 is the Rayleigh length of the beam

$$z_0 = \frac{\pi W_0^2}{\lambda} .$$

(2.94)

Finally $\zeta(z)$ is called Gouy phase and is

$$\zeta(z) = \tan^{-1}\left(\frac{z}{z_0}\right) .$$

(2.95)

The parameter of interest is the radius of curvature $R(z)$. It describes how well the phase fronts of a Gaussian beam match those of a spherical wave in paraxial approximation

$$\psi^{\text{SW}}(x, y, z) = \exp\left(-ikz\right)\exp\left[-\frac{ik}{2z}\left(x^2 + y^2\right)\right] .$$

(2.96)

For spherical waves, the radius of curvature is the distance of propagation z. For Gaussian beams at large propagation distances the factor $\left(\frac{z_0}{z}\right)^2$ can be neglected and $R(z) \to z$. The radius of curvature $R(z)$ is minimal for $z = z_0$, this corresponds to a spherical wave at distance $R(z) = 2z_0$. For distances smaller than z_0 the radius of wave front curvature rises again. Hence, a Gaussian beam can be approximated by a spherical wave at propagation distances larger than twice the Rayleigh length z_0. For a fixed focus-to-detector distance this limits the achievable magnification M: In order to record magnified holograms, the sample has to be placed at a

distances $\Delta_{f,1}$ larger than $R(z_0) = 2z_0$ where the wave fronts of a Gaussian beam can be approximated by those of a paraxial spherical wave (see Fig. 2.9(b)). Since the Rayleigh length $z_0 = \frac{\pi W_0^2}{\lambda}$ scales with the squared radius of the focal spot size, we see that for a fixed focus-to-detector distance $\Delta_{f,2}$ the maximum achievable magnification depends on the source size W_0 and is

$$M_{\max}(W_0) = \frac{\Delta_{f,2}\lambda}{2\pi W_0^2} \; . \tag{2.97}$$

Note that even though the magnification overcomes the strict dependence on the pixel size of the detector, it does not overcome the limitations imposed by the numerical aperture of the detector.

The analysis above shows that for high resolution imaging small spot sizes, i.e. focused beams are desirable. One can use Fresnel zone plates, compound refractive lenses or mirrors for focusing an X-ray beam to a small spot size. All of these optical elements are never aberration free. Focusing induces further challenges. X-ray mirrors are commonly used focusing devices at synchrotrons, i.e. at the GINIX-setup (Göttingen Instrument for Nano-Imaging with X-Rays) at DESY (Deutsches Elektronen Synchrotron) and at imaging beamlines of ESRF (European Synchrotron Radiation Facility). Surface roughness of these mirrors, which is of the order of nanometers, causes strong defocus interference fringes [29]. An example is depicted in Fig. 2.10(a). It shows the defocus intensity profile of an X-ray beam (flat field) which has been focused by two elliptical mirrors.

flat field   hologram   hologram/ flat field
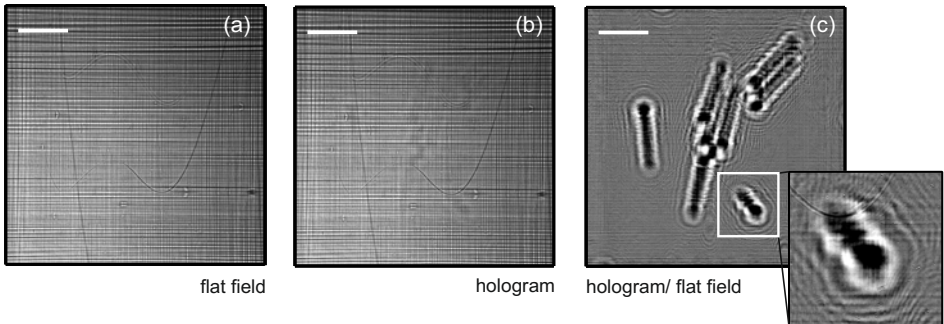
Figure 2.10: (a) Flat field recorded at ESRF/ID16A. (b) Hologram of a sample of low contrast. (c) Flat field corrected hologram of (b). Scale bars denote 2 µm.

When an illumination with such an intensity profile is used for holography, the interference fringes caused by the mirror surface can dominate the interference fringes caused by the sample which then are not visible anymore (see Fig. 2.10(b) which is the magnified hologram of a nanometer sized sample and which is basically indistinguishable from Fig. 2.10(a)). The classical procedure to deal with these kind of aberrations inherent the illumination is to divide the recorded hologram (Fig. 2.10(b)) by the intensity profile of the empty beam (Fig. 2.10(a)). The

result is depicted in Fig. 2.10(c). The interference fringes caused by the sample become visible. However, one can also see that this method of image correction does not fully eliminate all features resulting from the illumination (see magnified region of Fig. 2.10(c)). Even worse, the flat field correction is only justified in case that the following approximation holds:

$$
\begin{aligned}
\mathcal{D}_{\Delta_{i,j}}\left[o(x,y)\cdot p_i(x,y)\right] &\approx \mathcal{D}_{\Delta_{i,j}}\left[o(x,y)\right]\cdot \mathcal{D}_{\Delta_{i,j}}\left[p_i(x,y)\right] \\
&= \mathcal{D}_{\Delta_{i,j}}\left[o(x,y)\right]\cdot p_j(x,y)\ ,
\end{aligned}
\tag{2.98}
$$

where $\mathcal{D}_{\Delta_{i,j}}$ is the Fresnel propagation operator given in equation 2.45, $o(x,y)$ is the transmission function of the sample given in equation 2.76 and $p_i(x,y)$ is the illumination at defocus distance i. For a more rigorous treatment of the validity of the flat field correction, the reader is referred to [30,31].

Additionally, partial coherence limits the visibility of interference fringes between scattered and transmitted beam and contributes to a loss of information. Last but not least, resolution is always limited by the thermal and mechanical stability of the optical setup.

2.4.3 Determination of resolution from an image

Since this work is focusing on image reconstruction, a definition of image resolution is necessary. Resolution can be determined in a lot of different ways. Here, essentially two procedures of analysis were chosen: one in real space and one in reciprocal space. They will be explained in the following with the help of the image provided in Fig. 2.11(a). A first and intuitive method to estimate the resolution of Fig. 2.11(a) is to study edges appearing in the image. The steepness of these edges can serve as a measure of resolution. An edge $\Sigma(x)$ can be described by an error-function depending on the spatial coordinate x [32]

$$
\Sigma(x) = c\cdot \mathrm{erf}\left(\frac{\mu - x}{\sqrt{2}\sigma}\right) + d
\tag{2.99}
$$

$$
\text{where } \mathrm{erf}(y) = \frac{2}{\sqrt{\pi}}\int_0^y \exp(-\tau^2)d\tau\ .
\tag{2.100}
$$

The parameter c scales the height of the edge, d corrects for an offset and μ and σ are related to shifts along the x axis and the steepness of the edge. An error-function can be obtained from integration of a Gaussian function

$$
g(x) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\left(\frac{\mu - x}{\sqrt{2}\sigma}\right)^2\right]\ .
\tag{2.101}
$$

The full width at half maximum (FWHM) of a Gaussian function $g(x)$ is essentially determined by the parameter σ. It describes the steepness of the corresponding error-function

$$
\mathrm{FWHM}\left[g(x)\right] = 2\sqrt{2\ln 2}\cdot \sigma.
\tag{2.102}
$$

Thus the resolution can be estimated by fitting an error-function to selected edges and calculating $2\sqrt{2\ln 2}\,\sigma$. An example is shown in Fig. 2.11(b). Here a selected edge profile (blue line in Fig. 2.11(a)) is plotted. An error-function is adapted to this profile. With a real space pixel size in horizontal and vertical direction of $\delta x = \delta y = 100\,\mu$m, the full width at half maximum of the corresponding Gaussian function can be determined to 0.43 mm.
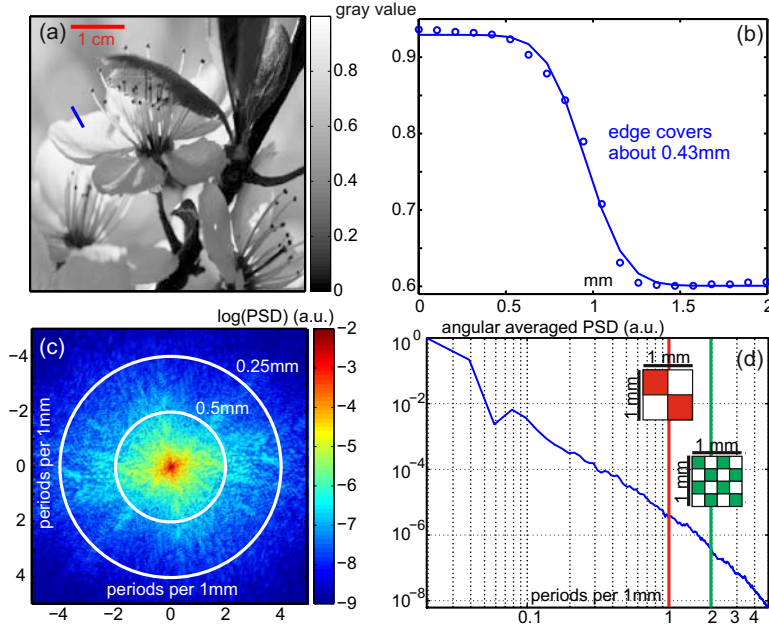
Figure 2.11: Resolution determination demonstrated for the image shown in (a). (b) An error function fitted to a selected region in (a). (c) Power spectral density of (a). (d) Angular averaged power spectral density of (a).

However, this method has a considerable drawback: It is restricted to very localized parts of the image. Detecting all of the edges that occur in the image and fitting error functions to these edges is laborious. For this reason, a different strategy is required revealing more global properties of the image. One possibility is to calculate the power spectral density[4] $\mathrm{PSD}(q_x, q_y)$ of the image $f(x,y)$ (see Fig. 2.11(c)). It is defined by the normalized, absolute value squared of the two dimensional Fourier transform $\mathcal{F}$ of $f(x,y)$

$$\mathrm{PSD}[f(x,y)](q_x, q_y) = \frac{1}{\max\left(|\mathcal{F}\left[f(x,y)\right]|^2\right)}\,|\mathcal{F}\left[f(x,y)\right]|^2. \qquad (2.103)$$

---

[4] Note that before calculation of the power spectral density, the data is usually multiplied by a window function. Here the window function is a Kaiser-Bessel window for $\beta = 8$ as defined in [33, 34].

The coordinates q_x and q_y are spatial frequency coordinates. For an image with $N \times N$ discrete pixels, q_x and q_y adopt discrete values $q_x = m\delta q_x$ and $q_y = m\delta q_y$ with

$$m = \frac{1}{2} \cdot [-N, -(N-1), ..., (N-1), N] \ ,$$

where the frequency step sizes are $\delta q_x = \frac{1}{N}$ and $\delta q_y = \frac{1}{N}$. Spatial frequencies $|(q_x, q_y)| = 0.5$ correspond to length scales of twice the pixel size of the image. Hence, they encode the maximum achievable resolution at a given sampling rate: namely edges that cover exactly two pixels. Frequencies $|(q_x, q_y)|$ within ± 0.5 encode edges that are spread over more than two pixels. Instead of calculating q_x and q_y in units of inverse pixels, we can multiply q_x and q_y by a prefactors α_x and α_y

$$\alpha_x = \frac{l_{\text{ref}}}{\delta x} \ , \quad \text{and} \quad \alpha_y = \frac{l_{\text{ref}}}{\delta y} \ , \tag{2.104}$$

where l_{ref} is a reference length. The reference length can be set to typical length scales occurring in the image. In the given example (Fig. 2.11), l_{ref} is set to 1 mm. The coordinates $(\alpha_x q_x, \alpha_y q_y)$ describe 'periods per reference length', i.e. one period per reference length corresponds to an edge (one complete black-white transition) spanned over the whole reference length (see red inlay of Fig. 2.11(d)); 2 periods per reference length correspond to two complete black-white transitions per reference length (green inlay of Fig. 2.11(d)).

To obtain a mean power spectrum independent of direction, an angular average of the power spectral density can be calculated as shown in Fig. 2.11(d). When the angular averaged power spectral density reaches a plateau, it indicates frequencies contributing to noise in the image. Here, no distinct plateau can be distinguished. However, when analyzing experimental data, this will be different.

2.5 The phase problem in coherent imaging

An optical field $\Psi(x, y, z, t)$ as defined in section 2.3.1 oscillates at $\approx 10^{19}$ Hz in the X-ray regime. No optical detection devices can follow such fast movements. Detectors measure intensities, i.e. the incoming field is integrated over time intervals much longer than 10^{-15} s:

$$I(x, y) = \langle \Psi(x, y, z, t)\Psi^*(x, y, z, t) \rangle \ , \tag{2.105}$$

where $\langle \cdot \rangle$ denotes the time average and $*$ is the complex conjugation[5]. As described in section 2.2, $I(x, y)$ is called intensity or irradiance. We are considering the ideal case of a monochromatic field of frequency ω. A monochromatic field (at position $z = i$ along the optical axis) can be separated into its time and space dependent components (see section 2.3.1):

$$\Psi_{i,\omega}(x, y, t) = \psi_{i,\omega}(x, y) \exp[-i\omega t] \ . \tag{2.106}$$

---

[5] The following argumentation until equation 2.109 can be found in [35].

In order to keep the notation simple, the index ω is skipped in the following; the lower index i denotes the position of ψ along the optical axis. The intensity $I(x, y)$ becomes

$$I(x, y) = \psi_i(x, y)\psi_i^*(x, y) \ . \tag{2.107}$$

In general, ψ_i is complex valued and can be separated into a real valued amplitude $a_i(x, y)$ and a complex exponential function including the phase $\phi_i(x, y)$:

$$\psi_i(x, y) = a_i(x, y) \exp\left(i\phi_i(x, y)\right) \ . \tag{2.108}$$

Inserting equation 2.108 into equation 2.105 results in the intensity of $\psi_i(x, y)$

$$I(x, y) = \psi_i(x, y)\psi_i^*(x, y) = a_i(x, y)^2 \ . \tag{2.109}$$

When $I(x, y)$ is measured, the phase information $\phi_i(x, y)$ is lost.
The goal of phase retrieval techniques is to recover the missing phase information from intensity measurements $I(x', y')$ at position j along the optical axis

$$
\begin{aligned}
I(x', y') &= \psi_j(x', y')\psi_j^*(x', y',) \\
&= \mathcal{P}_{\Delta_{i,j}}[\psi_i(x, y)]\mathcal{P}_{\Delta_{i,j}}^*[\psi_i(x, y)] \ .
\end{aligned}
\tag{2.110}
$$

The intensity $I(x', y')$ results from a wave field $\psi_i(x, y)$ at defocus position i that has been propagated along the distance $\Delta_{i,j}$ by an operator $\mathcal{P}_{\Delta_{i,j}}$ which is either the Fourier transform $\mathcal{F}$ or the Fresnel propagator $\mathcal{D}_{\Delta_{i,j}}$. The field $\psi_i(x, y)$ is the exit wave behind the sample placed at defocus position i and can be written as a product of the object's transmission function $o(x, y)$ and the illuminating beam or probe $p_i(x, y)$ (see section 2.3.4)

$$\psi_i(x, y) = o(x, y) \cdot p_i(x, y) \ . \tag{2.111}$$

Phase retrieval is only possible when the radiation that has transmitted an object and the radiation that has been scattered by an object is able to interfere during propagating along a distance $\Delta_{i,j}$. This is the case when the probe is sufficiently coherent (see 2.2).

2.5.1 Non-uniqueness in coherent diffractive imaging

In CDI microscopy, the properties of the absolute value of the Fourier transform

$$I(q_x, q_y) = \mathcal{F}[\psi_i(x, y)]\mathcal{F}^*[\psi_i(x, y)] \tag{2.112}$$

of a wave field $\psi_i(x, y)$ at defocus position i cause three kinds of non-uniqueness [9]. First, the recorded intensity $I(q_x, q_y)$ is insensitive to the absolute phase; only relative phases can be encoded by $I(q_x, q_y)$. Second, $I(q_x, q_y)$ is insensitive to absolute positions (x, y) of the sample. Third, $I(q_x, q_y) = \mathcal{F}[\psi_i(x, y)]\mathcal{F}^*[\psi_i(x, y)]$ is indistinguishable from $\tilde{I}(q_x, q_y) = \mathcal{F}[\psi_i(-x, -y)]\mathcal{F}^*[\psi_i^*(-x, -y)]$. These types of non-uniqueness shall be derived for the one dimensional $\psi_i(x)$ and $I(q_x)$. The extension to two dimensions is straightforward.

Absolute phase offset

To show that $I(q_x)$ is not influenced by phase offsets $\exp(i\alpha)$ with constant α, consider

$$
\begin{aligned}
\mathcal{F}\left[\psi_i(x)\exp\left(i\alpha\right)\right] &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\psi_i(x)\exp\left(i\alpha\right)\exp\left(-iq_x x\right)dx \\
&= \exp(i\alpha)\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\psi_i(x)\exp\left(-iq_x x\right)dx \\
&= \exp(i\alpha)\mathcal{F}\left[\psi_i(x)\right]\ .
\end{aligned}
\tag{2.113}
$$

For this reason,

$$
\begin{aligned}
I(q_x) &= \mathcal{F}\left[\psi_i(x)\exp(i\alpha)\right]\mathcal{F}^*\left[\psi_i(x)\exp(i\alpha)\right] \\
&= \exp(i\alpha)\mathcal{F}\left[\psi_i(x)\right]\exp(-i\alpha)\mathcal{F}^*\left[\psi_i(x)\right] \\
&= \mathcal{F}\left[\psi_i(x)\right]\mathcal{F}^*\left[\psi_i(x)\right].
\end{aligned}
\tag{2.114}
$$

Hence, intensities encode phase information up to a constant offset that cannot be determined.

Shifts of the sample

Consider the Fourier transform of a spatially shifted wave field $\psi_i(x-\xi)$:

$$
\begin{aligned}
\mathcal{F}\left[\psi_i(x-\xi)\right] &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\psi_i(x-\xi)\exp\left(-iq_x x\right)dx \\
&\overset{\tilde{x}=x-\xi}{=} \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\psi_i(\tilde{x})\exp\left(-iq_x\tilde{x}\right)\exp\left(-iq_x\xi\right)d\tilde{x} \\
&\overset{\tilde{x}\to x}{=} \exp\left(-iq_x\xi\right)\mathcal{F}\left[\psi_i(x)\right]\ .
\end{aligned}
\tag{2.115}
$$

Again, it follows that

$$
\begin{aligned}
I(q_x) &= \mathcal{F}\left[\psi_i(x-\xi)\right]\mathcal{F}^*\left[\psi_i(x-\xi)\right] \\
&= \exp\left(-i\xi\right)\mathcal{F}\left[\psi_i(x)\right]\cdot\exp\left(i\xi\right)\mathcal{F}^*\left[\psi_i(x)\right] \\
&= \mathcal{F}\left[\psi_i(x)\right]\cdot\mathcal{F}^*\left[\psi_i(x)\right]\ .
\end{aligned}
\tag{2.116}
$$

Hence, spatial shifts of the sample cannot be deduced from a single far-field intensity measurement.

Twin image

The wave field $\psi_i^*(-x)$ is called the twin image or twin signal with respect to $\psi_i(x)$.

When only intensities are recorded, intensities of the twin signal equal intensities of the signal. To see this, consider $\mathcal{F}^*\left[\psi_i(x)\right]$:

$$
\begin{aligned}
\mathcal{F}^*\left[\psi_i(x)\right] &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \psi_i^*(x) \exp^*\left(-iq_x x\right) dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \psi_i^*(x) \exp\left(-iq_x(-x)\right) dx \\
&\overset{\tilde{x}=-x}{=} -\frac{1}{\sqrt{2\pi}} \int_{\infty}^{-\infty} \psi_i^*(-\tilde{x}) \exp\left(-iq_x \tilde{x}\right) d\tilde{x} \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \psi_i^*(-\tilde{x}) \exp\left(-iq_x \tilde{x}\right) d\tilde{x} \\
&\overset{\tilde{x}\to x}{=} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \psi_i^*(-x) \exp\left(-iq_x x\right) dx = \mathcal{F}\left[\psi_i^*(-x)\right] \ .
\end{aligned}
$$

(2.117)

It follows that
$$
\mathcal{F}\left[\psi_i(x)\right] = \left(\mathcal{F}^*\left[\psi_i(x)\right]\right)^* = \mathcal{F}^*\left[\psi_i^*(-x)\right] \ .
$$
(2.118)

For this reason, the intensity is
$$
I(q_x) = \mathcal{F}\left[\psi_i(x)\right] \mathcal{F}^*\left[\psi_i(x)\right] = \mathcal{F}^*\left[\psi_i^*(-x)\right] \mathcal{F}\left[\psi_i^*(-x)\right] \ .
$$
(2.119)

Consequently, an intensity distribution $I(q_x)$ originates either from the signal $\psi_i(x)$ or the twin signal $\psi_i^*(-x)$.

2.5.2 Non-uniqueness in Fresnel coherent diffractive imaging

Holographic intensity distributions allow for less ambiguous interpretation. In particular, they are not insensitive to lateral shifts of the sample. However, absolute phase offsets are not measurable as well. In addition, the twin image is different from the twin image encountered in far-field CDI. The twin image problem in near-field holography is one of the very ancient problems arising when performing inline holography. It was already described in D. Gabor's first publication about holography [2]. Since the concept of the holographic twin image is important for this thesis, it shall be described in the remaining of this section. The argumentation basically follows the explanations given by L. Onural and P. D. Scott [36]. Propagation of wave fields in paraxial approximation is carried out using the Fresnel near-field propagator (see also equation 2.45)

$$
\mathcal{D}_{\Delta_{i,j}}[\psi_i] = \mathcal{F}^{-1}\left[\mathcal{F}[\psi_i] \cdot \exp\left\{-\frac{i\Delta_{i,j}}{2k} \cdot (q_x^2 + q_y^2)\right\}\right] \ .
$$
(2.120)

With the help of the convolution theorem [24]

$$
f(x,y) \otimes g(x,y) = 2\pi \, \mathcal{F}^{-1}\left[\mathcal{F}\left[f(x,y)\right] \cdot \mathcal{F}\left[g(x,y)\right]\right] \ ,
$$
(2.121)

equation 2.120 is

$$
\mathcal{D}_{\Delta_{i,j}}[\psi_i] = \psi_i \otimes \tilde{h}_{\Delta_{i,j}} \ ,
$$
(2.122)

where $\tilde{h}_{\Delta_{i,j}}(x,y)$ is the impulse response $h_{\Delta_{i,j}}(x,y)$ given in equation 2.48 without the constant phase factor $\exp(ik\Delta_{i,j})$.

$$\tilde{h}_{\Delta_{i,j}}(x,y) = -\frac{ik}{2\pi\Delta_{i,j}} \exp\left[\frac{ik(x^2 + y^2)}{2\Delta_{i,j}}\right] . \tag{2.123}$$

The Fourier transform of $\tilde{h}_{\Delta_{i,j}}(x,y)$ is the transfer function $\tilde{H}_{\Delta_{i,j}}(q_x, q_y)$ which again is related to the transfer function $H_{\Delta_{i,j}}$ (equation 2.48) by separating the constant phase offset $\exp(ik\Delta_{i,j})$:

$$\tilde{H}_{\Delta_{i,j}}(q_x, q_y) = \frac{1}{2\pi} \exp\left[-\frac{i\Delta_{i,j}(q_x^2 + q_y^2)}{2k}\right] . \tag{2.124}$$

The impulse response $\tilde{h}_{\Delta_{i,j}}$ has several interesting properties.

Complex conjugation of the impulse response

The complex conjugation $\tilde{h}^*_{\Delta_{i,j}}(x,y)$ is the same as the non-conjugated $\tilde{h}_{-\Delta_{i,j}}(x,y)$ with negative propagation distance:

$$\tilde{h}_{-\Delta_{i,j}}(x,y) = -\frac{ik}{2\pi(-\Delta_{i,j})} \exp\left[\frac{ik(x^2 + y^2)}{2(-\Delta_{i,j})}\right] = \tilde{h}^*_{\Delta_{i,j}}(x,y) . \tag{2.125}$$

Propagation performed by several steps

The convolution of two impulse responses with different propagation distance is the impulse response with the summed distance:

$$\begin{aligned}
\tilde{h}_{\Delta_{i,j}} \otimes \tilde{h}_{\Delta_{j,k}} &= 2\pi\mathcal{F}^{-1}\left[\tilde{H}_{\Delta_{i,j}}(q_x, q_y)\tilde{H}_{\Delta_{j,k}}(q_x, q_y)\right] \\
&= 2\pi\mathcal{F}^{-1}\left[\frac{1}{4\pi^2} \exp\left(\frac{-i(\Delta_{i,j} + \Delta_{j,k})(q_x^2 + q_y^2)}{2k}\right)\right] \\
&= \mathcal{F}^{-1}\left[\tilde{H}_{\Delta_{i,j}+\Delta_{j,k}}(q_x, q_y)\right] \\
&= \tilde{h}_{\Delta_{i,j}+\Delta_{j,k}}(x,y) = \tilde{h}_{\Delta_{i,k}} .
\end{aligned} \tag{2.126}$$

This illustrates that the propagation of light can be performed by either one single step of size $\Delta_{i,k}$ or by two (or several) successive steps $\Delta_{i,j}$ and $\Delta_{j,k}$.

Convolution of the impulse response with its complex conjugated

The convolution $\tilde{h}_{\Delta_{i,j}}(x,y) \otimes \tilde{h}^*_{\Delta_{i,j}}$ is the Dirac delta distribution $\delta(x,y)$:

$$\begin{aligned}
\tilde{h}_{\Delta_{i,j}}(x,y) \otimes \tilde{h}^*_{\Delta_{i,j}} &= \tilde{h}_{\Delta_{i,j}}(x,y) \otimes \tilde{h}_{-\Delta_{i,j}}(x,y) \\
&= 2\pi\mathcal{F}^{-1}\left[\tilde{H}_{\Delta_{i,j}}(q_x, q_y)\tilde{H}_{-\Delta_{i,j}}(q_x, q_y)\right] \\
&= 2\pi\mathcal{F}^{-1}\left[\frac{1}{4\pi^2}\right] \\
&= \mathcal{F}^{-1}\left[\frac{1}{2\pi}\right] = \delta(x,y) ,
\end{aligned} \tag{2.127}$$

where $\delta(x, y)$ is

$$\delta_{x,y} = \begin{cases} 1 & \text{if } x = y = 0 \\ 0 & \text{else ,} \end{cases} \tag{2.128}$$

and with the Fourier relations [32]

$$\mathcal{F}[\mathbb{I}] = 2\pi\delta(q_x, q_y) , \tag{2.129}$$

$$\mathcal{F}[\delta(x, y)] = \frac{1}{2\pi} , \tag{2.130}$$

where $\mathbb{I}$ is a two dimensional uniform function of amplitude one for all (x, y).

Propagation of a plane wave

Finally, one finds for the convolution of a plane wave $\mathbb{I}$ with constant phase set to zero

$$\begin{aligned} \mathbb{I} \otimes \tilde{h}_{\Delta_{i,j}}(x, y) &= 2\pi\mathcal{F}^{-1}\left[\mathcal{F}[\mathbb{I}] \, \tilde{H}_{\Delta_{i,j}}(q_x, q_y)\right] \\ &= 2\pi\mathcal{F}^{-1}\left[2\pi\delta(q_x, q_y)\tilde{H}_{\Delta_{i,j}}(q_x, q_y)\right] \\ &= 2\pi\frac{1}{2\pi}\int_{-\infty}^{\infty} 2\pi\delta(q_x, q_y)\tilde{H}_{\Delta_{i,j}}(q_x, q_y)\exp\left(i(q_x x + q_y y)\right) \, dq_x \, dq_y \\ &= 2\pi\underbrace{H_{\Delta_{i,j}}(0, 0)}_{=\frac{1}{2\pi}}\exp\left(i(0x + 0y)\right) = \mathbb{I} . \end{aligned} \tag{2.131}$$

The intensity of a hologram

The complex transmission function $o(x, y)$ (see equation 2.76) of an object can be expressed as $o(x, y) = 1 - a(x, y)$ where $a(x, y)$ is the complex opacity function, i.e. $a(x, y)$ is one where the object is fully absorbing (in this case $o(x, y)$ is zero) and $a(x, y)$ is zero where the object is fully transparent (in this case $o(x, y)$ is one). If a thin object $o(x, y)$ is positioned at defocus distance i, following the projection approximation, the exit surface wave $\psi_i(x, y)$ is

$$\psi_i(x, y) = o(x, y)p_i(x, y) = [\mathbb{I} - a(x, y)] \, p_i(x, y) , \tag{2.132}$$

where $p_i(x, y)$ is the illumination at defocus position i (i.e. at the position of the object); for simplicity, $p_i(x, y)$ shall be a plane wave with uniform amplitude $\mathbb{I}$ and zero phase. Propagation of the exit surface wave $\psi_i(x, y)$ results in

$$\begin{aligned} \mathcal{D}_{\Delta_{i,j}}[\psi_i(x, y)] &= \mathcal{D}_{\Delta_{i,j}}[\mathbb{I} - a(x, y)] \\ &= \mathcal{D}_{\Delta_{i,j}}[\mathbb{I}] - \mathcal{D}_{\Delta_{i,j}}[a(x, y)] \\ &= \mathbb{I} \otimes \tilde{h}_{\Delta_{i,j}}(x, y) - a(x, y) \otimes h_{\Delta_{i,j}}(x, y) \\ &= \mathbb{I} - a(x, y) \otimes \tilde{h}_{\Delta_{i,j}}(x, y) . \end{aligned} \tag{2.133}$$

The intensity $I(x, y) = \mathcal{D}_{\Delta_{i,j}}[\psi_i(x, y)] \mathcal{D}_{\Delta_{i,j}}^*[\psi_i(x, y)]$ can be approximated up to first order in the complex opacity function:

$$I(x, y) \approx \mathbb{I} - a(x, y) \otimes \tilde{h}_{\Delta_{i,j}}(x, y) - a^*(x, y) \otimes \tilde{h}_{\Delta_{i,j}}^*(x, y) . \tag{2.134}$$

The holographic twin image problem

Equation 2.134 describes the recording of a hologram. The hologram can be recorded on a photographic plate or with a two dimensional pixel detector. The straightforward analysis of such a hologram is to either illuminate the inverse photographic mask by the reference beam or to back propagate the digital image by the Fresnel propagator

$$
\begin{aligned}
\mathcal{D}_{-\Delta_{i,j}}[I(x,y)] \;=\;& \mathbb{I} \otimes \tilde{h}_{-\Delta_{i,j}}(x,y) \\
-\;& a(x,y) \otimes \underbrace{\tilde{h}_{\Delta_{i,j}}(x,y) \otimes \tilde{h}_{-\Delta_{i,j}}(x,y)}_{=\delta(x,y)} \\
-\;& a^{*}(x,y) \otimes \underbrace{\tilde{h}^{*}_{\Delta_{i,j}}(x,y) \otimes \tilde{h}_{-\Delta_{i,j}}(x,y)}_{=\tilde{h}_{-2\Delta_{i,j}}} = \\
=\;& \underbrace{\mathbb{I}}_{\text{direct beam}} - \underbrace{a(x,y)}_{\text{image}} - \underbrace{a^{*}(x,y) \otimes \tilde{h}_{-2\Delta_{i,j}}(x,y)}_{\text{defocused twin}} \;. \quad (2.135)
\end{aligned}
$$

The simple Fresnel back-transform of an inline-hologram will not result in an ideal image of the sample, but in a superposition of the direct beam, the focused image and a defocused complex conjugated version of the image – the so called twin image. The essential information used for phase retrieval is the measured intensity. For this reason it is clear that a non trivial problem in phase retrieval is to get rid of twin image components and to ideally reconstruct exclusively the object's transmission function.

Summary

This chapter had the main purpose to summarize the physical concepts needed in this work. Holography and coherent, lensless X-ray microscopy techniques were introduced. They rely on the use of coherent light. To this end, a brief outline of coherence theory was provided. Further important ingredients for the analysis of holographic or diffraction images are suitable descriptions of the propagation of light – in free space, in matter, in different imaging regimes and geometries. Furthermore, the notion of resolution is important to define. Resolution in lensless X-ray imaging is mainly limited by the numerical aperture of the detector and by the stability of the setup. Here, to analyze the resolution of an image, a real space method that measures the steepness of edges and a corresponding Fourier space technique that relies on the power spectral density are used. Lastly, different types of non-uniqueness that are inherent in the recorded far- and near-field intensity distributions are presented. These ambiguities will play an important role when it comes to interpret and understand the performance of phase retrieval algorithms.

3 Algorithmic of near-field propagation and phase retrieval

In this chapter different algorithmic techniques of near-field propagation and phase retrieval are presented. The chapter starts with a summary of discrete propagation techniques relying on the fast Fourier transform: the discrete version of the Fresnel-Kirchhoff integral (the so called single Fourier transform approach), the transfer function and impulse response propagation as well as the Fresnel-Bluestein ansatz.

The second part of this chapter is dedicated to iterative phase retrieval techniques. Without the claim of completeness, a summary is provided describing the mathematical techniques and concepts of algorithms, based on projections onto sets of functions fulfilling specific boundary conditions. The properties of the projection operators depend on the geometry of the sets onto which they project. Projections show useful characteristics when their underlying sets are closed and convex. These properties allow formulating iterative algorithms which converge to a solution that obeys all of the imposed boundary conditions simultaneously. Yet, it will be shown that some essential constraint sets used for phase retrieval are non-convex: a reason why convergence of the corresponding projection algorithms cannot be guaranteed.

Despite this mathematical obstacle, since the late 20th century phase retrieval techniques have been successfully applied in X-ray microscopy. Two such methods – the method of alternating projections and the difference map – shall be examined from a theoretical and an application oriented perspective. They reveal the basic ideas and problems of iterative algorithms. The reasons for the choice of these two algorithms are the following: First, the method of alternating projections is introduced, because it is broadly applied in X-ray imaging, here known as the Gerchberg-Saxton and the error reduction algorithm. Furthermore, its basic elements can be found in several more complicated projection algorithms. Second, the difference map algorithm is chosen for the reason, that it is one of the algorithms used in ptychographic coherent diffractive imaging.

Finally, in the remaining of this chapter, ptychographic phase retrieval is introduced. Ptychographic data can be analyzed using conjugate gradient optimization, a difference map formulation or an algorithm called (extended) ptychographical engine. The mathematical concepts provided in the first part of this chapter are connected to their application in ptychography. Last but not least, a motivation is presented how ptychography overcomes basic kinds of non-uniqueness connected to the far-field phase problem.

3.1 Propagation of discrete wave fields

Propagation of complex valued optical fields is both, challenging and essential for computational wave optics. It is challenging because accurate propagation simulation is accompanied by correct sampling of the phase chirp functions that are inherent in the Fresnel transform operators [37]. It is essential, since it contains the

fundamentals of wave optics necessary for digital holography. Here, four different implementations of near-field propagation shall be introduced: the single Fourier transform method [37, 38], the transfer function and the impulse response propagation [37] as well as the Fresnel-Bluestein transform [39, 40]. Extensive literature and practical implementation techniques of Fresnel diffraction can be found in the textbooks of D. G. Voelz [41] and J. D. Schmidt [42]. The section starts with the definition of the discrete Fourier transform and different formulations of the discrete Fresnel transform. Following the argumentation of [37, 41], sampling conditions for Fresnel diffraction are discussed. After that, the presented propagation algorithms are used to calculate the propagation of a plane wave passing through a square aperture. Since for this special propagation problem an analytic solution of the Fresnel propagation is available, the numeric results can be directly verified. Finally, the section closes by illustrating the performance of the different propagation techniques when used for holographic reconstruction.

3.1.1   Discrete near-field propagation

The Fresnel-Kirchhoff diffraction formula which has been introduced as a continuous integral, needs to be re-formulated for discretely sampled source and diffraction planes. To this end, also the continuous Fourier transform needs to be expressed in a discrete way. For simplicity, the theoretical considerations presented here are reduced to two spatial dimensions (one representing the lateral extend of the complex valued field and the other one the direction of propagation). The generalization to three dimensions is straightforward. Table 3.1 provides an overview of the notation used throughout this section.

Relation between continuous and discrete Fourier transform

parameter	continuous variable/ operation	discrete version
real space coordinate	x	$m\,\delta x,\ n \in \mathbb{Z}$
reciprocal space coordinate	q	$p\,\delta q,\ p \in \mathbb{Z}$
distance of propagation	Δ	Δ
integration	$\int_{-\infty}^{\infty} \dots dx$	$\sum_{n=0}^{n=N-1} \dots \delta x$
Fourier transform	$\mathcal{F}$ and $\mathcal{F}^{-1}$	$\mathbb{F}$ and $\mathbb{F}^{-1}$

Table 3.1: Continuous and discrete variables/operators.

The continuous Fourier transform of a function $f(x)$ and its inverse operation are given by

$$\mathcal{F}\left[f(x)\right] \quad = \quad F(q) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) \exp\left(-iqx\right) dx \ , \qquad (3.136)$$

$$\mathcal{F}^{-1}\left[F(q)\right] \quad = \quad f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} F(q) \exp\left(iqx\right) dq \ . \qquad (3.137)$$

The discrete Fourier transform $\mathbb{F}$ is defined as

$$\mathbb{F}\left[f(n\,\delta x)\right] = F(p\,\delta q) = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} f(n\,\delta x) \exp\left(-i\frac{2\pi}{N}pn\right) . \tag{3.138}$$

The corresponding inverse, discrete Fourier transform is defined as[6]

$$\mathbb{F}^{-1}\left[F(p\,\delta q)\right] = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} F(p\,\delta q) \exp\left(i\frac{2\pi}{N}pn\right) . \tag{3.139}$$

The exponential functions in the discrete and continuous Fourier transforms have to agree:

$$\underbrace{\exp\left(-iqx\right)}_{\text{from continuous }\mathcal{F}} = \exp\left(-ip\delta q\,n\delta x\right) \stackrel{!}{=} \underbrace{\exp\left(-i\frac{2\pi}{N}pn\right)}_{\text{from discrete }\mathbb{F}} . \tag{3.140}$$

It follows that

$$p\delta q\,n\delta x \stackrel{!}{=} \frac{2\pi}{N}pn . \tag{3.141}$$

Hence, the relation between discrete samples δx in real space and discrete samples δq in reciprocal space is

$$\delta q\,\delta x = \frac{2\pi}{N} , \tag{3.142}$$

which is a consequence of the definition of the discrete Fourier transform given in equation 3.138[7].

Periodic property of Fourier transformed discrete functions

A discrete function $f(n\delta x)$ can be obtained from a continuous function $f(x)$ by multiplication with a regularly spaced delta-comb $\mathrm{III}_{\delta x}(x)$

$$f(n\delta x) = f(x) \underbrace{\sum_{n=-\infty}^{+\infty} \delta(x - n\delta x)}_{:=\mathrm{III}_{\delta x}(x)} . \tag{3.143}$$

---

[6] Note that when replacing the integral of the continuous Fourier transform by the Riemann sum (see Table 3.1) for the discrete Fourier transform, the multiplicative factor δx is not included. The reason is, that $\mathbb{F}$ is defined on a discrete grid without special scaling of the sampling intervals [41]. Yet, when implementing propagation with the help of $\mathbb{F}$ and $\mathbb{F}^{-1}$, the scaling δx (or δq) should not be neglected.

[7] There exist several different variants for $\mathbb{F}$. Here, a definition was chosen that equally distributes the prefactor $\frac{1}{N}$ between $\mathbb{F}$ and $\mathbb{F}^{-1}$. This is analogous to the definition of the continuous Fourier transform, where the prefactor $\frac{1}{2\pi}$ was split between $\mathcal{F}$ and $\mathcal{F}^{-1}$ and both, $\mathcal{F}$ and $\mathcal{F}^{-1}$ get a prefactor $\frac{1}{\sqrt{2\pi}}$. However, it is also possible to attribute the full prefactor $\frac{1}{2\pi}$ to either $\mathcal{F}$ or $\mathcal{F}^{-1}$. Equivalently, one can combine the prefactor $\frac{1}{N}$ with either $\mathbb{F}$ or $\mathbb{F}^{-1}$.

The Fourier transform of a discrete function $f(n\delta x)$ is

$$
\begin{aligned}
\mathcal{F}\left[f(n\delta x)\right] &= \mathcal{F}\left[f(x)\cdot\mathrm{III}_{\delta x}(x)\right] \\
&= \frac{1}{\sqrt{2\pi}}\mathcal{F}\left[f(x)\right]\otimes\mathcal{F}\left[\mathrm{III}_{\delta x}(x)\right] \ ,
\end{aligned} \tag{3.144}
$$

where in the second step the (inverse) convolution theorem (equation 2.46) was applied. The Fourier transform of $\mathrm{III}_{\delta x}$ is (see for example [33])

$$
\mathcal{F}\left[\mathrm{III}_{\delta x}(x)\right] = \frac{\sqrt{2\pi}}{\delta x}\mathrm{III}_{\frac{2\pi}{\delta x}}(q) \ . \tag{3.145}
$$

For this reason the Fourier transformed, sampled version of $f(x)$ is

$$
\mathcal{F}\left[f(n\delta x)\right] = \frac{1}{\delta x}\mathcal{F}\left[f(x)\right]\otimes\mathrm{III}_{\frac{2\pi}{\delta x}}(q) \ . \tag{3.146}
$$

A convolution of a function with a delta-comb produces periodic copies of this function which are spaced by the period length of the delta-comb. Hence, when Fourier transforming a discrete, real space function, one will obtain a periodized, reciprocal space function. If $f(x)$ is bandlimited, i.e. if its Fourier transform is non-zero inside and zero outside a defined frequency region of size $[-b, b]$, then the periodic copies of $\mathcal{F}\left[f(x)\right]$ will not overlap if

$$
\frac{\pi}{\delta x} \geq b \ . \tag{3.147}
$$

If δx is smaller than $\frac{\pi}{b}$, the spacing between the periodic copies of $\mathcal{F}\left[f(x)\right]$ will be larger than necessary. Furthermore, if equation 3.147 is fulfilled, one can reconstruct the continuous $f(x)$ from its sampled version $\mathcal{F}\left[f(n\delta x)\right]$, which is known as the Whittaker-Shannon Theorem [24, 33]. Yet, choosing $\delta x > \frac{\pi}{b}$ leads to a superposition of the periodically repeated $\mathcal{F}\left[f(x)\right]$. The signal is aliased.

Single Fourier transform method
The Fresnel-Kirchhoff integral (see equation 2.50) in one dimension is

$$
\psi(x, \Delta) = \frac{1}{\sqrt{i\lambda\Delta}}\exp\left[\frac{ik}{2\Delta}x^2\right]\int_{-\infty}^{\infty}\psi(x', 0)\exp\left[\frac{ik}{2\Delta}x'^2\right]\exp\left[\frac{-ik}{\Delta}xx'\right]\ dx'. \tag{3.148}
$$

In this formulation, the constant phase prefactor $\exp\left(ik\Delta\right)$ has been separated. For a discrete equivalent of equation 3.148, the integral has to be replaced by a sum over discrete values as described in Table 3.1. In addition, the (real space) coordinates x' of the starting plane and the coordinates x of the destination plane of propagation have to be replaced by their discrete counterparts

$$
x' \rightarrow n\,\delta x' \ , \tag{3.149}
$$

$$
x \rightarrow m\,\delta x \ . \tag{3.150}
$$

This leads to

$$\psi(m\,\delta x, \Delta) = \frac{1}{\sqrt{i\lambda\Delta}} \exp\left[\frac{ik}{2\Delta}m^2\delta x^2\right]$$

$$\sum_{n=0}^{N-1} \psi(n\,\delta x', \Delta)\exp\left[\frac{ik}{2\Delta}n^2\delta x'^2\right]\exp\left[\frac{-ik}{\Delta}m\,\delta x\,n\,\delta x'\right]\delta x' \ . \quad (3.151)$$

Time consuming evaluation of the the summation in equation 3.151 can be avoided by expressing 3.151 in terms of the discrete Fourier transform. For this purpose, the last exponential function in equation 3.151 has to agree with the exponential function in equation 3.138:

$$\exp\left[\frac{-ik}{\Delta}m\,\delta x\,n\,\delta x'\right] = \exp\left[-i\frac{2\pi}{N}mn\right] \ . \quad (3.152)$$

Guaranteeing the equality of both exponents reveals a fixed relation between the pixel size of the input plane $\delta x'$ and the pixel size of the output plane δx:

$$\delta x\,\delta x' = \frac{\lambda\Delta}{N} \ . \quad (3.153)$$

Inserting equations 3.152 and 3.153 in equation 3.151 generates the discrete Fresnel-Kirchhoff integral expressed by one single Fourier transform:

$$\psi(m\,\delta x, \Delta) = \frac{1}{\sqrt{i\lambda\Delta}} \exp\left[\frac{i\pi\lambda\Delta m^2}{N^2\delta x'^2}\right]$$

$$\cdot\,\mathbb{F}\left[\psi(n\,\delta x', \Delta)\exp\left[\frac{ik}{2\Delta}n^2\delta x'^2\right]\delta x'\right] \ . \quad (3.154)$$

Fresnel transfer function propagation

In section 2.3.1, near-field propagation was deduced by the angular spectrum method. For one dimensional propagation and when separating constant phase factors, equation 2.45 is

$$\psi(x, \Delta) = \mathcal{F}^{-1}\left[\mathcal{F}[\psi(x,0)]\cdot\exp\left(-\frac{i\Delta}{2k}\cdot q^2\right)\right] \ . \quad (3.155)$$

The discrete equivalent of equation 3.155 is

$$\psi(m\,\delta x, \Delta) = \mathbb{F}^{-1}\left[\mathbb{F}[\psi(n\,\delta x, 0)]\cdot\exp\left(-\frac{i\Delta}{2k}\cdot n^2\delta q^2\right)\right] \ . \quad (3.156)$$

As detailed in section 2.5.2, the function

$$\tilde{H}_\Delta(n\,\delta q) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{i\Delta}{2k}\cdot n^2\delta q^2\right) \quad (3.157)$$

is called transfer function.

Fresnel impulse response propagation

Equation 3.155 can equally be formulated using the impulse response

$$\psi(x, \Delta) = \sqrt{2\pi} \mathcal{F}^{-1} \left[\mathcal{F}[\psi(x, 0)] \cdot \mathcal{F} \left[\frac{1}{\sqrt{i\lambda\Delta}} \exp\left(\frac{ikx^2}{2\Delta} \right) \right] \right] . \qquad (3.158)$$

The discrete version is

$$\psi(m\,\delta x, \Delta) = \mathbb{F}^{-1} \left[\mathbb{F}[\psi(n\,\delta x, 0)] \cdot \mathbb{F} \left[\frac{1}{\sqrt{i\lambda\Delta}} \exp\left(\frac{ikm^2\delta x^2}{2\Delta} \right) \right] \delta x \right] , \quad (3.159)$$

with the discrete impulse response

$$\tilde{h}_\Delta(m\,\delta x) = \frac{1}{\sqrt{i\lambda\Delta}} \exp\left(\frac{ikm^2\delta x^2}{2\Delta} \right). \qquad (3.160)$$

Fresnel-Bluestein propagation

One drawback of the single Fourier transform method given in equation 3.154 is the fixed relation between input and output pixel sizes (see equation 3.153). A slightly different formulation of the discrete Fresnel-Kirchhoff integral helps to overcome this drawback and allows scaling of input and output pixel sizes that is not restricted by equation 3.153 [39]. To this end, the complex exponential function of the discrete Fourier transform (see equation 3.152) is modified as proposed by L. I. Bluestein [43]:

$$\exp\left[\frac{-i2\pi}{N} mn \right] = \exp\left[\frac{-i\pi}{N} \left(m^2 + n^2 - (n-m)^2 \right) \right]$$

$$= \exp\left[\frac{-i\pi}{N} m^2 \right] \exp\left[\frac{-i\pi}{N} n^2 \right] \exp\left[\frac{i\pi}{N} (n-m)^2 \right] . \qquad (3.161)$$

With this formulation, equation 3.154 can be expressed in alternative way. Noting that the exponential functions defined in the output plane are not affected by the summation inherent in the discrete Fourier transform, the term $\exp\left[\frac{-i\pi}{N} m^2 \right]$ can be combined with the exponential function outside the Fourier transform in equation 3.154:

$$\exp\left[\frac{i\pi\lambda\Delta m^2}{N^2\delta x'^2} \right] \cdot \exp\left[\frac{-i\pi}{N} m^2 \right] =$$

$$\exp\left[\frac{i\pi}{\lambda\Delta} m^2\delta x^2 \right] \cdot \exp\left[\frac{-i\pi}{N} m^2 \right] =$$

$$\exp\left[-\frac{i\pi}{\lambda\Delta} m^2 \left(\frac{\lambda\Delta}{N} - \delta x^2 \right) \right] =$$

$$\exp\left[-\frac{i\pi}{\lambda\Delta} m^2\delta x \left(\delta x' - \delta x \right) \right] . \qquad (3.162)$$

In the first and last steps, relation 3.153 was used. An analogous combination of the exponential function depending on n^2 gives

$$\exp\left[\frac{ik}{2\Delta}n^2\delta x'^2\right]\cdot\exp\left[\frac{-i\pi}{N}n^2\right]=\exp\left[-\frac{i\pi}{\lambda\Delta}n^2\delta x'\left(\delta x-\delta x'\right)\right]\;. \tag{3.163}$$

Hence, again with equation 3.153 one finds

$$\begin{aligned}\psi(m\,\delta x,\Delta)\;=\;&\frac{1}{\sqrt{i\lambda\Delta}}\exp\left[-\frac{i\pi}{\lambda\Delta}m^2\delta x\left(\delta x'-\delta x\right)\right]\\&\sum_{m=1}^{N}\psi(n\,\delta x',\Delta)\exp\left[\frac{i\pi}{\lambda\Delta}n^2\delta x'\left(\delta x'-\delta x\right)\right]\\&\exp\left[\frac{i\pi}{\lambda\Delta}\delta x\,\delta x'(n-m)^2\right]\delta x'\;.\end{aligned} \tag{3.164}$$

This is a convolution of the functions [39]

$$f_1(n)\;=\;\psi(n\,\delta x',\Delta)\exp\left[\frac{i\pi}{\lambda\Delta}n^2\delta x'\left(\delta x'-\delta x\right)\right]\quad\text{and} \tag{3.165}$$

$$f_2(n)\;=\;\exp\left[\frac{i\pi}{\lambda\Delta}\delta x\,\delta x'n^2\right]\;. \tag{3.166}$$

It can be implemented as a multiplication in Fourier space

$$\boxed{\psi(m\,\delta x,\Delta)=\frac{1}{\sqrt{i\lambda\Delta}}\exp\left[-\frac{i\pi}{\lambda\Delta}m^2\delta x\left(\delta x'-\delta x\right)\right]\mathbb{F}^{-1}\left[\mathbb{F}\left[f_1\right]\cdot\mathbb{F}\left[f_2\right]\delta x'\right]\;.} \tag{3.167}$$

Note that the Fresnel-Bluestein transform with $M=\delta x/\delta x'=1$ (i.e. equal size of destination and source plane pixels) is the same as the propagation using the impulse response given in equation 3.159.

3.1.2   Sampling chirp functions

Discrete sampling of near-field propagation can lead to a large variety of artifacts; some are due to the periodicity of the discrete Fourier transform, others result from the fact that the complex valued chirp functions involved in near-field propagation are not band limited and cannot be sampled correctly [37, 41, 42].

Sampling the single Fourier transform method
The single Fourier transform method described in equation 3.148 contains two chirp functions

$$C_1(x')=\exp\left[\frac{ik}{2\Delta}x'^2\right]\quad\text{with}\;\;x'=n\delta x' \tag{3.168}$$

defined in the source plane x' and

$$C_2(x) = \exp\left[\frac{ik}{2\Delta}x^2\right] \text{ with } x = n\delta x \tag{3.169}$$

defined in the destination plane x. Both, C_1 and C_2 are 2π-periodic functions, such that the phases

$$\phi_{C_1}(x') = \frac{k}{2\Delta}x'^2 = \frac{\pi}{\lambda\Delta}x'^2 \text{ and } \phi_{C_2}(x) = \frac{k}{2\Delta}x^2 = \frac{\pi}{\lambda\Delta}x^2 \tag{3.170}$$

outside the interval

$$|\phi_{C_i}| \leq \pi \tag{3.171}$$

do not contain new information [44].

This means, in order to properly sample the chirp functions, the absolute phase difference between two neighboring pixels spaced by the frequency interval $\delta x'$ or δx has to be less than or equal to π [37, 41]. For C_1, this means

$$\delta x' \left|\frac{\partial\phi_{C_1}(x')}{\partial x'}\right|_{\max} \leq \pi \ . \tag{3.172}$$

It follows that

$$\delta x' \left|\frac{2\pi}{\lambda\Delta}x'\right|_{\max} \leq \pi \ . \tag{3.173}$$

Equation 3.173 is maximal at the limits of the field of view of the source plane, i.e. at $x'_{\max} = \pm\frac{1}{2}N\,\delta x'$, where N is the number of samples. This leads to the condition

$$\frac{\lambda\Delta}{N(\delta x')^2} \geq 1 \ . \tag{3.174}$$

The observation plane chirp C_2 is properly sampled if

$$\delta x \left|\frac{\partial\phi_{C_2}(x)}{\partial x}\right|_{\max} \leq \pi \ . \tag{3.175}$$

Calculating the derivative of C_2 with respect to x, considering that this derivative is maximal at the border of the destination plane at $x_{\max} \pm \frac{1}{2}N\delta x$ and making use of relation 3.153 results in

$$\frac{\lambda\Delta}{N(\delta x')^2} \leq 1 \ . \tag{3.176}$$

This means that strictly speaking, equation 3.154 is properly sampled only if [37, 41]

$$\frac{\lambda\Delta}{N(\delta x')^2} = 1. \tag{3.177}$$

If we are rather interested in the intensity of the propagated field instead of in its phase, the sampling condition of the destination plane chirp can be ignored and

only the source plane chirp has to fulfill relation 3.174.

Sample plane and source plane chirps can be reduced to their aliasing free regions

$$L = \frac{\lambda\Delta}{\delta x} \tag{3.178}$$

by multiplication with a filter that is one within $[-L/2, +L/2]$ and zero outside [45]. Hence, one reduces C_1 and C_2 to their undisturbed regions.

Sampling Fresnel transfer function propagation

The transfer function chirp C_{TF} in equation 3.155 is

$$C_{\text{TF}}(q) = \exp\left(-\frac{i\Delta}{2k} \cdot q^2\right) \;,\;\; \text{with}\;\; q = n\,\delta q \;. \tag{3.179}$$

It is formulated in Fourier space. The corresponding transfer function phase is

$$\phi_{\text{TR}} = -\frac{\Delta}{2k} \cdot q^2 \;. \tag{3.180}$$

The absolute maximal change of ϕ_{TR} has to be smaller than or equal to π

$$\delta q \left|\frac{\partial\phi_{\text{TR}}(q)}{\partial q}\right|_{\text{max}} \leq \pi \;. \tag{3.181}$$

It follows that

$$\delta q \left|\frac{\Delta}{k} q\right|_{\text{max}} \leq \pi \;. \tag{3.182}$$

The maximum value for the reciprocal coordinate q is $q_{\text{max}} = \pm\frac{1}{2}N\delta q$. This leads to

$$(\delta q)^2 \frac{N\Delta}{2k} \leq \pi \tag{3.183}$$

and finally with relation 3.142 and $k = \frac{2\pi}{\lambda}$ [37, 41]

$$\frac{\lambda\Delta}{N(\delta x)^2} \leq 1. \tag{3.184}$$

The transfer function propagator preserves the pixel size: Source plane and destination plane are of equal sampling $\delta x = \delta x'$.

When equation 3.184 is violated, in order to avoid aliasing, the number of pixels N of the field of view can be increased and thus again the sampling condition is fulfilled. Yet, extending the field of view leads to artifacts, because firstly, there is no way to predict the 'correct' extension and secondly, the information contained in the newly added pixels will influence the values of the original pixels during propagation.

Sampling Fresnel impulse response propagation

The chirp of the impulse response propagation given in equation 3.158 is

$$C_{\text{IR}}(x) = \exp\left[\frac{ikx^2}{2\Delta}\right] \ , \tag{3.185}$$

with the phase ϕ_{IR}

$$\phi_{\text{IR}} = \frac{kx^2}{2\Delta} \ . \tag{3.186}$$

Again, requiring C_{IR} to be 2π-periodic, leads to

$$\delta x \left|\frac{\partial C_{\text{IR}}(x)}{\partial x}\right|_{\text{max}} \leq \pi \ . \tag{3.187}$$

With $x_{\text{max}} = \frac{1}{2}N\delta x$, this is [37, 41]

$$\frac{\lambda\Delta}{N\delta x^2} \geq 1 \ . \tag{3.188}$$

Equation 3.188 states exactly the opposite condition than equation 3.184 for the transfer function sampling.

Similar to the transfer function propagation, the pixel sizes in source and destination plane are the same. In case that the sampling condition 3.188 is violated, one can reduce the impulse response to its aliasing free region

$$L = \frac{\lambda\Delta}{\delta x} \tag{3.189}$$

by setting values of C_{IR} outside $[-L/2, +L/2]$ to zero [45].

Sampling Fresnel-Bluestein propagation

When performing Fresnel-Bluestein propagation following equation 3.167, three exponential chirps have to be sampled. These are

$$C_{\text{FB1}}(m\,\delta x) = \exp\left[-\frac{i\pi}{\lambda\Delta}m^2\delta x\left(\delta x' - \delta x\right)\right] \ , \tag{3.190}$$

$$C_{\text{FB2}}(n\delta x') = \exp\left[\frac{i\pi}{\lambda\Delta}n^2\delta x'\left(\delta x' - \delta x\right)\right] \quad \text{and} \tag{3.191}$$

$$C_{\text{FB3}}(n\delta x') = \exp\left[\frac{i\pi}{\lambda\Delta}\delta x\,\delta x'n^2\right] \ . \tag{3.192}$$

If only intensities are of interest, the correct sampling of C_{FB1} is not critical. However, at least approximately correct sampling of C_{FB2} and C_{FB3} is important. Fresnel-Bluestein propagation does not require a fixed relation between input and output pixel sizes (like relation 3.153 for the Fresnel integral formulation):

$$\delta x' \cdot M = \delta x \ , \tag{3.193}$$

where M in principal can take any value [39]. When inserting equation 3.193 in C_{FB2}, one finds

$$
\begin{aligned}
C_{\text{FB2}}(n\delta x') &= \exp\left[\frac{i\pi}{\lambda\Delta}n^2\delta x'^2\right] \cdot \exp\left[-\frac{i\pi}{\lambda\Delta}n^2\delta x\,\delta x'^2\right] \\
&= \underbrace{\exp\left[\frac{ik}{2\Delta}n^2\delta x'^2\right]}_{=:C_{\text{FB2}_1}} \cdot \underbrace{\exp\left[-\frac{ik}{2\Delta}n^2 M\delta x'^2\right]}_{=:C_{\text{FB2}_2}} .
\end{aligned}
\tag{3.194}
$$

Again, by using equation 3.193, C_{FB3} can be reformulated:

$$
C_{\text{FB3}}(n\delta x') = \exp\left[\frac{ik}{2\Delta}Mn^2\delta x'^2\right] .
\tag{3.195}
$$

Recognizing that C_{FB2_1} is the same as C_{IR}, the sampling condition for C_{FB2_1} immediately follows and is

$$
\frac{\lambda\Delta}{N(\delta x')^2} \geq 1 .
\tag{3.196}
$$

The chirps C_{FB2_2} and C_{FB3} differ from C_{FB2_1} only by the factor M (and a sign), such that their sampling condition is

$$
\frac{\lambda\Delta}{MN(\delta x')^2} \geq 1 .
\tag{3.197}
$$

Unfortunately, it is not that easy to correctly filter C_{FB2} and C_{FB3}. It is in particular problematic, when the aliasing free areas of the chirps are smaller than the support of the complex valued field that should be propagated. Approaches to filter Fresnel-Bluestein propagation can be found in [40, 46].

3.1.3   Illustration of discrete near-field propagation

Accuracy of propagation is determined best via comparison with a known solution for Fresnel diffraction. An analytic solution of Fresnel diffraction is available for a square aperture. The field $U(x, \Delta)$ at distance Δ behind this aperture is given by [24]

$$
U(x, \Delta) = \frac{\exp(ik\Delta)}{2i}\left\{[C(\alpha_2) - C(\alpha_1)] + i[S(\alpha_2) - S(\alpha_1)]\right\} ,
\tag{3.198}
$$

where

$$
C(a) = \int_0^a \cos\left(\frac{\pi}{2}t^2\right)dt \quad\text{and}\quad S(a) = \int_0^a \sin\left(\frac{\pi}{2}t^2\right)dt
\tag{3.199}
$$

are the so called Fresnel integrals. The variables α_i contain wavelength, propagation distance and aperture width $2w$

$$
\alpha_1 = -\sqrt{\frac{2}{\lambda\Delta}}(w + x) \quad\text{and}\quad \alpha_2 = \sqrt{\frac{2}{\lambda\Delta}}(w - x) .
\tag{3.200}
$$

The Fresnel integrals were calculated following the approximation proposed by K. D. Mielenz [47]. A square aperture of 900 µm × 900 µm was simulated on a grid of 1000 × 1000 pixels with source plane pixels size $\delta x \times \delta y = 4.5$ µm × 4.5 µm. The wavelength λ was set to 600 nm and the propagation distance Δ was

$$\Delta = f \cdot \frac{N \delta x^2}{\lambda} \quad \text{with} \quad f = \{0.25,\, 1,\, 4\}. \tag{3.201}$$

The parameter f selects different sampling regimes. For $f = 1$ equations 3.174, 3.176, 3.184 and 3.188 are fulfilled. This means that the single Fourier transform method following equation 3.154, the transfer function propagation following equation 3.156 and the impulse response propagation following equation 3.159 should all be properly sampled. The Fresnel-Bluestein propagation given in equation 3.167 should at least be properly sampled for $\delta x / \delta x' = 1$. The short distance regime is described by $f < 1$. Here, the transfer function is sufficiently sampled while the impulse response is not. The destination plane chirp of the Fresnel propagator is better than Nyquist sampled while its source plane chirp is aliased. The sampling of Fresnel-Bluestein propagation is expected to be similar to the sampling of the impulse response and hence not enough.

Setting $f > 1$ marks the long distance regime. While the impulse response, the source chirp of the single Fourier transform method and the Fresnel-Bluestein propagation (depending on M) should be sampled sufficiently, the transfer function is aliased.

Figure 3.1 summarizes the results for the different propagation methods introduced in this section. The analytic propagation of the square aperture is plotted in black, whereas the respective numeric solutions are plotted in red (single Fourier transform propagation), green (propagation by the transfer function), orange (propagation by the impulse response) and blue (propagation by the Fresnel-Bluestein method with $M = 0.5$). On the left hand side of Fig. 3.1, unfiltered propagation is illustrated, while on the right hand side the respective chirp functions were filtered/the field of view was enlarged for propagation. Note that only intensities are analyzed; (probably) aliased phases are not shown.

The fixed scaling property between source and destination plane sampling described in equation 3.153 affects the single Fourier transform method: The longer the propagation distance, the coarser the sampling in the destination plane and the larger the field of view.

Propagation with the transfer function preserves the field of view. For the long distance regime ($f = 4$) aliasing of the intensity profile can be observed. This is removed successfully by extending the field of view before propagation.

As predicted by equation 3.188 the impulse response is undersampled in the short distance regime ($f < 1$). The consequence are periodic copies which disturb the propagated intensity profile and cause aliasing. They can be removed by filtering the transfer function chirp.

For $f = 1$ the result of the single Fourier transform method, the transfer function propagation and the impulse response approach all match the analytic intensity profile. For long distances, the transfer function is oversampled and reproduces

the analytic propagation result sufficiently well without aliasing.

Finally, Fresnel-Bluestein propagation shows comparable behavior as the impulse response propagation in the short distance regime: Periodic copies of the intensity profile appear around the centered solution and cause aliasing. The periodic copies can be avoided by reducing C_{FB2} to its aliasing free region. Care has to be taken, when this region is smaller or equal to the support of the wave field that should be propagated.

For $f = 1$ and in the long distance regime, Fresnel-Bluestein propagation (with $M = 0.5$) leads to non-aliased results. Note that the relation M between the source plane pixel size and the destination plane pixel size was arbitrarily set to a fixed value of 0.5 to illustrated that Fresnel-Bluestein propagation does neither require a fixed relation between input and output sampling intervals (as the single Fourier transform method), nor does it need to preserve the input sampling intervals (as the transfer function and the impulse response approach). The choice of M affects sampling of the chirp functions as well, such that no generalized distance regimes are determined here.
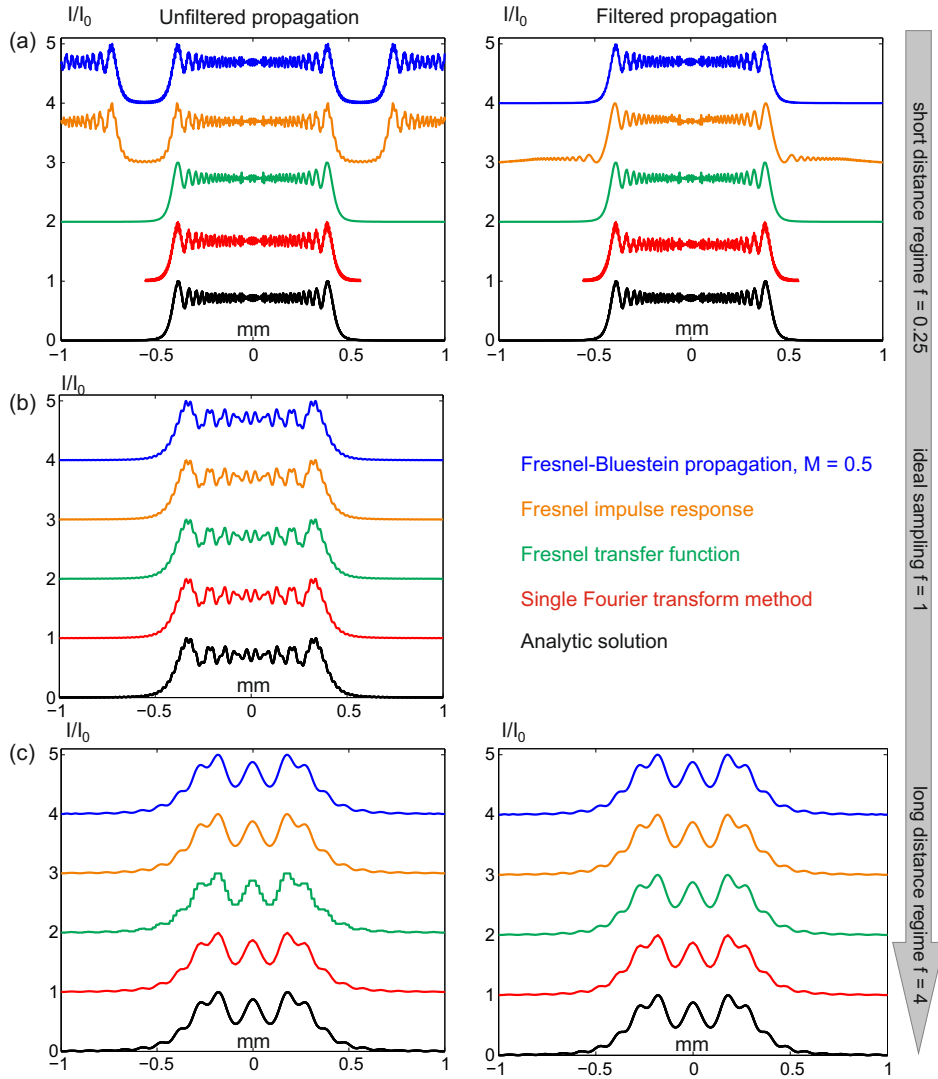
Figure 3.1: Comparison of the different propagation techniques (shifted for clarity) for a quadratic aperture and the distances given in equation 3.201. The analytic solution is shown in black. On the left, unfiltered propagation is illustrated, while on the right filtered propagation is presented. (a) Short distance regime. The impulse response propagation (orange) and the Fresnel-Bluestein propagation (blue) are aliased by periodic copies; these can be removed by filtering the chirp functions. (b) Propagation in the ideal sampling regime. No filters are necessary. (c) Propagation in the long distance regime. The propagation by the transfer function leads to aliasing (green curve). The intensities obtained from the other propagation methods are properly sampled. Extension of the field of view avoids aliasing of the transfer function propagation.

Next, the different propagation techniques are illustrated for back-propagation of a digitally recorded and flat field corrected hologram (Fig. 3.2(a)). The experimental setup used to measure hologram and flat field is described in detail in chapter 5.1.1. Important parameters are summarized in Table 3.2.

wavelength	633 nm
field of view in pixels	800×800
source plane sampling $\delta x'$	4.54 µm
propagation distance Δ	6.51 cm

Table 3.2: Parameters for the holographic reconstructions shown in Fig. 3.2.
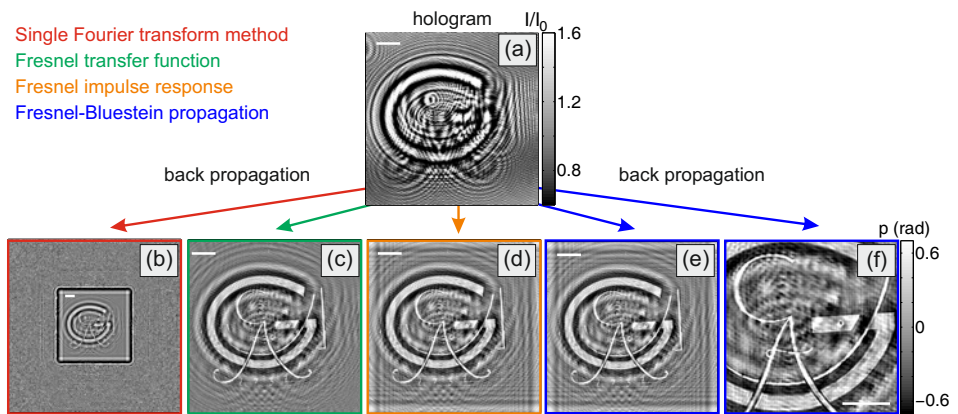
Figure 3.2: Holographic reconstruction with different propagation techniques. (a) Hologram. (b) Reconstruction using the single Fourier transform method. (c) Reconstruction using the transfer function. (d) Reconstruction using the impulse response. (e) Reconstruction using Fresnel-Bluestein propagation while preserving pixel sizes. (f) Reconstruction using Fresnel-Bluestein propagation requiring the destination plane sampling twice as fine as the source plane sampling. Images (a-f) contain the same number of pixels, all scale bars denote 0.5 mm.

The hologram is back-propagated by the single Fourier transform propagator (Fig. 3.2(b)), the use of the transfer function (Fig. 3.2(c)), the use of the impulse response (Fig. 3.2(d)) and by the Fresnel-Bluestein propagation for (e) equal source and destination pixel size and (f) $0.5\times$ finer sampling in the destination plane than in the source plane. Note that the number of pixels in (a-f) is kept the same, such that the scaling properties of the single Fourier transform method and the Fresnel-Bluestein propagation as well as the pixel size preservation of the transfer function and impulse response method become visible. In the remaining of this work, propagation is calculated by the transfer function method which performs best under several successive applications and for which it is compa-

rably easy to correct aliasing by extension of the field of view. The single Fourier transform method requires careful scaling of the pixel sizes of source and destination plane and is therefore not used for the phase retrieval techniques shown later. Fresnel-Bluestein and impulse response propagation are in particular suitable when propagation should be applied only once – like in the holographic reconstructions depicted in Fig. 3.2. Yet, since in most iterative phase retrieval techniques it is necessary to apply propagation several times, it turns out that both methods are not as well suited as the transfer function approach.

3.2 Theory of projection algorithms

The purpose of this section is to provide some mathematical terminology and a basic understanding for iterative algorithms that are designed to find elements in the intersection of two (or more) sets. The section starts with definitions of vector spaces, subspaces and subsets; properties (convexity and closedness) of subsets are explored. In the following, mappings are defined and certain properties of mappings are illustrated. A mapping called projection is examined: For closed and convex subsets, its characteristics are deduced. Projections can be applied in an iterative way to find the intersection of the subsets onto which they project.

The most intuitive combination of two projections – von Neumann's alternating projection algorithm – is introduced. A second technique – the difference map – is explained which overcomes certain difficulties of alternating projections for an appropriate choice of parameters [48]. The topics assembled in this section are addressed in more detail in the textbooks of C. Byrne [49], F. Deutsch [50], H. H. Bauschke and P. L. Combettes [51] as well as in [48, 52–55].

3.2.1 Definitions and properties of vector spaces, subspaces and subsets

Vector Space [32]

A linear space or vector space $\mathcal{V}$ over the field F of real/complex valued scalars contains an operation called addition $\mathcal{V} + \mathcal{V} \to \mathcal{V}$, an operation called scalar multiplication $F \cdot \mathcal{V} \to \mathcal{V}$, a zero element and an inverse element. The following properties hold:

$$(1) \ (\boldsymbol{u} + \boldsymbol{v}) + \boldsymbol{w} = \boldsymbol{u} + (\boldsymbol{v} + \boldsymbol{w}) \ \ \forall \, \boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in \mathcal{V}, \tag{3.202}$$

$$(2) \ \text{there is a vector } \boldsymbol{0} \in \mathcal{V} \text{ with } \ \boldsymbol{v} + \boldsymbol{0} = \boldsymbol{v} \ \ \forall \, \boldsymbol{v} \in \mathcal{V}, \tag{3.203}$$

$$(3) \ \text{for every vector } \boldsymbol{v} \text{ there is a vector } (-\boldsymbol{v}), \text{ such that } \boldsymbol{v} + (-\boldsymbol{v}) = \boldsymbol{0}, \tag{3.204}$$

$$(4) \ \boldsymbol{v} + \boldsymbol{w} = \boldsymbol{w} + \boldsymbol{v} \ \ \forall \, \boldsymbol{v}, \boldsymbol{w} \in \mathcal{V}, \tag{3.205}$$

$$(5) \ 1 \cdot \boldsymbol{v} = \boldsymbol{v} \ , \ \ \text{where 1 is the unity element of } F \ \text{ and } \boldsymbol{v} \in \mathcal{V}, \tag{3.206}$$

$$(6) \ \alpha \cdot (\beta \cdot \boldsymbol{v}) = (\alpha \cdot \beta) \cdot \boldsymbol{v} \ \ \forall \, \alpha, \beta \in F \text{ and } \boldsymbol{v} \in \mathcal{V}, \tag{3.207}$$

$$(7) \ (\alpha + \beta) \cdot \boldsymbol{v} = \alpha \cdot \boldsymbol{v} + \beta \cdot \boldsymbol{v} \ \ \forall \, \alpha, \beta \in F \text{ and } \boldsymbol{v} \in \mathcal{V}, \tag{3.208}$$

$$(8) \ \alpha \cdot (\boldsymbol{v} + \boldsymbol{w}) = \alpha \cdot \boldsymbol{v} + \alpha \cdot \boldsymbol{w} \ \ \forall \, \alpha \in F \text{ and } \boldsymbol{v}, \boldsymbol{w} \in \mathcal{V}. \tag{3.209}$$

Subspace [32]

Let $\mathcal{V}$ be a vector space over the field F and U a subset of $\mathcal{V}$. If the elements $\boldsymbol{u}$ of U fulfill conditions 3.202 to 3.209, U is called a subspace.

Affine subspace [49]

For a given subspace X_W with elements $\boldsymbol{x}_W$ of a vector space $\mathcal{V}$ and a fixed vector $\boldsymbol{d}$ of $\mathcal{V}$, the affine space W (containing elements $\boldsymbol{w}$) with respect to X_W is obtained by translating the elements of the subspace X_W by the vector $\boldsymbol{d}$:

$$W = X_W + \boldsymbol{d} = \{\boldsymbol{w} = \boldsymbol{x}_W + \boldsymbol{d} \mid \boldsymbol{x}_W \in X_W\} \ . \tag{3.210}$$

Euclidean space [49]

The Euclidean space $\mathbb{R}^N$ is a vector space that contains all N-dimensional vectors $\boldsymbol{x}$ with real valued entries x_i. For elements $\boldsymbol{x}$ and $\boldsymbol{y}$ of $\mathbb{R}^N$ the scalar product

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \sum_{i=1}^{N} x_i y_i \tag{3.211}$$

can be defined. With the help of the scalar product, the Euclidean length of a vector $\boldsymbol{x}$ is

$$||\boldsymbol{x}|| = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle} \ . \tag{3.212}$$

This immediately leads to the definition of the Euclidean distance $d(\boldsymbol{x}, \boldsymbol{y})$ of two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$:

$$d(\boldsymbol{x}, \boldsymbol{y}) = ||\boldsymbol{x} - \boldsymbol{y}|| = \sqrt{\langle \boldsymbol{x} - \boldsymbol{y}, \boldsymbol{x} - \boldsymbol{y} \rangle} \ . \tag{3.213}$$

The collection of all N-dimensional complex valued vectors $\boldsymbol{x}$ shall be denoted by $\mathbb{C}^N$. The scalar product can be defined by

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \sum_{i=1}^{N} x_i y_i^* \ , \tag{3.214}$$

where $*$ is the complex conjugation. Distances and norms are defined as for the $\mathbb{R}^N$. The spaces $\mathbb{R}^N$ and $\mathbb{C}^N$ are examples of a finite dimensional Hilbert space.

Inner product [49]

In a complex or real valued vector space $\mathcal{V}$ the scalar valued function $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ is called an inner product if for all $\boldsymbol{x}$, $\boldsymbol{y}$, $\boldsymbol{z}$ of $\mathcal{V}$ and all scalars α, the following properties are fulfilled:

$$\langle \boldsymbol{x} + \boldsymbol{y}, \boldsymbol{z} \rangle \quad = \quad \langle \boldsymbol{x}, \boldsymbol{z} \rangle + \langle \boldsymbol{y}, \boldsymbol{z} \rangle \ , \tag{3.215}$$

$$\langle \alpha \boldsymbol{x}, \boldsymbol{y} \rangle \quad = \quad \alpha \langle \boldsymbol{x}, \boldsymbol{y} \rangle \ , \tag{3.216}$$

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle \quad = \quad (\langle \boldsymbol{y}, \boldsymbol{x} \rangle)^* \ , \tag{3.217}$$

$$\langle \boldsymbol{x}, \boldsymbol{x} \rangle \quad \geq \quad 0 \text{ and if } \langle \boldsymbol{x}, \boldsymbol{x} \rangle = 0 \Leftrightarrow \boldsymbol{x} = \boldsymbol{0} \ . \tag{3.218}$$

The scalar product defined in equations 3.211 and 3.214 are examples of inner products [49].

Inner product space [56]

An inner product space $\mathcal{X}$ is a vector space $\mathcal{X}$ over a real or complex valued scalar field F equipped with an inner product $\langle \cdot, \cdot \rangle$ defined on $\mathcal{X}$. The Euclidean space $\mathbb{R}^N$ and the complex space $\mathbb{C}^N$ are inner product spaces.

Convex sets [49]

A subset C of $\mathcal{X}$ is called convex, if for every $\alpha \in \,]0, 1[$ and every two points $\boldsymbol{x}$, $\boldsymbol{y}$
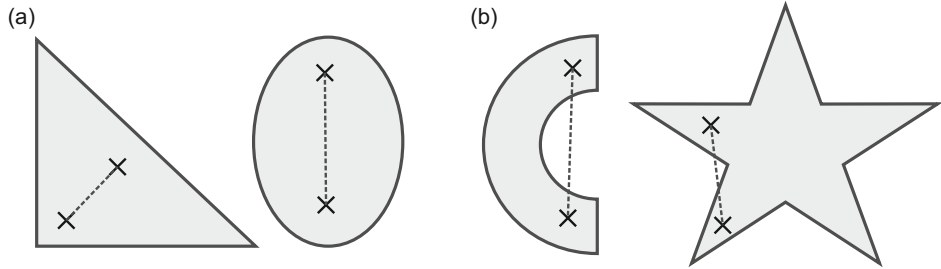
Figure 3.3: (a) Example of two convex sets. All points on the line connecting two elements of the set are inside the set as well. (b) illustration of non-convex sets. Not all points on the line connecting two elements of the set are inside the set.

of C, the combination

$$\alpha \cdot \boldsymbol{x} + (1 - \alpha) \cdot \boldsymbol{y} \tag{3.219}$$

is an element of C as well.

Equation 3.219 describes a line between the points $\boldsymbol{x}$ and $\boldsymbol{y}$. Convexity of a subset states that all points on the line in between $\boldsymbol{x}$ and $\boldsymbol{y}$ must also be elements of the subset. This is illustrated in Fig. 3.3.

Closed sets [53]

A set $C \subset \mathcal{X}$ is closed, if the limit point $\boldsymbol{c} \in \mathcal{X}$ of every sequence (c_n) in C that converges to $\boldsymbol{c}$ is within C. A closed set contains its boundary. Intersections of closed sets are closed. Unions of closed sets are closed. The empty set is also a closed set.

Distances between a point and a set [53]

The distance of a point $\boldsymbol{x} \in \mathcal{X}$ to a set $C \subset \mathcal{X}$ with elements $\boldsymbol{c}$ is defined by

$$d(\boldsymbol{x}, C) = \inf_{\boldsymbol{c} \in C} d(\boldsymbol{x}, \boldsymbol{c}) = \inf_{\boldsymbol{c} \in C} ||\boldsymbol{x} - \boldsymbol{c}|| \; . \tag{3.220}$$

3.2.2 Mappings

A mapping $T : \mathcal{X} \to \mathcal{X}$ associates every element $\boldsymbol{a} \in \mathcal{X}$ with an element $\boldsymbol{b} \in \mathcal{X}$. Here, we consider mappings that can be multi-valued (one single input value is mapped to several output values), or single valued (one single input value is mapped to one single output value).

Non-expansive mappings [51]

Let T be a mapping from the inner product space $\mathcal{X}$ to $\mathcal{X}$. T is non-expansive, if

$$||T(\boldsymbol{x}) - T(\boldsymbol{y})|| \leq ||\boldsymbol{x} - \boldsymbol{y}||, \;\; \forall \; \boldsymbol{x}, \boldsymbol{y} \in \mathcal{X} \; . \tag{3.221}$$
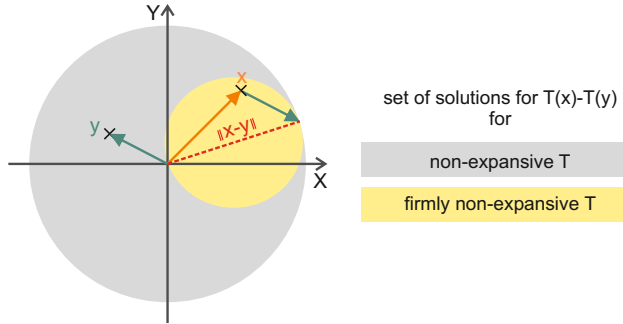
Figure 3.4: Illustration of the set of solutions for $T(\boldsymbol{x}) - T(\boldsymbol{y})$ for (gray) non-expansive T and (yellow) firmly non-expansive T. The illustration follows [57].

Firmly non-expansive mappings [51]
Let T be a mapping from the inner product space $\mathcal{X}$ to $\mathcal{X}$. T is firmly non-expansive, if

$$||T(\boldsymbol{x}) - T(\boldsymbol{y})||^2 + ||(Id - T)(\boldsymbol{x}) - (Id - T)(\boldsymbol{y})||^2 \leq ||\boldsymbol{x} - \boldsymbol{y}||^2, \ \forall \ \boldsymbol{x}, \boldsymbol{y} \in \mathcal{X} \ , \ (3.222)$$

where Id is the identity mapping. Equivalently, T is firmly non-expansive, if there is a non-expansive mapping $\tilde{T}$ such that

$$T = \frac{1}{2}(Id + \tilde{T}) \ , \tag{3.223}$$

where Id is the identity mapping which associates each input with itself [53].
Following the illustration by J. Eckstein [57] of a geometric interpretation for (firmly) non-expansive mappings, one can take the two dimensional Euclidean space and define two points $\boldsymbol{x}$ and $\boldsymbol{y}$ (see Fig. 3.4). A non-expansive mapping T guarantees that the distance between $\boldsymbol{x}$ and $\boldsymbol{y}$ is always larger than or equal to the distance between $T(\boldsymbol{x})$ and $T(\boldsymbol{y})$. While $T(\boldsymbol{x}) - T(\boldsymbol{y})$ is then found within the subset defined by the gray circle in Fig. 3.4, firmly non-expansive operators even reduce the distance between $T(\boldsymbol{x}) - T(\boldsymbol{y})$ by a factor of $\frac{1}{2}$. Their set of solutions $T(\boldsymbol{x}) - T(\boldsymbol{y})$ is shifted by half the vector $(\boldsymbol{x} - \boldsymbol{y})/2$. Solutions for a firmly non-expansive mapping T are found within the yellow circle of Fig. 3.4.

The property of firmly non-expansiveness given in equation 3.222 is equivalent to the following statement [51]:

$$||T(\boldsymbol{x}) - T(\boldsymbol{y})||^2 \leq \langle \boldsymbol{x} - \boldsymbol{y}, T(\boldsymbol{x}) - T(\boldsymbol{y}) \rangle \ . \tag{3.224}$$

To see this consider

$$
\begin{aligned}
||(Id - T)(\boldsymbol{x}) - (Id - T)(\boldsymbol{y})||^2 &= ||(\boldsymbol{x} - \boldsymbol{y}) - (T(\boldsymbol{x}) - T(\boldsymbol{y}))||^2 \\
&= ||\boldsymbol{x} - \boldsymbol{y}||^2 + ||T(\boldsymbol{x}) - T(\boldsymbol{y})||^2 \\
&\quad - 2 \langle \boldsymbol{x} - \boldsymbol{y}, T(\boldsymbol{x}) - T(\boldsymbol{y}) \rangle \ . \tag{3.225}
\end{aligned}
$$

With this restatement equation 3.222 becomes

$$||T(\boldsymbol{x}) - T(\boldsymbol{y})||^2 + ||\boldsymbol{x} - \boldsymbol{y}||^2 + ||T(\boldsymbol{x}) - T(\boldsymbol{y})||^2 - 2\langle \boldsymbol{x} - \boldsymbol{y}, T(\boldsymbol{x}) - T(\boldsymbol{y})\rangle$$
$$\leq ||\boldsymbol{x} - \boldsymbol{y}||^2$$
$$\Rightarrow \ ||T(\boldsymbol{x}) - T(\boldsymbol{y})||^2 \leq \langle \boldsymbol{x} - \boldsymbol{y}, T(\boldsymbol{x}) - T(\boldsymbol{y})\rangle . \qquad (3.226)$$

Iterative application of mappings and fixed points

Mappings can be applied in an iterative way

$$T^n(\boldsymbol{a}) = \underbrace{T(T(.....(T(\boldsymbol{a}))))}_{n \text{ times}} . \qquad (3.227)$$

If $T(T(\boldsymbol{a})) = T(\boldsymbol{a})$, $\forall \boldsymbol{a} \in \mathcal{X}$ the mapping T is called **idempotent**. This means, that elements of the set of solutions for T are mapped onto themselves. Furthermore, an element $\boldsymbol{x}$ of $\mathcal{X}$ for which $T(\boldsymbol{x}) = \boldsymbol{x}$ belongs to the **fixed point set** of T [53]

$$\text{Fix}\, T = \{\boldsymbol{x} \in \mathcal{X} : \ T(\boldsymbol{x}) = \boldsymbol{x}\} . \qquad (3.228)$$

Usually, fixed points are the limit points of sequences [53].

Weak convergence of a sequence [53]

Suppose $\boldsymbol{x}$ is a point in $\mathcal{X}$. A sequence $(\boldsymbol{x}_n)$ in $\mathcal{X}$ is said to be weakly convergent, if

$$\langle \boldsymbol{x}_n - \boldsymbol{x}, \boldsymbol{y}\rangle \to 0, \ \ \forall \ \boldsymbol{y} \in \mathcal{X}. \qquad (3.229)$$

The fixed points of a mapping can be found by iterative algorithms. This is not always an easy task. Yet, as shown by by Z. Opial [58] in the year 1967, when the mapping is firmly non-expansive, it is possible to find its fixed points by iterative application of the map (see equation 3.227).

Weak convergence of a sequence
of firmly non-expansive operators to a fixed point [58]

The sequence $(T^n(\boldsymbol{x}))$ generated by a firmly non-expansive mapping $T : \mathcal{X} \to \mathcal{X}$ with $\text{Fix}\, T \neq \emptyset$ converges weakly to a point in $\text{Fix}\, T$.

3.2.3 Projections

Projections are special kinds of mappings, that find the best approximation of a point $\boldsymbol{x}$ with respect to a subset A [53]. Depending on the geometry of A, they have different properties. In particular, it turns out that they point to a unique solution and are (firmly) non-expansive if their underlying subsets are closed and convex. Projections are the most important operators for iterative phase retrieval which is – speaking more abstractly – finding fixed points of a certain mapping. The mappings used for phase retrieval are compositions of projections. To this end, here the mathematical properties of projections are explored before special projection operators for the case of (Fresnel) coherent diffractive imaging are classified. This section is composed of material taken from [49, 50, 53, 54].
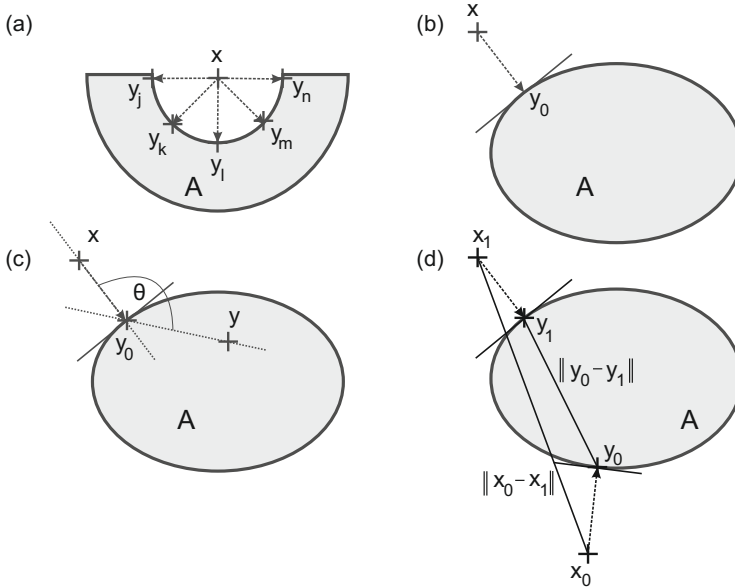
Figure 3.5: Characterizations of projections. (a) Projections onto closed sets can be multi-valued. (b) Projections onto closed, convex sets are unique. (c) The angle between the line defined by a point $\boldsymbol{x}$ and its projection $\boldsymbol{y_0}$ onto a subset A and the line defined by $\boldsymbol{y_0}$ and any other point $\boldsymbol{y}$ within A is at least 90 degrees. (d) Projections are non-expansive operators.

Best approximation [50]

Let A be a non-empty subset with elements $\boldsymbol{y}$ of an inner product space $\mathcal{X}$, that contains elements $\boldsymbol{x}$. An element $\overline{\boldsymbol{x}} \in A$ is called best approximation to $\boldsymbol{x} \in \mathcal{X}$ (or nearest point to $\boldsymbol{x}$ from A) if

$$||\boldsymbol{x} - \overline{\boldsymbol{x}}|| = d(\boldsymbol{x}, A) := \inf_{y \in A} ||\boldsymbol{x} - \boldsymbol{y}|| \ . \tag{3.230}$$

In words, the best approximation $\overline{\boldsymbol{x}}$ of a point $\boldsymbol{x}$ with respect to a subset A is found at infimum distance between $\boldsymbol{x}$ and A. Yet, since the infimum distance is not the minimal distance, but the lowest boundary distance, a best approximation does not forcibly have to exist.

Projection [50]

Let A be a non-empty subset of an inner product space $\mathcal{X}$, that contains elements $\boldsymbol{x}$. The set of all best approximation points $\overline{\boldsymbol{x}} \in A$ for $\boldsymbol{x}$ is denoted $P_A(\boldsymbol{x})$ and called projection:

$$P_A(\boldsymbol{x}) := \{\overline{\boldsymbol{x}} \in A \mid ||\boldsymbol{x} - \overline{\boldsymbol{x}}|| = d(\boldsymbol{x}, A)\} \ . \tag{3.231}$$

In a finite dimensional inner product space $\mathcal{X}$ with subset A, the set of all best approximation points of $\boldsymbol{x} \in \mathcal{X}$ with respect to A can be empty, contain a single element or more than one element. It is necessary for those projections which point to solutions different from the empty set that the subset onto which they map is closed. Fig. 3.5(a) illustrates the projection of $\boldsymbol{x}$ onto a set A. For the sketched case, this projection is multi-valued. As already indicated, projections have special properties, once the subsets onto which they project are closed and convex. In the following projections onto closed and convex sets are considered.

Uniqueness of best approximations [50]
Let A be a convex subset of the inner product space $\mathcal{X}$. Each $\boldsymbol{x} \in \mathcal{X}$ has at most one best approximation in A.

To see this, take a point $\boldsymbol{x} \in \mathcal{X}$ and two points $\boldsymbol{y_1}, \boldsymbol{y_2} \in P_A(x)$. As A is convex, $\frac{1}{2}(\boldsymbol{y_1} + \boldsymbol{y_2})$ is an element of A as well. The distance between $\boldsymbol{x}$ and A is smaller than or equal to the distance between $\boldsymbol{x}$ and $\frac{1}{2}(\boldsymbol{y_1} + \boldsymbol{y_2})$

$$d(\boldsymbol{x}, A) \quad \leq \quad \left\| \boldsymbol{x} - \frac{1}{2}(\boldsymbol{y_1} + \boldsymbol{y_2}) \right\| \tag{3.232}$$

$$= \quad \left\| \frac{1}{2}(\boldsymbol{x} - \boldsymbol{y_1}) + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{y_2}) \right\| \tag{3.233}$$

$$\leq \quad \frac{1}{2} \underbrace{||\boldsymbol{x} - \boldsymbol{y_1}||}_{=d(\boldsymbol{x}, A)} + \frac{1}{2} \underbrace{||\boldsymbol{x} - \boldsymbol{y_2}||}_{=d(\boldsymbol{x}, A)} = d(\boldsymbol{x}, A) . \tag{3.234}$$

The first and last equality symbols state that equality must be guaranteed throughout the inequalities. Equality for the triangle inequality holds if $\boldsymbol{x} - \boldsymbol{y_1} = \rho(\boldsymbol{x} - \boldsymbol{y_2})$ and with $||\boldsymbol{x} - \boldsymbol{y_1}|| = d(\boldsymbol{x}, A) = ||\boldsymbol{x} - \boldsymbol{y_2}||$ the parameter ρ must be $\rho = 1$. It follows that $\boldsymbol{y_1} = \boldsymbol{y_2}$, i.e. there is a unique projection of $\boldsymbol{x}$ onto A.
Fig. 3.5(b) shows such a projection onto a closed and convex set A. Only a single point $\boldsymbol{y} \in A$ is closest to $\boldsymbol{x}$.

Characterizations of best approximations from convex sets [50]
Let A be a non-empty, convex subset of the inner product space $\mathcal{X}$. The point $\boldsymbol{y_0} \in A$ is best approximation point of $\boldsymbol{x}$ with respect to A (i.e. $\boldsymbol{y_0} = P_A(\boldsymbol{x})$, $\boldsymbol{x} \in \mathcal{X}$) if and only if

$$\langle \boldsymbol{x} - \boldsymbol{y_0}, \boldsymbol{y} - \boldsymbol{y_0} \rangle \leq 0, \ \forall \ \boldsymbol{y} \in A . \tag{3.235}$$

Suppose $\langle \boldsymbol{x} - \boldsymbol{y_0}, \boldsymbol{y} - \boldsymbol{y_0} \rangle \leq 0$. Then it follows

$$||\boldsymbol{x} - \boldsymbol{y_0}||^2 \quad = \quad \langle \boldsymbol{x} - \boldsymbol{y_0}, \boldsymbol{x} - \boldsymbol{y_0} \rangle \tag{3.236}$$

$$\overset{\text{with relation 3.215}}{=} \quad \langle \boldsymbol{x} - \boldsymbol{y_0}, \boldsymbol{x} - \boldsymbol{y} \rangle + \underbrace{\langle \boldsymbol{x} - \boldsymbol{y_0}, \boldsymbol{y} - \boldsymbol{y_0} \rangle}_{\leq 0} \tag{3.237}$$

$$\leq \quad \langle \boldsymbol{x} - \boldsymbol{y_0}, \boldsymbol{x} - \boldsymbol{y} \rangle \tag{3.238}$$

$$\leq \quad ||\boldsymbol{x} - \boldsymbol{y_0}|| \ ||\boldsymbol{x} - \boldsymbol{y}|| . \tag{3.239}$$

The last inequality is the Cauchy-Schwarz inequality. This means that

$$\|\boldsymbol{x} - \boldsymbol{y_0}\|^2 \leq \|\boldsymbol{x} - \boldsymbol{y_0}\| \; \|\boldsymbol{x} - \boldsymbol{y}\| \quad \text{and} \tag{3.240}$$

$$\|\boldsymbol{x} - \boldsymbol{y_0}\| \leq \|\boldsymbol{x} - \boldsymbol{y}\| \; \forall \; \boldsymbol{y} \; . \tag{3.241}$$

Hence, $\boldsymbol{y_0} = P_A(\boldsymbol{x})$ is best approximation point.

Suppose that there are some $\boldsymbol{y} \in A$ with $\langle \boldsymbol{x} - \boldsymbol{y_0}, \boldsymbol{y} - \boldsymbol{y_0} \rangle \geq 0$. By convexity of A every point $\boldsymbol{y_\lambda} = \lambda \boldsymbol{y} + (1 - \lambda) \boldsymbol{y_0}$ with $\lambda \in]0, 1[$ is in A as well. Then

$$\|\boldsymbol{x} - \boldsymbol{y_\lambda}\|^2 = \langle \boldsymbol{x} - \boldsymbol{y_\lambda}, \boldsymbol{x} - \boldsymbol{y_\lambda} \rangle \tag{3.242}$$

$$= \langle \boldsymbol{x} - \lambda \boldsymbol{y} - (1 - \lambda) \boldsymbol{y_0}, \boldsymbol{x} - \lambda \boldsymbol{y} - (1 - \lambda) \boldsymbol{y_0} \rangle \tag{3.243}$$

$$= \langle \boldsymbol{x} - \boldsymbol{y_0} - \lambda(\boldsymbol{y} - \boldsymbol{y_0}), \boldsymbol{x} - \boldsymbol{y_0} - \lambda(\boldsymbol{y} - \boldsymbol{y_0}) \rangle \tag{3.244}$$

$$= \langle \boldsymbol{x} - \boldsymbol{y_0}, \boldsymbol{x} - \boldsymbol{y_0} \rangle + \lambda^2 \langle \boldsymbol{y} - \boldsymbol{y_0}, \boldsymbol{y} - \boldsymbol{y_0} \rangle$$
$$- 2\lambda \langle \boldsymbol{x} - \boldsymbol{y_0}, \boldsymbol{y} - \boldsymbol{y_0} \rangle \tag{3.245}$$

$$= \|\boldsymbol{x} - \boldsymbol{y_0}\|^2 + \lambda^2 \|\boldsymbol{y} - \boldsymbol{y_0}\|^2 - 2\lambda \langle \boldsymbol{x} - \boldsymbol{y_0}, \boldsymbol{y} - \boldsymbol{y_0} \rangle \tag{3.246}$$

$$= \|\boldsymbol{x} - \boldsymbol{y_0}\|^2 - \lambda \left[2 \langle \boldsymbol{x} - \boldsymbol{y_0}, \boldsymbol{y} - \boldsymbol{y_0} \rangle - \lambda \|\boldsymbol{y} - \boldsymbol{y_0}\|^2 \right] \; . \tag{3.247}$$

Choosing λ sufficiently small, the last bracket is greater zero. It follows that

$$\|\boldsymbol{x} - \boldsymbol{y_\lambda}\|^2 < \|\boldsymbol{x} - \boldsymbol{y_0}\|^2 \; . \tag{3.248}$$

That contradicts the assumption that $y_0 = P_A(\boldsymbol{x})$ is best approximation of $\boldsymbol{x}$. Equation 3.235 describes that the angle between the line defined by a point $\boldsymbol{x}$ and its projection $\boldsymbol{y_0}$ onto A and the line defined by $\boldsymbol{y_0}$ and any other point $\boldsymbol{y}$ within A is larger than or equal to 90 degrees (see Fig. 3.5(c)).

Projections are idempotent operators
It can be seen directly from the definition of a projection: The closest point of an element $P_A(\boldsymbol{x}) \in A$ with respect to A is the element itself:

$$P_A(P_A(\boldsymbol{x})) = P_A(\boldsymbol{x}) \; . \tag{3.249}$$

Projections onto convex sets are non-expansive operators [49, 59]
This means that

$$\|P_A(\boldsymbol{x}) - P_A(\boldsymbol{y})\| \leq \|\boldsymbol{x} - \boldsymbol{y}\| \; . \tag{3.250}$$

Equation 3.235 states that:

$$\langle \boldsymbol{x} - P_A(\boldsymbol{x}), P_A(\boldsymbol{y}) - P_A(\boldsymbol{x}) \rangle \leq 0 \; , \tag{3.251}$$

$$\langle \boldsymbol{y} - P_A(\boldsymbol{y}), P_A(\boldsymbol{x}) - P_A(\boldsymbol{y}) \rangle \leq 0 \; . \tag{3.252}$$

Combining both inequalities leads to

$$\langle \boldsymbol{x} - \boldsymbol{y} - P_A(\boldsymbol{x}) + P_A(\boldsymbol{y}), P_A(\boldsymbol{y}) - P_A(\boldsymbol{x}) \rangle \leq 0 \tag{3.253}$$

$$\langle \boldsymbol{x} - \boldsymbol{y}, P_A(\boldsymbol{y}) - P_A(\boldsymbol{x}) \rangle + \|P_A(\boldsymbol{y}) - P_A(\boldsymbol{x})\|^2 \leq 0 \tag{3.254}$$

$$- \langle \boldsymbol{x} - \boldsymbol{y}, P_A(\boldsymbol{x}) - P_A(\boldsymbol{y}) \rangle + \|P_A(\boldsymbol{x}) - P_A(\boldsymbol{y})\|^2 \leq 0 \; . \tag{3.255}$$

It follows that

$$
\begin{aligned}
||P_A(\boldsymbol{x}) - P_A(\boldsymbol{y})||^2 &\leq ||P_A(\boldsymbol{x}) - P_A(\boldsymbol{y})|| \, ||\boldsymbol{x} - \boldsymbol{y}|| && (3.256) \\
||P_A(\boldsymbol{x}) - P_A(\boldsymbol{y})|| &\leq ||\boldsymbol{x} - \boldsymbol{y}|| \, . && (3.257)
\end{aligned}
$$

Equation 3.257 characterizes a non-expansive operator (equation 3.221). Hence, projections are non-expansive operators. An illustration is provided in Fig. 3.5(d): The distance between two points $\boldsymbol{x_0}$ and $\boldsymbol{x_1}$ is longer than the distance between their corresponding mappings onto $\boldsymbol{y_0}$ and $\boldsymbol{y_1}$.

A projection onto a convex set is even more than non-expansive: It is a firmly non-expansive operator.

Projections onto convex sets are firmly non-expansive operators [49,51]
Consider

$$
\begin{aligned}
||P_A(\boldsymbol{x}) - P_A(\boldsymbol{y})||^2 &= \langle P_A(\boldsymbol{x}) - P_A(\boldsymbol{y}), P_A(\boldsymbol{x}) - P_A(\boldsymbol{y}) \rangle \\
&= \langle P_A(\boldsymbol{x}) - \boldsymbol{x} + \boldsymbol{x} - P_A(\boldsymbol{y}) + \boldsymbol{y} - \boldsymbol{y}, P_A(\boldsymbol{x}) - P_A(\boldsymbol{y}) \rangle \\
&= \langle \boldsymbol{x} - \boldsymbol{y}, P_A(\boldsymbol{x}) - P_A(\boldsymbol{y}) \rangle \\
&\quad + \underbrace{\langle P_A(\boldsymbol{x}) - \boldsymbol{x}, P_A(\boldsymbol{x}) - P_A(\boldsymbol{y}) \rangle}_{(I)} \\
&\quad + \underbrace{\langle \boldsymbol{y} - P_A(\boldsymbol{y}), P_A(\boldsymbol{x}) - P_A(\boldsymbol{y}) \rangle}_{(II)} \, . && (3.258)
\end{aligned}
$$

Part (I) in equation 3.258 can be written as

$$
\langle P_A(\boldsymbol{x}) - \boldsymbol{x}, P_A(\boldsymbol{x}) - P_A(\boldsymbol{y}) \rangle = \langle \boldsymbol{x} - P_A(\boldsymbol{x}), P_A(\boldsymbol{y}) - P_A(\boldsymbol{x}) \rangle \, . \tag{3.259}
$$

Due to the property of projections onto closed, convex sets given equation 3.235, (I) and (II) are both smaller than or equal to zero. It results that

$$
||P_A(\boldsymbol{x}) - P_A(\boldsymbol{y})||^2 \leq \langle \boldsymbol{x} - \boldsymbol{y}, P_A(\boldsymbol{x}) - P_A(\boldsymbol{y}) \rangle \, , \tag{3.260}
$$

which is the definition of firmly non-expansiveness (see equation 3.224).
Finally it should be emphasized once more, that all these characteristics (uniqueness, (firmly) non-expansiveness) only hold for those projections P_A that map onto closed and convex sets A. As we will see, in the special problem of phase retrieval, unfortunately not all underlying subsets A are closed and convex. For this reason convergence of phase retrieval algorithms in general cannot be guaranteed.

Relaxed projections [48,60]
A relaxed projection rP_A is a map $rP_A : \mathcal{X} \to \mathcal{X}$ (where $\mathcal{X}$ is an inner product space) that follows the following criterion:

$$
rP_A := (1 + \gamma)P_A - \gamma Id \, , \quad \text{with } \gamma \in \mathbb{R} \, . \tag{3.261}
$$
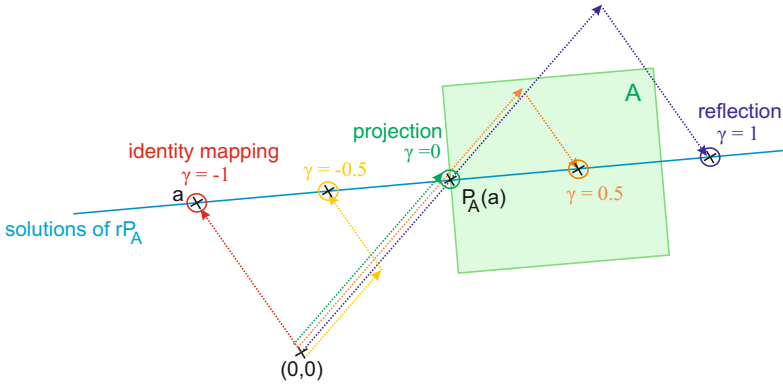
Figure 3.6: Illustration of the relaxed projection rP_A for different relaxation values $\gamma_i \in \mathbb{R}$.

The solution space of rP_A is a line defined by the point $\boldsymbol{a}$ and its projection $P_A(\boldsymbol{a})$ onto the constraint set A. The parameter γ selects a specific point on this line: one special result of rP_A. Hence, the parameter γ determines where the output of rP_A will be located with respect to its input value $\boldsymbol{a}$ and the projection of $\boldsymbol{a}$ onto the constraint set A. Instead of matching $\boldsymbol{a}$ directly onto $P_A(\boldsymbol{a})$, γ relaxes the position of the output value to points located on the line connecting $\boldsymbol{a}$ and $P_A(\boldsymbol{a})$. For this reason, γ is also called relaxation parameter. A special kind of relaxed projection is a reflection with $\gamma = 1$. The behavior of rP_A depends on the choice of γ and is illustrated in Fig. 3.6: P_A is the projection operator associated with the set A (shown in green). All possible solutions (i.e. solutions for all $\gamma \in \mathbb{R}$) are located on the line drawn in light blue. For $\gamma = 0$ the solution of rP_A will be the same as the solution of P_A (green circle). Choosing $\gamma_i = -1$ will match all input values $\boldsymbol{a}$ onto themselves (red circle): In this case, rP_A is the identity map Id. The reflection with $\gamma = 1$ is marked by the dark blue circle. Solutions of rP_A for $\gamma = \{-0.5, \ 0.5\}$ are also illustrated (yellow and orange circles). Of course, not only projections, but also every other kind of mappings can be relaxed.

3.3 Searching the intersection of two subsets

With the help of projections onto subsets C and D, elements should be found that simultaneously fulfill the properties of C and D. In other words, these elements are located in the intersection of C and D: $C \cap D$. Unfortunately, only one projection can be performed per time. Hence, it is not possible to directly project onto $C \cap D$. Hence, one combines these available one-step methods to a map such that by repetition of this combination finally an element of $C \cap D$ can be found. Further repetition ideally will not hop away from the element in $C \cap D$. There exists a variety of such combinations of projections. Two methods frequently used in phase retrieval are explained in this section, namely the map of alternating projections and the difference map. First, the algorithm of alternating projections

as the most simple projection algorithm shall be introduced. Second, preparing the section dealing with ptychographic algorithms, the difference map scheme will be explained. The behavior of alternating projections and the difference map will be examined for the ideal situation of projections onto convex, closed and orthogonal subsets. This is a special case but it will provide a basic understanding for the operations performed by the proposed maps when iterated more than once. The essential arguments presented below can be found in [48].

3.3.1   The map of alternating projections

The map of von Neumann's alternating projections AP is the most straight forward approach to find elements in the intersection of two constraint sets C and D. It is defined as

$$\mathrm{AP} := P_C\,P_D \ . \tag{3.262}$$

As proved in 1959 by W. Cheney and A. A. Goldstein, under certain conditions, especially under the condition of closed and convex sets this map converges to a fixed point.

Theorem of W. Cheney and A. A. Goldstein [59]
Let C and D be two closed, convex subsets of an inner product space $\mathcal{X}$ and let either one of the subsets be compact (closed and bounded), or let one of the subsets be finite dimensional and let the distance between the sets be attained, then the sequence $(\mathrm{AP})^n = (P_C\,P_D)^n$ of the projections P_C and P_D converges to a fixed point.

Furthermore, L. M. Brègman found that the map converges weakly to a fixed point which is in the intersection of both constraint sets, provided that the sets intersect.

Theorem of L. M. Brègman [53, 61]
Let $C \cap D \neq \emptyset$. Then both sequences (c_n) and (d_n) with

$$d_n = P_D(c_n) \ \ \text{and} \ \ c_{n+1} = P_C(d_n) \tag{3.263}$$

converge weakly to a point in $C \cap D$.

Next, following [48], a graphical motivation will be presented to describe what happens when alternating projections onto closed, convex, orthogonal and intersecting or non-intersecting subsets are performed.
We suppose that the constraint sets C and D are affine subspaces (defined in equation 3.210) in the n-dimensional Euclidean space $\mathbb{R}^N$. Their corresponding linear subspaces (i.e. which contain the origin) are denoted by X_C and X_D. The span of X_C and X_D is a linear subspace of $\mathbb{R}^N$ as well. It contains all linear combinations of elements of X_C and X_D. Its complement Y contains elements perpendicular to linear combinations of elements of X_C and X_D. Following equa-
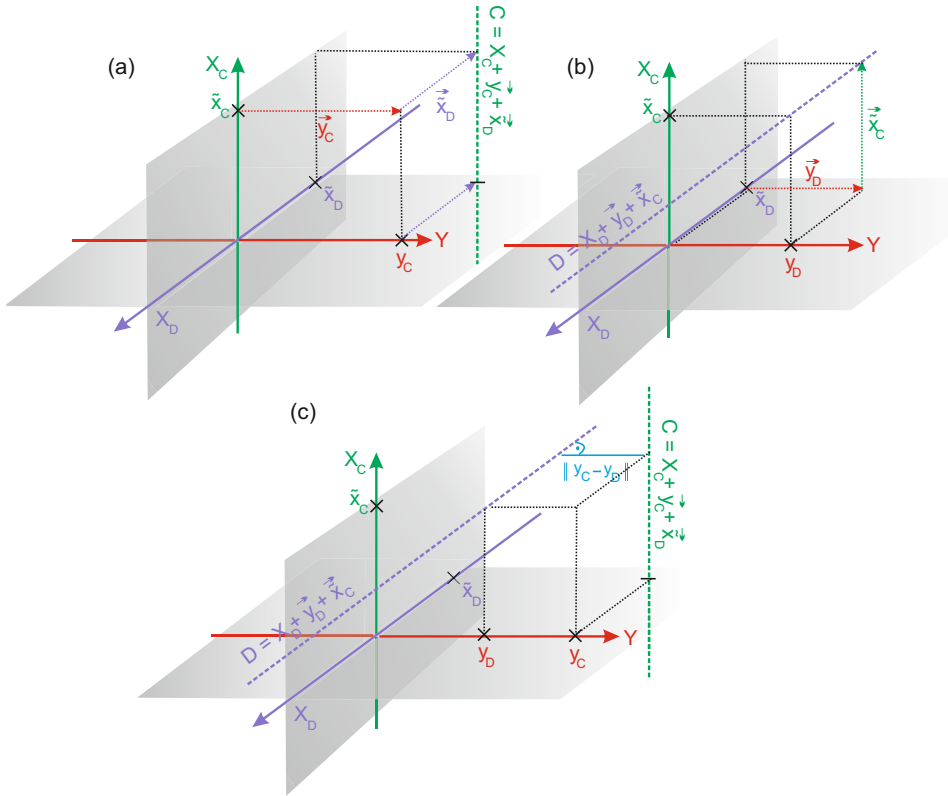
Figure 3.7: Illustration of equations 3.264 (a) and 3.265 (b) for $\mathbb{R}^3$. The orthogonal, affine subspaces C and D are generated by shifting the linear subspaces X_C and X_D in $\mathbb{R}^3$. The red axis Y is the complementary subspace to the plane defined by X_C and X_D. Y, X_C and X_D fully span the $\mathbb{R}^3$. (c) The minimal distance between C and D is $||\boldsymbol{y}_C \quad \boldsymbol{y}_D||$.

tion 3.210, the affine subspaces C and D are generated by shifting their underlying linear subspaces X_C and X_D:

$$C = X_C + \tilde{\boldsymbol{x}}_D + \boldsymbol{y}_C \quad \text{with } \boldsymbol{y}_C \in Y,\ \tilde{\boldsymbol{x}}_D \in X_D\ , \tag{3.264}$$

$$D = X_D + \tilde{\boldsymbol{x}}_C + \boldsymbol{y}_D \quad \text{with } \boldsymbol{y}_D \in Y,\ \tilde{\boldsymbol{x}}_C \in X_C\ . \tag{3.265}$$

Their minimal distance is $||\boldsymbol{y}_C \quad \boldsymbol{y}_D||$. If $\boldsymbol{y}_C$ equals $\boldsymbol{y}_D$, C and D intersect. An illustration is depicted in Fig. 3.7. An arbitrary point $\boldsymbol{a}$ in in $\mathbb{R}^N$ contains components of X_C, X_D and Y:

$$\boldsymbol{a} = \boldsymbol{x}_C + \boldsymbol{x}_D + \boldsymbol{y}\ . \tag{3.266}$$

If $\boldsymbol{a}$ is projected onto C via P_C it results in

$$P_C(\boldsymbol{a}) = \boldsymbol{x}_C + \tilde{\boldsymbol{x}}_D + \boldsymbol{y}_C\ . \tag{3.267}$$
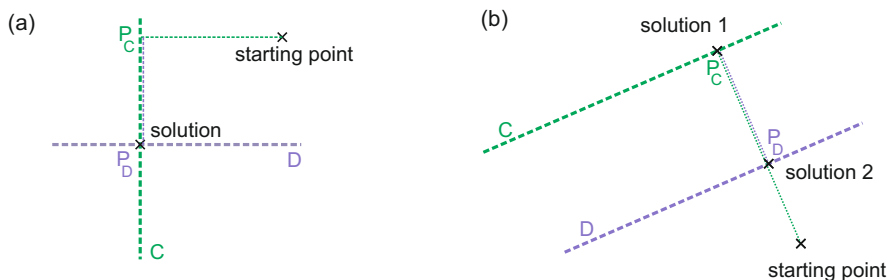
Figure 3.8: Alternating projections onto (a) intersecting, closed, orthogonal and convex subsets and (b) non-intersecting, closed and convex subsets. For case (a) the solution is found by successive projections onto the sets C and D. For case (b) no unique solution can be determined and the algorithm cycles between two points, one in C and one in D.

If $\boldsymbol{a}$ is projected onto D via P_D it results in

$$P_D(\boldsymbol{a}) = \tilde{\boldsymbol{x}}_C + \boldsymbol{x}_D + \boldsymbol{y}_D \ , \tag{3.268}$$

which is because of the distance minimizing properties of a projection (see equation 3.231). When P_C and P_D are applied in turns, as suggested by AP, this will result in

$$\mathrm{AP}(\boldsymbol{a}) = \tilde{\boldsymbol{x}}_C + \tilde{\boldsymbol{x}}_D + \boldsymbol{y}_C \ . \tag{3.269}$$

Further application of P_D on $\mathrm{AP}(\boldsymbol{a})$ will result in

$$P_D(\mathrm{AP}(\boldsymbol{a})) = \tilde{\boldsymbol{x}}_C + \tilde{\boldsymbol{x}}_D + \boldsymbol{y}_D \ .$$

For $\boldsymbol{y}_C$ equaling $\boldsymbol{y}_D$, the algorithm converges to the intersection of the subsets (Fig. 3.8 (a)). For $\boldsymbol{y}_C \neq \boldsymbol{y}_D$, i.e. in case that the two subsets C and D do not intersect, the map cycles between the points $\tilde{\boldsymbol{x}}_C + \tilde{\boldsymbol{x}}_D + \boldsymbol{y}_C$ and $\tilde{\boldsymbol{x}}_C + \tilde{\boldsymbol{x}}_D + \boldsymbol{y}_D$ of minimal separation $||\boldsymbol{y}_C - \boldsymbol{y}_D||$; stagnation occurs [48]. This is graphically depicted in Fig. 3.8 (b).

3.3.2   The difference map

Although the method of alternating projections is a very powerful, elegant and simple way of searching for intersections of constraint sets, it can easily lead to stagnation. To overcome this, in 2003, V. van Elser proposed an alternative algorithm to solve feasibility problems which is pulled away from local points of stagnation for an appropriate choice of parameters [48]. The method is called difference map. Since it found broad application in ptychographic phase retrieval, it will be explained in the following section. The difference map DM is defined as

$$\mathrm{DM} := Id + \beta \left[P_C \, r P_D - P_D \, r P_C \right] \ , \tag{3.270}$$

where Id is the identity mapping, $\beta \in \mathbb{R}$ is a constant and rP_i are the relaxed projections

$$rP_i := (1 + \gamma_i)P_i - \gamma_i Id, \ \ i = C, D, \ \gamma_i \in \mathbb{R} \ . \tag{3.271}$$
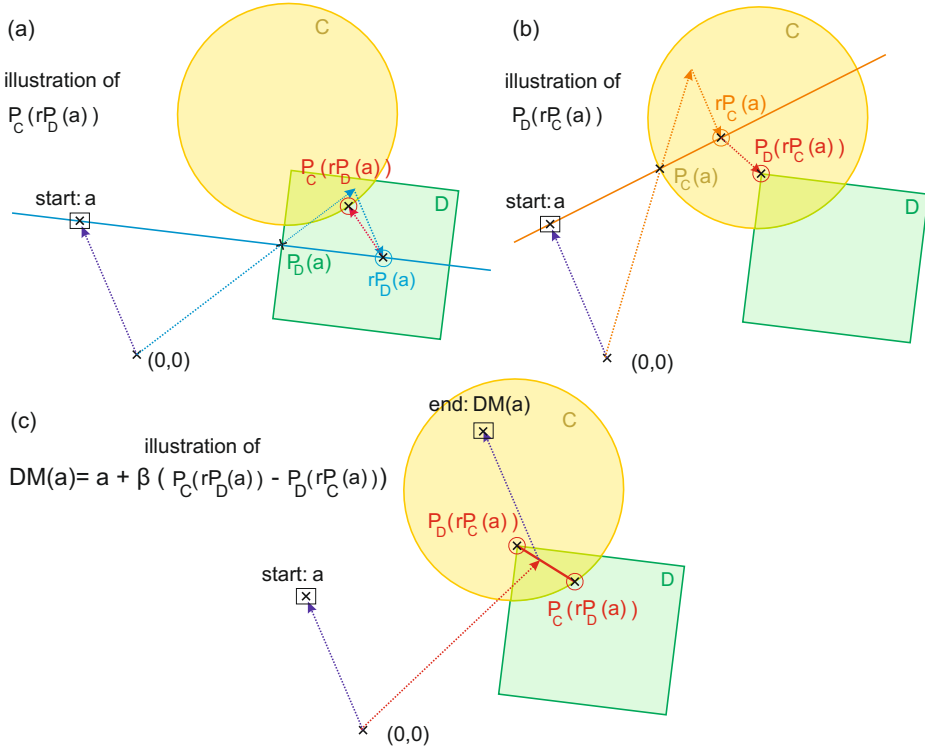
Figure 3.9: Step wise illustration of one complete iteration of the difference map starting at $\boldsymbol{a}$ and ending at DM($\boldsymbol{a}$). (a) Illustration of $P_C \, rP_D$. (b) Illustration of $P_D \, rP_C$. (c) Illustration of $Id + \beta \left[P_C \, r P_D - P_D \, r P_C \right]$ Further explanation can be found in the main text.

The operation DM searches for the intersection of two constraint sets. For this purpose, it combines two maps rP_i and two projections P_i. For illustration see Fig. 3.9: Two intersecting constraint sets (C, yellow circle and D, green rectangle) are drawn. The map rP_D (here with $\gamma_C = 0.5$) acts on $\boldsymbol{a}$ and results in the point $rP_D(\boldsymbol{a})$. This point is projected onto the set C. The result of the projection $P_C(rP_D(\boldsymbol{a}))$ is marked by a red circle (see Fig. 3.9(a)). The map rP_C (here with $\gamma_C = 0.5$) acts on $\boldsymbol{a}$ and results in the point $rP_C(\boldsymbol{a})$. This point is projected onto the set D. The result of the projection $P_D(rP_C(\boldsymbol{a}))$ is marked by a red circle (see Fig. 3.9(b)). The difference $||P_C(rP_D(\boldsymbol{a})) - P_D(rP_C(\boldsymbol{a}))||$ is the bold, red line in Fig. 3.9(c). Via the parameter β a point on this line is selected. The vector $\boldsymbol{a}$ is

added to the vector defining this point. The result is $\mathrm{DM}(\boldsymbol{a})$. One iteration of the difference map is complete.

The difference map stagnates for the point $\tilde{\boldsymbol{a}}$ that fulfills

$$P_C(rP_D(\tilde{\boldsymbol{a}})) - P_D(rP_C(\tilde{\boldsymbol{a}})) = 0 . \tag{3.272}$$

To find the corresponding point $\tilde{\boldsymbol{a}}_{C \cap D}$ in $C \cap D$, the map $P_D\, rP_C$ or $P_C\, rP_D$ has to be applied to the stagnation point $\tilde{\boldsymbol{a}}$:

$$P_C(rP_D(\tilde{\boldsymbol{a}})) = \tilde{\boldsymbol{a}}_{C \cap D} = P_D(rP_C(\tilde{\boldsymbol{a}})) . \tag{3.273}$$

To characterize the operation of DM, like in the previous section, the special case of two orthogonal, affine subspaces C and D defined as in equations (3.264) and (3.265) is observed once more. When DM is applied to a general point $\boldsymbol{a} = \boldsymbol{x}_C + \boldsymbol{x}_D + \boldsymbol{y}$, it results in [48]

$$
\begin{aligned}
\mathrm{DM}(\boldsymbol{x}_C + \boldsymbol{x}_D + \boldsymbol{y}) \quad = \quad & \tilde{\boldsymbol{x}}_C + \tilde{\boldsymbol{x}}_D + \boldsymbol{y} \\
+ \quad & (1 - \beta\gamma_D)(\boldsymbol{x}_C - \tilde{\boldsymbol{x}}_C) \\
+ \quad & (1 + \beta\gamma_C)(\boldsymbol{x}_D - \tilde{\boldsymbol{x}}_D) \\
+ \quad & \beta(\boldsymbol{y}_C - \boldsymbol{y}_D) .
\end{aligned}
\tag{3.274}
$$

Suppose C and D intersect and their intersection is not empty, i.e. $\boldsymbol{y}_C = \boldsymbol{y}_D = \tilde{\boldsymbol{y}}$. In this case DM is contractive for $(1 - \beta\gamma_D) \in\,]-1, 1[$ and $(1 + \beta\gamma_C) \in\,]-1, 1[$. The optimal choice of parameters would be $\gamma_D = \beta^{-1}$ and $\gamma_C = -\beta^{-1}$ [48]. Successive application of DM allows finding the fixed points of DM [48]

$$\mathrm{Fix}\,\mathrm{DM} = \{\boldsymbol{x} : \boldsymbol{x} = \tilde{\boldsymbol{x}}_C + \tilde{\boldsymbol{x}}_D + \boldsymbol{y}, \ \forall\, \boldsymbol{y} \in Y\} . \tag{3.275}$$

Fixed points of DM are not yet in $C \cap D$. The application of $P_C\, rP_D$ or $P_D\, rP_C$ to these fixed points leads to the solution

$$\boldsymbol{x}_{C \cap D} = \tilde{\boldsymbol{x}}_C + \tilde{\boldsymbol{x}}_D + \tilde{\boldsymbol{y}} , \tag{3.276}$$

which is in the intersection $C \cap D$. This is graphically demonstrated in Fig. 3.10(a). The dashed green and violet lines are the orthogonal, affine subspaces C and D which intersect in the point $\tilde{\boldsymbol{x}}_C + \tilde{\boldsymbol{x}}_D + \tilde{\boldsymbol{y}}$. The orange line is the set of fixed points of DM. By the performance of $P_C\, rP_D$ or $P_D\, rP_C$, these fixed points are all mapped onto $\tilde{\boldsymbol{x}}_C + \tilde{\boldsymbol{x}}_D + \tilde{\boldsymbol{y}}$.

Suppose that C and D do not intersect, i.e. $\boldsymbol{y}_C \neq \boldsymbol{y}_D$. For $\gamma_D = \beta^{-1}$ and $\gamma_C = -\beta^{-1}$, iterative application of DM leads to [48]

$$
\begin{aligned}
\mathrm{DM}^n(\boldsymbol{x}_C + \boldsymbol{x}_D + \boldsymbol{y}) \quad = \quad & \tilde{\boldsymbol{x}}_C + \tilde{\boldsymbol{x}}_D + \boldsymbol{y} \\
+ \quad & n\beta(\boldsymbol{y}_C - \boldsymbol{y}_D).
\end{aligned}
\tag{3.277}
$$

Successive iterations of DM hence move the output away from points of stagnation. In Fig. 3.10(b) this behavior is illustrated by the orange line and arrows pointing away from elements of minimal separation between C and D.
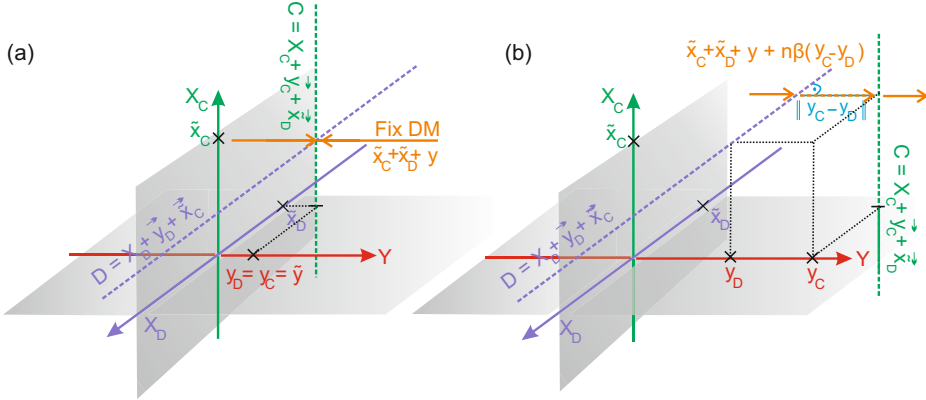
Figure 3.10: (a) Illustration of the fixed point set (orange line) of DM for two intersecting, orthogonal constraint sets. By the action of $P_C\, r P_D$ or $P_D\, r P_C$ these fixed points are mapped to $C \cap D$. (b) Where algorithms like alternating projections would stagnate, the solution of DM is pulled away from points of minimal separation between C and D.

A further note on the difference-map algorithm: The optimal choice of the relaxation parameters γ_C and γ_D turned out to be $\gamma_C = \frac{1}{\beta}$ and $\gamma_D = \frac{1}{\beta}$ [48]. Yet, when the outer projection/reflection is onto a linear subspace, for this choice of parameters and for $\beta = 1$, the difference map transforms to [62]

$$\tilde{\mathrm{DM}} = Id + [P_C\,(2P_D \quad Id) \quad P_D] = P_C\,(2P_D \quad Id) + (Id \quad P_D)\,. \qquad (3.278)$$

This is identical to the much older Douglas-Rachford algorithm, for which convergence in the convex setting is known [53, 62]. Yet, since this work will concentrate on ptychographic phase retrieval and since one of the famous ptychgraphic algorithms is referred to as difference map ptychography (with the specific parameter setting of $\gamma_D = \beta^{-1}$, $\gamma_C = \beta^{-1}$ and $\beta = 1$), it shall still be denoted as difference map rather than as Douglas-Rachford.

3.4   Selected projections for phase retrieval

Phase retrieval is a method to solve the phase problem described in section 2.5 by finding the complex valued envelope $\psi_i(x, y)$ of a paraxial and monochromatic wave function $\Psi_i(x, y, z, t) = \psi(x, y)\exp(ikz)\exp(-i\omega t)$. Now, the envelope $\psi(x, y)$ shall get a lower index i ($\psi_i(x, y)$) which refers to the z coordinate oriented along the axis of paraxial propagation. Hence, a wave front $\psi_i(x, y)$ is a complex valued two dimensional function at fixed longitudinal position i. By propagating $\psi_i(x, y)$ forwards and backwards between two (or more) positions i, j along the optical axis and by applying boundary conditions to $\psi_i(x, y)$ which are characteristic for the wave field at position i and boundary conditions to $\psi_j(x, y)$ which are characteristic for the wave field at position j, a wave field should be found that is consistent

with all of the imposed boundary conditions. Typically these boundary conditions are the agreement with the measured data in the observation plane j and additional knowledge about the optical setup or the imaged object in the starting plane i.

The connection to the iterative projection methods described before is the following: The named boundary conditions are imposed onto $\psi_i(x, y)$ and $\psi_j(x, y)$ by projection operators. These projection operators modify $\psi_i(x, y)$ and $\psi_j(x, y)$ by minimum changes, such that the new $\psi_i(x, y)$ and $\psi_j(x, y)$ are consistent with the respective boundary conditions. Projections are chosen to find fixed points of the phase retrieval mapping. Yet, therefore, the constraint sets underlying the projection operators have to be closed and convex. In particular, when these sets are not convex, projections are multi-valued (see Fig. 3.5(a)) which is an obstacle to prove convergence of the respective projection algorithms [54]. In this section the properties of projections used for phase retrieval are explored; the geometry of their underlying subsets is explained. Extensive literature and a detailed analysis of this subject can be found in [54].

Beforehand a few words on notation: Since in the following not only projections, but their iterative application, as well as wave fields and objects at different longitudinal and lateral positions will be of interest, several indices will be necessary to fully characterize the involved complex valued functions. Lower indices i will always refer to the longitudinal position (z-direction, i.e. direction of propagation) of a wave field. Upper indices in round brackets (i) indicate the lateral position of object or probe. Upper indices which are not in brackets denote the current iteration n or k. Despite these many indices, to keep notation short, lateral coordinates $\boldsymbol{r} = (x, y)$ are not always explicitly mentioned, but the dependence of the object, the illumination or probe and the exit wave on these coordinates is implicitly assumed.

The modulus or magnitude constraint:
Reconstructed intensities have to match measured intensities

Among all discrete wave fields $\psi_i(x, y)$, there is a subset M of discrete wave fields $\psi_{i,\mathrm{mod}}(x, y)$ that matches the measured (and – with this in mind – discrete) intensity distribution. The best approximation $P_M(\psi_i(x, y))$ of $\psi_i(x, y)$ is found by

$$P_M(\psi_i) := \mathcal{P}_{-\Delta_{i,j}} \left[\frac{\mathcal{P}_{\Delta_{i,j}}[\psi_i]}{|\mathcal{P}_{\Delta_{i,j}}[\psi_i]|} \cdot \sqrt{I_j} \right] , \qquad (3.279)$$

where I_j is the measured intensity distribution at defocus or z-distance j. The operator $\mathcal{P}_{\Delta_{i,j}}$ is either the Fourier transform $\mathcal{F}$ or the Fresnel transform $\mathcal{D}_{\Delta_{i,j}}$ referring to the propagation of light in the optical far- or near-field (see section 2.3.2).

Considering a single pixel at (x, y), the geometry of the modulus constraint is the boundary line of a circle with radius $\sqrt{I_j(x, y)}$ [48, 53]: Phase values from $[-\pi, \pi[$ can be combined with $\sqrt{I_j(x, y)}$, see Fig. 3.11(a). The projection onto a circle is unique with the exception of one single point: The center of the circle corresponds
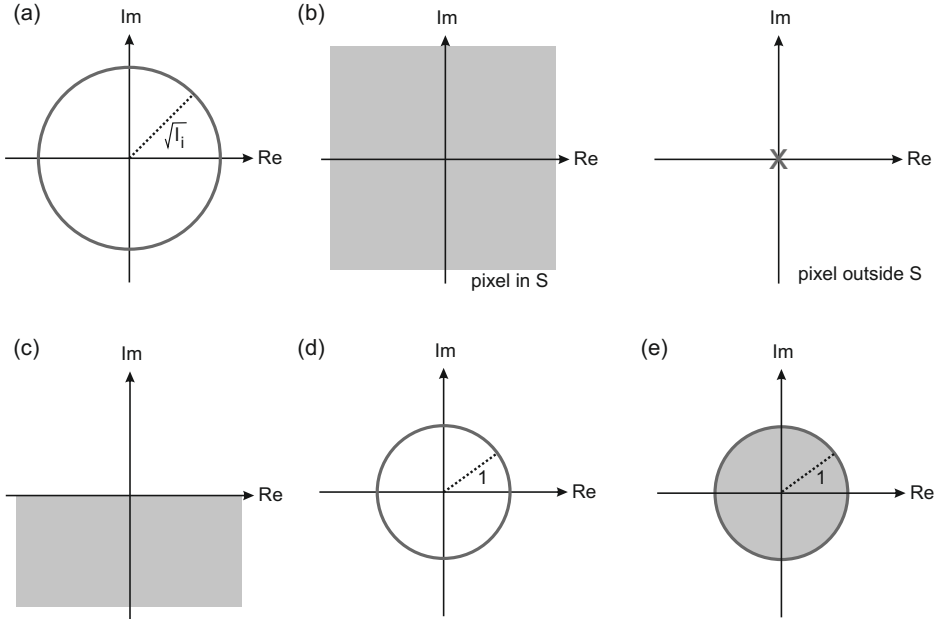
Figure 3.11: Projections for phase retrieval and the geometry of their underlying sub-
sets. (a) The modulus constraint replaces the reconstructed intensities in each pixel
by the measured ones and keeps the reconstructed phases. The underlying subset
for each pixel is a ring of radius $\sqrt{I_j(x,y)}$. (b) The support constraint sets pixels
outside σ to zero. Pixels inside σ are allowed to have any value in the complex plane.
(c) In the case of X-rays, objects have negative phase shift. Functions with negative
phase shift are elements of the lower half of the complex plane. (d) Pure phase ob-
jects have full transmission with amplitude equal to one. Phases are left untouched.
The geometry of this set is the circumference of the unit circle in the complex plane.
(e) Transmission cannot exceed the value of one. Complex functions fulfilling this
constraint are elements of the unit circle in the complex plane.

to a zero valued amplitude and is projected onto each and every point of the circle's
boundary. The circumference of a circle is not convex. This means that conver-
gence and in particular convergence to a unique solution cannot be guaranteed [54].

The support constraint: Objects limited to a finite support
The measured intensity is not the only boundary condition that can be deduced
from the experimental configuration. If the object is restricted to a certain area
(or in discrete formulation to a certain number of image points), the so called
support projection acting onto each image point can be formulated as

$$P_S(\psi_i(x,y)) := \begin{cases} \psi_i(x,y) & \text{if } (x,y) \in \sigma \ , \\ 0 & \text{else,} \end{cases} \tag{3.280}$$

where S is the set of all complex valued functions which are zero or non-zero within and zero outside their support σ. Since S is a linear subspace [48], it contains all linear combinations of its elements as well. As it contains all linear combinations of its elements, it contains also affine combinations of its elements and in particular complex combinations. It follows that S is convex [53]. The projection onto S for a single pixel is the projection onto the complex valued plane in case that the specific pixel is inside σ and zero in case that it is outside σ, see Fig. 3.11(b).

The negativity constraint: Objects with negative phase shift
A projection onto transmission functions $o(x, y)$ of small negative phase shift can be formulated as

$$P_N(o(x,y)) := \begin{cases} o(x,y) & \text{if } \arg(o(x,y)) \leq 0 \; , \\ |o(x,y)| & \text{else.} \end{cases} \tag{3.281}$$

The underlying subset of P_N is the lower half of the complex plane and hence convex, see Fig. 3.11(c). This projection makes sense for thin, weak objects with a complex index of refraction of $n = 1 - \delta + i\beta$, where δ is positive and $\delta << 1$, i.e. objects imaged by X-rays.

The phase object constraint: Objects with no absorption
Transparent objects only shift the phase of the incoming light. They (approximately) do not absorb any photons. Consequently the absolute value of their transmission function is always one:

$$P_P(o(x,y)) := \frac{o(x,y)}{|o(x,y)|} \; . \tag{3.282}$$

The operator $P_P(o(x,y))$ projects onto a ring with radius one in the complex valued plane. Similar to the magnitude constraint, its underlying subset is of non-convex geometry. For illustration, see Fig. 3.11(d).

The transmission constraint:
Objects with full or less than full transmission
The objects considered here can only absorb (and, of course, shift the phase), but they cannot generate additional photons with the same wavelength like the illuminating beam. The absolute value of the transmission function cannot exceed the value of one. A projection onto the set of functions with absolute value lower than one can be expressed as

$$P_T(o(x,y)) := \begin{cases} o(x,y) & \text{if } |o(x,y)| \leq 1 \; , \\ \frac{o(x,y)}{|o(x,y)|} & \text{else.} \end{cases} \tag{3.283}$$

The corresponding set of functions is of convex nature: It contains all functions with absolute value inside the unit circle in the complex plane, see Fig. 3.11(e).

3.5   The Gerchberg-Saxton algorithm

The Gerchberg-Saxton map is the most intuitive, iterative phase retrieval algorithm. It was developed by R. W. Gerchberg and W. O. Saxton, published in 1972 and introduced as a practical algorithm to determine phases from intensity measurements in the diffraction and the image plane [63]. It is an example of von Neumann's alternating projections applied to phase retrieval [53, 60]. Until today, it has found a broad range of applications [64].

The two constraints chosen for the Gerchberg-Saxton algorithm are two modulus constraints (see equation 3.279): one in the image plane (real space, longitudinal position i) and one in the diffraction plane (Fourier space, longitudinal position j). Hence, two intensity recordings at different positions along the optical axis (one in the far-field, and one at the position of the sample) are required. The corresponding subsets shall be called M_{ff} and M_{nf}. Projections onto these sets are performed by

$$P_{M_{\mathrm{ff}}}(\psi_i) \quad := \quad \mathcal{F}^{-1}\left[\frac{\mathcal{F}[\psi_i]}{|\mathcal{F}[\psi_i]|} \cdot \sqrt{I_j}\right] \ , \tag{3.284}$$

$$P_{M_{\mathrm{nf}}}(\psi_i) \quad := \quad \frac{\psi_i}{|\psi_i|} \cdot \sqrt{I_i} \ , \tag{3.285}$$

where I_j are the measured intensities in the Fourier domain and I_i the measured intensities in the image plane. Using this terminology, the Gerchberg-Saxton map is

$$\psi_i^{k+1} = \mathrm{GS}(\psi_i^k) := P_{M_{\mathrm{ff}}}(P_{M_{\mathrm{nf}}}(\psi_i^k)) \ , \tag{3.286}$$

where k denotes current iteration of the map.

3.6   The error reduction algorithm

In 1982, J. R. Fienup generalized the Gerchberg-Saxton map from its special application for imaging using one intensity recording in the image plane and one in the optical far-field to a larger class of problems. These problems are constrained by at least two different a priori known or measurable boundary conditions [65]. He named this kind of algorithm 'error reduction' , because he showed that the difference between successive iterates in real space and in Fourier space decreased or stagnated when applying the error reduction map in an iterative way [65].

For the special problem of phase retrieval from measured diffraction intensities, J. R. Fienup made use of the modulus constraint $P_{M_{\mathrm{ff}}}$ in the observation plane (see equation 3.279) and a real space support constraint P_S in the source plane (see equation 3.280) combined to an alternating projections algorithm:

$$\psi_i^{k+1} = \mathrm{ER}(\psi_i^k) := P_S(P_{M_{\mathrm{ff}}}(\psi_i^k)) \ . \tag{3.287}$$

Since both of these very early phase retrieval algorithms (Gerchberg-Saxton and error reduction method) are based on alternating projections, they show all the advantages and disadvantages of this technique. In case that the underlying subsets
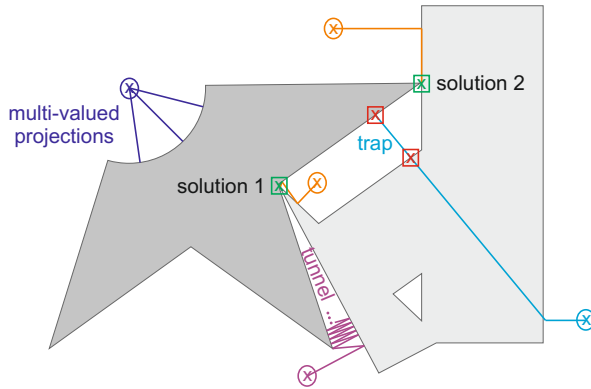
Figure 3.12: Problems of alternating projections onto subsets of arbitrary, closed geometry. Dark blue: Projections might not be unique. Orange: The initial guess determines the solution. Violet: Extremely slow convergence can occur when successive iterates are entering a tunnel. Light blue: Endless cycling between elements that are not in the intersection of the two subsets is possible, the algorithm is trapped.

of the used projectors are closed, convex and intersecting, the method of alternating projections converges to an element within this intersection [61]. We saw that not all of the subsets used in phase retrieval are of closed and convex geometry. Especially the subset underlying the important magnitude constraint is not convex. Furthermore, since the measured intensities are necessarily bounded by a compact support (namely the field of view of the detector), they actually cannot result from a confined object in real space. The same is true for holographic intensity distributions: Strictly speaking, the intersection of the two subsets (support and magnitude constraint) does not exist.

In particular for two subsets of arbitrary geometry, the following can happen [60]: (1) Projections might be multi-valued (see dark blue paths in Fig. 3.12). (2) Multiple points of intersections can exist. The choice of the initial guess determines the solution (an illustration is provided in Fig. 3.12, orange paths). (3) During successive iterations the algorithm might get trapped between two elements which are not in the intersection of the constraint sets, but at minimal distance with respect to each other (red points in Fig. 3.12 and light blue path). (4) The algorithm might get stuck in a tunnel, i.e. the change from one iteration to the next is negligible (see Fig. 3.12, violet path). While case (3) cannot occur for intersecting, convex sets, case (4) can happen for convex sets as well [60]. Especially due to traps, the error reduction and the Gerchberg-Saxton algorithm can produce poor solutions.

J. R. Fienup and C. C. Wackerman observed three typical modes of stagnation for the special case of phase retrieval based on alternating projections [66]. The algorithm stagnates when (1) its output contains features of the image and the twin image simultaneously, (2) when the reconstruction is superimposed by stripes of low contrast or when (3) unintentional truncation of the object occurs caused

by a misaligned support.

Twin images $\psi_i^*(\;x,\;y)$ of $\psi_i(x,y)$ typically spoil those reconstructed objects that have a centrosymmetric support. The problem can be overcome by dynamically modulating the support in particular by starting with a non-symmetric support [66]. Similarly unintentional truncation (due to a misaligned or too restrictive support) can be avoided by shifting the support during reconstruction [66]. All of these methods however do not touch the underlying concept of alternating projections. A different approach to get rid of the described problems would be to use phase retrieval methods including different combinations of projections and relaxation techniques and/or to make use of more than one intensity recording in the image or diffraction plane with slightly different information content.

3.7   Ptychography in the optical far-field

In the last ten years, an imaging method became famous that uses multiple intensity recordings in the diffraction plane generated by shifting an object through a confined beam. It is called ptychography and despite that the underlying concepts date back to the year of 1968 [67], first application in X-ray coherent diffractive imaging was roughly forty years later.
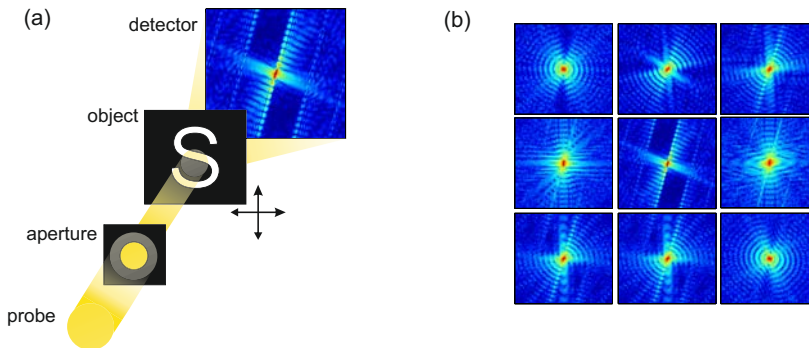
Figure 3.13: Illustration of Ptychography. (a) An object is shifted inside a confined, coherent beam denoted probe. Intensities are recorded in the optical far-field. (b) For each position of the object, one diffraction image is recorded. Using these images phase retrieval can be performed.

Basically, ptychography is the same as scanning coherent diffractive X-ray imaging with small step sizes: An (extended) object of interest is illuminated by a coherent X-ray beam confined by some aperture. The intensity distribution of the transmitting beam is measured in the optical far-field. The beam is scanned to several positions illuminating different regions of the object, all partially overlapping with the previous ones. Equivalently the object can be scanned through a fixed beam. An illustration is depicted in Fig. 3.13(a). For each position of the beam (or the

object) one image is recorded in the diffraction plane (see Fig. 3.13(b)). It is worth to notice that ptychography is not restricted to the use of X-rays: Also coherent visible light can be utilized [68]. However, ptychography is particularly useful for X-ray microscopy, as it does not necessarily require focusing lenses, which are hard to fabricate for this spectral range.

3.7.1   The ptychographical iterative engine

In 2004, the first ptychographic algorithm for X-ray imaging was designed by J. M. Rodenburg and H. M. L. Faulkner [68, 69]. It is called PIE (ptychographical iterative engine) and conceived to find a single transmission function of an object, that together with a known illumination reproduces all of the measured diffraction images of the partially illuminated object.

In this early states of ptychographic algorithm development, it was essential to fully characterize the probe $p_j(x, y)$ in amplitude and phase before trying to reconstruct the object $o(x, y)$ which had been scanned through the beam at defocus position j. The longitudinal position j of the object is not changed. With this in mind the lower index j is skipped in the following, however p and o still refer to object and probe at the same, constant defocus position j. Instead of scanning the object, the PIE historically is designed such that the probe is scanned to $i = [1, 2, ...m]$ positions across the object. The probe at position $\boldsymbol{r}^{(i)} = (x^{(i)}, y^{(i)})$ with respect to the object $o(x, y) = o(\boldsymbol{r})$ is

$$p^{(i)}(\boldsymbol{r}) = p(\boldsymbol{r} - \boldsymbol{r}^{(i)}) \; , \tag{3.288}$$

and the exit wave $\psi^{(i)}(x, y) = \psi^{(i)}(\boldsymbol{r})$ following the projection approximation is

$$\psi^{(i)}(\boldsymbol{r}) = o(\boldsymbol{r}) \, p(\boldsymbol{r} - \boldsymbol{r}^{(i)}) = o(\boldsymbol{r}) \, p^{(i)}(\boldsymbol{r}) \; . \tag{3.289}$$

The PIE algorithm [68, 69] is

$$o^{n+1}(\boldsymbol{r}) = o^n(\boldsymbol{r}) + \beta w \frac{\left(p^{(i)}\right)^*(\boldsymbol{r})}{|p^{(i)}(\boldsymbol{r})|^2 + \alpha} \left[P_{M_{\mathrm{ff}}}\left(\psi^{(i),n}(\boldsymbol{r})\right) - \psi^{(i),n}(\boldsymbol{r}) \right] \; . \tag{3.290}$$

The factor w is a weighting factor

$$w = \frac{|p^{(i)}(\boldsymbol{r})|}{\max(|p^{(i)}(\boldsymbol{r})|)} \; . \tag{3.291}$$

The purpose of w is to suppress noise; β is a relaxation coefficient and α should prevent division by zero. The PIE algorithm acts sequentially: It adapts the reconstructed intensities to the measured ones for one single lateral position i of the probe and calculates an update for the object. Then it takes the corrected object field function, shifts the probe to the next lateral position j and again adapts the intensities. One cycle k of PIE is complete when the object has been shifted once to every position $i = [1, 2, ...m]$.

Setting the weighting factor w to $w = 1$, and choosing a spatially confined probe within a region of radius σ

$$p(x,y) := \begin{cases} 1 & \text{if } \sqrt{x^2 + y^2} \leq \sigma \ , \\ 0 & \text{else} \ , \end{cases} \tag{3.292}$$

the factor $\frac{\left(p^{(i)}\right)^*(\boldsymbol{r})}{|p^{(i)}(\boldsymbol{r})|^2 + \alpha}$ is nothing else but a support constraint with the support shifted to position i [69]. Further simplification would be to take only one single lateral position into account. Since the probe is confined to a region of radius σ, the reconstructed object will be limited to this area as well. With the definition of the probe by equation 3.292 and an object confined to the circular region with radius σ, the following holds

$$\psi(\boldsymbol{r}) = p(\boldsymbol{r})\, o(\boldsymbol{r}) = P_S(o(\boldsymbol{r})) = o(\boldsymbol{r}) \ . \tag{3.293}$$

In this case, the PIE scheme of equation 3.290 reduces to

$$\psi^{n+1}(\boldsymbol{r}) = (1 + \gamma)P_S\left[P_{M_\text{ff}}\left(\psi^n(\boldsymbol{r})\right)\right] - \gamma\psi^n(\boldsymbol{r}) \ . \tag{3.294}$$

where $\gamma = \beta - 1$. The performance of a relaxed operator (a relaxed projection) was illustrated in section 3.2.3 and Fig. 3.6: Equation 3.294 is a line connecting the starting guess for ψ^n with a new estimate $P_S\left(P_{M_\text{ff}}(\psi^{(n)})\right)$. Choosing $\gamma = 0$ will select the new estimate as update for ψ^n. Choosing $\gamma = -1$ will match the output onto the input and will not change ψ^n at all.

Equation 3.294 can be interpreted as a relaxed version of the error reduction algorithm. For $\gamma = 0$, it is exactly the error reduction algorithm [69]. However, taking different positions of the confined probe into account breaks the symmetry imposed by an (in most cases symmetric) support. Hence, even for a uniform probe as defined in (3.292) within a circular region, the PIE shows considerable advantages compared to the error reduction method with the object located at one single position [69, 70]; it especially does not run into stagnation caused by twin image artifacts, as detailed later. Yet, the main drawbacks of the PIE method are: (1) It requires a fully characterized probe in amplitude and phase; (2) it requires accurate knowledge about the positions of the object (or the probe) [70].

3.7.2 Ptychography with the difference map

The main drawback of the PIE is the need for accurate knowledge of the probe. The first algorithm to overcome this restriction was published by M. Guizar-Sicairos and J. R. Fienup in May 2008. It is based on a nonlinear optimization approach using the conjugate gradient method [70]. For the first time, it was possible to simultaneously reconstruct object and probe. In addition, inaccuracies in the shifts of the object could be corrected as well.

Shortly after the non-linear optimization approach described in [70], in July 2008, P. Thibault *et. al.* [71] published a method based on the difference map DM introduced in section 3.3.2 that is also capable of reconstructing object *and* probe.

However, in contrast to [70], still accurate knowledge about the translations imposed on the object is required.

Recall the definition of the difference map DM:

$$\text{DM} := Id + \beta \left[P_C \, rP_D - P_D \, rP_C \right] \; , \tag{3.295}$$

where Id is the identity mapping, $\beta \in \mathbb{R}$ is a constant and rP_i are the relaxed projections

$$rP_i := (1 + \gamma_i)P_i - \gamma_i Id, \; i = C, D, \; \gamma_i \in \mathbb{R} \; .$$

In the special case of ptychography, P_D is the projection $P_{M_{\text{ff}}}$ onto the set of functions M_{ff} that reproduce the m measured intensities I_i ($i = [1, 2, ...m]$) in the diffraction plane. The second constraint set is the set O of functions

$$\hat{\psi}(\boldsymbol{r}) = \hat{o}(\boldsymbol{r}) \cdot \hat{p}^{(i)}(\boldsymbol{r}) \; , \tag{3.296}$$

which minimize the objective function

$$F := \sum_{i=1}^{m} \sum_{x,y}^{M,N} |\Psi - \hat{\Psi}|^2 \; , \tag{3.297}$$

where M and N are the number of samples/pixels in horizontal and vertical direction and where

$$\Psi \quad := \quad \left\{ \psi^{(1)}(\boldsymbol{r}), \; \psi^{(2)}(\boldsymbol{r}), ..., \psi^{(m)}(\boldsymbol{r}) \right\} \; \text{and} \tag{3.298}$$

$$\hat{\Psi} \quad := \quad \left\{ \hat{\psi}^{(1)}(\boldsymbol{r}), \; \hat{\psi}^{(2)}(\boldsymbol{r}), ..., \hat{\psi}^{(m)}(\boldsymbol{r}) \right\} \; . \tag{3.299}$$

The single $\psi^{(i)}(\boldsymbol{r})$ are the current iterates for the exit wave at position i. Minimization of F can be achieved by setting the derivative of F with respect to $\hat{p}^{(i)}(\boldsymbol{r})$ and $\hat{o}(\boldsymbol{r})$ to zero. It results in two coupled equations for object and probe [72]:

$$\hat{o}(x,y) \quad = \quad \frac{\sum_{i=1}^{m} \hat{p}^*(\boldsymbol{r} - \boldsymbol{r}^{(i)}) \psi^{(i)}(\boldsymbol{r})}{\sum_{i=1}^{m} |\hat{p}(\boldsymbol{r} - \boldsymbol{r}^{(i)})|^2} \; , \tag{3.300}$$

$$\hat{p}(x,y) \quad = \quad \frac{\sum_{i=1}^{m} \hat{o}^*(\boldsymbol{r} + \boldsymbol{r}^{(i)}) \psi^{(j)}(\boldsymbol{r} + \boldsymbol{r}^{(i)})}{\sum_{i=1}^{m} |\hat{o}(\boldsymbol{r} + \boldsymbol{r}^{(i)})|^2} \; . \tag{3.301}$$

The operator pointing onto the set O cannot exactly be expressed as a projector [73]. For this reason, it shall be named Π_O instead of P_O. The wave function $\hat{\psi}(\boldsymbol{r})$ shall be separable in object and probe, hence we denote Π_O as separability constraint.

When replacing P_C by Π_O and P_D by $P_{M_{\text{ff}}}$, equation (3.295) is

$$\text{DM}_{\text{pty}} = Id + \beta \left[\Pi_O \, rP_{M_{\text{ff}}} - P_{M_{\text{ff}}} \, r\Pi_O \right] \quad =$$
$$Id + \beta \left\{ \Pi_O \left[(1 + \gamma_{M_{\text{ff}}}) P_{M_{\text{ff}}} - \gamma_{M_{\text{ff}}} Id \right] - P_{M_{\text{ff}}} \left[(1 + \gamma_{\text{o}}) \Pi_O - \gamma_{\text{o}} Id \right] \right\} \quad =$$
$$Id + \beta(1 + \gamma_{M_{\text{ff}}}) \Pi_O \, P_{M_{\text{ff}}} - \beta \gamma_{M_{\text{ff}}} \Pi_O - \beta(1 + \gamma_{\text{o}}) P_{M_{\text{ff}}} \, \Pi_O + \beta \gamma_{\text{o}} P_{M_{\text{ff}}} \; . \tag{3.302}$$
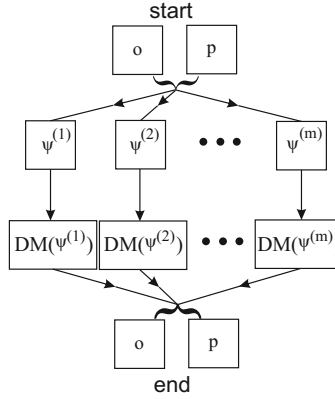
Figure 3.14: One full cycle of pPIE or difference map based ptychography.

Choosing $\beta = \gamma_{M_{\text{ff}}} = -1$, and $\gamma_o = 1$ leads to the mapping provided in [72]:

$$\text{DM}_{\text{pty}}(\Psi^k) = \Psi^{k+1} = \Psi^k + P_{M_{\text{ff}}}\left[2\Pi_O(\Psi^k) - \Psi^k\right] - \Pi_O(\Psi^k) , \qquad (3.303)$$

where upper indices refer to the current/the next iteration. Note that this algorithm is performing in a parallel way: In contrast to the PIE algorithm, DM_{pty} addresses the single elements of Ψ^k separately, i.e. the update $\psi^{(i),k}$ does not influence the update of $\psi^{(j),k}$ with $i \neq j$ (see Fig. 3.14). For this reason the difference map ptychography is also called pPIE (parallel ptychographical iterative engine) [74].

3.7.3 The extended ptychographical iterative engine

Some months after publication of the nonlinear optimization and the difference map approaches to simultaneously reconstruct object and probe, A. M. Maiden and J. M. Rodenburg provided an extension of the PIE algorithm, capable of retrieving information of both, object and illumination at the same time. This extended ptychographical iterative engine (ePIE) consists of the two maps ePIE_o (update of o) and ePIE_p (update of p) with reversed roles of object and probe [74]:

$$
\begin{aligned}
\text{ePIE}_o(o^n(\boldsymbol{r})) &= o^{n+1}(\boldsymbol{r}) \\
&= o^n(\boldsymbol{r}) + \alpha \frac{\left(p^n(\boldsymbol{r} - \boldsymbol{r}^{(i)})\right)^*}{\max(|p^n(\boldsymbol{r} - \boldsymbol{r}^{(i)})|^2)} \\
&\quad \cdot \left[P_{M_{\text{ff}}}\left(\psi^{(i),n}(\boldsymbol{r})\right) - \psi^{(i),n}(\boldsymbol{r})\right] ,
\end{aligned} \qquad (3.304)
$$

$$\begin{aligned}
\text{ePIE}_p(p^n(\boldsymbol{r})) &= p^{n+1}(\boldsymbol{r}) \\
&= p^n(\boldsymbol{r}) + \beta \frac{\left(o^n(\boldsymbol{r}+\boldsymbol{r}^{(i)})\right)^*}{\max(|o^n(\boldsymbol{r}+\boldsymbol{r}^{(i)})|^2)} \\
&\quad \cdot \left[P_{M_{\text{ff}}}\left(\psi^{(i),n}(\boldsymbol{r})\right) - \psi^{(i),n}(\boldsymbol{r})\right] ,
\end{aligned} \tag{3.305}$$

where the parameters $\alpha, \beta \in \mathbb{R}$ adjust the step size of the update and where the exit wave for the probe at position i is described by

$$\psi^{(i),n}(\boldsymbol{r}) = o^n(\boldsymbol{r})\, p^n(\boldsymbol{r}-\boldsymbol{r}^{(i)}) . \tag{3.306}$$

One full iteration k is complete, when $n = m$, where m is the number of recorded diffraction patterns. After that, the following iteration starts with (sub-)index $n = 1$ addressing the first (randomly chosen) diffraction pattern again. It was observed that this method is more robust to noise in the diffraction images and provides faster convergence than the difference map approach [74]. As the PIE method, the ePIE operates sequentially: The updates of object and probe at position i in sub-iteration n will influence the updates of object and probe at position $j > i$. A graphical visualization is depicted in Fig. 3.15. Whereas single mappings inside one iteration of the pPIE (see Fig. 3.14) have no knowledge about each other and hence can be performed in parallel, the ePIE intertwines its single sub-iterates within one global iteration (see Fig. 3.15).
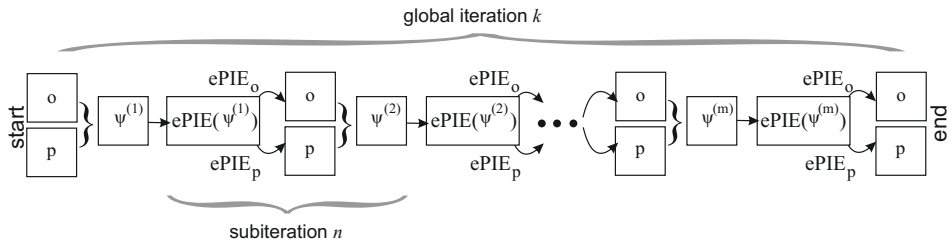
Figure 3.15: One full cycle of ePIE.

3.7.4 Lateral diversity as a solution to the twin image problem in far-field imaging

Following the Frauenhofer approximation, the intensities measured in the diffraction plane are the squared modulus of the Fourier transform of the signal $\psi(\boldsymbol{r})$ in the image plane. Unfortunately, there are two, symmetric signals $\psi(\boldsymbol{r})$ and $\psi^*(-\boldsymbol{r})$ (the so called twin signal) that produce the same far-field intensity (see section 2.5.1). Both, $\psi(\boldsymbol{r})$ and $\psi^*(-\boldsymbol{r})$ fulfill the boundary conditions of the magnitude constraint correctly. If in addition the support of $\psi(\boldsymbol{r})$ is centrosymmetric about the origin, then the support of $\psi(\boldsymbol{r})$ is the same as the support of $\psi^*(-\boldsymbol{r})$. Consequently, both functions $\psi(\boldsymbol{r})$ and $\psi^*(-\boldsymbol{r})$ also fulfill the boundary conditions

imposed by the support constraint [70]. They are both true solutions of the phase retrieval problem. When during the first iterations of an alternating projections algorithm image and twin are simultaneously reconstructed, stagnation occurs [66].
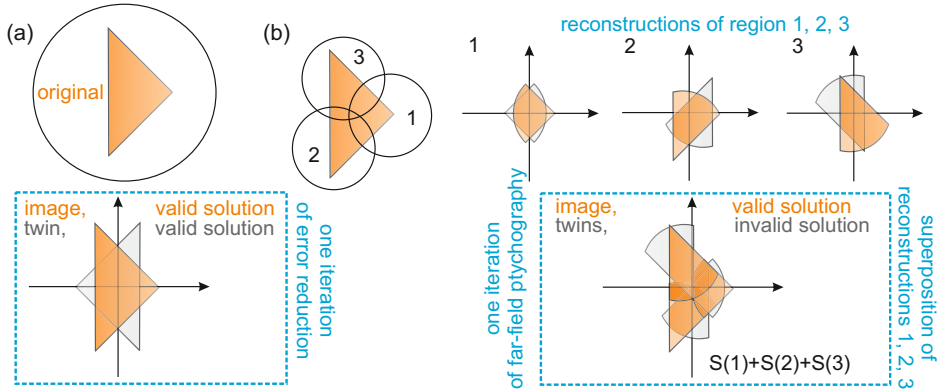
Figure 3.16: (a) The error reduction cannot distinguish between the image (orange triangle) and its twin (gray triangle), when an object is fully illuminated inside a centrosymmetric area. Reconstructions are likely to stagnate in mixtures of both - image and twin image. (b) Illuminating different, slightly overlapping regions (1, 2, 3) of the object, reconstructing them separately, shifting them by the known translations S and superimposing these shifted single reconstructions (S(1), S(2), S(3)) will assemble the image in the correct and the twin image in an incorrect way. Whereas in situation (a) further iterations of phase retrieval will stagnate, in situation (b) the solution corresponding to the image is favored.

A non-symmetric support can be used to avoid reconstructions showing features of $\psi(\boldsymbol{r})$ and $\psi^*(\boldsymbol{-r})$ [66]. But this is not the only way how to get rid of the twin image. One reason why far-field ptychography is so powerful is that translational diversity in the recorded diffraction patterns reduces twin image artifacts and after some iterations pulls the solution towards $\psi(\boldsymbol{r})$ [70]. In Fig. 3.16(a) a sketch is provided to demonstrate the result of a simple step of error reduction. An object (here the orange triangle) is fully exposed within the region of the circle. Performing one step of error reduction will mix the image (lower orange triangle) and its twin (gray triangle), each for itself a valid solution. Restricting the illuminated area and exposing several regions (regions 1, 2 and 3 in Fig. 3.16(b)) will result in partial reconstructions, all mixed by their twins and all centered (as the magnitude of the Fourier transform is insensitive to spatial shifts). Yet, with the information at hand, where regions 1, 2, and 3 are located with respect to the origin, it is possible to add up the shifted, partial reconstructions in the correct way (blue box in Fig. 3.16(b)). The image is restored correctly while the twin image parts are not assembled appropriately [70]. Using this result as the input for the next iteration,

phase retrieval will favor the image instead of its twin and will eventually pull the solution towards $\psi(\boldsymbol{r})$[8].

Summary and outline

This chapter summarized algorithmic tools necessary for phase retrieval. First, a review of discrete near-field propagation was presented. Among these techniques, the transfer function implementation was chosen as preferred propagator for the implementation of iterative phase retrieval.

Second, starting from basic mathematical tools, projection algorithms were introduced. To this end, firstly, projections were described. Their properties depend on the set of functions onto which they project. Once these sets are closed, convex and intersecting, alternating projections converge to a fixed point that has the properties of all of the sets simultaneously. Iterative phase retrieval relies on projections onto sets of functions that fulfill boundary conditions of the experimental setup and reproduce the measured data. Different phase retrieval techniques were introduced: among them the method of Gerchberg and Saxton and the error reduction algorithm. Both of them belong to the class of alternating projections algorithms. Furthermore, the difference map was described in a theoretical way and in an application oriented version in the context of far-field ptychography. Finally, different algorithms to analyze ptychographic data were summarized. In the following chapter, one of these algorithms, namely the ePIE method, will be modified and applied for a generalized version of ptychography for the optical near-field.

---

[8] Note that the overlap between neighboring scan positions has to be considerably larger than sketched in Fig. 3.16. A large overlap further results in an 'incoherent' mixture of neighboring twin image components and impedes stagnation.

4 Near-field ptychography: Theory, implementation, simulations

The situation of near-field X-ray holography deviates in many aspects from far-field diffraction imaging. In particular, two main differences have to be considered: (1) The propagation of optical wave fronts is carried out by the Fresnel-Kirchhoff-integral instead of the Fourier transform. (2) The probe is not confined by any kind of aperture and the entire field of view of the detector is illuminated.
While the first aspect is a technical detail that affects implementation and experimental realization, such that the sampling condition of the near-field propagator is fulfilled, the second aspect is crucial for successful reconstruction of the correct object and probe. It becomes necessary to adapt and generalize the whole procedure of ptychography and thereby to include longitudinal translations of the sample.
Without the claim of completeness, a short review of relevant literature will be given. A motivation is provided why this additional degree of freedom – namely translations of the sample along the optical axis – is necessary. It follows a detailed description of the constraints and the implementation of near-field ptychography. Subsequently, simulations are presented to illustrate performance, capabilities and limits of near-field ptychography. The chapter closes with a summary.

4.1 Ptychography in the optical near-field: Concepts and developement

The history of Fresnel X-ray imaging dates back to the late 20th century. In 1996, S. W. Wilkins *et al.* demonstrated phase contrast X-ray imaging with a polychromatic and confined source [12]. Lateral confinement of the optical field leads to an increase in the lateral coherence length (which is inversely proportional to the source size) and due to the magnifying cone beam geometry allows for high resolution imaging basically independent of the detector resolution.
One year later, the concept of cone beam X-ray imaging with a one dimensional X-ray waveguide was successfully performed by S. Lagomarsino *et al.* at the European Synchrotron Radiation Facility (ESRF). The image quality was promising enough that the authors already pointed out the possibility to reconstruct the object's transmission function by numerical algorithms [13].
In the year 2006, G. J. Williams *et al.* imaged a solid state test structure with an X-ray beam of highly curved phase fronts. Focusing was performed by the use of a Fresnel zone-plate [15]. In the focal plane, the beam was confined by an order sorting aperture. A central stop prevented the zeroth order diffraction from hitting the detector. The measured diffraction patterns include a Fresnel diffraction region in the center as well as intensity scattered to higher angles. Thus, the primary beam decayed considerably within the field of view. This enabled a probe reconstruction prior to phase retrieval of the object's transmission function. The probe was determined by the method of H. M. Quiney *et al.* – a technique that allows reconstructing a highly focused beam by iterating between the pupil plane

of the zone-plate, the focal plane and the detection plane [75]. Hence, two data sets were needed – one that contained the Fresnel diffraction patterns and one that contained independent recordings of the probe. After separate probe retrieval, an error-reduction algorithm was used to reconstruct the complex valued object.

In the same year, magnifying cone beam off-axis holography with two dimensional X-ray waveguides was demonstrated by C. Fuhse *et al.* [76]. Instead of an error reduction scheme as in [15], here a single step holographic phase retrieval was performed, using the approach of off-axis illumination to avoid twin image artifacts. Apart from cone beam X-ray imaging that also records the decay of the illumination, X-ray diffraction in the optical near-field has been under investigation as well. Here, the entire field of view of the detector is illuminated. Phase retrieval based on the contrast transfer function and a series of images defocused to different degrees was successfully applied and finally, in 1999, lead to the development of holo-tomography by P. Cloetens *et al.* [10, 11].

In the following years, cone beam holography was further developed and in 2008, the technique of keyhole coherent diffractive imaging [77] was invented. The setup is basically the same as for Fresnel coherent diffractive imaging: A highly curved X-ray beam, focused by a zone-plate is impinging on a sample. Effective near-field diffraction, as well as scattering to higher angles is recorded by the detector. The difference with the experiment published in [15] is that in [77] an extended test structure was imaged by separate recordings of different regions of the sample. These regions were reconstructed independently (error reduction with probe retrieval performed beforehand following the method of [75]) and finally composed to a single image.

One year later, D. J. Vine *et al.* [78] published a method called ptychographic Fresnel coherent diffractive imaging. It is a combination of Fresnel coherent diffractive imaging and ptychography following the PIE method without probe retrieval. Again, prior to reconstruction, phase retrieval for the probe is performed by the algorithm of H. M. Quiney. The scattered field is determined by alternating projections using a support and a magnitude constraint. After that, the object's transmission function is calculated similarly as in the PIE method of J. M. Rodenburg and H. M. L. Faulkner [68, 69].

In 2013, M. Stockmar *et al.* published near-field ptychography based on the difference map and – as in far-field ptychography – on exclusively lateral diversity [79] and which finally was also extended to tomography [80]. The main differences with the methods which were just described are that firstly, an illumination extending over the entire field of view is used. With the primary beam covering the whole active area of the detector, no higher angle scattering signal can be recorded. The data consists of magnified Fresnel holograms. Secondly, probe retrieval is performed simultaneously to the reconstruction of the object. However, phase retrieval did not work as well as for far-field ptychography: Reconstructions were noisy. This noise was attributed to a lack of diversity in the data. By putting a wave front diffusing material (i.e. a piece of paper) in the beam path upstream the sample, the quality of the reconstructions could be considerably improved. Yet, now, the information about the probe is lost: Instead of the pure, reconstructed

X-ray beam, the deteriorated probe is useless for further analysis like beam characterization.

This first real near-field ptychography attempt was mainly inspired by far-field imaging techniques, where indeed diversity in the data is most efficiently generated by laterally shifting the sample through the beam. Yet, contrast and diversity in near-field imaging mainly emerge from defocus translations of the sample as applied in electron microscopy and in holo-tomography [11].
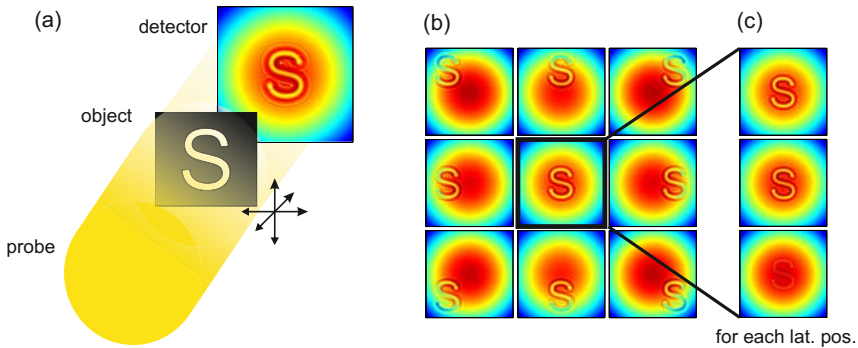
Figure 4.1: Near-field ptychography scheme. (a) An object is illuminated by an extended wave front. Intensities are recorded in the optical near-field/the holographic regime. (b) A conventional ptychographic data set as for the far-field case. (c) In addition, the object is translated to several positions along the optical axis. At each lateral position of (b) a dataset as illustrated in (c) is recorded.

In 2011, the fact that longitudinal shifts help to enhance the quality of phase retrieval has also been detailed on by C. Putkunz *et al.* [81]: An increase in phase resolution and a decrease in artifacts were observed . Translations of the sample in all three dimensions of space are required. Phase retrieval was implemented by adaption of the difference map scheme. However, despite the fact that in this work the application of the difference map for simultaneous phase retrieval of object and probe (see [71]) is cited it was not made use of the update of the probe as proposed in [71]. As in [15, 77, 78], C. Putkunz *et al.* applied the phase retrieval algorithm of H. M. Quiney *et al.* [75] to reconstruct the probe prior to the object. Note that whereas the method of [75] for probe retrieval works for the special configuration of a focused beam cleaned up by an order sorting aperture, it does not work for reconstructing a full field illumination.

The main task of this thesis was to establish ptychography for the optical near-field without the need of any wave front diffusing material or the recording of flat fields, but with the integration of defocus variations of the sample in order to enhance diversity in the data. It will become clear that a full ptychographic data set for the optical near-field does not only require lateral translations of the sample

(as in the far-field case and in [79]), but also longitudinal translations along the optical axis. In other words, in order to collect holograms with sufficient diversity, a conventional ptychographic data set consisting of exclusively lateral shifts has to be recorded at several positions along the optical axis (see Fig. 4.1).

4.2   Longitudinal diversity as a solution to the twin image problem in near-field imaging

In section 2.5.2, it was deduced that Fresnel back propagation of an inline hologram results in a mixture of a focused, sharp image and an out of focus complex conjugated image (the twin image), see Fig. 4.2(a) and equation 2.135. Assume that an object is shifted only laterally in an extended, clean illumination. A hologram is recorded and reconstructed by simple back propagation for every position of the object. Each of the reconstructions will contain the same, but laterally shifted, defocused twin image (Fig. 4.2(b)). When the reconstructions are superimposed according to their lateral shifts S, the reconstructed twin features will superimpose in the same (correct) way as the reconstructed images (Fig. 4.2(c)). There is no difference with a single and simple holographic reconstruction (Fig. 4.2(a)). Further iterations will most likely stagnate at this solution.
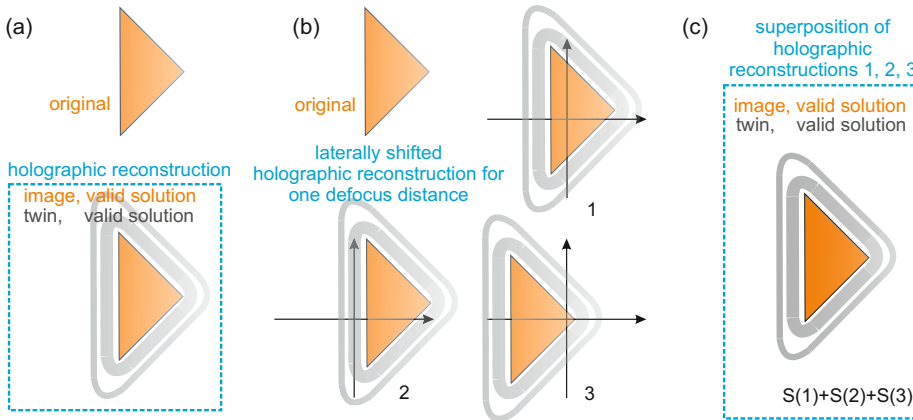
Figure 4.2: Near-field twin image features spoiling the holographic reconstruction (a) cannot be eliminated by lateral translations of an object (orange triangle) in a full-field illumination (b). (c) Images and twin images will add up to correct solutions of the magnitude constraint. Further iterations of phase retrieval will lead to stagnation.

If in contrast the distance between object and detector is alternated, holographic reconstructions will show sharp objects superimposed by twin images defocused to different degrees (Fig. 4.3). Adding up these reconstructions leads to a valid

solution for the object. The superimposed twins, however, cannot be attributed to a single distance (Fig. 4.3(b)). They are invalid solutions of the magnitude constraint. Using this superposition as initial guess for further iterations of phase retrieval, the respective algorithm will most likely converge to the image rather than to a mixture of image and twin.
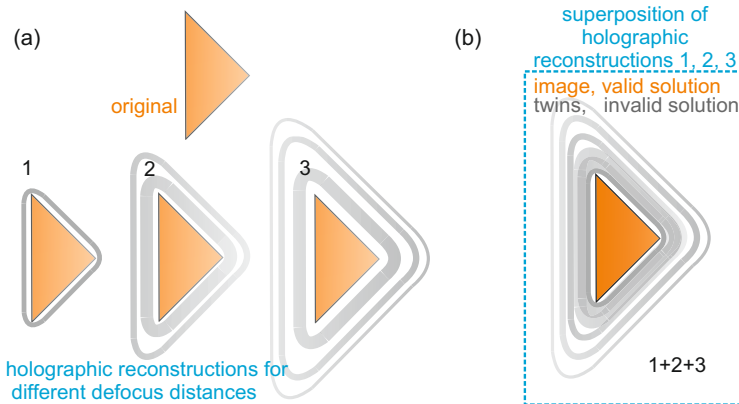
Figure 4.3: Successful twin image elimination in the optical near-field. (a) Holographic reconstructions for the object (orange triangle) at different longitudinal positions with respect to the detector. (b) The superposition of the reconstructions sketched in (a) will add up to a combination of a valid solution of the magnitude constraint and a mixture of twin images defocused to different degrees, which does not fulfill the magnitude constraint. Further iterations of phase retrieval will most likely favor the image rather than a combination of image and twin.

From this analysis, a fundamental difference between the special needs of far- and near-field ptychography becomes clear: In the near-field, it cannot be sufficient to shift an object only perpendicular to the optical axis. Further translations along the optical axis are required to accomplish the task of lateral shifts in far-field ptychography, namely to eliminate twin image artifacts and to provide diversity to the measured data.

If longitudinal diversity is of such importance for near-field imaging, what is the purpose of lateral diversity in this context? We have shown that simultaneous phase retrieval of an object and a non-uniform probe is not possible by exclusively using on-axis shifts [82]. We will see that lateral translations of the object are important for a successful reconstruction of the probe – unperturbed by twin image artifacts.

4.3 Constraints for near-field-ptychography

Near-field ptychographic phase retrieval allows including different constraints. In general, three types of constraints are to be distinguished: (1) Constraints related to the exit wave, (2) constraints related to the probe, (3) constraints related to

the object. Most of these constraints can be formulated as projections onto sets of functions fulfilling the imposed boundary conditions (see section 3.2.3). Finally, the update of object and probe is formulated as a combination of all of these constraints.

Due to the extended probe used for near-field ptychography and in contrast to far-field ptychography, it makes more sense to talk about shifts of the object, rather than to consider shifts of the probe. The object $o^{(i)}$ at lateral position i is

$$o^{(i)}(\boldsymbol{r}) = o(\boldsymbol{r} + \boldsymbol{r}^{(i)}) \ , \tag{4.307}$$

where $\boldsymbol{r}^{(i)}$ is the vector pointing to the center of mass of the object at position i. Shifts of an object from lateral position i to lateral position j will be expressed by

$$\begin{aligned} S^{(i \to j)}\left(o^{(i)}(\boldsymbol{r})\right) &= o(\boldsymbol{r} + \boldsymbol{r}^{(i)} + (\boldsymbol{r}^{(j)} - \boldsymbol{r}^{(i)})) \\ &= o(\boldsymbol{r} + \boldsymbol{r}^{(j)}) = o^{(j)}(\boldsymbol{r}) \ . \end{aligned} \tag{4.308}$$

The shift operator $S^{(i \to j)}$ shall always act on the transmission function of the object – it shall not act on the probe, hence

$$\psi_j^{(j)} = S^{(i \to j)}\left(\psi_j^{(i)}\right) = S^{(i \to j)}\left(o^{(i)}\right) p_j \ . \tag{4.309}$$

Constraints for the exit wave

Magnitude constraint for the exit wave

Most of the knowledge about object and probe is taken from the measured data. To be more precise, the most important information is encoded in the recorded holograms. The current iterate of the exit wave $\psi_j^{(i)}$ has to reproduce these holograms. For this purpose the magnitude constraint $P_{M_{\mathrm{nf}}}$ (see also description related to equation 3.279) for the optical near-field is used

$$P_{M_{\mathrm{nf}}}(\psi_j^{(i)}) = \mathcal{D}_{-\Delta_{j,\mathrm{det}}}\left[\frac{\mathcal{D}_{\Delta_{j,\mathrm{det}}}(\psi_j^{(i)})}{\left|\mathcal{D}_{\Delta_{j,\mathrm{det}}}(\psi_j^{(i)})\right|} \sqrt{I_{j,\mathrm{det}}^{(i)}} \right] \ . \tag{4.310}$$

As in the previous chapter, upper indices in brackets always denote lateral positions of the sample, lower indices always refer to the longitudinal position of the sample. Consequently, $\psi_j^{(i)}$ is the exit wave with the sample at lateral position i and longitudinal position j. The notation of the intensity $I_{j,\mathrm{det}}^{(i)}$ has to be read as intensity of $\psi_j^{(i)}$ recorded at fixed defocus position det.

Separability constraint for the exit wave

The exit wave function $\psi_j^{(i)}$ needs to be separable into the complex transmission

function of the object at lateral position i and the complex wave field of the probe at the specific defocus position j. To this end, the objective function

$$F := \sum_{\underbrace{x,y}_{\text{pixels}}} \sum_{\underbrace{j}_{\text{long. shifts}}} \sum_{\underbrace{i}_{\text{lat. shifts}}} |\psi_j^{(i)}(x,y) - \hat{\psi}_j^{(i)}(x,y)|^2, \tag{4.311}$$

where

$$\hat{\psi}_j^{(i)} = \hat{o}^{(i)} \hat{p}_j = S^{(1 \to i)} \left[\hat{o}^{(1)} \right] \mathcal{D}_{\Delta_{1,j}} [\hat{p}_1]$$

is minimized (see also [82]). Since F contains only absolute values, it is minimal when each of its single summands $f_j^{(i)}$ is minimal. It suffices to minimize

$$
\begin{align}
f_j^{(i)} &:= |\psi_j^{(i)} - \hat{\psi}_j^{(i)}|^2 \tag{4.312} \\
&= \left(\psi_j^{(i)} - \hat{o}^{(i)} \hat{p}_j \right) \left(\psi_j^{(i)} - \hat{o}^{(i)} \hat{p}_j \right)^* \tag{4.313} \\
&= \psi_j^{(i)} \psi_j^{(i)*} - \hat{o}^{(i)*} \hat{p}_j^* \psi_j^{(i)} - \hat{o}^{(i)} \hat{p}_j \psi_j^{(i)*} + \hat{o}^{(i)} \hat{p}_j \, \hat{o}^{(i)*} \hat{p}_j^* . \tag{4.314}
\end{align}
$$

The function $f_j^{(i)}$ depends on $\hat{o}^{(i)}$ and $\hat{p}_j$. The point wise partial derivatives with respect to these variables are set to zero:

$$\frac{\partial f_j^{(i)}}{\partial \hat{o}^{(i)}} = -\hat{p}_j \psi_j^{(i)*} + |\hat{p}_j|^2 \hat{o}^{(i)*} = 0 , \tag{4.315}$$

$$\frac{\partial f_j^{(i)}}{\partial \hat{p}_j} = -\hat{o}^{(i)} \psi_j^{(i)*} + |\hat{o}^{(i)}|^2 \hat{p}_j^* = 0 . \tag{4.316}$$

Solving equation 4.315 for $\hat{o}^{(i)}$ results in

$$\hat{o}^{(i)} = \frac{\hat{p}_j^* \psi_j^{(i)}}{|\hat{p}_j|^2} . \tag{4.317}$$

Solving equation 4.316 for $\hat{p}_j$ leads to

$$\hat{p}_j = \frac{\hat{o}^{(i)*} \psi_j^{(i)}}{|\hat{o}^{(i)}|^2} . \tag{4.318}$$

Equation (4.317) can be interpreted in an intuitive way:
Noting that $|\hat{p}_j|^2 = \hat{p}_j \hat{p}_j^*$, the division by $|\hat{p}_j|^2$ combined with the multiplication with $\hat{p}_j^*$ is similar to the division by $\hat{p}_j$. Assuming that $\psi_j^{(i)}$ can be formulated as a product of object and probe, only the component referring to the object is left. An analogous argumentation is possible for the evaluation of the probe by equation (4.318).
Note that both $\hat{o}^{(i)}$ and $\hat{p}_j$ should be non-zero in order to provide reasonable values for equations 4.317 and 4.318. For near-field ptychography, in contrast to far-field

ptychography, this can be guaranteed: Thin objects of high transmission and low absorption in an extended wave field are considered. For this reason, the absolute value of the object's transmission function should be close to one. The absolute value of $\hat{p}_j$ is far from zero: The entire field of view of the detector is illuminated by a partially coherent p_j. Yet, if one would account for a probe covering only a finite section of the detector and for an object that has regions of zero transmission, it would be more accurate to describe equations 4.317 and 4.318 in a similar way as shown in [73] by

$$\hat{o}^{(i)} \approx o^{(i)} - \lambda_o \frac{\partial f_j^{(i)}}{\partial \hat{o}^{(i)}} \ , \tag{4.319}$$

$$\hat{p}_j \approx p_j - \lambda_p \frac{\partial f_j^{(i)}}{\partial \hat{p}_j} \ , \tag{4.320}$$

where λ_o and λ_p are parameters that determine the step size in direction of the negative gradients $\frac{\partial f_j^{(i)}}{\partial \hat{o}^{(i)}}$ and $\frac{\partial f_j^{(i)}}{\partial \hat{p}_j}$.

Constraints for the object

Constraints which only affect the object can be deduced from pre-knowledge of the sample without considering the measured holograms.

Phase object constraint

In most cases, due to the complex index of refraction with $(\delta \gg \beta)$, contrast caused by phase shift dominates with respect to contrast caused by absorption. This is in particular true for weakly absorbing materials like biological tissues. If the sample is approximately transparent, a phase object constraint can be applied which projects onto functions of uniform and full transmission:

$$P_P(o^{(i)}(\boldsymbol{r})) := \frac{o^{(i)}}{|o^{(i)}|} \ . \tag{4.321}$$

As deduced in section 3.4, the phase object constraint is related to a non-convex set.

Amplitude constraint

A less restrictive version of equation 4.321 is the amplitude or transmission constraint (see section 3.4). It can also be applied for objects with arbitrary absorption. Transmission values larger than one are set to one, transmission values lower than one are left unchanged. This procedure is similar to the method proposed by T. Latychevskaia and H.-W. Fink [83] for solving the twin image problem. However, assuming an object that extends over the entire field of view and that consists of a combination of different materials, it makes more sense to correct

transmission values that increase a certain value $\epsilon \leq 1$, because full transmission might still be too large

$$P_T(o^{(i)}(\boldsymbol{r})) := \begin{cases} o^{(i)}(\boldsymbol{r}) & \text{if } |o^{(i)}(\boldsymbol{r})| \leq \epsilon , \\ \epsilon \frac{o^{(i)}(\boldsymbol{r})}{|o^{(i)}(\boldsymbol{r})|} & \text{else.} \end{cases} \qquad (4.322)$$

For example, ϵ can be set to the average transmission of the object. The operation performed by equation 4.322 suppresses features of the amplitude exceeding a certain value ϵ. It is likely that these features result from interference of image and twin. The underlying set of functions fulfilling the transmission constraint is convex.

Negativitiy constraint
Thin objects illuminated by X-rays cause a negative phase shift. Positive phase values can be set to zero (see 3.4).

$$P_N(o^{(i)}(\boldsymbol{r})) := \begin{cases} o^{(i)}(\boldsymbol{r}) & \text{if } \arg(o^{(i)}(\boldsymbol{r})) \leq 0 , \\ |o(x,y)| & \text{else.} \end{cases} \qquad (4.323)$$

Automatic focusing
Since there is always uncertainty about the knowledge of the exact propagation distances, tools for automatic adjustment are beneficial. The automatic focusing algorithm used here is inspired by the method of P. Langehanenberg *et al.* [84]. The recorded hologram $\sqrt{I_{j,\text{det}}}$ with the detector at distance $\Delta_{j,\text{det}}$ is

$$I_{j,\text{det}} = |\mathcal{D}_{\Delta_{j,\text{det}}}[\psi_j]|^2 = |\mathcal{D}_{\Delta_{j,\text{det}}}[o_j\, p_j]|^2 .$$

Here the object $o_j(x,y)$ also gets an index j indicating its position $z = z_j$ along the optical axis. Consider a uniform illumination $p_j(x,y) = \mathbb{I}$, i.e. $\psi_j = o_j$. The absolute value of the holographic reconstruction

$$o_{\tilde{j}}(x,y) = \left| \mathcal{D}_{-\Delta_{\tilde{j},\text{det}}}[I_{j,\text{det}}] \right| \qquad (4.324)$$

with incorrect distance $z_{\tilde{j}} = z_j + \delta$ is focused for the distance $z_{\tilde{j}}$ where the logarithmically weighted, accumulated Fourier spectrum

$$f(z_{\tilde{j}}) = \sum_{q_x, q_y} \log\left\{ 1 + \left| \mathcal{F}\left[o_{\tilde{j}}(x,y) \right] \right| \right\} \qquad (4.325)$$

is minimal in case that the contrast due to phase shift dominates the contrast due to absorption, or maximal in case that absorption contrast is more pronounced [84].
 To find the minimum/maximum of $f(z_{\tilde{j}})$, a climbing-search strategy [84] is used (see Fig. 4.4).
This procedure is similar to manually adjusting a microscope. At first, the sample-detector distance is changed using a coarse step size $\Delta z^{(1)}$. The position where the sample appears focused is determined. After that, the sample is translated
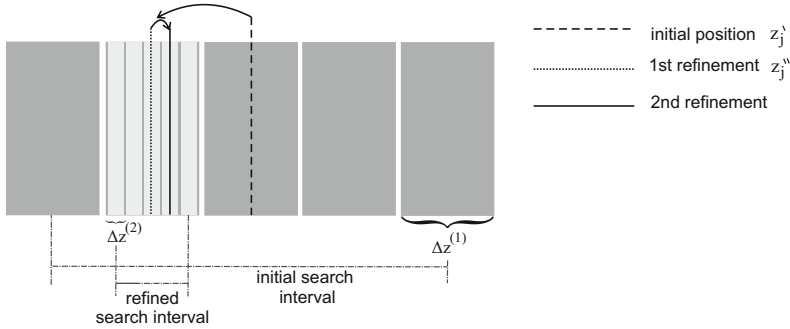
Figure 4.4: Illustration of the climbing search strategy.

again with respect to the detector. This time a smaller step size $\Delta z^{(2)}$ is used and the position of the object is further refined.

The climbing-search strategy does essentially the same: In a first step, $f(z_{\tilde{j}})$ is calculated for values of $z_{\tilde{j}}$ in a large interval around an initial guess $z'_{\tilde{j}}$. For this purpose, a coarse step size $\Delta z^{(1)}$ is used. The minimum (maximum) of $f(z_{\tilde{j}})$ within this interval is determined and a refined distance $z''_{\tilde{j}}$ is found. Next, $f(z_{\tilde{j}})$ is minimized (maximized) for values within a narrower interval around $z''_{\tilde{j}}$. This time, a finer step size $\Delta z^{(2)}$ is chosen. The minimum of $f(z_{\tilde{j}})$ within this narrow interval is determined. The process is iterated until the minimum (maximum) of $f(z_{\tilde{j}})$ is found up to a desired precision.
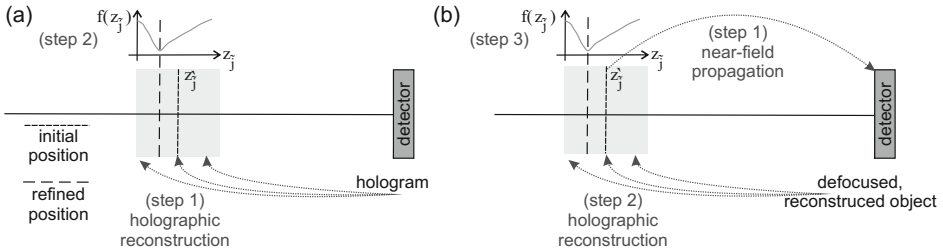
Figure 4.5: Application of automatic focusing prior to reconstruction (a) and during phase retrieval (b). (a) The recorded holograms are back propagated to different distances $z_{\tilde{j}}$ around an initial position at z'_j. Equation 4.325 is evaluated and its minimum (maximum) is determined. A refined distance is found. (b) The transmission function is propagated over a (probably wrong) distance. It is back propagated to several positions around its starting point. Equation 4.325 is evaluated and an updated position of the sample is found at the minimum/maximum of f.

The automatic focusing procedure can be applied to the recorded holograms and to the reconstructed, defocused transmission function of the object in iteration i. Before phase retrieval, distances are roughly determined with the help of the

recorded (and if possible for this purpose filtered and flat field corrected) holograms (see Fig. 4.5(a)). These distances are used for the first few iterations of phase retrieval.

It is the magnitude constraint, which requires information about the propagation distances. Since by this constraint the reconstructed intensities are replaced by the measured ones, the intensity of the defocused object *at the detector* should be correct. This does *not yet* mean that the absolute position where the object should be focused is correct, too. For this reason, the sample-detector distance is further refined during phase retrieval starting from the defocused, current iterate of the transmission function in the detection plane (see Fig. 4.5(b)).

Constraints for the probe

The probe is the most uncertain part; we almost have no knowledge about what constraints the probe despite the average number of photons and eventually a measured flat field. The following options to constrain the probe can be applied, if at least one measured flat field is available. (1) The number of photons can be adjusted by scaling the total reconstructed intensity to the intensity of the flat field. (2) Using the flat field, the magnitude projection can be applied to the probe. (3) Having access to two flat fields at different positions along the optical axis, a Gerchberg-Saxtion step can be used to correct the reconstructed illumination.

Adjustment of total flux
Adjustment of the flux is done by

$$\mathrm{F_{ad}}(|p_1(x,y)|^2) = \left(\sum_{x,y} I_{\mathrm{flat}}(x,y) \right) \cdot \frac{|p_1(x,y)|^2}{\sum_{x,y} p_1(x,y)p_1^*(x,y)} \ , \qquad (4.326)$$

where I_{flat} is the measured flat field.

Magnitude constraint for the probe
Similar to the magnitude constraint for the exit surface wave, a magnitude constraint can be formulated that adapts the reconstructed flat field to the measured one – provided that a flat field has been recorded:

$$P_{M_{\mathrm{flat}}}(p_1) = \mathcal{D}_{-\Delta_{1,\mathrm{det}}} \left[\frac{\mathcal{D}_{\Delta_{1,\mathrm{det}}}(p_1)}{|\mathcal{D}_{\Delta_{1,\mathrm{det}}}(p_1)|} \sqrt{I_{\mathrm{flat}}} \right] \ . \qquad (4.327)$$

Here, the total intensity is automatically adjusted.

Gerchberg-Saxton step
A Gerchberg-Saxton step can be performed, if flat fields at two different positions along the optical axis are available. It corrects the reconstructed intensities of the probe at two different defocus distances

$$\mathrm{GS}(p_1) = P_{M_{\mathrm{flat}_2}} \left[P_{M_{\mathrm{flat}_1}}(p_1) \right] \ . \qquad (4.328)$$

In experiments at synchrotron beamlines, it is often not possible or not foreseen to shift the detector along the optical axis. In general, the detector is translated under a certain angle with respect to the beam path. Additional alignment of the flat fields – one with respect to the other and with respect to all of the measured holograms – has to be performed before phase retrieval. For this reason, the recording and usage of two flat fields at different positions is not favored.

Additional tools to improve reconstructions

Low pass filter

In case of a strongly aberrated probe that contains features of high frequency, it is helpful to filter the reconstructed object during the first iterations and thereby to force high frequencies to merge into the probe. Low pass filtering is done in Fourier space by multiplication with a Gaussian function $g(q_x, q_y)$ of adjustable full width at half maximum of $2\sqrt{2\ln 2}\,\sigma$.

$$\text{LP}\left[o_j^{(i)}\right] = \mathcal{F}^{-1}\left[\mathcal{F}\left(o_j^{(i)}\right) \cdot g(q_x, q_y)\right]\;, \tag{4.329}$$

where

$$g(q_x, q_y) := \exp\left(-\frac{q_x^2 + q_y^2}{2\sigma_k^2}\right)\;. \tag{4.330}$$

The parameter σ_k depends on the current iteration k

$$\sigma_k = \sigma \cdot k\;. \tag{4.331}$$

It is linearly increased during phase retrieval. Hence, filtering is gradually relaxed and finally switched off completely.

Linear blending

The technique of linear blending is used to prevent artificial edges that result from the recombination of the current iterate of the object, which shows a single part of the object, and the previous guess, which shows the entire object. The method is commonly found in software designed for image processing.

Two images A (for example the photograph shown in Fig. 4.6(a), left part) and B (for example the photograph shown in Fig. 4.6(b), left part) are linearly blended into each other by the following procedure: (1) The mean absolute value $\overline{A}$ of image A is adjusted to the mean absolute value $\overline{B}$ of image B. (2) The adjusted image A is multiplied by an inverse truncated pyramid mask $(1 - \gamma)$ which is zero within its central region and linearly increasing towards the edges, where it is one (Fig. 4.6(a), right part and Fig. 4.6(c)). Image B is multiplied by a truncated pyramid mask γ which is one within its central region and linearly decaying towards the edges, where it is zero (Fig. 4.6(b), right part and Fig. 4.6(d)). Finally, both masked images are added

$$C = (1 - \gamma)\left[A + (\overline{B} - \overline{A})\right] + \gamma B\;. \tag{4.332}$$
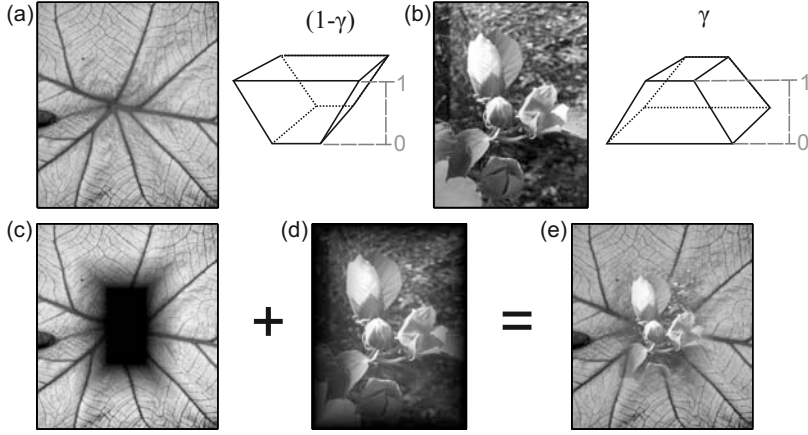
Figure 4.6: Illustration of linear blending. (a) Image A and inverse, truncated pyramid mask $(1 - \gamma)$. (b) Image B and truncated pyramid mask γ. (c) Image A multiplied by $(1 - \gamma)$. (d) Image B multiplied by γ. (e) Result: sum of (c) and (d).

The resulting image C does not show sharp edges as a consequence of the combination of A and B (Fig. 4.6(e)). The steepness of the truncated pyramid mask γ has to be chosen such that the transition between A and B appears quite natural and smooth.

4.4 Implementation of near-field ptychography

The implementation of near-field ptychography follows the ePIE scheme proposed in [74]. The combination of the techniques introduced in this chapter in an appropriate way results in an algorithm for near-field ptychography which was designed and implemented in the scope of this work.

For $k = 1 : it$, where it is the number of iterations,
  for $n(i, j) = 1 : m$ where m is the number of recorded holograms,
    choose $(i, j) \in [\underbrace{(1, 1), (1, 2), ..., (2, 1), (2, 2), ...}_{m \text{ pairs of } (i,j)}]$.

Calculate the exit surface wave:

$$\psi_j^{(i),n} = o^{(i),n} \, p_j^n = D_{\Delta_{1,j}} \left[p_1^n \right] \, S^{(1 \to i)} \left[o^{(1),n} \right] \; . \tag{4.333}$$

Separate object and probe:

$$\tilde{p_1}^{n+1} = F_{\mathrm{ad}}\left[p_1^n + \alpha D_{-\Delta_{1,j}}\left[\frac{\left(P_{M_{\mathrm{nf}}}\left[\psi_j^{(i),n}\right] - \psi_j^{(i),n}\right)\, S^{(1\to i)}\left[o^{(1),n}\right]^*}{|S^{(1\to i)}\left[o^{(1),n}\right]|^2}\right]\right],$$

(4.334)

and

$$\tilde{o}^{(1),n+1} = P_X\left\{\mathrm{LP}\left[\tilde{\tilde{o}}^{(1),n+1}\right]\right\},$$

(4.335)

where

$$\tilde{\tilde{o}}^{(1),n+1} = S^{(i\to 1)}\left[\underbrace{S^{(1\to i)}\left[o^{(1),n}\right] + \beta\frac{\left(P_{M_{\mathrm{nf}}}\left[\psi_j^{(i),n}\right] - \psi_j^{(i),n}\right)\, D_{\Delta_{1,j}}\left[\tilde{p_1}^{n+1}\right]^*}{|D_{\Delta_{1,j}}\left[\tilde{p_1}^{n+1}\right]|^2}}_{\text{summed by linear blending}}\right]$$

(4.336)

and where

$$\mathrm{LP} = \begin{cases} \mathrm{LP} & \text{for } 1 \le k \le a \\ Id & \text{else,} \end{cases}$$

(4.337)

and where further constraints are

$$P_X = \begin{cases} Id & \\ P_T & \text{for } b \le k \le c \\ P_P & \text{for } b \le k \le c \\ P_N & \text{for } b \le k \le c \\ P_N\, P_T & \text{for } b \le k \le c \\ P_N\, P_P & \text{for } b \le k \le c \,. \end{cases}$$

(4.338)

Finally, automatic focusing can be applied

$$\tilde{o}^{(1),n+1} \xrightarrow{\text{automatic focusing}} o^{(1),n+1},$$

(4.339)

and different constraints for the probe can be used

$$p_1^{n+1} = P_Y\left[\tilde{p_1}^{n+1}\right],$$

(4.340)

where

$$P_Y = \begin{cases} \mathrm{GS} \\ P_{M_{\mathrm{flat}}} \\ Id. \end{cases}$$

(4.341)

The relaxation parameters α and β can be changed during the reconstruction. The advantage is that in the first iterations, the reconstruction of either the object or the probe can be favored while the other one is kept almost constant. In fact, addressing special weight to the reconstruction of the object at the beginning has turned out to speed up convergence to reasonable solutions.

In cone beam X-ray microscopy, bringing the sample close to the focus provides

an increase in resolution. For this reason it could be of interest to attribute more influence to the reconstructions with finest pixel size. To this end, the reconstructed object resulting from the plane closest to the focus can be favored by a larger value of β.

The thresholds a, b and c for switching additional constraints on and off can be chosen independently for each constraint. It helps to highly constrain the system for the first iterations and relax or turn down the additional boundary conditions for succeeding iterations.
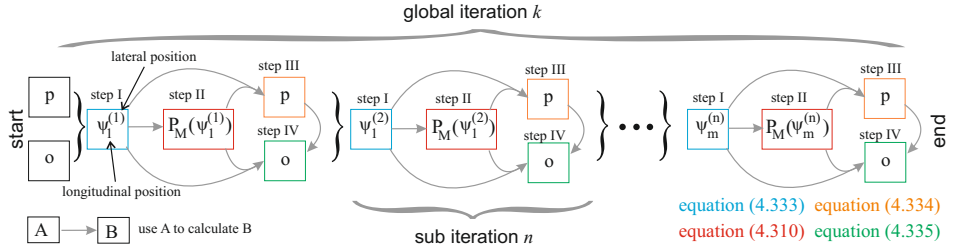
Figure 4.7: Sketch of one global iteration k of the near-field ptychography algorithm.

Figure 4.7 (see also [85]) visualizes the technical concept of near-field ptychography, and especially its differences and similarities with respect to the ePIE method (see also Fig. 3.15). One global iteration k is shown, which consists of subiterations n that address the single holograms.

Each subiteration contains the following four steps: (1) The guessed exit wave $\psi_j^{(i)}$ is calculated by multiplication of o and p. (2) The magnitude constraint is applied to the exit wave. This results in a modified exit wave $P_{M_{\mathrm{nf}}}\left[\psi_j^{(i),n}\right]$. (3) Next, $\psi_j^{(i)}$ and $P_{M_{\mathrm{nf}}}\left[\psi_j^{(i),n}\right]$ are required to calculate an update of p. (4) Finally, the new p, $\psi_j^{(i)}$ and $P_{M_{\mathrm{nf}}}\left[\psi_j^{(i),n}\right]$ are used for the update of o.

The main differences with the classical ePIE scheme are: First, the separation of object and probe is obtained from division by $|S^{(1\to i)}\left[o^{(1),n}\right]|^2$ and division by $|D_{\Delta_{1,j}}\left[p_1^{n+1}\right]|^2$. In a formulation identical to the ePIE, it should be a division by $\max\left(|S^{(1\to i)}\left[o^{(1),n}\right]|^2\right)$ and $\max\left(|D_{\Delta_{1,j}}\left[p_1^{n+1}\right]|^2\right)$, respectively (see equations 3.304 and 3.305).

Second, the updates of probe and object are intertwined with each other, which is not the case for the ePIE, where the newly calculated probe is not used for the update of the object within the same sub-iteration (see Fig. 3.15).

4.5 Pre-processing of the recorded holograms

Before starting phase retrieval, there are some important pre-processing steps to do, namely (1) averaging and dark-field correction of the recorded images, (2) refinement of distances by automatic focusing, (3) re-ordering of the measured data,

(4) determination of shifts by sub-pixel image registration [86].

If a dark-field is available, every hologram is corrected by subtraction of the dark current. In addition, to reduce noise, several redundant holograms can be averaged.

Since the distances between sample and detector (for each defocus position) are always affected by errors, it helps to apply the automatic focusing procedure introduced in section 4.3 to the measured holograms and thereby to refine the experimentally determined distances.

A crucial step in pre-processing is the precise calculation of lateral and longitudinal shifts in sub-pixel precision.
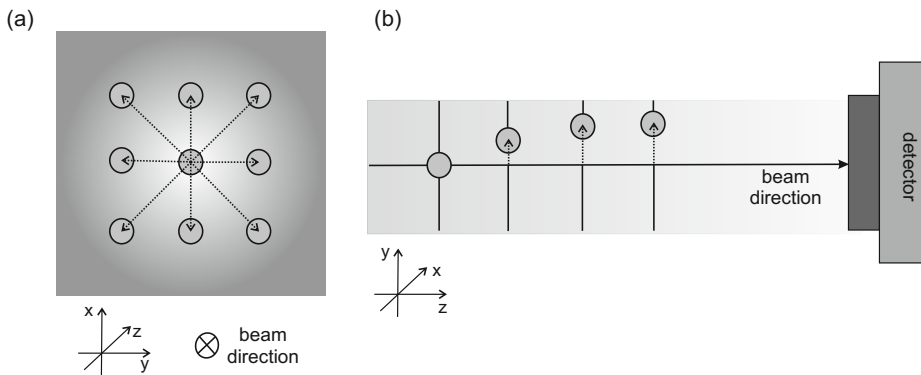
Figure 4.8: (a) Lateral alignment. Shifts in (x,y) (dotted arrows) between the different lateral positions are calculated for each plane. (b) Longitudinal alignment. Shifts in (x,y) between the different defocus positions are determined.

Relative shifts of the holograms are extracted from the measured data by using the image registration technique of M. Guizar-Sicairos *et al.* [86] for sub-pixel alignment.

Lateral shifts are always calculated with respect to the first, stored image of a lateral scan (see Fig. 4.8(a)). Longitudinal alignment is performed with respect to the object found at the largest distance of sample and detector (see Fig. 4.8(b)). While longitudinal alignment due to a considerable large amount of overlap is not difficult, lateral alignment can be challenging whenever not all of the recorded holograms overlap. This is illustrated in Fig. 4.9: The first image (Arabic index (1)) does not overlap with the last image (Arabic index (9)) of the lateral scan. Yet, the central image (Arabic index (5)) overlaps with each and every image of the scan. The re-arrangement of the holograms in a spiral order (following the Roman numbers and arrows) guarantees that the first, stored image (Roman index (I)) overlaps with every other image.

Proper alignment is the most crucial step in data pre-processing. There are two options implemented to perform this task. The optimal method has to be chosen regarding the measured data and/or the properties of the sample.
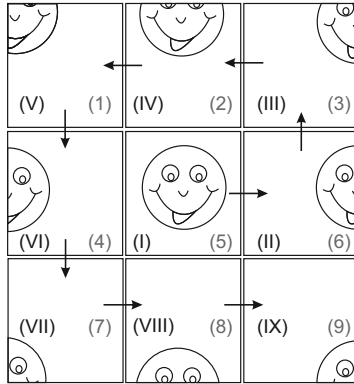
Figure 4.9: Images (1) to (9) are created by shifting a sample horizontally and vertically. Image (1) has no overlap with image (9). However, the central image (Arabic index (5)) overlaps with all of the images (1)-(9). When calculating transversal shifts, it is advantageous to store the images according to the roman indices. This guarantees that for all images shifts can be calculated with respect to the central image.

(1) It is possible to align the *holograms* for each lateral scan with respect to the first, stored hologram.

(2) Another option is to align the *holographic reconstructions* of the raw data using either the reconstructed phases (recommended procedure for phase objects), or the reconstructed amplitudes (recommended for mainly absorbing objects).

Both options allow to pre-filter the raw data before alignment by application of a mask in Fourier space, that needs to be specially designed for each experiment.

If a flat field is available, the holograms are flat field corrected before alignment (but, of course, not again before phase retrieval). Flat field correction combined with Fourier space filtering can considerable facilitate alignment.

4.6 Modification of near-field ptychography for cone beam geometry

As detailed in section 2.3.3 and section 2.4, curved wave fronts enable the recording of magnified holograms. This overcomes limitations imposed by the physical pixel size of the detector. Next, the near-field ptychography algorithm shall be adapted to cone beam geometry (see also supplemental material of [85]). In section 2.3.3, it was deduced that propagation of a cone beam is equivalent to propagation of a magnified plane wave.

Let ψ_1 be a modulated plane wave at distance $\Delta_{f,1}$ with complex amplitude $\mathcal{A}$

$$\psi_1(x,y) = \mathcal{A}\exp\left(ik\Delta_{f,1}\right) \ . \tag{4.342}$$
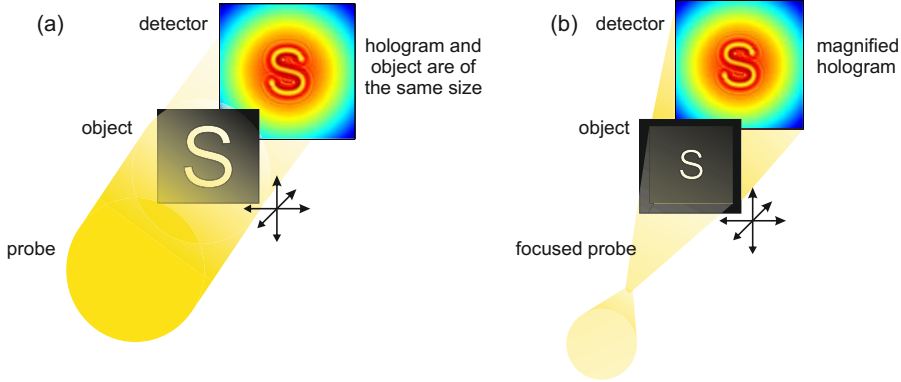
Figure 4.10: (a) Near-field ptychography setup for parallel beam geometry. (b) Near-field ptychography setup for magnifying cone beam geometry.

A respective cone beam illumination in paraxial approximation at distance $\Delta_{f,1}$ behind the focus is then described by

$$\psi^{\mathrm{SW}}(x, y, z = \Delta_{f,1}) = \psi_1(x, y) \exp\left[\frac{ik}{2\Delta_{f,1}} \; x^2 + y^2)\right] \; . \qquad (4.343)$$

Propagation of ψ^{SW} using equation 2.45 is performed by magnification of ψ_1 by a factor of

$$M = \frac{\Delta_{f,1} + \Delta_{1,\mathrm{det}}}{\Delta_{f,1}} \; . \qquad (4.344)$$

The propagation distance has to be transformed as well, resulting in an effective length

$$z_{\mathrm{eff}} = \frac{\Delta_{1,\mathrm{det}}}{M} \; , \qquad (4.345)$$

where $\Delta_{1,\mathrm{det}}$ is the distance from plane 1 to the detector.

Cone beam near-field ptychography requires transformation of both, object and probe to effective geometry. The exit wave has to be propagated from any defocus plane to the detector. Each distance corresponds to a different magnification. Furthermore, the probe has to be propagated between the different defocus positions. In fact, it is necessary to consider two settings: one for the exit wave, i.e. the product of object and probe and one for the probe. This is depicted in Fig. 4.11, which illustrates effective propagation of (a) the exit wave, and (b) the probe.

The magnification of the object (or equivalently the exit wave) $M_o^{(n)}$ is determined by the distance of the object at longitudinal position n with respect to the focus $(\Delta_{f,1} + \Delta_{1,n})$ and the distance between focus and detector $(\Delta_{f,1} + \Delta_{1,\mathrm{det}})$

$$M_o^{(n)} = \frac{\Delta_{f,1} + \Delta_{1,\mathrm{det}}}{\Delta_{f,1} + \Delta_{1,n}} \; . \qquad (4.346)$$
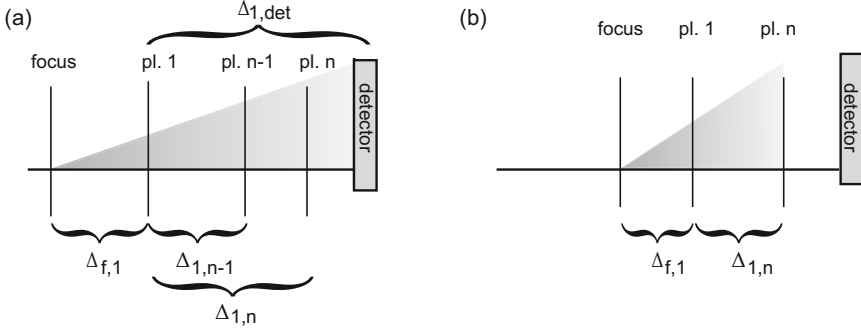
Figure 4.11: (a) Effective geometry for propagation of the exit wave. (b) Effective geometry for propagation of the probe.

With this information at hand, effective pixel sizes and effective propagation distances for the exit wave can be calculated by

$$d_{\text{eff},o}^{(n)} = \frac{\delta}{M_o^{(n)}} \quad \text{and} \quad z_{\text{eff},o}^{(n)} = \frac{\Delta_{1,\text{det}}}{M_o^{(n)}} \frac{\Delta_{1,n}}{} \; , \tag{4.347}$$

where δ is the physical pixel size of the detector. According to Fig. 4.11(b), the propagation of the probe requires the magnification $M_p^{(n)}$

$$M_p^{(n)} = \frac{\Delta_{f,1} + \Delta_{1,n}}{\Delta_{f,1}} \; . \tag{4.348}$$

Effective pixel sizes and propagation distances become

$$d_{\text{eff},p}^{(n)} = \frac{\delta}{M_p^{(n)}} \quad \text{and} \quad z_{\text{eff},p}^{(n)} = \frac{\Delta_{1,n}}{M_p^{(n)}} \; . \tag{4.349}$$

The product of both magnifications – $M_o^{(n)}$ and $M_p^{(n)}$ – is constant for every position n

$$M_p^{(n)} \cdot M_o^{(n)} = \max\left[M_o^{(n)} \right] = \text{const.} \tag{4.350}$$

This illustrates that a step wise effective propagation from one plane to the next and then to the detector results in the same magnification as a one-step propagation directly to the detector.

When further examining the effective propagation of the probe, one can observe that the effective field of view $N_x^{(n)} \cdot d_{x,\text{eff},p}^{(n)} \times N_y^{(n)} \cdot d_{y,\text{eff},p}^{(n)}$ (yellow parts in Fig. 4.12) is constant for each defocus plane. The detector is always at constant distance $\Delta_{1,\text{det}}$ with respect to the focus. In this configuration, there is no way to access information about the regions outside the effective field of view (gray regions in Fig. 4.12).
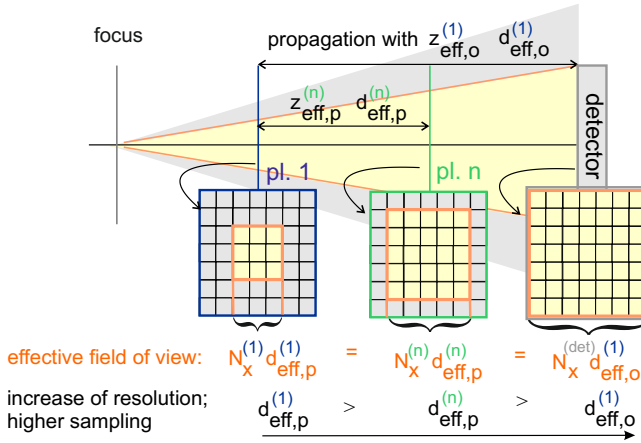
Figure 4.12: For each longitudinal position n, the effective field of view (yellow) is constant. Propagation of the probe in effective geometry from plane 1 to plane n requires larger pixel size and coarser sampling than propagation from plane 1 to the stationary detector. Decreasing and increasing of the resolution is avoided by using a scaling factor $s^{(n)}$. The probe is always propagated with highest resolution $d_{\text{eff},o}^{(1)}$, keeping the number of pixels as well as the effective field of view constant.

The effective pixel size of the probe decreases for defocus positions closer to the detector. To keep the effective field of view constant, the number of pixels, carrying information has to increase (see lower part of Fig. 4.12).

Following these insights, one can think of two ways to propagate the probe from plane 1 to plane n using effective geometry: (1) In plane 1 the complex valued field of the probe is of highest resolution, i.e. of smallest effective pixel size. For propagation from plane 1 to plane $n > 1$ the resolution of the probe has to be decreased to $d_{\text{eff},p}^{(n)} < d_{\text{eff},o}^{(1)}$. Starting from plane n for propagation to the detector, the resolution has to be increased again. Interpolation and binning of the probe is always accompanied by a loss of information and requires computation time.

(2) A more efficient technique is to keep the number of pixels with information content constant. This is equal to scaling the effective pixel size of the probe

$$\tilde{d}_{\text{eff},p}^{(n)} = d_{\text{eff},p}^{(n)} \cdot s^{(n)} \tag{4.351}$$

by the factor

$$s^{(n)} = \frac{M_p^{(n)}}{\max\left[M_o^{(n)}\right]} \quad . \tag{4.352}$$

No interpolation or reduction of the probe is necessary. Instead, the probe is assumed to be available at highest possible resolution $d_{\text{eff},o}^{(1)}$ in each defocus plane.

4.7   A study based on simulations for a parallel beam setup

Before turning to experimental realization of near-field ptychography, a preliminary study is presented in this section. The following issues are addressed: (1) Which artifacts do typically occur in the reconstructions? (2) What is the main purpose of longitudinal diversity? (3) What kind of information is added by lateral diversity?

Table 4.1: Parameters for the simulations. Fresnel numbers $F_{10} = a^2/z\lambda$ are defined by a characteristic structure of $a = 10\delta$, where δ is the pixel size and λ is the wavelength.

wavelength [nm]	0.1
pixel size [µm]	0.1
number of pixels	1024×1024
F_{10}	0.10, 0.11, 0.14, 0.33
# holograms	64 (4 defocus planes, 16 lat. positions)

To answer these questions, different simulations were carried out using the same values for wavelength, pixel size, and the same size of the field of view (see Table 4.1). The simulated holograms (examples depicted in Fig. 4.13) are free of noise such that artifacts in the reconstruction can be attributed to the phase retrieval algorithm rather than to the influence of noise. Oversampling of the Fresnel-propagator (see equation 2.45 and section 3.1) was fulfilled for each propagation distance. The essential difference between the three types of simulations is the choice of object and probe.

4.7.1   A first proof of principle

To determine the characteristics and the extent of errors in the reconstructions, an artificial case was studied: a probe consisting of two different photographs (one for the amplitude and one for the phase) as well as an object modeled by two other photographs for amplitude and phase. Figure 4.14(a) shows the modeled object in amplitude a (first picture from left) and phase p (second picture from left), as well as the modeled probe in amplitude (second picture from right) and phase (first picture from right).

The artificial object was shifted to four positions along the optical axis with Fresnel numbers (calculated for 10 pixels) of 0.10, 0.11, 0.14, 0.33. A lateral scan (four by four scan points) was simulated at each of these longitudinal positions. In total 64 simulated holograms were used for phase retrieval. The step size for the lateral scans was set to 50 pixels in horizontal and vertical direction. Fig. 4.13(a) depicts one of the artificial holograms as an example.

Phase retrieval was initialized with a uniform amplitude distribution and zero phase (both for object and probe). No other constraints than the magnitude constraint and the separation of object and probe were used. During the first iteration the update of the probe was almost kept constant ($\alpha = 0.01$) and the update of the object was performed with $\beta = 0.3$. After the first iteration both relaxation
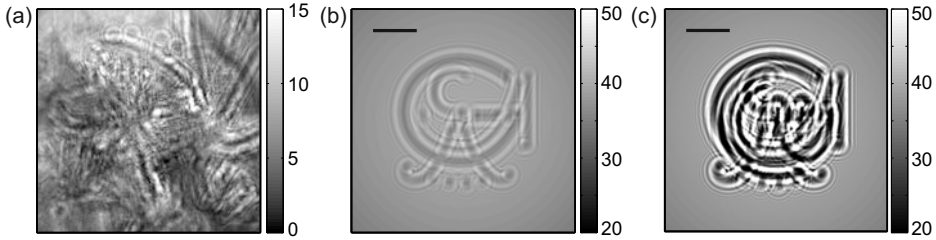
Figure 4.13: Intensity (a.u.) of simulated holograms for object and probe modeled by four different photographs (a); the object modeled by a structure forming the logo 'GAU' and a clean probe (softly decaying amplitude and zero phases) (b); the object modeled by the 'GAU' structure (stronger contrast than in (b)) and the probe modulated by a second structure ('IRP'-logo) of equal contrast as used for the object (c). Scale bars denote 20 µm.
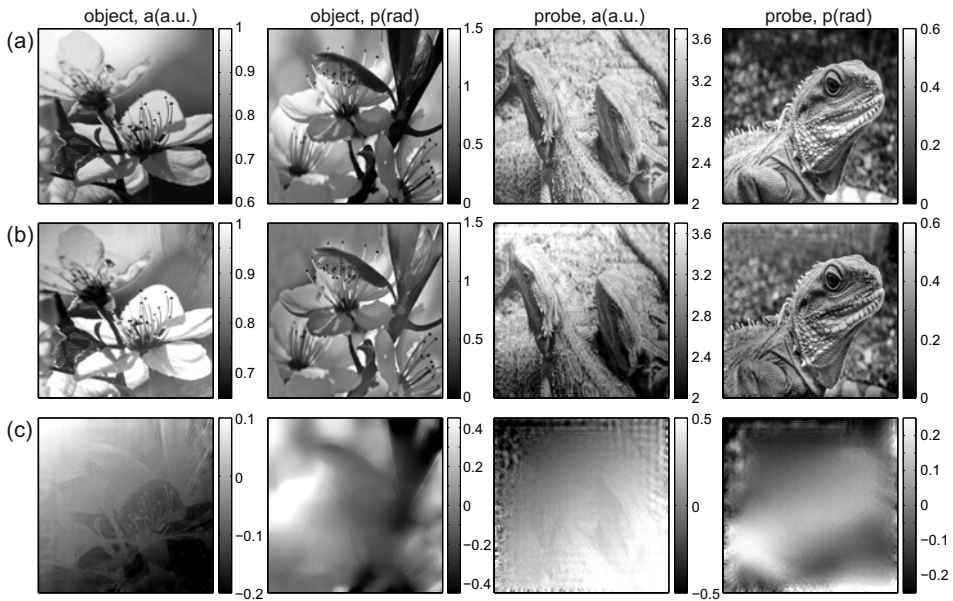
Figure 4.14: (a) Models for object and probe in amplitude and phase. (b) Reconstructed object and probe. (c) Difference between the images shown in (a) and (b). Global modulations of contrast as well as local artifacts at the limits of the field of view and in low frequency regions of the amplitudes become visible.

parameters were set to 0.3^9. After 40 iterations the algorithm was terminated;

---

[9] These are typical values for α and β which were chosen for most of the presented reconstructions. However, appropriate values have to be found by trial and error for each data set independently.

reconstructions of object and probe are depicted in Fig. 4.14(b).

A first inspection reveals that – despite some differences in contrast – the results for object and probe mirror the originals quite well. Even small details in the probe are found correctly.

However, by closer examination of the object's amplitude, it turns out that especially in regions of low spatial frequencies and hence less structure, sharp components of the reconstructed phases become visible (for example in the upper right corner). This might be induced by numerical artifacts which occur when performing propagation of artificially extended wave fronts. Even though the near-field propagator was oversampled for each distance, and thus extension of the field of view would not have been necessary, it turns out that enlarging the field of view for propagation reduces in particular disturbances at the edges of the reconstructions. Here, still some of these named edge effects are visible in the corners of the reconstructed amplitude of the probe.

In addition to these local errors, global differences between the originals and the reconstructions become visible when subtracting the images shown in Fig. 4.14(b) from those shown in Fig. 4.14(a). The differences in contrast are either a low frequency background consisting of an approximately linear gradient that modulates the reconstructions, or a more complex background that in particular reduces the contrast between extended neighboring and nearly homogeneous regions of different gray values.

While imprints of the phases in the amplitudes are problems that are very unlikely to negatively influence the reconstructions of physical objects (edges in the amplitude will also occur as edges in the phase – simply by the nature of physical objects), the reasons for global modulations in contrast and techniques to suppress them have to be studied in future work.

Another way to judge the performance of an algorithm is to calculate the mean squared contrast gap G_f of amplitude and phase of object and probe within one sub-iteration n

$$G_f^n = \sqrt{\frac{1}{M \cdot N} \sum_{x,y} \left(f^{n-1}(x,y) - f^n(x,y)\right)^2}, \qquad (4.353)$$

where f^n is either $|o^n|$, $\arg(o^n)$, $|p_1^n|$ or $\arg(p_1^n)$ and M, N are the numbers of pixels in horizontal/vertical direction. In addition, the mean squared error E^n in iteration n between the reconstructed hologram $\sqrt{\tilde{I}_{j,\mathrm{det}}^{(i),n}(x,y)}$ and the measured hologram $\sqrt{I_{j,\mathrm{det}}^{(i)}(x,y)}$ can be calculated

$$E^n = \sqrt{\frac{1}{M \cdot N} \sum_{x,y} \left(\sqrt{\tilde{I}_{j,\mathrm{det}}^{(i),n}(x,y)} - \sqrt{I_{j,\mathrm{det}}^{(i)}(x,y)}\right)^2}. \qquad (4.354)$$

The error metric E^n is evaluated each time when the magnitude constraint is applied; G_f^n is calculated each time when object and probe are updated. This means, within one global iteration k, E^n and G_f^n are determined m times, where
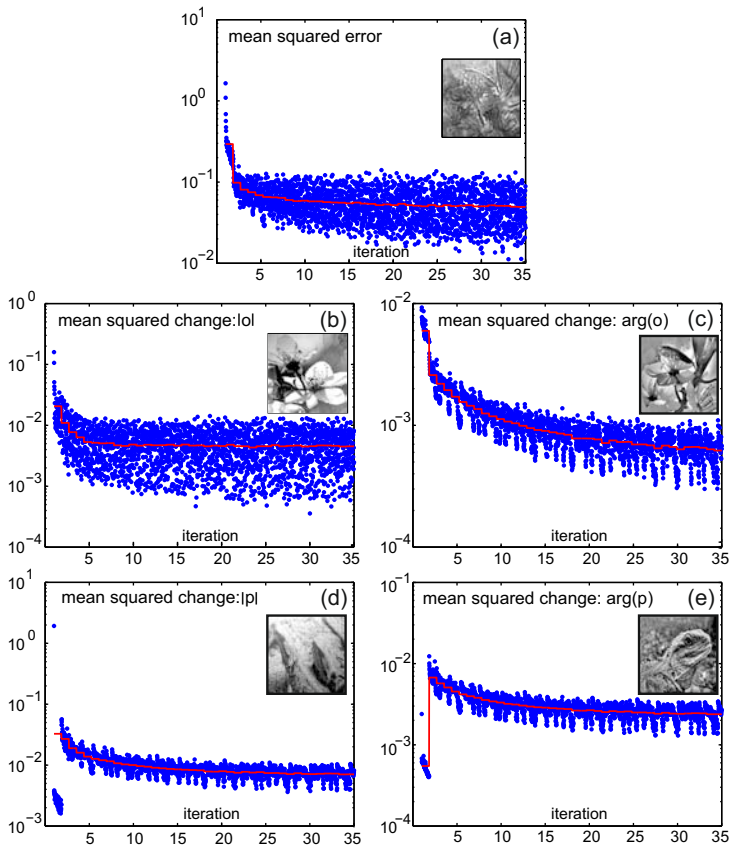
Figure 4.15: (a) Mean squared error of the reconstructed holograms with respect to measured/simulated holograms. (b) Mean squared change of $|o^n|$. (c) Mean squared change of $\arg(o^n)$. (d) Mean squared change of $|p_1^n|$. (e) Mean squared change of $\arg(p_1^n)$.

m is the number of recorded holograms (see also the sketch of the algorithm in Fig. 4.7).

The values of E^n and G_f^n for the reconstructions shown in Fig. 4.14 are depicted in Fig. 4.15. The blue dots indicate the respective E^n or G_f^n, whereas the red lines are the average value of E^n or G_f^n within one global iteration k. By observing E^n (Fig. 4.15(a)), it can be seen that while during sub-iterations n the differences with the measured data fluctuate, the mean fluctuations (red line in Fig. 4.15(a)) of the respective global iterations k decrease. The same statement can be given regarding the mean squared change of $|o^n|$, $\arg(o^n)$, $|p_1^n|$ and $\arg(p_1^n)$ shown in Fig. 4.15(b)-(e). The jumps in Fig. 4.15(d) and (e) after the second iteration reflect the change of α – the relaxation parameter for the update of the probe.

4.7.2   Clean illumination

Whereas the simulation presented in the previous section reveals numerical difficulties, in the following, more realistic simulations were performed to illustrate challenges based on physical non-uniqueness - such as the elimination of twin image features. For this purpose an object with refractive index $n = 1 - \delta + i\beta$ (where $\beta = 1 \cdot 10^{-6}$, $\delta = 1 \cdot 10^{-5}$) and 0.2 µm thickness was simulated (see Fig. 4.16(a), images on the left hand side). It was exposed to a basically uniform probe with a Gaussian amplitude profile (amplitude values varying from 6.3 to 5.9) and zero phase (see Fig. 4.16(a), images on the right hand side). Wavelength, field of view and Fresnel numbers are listed in Table 4.1 and are identical to those of the previous simulation. Same is true for the initialization of phase retrieval. To save computation time, the artificial holograms were binned by a factor of two[10]. The lateral step size covered 20 binned pixels in horizontal and vertical direction. The reconstructions after 35 iterations are depicted in Fig. 4.16(b).
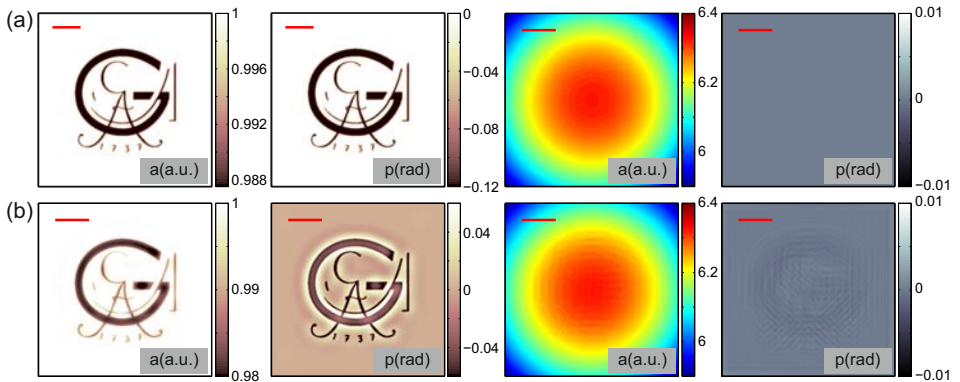
Figure 4.16: a) Models for object and clean probe with uniform, flat phases. (b) Reconstructed object and probe. Scale bars denote 20 µm.

The reconstructed object appears sharp and clean. Concerning the reconstructed amplitudes, the difference in contrast between dark and light regions can be sufficiently reproduced. Yet, especially the phases show a considerable enhancement of contrast around dominant edges. Finer structures are less affected by this phenomenon. Up to a global phase shift and a loss in contrast, the phases of the object can be found reliably. The reconstructed probe reveals some modulations in amplitude and phase which can be attributed to propagation as well as to separation of object and probe. However, these modulations are weak enough such that they basically do not contribute and the reconstruction of the probe was successful.
In chapter 4.2 it was demonstrated that near-field ptychography using an extended illumination requires a different strategy to reduce and eliminate twin image artifacts spoiling the reconstruction of the object. It was graphically motivated in

---

[10]Note that binning by a factor of two reduces the total number of pixels to 512×512. Fresnel numbers over 10 (binned) pixels are [0.40, 0.44, 0.56, 1.32].
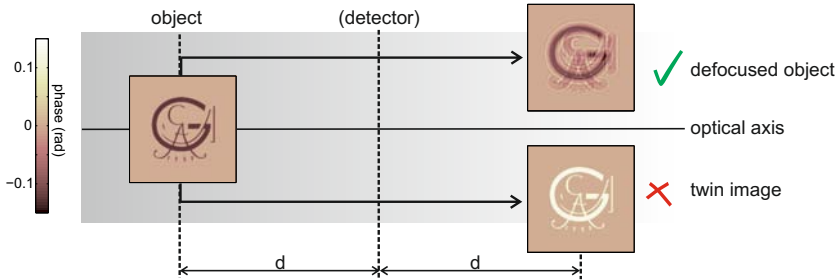
Figure 4.17: The reconstructed object is propagated to its twin image plane. Phase retrieval has been successful, if a defocused object becomes visible. Phase retrieval has failed to eliminate the twin image, if a sharp complex conjugated image of the object appears.

Fig. 4.2 and 4.3 that by using holograms with exclusively lateral diversity, twin image features will be combined in a functional way, whereas by using holograms with longitudinal diversity, differently defocused twin images will add up such that no single distance can be addressed to the resulting defocused components in the reconstructed object. Further iterations of phase retrieval will favor the image instead of its twin. To strengthen this statement, nine simulations were performed, each with a different amount of lateral and longitudinal diversity.

The reconstructed objects were defocused by twice the sample-detector distance, i.e. they were propagated to the twin image plane. Figure 4.17 illustrates that in the twin image plane we expect to either observe the phases of the properly defocused object, or the phases of the complex conjugated, focused object. In the first case, phase retrieval has successfully eliminated twin image components. In the second case, phase retrieval has not been able to reconstruct the object without its twin.
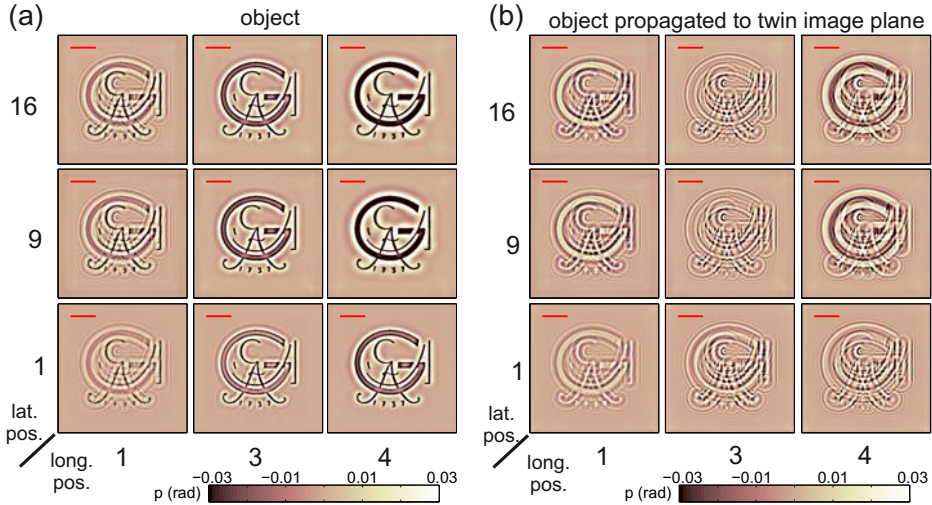
Figure 4.18: Illustration of the importance of longitudinal diversity for near-field imaging. (a) Phases of the reconstructed objects in clean illumination using holograms of different amount of lateral and longitudinal diversity. (b) Phases of (a) in the twin image plane. Scale bars denote 20 µm.

Fig. 4.18 summarizes the following results:

(1) In case of an extended, clean illumination covering the whole object, lateral diversity does not enhance the quality of the reconstructed object. No differences can be noticed when comparing the simple holographic reconstruction (Fig. 4.18,(a) one lateral, one longitudinal position) with reconstructions based on exclusively lateral diversity (Fig. 4.18(a), 9 and 16 lateral, one longitudinal position).

(2) To demonstrate that the unwanted artifacts in the reconstructed object are indeed features resulting from the twin image, Fig. 4.18(b) depicts the propagation of the objects shown in (a) to the twin image plane at twice the sample-detector distance. If the complex conjugated of (a) is visible in (b), the blurred components of (a) belong to the virtual image (sharp in (b)). This behavior is distinctly visible for reconstructions based on holograms with exclusively lateral diversity.

(3) Adding longitudinal diversity to the holograms almost instantaneously reduces twin image components and the propagation of the reconstructed object to the conjugated plane does not contain sharp complex conjugated features anymore.

(4) Further increase of (lateral and longitudinal) diversity increases the contrast of the reconstructions, but on the other hand, it slows down computation.

To complete the above discussion, the change of $\arg\left(o^{(n)}\right)$ for the nine simulations with different lateral and longitudinal diversity is presented in Fig. 4.19. Whereas the increase of lateral diversity leads to stagnation of phase retrieval, longitudinal diversity contributes to larger changes of $\arg\left(o^{(n)}\right)$ in successive iterations. As in Fig. 4.15, the red line describes the mean change within one global
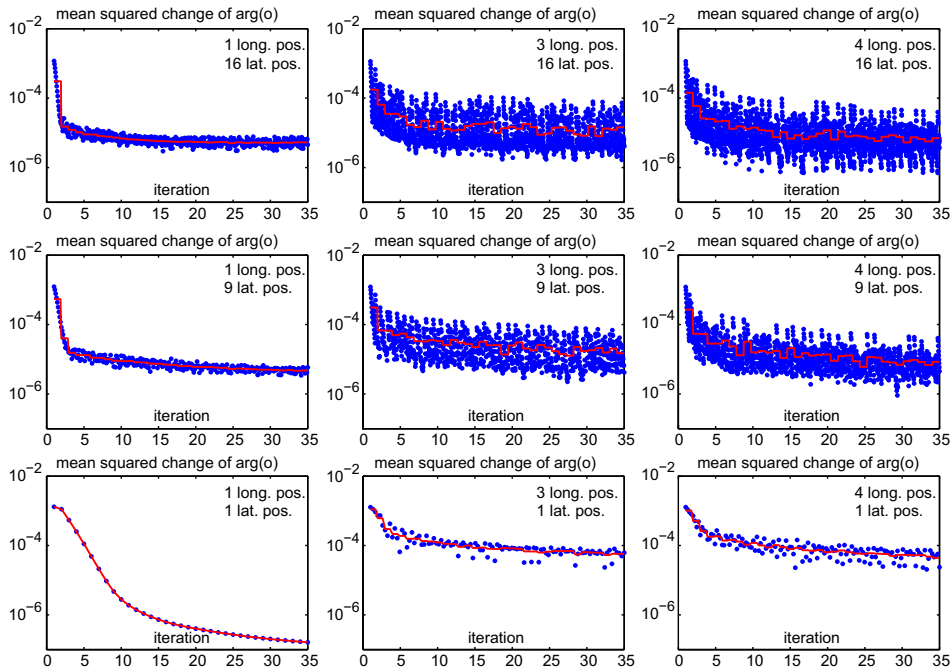
Figure 4.19: Mean squared change of arg(o) for simulations with different amount of lateral and longitudinal diversity.

iteration containing m sub-iterations, where m is the total number of lateral and longitudinal scan points[11].

In Fig. 4.18, it can be seen that the quality of the reconstructed object obtained from holograms with only longitudinal diversity is already satisfying. Under these circumstances, what is the purpose of lateral diversity for near-field imaging?

One answer is that it considerably improves the quality of the reconstructed probe. This is laid out in Fig. 4.20: Whereas the separation of object and probe does not work for a single hologram or holograms of only longitudinal diversity (left hand side in Fig. 4.20(a) and (b)) [82], it works very well for holograms containing lateral diversity and a mixture of lateral and longitudinal diversity (right hand side in Fig. 4.20(a) and (b)).

Summarizing these insights, it requires lateral diversity to reconstruct the probe and longitudinal diversity to reconstruct the object.

Still, the analysis based on the visualization of the twin image could be contestable: The amount of holograms and consequently the amount of diversity that entered the simulations with exclusively lateral diversity (16 lateral positions, i.e. 16 holograms or 9 lateral positions and 9 holograms) is much less than the amount of

---

[11]Note that a very smooth curve of G_f^n (or E^n) does not forcibly indicate reconstructions of better quality.
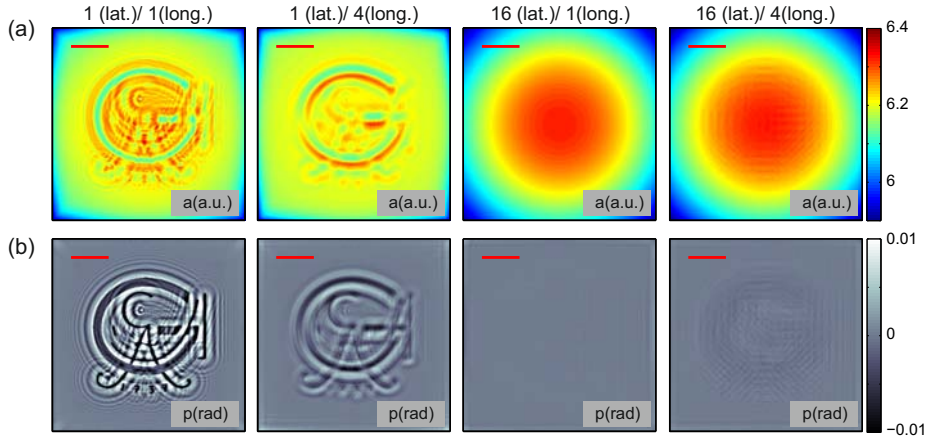
Figure 4.20: Illustration of the purpose of lateral diversity. (a) Amplitudes of the reconstructed clean probe using holograms of different amount of lateral and longitudinal diversity. (b) Phases of the reconstructed clean probe using holograms of different amount of lateral and longitudinal diversity. Scale bars denote 20 μm.

holograms (64 in total) used for the most successful simulation where lateral and longitudinal diversity were combined.

It is possible to argue that once the total amount of holograms containing no matter what kind of diversity is the same, exclusively lateral diversity should lead to the same results as longitudinal and lateral diversity together. Yet, this is not the case: An additional simulation was carried out with 64 artificial holograms, all with only lateral diversity. The lateral step size was set to 10 (binned) pixels, while the whole field of view consisted of 512×512 (binned) pixels. The reconstructions obtained from this simulations are directly comparable to the reconstructions obtained from the simulation based on 16 lateral scan points (step size of 20 (binned) pixels) for each of the four defocus positions. Results are depicted in Fig. 4.21 which shows (a) the reconstructions from data using 64 holograms that contain exclusively lateral diversity and (b) the results of reconstructions from 64 holograms of mixed lateral and longitudinal diversity. Again, the superiority of near-field ptychography using lateral and longitudinal shifts over near-field ptychography relying only on lateral diversity is clearly emphasized.
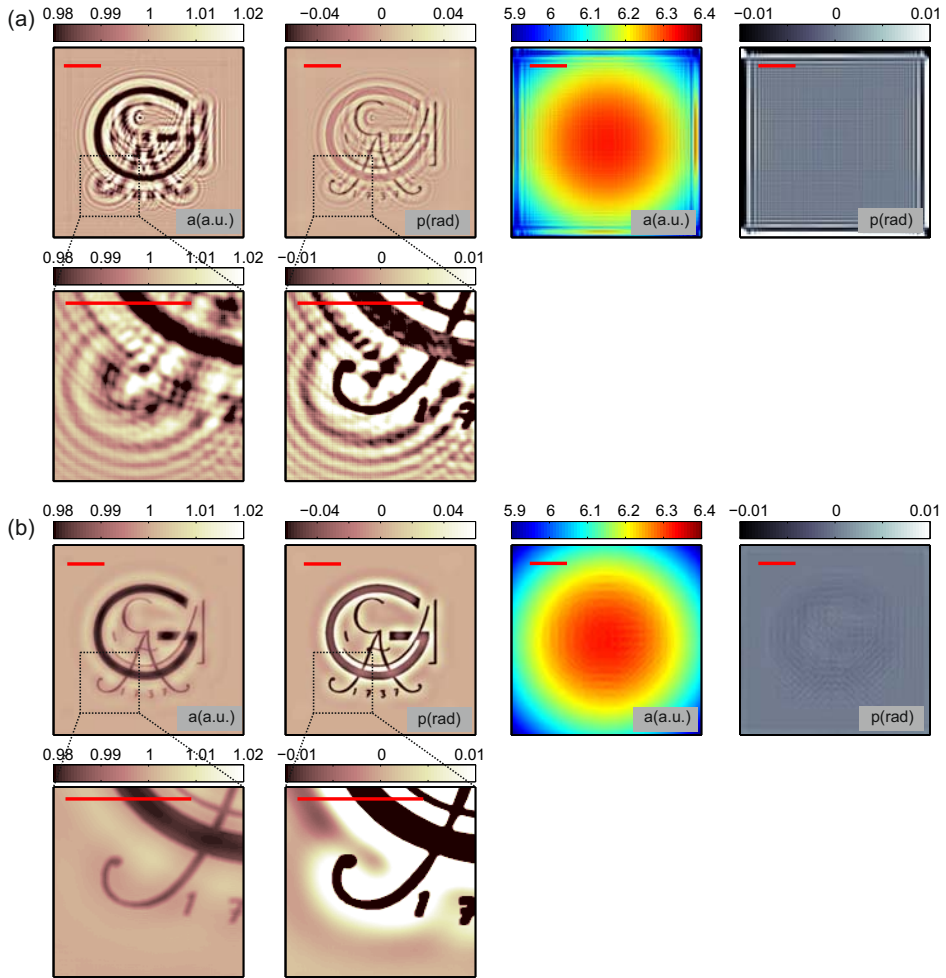
Figure 4.21: Simulations with 64 holograms in total. (a) Near-field ptychography with only lateral diversity. (b) Near-field ptychography with lateral and longitudinal diversity. Scale bars denote 20 μm.

4.7.3   Modulated illumination

In reality, the probe is never perfectly clean. It was observed in [79] that a strongly structured probe enhances the quality of the reconstructed object. In the following, this behavior shall be motivated and additional insights will be given.
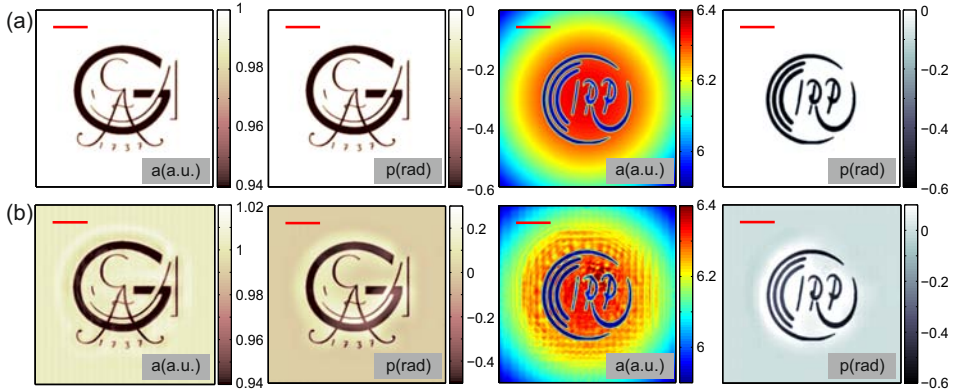
Figure 4.22: a) Models for object and modulated probe. (b) Reconstructed object and probe. Scale bars denote 20 μm.

The simulated object (Fig. 4.22(a), first two images from the left) is assumed to have a refractive index of $n = 1 - \delta + i\beta$ (where $\beta = 1 \cdot 10^{-6}$, $\delta = 1 \cdot 10^{-5}$) and 1 μm thickness. The probe (essentially the same as in chapter 4.7.2) was modulated by a second structure of identical refractive index and also 1 μm thickness (see Fig. 4.22(a), first two images from the right). Distances and further parameters were chosen as in section 4.7.2 (see Table 4.1) for the binned configuration. An example of a hologram is depicted in Fig. 4.13(c). Besides the modulus projection for the exit wave and the separation of object and probe, no further constraints were used for phase retrieval.

Figure 4.22(b) displays the reconstructions after 35 iterations of near-field ptychography. The conclusions which were obtained by studying the clean beam setup and which are related to the overall contrast of the object are equally true for the simulation with modulated probe. The amplitude of the reconstructed probe displays stripes (probably an artifact of propagation of a limited field of view and of the use of several lateral shift operations). The reconstructed phase of the probe is of slightly reduced contrast but shows an enhancement of contrast around edges. Besides this, both, the reconstructed amplitude and the reconstructed phase of the probe are free of artifacts.

Like in the previous section, the significance of lateral and longitudinal diversity is illustrated by comparing nine reconstructions that differ by the amount of holograms which were used for phase retrieval. The reconstructions of object and modulated probe are propagated to their respective twin image planes. A reasonable solution for object and probe is obtained, if in these twin image planes, properly defocused images are visible (see Fig. 4.23).
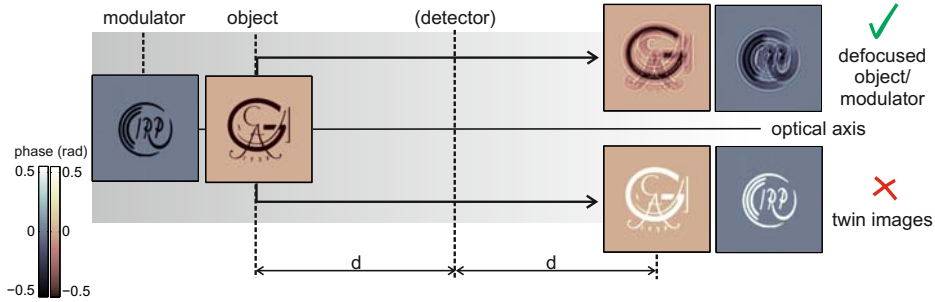
Figure 4.23: The reconstructed object and the reconstructed, modulated probe are propagated to the respective twin image planes. Phase retrieval has been successful, if a defocused object and a defocused modulated probe become visible. Phase retrieval has failed to eliminate the twin image, inherent in the object and inherent in the probe, if a sharp complex conjugated image of the object/the modulated probe appears.
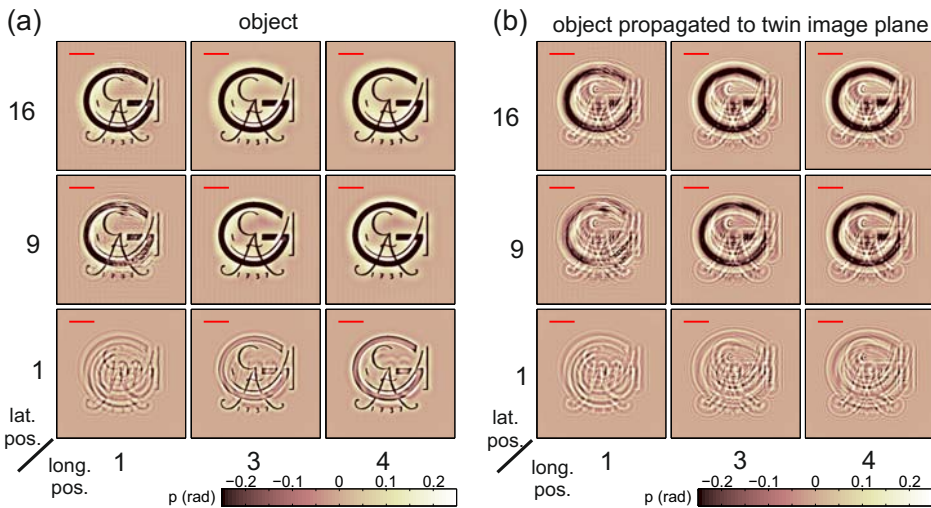
Figure 4.24: Illustration of the importance of longitudinal diversity under the influence of a modulated probe for near-field imaging. (a) Phases of the reconstructed objects using holograms of different amount of lateral and longitudinal diversity. (b) Phases of (a) in the twin image plane. Scale bars denote 20 µm.

In agreement with [79], we find that if the holograms used for phase retrieval contain only lateral diversity, and if the probe is considerably modulated, an improved quality of the reconstructed object can be observed (see Fig. 4.24(a), several lateral positions, one longitudinal position). However, as it can be seen in Fig. 4.24(b) (several lateral positions, one longitudinal position), some twin image details, especially in finely structured regions are still visible. In accordance with

the analogous examination presented in section 4.7.2, twin image components in the reconstructed object completely vanish when holograms with longitudinal diversity are added to the reconstruction.
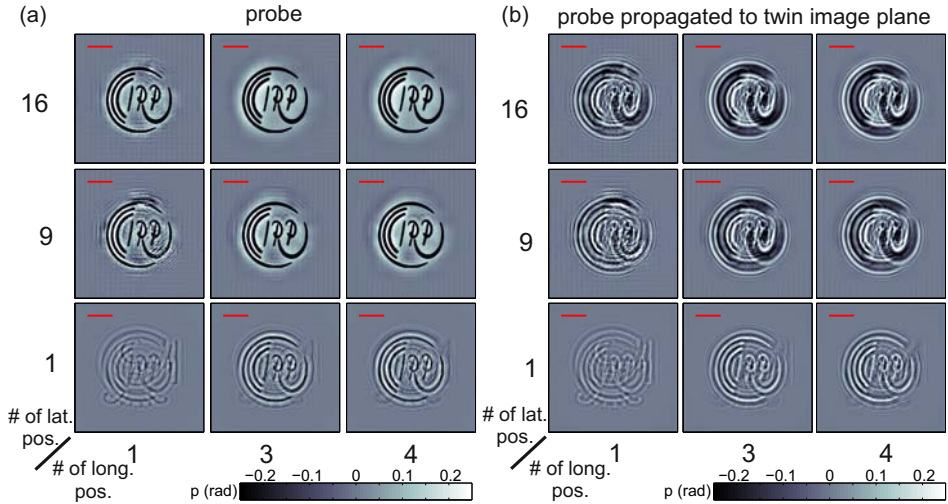
Figure 4.25: Illustration of the importance of lateral diversity for near-field imaging with simultaneous probe retrieval. (a) Phases of the reconstructed, modulated probes using holograms of different amount of lateral and longitudinal diversity. (b) Phases of (a) in the twin image plane. Scale bars denote 20 µm.

An interesting finding can be made when observing the corresponding reconstructed probes for these nine simulations with different amount of data (see Fig. 4.25). In accordance with the insights gained in the previous section, longitudinal diversity alone does not suffice to reconstruct the probe. The main features deteriorating the probe can be addressed to the twin image of the illumination (Fig. 4.25(b)). Lateral diversity in the holograms reduces twin image components of the probe in almost every case. It seems that lateral diversity fulfills an analogous task for the probe as longitudinal diversity for the object: The removal of twin image artifacts. This is the reason why lateral *and* longitudinal diversity are equally important for a successful near-field ptychography reconstruction.

In addition, for the nine simulations, in Fig. 4.26 the mean squared differences E^n between the measured/simulated holograms and the reconstructed holograms are shown. Essentially two messages are provided: (1) A very smoothly decreasing error does not necessarily indicate high quality reconstructions for object and probe.

(2) Adding further and further lateral diversity to the data leads to stagnation, longitudinal diversity slowly pulls the error towards lower values.

Yet, as it can be seen from Fig. 4.26, the number of applied magnitude constraints
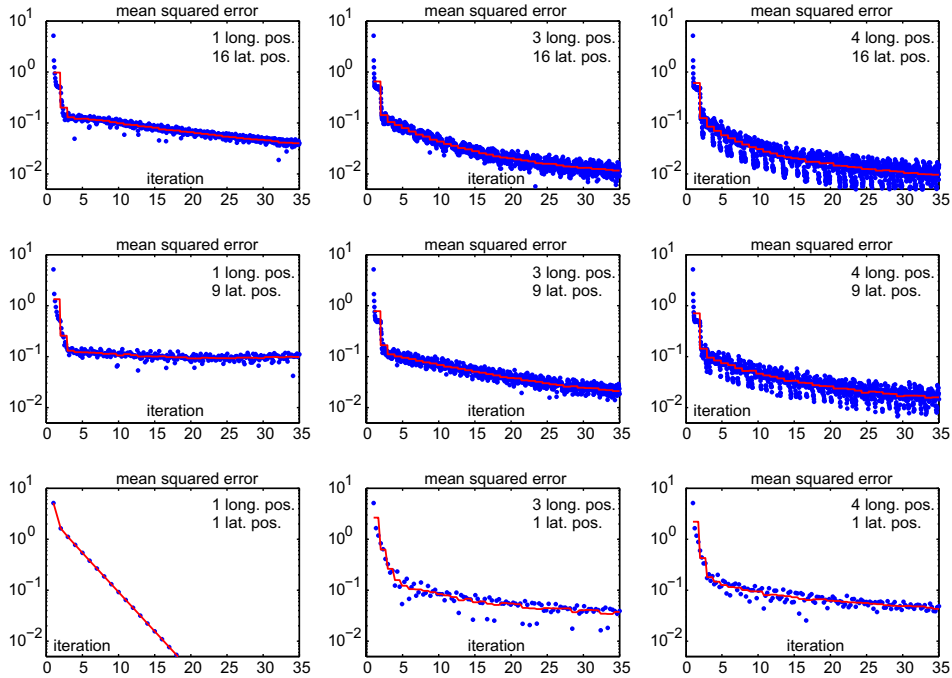
Figure 4.26: Mean squared error E^n in (sub-) iteration n, i.e. each time when a magnitude constraint was applied, for simulations with different amount of lateral and longitudinal diversity. The red line shows E^n averaged over one global iteration k.

depends on the number of available holograms.

To illustrate the problems and benefits arising from the use of holograms with only lateral diversity/mixed longitudinal and lateral diversity and to present a fair comparison between near-field ptychography using lateral diversity and near-field ptychography using lateral and longitudinal diversity for phase retrieval, 64 holograms with exclusively lateral diversity were simulated in the given setup. The lateral step size was 10 (binned) pixels, with a total number of 512 (binned) pixels in each dimension. Phase retrieval was performed.

The results are depicted in Fig. 4.27(a). In agreement with [79], we find that the reconstruction of the object looks quite promising. However, a closer look still reveals weak twin image features. The reconstructed probe shows several artifacts, that may either result from numerous shifting operations or from numerical difficulties with near-field propagation. For comparison, the reconstructions of object and probe using 64 holograms of mixed lateral and longitudinal diversity are given in Fig. 4.27(b). Twin image features are completely removed and the reconstructed probe is much closer to the original (see Fig. 4.22(a)). Hence, despite the fact that the number of holograms was the same for both simulations and despite a structured illumination, the quality of the reconstructions is considerably better when

using holograms of lateral and longitudinal diversity. Again, the message is that lateral diversity alone is not sufficient for ptychography in the optical near-field.
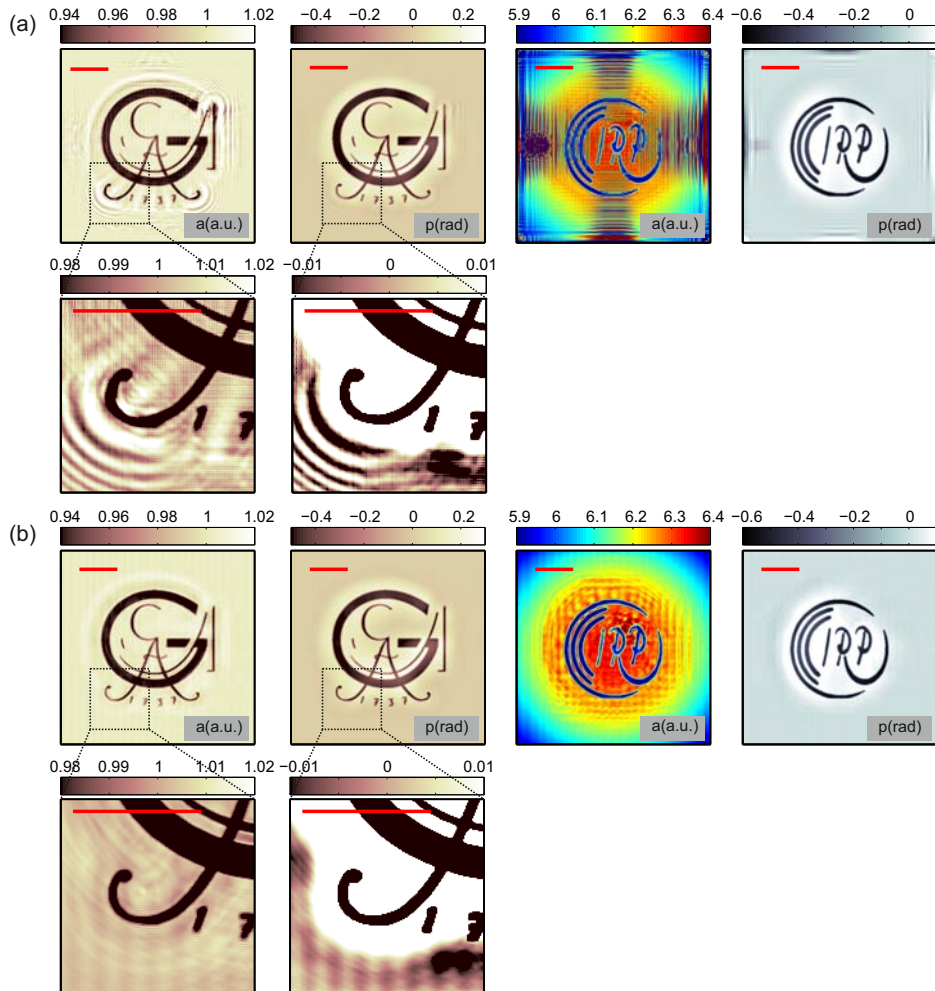
Figure 4.27: Comparison between ptychography using 64 holograms with lateral diversity (a) and ptychography using 64 holograms with lateral and longitudinal diversity (b). In both cases phase retrieval was performed for 35 iterations. The number of magnitude projections per iteration are the same for (a) and (b). Order of images from left to right: Amplitude of the reconstructed object, phase of the reconstructed object, amplitude of the reconstructed probe, phase of the reconstructed probe. Scale bars denote 20 µm.

4.7.4   Exploring limits: Objects with large phase shift

Near-field ptychography with lateral and longitudinal diversity has another important advantage: It is possible to reconstruct objects with large phase shifts. For this purpose, the thickness of the simulated object ('GAU'-logo) was changed from 1 µm to 3 µm resulting in a phase shift of 1.89 rad, which is larger than $\pi/2$ rad. The thickness of the beam modulating object ('IRP'-logo) was left unchanged at 1 µm. Two simulations of 35 iterations each were performed, using either 64 holograms (512×512 pixels) of exclusively lateral diversity (with lateral step size of 10 pixels and Fresnel number over 10 pixels of 0.40) or 64 holograms (512×512 pixels) containing lateral and longitudinal diversity (4 defocus positions and 16 lateral scan points per defocus position with lateral step size of 20 pixels and defocus shifts attributed to Fresnel numbers of [0.40, 0.44, 0.56, 1.32].
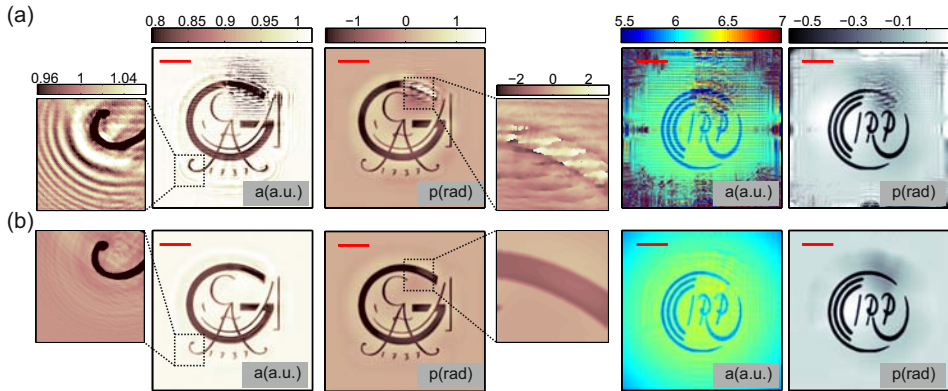
Figure 4.28: (a) Ptychography (35 iterations) of a thick object based on 64 holograms containing only lateral diversity. (b) Ptychography (35 iterations) of a thick object based on 64 holograms containing lateral and longitudinal diversity. Order of images from left to right: Amplitude of the reconstructed object, phase of the reconstructed object, amplitude of the reconstructed probe, phase of the reconstructed probe. Scale bars denote 20 µm.

The reconstructions are depicted in Fig. 4.28. Ptychography based on lateral diversity (Fig. 4.28(a)) shows two main problems, namely twin image artifacts (see zoomed region of the reconstructed amplitude of the object with adjusted contrast) and phase jumps between positive and negative values which can finally lead to phase wrapping artifacts (phase jumps between $-\pi$ rad and π rad). This can be seen in the zoomed region of the reconstructed phase of the object. Note that this is in contrast to the findings presented in [87]. The reason for this might be the choice of the wave front diffusing structure which for this simulation might not have been as strong as in [87]. Ptychography using lateral and longitudinal diversity (Fig. 4.28(b)) is much more robust with respect to both described prob-

lems: Twin image components are almost completely eliminated and phase jumps do not occur that quickly.

Summary

In this chapter, a thorough derivation of near-field ptychography and the underlying concepts, as well as technical details concerning the implementation were provided. To this end, an adaption of the ePIE scheme presented in the previous chapter was chosen. Essential ingredients of near-field ptychography are holograms that contain lateral *and* longitudinal diversity. Hence, a ptychographic data set for the holographic near-field is recorded by shifting the sample along and perpendicular to the optical axis. In contrast to far-field imaging, shifting the sample along the longitudinal direction provides more diversity to the data and allows to eliminate holographic twin image ambiguities. Furthermore, near-field ptychography enables reconstructions of objects with strong phase shifting properties. These findings were illustrated by selected simulations with completely artificial and more realistic objects and probes.

While simulations based on artificial phantoms are used to visualize technical problems influencing the reconstructions, realistic objects and probes are used to provide a preview of what can be expected when near-field ptychography is applied to experimental data.

5 Experimental results

This chapter presents near-field ptychographic imaging experiments. Starting from preliminary studies with coherent, visible light, near-field ptychography is used to reconstruct object and probe from holograms recorded with coherent, parallel and cone beam X-rays at synchrotron sources. Parts of the content of this chapter – especially section 5.1 and section 5.3.1 – can be found in [85] and are reviewed here in more detail.

Extended and confined solid state test structures serve as samples. They are illuminated by clean and intentionally modulated wave fronts. An analysis of the reconstructed refractive index of one of the test structures is given. Furthermore, near-field ptychography is compared to reconstructions obtained from inversion of the contrast transfer function. It will turn out that near-field ptychography allows for imaging with high resolution and low noise.

5.1 Near-field ptychography with coherent visible light

5.1.1 Experimental setup
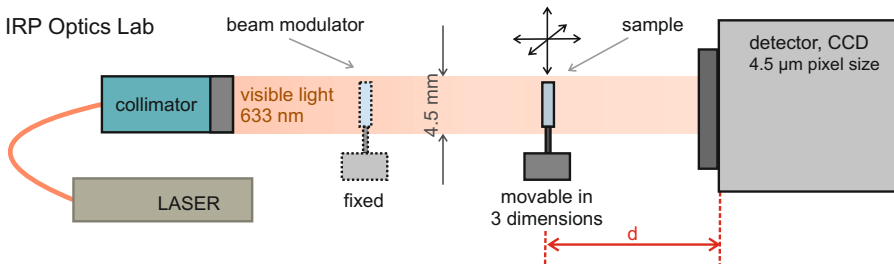
Figure 5.1: Setup for visible light near-field ptychography: A laser beam of 633 nm wavelength is coupled into a monomode optical fiber attached to a collimator. Two objects – one fixed for intentional beam modulation and one movable in all three dimensions of space – can be inserted in the beam path. The detector is a CCD camera with pixel size of 4.5 µm. The distances d between the movable sample and the detector were chosen such that the sampling condition of the Fresnel near-field propagator was fulfilled.

Near-field ptychography was experimentally tested using the setup for visible light holography shown in Fig. 5.1. The probe (a laser beam of 633 nm wavelength) is coupled into a monomode optical fiber attached to a collimator of numerical aperture NA=0.14 (model 60FC-L-0-M60-33, Schäfter+Kirchhoff, Germany). This leads to a coherent, monochromatic, extended and parallel beam of about 4.5 mm in diameter. As sketched in Fig. 5.1, there are two possibilities to insert an object into the beam path: (1) A fixed sample holder allows for intentional beam modulation by a suitable test structure; (2) at a distance of 13 cm behind this modulation position a motorized xyz-stage is installed to enable translations of

the object of interest in all three dimensions of space. The zero point of the sample stage is at a distance of 6.5 cm with respect to a stationary CCD detector (CoolSnap Myo, Photometrics, USA). Distances are chosen such that the sampling condition of the Fresnel propagator (see equation 3.184) is fulfilled.
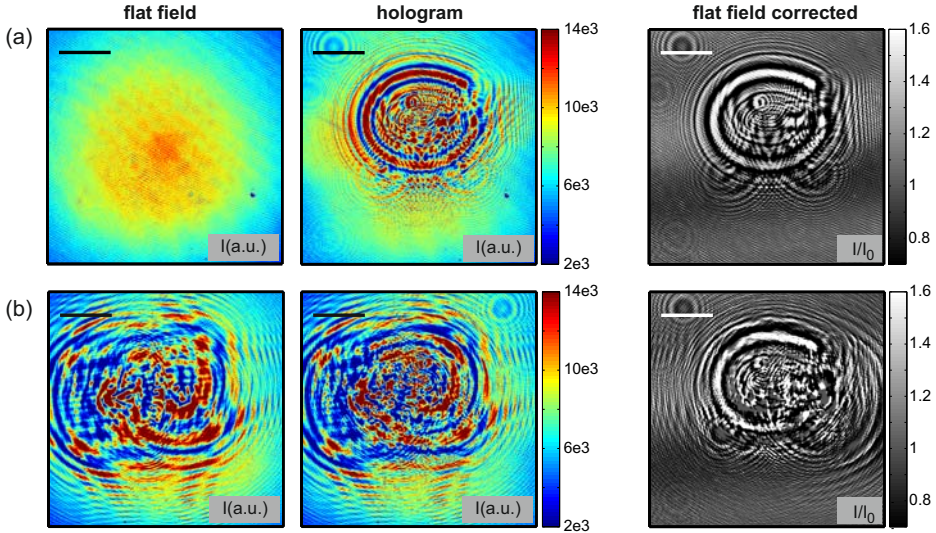
Figure 5.2: (a) Left: Flat field of unmodulated laser beam. Center: Hologram of GAU-logo. Right: Hologram of GAU-logo after conventional flat field correction. (b) Left: Flat field of modulated laser beam. Center: Hologram of GAU-logo. Right: Hologram of GAU-logo after conventional flat field correction. Scale bars denote 0.5 mm.

Two separate experiments were carried out to demonstrate near-field ptychographic phase retrieval. At first, a sample (photo resist on glass with logo 'GAU'; the structured area covered approximately $2 \times 2\,\mathrm{mm}^2$) was attached to the movable sample stage and shifted to distances $d \in [6.5, 6.0, 5.6, 5.2, 4.9]$ cm with respect to the detector. At each of these positions along the optical axis, the sample was scanned to 16 positions inside the plane perpendicular to the beam path with a lateral step size of 0.85 mm in horizontal and 0.68 mm in vertical direction. Holograms were recorded at each scan point as detailed in Table 5.1.

Next, another test pattern (photo resist on glass with logo 'IRP'; the structured area covered approximately $2.5 \times 2\,\mathrm{mm}^2$) was additionally positioned in the beam behind the collimator to intentionally modulate the probe. Similar to the first experiment, the movable sample (logo 'GAU') was translated along and perpendicular to the optical axis. Details can also be found in Table 5.1.

Figure 5.2 depicts selected raw data and a flat field corrected hologram of the first experiment (a) and of the second experiment (b), respectively. The first column shows the intensity profile of the probe (top: non-modulated, bottom: modulated); the second column provides illustrative examples of holograms for both experiments. The third column shows the same holograms after conventional

object	photo resist (1.4 µm) on glass
	structured area approximately 2×2 mm^2
beam modulation	no/ photo resist (1.4 µm) on glass
	structured area approximately 2.5×2 mm^2
geometry	parallel beam
wavelength [nm]	633
pixel size [µm]	4.54
number of pixels	1940×1460
accumulation time [ms]	10×6
F_{10}	0.051, 0.054, 0.057, 0.062, 0.067
lateral step size [mm]	0.85 (horiz.), 0.68 (vert.)
sample-detector distances [cm]	6.5, 6.0, 5.6, 5.2, 4.9
# holograms	80 (5 defocus planes,
	each with 16 laterally shifted holograms)
image corrections	dark image, faulty pixel mask

Table 5.1: Parameters for the visible light in-line holography experiments. Fresnel numbers $F_{10} = a^2/z\lambda$ are calculated for a structure size of $a = 10\delta$ with pixel size δ and wavelength λ.

flat field correction. In both cases (clean and modulated probe) not all features resulting from the illumination can be eliminated by division of the hologram by the flat field.

5.1.2 Results of near-field ptychography for visible light

To reconstruct object and probe, the ptychographic near-field algorithm described in chapter 4 was initialized with a homogeneous amplitude distribution for object and probe with zero phases. During the first three iterations only small changes were allowed to influence the probe. Twin suppression (see equation 4.322) was applied during the first seven iterations. Automatic focusing (see equation 4.325) was used at the beginning of iteration two, four and six assuming the object to be weakly absorbing. Phase retrieval was performed for 20 iterations. Results are depicted in Fig. 5.3, 5.5 and 5.7.

Phase retrieval with high image quality and separation of object and probe could be demonstrated for a clean as well as for a modulated probe. Note that whereas the reconstructed amplitude of the object is of rather weak contrast, the retrieved phase is strongly varying (the contrast in the reconstructed phase is about $\pi/2$ rad). The resolution of the reconstructed object was determined as described in section 2.4.3 by (1) selecting a region of interest and fitting step functions to edges of this region, (2) calculating the power spectral density and (3) angular averaging the power spectral density. Results are shown in Fig. 5.4. For both cases – the experiment with clean illumination and the experiment with modulated beam – the achieved resolution is between 20 µm and 10 µm.
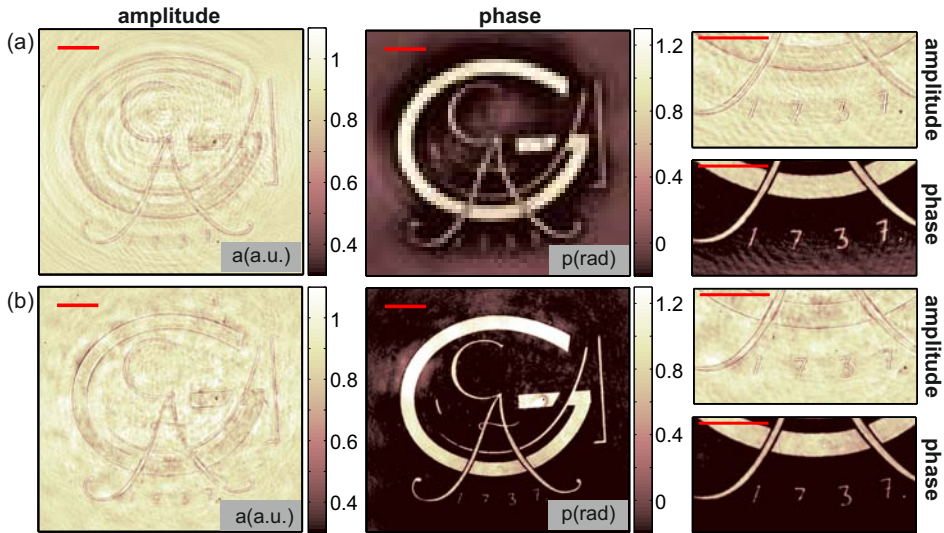
Figure 5.3: (a) Reconstructed object in amplitude (left) and phase (center) for a clean illumination. Right: Small detail of the reconstruction. (b) Reconstructed object in amplitude (left) and phase (center) for a modulated illumination. Right: Small detail of the reconstruction. Scale bars denote 0.5 mm
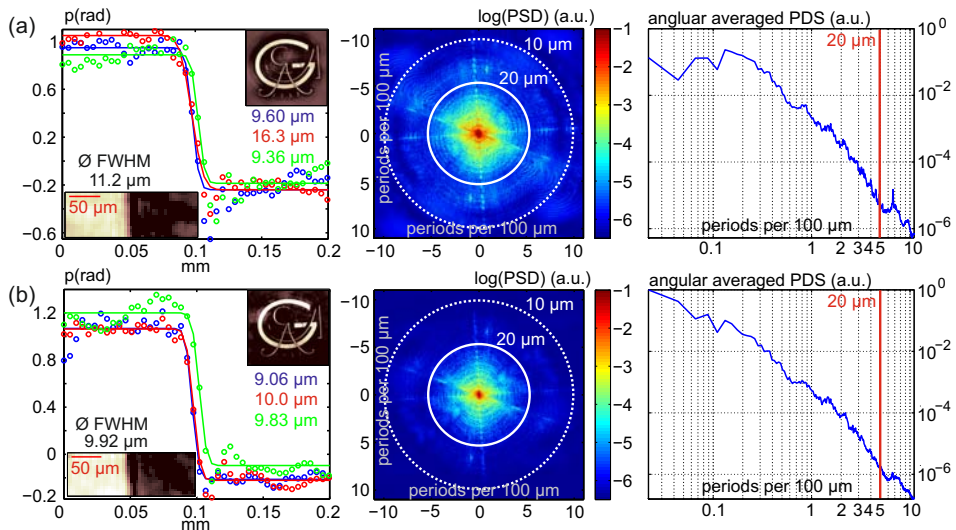
Figure 5.4: (a) Determination of the object's resolution (full period) for the experiment with clean illumination. Left: A region of interest was extracted and error functions were adapted to the edges. Three examples are plotted. Center: The power spectral density of the reconstructed phases. Right: Angular averaged power spectral density. (b) Same analysis as in (a) for the experiment with modulated illumination.

The power spectral density in Fig. 5.4(a) reveals a slightly broader intensity distribution in the high frequency range than the power spectral density shown in (b).
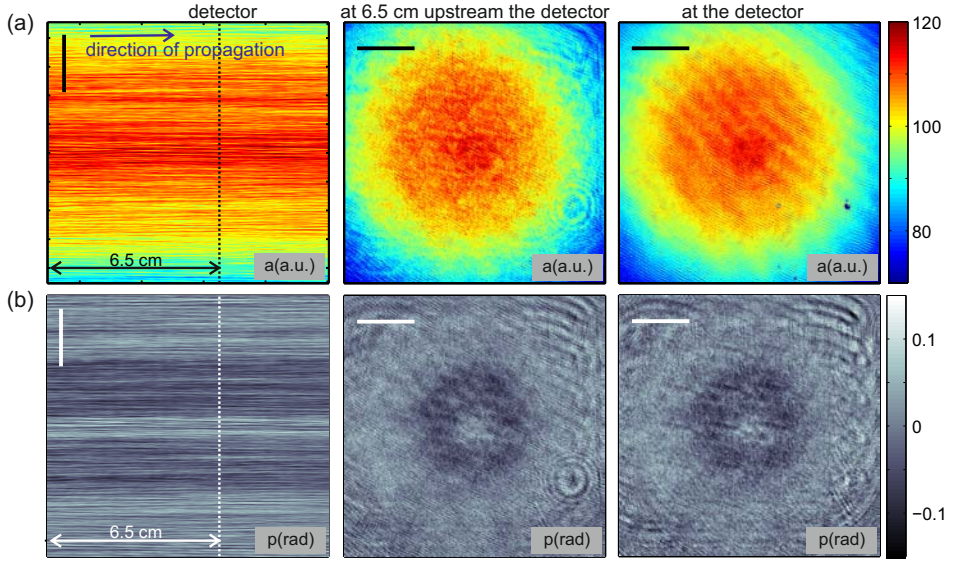
Figure 5.5: Reconstructed clean probe in (a) amplitude and (b) phase. Left: One-dimensional cut through the reconstructed amplitude (a) and phase (b) profile for different distances along the optical axis. Center: Reconstructed amplitude (a) and phase (b) of the probe at distance $d = 6.5$ cm upstream the detector (at the starting position of the sample). Right: Reconstructed amplitude (a) and phase (b) at the detector. Scale bars denote 0.5 mm.

To characterize the reconstructed clean probe (see Fig. 5.5), it was propagated for 10 cm along the optical axis starting from the farthest position d of the object with respect to the detector (Fig. 5.5, left part). The overall amplitude and phase profile stays the same (as expected for a parallel beam illumination with nearly flat amplitudes and phases). The reconstructed amplitude at the detector (depicted in the left part of Fig. 5.5) qualitatively matches the measured flat field (depicted in Fig. 5.2). The phase distribution of the probe shows a torus shaped profile in the region of highest intensity (Fig. 5.5 (b)). Since light that has passed through a monomodal fiber should not reveal these aberrations, it is most likely that they result from the collimating lens. To further characterize the phase profile, the center of mass of the reconstructed intensity was determined. Around the center of mass, a region of interest was extracted. The reconstructed phase within this region of interest was low pass filtered (see Fig. 5.6(b), left image) and angular averaged.
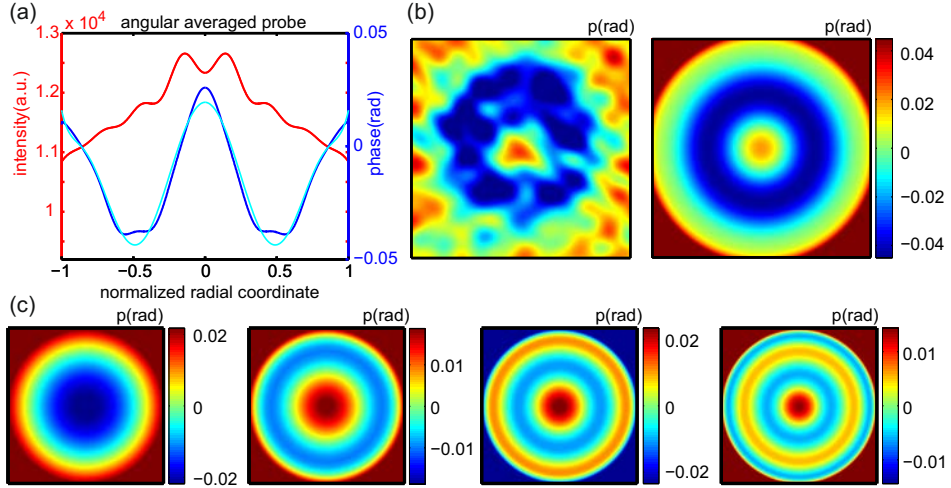
Figure 5.6: (a) Intensity profile (red) and phase profile (dark blue) of the reconstructed probe at $d = 6.5$ cm upstream the detector. The phase profile was fitted to a linear combination of radial Zernike polynomials (light blue curve). (b) Low pass filtered, reconstructed phase profile of the probe (left) and two-dimensional representation of the corresponding linear combination of Zernike polynomials (right). (c) Single radial Zernike polynomials with their respective weights used to explain the reconstructed phase profile.

A linear combination of Zernike polynomials [28] with radial coordinate ρ

$$f(\rho) = a + b \cdot R_2^0(\rho) + c \cdot R_4^0(\rho) + d \cdot R_6^0(\rho) + e \cdot R_8^0(\rho) , \tag{5.355}$$

with

$$R_2^0(\rho) = 2\rho^2 - 1 , \tag{5.356}$$
$$R_4^0(\rho) = 6\rho^4 - 6\rho^2 + 1 , \tag{5.357}$$
$$R_6^0(\rho) = 20\rho^6 - 30\rho^4 + 12\rho^2 - 1 , \tag{5.358}$$
$$R_8^0(\rho) = 70\rho^8 - 140\rho^6 + 90\rho^4 - 20\rho^2 + 1 \tag{5.359}$$

was then fitted to the low pass filtered and angular averaged phase profile. The result is depicted in Fig. 5.6(a), along with the two-dimensional representation of the adjusted phase in the right image of Fig. 5.6(b). The single Zernike polynomials (equations 5.356 - 5.359) are plotted in Fig. 5.6(c) with their respective weights. The weights a, b, c, d, e were determined to

$$a = -0.015 \pm 2.1 \cdot 10^{-4} ,$$
$$b = 0.022 \pm 7.0 \cdot 10^{-4} ,$$
$$c = 0.019 \pm 9.0 \cdot 10^{-4} ,$$
$$d = -0.024 \pm 1.0 \cdot 10^{-3} ,$$
$$e = 0.014 \pm 1.1 \cdot 10^{-3} .$$

Hence, the main aberrations are due to a defocus term (R_2^0) and the spherical aberrations R_4^0 and R_6^0.
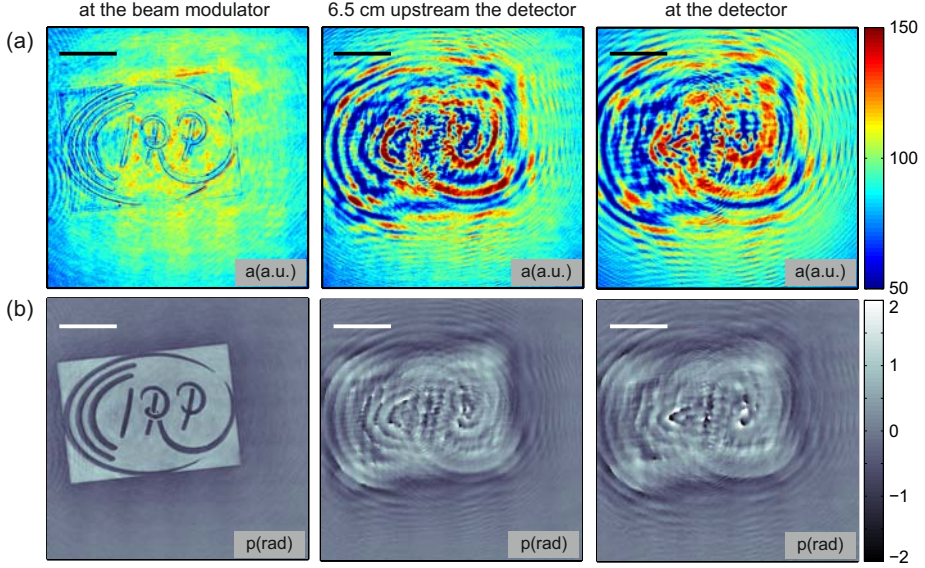
Figure 5.7: Reconstructed modulated probe in (a) amplitude and (b) phase in the modulation plane (left), at distance $d = 6.5$ cm upstream the detector (center) and directly at the detector (right). Scale bars denote 0.5 mm.

Next, the reconstructed probe of the data set with beam modulator was analyzed. To this end, the reconstructed probe was back propagated to the position of the beam modulator. Figure 5.7 depicts the amplitude (a) and phase (b) of the re-constructed, modulated probe directly in the modulation plane. As expected, the beam modulator is clearly visible in both amplitude and phase. Furthermore, amplitude and phase are shown at $d = 6.5$ cm upstream with respect to the detector (center column of Fig. 5.7). Finally, amplitude and phase directly at the detector are depicted in the left column.

5.1.3   An analysis regarding the twin image problem

Next, the following question should be addressed: What is the main contribution of lateral and longitudinal diversity to the reconstruction? Is it possible to verify the theoretical findings presented in section 4.7 in experiments? The suitability of lateral and longitudinal shifts to create diversity is investigated by further analysis of the data presented in part 5.1.2. In particular and in agreement with the simulations presented in section 4.7, it can be shown that both, lateral and longitudinal shifts are required to remove twin image artifacts in object and probe.
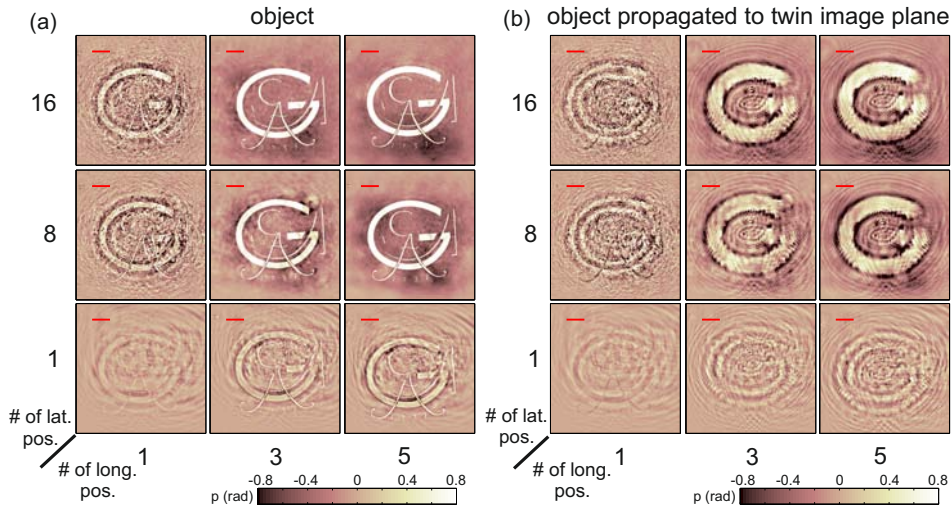
Figure 5.8: (a) Objects (phases) reconstructed by using holograms with varying amount of lateral and longitudinal diversity. (b) Reconstructed objects propagated to the twin image plane at a distance of 2×6.5 cm with respect to the farthest sample-detector position. Scale bars denote 0.5 mm.
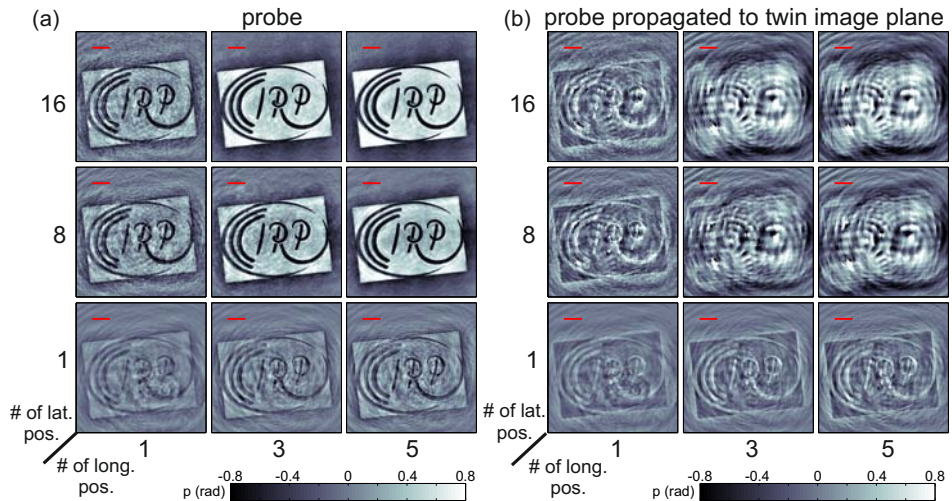
Figure 5.9: (a) Probes (phases) reconstructed by using holograms with varying amount of lateral and longitudinal diversity. (b) Reconstructed probes propagated to the twin image plane at a distance of $2 \times (13 + 6.5)$ cm downstream the position of the beam modulator. Scale bars denote 0.5 mm.

For this purpose, the amount of laterally and longitudinally shifted holograms (movable object, modulated probe) contributing to phase retrieval was successively reduced. For each of these reduced data sets, phase retrieval (without twin image suppression) was performed. The reconstructed object was propagated to its twin image plane, the reconstructed probe to the modulation plane and the twin image plane with respect to the beam modulator.

Results for the object are detailed in Fig. 5.8; results for the probe are depicted in Fig. 5.9. On closer examination of the reconstructed objects, the following points can be stated: (1) It is not possible to separate object and probe by using the information of a single hologram. (2) Using holograms containing only lateral diversity leads to high frequency artifacts and phase wrapping. This is in agreement with the simulation of strongly phase shifting objects presented in section 4.7.4. (3) Sharp complex conjugated components of the object can be observed in the twin image plane – an indication that lateral shifts alone do not contain enough diversity to solve the twin image problem. Again, this is in agreement with the simulations shown in section 4.7. (4) Exploiting only longitudinal diversity enhances the quality of the reconstructed object: Twin image features are reduced. However, longitudinal diversity alone does not allow for separation of object and probe. Features resulting from the probe are still visible in the reconstructed object. (5) It is the combination of longitudinal and lateral shifts that solves the twin image problem and at the same time separates object and probe.

Inspection of the reconstructed probes provides further insight into the importance of especially lateral diversity: (1) Longitudinal diversity alone is not able to remove twin image artifacts in the probe (see also [85]). (2) In contrast, lateral shifts of the object 'sample' the illumination inside an extended region perpendicular to the optical axis. The quality of the reconstructed probe is considerably enhanced by adding lateral diversity. (3) However, because the equations for separation of object and probe (see equations 4.334 and 4.335) are coupled, a convincing reconstruction of the probe can only be achieved along with a convincing reconstruction of the object and contrariwise. Also these findings correspond well to the insights provided by the simulations of section 4.7.

5.2   X-ray near-field ptychography in a parallel beam setting

Generalized ptychographic phase retrieval is demonstrated for full-field phase contrast radiography. Experiments were performed at the insertion device beam line ESRF/ID19 [88]. A sketch of the setup can be found in Fig. 5.10(a). The pink beam mode at a photon energy $E = 18.77 \, \text{keV}$ (first harmonic) was used.
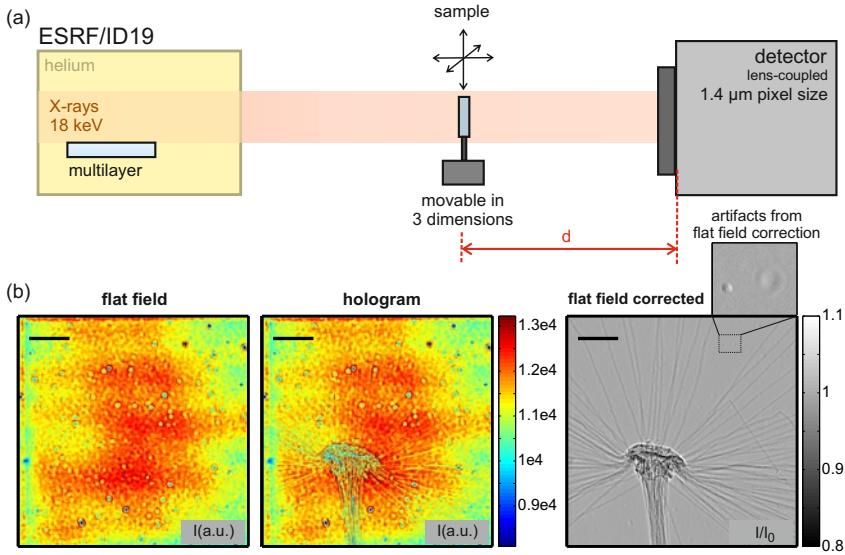
Figure 5.10: (a) Setup of the ESRF/ID19 imaging beamline with parallel probe. (b) Left: Flat field. Center: Hologram of dandelion seed. Right: Flat field corrected hologram. Scale bars denote 250 µm.

For this experiment a seed of dandelion attached between two polypropylene foils served as a sample.

It was translated to variable distances $d \in [20.56, 19.96, 19.36, 18.96]$ cm with respect to the stationary detector, a $Gd_3Ga_5O_{12}$:Eu single-crystal scintillator of 10 µm thickness with an optical microscope of $20\times$ magnification (NA=0.4) and a fast-read out CCD detector (FReLoN 2k, ESRF) [89]. At each position along the optical axis, 16 holograms with different lateral translations of the sample were recorded. The lateral step size was 0.21 mm. Further details can be found in Table 5.2. Figure 5.10(b) shows a measured flat field at ID19, an example of a hologram of the dandelion seed recorded in the regime of direct contrast and the flat field corrected hologram. Although the beam was comparably clean, features resulting from the probe are remaining in the flat field corrected hologram.

Near-field ptychography was initialized using a back-propagated flat field for the probe and a uniform distribution of amplitudes (zero phases) for the object. Holograms were binned by a factor of two to speed up computation. During the first iterations, automatic focusing was applied (see equation 4.325). The phases of the

Table 5.2: Parameters for x-ray holography at ESRF/ID19.

object	seed of dandelion
	attached to a polypropylene foil
geometry	parallel beam
wavelength [nm]	0.066
pixel size (after binning) [µm]	1.40
number of pixels (after binning)	1024×1024
accumulation time [s]	0.1
F_{10}	14.4, 14.8, 15.3, 15.6
sample-detector distances [cm]	20.56, 19.96, 19.36, 18.96
lateral step size [mm]	0.21
# holograms	64 (4 defocus planes,
	16 lat. scan positions)

object were projected to negative values (see equation 4.323). Phase retrieval was performed for 10 iterations in total.

The reconstructed, absolute value of the object's transmission function is depicted in Fig. 5.11(a). Two regions of interest revealing little barbs are shown next to the overview image. The full period resolution estimated by the power spectral density of the reconstructed amplitudes shown in (a) is between 5 µm and 10 µm. The reconstructed, complex valued probe is depicted in Fig. 5.12 (left and center). It results in an intensity distribution at the detector which is shown in the right part of Fig. 5.12 and compares well with the measured flat field (Fig. 5.10(b, left image)).
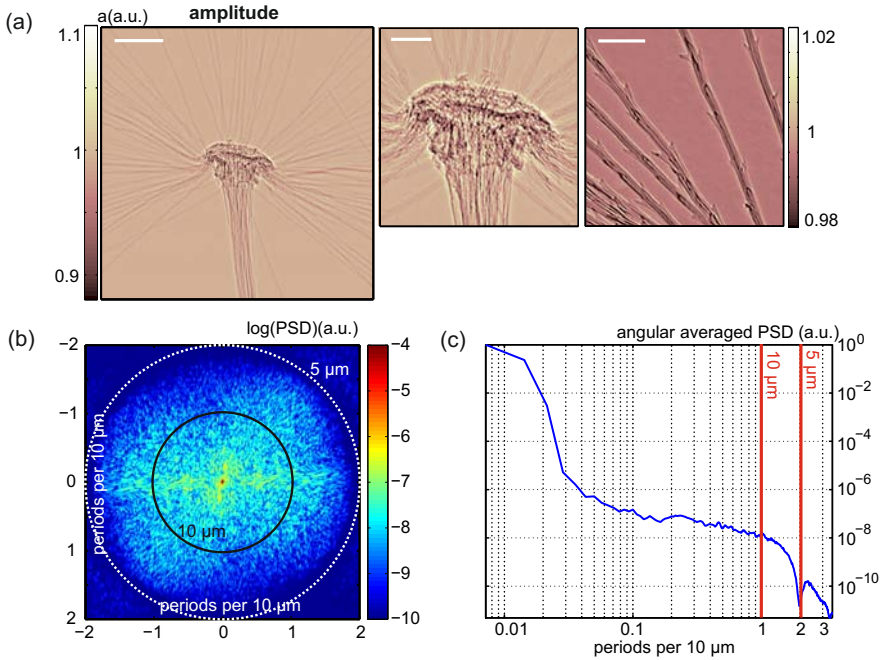
Figure 5.11: (a) Reconstructed amplitude of the object with two regions of interest. Scale bars denote 250 µm (overview) and 100 µm (regions of interest). (b) Power spectral density of (a). (c) Angular averaged power spectral density of (a).
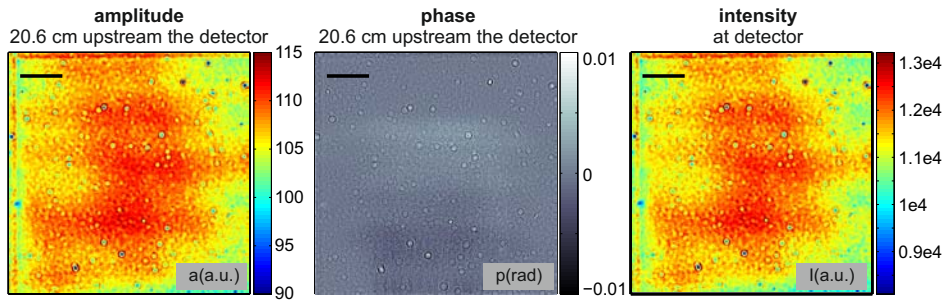
Figure 5.12: Reconstructed probe. Left: Amplitude at 20.6 cm upstream with respect to the detector. Center: Phase at 20.6 cm upstream with respect to the detector. Right: Reconstructed flat field. Scale bars denote 250 µm.

5.3 Waveguide based cone beam X-ray near-field ptycho-graphy

Next, two series of experiments are presented. Both were performed at the DESY/-P10/GINIX-setup [90–92]. The main difference was the choice of samples. In the first experiment, a confined sample was studied. In the second experiment, an extended sample covering the whole field of view was used. The probe was shaped by an X-ray waveguide which results in a smooth and clean illumination. In both series of experiments the main contribution of lateral diversity was due to (mis-) alignment of the sample in the beam path: Lateral shifts in the order of micrometers appear quite naturally, because defocus shifts are difficult to be performed exactly parallel with the optical axis.

5.3.1 Imaging confined objects in a deeply holographic regime

The main components of the X-ray holographic microscopy setup GINIX at P10 are depicted in Fig. 5.13[12].

The undulator beam with photon energy $E = 7.9$ keV is focused by Kirkpatrick-Baez (KB) mirrors onto the entrance side of an X-ray waveguide (lithographically defined air channel in silicon, 91 nm $\times$ 70 nm $\times$ 1 mm) [93]. The result is a clean, mode filtered and highly divergent X-ray beam as depicted in Fig. 5.13(b), (c) and (d) (left column). During propagation, the probe diverges and forms a cone-like illumination. Hence, it is possible to collect images of the object with different magnification and resolution, covering an extended field of view.

The sample was a nano-fabricated test structure containing patterns of different sizes and layouts produced by focused ion beam on a 200 nm gold substrate and attached to a Si_3N_4 membrane. Three different test patterns were imaged: (1) a structure in the following called 'Nano-World' (electron micrograph depicted in the right part of Fig. 5.14(a)), (2) a structure in the following called 'IRP-Logo' (electron micrograph depicted in the right part of Fig. 5.14(b)) and (3) a structure in the following called 'Micro-World' (electron micrograph depicted in the right part of Fig. 5.14(c)). Structure (1) was positioned at a distance of 5.1 mm behind the waveguide exit and shifted by [0.38, 0.89, 1.40] mm with respect to its initial position towards the detector (sCMOS, Photonic Science, UK; 6.54 µm pixel size) which was placed at 5.12 m downstream the waveguide. In each defocus plane four exposures at slightly different lateral positions were taken (three of them with an exposure time of 1 s, one of them (the sample was centered in the beam) with an exposure time of 3 s). Details can be accessed in Table 5.3.

With the sample positioned 5.1 mm behind the waveguide, the effective pixel size

---

[12]Scale bars in figures showing the results of cone beam microscopy are calculated with respect to the smallest effective pixel size (in Fig. 5.13 this pixel size would be 6.5 nm for the images shown in (b) and 19.1 nm for the images shown in (c) and (d)). Effective pixel sizes are listed for each experiment in the corresponding tables.
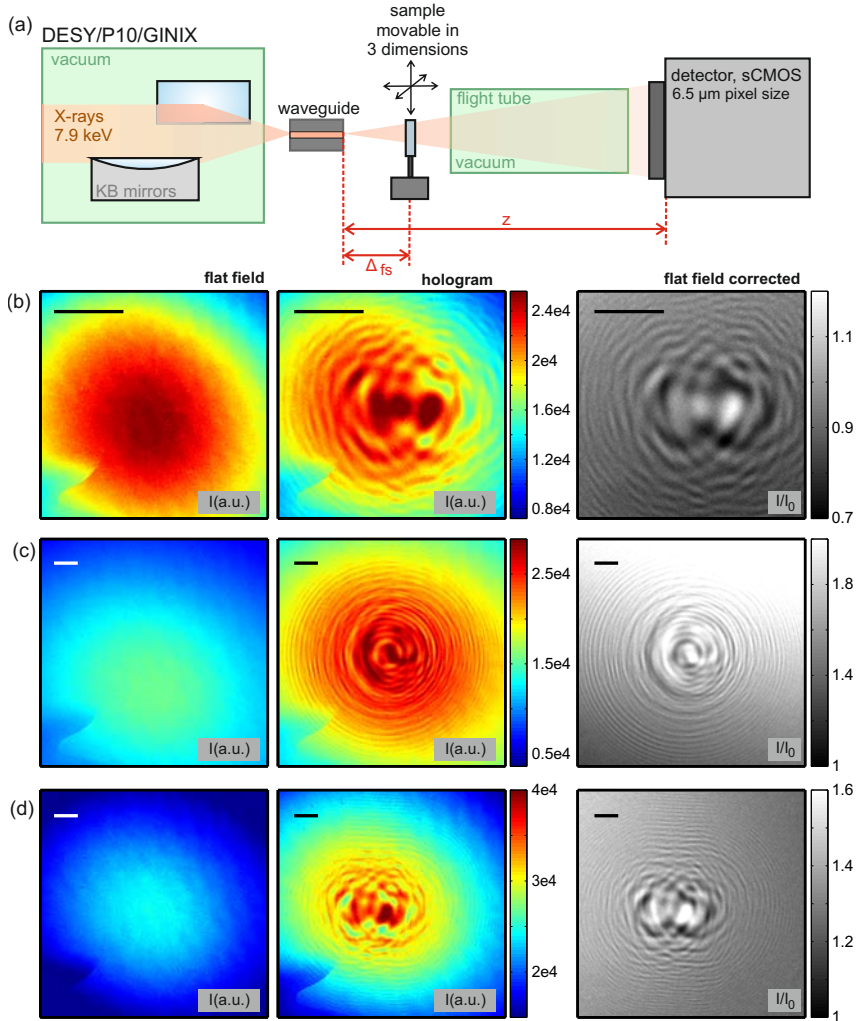
Figure 5.13: (a) Setup of DESY/P10/GINIX for waveguide based X-ray holographic microscopy. (b) Test structure 'Nano-World'. Left: Flat field. Center: Exemplary hologram. Right: Hologram divided by flat field. (c) Test structure 'IRP-Logo', same layout as in (b). (d) Test structure 'Micro-World', same layout as in (b) and (c). Scale bars denote 2 μm.

in the sample plane is 6.5 nm. However, this comes at a price: The near-field propagator is considerably under sampled (see section 3.1):

$$\frac{\lambda \cdot z_{\text{eff}}}{N \cdot d_{\text{eff}}^2} = \frac{0.16\text{nm} \cdot 5.1\text{mm}}{1080 \cdot (6.5\text{nm})^2} = 17.9 \ , \tag{5.360}$$

with wavelength λ, effective propagation distance z_{eff} for the largest sample to detector spacing, number of pixels N and effective pixel size d_{eff} in one dimension. To avoid aliasing, the field of view of the propagated field needs to be enlarged by a factor of 17.9. However, this requires a long computation time and the storage of a large amount of data. For this reason, the field of view was still kept smaller than necessary for Nyquist sampling.

In the next two experiments, the object was placed at 15.13 mm behind the waveguide exit and translated towards the detector by distances of [0.1, 0.5, 1.0, 1.5, 2.0, 2.5] mm with respect to the initial position. Further details can be found in Table 5.3. Compared to the first experiment, the distance between waveguide and detector was left unchanged. Two images per plane were recorded. An example of a flat field, a hologram and the flat field corrected hologram can be seen in Fig. 5.13(b) for the structure 'Nano-World', in Fig. 5.13 (c) for the structure 'IRP-Logo' and in Fig. 5.13(d) for the structure 'Micro-World'. In (c) and (d) the intensity of the flat field is considerably lower than the intensity of the hologram. The reason is that during the experiment the waveguide slightly drifted in the beam. Nevertheless, since the flat fields are not used for phase retrieval, this does not impede data analysis.

Reconstructions were carried out using the same stetting of parameters for all three cases. Near-field ptychography was initialized with a back-propagated version of the measured intensities (for the probe) and a homogeneous distribution of uniform amplitude and phase corresponding to 200 nm gold illuminated by X-rays of 7.9 keV. In total, phase retrieval was run for 20 iterations. During the first eight iterations, special weight was attributed to the update of the object while keeping the probe rather stable. The update of the object in the plane of highest resolution was enhanced with respect to the updates resulting from defocus positions with lower resolution. Twin image artifacts were suppressed during the first seven iterations (see equation 4.322). Following equation 4.323, the phases of the object were projected to negative values. Automatic focusing (equation 4.325) was applied during reconstruction.

Fig. 5.14 shows the reconstructed objects for all three experiments. The amplitudes of the reconstructed transmission functions are depicted on the left hand side of Fig. 5.14. As expected, slight aliasing artifacts can be seen in the amplitude of 'Nano-World'. A transmission larger than one can be observed in Fig. 5.14(c). For this reason, the reconstructed amplitude is more qualitative than quantitative. The reconstructed phase shifts match the expectations better than the reconstructed amplitudes. A more detailed analysis of the reconstructed refractive index is given in section 5.3.1.

Table 5.3: Parameters for X-ray waveguide holography at DESY/P10/GINIX.

object	'Nano-World', Si_3N_4+Ti+Au
	(200 nm + 3 nm + 205 nm)
geometry	cone beam
wavelength [nm]	0.16
eff. pixel size [nm]	6.5, 7.0, 7.7, 8.3
number of pixels	1920×1080
accumulation time [s]	1 and 3
eff. F_{10}	$5.3e-3$, $5.7e-3$, $6.2e-3$, $6.8e-3$
waveguide-sample distances [mm]	5.10, 5.50, 6.00, 6.50
waveguide-detector distance [m]	5.12
maximum lateral displacement [µm]	2.6
# holograms	16 (4 defocus planes, 4 exposures each)
object	'IRP-Logo', Si_3N_4+Ti+Au
	(200 nm + 3 nm + 205 nm)
geometry	cone beam
wavelength [nm]	0.16
eff. pixel size [nm]	19.1, 19.3, 19.8, 20.4, 21.0, 21.7, 22.3
number of pixels	1920×1080
accumulation time [s]	3
eff. F_{10}	$15.6e-3$, $15.7e-3$, $16.1e-3$,
	$16.6e-3$, $17.1e-3$,
	$17.7e-3$, $18.2e-3$
waveguide-sample distances [mm]	15.13, 15.21, 15.63,
	16.12, 16.62, 17.13, 17.62
waveguide-detector distance [m]	5.13
maximum lateral displacement [µm]	1.4
# holograms	14 (7 defocus planes, 2 exposures each)
object	'Micro-World', Si_3N_4+Ti+Au
	(200 nm + 3 nm + 205 nm)
geometry	cone beam
wavelength [nm]	0.16
eff. pixel size [nm]	19.1, 19.3, 19.8, 20.4, 21.1, 21.7, 22.3
number of pixels	1920×1080
accumulation time [s]	3
eff. F_{10}	$15.6e-3$, $15.7e-3$, $16.1e-3$,
	$16.6e-3$, $17.2e-3$,
	$17.7e-3$, $18.2e-3$
waveguide-sample distances [mm]	15.13, 15.23, 15.63,
	16.12, 16.63, 17.13, 17.62
waveguide-detector distance [m]	5.13
maximum lateral displacement [µm]	1.4
# holograms	14 (7 defocus planes, 2 exposures each)
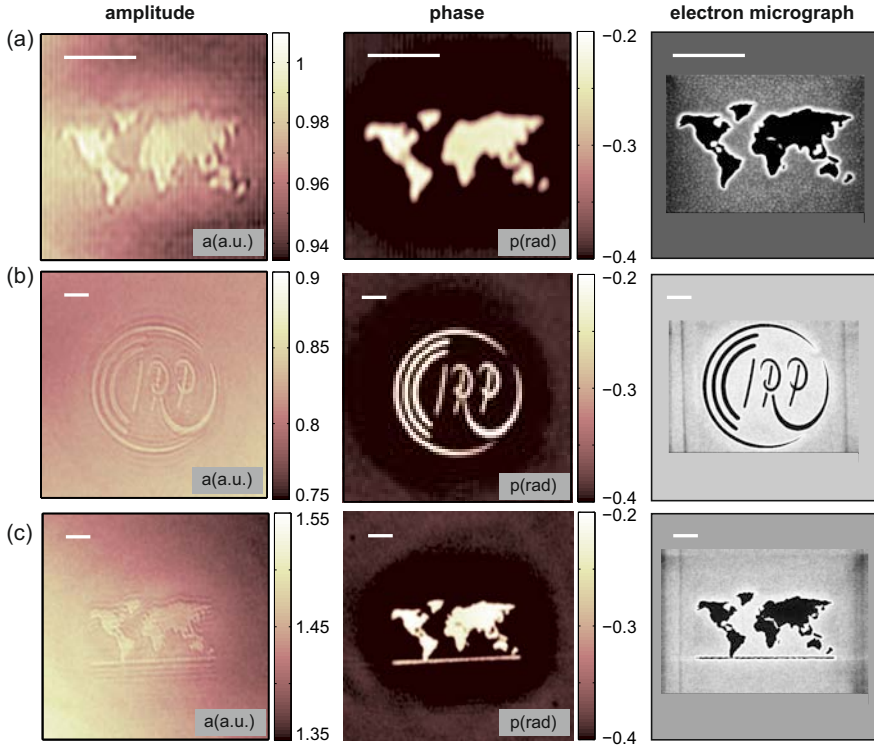
Figure 5.14: Reconstructed objects and scanning transmission electron micrographs of the test structures. (a) Test structure 'Nano-World', (b) test structure 'IRP-Logo' and (c) test structure 'Micro-World'. Left: Reconstructed amplitude. Center: Reconstructed phase. Right: Electron micrograph. Scale bars denote 1 µm.

One of the reconstructed probes is depicted in Fig. 5.15. Amplitude (a) and phase (b) of the reconstructed probe are shown at 15 mm behind the waveguide exit. The reconstructed intensities at the detector are shown in (c). Note that spoiled regions of the camera (see lower left part of the flat fields) are not addressed to the object but are merged into the probe. As expected, the reconstructed probes are essentially the same for all three experiments. Due to the unstable waveguide and hence also unstable illumination of the object during the scan, a quantitative comparison with the flat field is difficult.

The full period resolution of the reconstructed objects was determined (1) by fitting step functions to selected regions (left part of Fig. 5.16), (2) by calculating the power spectral density (central part of Fig. 5.16) and (3) by angular averaging of the power spectral density (right part of Fig. 5.16). Following this analysis, the resolution was between 140 nm and 100 nm (full period) with lowest effective pixel sizes between 6.5 nm ('Nano-World') and 19.1 nm ('Micro-World' and 'IRP-Logo'). Despite the low effective pixel size, the resolution of 'Nano-World' is not better than the resolution of the other two test patterns. The reasons might be
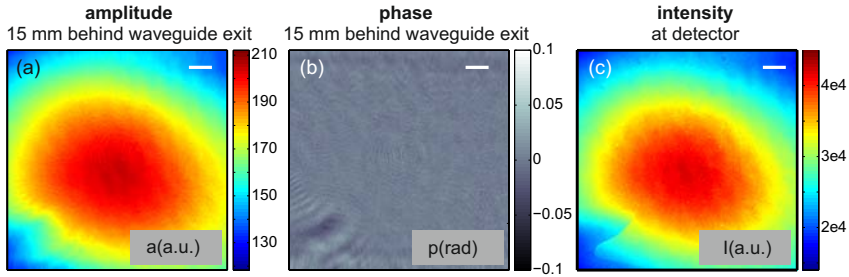
Figure 5.15: Reconstructed probe. (a) Amplitude at 15 mm behind the waveguide. (b) Phase at 15 mm behind the waveguide. (c) Reconstructed flat field. Scale bars denote 2 μm (with respect to the smallest effective pixel size).
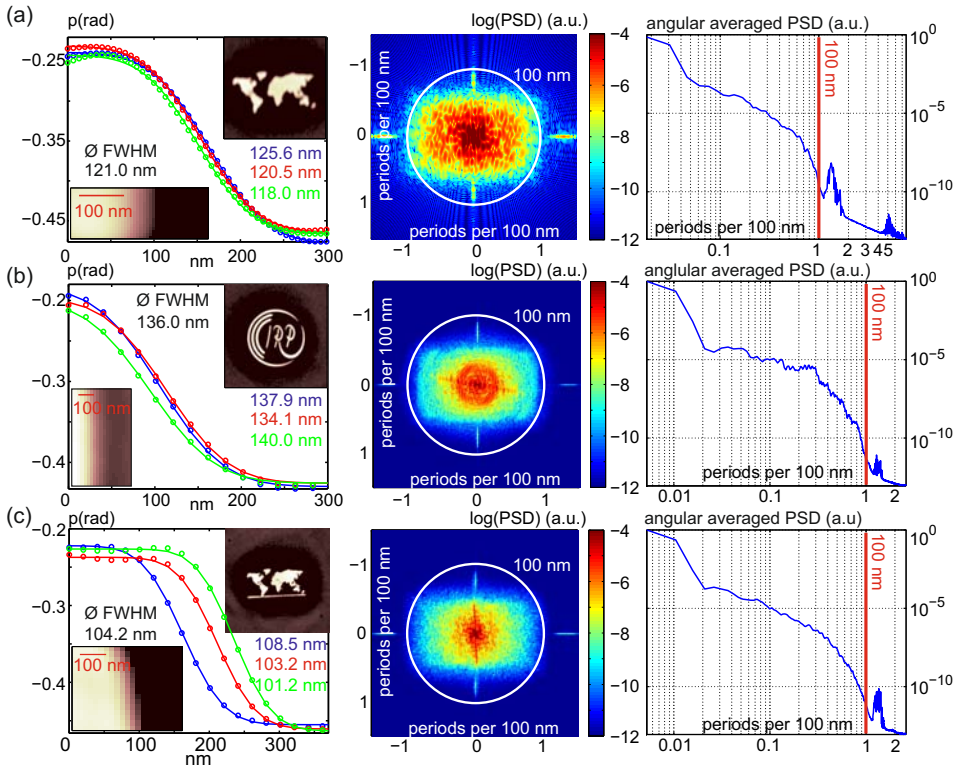
Figure 5.16: Resolution analysis for the reconstructed phase of (a) 'Nano-World', (b) 'IRP-Logo' and (c) 'Micro-World'. Left: Adaption of error functions to selected edges. Center: Power spectral density. Right: Angular averaged power spectral density.

instabilities in the setup due to vibrations at the sample during and right after scanning the test structure.

Analysis of the refractive index

The test structure was made of 205 nm Au on a Si_3N_4 membrane of 200 nm thickness with an interlayer of 3 nm Ti. The reconstructed difference in the refractive index for structured and unstructured regions shall be determined. For this analysis, the small layer of titanium is neglected. For an energy of 7.9 keV the refractive index of Au is [94]

$$n_{Au} = 1 - \delta + i\,\beta = 1 - 4.9 \cdot 10^{-5} + i\,5.2 \cdot 10^{-6},$$

and the refractive index of Si_3N_4 is [94]

$$n_{Si_3N_4} = 1 - \delta + i\,\beta = 1 - 1.2e \cdot 10^{-5} + i\,1.8 \cdot 10^{-7}.$$

With a layer of 205 ± 20 nm of Au[13], this leads to a transmission ranging between 95.4% and 96.2%. The expected phase shift is between -0.44 rad and -0.36 rad. For a layer of 200 ± 20 nm Si_3N_4 we expect a transmission of about 99.8% and a phase shift between -0.10 rad and -0.08 rad. Since only relative phase shifts can be measured by phase retrieval techniques, the relative phase shift between regions of Au and regions of Si_3N_4 should be between -0.36 rad and -0.26 rad. The difference in transmission should be between 3.62% and 4.44%.

To determine the experimental phase shift and transmission of the reconstruction shown in Fig. 5.17(a) and (b), line profiles were extracted (depicted in Fig. 5.17(d-e)). They indicate that the differences in transmission and phase (about 2% difference in transmission and 0.22 rad in phase shift) are slightly lower than expected. To nail down more precise values, a threshold was applied to the reconstructed object separating regions of Au (gray areas in Fig. 5.17(c)) and Si_3N_4 (white areas in Fig. 5.17(c)) while neglecting the edges (black areas in Fig. 5.17(c)). By averaging over all pixels of regions consisting of Au and over all pixels of regions consisting of Si_3N_4, a mean phase shift between structured and unstructured regions of -0.21 rad $\pm\, 0.02$ rad and a mean difference in transmission of $1.4\% \pm 1.2\%$ was found (see histograms in Fig. 5.17 (f-i)). Whereas the validity of the measured relative transmission is still questionable, the relative phase shift between Au and Si_3N_4 is quite realistic.

---

[13]The thickness was determined with the help of a profilometer (*Veeco Dektak 6*); an error of ± 20 nm is justified.
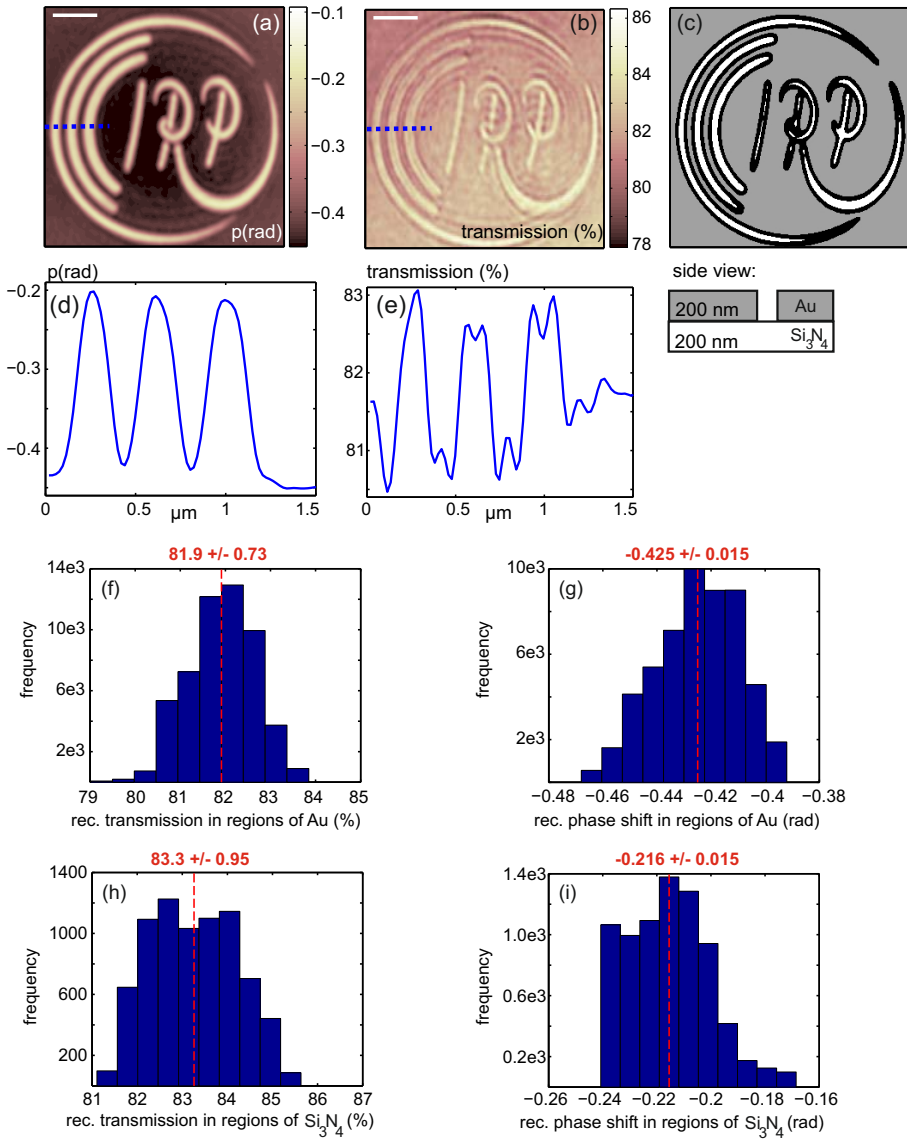
Figure 5.17: Determination of (a) the reconstructed phase shift (scale bar denotes 1 µm) and (b) the reconstructed transmission (scale bar denotes 1 µm). (c) Mask indicating the layout of the test structure as well as regions of Au and Si_3N_4. (d) Line profile through (a). (e) Line profile through (b). (f) Histogram and mean value with standard deviation of the reconstructed transmission in regions of Au. (g) Histogram and mean value with standard deviation of the reconstructed phase shift in regions of Au. (h-i) Respective histograms for the regions of Si_3N_4.

5.3.2 Imaging extended objects in a deeply holographic regime

The experiments shown next demonstrate reconstructions of extended objects in a clean, as well as in a disturbed probe. In contrast to methods, where only overlapping regions are reconstructed, the near-field ptychographic algorithm takes all of the measured data into account: Instead of interpolating the holograms to the same resolution before phase retrieval, *the reconstructed object* is interpolated during phase retrieval to the highest possible resolution determined by the experiment. The consequences are reconstructions of high resolution *and* large field of view.
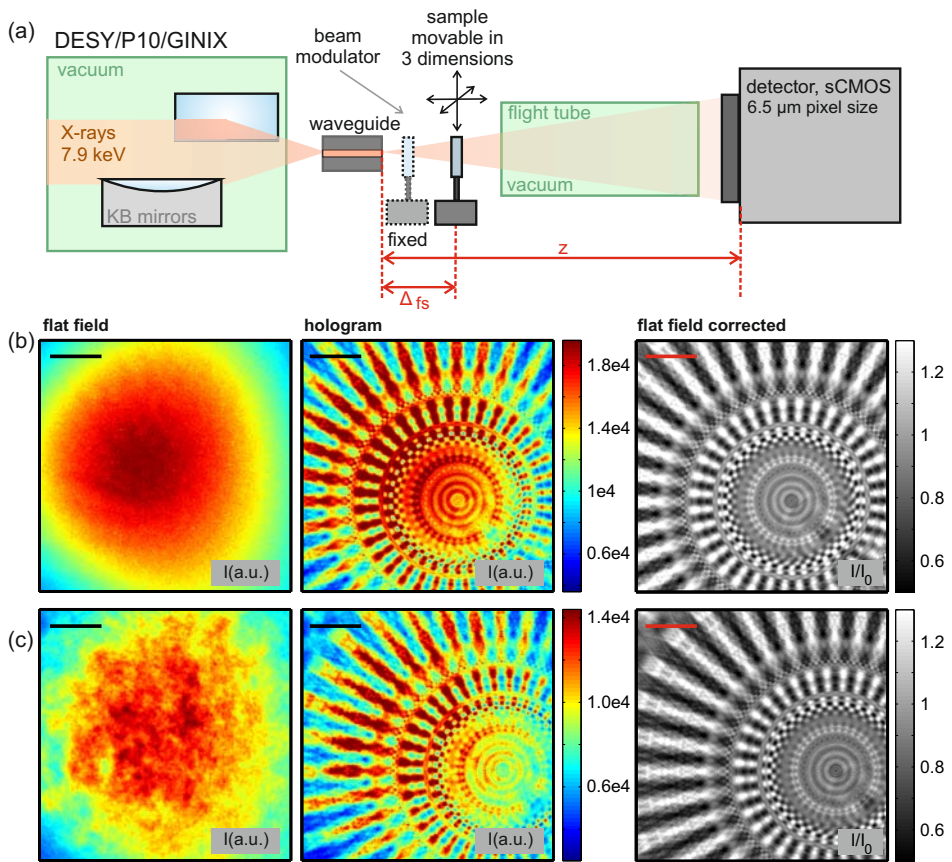
Figure 5.18: (a) Setup of DESY/P10/GINIX. (b) Test structure 'Siemens-Star' in clean probe. Left: Flat field. Center: Exemplary hologram. Right: Hologram divided by flat field. (c) Test structure 'Siemens-Star' in modulated probe, same layout as in (b). Scale bars denote 5 μm.

The experiment was carried out at DESY/P10/GINIX. Hence, the setup of the

X-ray microscope is basically the same as described above – with one modification: a fixed object for intentional beam modulation can be inserted between waveguide exit and sample (see Fig. 5.18). Hard X-rays ($\lambda = 1.6 \cdot 10^{-10}$ m) were focused by Kirkpatrick-Beatz mirrors (KB-mirrors) and passed through an X-ray waveguide (91 nm $\times$ 70 nm) [93]. Holograms of a 'Siemens-Star' (ANT/XRESO-50HC, NTT-AT, Japan, 200 nm Ta) initially placed at $\Delta_{\mathrm{fs}} = 19.9$ mm behind the waveguide were recorded at $z = 5.34$ m with respect to the waveguide exit by an sCMOS detector (Photonic Science, UK; 6.54 µm pixel size). The 'Siemens-Star' was shifted to four additional defocus positions at $\Delta_{\mathrm{fs}} = [20.8, 25.8, 30.8, 35.8]$ mm behind the waveguide exit. For better statistics, at each of these five defocus positions 10 holograms of the sample were collected (each with an exposure time of 8 s). The object was not scanned laterally such that the main contribution of lateral diversity was included in the defocus shifts.

Two experiments were performed using this configuration. First, the probe was not intentionally modulated. A measured flat field, one exemplary hologram and the flat field corrected hologram are shown in Fig. 5.18(b). Second, the probe was disturbed by 100 nm phase shifting latex spheres spread on a $\mathrm{Si_3N_4}$ frame. A flat field, an exemplary hologram and the flat field corrected hologram are depicted in Fig. 5.18(c). Further details concerning the experimental setup as well as the effective propagation distances and pixel sizes are listed in Table 5.4.

Table 5.4: Experimental parameters P10.

object	'Siemens-Star', Ta (200 nm)
beam modulation	none/ 100 nm latex spheres
geometry	cone beam
wavelength [nm]	0.16
eff. pixel size [nm]	24.3, 25.5, 31.7, 37.8, 43.9
number of pixels	1920×1080
accumulation time [s]	10×8
eff. F_{10}	$19.0\mathrm{e}-3$, $20.0\mathrm{e}-3$,
	$24.9\mathrm{e}-3$, $29.7\mathrm{e}-3$, $34.5\mathrm{e}-3$
waveguide-sample distances [mm]	19.8, 20.8, 25.8, 30.8, 35.8
waveguide-detector distance [m]	5.33
maximum lateral displacement [µm]	2.7/ 3.4
# holograms	50 (5 defocus planes,
	10 exposures each)

Phase retrieval was initialized by a back-propagated flat field for the probe and a homogeneous amplitude and phase distribution of Ta at 7.9 keV for the object. The algorithm was performed for 20 iterations (each experiment). Twin images were suppressed during the first iterations (equation 4.322); the phases of the object's transmission function were projected to negative values (equation 4.323) and automatic focusing (equation 4.325) was included.

The reconstructed probes are depicted in Fig. 5.19. The clean probe shows weak

artifacts resulting from the object and visible in the reconstructed phases which are between -0.05 rad and 0.05 rad. These artifacts could have probably been avoided with more lateral diversity in the holograms. The reconstructed intensities at $z = 5.34$ m compare well with the measured flat field. Same is true for the disturbed probe.
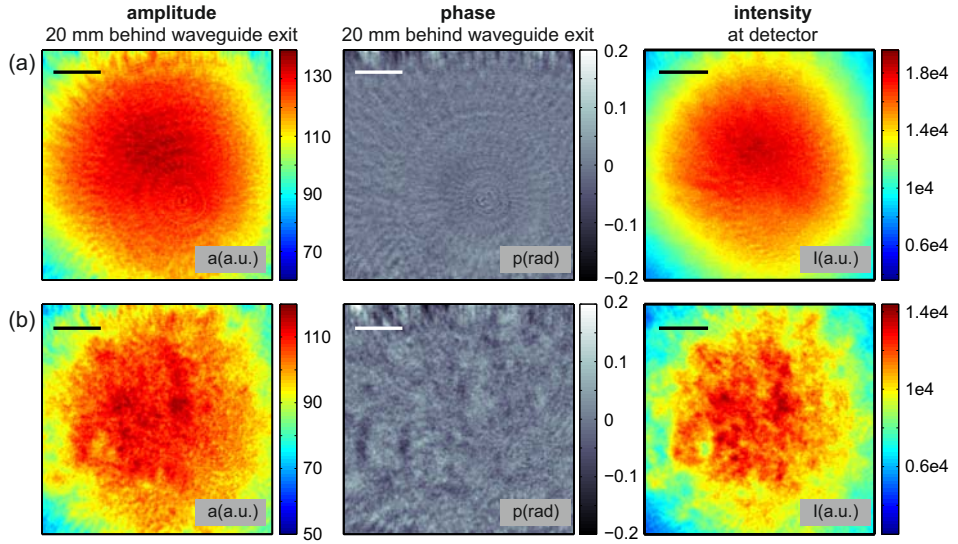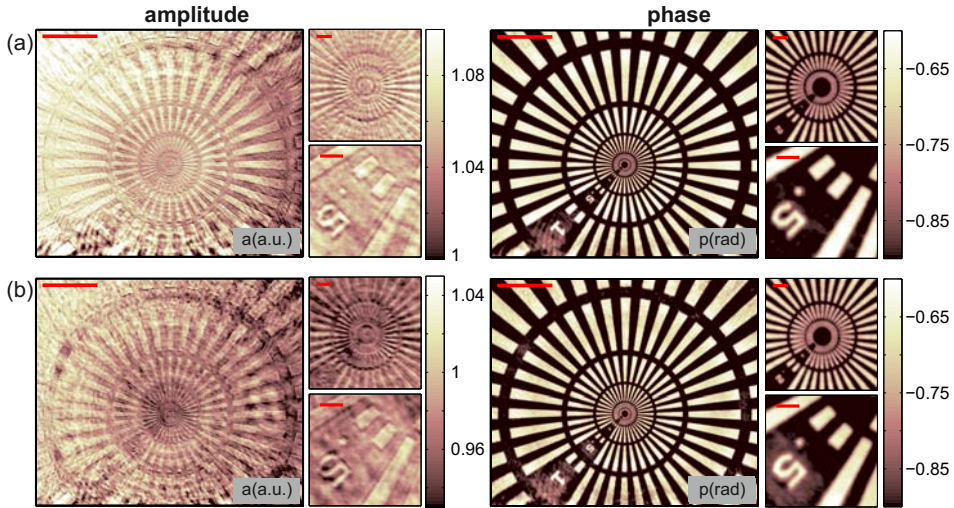

Figure 5.19: Reconstructed probe for (a) non-modulated waveguide beam and (b) waveguide beam modulated by phase shifting latex spheres. Left: Reconstructed amplitude at largest sample to detector distance. Center: Reconstructed phase at largest sample to detector distance. Right: Reconstructed flat field. Scale bars denote 5 µm

The reconstructed objects for both experiments are summarized in Fig. 5.20. The main difference in quality of the reconstructions shown in Fig. 5.20(a) and (b) can be seen in the amplitudes (the left part of Fig. 5.20). For a clean illumination compared to a modulated illumination, the reconstructed amplitudes contain less noise and slightly sharper edges. These differences are not visible for the reconstructed and much more dominant phases. In fact, the reconstructed phases are clean and artifact free.

Fig. 5.21 details the resolution of the reconstructed phases of the objects. The resolution was analyzed by edge fits (left part of Fig. 5.21), the power spectral density (central part of Fig. 5.21) as well as the angular averaged power spectral density (right part of Fig. 5.21). With an effective pixel size of 25 nm, features down to 100 nm lines and spaces (full period resolution) could be resolved in both experiments.

Figure 5.20: Reconstructed object for (a) non-modulated waveguide beam and (b) waveguide beam modulated by phase shifting latex spheres. Left: Reconstructed amplitude. Right: Reconstructed phase. Scale bars denote 10 μm in the overview part and 1 μm in the zoomed regions.
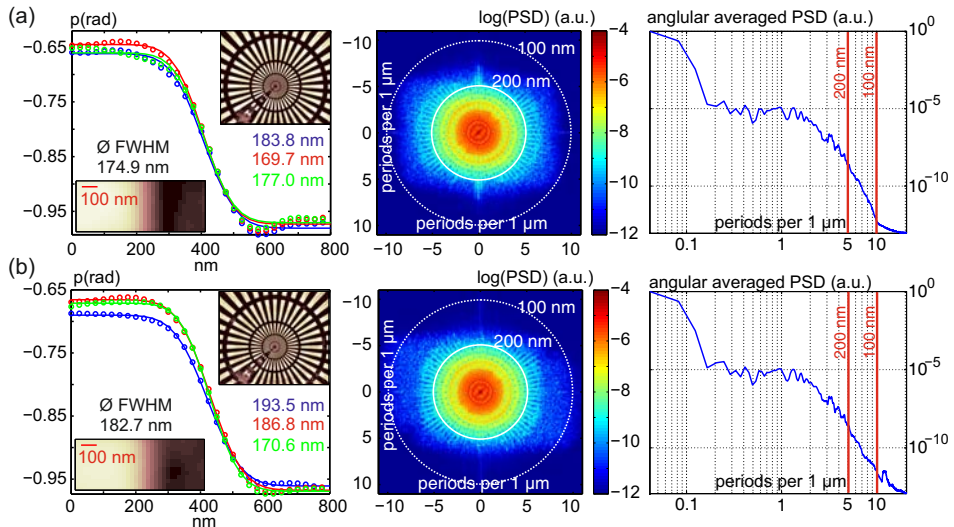
Figure 5.21: Resolution analysis for the reconstructed phase of (a) 'Siemens-Star' imaged by a clean probe and (b) 'Siemens-Star' imaged by a modulated probe. Left: Adaption of error functions to selected edges. Center: Power spectral density. Right: Angular averaged power spectral density.

Comparison between CTF based reconstruction and near-field ptychography
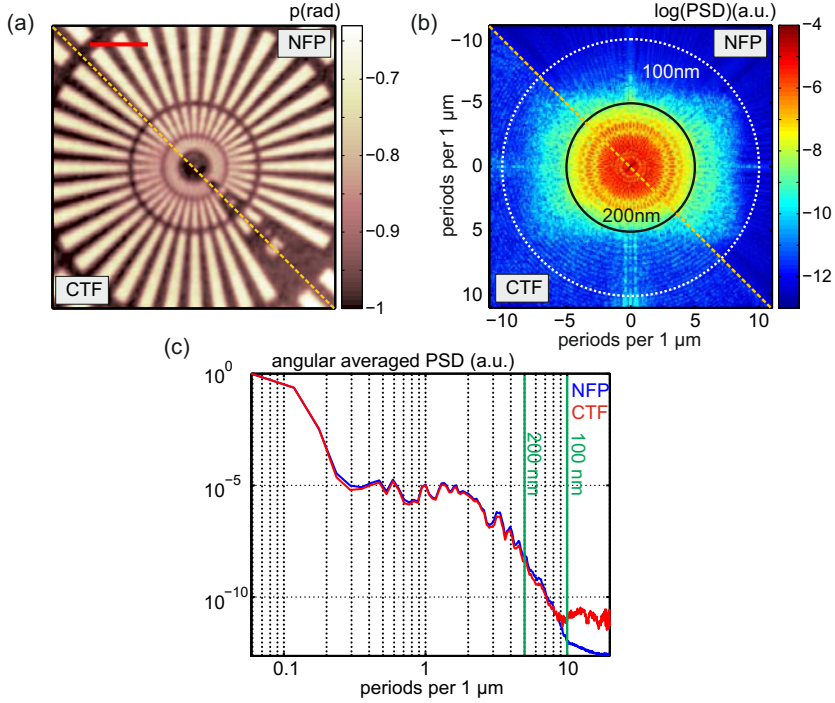
Figure 5.22: Comparison between CTF based reconstruction and near-field ptychography (NFP). (a) Composite image of CTF based reconstruction (lower left part) and near-field ptychographic reconstruction (upper right part). (b) Composite image of power spectral density of (a). (c) Angular averaged power spectral density of the CTF based reconstruction (red) and the near-field ptychographic reconstruction (blue).

Near-field ptychography is an iterative technique that does not need any linearization of the propagator or the object's transmission function. The flat field of the illumination does not need to be known as near-field ptychography does not rely on the faulty flat field correction. However, in the case of a clean probe, the flat field correction is approximately correct and high quality reconstructions of the object (without the probe) can also be obtained by quicker single-step techniques that are for example based on the contrast transfer function (CTF) [11, 95–97].
A flat field corrected and Fourier transformed hologram $\mathcal{F}[I_\Delta](q_x, q_y)$ which was recorded at some distance Δ with respect to the sample can be approximately described by [95]

$$\mathcal{F}[I_\Delta](q_x, q_y) \approx 2\pi \delta_D(q_x, q_y) + 2\mathcal{F}[\phi](q_x, q_y)\sin(\mathcal{X}) \,, \qquad (5.361)$$

where δ_D is the Dirac delta, ϕ is the phase shift of the transmission function of the object and

$$\mathcal{X} = \frac{\Delta\lambda}{4\pi}\left(q_x^2 + q_y^2\right) \ . \tag{5.362}$$

Since the CTF function contains zero crossings which depend on the propagation distance Δ, several holograms with different Δ are used for phase retrieval. The phase is determined by minimization of a cost function: The summed distance between the approximated and measured holograms is analytically minimized by setting the derivative with respect to ϕ to zero and after that by directly calculating the phase shift ϕ [95].

Here, the same data set (in particular the same number of holograms) was fed in the CTF [11,95,97] and in the near-field ptychography algorithm. The reconstructions are analyzed regarding their resolution. Fig. 5.22 depicts the reconstructed object (a). The image shown there is a composition of the reconstructions achieved by CTF (lower left part) and by near-field ptychography (upper right part). The reconstructed phase shift is comparable. However the contrast in the central region of the reconstruction obtained by near-field ptychography is slightly enhanced. At an effective pixel size of 24 nm, both reconstructions reveal the 100 nm lines and spaces. The 50 nm lines and spaces are not resolved anymore. Both reconstructions look very clean and artifact free. A closer inspection is necessary to mark the differences between the two techniques.
To this end, the power spectral density (Fig. 5.22(b)) was calculated. Although the power spectral density is essentially the same for frequencies corresponding to up to 150 nm lines and spaces, a larger amount of noise is found in the high frequency domain for the CTF reconstruction than for the near-field ptychographic reconstruction.
The same effect can be seen in the angular averaged power spectral density depicted in Fig. 5.22(c). At a resolution of 120 nm lines and spaces, noise starts to interfere with the red curve (referring to the CTF reconstruction), whereas the blue curve (referring to the near-field ptychography reconstruction) is deteriorated by noise starting at a resolution of 100 nm lines and spaces.
Hence, even if this particular data set does not allow to prove a significant increase in resolution when turning from CTF-based reconstruction techniques to near-field ptychography, it provides a strong indication that near-field ptychography offers the potential for high resolution imaging at spatial frequencies where techniques requiring a flat field correction suffer from noise.

The results of the last sections can be summarized by three main points. (1) Near-field ptychography can be applied to image confined objects with unknown support; it successfully eliminates the twin image without the need of a support constraint or restriction to phase objects.
(2) Phase shifting properties of a material can be reliably determined.
(3) Compared to CTF techniques, large objects covering the whole field of view can be reconstructed with lower noise in the high frequency domain.

5.4 KB-cone beam X-ray near-field ptychography

To obtain a small focal spot and hence a highly divergent beam, a pair of curved Kirkpatrick-Baez (KB) mirrors is commonly used for X-rays. Focusing X-rays however comes at a price: Strong aberrations caused by surface defects of the mirrors on the nanometer scale cause a highly structured probe [29]. An example can be seen in the left part of Fig. 5.23(b). A flat field measured at the newly build ID16A beamline at ESRF (the setup of this beamline is sketched in Fig. 5.23(a)) is depicted. The horizontal and vertical stripes are the mentioned figure errors. Holograms recorded with such an illumination are basically indistinguishable from the flat field – at least, they are not distinguishable by eye and certainly not at first glance (see central part of Fig. 5.23(b)). The standard way to correct raw data, i.e. the division of the hologram by the flat field, reveals the hologram caused by the transmission function of the sample in the beam path (right part of Fig. 5.23(b)). As already detailed, the flat field correction is only applicable in case of a homogeneously or mildly disturbed probe. It breaks down when aberrations become too dominant or are of higher, spatial frequency than the features caused by the object of interest [30].

Here, near-field ptychography in a cone beam that was focused by KB-mirrors is shown.
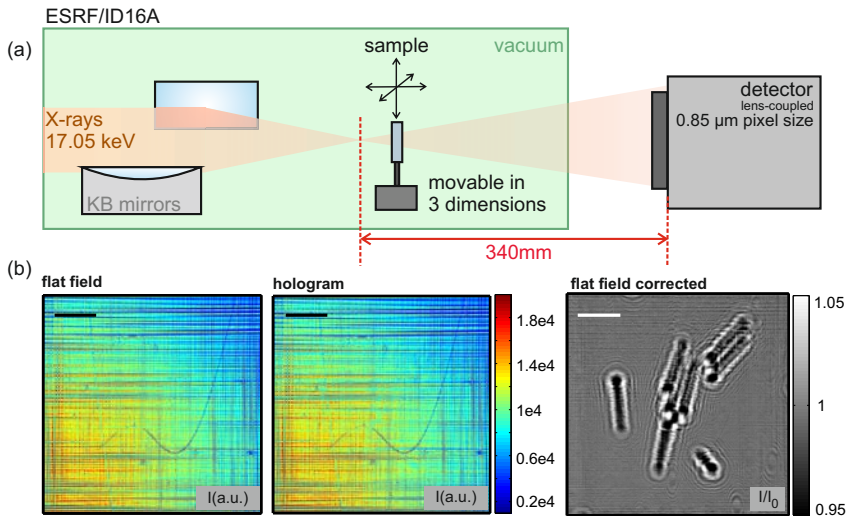
Figure 5.23: (a) Setup at ESRF/ID16A dedicated to high resolution imaging by nano-focused hard X-ray beams. (b) Left: Flat field. Center: Hologram of nanowires. Right: Flat field corrected hologram. Scale bars denote 2 μm.

The experiment was performed at the ID16A beamline of ESRF which is dedicated to high resolution imaging by nano-focused hard X-ray beams (spot sizes down to 40 nm in two dimensions). As described in Fig. 5.23(a) X-rays of 17.05 keV

Table 5.5: Experimental parameters ID16A.

object	'nanowires', InP substrate, insulating SiO_2 layer, ITO cover
geometry	cone beam
wavelength [nm]	0.073
pixel size of the detector (2x binned) [µm]	16.9
eff. pixel size (2x binned) [nm]	10.0, 10.5, 12.2, 15.8
number of pixels (2x binned)	1024×1024
accumulation time [s]	2
eff. F_{10}	$6.90e-2$, $7.19e-2$, $8.36e-2$, $10.88e-2$
distances focus-sample [mm]	2.02, 2.11, 2.45, 3.18
distance focus-detector [cm]	34.04
lateral step size [µm]	2.5
# holograms	64 (4 defocus planes, 16 lat. positions)

are confined by KB-mirrors. Since not only the mirrors, but also the sample is kept in vacuum, the object of interest can be moved right behind the focal spot. The given configuration enables imaging at an effective pixel size down to 5 nm. For this experiment a sample consisting of nanowires (InP substrate, insulating SiO_2 layer and ITO cover [98]) spread onto a Si_3N_4 membrane was chosen. The sample was inserted in the beam at a defocus distance of 2.02 mm with respect to the focus and 34.04 cm apart from a high-resolution imaging camera. The detector was lens-coupled to a FReLoN F_K4320 (0.85 µm pixel size). The sample was translated towards the detector with defocus distances of [2.02, 2.11, 2.45, 3.18] mm. At each defocus position 16 holograms were recorded with the object shifted inside the plane perpendicular to the optical axis by a lateral step size of 2.5 µm. Further details can be found in Table 5.5.

Near-field ptychographic phase retrieval was performed for ten iterations, including the projection of the object onto a pure phase object (equation 4.321) with negative phase shift (equation 4.323). Automatic focusing (equation 4.325) was applied during reconstruction. Phase retrieval was initialized by a uniform amplitude distribution with zero phases for object and probe. The reconstructed object is presented in Fig. 5.24. Since the recorded holograms were binned by a factor of two before reconstruction, the smallest effective pixel size is 10 nm.

Single nanowires are clearly distinguishable. Even internal density variations become visible. The reconstructed object is free of artifacts resulting from the probe. To estimate the resolution, step functions were fit to selected regions (shown in Fig. 5.24(b)) indicating an average, full period resolution of 89.7 nm. The power spectral density (Fig. 5.24(c)) reveals an even higher frequency content and a full
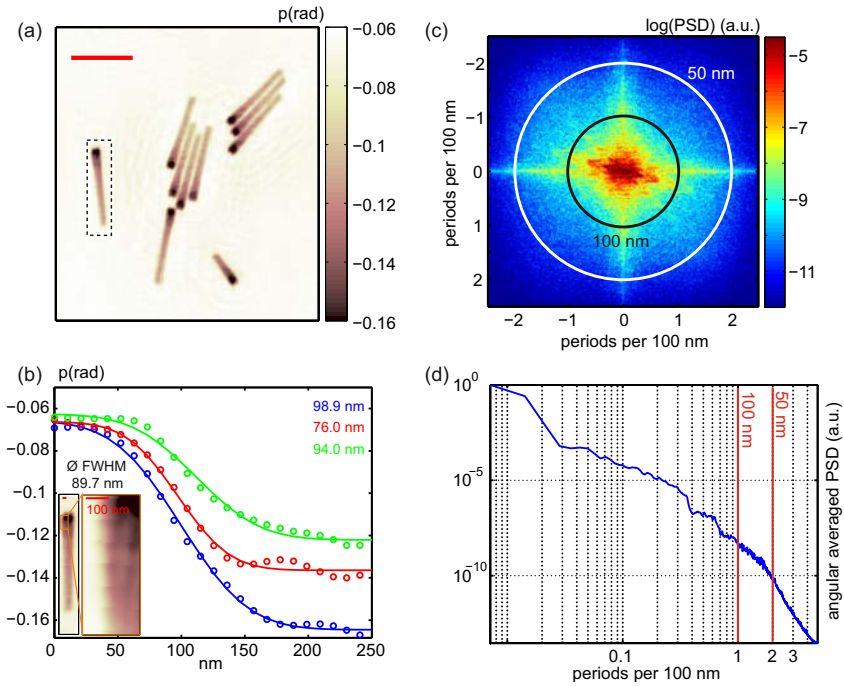
Figure 5.24: Reconstructed nanowires. (a) Reconstructed phase. Scale bar denotes 2 µm. (b) Step functions were fitted to selected regions to determine the resolution. Scale bars denote 0.1 µm. (c) Power spectral density of (a). (d) Radially averaged power spectral density of (a).

period resolution down to 50 nm. In addition, the angular averaged power spectral density was calculated (Fig. 5.24(d)) and again a resolution was found between 100 nm and 50 nm.

The reconstructed probe is depicted in Fig. 5.25 showing amplitude and phase at 2.02 mm behind the focal plane (left and central part of Fig. 5.25) as well as the reconstructed flat field (right part) which compares well with the measured flat field presented in Fig. 5.23(b, right part).

Clear separation of object and probe demonstrated by inspection of holograms

The distinct separation of object and probe can be illustrated by inspecting the raw data (left part of Fig. 5.26), the flat field corrected raw data (central part of Fig. 5.26) and the hologram obtained from propagation of the reconstructed transmission function of the object (right part of Fig. 5.26).

The hologram of the object's transmission function is basically not visible in the recorded raw data. Figure errors resulting from the surface of the KB-mirrors and
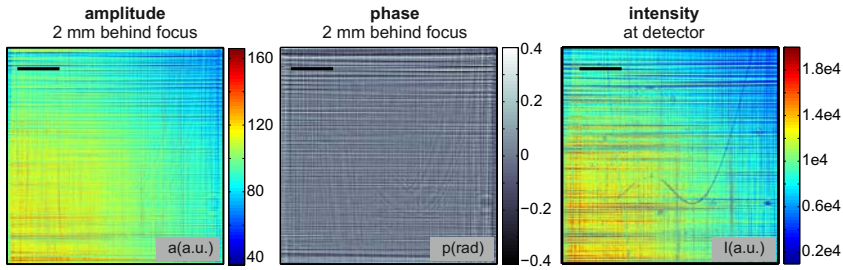
Figure 5.25: Reconstructed probe. Left: Amplitudes at 2.02 mm behind the focal plane. Center: Phases of the probe shown in the image on the left hand side. Right: Reconstructed flat field. Scale bars denote 2 µm.
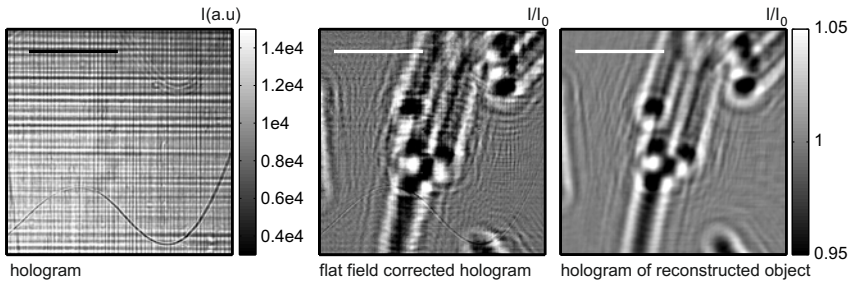
Figure 5.26: Comparison of raw data (left), flat field corrected raw data (center) and reconstructed, propagated transmission function of the sample. Scale bars denote 2 µm.

dirt in the beam path are of much stronger contrast than the object of interest. Since these artifacts are not only very dominant but also in the high frequency domain, the flat field corrected hologram still reveals features which can be attributed to the probe. It is obvious that here, flat field correction cannot be the method of choice to obtain the hologram of the transmission function, which is necessary for those phase retrieval techniques that cannot reconstruct the object along with the probe. Phase retrieval methods relying on defective holograms (like in the central part of Fig. 5.26) will provide reconstructions disturbed by artifacts of the illumination.

The right part of Fig. 5.26 shows the propagated transmission function of the object reconstructed by near-field ptychography. It neither contains features of dirt, nor artifacts of the probe. Separating object and probe during phase retrieval results in more convincing results than methods relying on the standard flat field correction. Interestingly, artifacts resulting from dirt are completely merged into the reconstructed probe and do not spoil the object.

Comparison of CTF based reconstruction and near-field ptychography
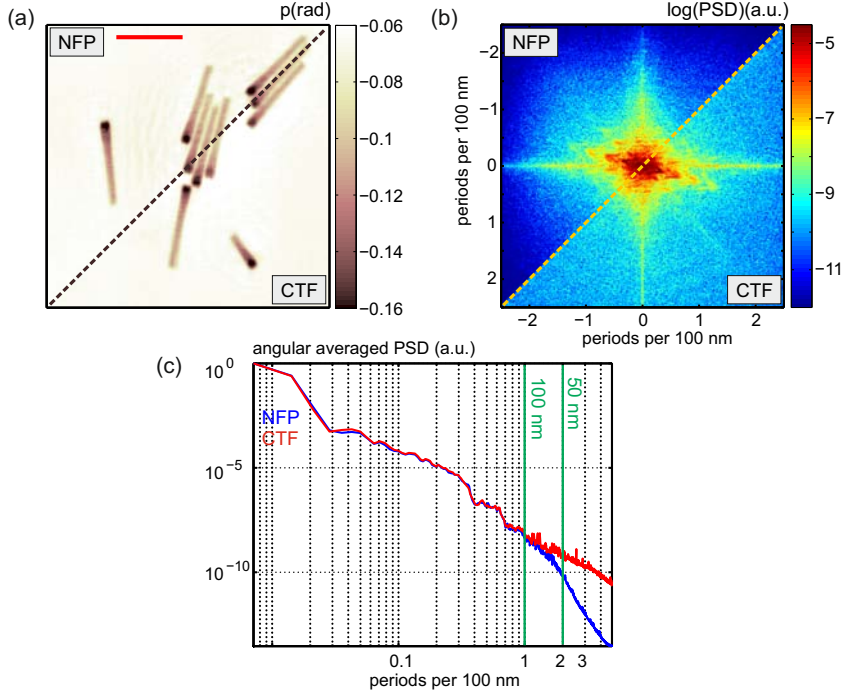
Figure 5.27: Comparison between CTF based reconstruction and near-field ptychography. (a) Composite image of CTF based reconstruction (lower left part) and near-field ptychographic reconstruction (upper right part); scale bar denotes 2 µm. (b) Composite image of power spectral density of (a). (c) Angular averaged power spectral density of CTF based reconstruction (red) and near-field ptychographic reconstruction (blue).

Next, it shall be illustrated, to what extend reconstructions from flat field corrected holograms differ from reconstruction achieved by near-field ptychography. For this purpose, the data collected for near-field ptychography was analyzed using CTF based phase retrieval [11, 95, 97] for each lateral position including the corresponding four defocus shifts. After that, a composite image was formed by sub-pixel alignment and linear blending of all the single CTF reconstructions. Thus the amount of data used for both phase retrieval algorithms – CTF and near-field ptychography – was exactly the same. This is inevitable for a legitimate comparison judging the quality of the reconstruction.

Fig. 5.27(a) reveals that in fact, it is hard to distinguish the CTF phase retrieval result from near-field ptychography. Again, a closer look from a different perspective is useful.

To this end, the power spectral density of both reconstructions was calculated (de-

picted in Fig. 5.27(b)). Despite very similar appearance in real space, CTF based techniques suffer from high frequency noise visible in Fourier space and impeding high resolution.

In contrast, near-field ptychography shows less noise. The same result is also illustrated by the angular averaged power spectral density (Fig. 5.27(c)): Whereas the signals of the CTF and the near-field ptychographic reconstructions are essentially the same for the lower spatial frequencies, both reconstructions start to deviate in the high frequency range. The signal obtained by the CTF reconstruction shows a noisy behavior for frequencies encoding a resolution beyond the 100 nm lines and spaces. In the same region the frequency content of the reconstruction obtained from near-field ptychography decays considerably. This is comparable to the results presented for waveguide imaging: CTF based reconstructions are more strongly affected by noise than reconstructions achieved by near-field ptychography.

Near-field ptychography is an iterative method and compared to the one step approach of CTF, it requires much longer computation time for phase retrieval. Certainly, it is not the method of choice to provide a quick overview of the measured data. However, if reconstructions of high resolution and low noise along with the information of the probe are needed, it has the tendency to show more details than CTF based techniques.

6 Conclusion

The need to correct artifacts resulting from an imperfect illumination was the main motivation of this work. Since conventional flat field correction is strictly speaking not correct and impedes obtaining highly resolved micrographs, an alternative approach is advantageous. For far-field coherent X-ray imaging techniques, simultaneous reconstruction of object and probe is feasible. Yet, so far, for the optical near-field, phase retrieval that separates object and (undisturbed) probe has not been achieved.

The main goal of this thesis was to design a phase retrieval algorithm that closes this gap and can reconstruct object and probe in the optical near-field in a similar way as it is done in far-field ptychography [70, 71, 74]. To achieve this aim, it is necessary to understand the fundamental differences of far- and near-field images: Whereas non-uniqueness inherent in far-field intensity distributions can be overcome by scanning a confined illumination across an extended sample, non-uniqueness in the extended optical near-field is eliminated by translating the sample along and perpendicular to the optical axis.

To this end, the experimental and the algorithmic concept of far-field ptychography was generalized by adding an additional degree of freedom to the displacements of the sample, namely translations to different defocus positions. The benefits for phase retrieval from defocus translations is well known for the optical near-field [11]. Here, longitudinal diversity was combined with lateral diversity for the purpose of phase retrieval for object *and* probe. It was shown that neither with longitudinal diversity alone, nor with exclusively lateral diversity as in far-field imaging methods, it is possible to calculate object and probe from a series of holograms. It is the combination of both, that enables complete separation and artifact free phase retrieval. Hence, a full data set for near-field ptychography consists of holograms that show lateral and longitudinal displacements of the sample.

Experiments were performed with visible light in a parallel beam geometry using a collimated laser and with X-rays at different synchrotron sources. It could be demonstrated successfully that the developed method works equally well for a parallel beam (Fig. 6.1(a)) and a magnifying cone beam configuration (Fig. 6.1(b)). Simulations and experiments indicate that strong phase objects can be reconstructed without the need of linearization of neither the object's transmission function nor the propagator. Furthermore realistic phase shifting properties of a test structure were determined by analyzing the reconstructed transmission function of the object. Finally, the comparison of reconstructions obtained by near-field ptychography and reconstructions obtained by a method that inverts the contrast transfer function reveals the potential of near-field ptychography to provide highly resolved images with low noise.
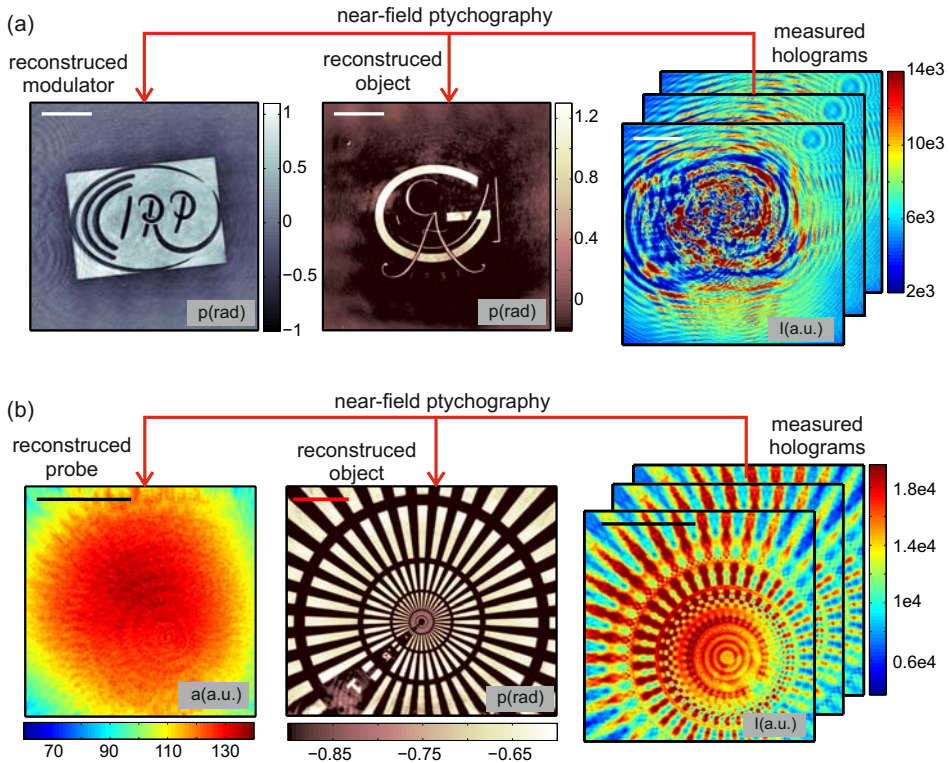
Figure 6.1: (a) Parallel beam near-field ptychography with visible light and modulated probe. Phase retrieval separates object and beam modulator. Scale bars denote 1 mm. (b) Magnifying cone beam near-field ptychography with hard X-rays. Object and clean probe are reconstructed. Scale bars denote 10 μm.

Besides information about the sample, simultaneous reconstruction of object and probe is a method to characterize extended wave fronts. To this end, it could be beneficial to also record the decay of the illumination and to not use an implementation of phase retrieval for the effective, magnifying geometry. Instead, a direct approach to retrieve curved wave fronts could provide new insights.

The experiments with visible light and modulated probe anticipate a further application of near-field ptychography by swapping the roles of sample and beam modulator. A specially designed 'reporter' object could be used to divert radiation scattered by the beam modulator to higher angles than covered by the numerical aperture of the detector and thereby enabling phase retrieval of the beam modulating structure at super-resolution.

References

[1] D. Gabor. Holography, 1948-1971. *Proceedings of the IEEE*, 60(6):655–668, June 1972.

[2] D. Gabor. A New Microscopic Principle. *Nature*, 161(4098):777–778, May 1948.

[3] W. L. Bragg. A new type of x-ray microscope. *Nature*, 143:678, April 1939.

[4] M. K. Kim. *Digital Holographic Microscopy, Principles, Techniques, and Applications*. Springer Series in Optical Sciences 162. Springer Science+Business Media, LLC, 2011.

[5] D. Sayre. Prospects for long-wavelength x-ray microscopy and diffraction. In *Imaging Processes and Coherence in Physics. Proceedings of a Workshop, Held at the Centre de Physique, Les Houches, France, March 1979. Lecture Notes in Physics, Vol. 112, edited by M. Schlenker, M. Fink, J. P. Goedgebuer, C. Malgrange, J. Ch. Viénot, and R. H. Wade*, pages 229–235. Springer-Verlag, Berlin, 1980.

[6] H. N. Chapman and K. A. Nugent. Coherent lensless X-ray imaging. *Nat Photon*, 4(12):833–839, December 2010.

[7] K. A. Nugent. Coherent methods in the X-ray sciences. *Adv. Phys.*, 59(1):1–99, 2010.

[8] J. Miao, P. Charalambous, and D. Sayre. Extending the methodology of X-ray crystallography to allow imaging of micrometre-sized non-crystalline specimens. *Nature*, 400(6742):342–344, July 1999.

[9] H. M. Quiney. Coherent diffractive imaging using short wavelength light sources. *Journal of Modern Optics*, 57(13):1109–1149, 2010.

[10] P. Cloetens, W. Ludwig, J. Baruchel, J.-P. Guigay, P. Pernot-Rejmánková, M. Salomé-Pateyron, M. Schlenker, J.-Y. Buffière, E. Maire, and G. Peix. Hard x-ray phase imaging using simple propagation of a coherent synchrotron radiation beam. *J. Phys. D: Appl. Phys.*, 32(10A):A145–A151, 1999.

[11] P. Cloetens, W. Ludwig, J. Baruchel, D. Van Dyck, J. Van Landuyt, J. P. Guigay, and M. Schlenker. Holotomography: Quantitative phase tomography with micrometer resolution using hard synchrotron radiation x rays. *Appl. Phys. Lett.*, 75(19):2912–2914, 1999.

[12] S. W. Wilkins, T. E. Gureyev, D. Gao, A. Pogany, and A. W. Stevenson. Phase-contrast imaging using polychromatic hard X-rays. *Nature*, 384(6607):335–338, November 1996.

[13] S. Lagomarsino, A. Cedola, P. Cloetens, S. Di Fonzo, W. Jark, G. Soullié, and C. Riekel. Phase contrast hard x-ray microscopy with submicron resolution. *Appl. Phys. Lett.*, 71(18):2557–2559, November 1997.

[14] K. A. Nugent, A. G. Peele, H. N. Chapman, and A. P. Mancuso. Unique Phase Recovery for Nonperiodic Objects. *Phys. Rev. Lett.*, 91(20):203902, 2003.

[15] G. J. Williams, H. M. Quiney, B. B. Dhal, C. Q. Tran, K. A. Nugent, A. G. Peele, D. Paterson, and M. D. de Jonge. Fresnel coherent diffractive imaging. *Phys. Rev. Lett.*, 97(2):025506, 2006.

[16] L. W. Whitehead, G. J. Williams, H. M. Quiney, K. A. Nugent, A. G. Peele, D. Paterson, M. D. de Jonge, and I. McNulty. Fresnel diffractive imaging: Experimental study of coherence and curvature. *Phys. Rev. B*, 77(10):104112, March 2008.

[17] D. M. Paganin. *Coherent X-Ray Optics*. Oxford University Press, New York, USA, 2006.

[18] E. Hecht. *Optics - Fourth Edition*. Addison, Wesley, San Francisco, USA, 2002.

[19] J. W. Goodman. *Statistical Optics*. Wiley Classics Libary, New York, USA, 2000.

[20] F. van der Veen and F. Pfeiffer. Coherent x-ray scattering. *J. Phys.: Condens. Matter*, 16(28):5003, 2004.

[21] B. E. A. Saleh and M. C. Teich. *Grundlagen der Photonik*. WILEY-VCH Verlag, Weinheim, Germany, 2008.

[22] G. Brooker. *Modern Classical Optics*. Oxford University Press, New York, USA, 2011.

[23] P. Willmott. *An Introduction to Synchrotron Radiation*. Wiley and Sons, Ltd., Publication, Chichester, West Sussex, UK, 2011.

[24] J. W. Goodman. *Introduction to Fourier Optics*. Roberts & Company: Englewood, Colorado, 2005.

[25] J. Als-Nielsen and D. McMorrow. *Elements of Modern X-ray Physics*. 2nd edition edition, 2011.

[26] L. De Caro, C. Giannini, A. Cedola, S. Lagomarsino, and I. Bukreeva. X-ray point- and line-projection microscopy and diffraction. *Opt. Commun.*, 265(1):18–28, September 2006.

[27] W. Demtröder. *Experimentalphysik 2, 3. Auflage*. Springer Verlag Berlin Heidelberg, Berlin, Germany, 2004.

[28] M. Born and E. Wolf. *Principles of Optics, 7th expanded edition*. Cambridge University Press, Cambridge, UK, 2002.

[29] H. Mimura, K. Yamauchi, K. Yamamura, A. Kubota, S. Matsuyama, Y. Sano, K. Ueno, K. Endo, Y. Nishino, K. Tamasaku, M. Yabashi, T. Ishikawa, and Y. Mori. Image quality improvement in a hard X-ray projection microscope using total reflection mirror optics. *J. Synchrotron Radiat.*, 11(4):343–346, July 2004.

[30] J. Hagemann, A.-L. Robisch, D. R. Luke, C. Homann, T. Hohage, P. Cloetens, H. Suhonen, and T. Salditt. Reconstruction of wave front and object for inline holography from a set of detection planes. *Opt. Express*, 22(10):11552–11569, May 2014.

[31] C. Homann, T. Hohage, J. Hagemann, A.-L. Robisch, and T. Salditt. Validity of the empty-beam correction in near-field imaging. *Phys. Rev. A*, 91:013821, January 2015.

[32] I. N. Bronstein, K. A. Semendjajew, G. Musiol, and H. Mühlig. *Taschenbuch der Mathematik*. Verlag Harri Deutsch, Frankfurt am Main, Germany, 2006.

[33] K. Giewekemeyer. *A study on new approaches in coherent x-ray microscopy of biological specimens*. Phd thesis, Georg-August-Universität Göttingen, 2011.

[34] T. Butz. *Fouriertransformation für Fußgänger*. Vieweg+Teubner Verlag / Springer Fachmedien Wiesbaden GmbH, Wiesbaden, 2012.

[35] B. J. Thompson. Interference and diffraction. In H. J. Caulfield, editor, *Handbook of Optical Holography*, chapter 2.2, pages 29–30. Academic Press, Inc., New York, USA, 1979.

[36] L. Onural and P. D. Scott. Digital decoding of in-line holograms. *Opt. Eng.*, 26(11):1124–1132, November 1987.

[37] D. G. Voelz and M. C. Roggemann. Digital simulation of scalar optical diffraction: revisiting chirp function sampling criteria and consequences. *Appl. Opt.*, 48(32):6132–6142, November 2009.

[38] D. Mas, J. Garcia, C. Ferreira, L. M. Bernardo, and F. Marinho. Fast algorithms for free-space diffraction patterns calculation. *Opt. Commun.*, 164(4–6):233 – 245, June 1999.

[39] J. F. Restrepo and J. Garcia-Sucerquia. Magnified reconstruction of digitally recorded holograms by Fresnel–Bluestein transform. *Appl. Opt.*, 49(33):6430–6435, November 2010.

[40] N. Verrier and M. Atlan. Off-axis digital hologram reconstruction: some practical considerations. *Appl. Opt.*, 50(34):H136–H146, December 2011.

[41] D. G. Voelz. *Computational Fourier optics : A MATLAB tutorial*. Society of Photo-Optical Instrumentation Engineers (SPIE), Bellingham, Washington, USA, 2011.

[42] J. D. Schmidt. *Numerical simulation of optical wave propagation with examples in MATLAB*. Society of Photo-Optical Instrumentation Engineers (SPIE), Bellingham, Washington, USA, 2010.

[43] L. I. Bluestein. A linear filtering approach to the computation of discrete Fourier transform. *IEEE Transactions on Audio and Electroacoustics*, 18(4):451–455, December 1970.

[44] U. Spagnolini. 2-D phase unwrapping and instantaneous frequency estimation. *IEEE Transactions on Geoscience and Remote Sensing*, 33(3):579–589, May 1995.

[45] K. Matsushima and T. Shimobaba. Band-Limited Angular Spectrum Method for Numerical Simulation of Free-Space Propagation in Far and Near Fields. *Opt. Express*, 17(22):19662–19673, October 2009.

[46] T. Shimobaba, T. Kakue, N. Okada, M. Oikawa, Y. Yamaguchi, and T. Ito. Aliasing-reduced Fresnel diffraction with scale and shift operations. *Journal of Optics*, 15(7):075405, June 2013.

[47] K. D. Mielenz. Computation of Fresnel Integrals. II. *J. Res. Natl. Inst. Stand. Technol.*, 105:589–590, July 2000.

[48] V. Elser. Phase retrieval by iterated projections. *J. Opt. Soc. Am. A*, 20(1):40–55, January 2003.

[49] C. L. Byrne. *Applied iterative methods*. AK Peters Wellesley, 2008.

[50] F. Deutsch. *Best Approximation in Inner Product Spaces*. Springer-Verlag New York, 2001.

[51] H. H. Bauschke and P. L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer Science & Business Media, New York, USA, 2011.

[52] V. Elser, I. Rankenburg, and P. Thibault. Searching with iterated maps. *PNAS*, 104(2):418–423, January 2007.

[53] H. H. Bauschke, P. L. Combettes, and D. R. Luke. Phase retrieval, error reduction algorithm, and Fienup variants: a view from convex optimization. *J. Opt. Soc. Am. A*, 19(7):1334–1345, July 2002.

[54] D. R. Luke, J. V. Burke, and R. G. Lyon. Optical Wavefront Reconstruction: Theory and Numerical Methods. *SIAM Review*, 44(2):169–224, 2002.

[55] D. R. Luke, R. Hesse, and P. Neumann. *Lecture notes: Projection methods and applications to image processing*. 2012.

[56] P. K. Jain, O. P. Ahuja, and K. Ahmad. *Functional Analysis*. New Age International (P) Ltd., New Delhi, India, 1995.

[57] J. Eckstein. *Splitting Methods for Monotone Operators with Applications to Parallel Optimization*. Phd thesis, Massachusetts Institute of Technology, June 1989.

[58] Zdzislaw Opial. Weak convergence of the sequence of successive approximations for nonexpansive mappings. *B. Am. Math. Soc.*, 73(4):591–597, July 1967.

[59] W. Cheney and A. A. Goldstein. Proximity Maps for Convex Sets. *P. Am. Math. Soc.*, 10(3):448–450, June 1959.

[60] A. Levi and H. Stark. Image restoration by the method of generalized projections with application to restoration from magnitude. *J. Opt. Soc. Am. A*, 1(9):932–943, September 1984.

[61] L. M. Brègman. The method of successive projection for finding a common point of convex sets. *Sov. Math. Dokl.*, 6:688–692, 1965.

[62] D. R. Luke. Relaxed averaged alternating reflections for diffraction imaging. *Inverse Problems*, 21(1):37, 2005.

[63] R. W. Gerchberg and W. O. Saxton. A Practical Algorithm for the Determination of Phase from Image and Diffraction Plane Pictures. *OPTIK*, 35(2):237–246, 1972.

[64] M. A. Fiddy and U. Shahid. Legacies of the Gerchberg-Saxton algorithm. *Ultramicroscopy*, 134:48 – 54, 2013. 65th Birthdays of W Owen Saxton, David J Smith and Dirk Van Dyck / {PICO} 2013 & From Multislice to Big Bang.

[65] J. R. Fienup. Phase retrieval algorithms: a comparison. *Appl. Opt.*, 21(15):2758–2769, August 1982.

[66] J. R. Fienup and C. C. Wackerman. Phase-retrieval stagnation problems and solutions. *J. Opt. Soc. Am. A*, 3(11):1897–1907, November 1986.

[67] W. Hoppe. Beugung im inhomogenen Primärstrahlwellenfeld. I. Prinzip einer Phasenmessung von Elektronenbeungungsinterferenzen. *Acta Crystallogr. A*, 25(4):495–501, 1969.

[68] H. M. L. Faulkner and J. M. Rodenburg. Movable Aperture Lensless Transmission Microscopy: A Novel Phase Retrieval Algorithm. *Phys. Rev. Lett.*, 93:023903, July 2004.

[69] J. M. Rodenburg and H. M. L. Faulkner. A phase retrieval algorithm for shifting illumination. *Appl. Phys. Lett.*, 85(20):4795–4797, November 2004.

[70] M. Guizar-Sicairos and J. R. Fienup. Phase retrieval with transverse translation diversity: a nonlinear optimization approach. *Opt. Express*, 16(10):7264–7278, May 2008.

[71] P. Thibault, M. Dierolf, A. Menzel, O. Bunk, C. David, and F. Pfeiffer. High-Resolution Scanning X-ray Diffraction Microscopy. *Science*, 321(5887):379–382, July 2008.

[72] P. Thibault, M. Dierolf, O. Bunk, A. Menzel, and F. Pfeiffer. Probe retrieval in ptychographic coherent diffractive imaging. *Ultramicroscopy*, 109(4):338–343, March 2009.

[73] R. Hesse, D. R. Luke, S. Sabach, and M. K. Tam. Proximal Heterogeneous Block Input-Output Method and application to Blind Ptychographic Diffraction Imaging. *arXiv:1408.1887v1*, pages 1 – 31, August 2014.

[74] A. M. Maiden and J. M. Rodenburg. An improved ptychographical phase retrieval algorithm for diffractive imaging. *Ultramicroscopy*, 109(10):1256 – 1262, September 2009.

[75] H. M. Quiney, A. G. Peele, Z. Cai, D. Paterson, and K. A. Nugent. Diffractive imaging of highly focused X-ray fields. *Nat. Phys.*, 2:101–104, February 2006.

[76] C. Fuhse, C. Ollinger, and T. Salditt. Waveguide-Based Off-Axis Holography with Hard X-Rays. *Phys. Rev. Lett.*, 97:254801, December 2006.

[77] B. Abbey, K. A. Nugent, G. J. Williams, J. N. Clark, A. G. Peele, M. A. Pfeifer, M. de Jonge, and I. McNulty. Keyhole coherent diffractive imaging. *Nat. Phys.*, 4:394–398, May 2008.

[78] D. J. Vine, G. J. Williams, B. Abbey, M. A. Pfeifer, J. N. Clark, M. D. de Jonge, I. McNulty, A. G. Peele, and K. A. Nugent. Ptychographic Fresnel coherent diffractive imaging. *Phys. Rev. A*, 80:063823, December 2009.

[79] M. Stockmar, P. Cloetens, I. Zanette, B. Enders, M. Dierolf, F. Pfeiffer, and P. Thibault. Near-field ptychography: phase retrieval for inline holography using a structured illumination. *Sci. Rep.*, 3(1927):1–6, May 2013.

[80] M. Stockmar, M. Hubert, M. Dierolf, B. Enders, R. Clare, S. Allner, A. Fehringer, I. Zanette, J. Villanova, J. Laurencin, P. Cloetens, F. Pfeiffer, and P. Thibault. X-ray nanotomography using near-field ptychography. *Opt. Express*, 23(10):12720–12731, May 2015.

[81] C. T. Putkunz, J. N. Clark, D. J. Vine, G. J. Williams, M. A. Pfeifer, E. Balaur, I. McNulty, K. A. Nugent, and A. G. Peele. Phase-Diverse Coherent Diffractive Imaging: High Sensitivity with Low Dose. *Phys. Rev. Lett.*, 106:013903, January 2011.

[82] A.-L. Robisch and T. Salditt. Phase retrieval for object and probe using a series of defocus near-field images. *Opt. Express*, 21(20):23345–23357, October 2013.

[83] T. Latychevskaia and H.-W. Fink. Solution to the Twin Image Problem in Holography. *Phys. Rev. Lett.*, 98:233901, June 2007.

[84] P. Langehanenberg, B. Kemper, and G. von Bally. Autofocus algorithms for digital-holographic microscopy. In *Biophotonics 2007: Optics in Life Science*, page 6633_13. Optical Society of America, 2007.

[85] A.-L. Robisch, K. Kröger, A. Rack, and T. Salditt. Near-field ptychography using lateral and longitudinal shifts. *New J. Phys.*, 17(7):073033, 2015.

[86] M. Guizar-Sicairos, S. T. Thurman, and J. R. Fienup. Efficient subpixel image registration algorithms. *Opt. Lett.*, 33(2):156–158, January 2008.

[87] M. Stockmar, I. Zanette M., Dierolf, B. Enders, R. Clare, F. Pfeiffer, P. Cloetens, A. Bonnin, and P. Thibault. X-ray Near-Field Ptychography for Optically Thick Specimens. *Phys. Rev. Applied*, 3:014005, January 2015.

[88] T. Weitkamp, P. Tafforeau, E. Boller, P. Cloetens, J.-P. Valade, P. Bernard, F. Peyrin, W. Ludwig, L. Helfen, and J. Baruchel. Status and evolution of the ESRF beamline ID19. In *X-ray optics and microanalysis: Proceedings of the 20th International Congress, Karlsruhe (Germany), 15-18 September 2009*, volume 1221, pages 33–38, 2010.

[89] J.-C. Labiche, O. Mathon, S. Pascarelli, M. A. Newton, G. G. Ferre, C. Curfs, G. Vaughan, A. Homs, and D. F. Carreiras. Invited article: The fast readout low noise camera as a versatile x-ray detector for time resolved dispersive extended x-ray absorption fine structure and diffraction studies of dynamic problems in materials science, chemistry, and catalysis. *Rev. Sci. Instrum.*, 78(9):091301–11, September 2007.

[90] S. Kalbfleisch, H. Neubauer, S. P. Krüger, M. Bartels, M. Osterhoff, D. D. Mai, K. Giewekemeyer, B. Hartmann, M. Sprung, and T. Salditt. The Göttingen Holography Endstation of Beamline P10 at PETRA III/DESY. In I. McNulty, C. Eyberger, and B. Lai, editors, *The 10th international conference on X-ray microscopy, Chicago, Illinois (USA), 15-20 August 2010*, volume 1365, pages 96–99, 2011.

[91] T. Salditt, S. Kalbfleisch, M. Osterhoff, S. P. Krüger, M. Bartels, K. Giewekemeyer, H. Neubauer, and M. Sprung. Partially coherent nano-focused x-ray radiation characterized by Talbot interferometry. *Opt. Express*, 19(10):9656–9675, May 2011.

[92] T. Salditt, M. Osterhoff, M. Krenkel, R. N. Wilke, M. Priebe, M. Bartels, S. Kalbfleisch, and M. Sprung. Compound focusing mirror and X-ray

waveguide optics for coherent imaging and nano-diffraction. *J. Synchrotron Radiat.*, 22(4):867–878, July 2015.

[93] H. Neubauer, S. Hoffmann, M. Kanbach, J. Haber, S. Kalbfleisch, S. P. Krüger, and T. Salditt. High aspect ratio x-ray waveguide channels fabricated by e-beam lithography and wafer bonding. *J. Appl. Phys.*, 115(21):214305, June 2014.

[94] CXRO – X-Ray Properties of the Elements . `http://henke.lbl.gov/optical_constants/pert_form.html`. Accessed: 2015-05-28.

[95] S. Zabler, P. Cloetens, J.-P. Guigay, J. Baruchel, and M. Schlenker. Optimization of phase contrast imaging using hard x rays. *Rev. Sci. Instrum.*, 76(7):073705, June 2005.

[96] J. P. Guigay, M. Langer, R. Boistel, and P. Cloetens. Mixed transfer function and transport of intensity approach for phase retrieval in the Fresnel region. *Opt. Lett.*, 32(12):1617–1619, June 2007.

[97] M. Krenkel, M. Bartels, and T. Salditt. Transport of intensity phase reconstruction to solve the twin image problem in holographic x-ray imaging. *Opt. Express*, 21(2):2220–2235, January 2013.

[98] R. N. Wilke, J. Wallentin, M. Osterhoff, D. Pennicard, A. Zozulya, M. Sprung, and T. Salditt. High Flux Ptychographic Imaging Using the New 55 μm-Pixel Detector 'Lambda' Based on the Medipix3 Readout Chip. *Acta Crystallogr.A*, 70(6):552–562, 2014.

List of publications

A.-L. Robisch, K. Kröger, A. Rack, and T. Salditt. Near-field ptychography using lateral and longitudinal shifts. *New J. Phys.*, 17(7):073033, August 2015.

C. Homann, T. Hohage, J. Hagemann, A.-L. Robisch and T. Salditt. Validity of the empty-beam correction in near-field imaging. *Phys. Rev. A*, 91:013821, January 2015.

K. Keskinbora, A.-L. Robisch, M. Mayer, U. T. Sanli, C. Grévent, C. Wolter, M. Weigand, A. Szeghalmi, M. Knez, T. Salditt and G. Schütz. Multilayer Fresnel zone plates for high energy radiation resolve 21 nm features at 1.2 keV. *Opt. Express*, 22(15):18440–18453, July 2014.

J. Hagemann, A.-L. Robisch, D. R. Luke, C. Homann, T. Hohage, P. Cloetens, H. Suhonen and T. Salditt. Reconstruction of wave front and object for inline holography from a set of detection planes. *Opt. Express*, 22(10):11552–11569, May 2014.

A.-L. Robisch and T. Salditt. Phase retrieval for object and probe using a series of defocus near-field images. *Opt. Express*, 21(20):23345–23357, October 2013.

M. Osterhoff, M. Bartels, F. Döring, C. Eberl, T. Hoinkes, S. Hoffmann, T. Liese, V. Radisch, A. Rauschenbeutel, A.-L. Robisch, A. Ruhlandt, F. Schlenkrich, T. Salditt and H.-U. Krebs. Two-dimensional sub-5-nm hard x-ray focusing with MZP. *Proc. of SPIE* 8848, Advances in X-Ray/EUV Optics and Components VIII, 884802, September 27, 2013.

K. Keskinbora, A.-L. Robisch, M. Mayer, C. Grévent, A. V. Szeghalmi, M. Knez, M. Weigand, I. Snigireva, A. Snigirev, T. Salditt and G. Schütz. Recent advances in use of atomic layer deposition and focused ion beams for fabrication of Fresnel zone plates for hard x-rays. *Proc. of SPIE* 8851, X-Ray Nanoimaging: Instruments and Methods, 885119, September 26, 2013.

F. Döring, A.L. Robisch, C. Eberl, M. Osterhoff, A. Ruhlandt, T. Liese, F. Schlenkrich, S. Hoffmann, M. Bartels, T. Salditt and H.U. Krebs. Sub-5 nm hard x-ray point focusing by a combined Kirkpatrick-Baez mirror and multilayer zone plate. *Opt. Express*, 21(16):19311–19323, August 2013.

Danksagung – Acknowledgments

Die schönen Fotos, die ich für die Simulationen verwenden durfte, stammen von Lara Wimmer. Vielen Dank dafür!

Für hilfreiche Korrekturen und geduldiges Lesen dieser Arbeit bedanke ich mich herzlichst bei Johannes Hagemann, Sarah Hoffmann, Dr. Henrike Neubauer und Lara Wimmer.

Allen KollegInnen aus dem Institut für Röntgenphysik danke ich für ihre freundschaftliche und kameradschaftliche Arbeitsweise.

Nicht zuletzt danke ich meiner Familie, meinen Eltern, meinen FreundInnen und besonders Iwan Fritsch für ihre Rücksichtnahme und ihr Vertrauen.

Curriculum Vitae

Anna-Lena Robisch

geboren am	02.07.1987
in	Marktredwitz
Staatsangehörigkeit	deutsch

Wissenschaftlicher Werdegang

2012–2015	Promotionsstudium der Physik an der Georg-August-Universität Göttingen
seit 2012	Wissenschaftliche Mitarbeiterin am Institut für Röntgenphysik der Georg-August-Universität Göttingen
08/2012	Masterarbeit: *Phasenrekonstruktion von Objekttransmissions- und Beleuchtungsfunktion auf der Basis nichtlinearer Optimierung*
2010–2012	Studium der Physik an der Georg-August-Universität Göttingen
2012	Wissenschaftliche Hilfskraft am Institut für Röntgenphysik der Georg-August-Universität Göttingen
07/2010	Bachelorarbeit: *Bikontinuierliche C(Y)-Strukturen als photonische Kristallmodule mit Händigkeit*
2007–2010	Studium der Physik an der Friedrich-Alexander-Universität Erlangen-Nürnberg
06/2007	Abitur
1998–2007	Schulbildung am Walter-Gropius-Gymnasium Selb

Lensless, holographic X-ray microscopy is a non-invasive imaging technique that provides resolution on the nanometer scale. Therefore, a divergent, coherent and especially clean wave front impinging on the sample is needed. Yet, focusing X-rays by even the most advanced X-ray mirrors causes so called figure errors of high spatial frequency content. The results are strongly deteriorated intensity profiles that are often even more pronounced than the holographic image of the sample itself.

A common strategy to compensate these figure errors is to divide the hologram by the pure intensity profile of the beam (the so called flat field). However, this division is only valid in the limiting case of an illumination focused down to a point source. In reality, as a consequence of a finite spot size, one has to accept a loss in resolution when performing the flat field correction. An approach different from the described straightforward procedure is necessary. Here, the simultaneous reconstruction of object and probe is proposed using holograms which were not flat field corrected before phase retrieval.

To this end, a method has been developed that allows simultaneously reconstructing object and probe in amplitude and phase from holographic intensity recordings. The experimental way of proceeding was mainly inspired by well-established holographic full-field X-ray imaging techniques that require holograms defocused to different degrees. Consequently, the conclusion seems reasonable that diversity in the optical near-field arises mainly from variation of the propagation distance of light. This so called longitudinal diversity is used to properly phase the transmission function of the sample of interest. The algorithmic strategy of simultaneous phase retrieval for object and probe draws on far-field ptychography where lateral translations of the sample create diverse diffraction patterns. In view of the need for longitudinal diversity realized by shifts of the sample along the optical axis, ptychography has been generalized and adapted for the optical near-field. Hence, translations of the sample in all three dimensions of space need to be exploited to collect enough information about object and probe such that both can be reconstructed simultaneously in amplitude and phase. Concepts have been put into practice by simulations as well as by experiments with coherent visible light and hard X-rays from synchrotron sources.

The presented approach offers the opportunity to perform high resolution imaging, to be extended to tomography and to be adapted to super-resolution experiments.

GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

Universitätsverlag Göttingen