Amplitude and phase reconstruction for Low-Energy Electron Holography of individual proteins

Hannah Ochner

EPFL

These transient facts,
These fugitive impressions,
Must be transformed by mental acts,
To permanent possessions.
Then summon up your grasp of mind,
Your fancy scientific,
Till sights and sounds with thought combine
Become of truth prolific.

James Clerk Maxwell: To the Chief Musician upon Nabla: A Tyndallic Ode

Abstract

Single-molecule imaging methods are of importance in structural biology, and specifically in the imaging of proteins, since they can elucidate conformational variability and structural changes that might be lost in imaging methods relying on averaging processes. Low-energy electron holography (LEEH) is a promising technique for imaging individual proteins with negligible radiation damage, which can be combined with a sample preparation process by native electrospray ion beam deposition (native ES-IBD) ensuring the creation of chemically pure samples suitable for holographic imaging.

The central step in the analysis of data measured by LEEH is the numerical reconstruction of the object from the experimentally acquired holograms. Full information about the imaged object consists of both amplitude and phase information imprinted on the scattered wave during the interaction of the electron beam with the molecule and encoded in the hologram created by the interference of the scattered wave and the transmitted incident wave. A propagation-based algorithm with the goal of reconstructing the wave field in the plane of the object is presented and applied to holograms of individual proteins prepared by ES-IBD and measured with a low-energy electron holography microscope. The information retrieved from the reconstruction of the amplitude distribution in the object plane is discussed by analysing holograms of antibodies, demonstrating that the inherent conformational variability of these molecules can be mapped by LEEH. The influence of the sample preparation process on the surface conformations is tracked by tuning the landing energy of the proteins during deposition.

To complement the amplitude data, an iterative phase retrieval algorithm is implemented to reconstruct the phase distribution in the object plane along with the amplitude distribution. The algorithm's performance and robustness is thoroughly evaluated and simulations regarding multiple scattering effects and element-dependent variations in scattering strength are carried out to provide a reference for the interpretation of phase data retrieved from experimentally acquired holograms. The iterative phase retrieval algorithm is then applied to protein data, indicating that both molecular density and charges can be related to features in the phase reconstructions, while the presence of metals does not correlate with specific phase signals at the current resolution obtainable from the experimental data.

Since proteins are inherently three-dimensional, approaches towards three-dimensional reconstruction schemes are discussed, which will be the focus of future work.

Keywords: low-energy electron holography, hologram reconstruction, phase retrieval, single-molecule imaging, native electrospray ion beam deposition, protein imaging

Zusammenfassung

Einzelmolekülmethoden spielen eine wichtige Rolle in der Strukturbiologie, und insbesondere bei der Abbildung von Proteinstrukturen, da sie die Darstellung von konformationeller Variabilität und von Strukturänderungen erlauben, welche von Methoden, die Mittelungsprozesse über eine große Anzahl von Molekülen erforden, nicht abgebildet werden können. Niederenergetische Elektronenholografie (engl. low-energy electron holography, LEEH) ist eine vielversprechende Methode zur Einzelmolekülabbildung von Proteinen unter Vermeidung von Strahlenschäden, die sich gut mit einer Probenpräparation durch Elektrospray-Ionenstrahldeposition (engl: electrospay-ion beam deposition, ES-IBD) verbinden lässt, wodurch die Produktion chemisch reiner, für die Abbildung durch LEEH geeigneter Proben gewährleistet werden kann.

Der zentrale Schritt in der Analyse von LEEH-Daten besteht in der numerischen Rekonstruktion des Objekts aus dem gemessenen Hologramm. Die vollständige Information über das abgebildete Objekt geht während des Streuprozesses in Form von Amplituden- und Phasenänderungen auf die einfallende Welle über und wird durch die Intereferenz zwischen gestreuter und unverändert transmittierter Welle im Hologramm gespeichert. Ein auf Wellenpropagation basierender Algorithmus zur Rekonstruktion des komplexen Wellenfelds in der Objektebene wird vorgestellt und auf Hologramme einzelner Proteine, die durch ES-IBD präpariert und mit einem LEEH-Mikroskop gemessen wurden, angewandt. Die in diesem Prozess gewonnenen Amplitudendaten werden anhand von Antikörperhologrammen diskutiert, wobei gezeigt wird, dass die dieser Art von Molekülen inhärente konfomationelle Variabilität durch LEEH abgebildet werden kann. Der Einfluss des Probenpräparationsprozesses auf die Oberflächenkonformationen wird durch Änderung der kinetischen Energie der Moleküle während des Landungsprozesses untersucht.

Um zusätzlich zu den Amplitudendaten die Phasendistribution in der Objektebene zu rekonstruieren, wird ein iterativer Algorithmus eingesetzt. Das Verhalten und die Fehlerrobustheit des Algorithmus werden evaluiert und Effekte multiplen Streuens und elementabhängiger Variationen im Streupotential werden simuliert, um einen Referenzrahmen für die Interpretation der Phasenrekonstruktion experimentell gemessener Hologramme zu schaffen. Die Anwendung des iterativen Phasenrekonstruktionsalgorithmus auf Proteindaten zeigt, dass sowohl Dichtevariationen im Molekül als auch Ladungen mit spezifischen Merkmalen in der Phasenrekonstruktion in Verbindung gebracht werden können; die derzeitige Auflösung erlaubt es jedoch nicht, Phasensignale der Präsenz von Metalatomen im Molekül zuzuschreiben.

Da Proteine dreidimensionale Objekte sind, werden abschließend verschiedene Ansätze für eine dreidimensionale Rekonstruktion diskutiert.

Zusammenfassung

Stichwörter: Niederenergetische Elektronenholografie (LEEH), Hologrammrekonstruktion, Phasenrekonstruktion, Einzelmolekülabbildung, native Elektrospray-Ionenstrahldeposition (ES-IBD), Proteinabbildung

Contents

Contents

List of Figures

List of Abbreviations

2D	two-dimensional
3D	three-dimensional
ADH	Alcohol dehydrogenase
cryo-EM	cryogenic Electron Microscopy
cyt C	cytochrome C
ESI	Electrospray Ionization
ES-IBD	Electrospray Ion Beam Deposition
FFT	Fast Fourier Transform
FIM	Field Ion Microscope
IgG	Immunoglobulin G
LEEH	Low-Energy Electron Holography
MAdCAM-1	Mucosal vascular Adressin Cell Adhesion Molecule 1
PDB	Protein Data Bank
PMMA	Polymethylmethacrylate
PTA	Phosphotungstic acid
SEM	Scanning Electron Microscopy
SLG	Single Layer Graphene
TEM	Transmission Electron Microscopy
UHV	Ultra High Vacuum
XFEL	X-ray Free Electron Laser

Introduction

Biological function of proteins is often related to structure and structural changes [1, 2, 3, 4, 5]. Relevant changes can occur on all levels of the structural hierarchy, from the atomic level to the level of subunits; often, conformational changes in higher levels of the structural hierarchy (structural motifs, domains, entire subunits) are the result of small-scale structural alterations [1, 6]. Thus, high-resolution biomolecular imaging techniques capable of resolving sub-molecular detail play an essential role in answering biological and biomedical questions on the molecular level [7], with direct consequences for applications as for example in drug discovery [8, 9, 10].

Currently, the leading techniques for atomically resolved structure determination from protein ensembles are X-ray crystallography, cryogenic electron microscopy (cryo-EM) and nuclear magnetic resonance spectroscopy (NMR) [11, 12, 13, 14, 15]. X-ray crystallography calculates the electron density of a protein from X-ray diffraction patterns of small protein crystals. Because of streamlined, automated processes, it allows fast, high-throughput structure determination of a large range of proteins [15]. While X-ray crystallography is not limited in terms of protein size, NMR spectroscopy performs better for smaller proteins [16]. In NMR spectroscopy, structural detail can be inferred from the measurement of chemical shifts, the nuclear magnetic resonance frequencies relative to a standard in an external magnetic field. The distribution of chemical shifts can be related to the electron distribution, from which distances between the nuclei, and in turn the molecular structure, can be determined [17]. NMR additionally provides insights into protein dynamics on various time scales [18, 19].

Recent advances in instrumentation resulted in a growing number of high-resolution protein structures from cryo-EM imaging [20]. Protein samples are prepared by plunge freezing [21] and imaged by transmission electron microscopy at low radiation doses, resulting in a reduced radiation damage, but also reduced contrast, which is compensated by sophisticated alignment and averaging workflows. While the method can be applied to proteins that cannot be crystallized, sample preparation is often challenging, and successful structure determination depends on the performance of the classification algorithms employed to retrieve a molecular structure from a multitude of imaged proteins in a sample [15].

The large amount of protein structures in the Protein Data Bank (PDB) bears witness to the success of these techniques regarding protein structure determination. There are, however, proteins with properties that make them difficult to study with the aforemen-

tioned methods, either because they are not suitable in terms of sample preparation (e. g. they are not crystallizable) or because they exhibit flexible or disordered domains or a high degree of conformational variability.

Since all of the techniques mentioned above rely on an averaging step that can involve thousands of molecules, such structural variability, which can lead to significant differences in appearance between instances of molecules, complicates the imaging process. Specifically, features in which the individual molecules differ from one another, will be averaged out. Thus, the mapping of conformational variability is a major challenge in protein structure determination [22, 23]. Single-molecule imaging methods, which do not require averaging, could hence provide relevant complementary information, especially regarding the imaging of conformationally variable proteins.

Low-energy electron holography (LEEH) is a promising single-molecule technique for protein imaging since it provides the high contrast required for the imaging of individual proteins at negligible radiation damage [24, 25, 26]. It operates at low electron energies in the range of 50-150 eV and is experimentally implemented as a lens-free in-line holography set-up. The design is based on the imaging geometry originally suggested by Gabor [27], who first proposed holography as an imaging technique in 1948. In in-line holography, the electron emitter, the sample and the detector are arranged along the same axis. The incident wave produced by the emitter (the reference wave) interacts with the sample, resulting in a scattered wave (the object wave). The interference pattern created by the object wave and the reference wave at the detector is called the hologram. It encodes the information about the amplitude and phase distribution of the imaged object, which, since the hologram is not a real-space image, has to be retrieved by a numerical reconstruction process.

The fabrication of suitable electron emitters enabled the development of a low-energy electron in-line holography set-up [28, 29, 30] which allowed for the imaging of suspended macromolecules [25]. The presence of charging artefacts due to the lack of a substrate initially hindered the hologram reconstruction. However, with the introduction of single layer graphene as a substrate [31], the imaging and reconstruction of biomolecules was facilitated [32], culminating in the demonstration of the LEEH imaging of small, globular proteins at a resolution of approximately 1 nm [24] when combining LEEH with sample preparation by native electrospray ion beam deposition (ES-IBD) [33].

Since the image of the object is obtained by numerical reconstruction from the measured holograms, the development and optimization of reconstruction algorithms is crucial for successful imaging.

The interaction of low-energy electrons with proteins will lead to both elastic and inelastic scattering events, resulting in changes to both the amplitude and phase of the incident wave. The information imprinted on the wave by the scattering processes is stored in the hologram, which is ultimately recorded in the holographic imaging process. Hence, data processing and reconstruction to extract the information encoded in the hologram plays a central role in LEEH. This requires the development of an image analysis workflow consisting of the image reconstruction and its interpretation in terms of the molecular structure. Amplitude information can be reconstructed by a one-step propagation-based algorithm, while phase retrieval and the extraction of three-dimensional (3D) information require more advanced techniques. So far, only the

application of the one-step reconstruction has been demonstrated for experimental LEEH data, while algorithms for phase and 3D analysis have solely been discussed in the context of simulated examples.

Next to the optimization of the reconstruction process itself, a statistical analysis of the measured data is essential for the interpretation of the features emerging in both the amplitude and phase reconstructions since this allows the evaluation of structural details elucidated on the level of an individual protein by LEEH imaging in comparison to other molecules.

The goal of this thesis is to apply LEEH imaging to a wide range of proteins, including proteins with a high degree of structural flexibility, and to explore the structural information encoded in both the amplitude and phase reconstructions of protein holograms. This is achieved by optimizing amplitude and phase reconstruction algorithms and by a subsequent detailed analysis of the results of applying those algorithms to experimentally acquired holograms of proteins.

Chapter 1 provides the mathematical background of hologram formation and reconstruction and details both the numerical implementation of the reconstruction algorithm as well as the experimental set-up.

In Chapter 2, the imaging of proteins exhibiting conformational variability is studied using the example of antibody molecules. LEEH's capability of distinguishing different antibody conformations and of mapping the structural variability of flexible proteins is demonstrated. Furthermore, the relationship between the ES-IBD sample preparation process and the observed surface conformations is explored by an in-depth statistical analysis. Different classes of surface conformations are related to processes occurring in the gas phase as well as during the landing on the graphene surface, and the role of the landing energy regarding the abundance of certain types of surface conformations is elucidated.

Chapters 3 and 4 address phase reconstruction and the interpretation of phase data. An iterative phase retrieval algorithm and its numerical implementation are presented and characterized in detail in Chapter 3. Its performance and robustness with respect to artefacts is quantified. Simulations modelling the effects of multiple scattering and locally different scattering strengths on the phase reconstruction are performed and discussed in terms of possible interpretations of phase data.

In Chapter 4, the iterative phase retrieval algorithm is applied to experimentally acquired holograms. Different possible contributions to the phase reconstructions are considered. The experimental evidence suggests that the reconstructed phase distribution is related to the molecular density. Charges and local changes in electric potential can also affect the reconstructed phase distribution, whereas a change in local scattering strength due to the presence of metals could not be detected at the current resolution.

Chapter 5 explores first steps towards a three-dimensional (3D) reconstruction. An approach towards 3D reconstruction based on an iterative deconvolution algorithm with a single-hologram input is discussed with the aid of simulated examples, and general considerations regarding a tomographic reconstruction within the LEEH framework are outlined.

Chapter 6 summarizes the results presented in this thesis and provides an outlook onto future projects.

1 Low-energy electron holography: theory and experimental set-up

Holography as an imaging method was developed by Dennis Gabor in 1947 [34] as part of his quest to achieve atomic resolution with electron microscopes. In 1933, the electron microscope, first introduced by Max Knoll and Ernst Ruska in 1931 [35], was shown to be capable of achieving higher resolution than traditional optical microscopes [36]. The resolution was, however, still far from atomic resolution. While the deBroglie wavelength of high energy electrons ($100\,\mathrm{keV}$ - $1\,\mathrm{MeV}$) as used in electron microscopy is in the picometer range, i.e. much smaller than the atomic diameter, and would thus in principle allow for atomic resolution, the aberrations of the electron lenses, specifically spherical and chromatic aberrations [37], led to a theoretical resolution limit of $4\,\text{Å}$ and to a practical resolution limit of approximately $12\,\text{Å}$ by the mid 1940s, which made atomic resolution impossible to attain. Gabor realized that an improvement of the electron lenses sufficient to allow for atomic resolution would be extremely difficult to achieve [38]. Indeed, it took another 50 years until the development of spherical-aberration-corrected electron lenses [39].

Instead of working on improving the performance of electron lenses, Gabor suggested to overcome the limitations of electron optics by making use of the wave nature of electrons, in particular the ability of coherent waves to create interference patterns. This allowed him to propose a microscopic set-up that did not require any electrostatic or electromagnetic lenses [27, 38]. By superimposing the wave scattered by the object with a coherent reference wave, an interference pattern that stores the full information about the object wave is created. This property of the imaging process inspired the name Gabor chose for the new technique: the word *holography* is a combination of the Greek words *holos* (whole) and *graphē* (to write, draw). Gabor predicted that from this interference pattern, when recorded and subsequently illuminated by the reference wave, the object wave could be reconstructed [27, 38, 34]. He envisioned this "new microscopic principle" [27] as a two-step process. The first step consists of the creation of the interference pattern using an electron beam, in the second step, the object would be reconstructed via a light-optical process. Since low-aberration optical lenses are much easier to fabricate than low-aberration electron lenses, using a light-optical reconstruction process can circumvent the problems induced by the electron lens aberrations [27, 38]. In order to employ light in the second step, the hologram produced by the electron beam has to be scaled up to account for the difference in wavelength between electrons and visible light.

The incorporation of the reference wave into the imaging process, which is the main difference between holographic and diffractive imaging, allows the storage of spatial information in the interference pattern, which makes this technique suitable for imaging any kind of object without requirements regarding symmetry or periodicity [40].

Gabor provided a proof of principle of the holography method in 1947 using an optical setup consisting of a mercury lamp and a pinhole as a partially coherent source, microphotographs as objects and a photographic plate to record the hologram [27, 38, 34]. Since holography requires coherent sources, which are needed to create interference patterns, the method only became widely applied in the early 1960s, when the advent of the laser [41, 42] provided a coherent and bright radiation source.

Dennis Gabor was awarded the Nobel prize in physics for the development of holography in 1971. Today, the holographic imaging method is used in a wide range of experimental set-ups: in optical settings with lasers [43, 44], in X-ray [45, 46, 47] and free-electron laser [48] applications, and in electron microscopy, both at high [49, 50, 51, 52] and low electron energies [24, 53]. Two different imaging geometries are commonly used, referred to as in-line geometry and off-axis geometry. In an in-line geometry, object wave and reference wave share the same optical axis, while in an off-axis geometry, the two waves propagate along different axes and are only superimposed in the detector plane. While the former geometry can be implemented without the use of lenses, the latter requires lenses. Thus, in imaging situations for which aberration-corrected lenses are available, an off-axis geometry, as first demonstrated by Leith and Upatnieks in 1962 [54], is often employed, since it facilitates the removal of the twin image [49, 55], a contribution to the hologram that prevents perfect reconstruction (see section 1.1.1). An in-line geometry for holographic imaging is preferable when no suitable aberration-corrected lenses are available for the radiation type used for imaging.

The goal of this thesis is to investigate electron holography as an imaging method for biomolecules at the single-molecule level. When applying holography to the single-molecule imaging of biomolecules, such as proteins, several factors have to be taken into account in choosing a suitable imaging set-up and radiation source. Specifically, three main parameters have to be considered: the desired resolution, which, since the length scales of interest in protein imaging are in the angstrom range, would ideally be in the size range of a single atom, the achievable image contrast and the radiation damage induced by the irradiating beam. These parameters cannot be independently selected, as will be discussed in the following section.

To obtain sufficient contrast for imaging proteins at the single-molecule level without averaging or staining, high scattering cross sections characterising the interaction between the molecules and the beam are necessary. The electron scattering cross sections increase with decreasing energy, reaching a maximum around 100 eV [56, 57, 58]. Diffraction-limited resolution depends on the wavelength, hence, in that respect, high energies, which correspond to short wavelengths, are advantageous.

In theory, resolution is only limited by the wavelength and the optical system (e.g. by electron lenses), in practice, radiation damage is an additional limiting factor regarding the attainable resolution of an imaging technique, especially when investigating radiation-sensitive objects such as biomolecules [59, 60, 61]. The main types of radiation damage are atomic displacement (knock-on damage), associated with high-angle elastic scattering

events [61, 62, 63], and radiolysis [61, 62], the breaking of bonds by ionising radiation. Radiolysis is the result of inelastic scattering processes and can both be induced by primary radiation, i. e. the incoming beam, as well as by secondary electrons generated by the primary damage [60, 64]. Radiation damage during imaging can significantly reduce the achievable resolution. For high resolution imaging, radiation damage must thus either be minimized, e. g. by introducing cryogenic conditions, which reduces damage by secondary electrons [60, 62], or "outrun", i. e. the imaging process has to happen on a shorter timescale than the damage ("diffraction before destruction" [65, 66]). Since radiation damage is more severe in photon irradiation than in electron irradiation, as the energy transferred per inelastic scattering event is larger [67], the latter concept is mainly pursued in XFEL imaging [65, 66].

Atomic displacement does only significantly contribute to radiation damage at high energies because it requires incident energies above a threshold energy [62], which depends on the atomic number Z of the target atom; energies sufficient for hydrogen displacement are in the range of $5 \, \text{keV}$ [67]. Higher scattering cross sections at lower energies yield higher image contrast, and hence an increased information content per image at a fixed dose. While this speaks in favour of employing low-energy electrons for imaging single biomolecules, it would seem that the drawback would be an increase in radiation damage by radiolysis. However, experimental studies with low-energy electrons ($\sim 100 \, \text{eV}$) have shown that proteins appear to be stable on the time scale of minutes [24, 25] even at electron doses [25, 26] exceeding those typically employed in electron microscopy [59, 11]. The interaction of low-energy electrons and biological matter is complex, hence radiation damage in this energy range is not yet fully understood. Despite the high cross sections and the corresponding short inelastic mean free paths in the nanometer range [57], there are indications that mechanisms exist, which limit the radiation damage by beams of very low incident energies in the range of $100 \, \text{eV}$ [61, 26], especially when imaging thin samples. In thin samples, such as individual proteins, whose thickness is in the order of magnitude of the inelastic mean free path, smaller amounts of secondary electrons are produced. Additionally, in the case of thin samples, these secondary electrons can escape into vacuum before interacting with the object [61, 68], which reduces the amount of multiple scattering processes within the sample. Moreover, at very low beam energies, the incident electrons may not have sufficient energy to create a large number of secondary electrons, thereby limiting radiation damage [61, 68]. Depending on the dominant ionization mechanism underlying radiolysis, there could be energy thresholds below which radiation damage would decrease dramatically, e. g. if only K-shell ionization, which requires energies exceeding certain, element-dependent thresholds, would significantly contribute to ionization-related damage [61, 69, 70, 71].

While experimental evidence of low-energy electron imaging shows that radiation damage does not seem to be critical for imaging at a resolution of approximately $1 \, \text{nm}$ [24], since no structural changes are observed at this length scale, it is of course possible that the radiation damage occurs on smaller lengths scales while leaving the overall molecular structure intact.

These considerations imply that thin samples, such as single proteins, could be imaged by low-energy electrons with minimal radiation damage. Holography with low-energy electrons in the range of 50-200 eV hence appears to be a promising candidate for non-destructive biomolecular imaging, even over longer time scales in the range of minutes.

Since lens corrections at these energies are very difficult to engineer as aberrations are energy-dependent and are stronger at low energies [72], the lensless in-line geometry originally proposed by Gabor [27, 38] is the preferable imaging geometry for this energy range.

This chapter will be divided into two parts: the first part will expound the theory of in-line holography, starting from general principles and covering both the mathematical formulation and the numerical implementation of hologram formation and reconstruction. Furthermore, coherence of the source as a prerequisite of holography as well as the achievable resolution will be discussed. The second part will focus on the application of holography to the specific case of single-molecule imaging of biomolecules with low-energy electrons in an in-line holography microscope. The experimental set-up as well as the sample preparation steps required for successful imaging of biomolecules with low-energy electron holography will be explained in detail.

1.1 Holography Theory

1.1.1 General principle of holography

Holographic imaging is based on the superposition of the wave front diffracted by an object (the object wave) with a coherent background wave front that has not interacted with the object (the reference wave). This superposition leads to the interference of the reference wave with the coherent part of the object wave, creating an interference pattern that stores information about both the amplitude and phase distribution of the object [27, 38]. A record of this interference pattern on a detector or a photographic plate, which encodes the modulus squared of the complex wave field in the detector plane, is referred to as the hologram. Maxima in the hologram correspond to regions in which the object wave and the reference wave had the same phase, which resulted in a constructive interference of the two waves. Thus, if the recorded hologram, e.g. in the form of a photographic plate, is in turn illuminated by the reference wave, the phases in the regions of maximum transmission will coincide with the phases of the object, and the variation of transmissibility will map the amplitude modifications induced by the object [27]. As a result, the object can be reconstructed from the hologram by illumination with the reference wave.

In in-line holography, the experimental geometry for holographic imaging discussed in this thesis, reference wave and object wave share the same optical axis. Hence, the reference wave itself is used to illuminate the object during hologram generation [27, 38, 73] (see Fig. 1.1). When using a point source to illuminate an object, the reference wave has the form of a spherical wave, which results in the geometric magnification of the object without the need for any optical elements. While part of the incident reference wave is scattered due to the interaction with the object, resulting in the creation of the object wave, most of the reference wave will pass through the object plane without scattering. Behind the object, the reference wave and the object wave interfere, and the resulting wavefront propagates to the detector, where it is recorded as a hologram (see Fig. 1.1). The hologram formation comprises the first step of the two-step holographic imaging

process [38]. As a second step, the wave front in the object plane, which is usually referred to as the exit wave [73], is reconstructed from the hologram, retrieving information about the imaged object.

The superposition of object wave and reference wave in the detector plane yields several contributions to the hologram. The complex wave field U created by the interference has the form

$$U = \psi_R + \psi_O, \tag{1.1}$$

where ψ_R and ψ_O denote the reference wave and the object wave, respectively. Since the hologram H actually recorded by a detector is a real-valued quantity, it takes the form of the modulus squared of the complex wave field U,

$$H = |U|^2 = |\psi_R + \psi_O|^2 = \psi_R\psi_R^* + \psi_O\psi_O^* + \psi_R\psi_O^* + \psi_R^*\psi_O, \tag{1.2}$$

where * denotes the complex conjugate.

The reference wave is assumed to be a monochromatic wave of amplitude a_R and phase φ_R, i.e. $\psi_R = a_R e^{i\varphi_R}$. The object wave can analogously be written as $\psi_O = a_O e^{i\varphi_O}$. a_R, a_O, φ_R and φ_O are real-valued. In the following, the letter U will be used to denote complex-valued wave fields, whereas the letter a is used for real-valued amplitudes. Substituting the definitions of ψ_R and ψ_O into eq. (1.2) yields

$$H = |a_R|^2 + |a_O|^2 + a_R e^{i\varphi_R} a_O^* e^{-i\varphi_O} + a_R^* e^{-i\varphi_R} a_O e^{i\varphi_O} \tag{1.3}$$

$$= a_R^2 + a_O^2 + a_R a_O e^{-i(\varphi_O - \varphi_R)} + a_R a_O e^{i(\varphi_O - \varphi_R)}. \tag{1.4}$$

Since the amplitudes are real-valued, i.e. $a_i^* = a_i$ with $i = \{R, O\}$, the modulus squared of the amplitudes has been replaced by the squares of the amplitudes in the second step and the * has been dropped. The first term, a_R^2, is the intensity contribution of the reference wave, which, in the ideal case of uniform illumination, is constant. The second term, a_O^2, is the intensity contribution of the object wave, and the last two terms, which are complex conjugates of one another, encode the interference effects.

An illumination of the hologram with the reference wave thus results in the reconstructed wave field

$$U_{rec} = \psi_R H = a_R e^{i\varphi_R} \left[a_R^2 + a_O^2 + a_R a_O e^{-i(\varphi_O - \varphi_R)} + a_R a_O e^{i(\varphi_O - \varphi_R)} \right] \tag{1.5}$$

$$= a_R^2 e^{i\varphi_R} \left[a_R + \frac{a_O^2}{a_R} + a_O e^{-i(\varphi_O - \varphi_R)} + a_O e^{i(\varphi_O - \varphi_R)} \right]. \tag{1.6}$$

Comparing this to the complex wave field created by the interference of the object wave with the reference wave

$$U = \psi_R + \psi_O = a_R e^{i\varphi_R} + a_O e^{i\varphi_O} = e^{i\varphi_R} \left[a_R + a_O e^{i(\varphi_O - \varphi_R)} \right], \tag{1.7}$$

it is evident that the first and the last term in the brackets of eq. (1.6) are the same as in eq. (1.7) apart from the factor a_R^2, which is constant if the reference wave provides a uniform background. Thus, if eq. (1.6) is to reconstruct U as given in eq. (1.7), it needs to be shown that the other two terms appearing in eq. (1.6) do not significantly contribute to the reconstructed wave field [38]. The second term, which originates from the scattered intensity, is small if the scattered amplitude a_O is much smaller than the
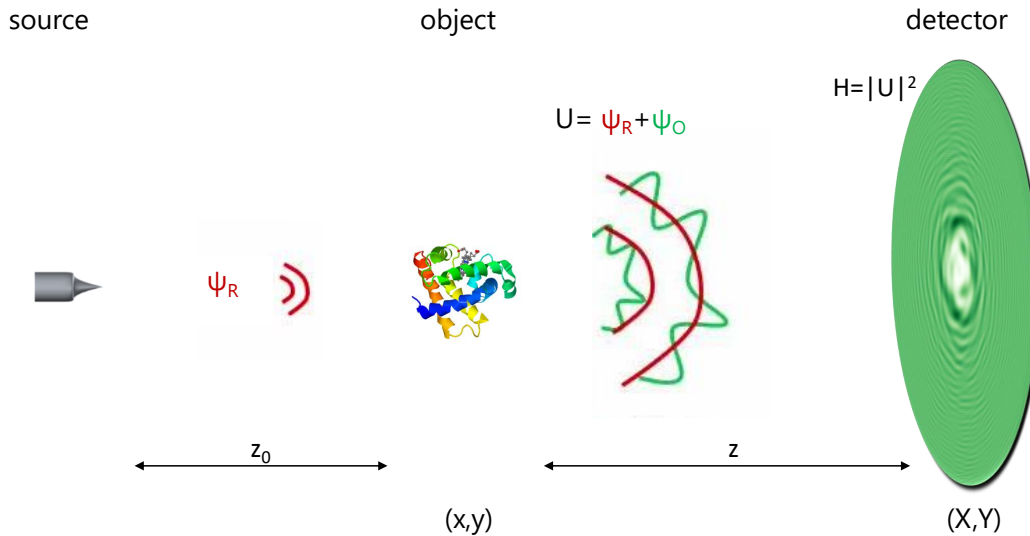
Figure 1.1: **Schematics of in-line holography:** In-line holography requires a source producing the reference wave ψ_R, which subsequently interacts with the object. This leads to a scattered wave ψ_O which interferes with the reference wave to form the wave front $U = \psi_R + \psi_O$ that propagates to the detector. Lowercase letters (x, y) denote the spatial coordinates in the object plane, while uppercase letters (X, Y) represent the coordinates in the detector plane. In in-line holography with a spherical incident wave, the source-to-sample distance, z_0, is in general several orders of magnitude smaller than the sample-to-detector distance, z, which allows for large geometric magnification factors.

amplitude of the reference wave a_R. By using a bright source, this condition can be fulfilled.

The third term in eq. (1.6) is the complex conjugate of the fourth term, i. e. it has the same amplitude, but phases of opposite sign. While the fourth term represents the wave originally scattered by the object (the exit wave), the third term constitutes the so-called twin image, which appears in focus in a plane that is symmetric to the object plane with respect to the source [38, 74] (see Fig. 1.2). The ambiguity resulting in the twin image is due to the lack of absolute phases measured at the detector. While the relative phases are encoded in the interference pattern, they do not allow a unique determination of the sign of the reconstructed object's phase, i. e. both the object and the point-symmetric twin object with phases of different sign would produce the same hologram. Thus, the third term in eq. (1.6) describes an out-of-focus contribution of the twin image to the reconstruction of the exit wave.

To eliminate the contributions by the twin image and hence retrieve the accurate reconstruction of the object, one has to find experimental or numerical schemes to separate the two signals. Experimentally, this is possible by increasing the source-to-sample distance z_0, which in turn increases the separation of object and twin image and thus reduces the out-of-focus contribution. High-magnification imaging in an in-line geometry, however, requires small source-to-sample distances. By changing the geometry to an off-axis set-up [54], the twin-image contribution can easily be separated from the object reconstruction [49, 55], this, however, requires the use of optical elements. A

numerical scheme for eliminating the twin image contribution and thereby recovering the object's phase distribution is presented in Chapter 3.
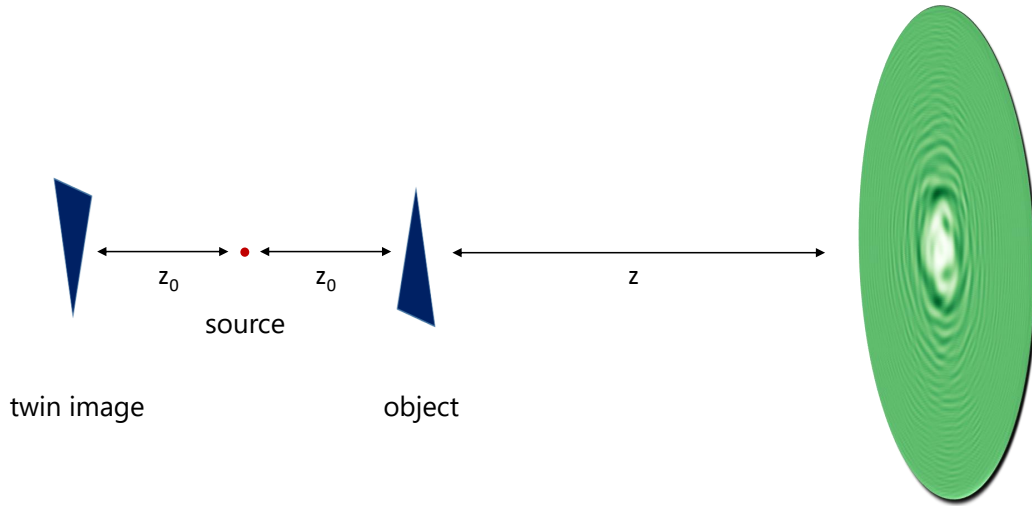
Figure 1.2: **Twin image:** Because the absolute phases at the detector are not known, the reconstruction process yields two images, the object and its twin image with identical amplitudes and phases of opposite sign. The twin image appears in point symmetry to the object with respect to the source.

1.1.2 Mathematical formulation of holography

In the previous section, the general principle underlying holography has been described. In order to formulate an algorithm for reconstruction, the mathematical form of the quantities involved has to be understood, specifically, the relation between the complex wave front U and the properties of the object needs to be elucidated. Since U is formed by diffraction at the object and subsequent wave front interference, the general principles of diffraction are reviewed and subsequently applied to in-line holography.

Diffraction

Diffraction is the result of the interaction of a wave with an object or an aperture which leads to a change of the wave front's propagation direction that allows the wave to spread into regions defined by the geometrical shadow of the object, i.e. regions that would not be accessible to the incident wave. Such phenomena have first been described by Grimaldi in the 17th century, who also first used the term *diffraction* [75, 40]. In 1818, Fresnel [76] was able to explain diffraction effects by combining Huygens' proposal [77] that every point on a wave front can be viewed as the source of a spherical wave, whose envelope forms the wavefront at any later time, with the principle of interference [40]. This is know as the *Huygens-Fresnel principle.*
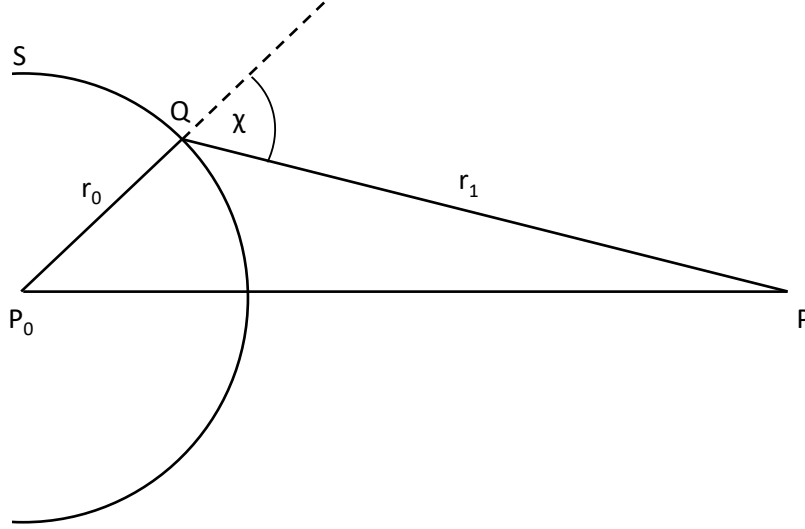
Figure 1.3: **Huygens-Fresnel principle for free-space propagation:** To calculate the complex wave field at a point P, the contributions of all spherical waves emanating from points on the wavefront S (such as Q) have to be integrated. The wavefront S is formed by the spherical wave emitted from the source at P_0.

Mathematically, it can be expressed in the following way [40]: Starting from a point source at P$_0$, which emits a monochromatic spherical wave $U_0(r_0) = \frac{A_0 e^{ikr_0}}{r_0}$ with amplitude A_0 and wave number k, a wave front S with radius r_0 can be constructed (see Fig. 1.3). Following Huygens, each point on the wavefront S can now be considered as the origin of another spherical wave. As an example, the point Q on S is considered, which is the origin of the spherical wave of unit amplitude $U_Q = \frac{e^{ikr_1}}{r_1}$, where r_1 is the distance to the point P at which the complex wave field is to be evaluated (see Fig. 1.3). Thus, the contribution to the complex wave field $U(P)$ at P originating from the wave front element dS at Q is given by

$$\mathrm{d}U(P) = K(\chi)\frac{A_0 e^{ikr_0}}{r_0}\frac{e^{ikr_1}}{r_1}\mathrm{d}S. \tag{1.8}$$

$K(\chi)$ is called the *inclination factor* and encodes the influence of the diffraction angle χ, with K being maximal for $\chi = 0$ and zero for $\chi = \frac{\pi}{2}$, which would correspond to $\overline{QP}$ being tangential to S [40]. Since not only Q, but all points on S contribute to the wave field at P, $U(P)$ takes the form

$$U(P) = \frac{A_0 e^{ikr_0}}{r_0} \iint_S K(\chi)\frac{e^{ikr}}{r}\mathrm{d}S, \tag{1.9}$$

where the integration over S takes into account the different distances r between the points on the wave front and P, as well as the respective diffraction angles χ.

By comparing this result to the experiment, Fresnel concluded that an additional factor $-\frac{i}{\lambda}$ must be included in eq. (1.9) [40], yielding

$$U(P) = -\frac{i}{\lambda} \frac{A_0 e^{ikr_0}}{r_0} \iint_S K(\chi) \frac{e^{ikr}}{r} \mathrm{d}S. \tag{1.10}$$

The origin of this term becomes clear when looking at Kirchhoff's mathematically rigorous formulation of the theory of diffraction [40].

Kirchhoff demonstrated that the Huygens-Fresnel principle is an approximation of an integral theorem, known as the *Kirchhoff integral theorem*. This theorem provides the solution of the homogeneous wave equation

$$(\nabla^2 + k^2)U = 0 \tag{1.11}$$

at a point P as a function of both the solution of the wave equation and its first derivative at all points of a closed surface encompassing P [40]. In the following, U is considered to be the spatial part of a monochromatic scalar wave with wave number k. For two wave functions U and U', Green's second identity holds, i.e.

$$\iiint_V (U\nabla^2 U' - U'\nabla^2 U)\mathrm{d}V = -\iint_S \left(U\frac{\partial U'}{\partial \hat{n}} - U'\frac{\partial U}{\partial \hat{n}}\right)\mathrm{d}S, \tag{1.12}$$

if U and U' have continuous first and second partial derivatives within the integration volume V and on its surface $S = \partial V$ [40]. $\frac{\partial}{\partial \hat{n}}$ designates the derivative with respect to the inward normal of S. If U and U' both satisfy the homogeneous wave equation, the left side of eq. (1.12) is zero, which yields

$$\iint_S \left(U\frac{\partial U'}{\partial \hat{n}} - U'\frac{\partial U}{\partial \hat{n}}\right)\mathrm{d}S = 0. \tag{1.13}$$

In order to derive a solution for U, U' can be set to $U' = \frac{e^{ikr}}{r}$, where r is the distance to a point P within the volume V at which U is to be evaluated. U' has a singularity at P ($r = 0$), hence P cannot be part of the integration domain. This can be taken into account by including a small sphere around P as a second integration surface S_1. This results in

$$\iint_S \left(U\frac{\partial}{\partial \hat{n}}\frac{e^{ikr}}{r} - \frac{e^{ikr}}{r}\frac{\partial U}{\partial \hat{n}}\right)\mathrm{d}S + \iint_{S_1} \left(U\frac{\partial}{\partial \hat{n}}\frac{e^{ikr}}{r} - \frac{e^{ikr}}{r}\frac{\partial U}{\partial \hat{n}}\right)\mathrm{d}S = 0. \tag{1.14}$$

Evaluating the partial derivative of U' in the second term of eq. (1.14) leads to

$$\iint_S \left(U\frac{\partial}{\partial \hat{n}}\frac{e^{ikr}}{r} - \frac{e^{ikr}}{r}\frac{\partial U}{\partial \hat{n}}\right)\mathrm{d}S = -\iint_{S_1} \left(U\frac{e^{ikr}}{r}\left(ik - \frac{1}{r}\right) - \frac{e^{ikr}}{r}\frac{\partial U}{\partial \hat{n}}\right)\mathrm{d}S, \tag{1.15}$$

with $\frac{\partial}{\partial \hat{n}}\frac{e^{ikr}}{r} = \nabla\frac{e^{ikr}}{r}\cdot\hat{n} = \frac{e^{ikr}}{r}(ik - \frac{1}{r})$. The last equality holds because the surface S_1 that $\hat{n}$ is normal to is a sphere, hence in spherical coordinates only the radial derivative remains, i.e. $\nabla\frac{e^{ikr}}{r}\cdot\hat{n} = \frac{\partial}{\partial r}\frac{e^{ikr}}{r}\cdot\hat{r}$.
Since the surface S_1 has been chosen as a sphere, the integral can be rewritten in

spherical coordinates by expressing the surface element $\mathrm{d}S$ in terms of the solid angle $\mathrm{d}\Omega$, $\mathrm{d}S = s^2\mathrm{d}\Omega$. With this, the integral takes the form

$$\iint_S \left(U\frac{\partial}{\partial \hat{n}}\frac{e^{ikr}}{r} - \frac{e^{ikr}}{r}\frac{\partial U}{\partial \hat{n}} \right)\mathrm{d}S = -\iint_\Omega \left(U\frac{e^{iks}}{s}(ik - \frac{1}{s}) - \frac{e^{iks}}{s}\frac{\partial U}{\partial \hat{n}} \right)s^2\mathrm{d}\Omega \qquad (1.16)$$

$$= -\iint_\Omega e^{iks}\left(iksU - U - s\frac{\partial U}{\partial \hat{n}} \right)\mathrm{d}\Omega. \qquad (1.17)$$

In the limit $s \to 0$, the first and the third term of eq. (1.17) vanish, thus the right hand side of the integral yields $4\pi U$. This leads to the final form of the Kirchhoff integral theorem

$$U(P) = \frac{1}{4\pi}\iint_S \left(U\frac{\partial}{\partial \hat{n}}\frac{e^{ikr}}{r} - \frac{e^{ikr}}{r}\frac{\partial U}{\partial \hat{n}} \right)\mathrm{d}S. \qquad (1.18)$$

From the Kirchhoff integral theorem (eq. (1.18)), an expression for $U(P)$ can be derived that closely resembles the integral form of the Huygens-Fresnel principle (eq. (1.9)), which is useful for calculating the diffraction resulting from the interaction with an object and can hence be applied in holography theory. This form of the integral theorem is known as the *Fresnel-Kirchhoff diffraction formula* [40, 38].

To derive the Fresnel-Kirchhoff diffraction formula, one can consider the situation depicted in Fig. 1.4: a monochromatic, spherical wave $U(r_0) = a\frac{e^{ikr_0}}{r_0}$ emitted from a source at P_0 that propagates towards a barrier with an aperture. The value of the complex wave field is evaluated at a point P behind the barrier. The aperture is assumed to be large in comparison to the wavelength $\lambda = \frac{2\pi}{k}$, but small in comparison to the distances r_0 and r of P_0 and P to the barrier. $U(P)$ can be calculated by evaluating the Kirchhoff integral theorem on a closed surface formed by the aperture, the barrier and a sphere of radius R centred around P that intersects the barrier and thereby closes the surface. The integral can hence be split into three parts which are integrated over the three different contributions to the closed surface: the aperture (S_1), the barrier (S_2), and the section of the sphere (S_3) (see Fig. 1.4). This yields:

$$U(P) = \frac{1}{4\pi}\iint_{S_1} + \iint_{S_2} + \iint_{S_3} \left(U\frac{\partial}{\partial \hat{n}}\frac{e^{ikr}}{r} - \frac{e^{ikr}}{r}\frac{\partial U}{\partial \hat{n}} \right)\mathrm{d}S. \qquad (1.19)$$

For an exact solution of the Kirchhoff integral theorem, the values of both U and $\frac{\partial U}{\partial \hat{n}}$ would have to be known for all three parts of the integration surface. This is in general not the case, however, reasonable approximations can be made.

For the segment S_1, it can be assumed that U and $\frac{\partial U}{\partial \hat{n}}$ are the same as in the case of free propagation without an object present:

$$U_{S_1} = a\frac{e^{ikr_0}}{r_0} \quad \text{and} \quad \frac{\partial U_{S_1}}{\partial \hat{n}} = a\frac{e^{ikr_0}}{r_0}\left(ik - \frac{1}{r_0} \right)\cos(\theta_0), \qquad (1.20)$$

where the factor $\cos(\theta_0)$ comes in since $\frac{\partial U_{S_1}}{\partial \hat{n}} = \nabla U_{S_1} \cdot \hat{n}$.

In the segment S_2, at the barrier, both quantities can be assumed to be zero, i.e.

$$U_{S_2} = 0 \quad \text{and} \quad \frac{\partial U_{S_2}}{\partial \hat{n}} = 0. \qquad (1.21)$$
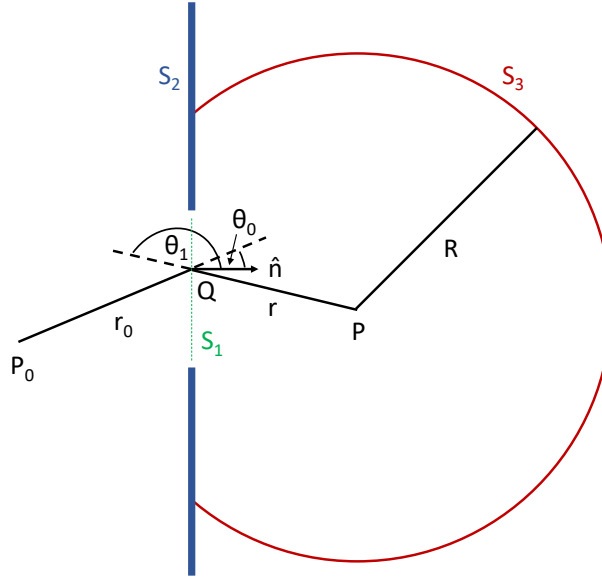
Figure 1.4: **Derivation of the Fresnel-Kirchhoff diffraction formula:** The Kirchhoff diffraction theorem is evaluated over the combined surface $S_1 + S_2 + S_3$, where S_1 is the aperture (green), S_2 is the barrier (blue) and S_3 is a sphere of radius R centred on P (red). P_0 is the source of a spherical wave, $\hat{n}$ is the normal to the barrier, θ_0 and θ_1 are the angles of r_0 and r form with the normal, respectively.

Equations (1.20) and (1.21) comprise *Kirchhoff's boundary conditions* [40].
The remaining contribution from S_3 becomes negligible if R is chosen so large that at the time the wave field at P is evaluated, no contribution from S_3 could have reached P. This implies that the wave field comes into existence at a certain point in time t_0 and then takes time t to reach a radial extension of R. Since strictly monochromatic waves exist at all times, the argument only holds for not strictly monochromatic waves, hence the resulting formula is only an approximation of the integral theorem.
With

$$\frac{\partial}{\partial \hat{n}} \frac{e^{ikr}}{r} = \frac{e^{ikr}}{r} \left(ik - \frac{1}{r} \right) \cos(\theta_1), \tag{1.22}$$

the Fresnel-Kirchhoff diffraction formula takes the form

$$U(P) = \frac{1}{4\pi} \iint_{S_1} \left[a\frac{e^{ikr_0}}{r_0} \frac{e^{ikr}}{r} \left(ik - \frac{1}{r} \right) \cos(\theta_1) - \frac{e^{ikr}}{r} a\frac{e^{ikr_0}}{r_0} \left(ik - \frac{1}{r_0} \right) \cos(\theta_0) \right] \mathrm{d}S. \tag{1.23}$$

$$= -\frac{i}{2\lambda} \iint_{S_1} \left[a\frac{e^{ik(r_0+r)}}{r_0 r} \left(\cos(\theta_0) - \cos(\theta_1) \right) \right] \mathrm{d}S. \tag{1.24}$$

In the last step, the assumption that the wavelength is much smaller than the distances r_0 and r, $\lambda \ll r_0$ and $\lambda \ll r$, which, with $k = \frac{2\pi}{\lambda}$, implies $k \gg \frac{1}{r_0}$ and $k \gg \frac{1}{r}$, has been used to drop the terms involving $\frac{1}{r_0}$ and $\frac{1}{r}$.

15

Comparing to eq. (1.9), the prefactor has been mathematically justified, and an expression for $K(\chi)$ has been found in terms of the angles θ_0 and θ_1.

Application to holography

The Fresnel-Kirchhoff diffraction formula can be directly applied in the context of in-line holographic imaging [38, 73] as depicted in Fig. 1.1. While in general, the objects imaged by holography are three-dimensional, the discussion in this section focuses on two-dimensional objects, corresponding to a projection of a three-dimensional object into a plane, for enhanced mathematical clarity. Such an object can be described by a complex-valued transmission function $t(x, y)$, where x and y are the coordinates in the object plane. For the conditions utilized in the derivation of the Fresnel-Kirchhoff diffraction formula to hold, t should not vary noticeably on length scales smaller than the wave length of the incident radiation [38]. Assuming a spherical wave as incident illumination, the application of the Fresnel-Kirchhoff diffraction formula yields the wave field at a point $P = (X, Y)$ in the detector plane:

$$U(P) = -\frac{i}{2\lambda} \iint \left[t(x, y) \frac{e^{ik(r_0+r_1)}}{r_0 r_1} (\cos(\theta_0) - \cos(\theta_1)) \right] \mathrm{d}x\mathrm{d}y, \qquad (1.25)$$

where r_0 is the source-to-sample distance and r_1 the sample-to-detector distance and the amplitude a in eq. (1.24) has been replaced by the transmission function $t(x, y)$. The integration is carried out over the coordinates of the object plane. For small opening angles of the incident beam, which is the case in most experimental in-line holography set-ups, $\cos(\theta_0) = -\cos(\theta_1) \approx 1$, hence

$$U(X, Y) = -\frac{i}{\lambda} \iint \left[t(x, y) \frac{e^{ik(r_0+r_1)}}{r_0 r_1} \right] \mathrm{d}x\mathrm{d}y. \qquad (1.26)$$

This integral can also be interpreted as the exit wave $U_{exit}(x, y) = U_{incident}(x, y)t(x, y)$ being propagated to the detector by a propagator function of the form [73]

$$\frac{\exp\left(ik|\mathbf{r} - \mathbf{R}| \right)}{|\mathbf{r} - \mathbf{R}|}, \qquad (1.27)$$

where $|\mathbf{r} - \mathbf{R}|$ denotes the distance between a point $\mathbf{r} = (x, y, z)$ in the object plane and a point $\mathbf{R} = (X, Y, Z)$ in the detector plane.
Representing the coordinates in that way, the complex wave field U_{det} at a point (X, Y) in the detector plane for an incident spherical wave $U_{incident} = \frac{e^{ikr}}{r}$ can be evaluated as

$$U_{det}(X, Y) = -\frac{i}{\lambda} \iint \left[t(x, y) \frac{e^{ikr}}{r} \frac{e^{(ik|\mathbf{r}-\mathbf{R}|)}}{|\mathbf{r} - \mathbf{R}|} \right] \mathrm{d}x\mathrm{d}y. \qquad (1.28)$$

For small opening angles, the paraxial approximation is valid, hence r and $|\mathbf{r} - \mathbf{R}|$, which, when calculated exactly takes the form $|\mathbf{r} - \mathbf{R}| = \sqrt{(x - X)^2 + (y - Y)^2 + (z - Z)^2}$, can

be approximated by the expressions [38, 73]

$$r \approx z + \frac{x^2 + y^2}{2z} \tag{1.29}$$

$$|\mathbf{r} - \mathbf{R}| \approx Z + \frac{(x - X)^2 + (y - Y)^2}{2Z}, \tag{1.30}$$

which can be derived by expanding the square root. In eq. (1.30), the fact that $z \ll Z$ has additionally been taken into account, a condition necessary for magnifying the image in the lens-free in-line set-up. The use of the paraxial approximation significantly simplifies the evaluation of the diffraction integral and is common in many applications of in-line holography [38, 73, 78, 79].

Inserting those expressions into the arguments of the exponentials in eq. (1.28) and approximating the denominators by the first-order terms of the respective expansions results in

$$U_{det} = -\frac{i}{\lambda} \iint \left[t(x,y) \frac{\exp\left(ik\left(z + \frac{x^2+y^2}{2z}\right)\right)}{z} \frac{\exp\left(ik\left(Z + \frac{(x-X)^2+(y-Y)^2}{2Z}\right)\right)}{Z} \right] \mathrm{d}x\mathrm{d}y \tag{1.31}$$

$$= -\frac{i}{\lambda z Z} e^{\left(\frac{2\pi i}{\lambda}(Z+z)\right)} \iint t(x,y) e^{\left(\frac{i\pi}{\lambda z}(x^2+y^2)\right)} e^{\left(\frac{i\pi}{\lambda Z}((x-X)^2+(y-Y)^2)\right)} \mathrm{d}x\mathrm{d}y \tag{1.32}$$

$$= -\frac{i}{\lambda z Z} e^{\left(\frac{2\pi i}{\lambda}(Z+z)\right)} e^{\left(\frac{i\pi}{\lambda Z}(X^2+Y^2)\right)} \iint t(x,y) e^{\left(\frac{i\pi}{\lambda z}(x^2+y^2)\right)} e^{\left(\frac{i\pi}{\lambda Z}(x^2+y^2)\right)} e^{\left(-\frac{2\pi i}{\lambda Z}(xX+yY)\right)} \mathrm{d}x\mathrm{d}y. \tag{1.33}$$

Since $z \ll Z$ holds, the contribution of the term $\exp\left(\frac{i\pi}{\lambda Z}(x^2+y^2)\right)$ is much smaller than that of $\exp\left(\frac{i\pi}{\lambda z}(x^2+y^2)\right)$, hence it can be neglected, yielding

$$U_{det}(X,Y) = -\frac{i}{\lambda z Z} e^{\left(\frac{2\pi i}{\lambda}(Z+z)\right)} e^{\left(\frac{i\pi}{\lambda Z}(X^2+Y^2)\right)} \iint t(x,y) e^{\left(\frac{i\pi}{\lambda z}(x^2+y^2)\right)} e^{\left(-\frac{2\pi i}{\lambda Z}(xX+yY)\right)} \mathrm{d}x\mathrm{d}y. \tag{1.34}$$

This equation takes the form of a Fourier transform as defined in eq. (1.44).

While the complex wave field at the detector can in principle be evaluated this way, it is useful to bring this integral into the form of a convolution since this facilitates its numerical evaluation by allowing the use of the convolution theorem.

The convolution $f * g$ of two functions f and g is defined as

$$f * g(t) = \int_{-\infty}^{\infty} f(\tau)g(t-\tau)\mathrm{d}\tau. \tag{1.35}$$

Eq. (1.34) can be rewritten as

$$U_{det}(X,Y) = -\frac{i}{\lambda z Z} e^{\left(\frac{2\pi i}{\lambda}(Z+z)\right)} e^{\left(\frac{i\pi}{\lambda Z}(X^2+Y^2)(1-\frac{z}{Z})\right)} \iint t(x,y) e^{\left(\frac{i\pi}{\lambda z}((x-X\frac{z}{Z})^2+(y-Y\frac{z}{Z})^2)\right)} \mathrm{d}x\mathrm{d}y. \tag{1.36}$$

17

Setting $X' = X\frac{z}{Z}$ and $Y' = Y\frac{z}{Z}$, the integral takes the form of a convolution:

$$U_{det}(X,Y) \propto -\frac{i}{\lambda z} \iint t(x,y) e^{\left(\frac{i\pi}{\lambda z}((x-X')^2+(y-Y')^2)\right)} \mathrm{d}x\mathrm{d}y \tag{1.37}$$

$$= \iint t(x,y)s(x-X',y-Y')\mathrm{d}x\mathrm{d}y \tag{1.38}$$

$$= (t*s)(X',Y'), \tag{1.39}$$

with

$$s(x,y) = -\frac{i}{\lambda z} e^{\left(\frac{i\pi}{\lambda z}(x^2+y^2)\right)}. \tag{1.40}$$

Ignoring the constant prefactors, the hologram H generated from a transmission function t thus takes the form

$$H(X,Y) = |U_{det}(X,Y)|^2 = |(t*s)(X,Y)|^2. \tag{1.41}$$

Equations (1.28) - (1.41) describe the process of hologram generation. Hologram reconstruction can be treated in the same way, by replacing $t(x,y)$ by $H(X,Y)$ in eq. (1.39), changing the propagation direction by reversing the sign of the argument of the propagator function s, and integrating over the detector plane instead of the object plane [73]. This yields the reconstructed exit wave

$$U_{exit}(x,y) \approx \frac{i}{\lambda z} \iint H(X,Y) e^{\left(-\frac{i\pi}{\lambda z}(x^2+y^2)\right)} \mathrm{d}X\mathrm{d}Y = (H*s^*)(x,y), \tag{1.42}$$

where $s^*(x,y) = \frac{i}{\lambda z} \exp\left(-\frac{i\pi}{\lambda z}(x^2+y^2)\right)$ denotes the complex conjugate of the propagator function s (eq. (1.40)).

The main difference between hologram reconstruction and hologram simulation is that the transmission function t is in general complex-valued, hence full knowledge of t will yield an exact value of U_{det}, while the hologram H is always a real-valued. Thus, the exit wave calculated by eq. (1.42) is not an exact reconstruction of the transmission function since the absolute phases in the detector plane are unknown. Eq. (1.42), however, works well in many situations, approaches overcoming this ambiguity will be discussed in Chapter 3. For the numerical implementation of hologram simulation and reconstruction, the convolution form is advantageous since it allows the integral to be expressed in terms of two Fourier transforms, which are very easy to evaluate numerically via the convolution theorem.

The convolution theorem states that the Fourier transform of a convolution of two functions is equal to the product of the Fourier transforms of the individual functions:

$$\mathcal{F}(f*g) = \mathcal{F}(f)\mathcal{F}(g), \tag{1.43}$$

where $\mathcal{F}(f)$ denotes the Fourier transform of a function f, which is defined as

$$\mathcal{F}(f)(p) = \int_{-\infty}^{\infty} f(x)e^{-2\pi ipx}\mathrm{d}x. \tag{1.44}$$

The theorem can be derived by inserting the definition (1.44) into the left hand side of eq. (1.35):

$$\mathcal{F}(f * g) = \int_{-\infty}^{\infty} e^{-2\pi i p t} \left[\int_{-\infty}^{\infty} f(\tau) g(t - \tau) \mathrm{d}\tau \right] \mathrm{d}t. \tag{1.45}$$

Since the integration limits are infinite, setting $x = t - \tau$ does not change them, hence this yields

$$\mathcal{F}(f * g) = \int_{-\infty}^{\infty} e^{-2\pi i p x} e^{-2\pi i p \tau} \int_{-\infty}^{\infty} f(\tau) g(x) \mathrm{d}\tau \mathrm{d}x \tag{1.46}$$

$$= \int_{-\infty}^{\infty} f(\tau) e^{-2\pi i p \tau} \mathrm{d}\tau \int_{-\infty}^{\infty} g(x) e^{-2\pi i p x} \mathrm{d}x \tag{1.47}$$

$$= \mathcal{F}(f) \mathcal{F}(g), \tag{1.48}$$

which proves the theorem.

With the help of the convolution theorem, a convolution of two functions can be expressed as

$$f * g = \mathcal{F}^{-1}(\mathcal{F}(f)\mathcal{F}(g)), \tag{1.49}$$

where $\mathcal{F}^{-1}$ denotes the inverse Fourier transform, i. e.

$$f(x) = \mathcal{F}^{-1}(\mathcal{F}(f)(p)) = \int_{-\infty}^{\infty} \mathcal{F}(f)(p) e^{2\pi i p x} \mathrm{d}p. \tag{1.50}$$

Hence U_{det} and U_{exit} can be written as

$$U_{det} = \mathcal{F}^{-1}(\mathcal{F}(t)\mathcal{F}(s)) \tag{1.51}$$

$$U_{exit} = \mathcal{F}^{-1}(\mathcal{F}(H)\mathcal{F}(s^*)) \tag{1.52}$$

in terms of the transmission function t, the hologram H and the propagator function s and its complex conjugate s^*.

Equations (1.51) and (1.52) can directly be used to develop an algorithm for hologram simulation and reconstruction (see section 1.1.3). Thus, as a last step before turning to the numerical implementation, $\mathcal{F}(s)$ needs to be calculated.

Since s (eq. (1.40)) has the form of a Gaussian function, its Fourier transform is straightforward to compute by completing the square in the argument of the exponential:

$$\mathcal{F}(s)(u, v) = -\frac{i}{\lambda z} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{\frac{i\pi}{\lambda z}(x^2 + y^2)} e^{-2\pi i (ux + vy)} \mathrm{d}x \mathrm{d}y \tag{1.53}$$

$$= -\frac{i}{\lambda z} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{\frac{i\pi}{\lambda z}\left((x - \lambda z u)^2 + (y - \lambda z v)^2\right)} e^{-i\pi \lambda z (u^2 + v^2)} \mathrm{d}x \mathrm{d}y \tag{1.54}$$

$$= -\frac{i}{\lambda z} e^{-i\pi \lambda z (u^2 + v^2)} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{\frac{i\pi}{\lambda z}\left(x'^2 + y'^2\right)} \mathrm{d}x' \mathrm{d}y', \tag{1.55}$$

where u and v are the coordinates in Fourier space. In the last step, a change of variables with $x' = x - \lambda z u$ and $y' = y - \lambda z v$ has been implemented. The remaining integral term is a product of two identical Gaussian integrals, each of which has the solution

$$\int_{-\infty}^{\infty} e^{-ax^2} \mathrm{d}x = \sqrt{\frac{\pi}{a}}, \tag{1.56}$$

with $a = -\frac{i\pi}{\lambda z}$. Inserting this, $\mathcal{F}(s)$ becomes

$$\mathcal{F}(s) = -\frac{i}{\lambda z} e^{-i\pi\lambda z(u^2+v^2)} \sqrt{\frac{\pi}{-\frac{i\pi}{\lambda z}}} \sqrt{\frac{\pi}{-\frac{i\pi}{\lambda z}}} = -\frac{i}{\lambda z} e^{-i\pi\lambda z(u^2+v^2)} \frac{\pi}{-\frac{i\pi}{\lambda z}} = e^{-i\pi\lambda z(u^2+v^2)}. \tag{1.57}$$

Although the convolution approach has so far been discussed in the paraxial approximation, it is not restricted to the paraxially approximated integral, since the general form of the diffraction integral used to calculate the exit wave can also be written in the form of a convolution, i. e.

$$U(x,y) = -\frac{i}{\lambda} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} H(X,Y) R(X,Y) \frac{e^{-ik\rho}}{\rho} \mathrm{d}X \mathrm{d}Y \tag{1.58}$$

with $\rho = \sqrt{(X-x)^2 + (Y-y)^2 + (Z-z)^2}$, which yields

$$U(x,y) = (H \cdot R) * S = \mathcal{F}^{-1}(\mathcal{F}(H \cdot R) \cdot \mathcal{F}(S)) \tag{1.59}$$

with the propagation function

$$S(X,Y) = \frac{e^{-ik\sqrt{X^2+Y^2+(Z-z)^2}}}{\sqrt{X^2+Y^2+(Z-z)^2}}. \tag{1.60}$$

The Fourier transform of the propagator function, however, is much harder to evaluate in the general case.

Applicability of a wave-optical description to electron holography

In principle, in-line holography can be performed with different types of radiation, most commonly, photons and electrons are used. The mathematical description of holographic imaging provided in this chapter has been based on wave-optical principles, specifically on Kirchhoff's diffraction theorem, which in turn is derived from the Helmholtz equation. The Helmholtz equation is a form of the time-independent homogeneous wave equation and can be expressed as follows:

$$\left(\Delta + k^2\right) U(\mathbf{r}) = 0, \tag{1.61}$$

where Δ is the Laplace operator, k is the wave number, and $U(\mathbf{r})$ is the spatial part of a wave field.

Since the wave equation can directly be derived from Maxwell's equations, the wave equation is an appropriate description of the propagation of electromagnetic waves and can thus be applied in a wave-optical description of photon scattering.

The experimentally acquired protein holograms discussed in this thesis, however, are generated using a beam of low-energy electrons. Since electron wave functions are governed by the Schrödinger equation, the validity of applying the Helmholtz equation-based wave-optical description of holography outlined in the previous sections needs to be discussed.

Like the wave equation, the Schrödinger equation can be split into a time-dependent and a time-independent part using a separation ansatz for the wave function. The resulting time-independent Schrödinger equation has the form

$$\hat{H}\psi(\mathbf{r}) = E\psi(\mathbf{r}), \tag{1.62}$$

with the Hamiltonian $\hat{H} = -\frac{\hbar^2}{2m}\Delta + V(\mathbf{r})$, where m is the electron mass, E denotes the energy of the system, and $V(\mathbf{r})$ describes a time-independent potential through which the electron is moving.

Inserting the expression for the Hamiltonian, the time-independent Schrödinger equation can be rewritten as

$$\left(-\frac{\hbar^2}{2m}\Delta + V(\mathbf{r})\right)\psi(\mathbf{r}) = E\psi(\mathbf{r}) \tag{1.63}$$

$$\Leftrightarrow -\frac{\hbar^2}{2m}\Delta\psi(\mathbf{r}) + (V(\mathbf{r}) - E)\,\psi(\mathbf{r}) = 0 \tag{1.64}$$

$$\Leftrightarrow \quad \left(\Delta + k^2\right)\psi(\mathbf{r}) = 0, \quad k(\mathbf{r}) = \sqrt{\frac{2m}{\hbar^2}(E - V(\mathbf{r}))}. \tag{1.65}$$

With this, the time-independent Schrödinger equation takes the same form as the Helmholtz equation. Thus, in scattering situations with time-independent potentials, as can be assumed to be the case for protein potentials in our in-line holography set-up since radiation damage and diffusion effects are negligible, electron scattering can be approximately described by a wave-optical formalism as presented in this chapter. To incorporate effects specific to electron scattering at the energies employed in our experimental scheme, different approaches need to be pursued, a first attempt towards this is discussed in section 3.3.2.

1.1.3 Numerical implementation

Gabor's original idea for reconstruction was an experimental scheme involving illumination by the reference wave. In many situations, however, the use of a numerical reconstruction algorithm is advantageous. While a numerical implementation of the general integral form of the reconstructed exit wave (eq. (1.28)) can in principle be found, the expression derived via the convolution theorem (eq. (1.52)) is useful from a numerical point of view, since instead of an integral, it employs two Fourier transforms. The evaluation of integrals can be numerically costly, but Fourier transforms can be implemented via fast Fourier transform (FFT) algorithms which reduce the necessary computation time considerably; a direct evaluation of the integral is approximately 600 times slower than a convolution approach (see section 3.3.2). Additionally, reconstruction schemes utilizing two Fourier transforms instead of one have the benefit that hologram and object are sampled with a similar amount of pixels [73].

Based on eq. (1.52), an algorithm for hologram simulation and reconstruction can be formulated. In the case of hologram simulation, the following steps need to be implemented [73]:

1. Calculation of the Fourier transform of the transmission function t. As discussed below, it is necessary to centre the Fourier transform on the optical axis.

2. Calculation of the Fourier transform of s, $\mathcal{F}(s) = e^{-i\pi\lambda z(u^2+v^2)}$, where u and v are the coordinates in Fourier space.

3. Multiplication of $\mathcal{F}(t)$ and $\mathcal{F}(s)$ and calculation of the inverse Fourier transform of the product. Taking the modulus squared of the result yields the real-valued hologram.

The corresponding procedure for hologram reconstruction is:

1. Calculation of the Fourier transform of the hologram H.

2. Calculation of the Fourier transform of s^*, $\mathcal{F}(s^*) = e^{i\pi\lambda z(u^2+v^2)}$.

3. Multiplication of $\mathcal{F}(H)$ and $\mathcal{F}(s^*)$ and calculation of the inverse Fourier transform of the product.

In order to translate this into code, as a first step, a routine for the discrete Fourier transforms needs to be implemented. The images of the experimentally obtained holograms are taken as digital photographs with a number of pixels N^2, hence a discretized version of the continuous Fourier transform discussed above has to be employed. Since the imaging system is centred on the optical axis, the Fourier transform has to be centred on the optical axis as well to obtain correctly transformed images. Mathematically, this can be expressed as follows [73]:
Fast Fourier transforms are based on discrete Fourier transforms, which take the form

$$F(p,q) = \text{FT}(f(m,n)) = \sum_{m,n=1}^{N} f(m,n)e^{-\frac{2\pi i}{N}(mp+nq)}, \tag{1.66}$$

where (p,q) denote the pixel numbers in the Fourier domain, (m,n) denote the pixel numbers in the spatial domain and FT denotes the discrete Fourier transform. To centre the Fourier transform, coordinate transformations have to be made both in the spatial domain (coordinates (x,y), pixel size Δ_O) and in the Fourier domain (coordinates (u,v), pixel size Δ_F) to align the centre of the distribution $f(x,y)$ with the centre of the (x,y)-coordinate system:

$$x = \left(m - \frac{N}{2}\right)\Delta_O, \quad m = 1,...,N \tag{1.67}$$

$$y = \left(n - \frac{N}{2}\right)\Delta_O, \quad n = 1,...,N \tag{1.68}$$

$$u = \left(p - \frac{N}{2}\right)\Delta_F, \quad p = 1,...,N \tag{1.69}$$

$$v = \left(q - \frac{N}{2}\right)\Delta_F, \quad q = 1,...,N, \tag{1.70}$$

$$\tag{1.71}$$

where the pixel sizes in the respective domains are defined via the sizes of the corresponding image areas, S_O and S_F, as

$$\Delta_O = \frac{S_O}{N} \quad \text{and} \quad \Delta_F = \frac{S_F}{N}. \tag{1.72}$$

The centred phase term for eq. (1.66) thus takes the form

$$\frac{2\pi}{N}(xu + yv) = \frac{2\pi}{N}\left(mp + nq - \frac{N}{2}(m + p + n + q) + \frac{N^2}{2}\right). \tag{1.73}$$

Inserting this into eq. (1.66) while dropping the last, constant term, yields the centred discrete Fourier transform

$$F(p, q) = e^{i\pi(p+q)} \sum_{m,n=1}^{N} f(m, n)e^{-\frac{2\pi i}{N}(mp+nq)}e^{i\pi(m+n)} \tag{1.74}$$

$$= e^{i\pi(p+q)}\mathrm{FT}\left(f(m, n)e^{i\pi(m+n)}\right), \tag{1.75}$$

and, analogously, its inverse

$$f(m, n) = e^{-i\pi(m+n)}\mathrm{FT}^{-1}\left(F(p, q)e^{-i\pi(p+q)}\right). \tag{1.76}$$

Using the programming language Python, this can be implemented into code in two ways, either functions following the form of equations (1.75) and (1.76), or the NumPy function `fftshift` can be used.

The former case can be written as functions of an input hologram `holin` that represents $f(m, n)$:

```
1  def FT2D(holin):                    #2D centred Fourier transform
2      (Nx,Ny)=holin.shape[:2]
3      sum_ij=[i+j for i in range(Nx) for j in range(Ny)]
4      sum_ij = np.array(sum_ij, dtype=complex)
5      f1=np.empty(Nx*Ny, dtype=complex)
6      f1=np.exp(1j*np.pi*sum_ij)
7      f1=np.reshape(f1, (Nx, Ny))
8      FT=np.fft.fft2(f1*holin)
9      return f1*FT
10
11
12 def IFT2D(holin):                   #2D centred inverse Fourier transform
13     (Nx,Ny)=holin.shape[:2]
14     sum_ij=[i+j for i in range(Nx) for j in range(Ny)]
15     sum_ij = np.array(sum_ij, dtype=complex)
16     f2=np.empty((Nx,Ny), dtype=complex)
17     f2=np.exp(-1j*np.pi*sum_ij)
18     f2=np.reshape(f2, (Nx, Ny))
19     FT=np.fft.ifft2(f2*holin)
20     return f2*FT
```

In line 2, the shape of the input array, which determines the shape of the Fourier transformed array, is read out. In line 3 and 4, an array with entries $(i + j)$ for all pixel numbers (i, j) is created. To improve the performance of the algorithm, i.e. to decrease computation time, the for-loop in line 3 has been written in matrix form. Line 4 initialises an empty array of the same shape as the input array. This array is filled

in line 6, corresponding to an array with entries $e^{i\pi(i+j)}$ for all (i,j) ((i,j) represents (m,n) in the code). The last two lines of code correspond to eq. (1.75). First, the Fourier transform of $f(m,n)e^{i\pi(m+n)}$ is calculated using the NumPy function `fft2` in line 8. Since the arrays in the spatial and the Fourier domain have the same shape, `sum_ij` represents both $(m+n)$ and $(p+q)$. Thus, in line 9 the result is multiplied with the array with the $e^{i\pi(i+j)}$ entries to yield the final result.

Alternatively, the centred Fourier transform can be based on the NumPy functions `fft2` and `fftshift` and the corresponding inverse functions:

```
1  def FFT2 (holin):
2      F = ifftshift(fft2(fftshift(holin)))
3      return F
```

```
1  def iFFT2 (holin):
2      F = fftshift(ifft2(ifftshift(holin)))
3      return F
```

For the second step, the calculation of the Fourier transform of s, a propagator function has to be defined. Since an analytical expression for $\mathcal{F}(s)$ is known, this can be directly implemented.

```
1  def Propagator(ar, distance, holin, wavelength):
2      delta0 = 1/ar
3      (Nx,Ny)=holin.shape[:2]
4      uv=[(ii-Nx/2)**2+(jj-Ny/2)**2 for ii in range(Nx) for jj in range(Ny)]
5      uv=np.array(uv, dtype=complex)
6      p=np.empty((Nx,Ny), dtype=complex)
7      p= np.exp(1j*np.pi*wavelength*distance*(delta0**2)*uv)
8      p=np.reshape(p, (Nx, Ny))
9      return p
```

The propagator function takes four arguments: the physical object size (`ar`), corresponding to the physical side length of the input image, the source-to-sample distance (`distance`), the input hologram or object (`holin`) and the wavelength (`wavelength`). `delta0`, defined in line 2, is the pixel size in Fourier space. As in the definition of the centred Fourier transform above, line 3 extracts the shape of the input array, which is then used to set the shape of the output array (line 6). Line 4 and 5 calculate an array with entries $u^2 + v^2$, where u and v are centred coordinates. In line 7, the expression $e^{i\pi\lambda z(u^2+v^2)}$ is calculated. Since u and v are Fourier space coordinates, they have to be multiplied by the Fourier space pixel size `delta0`.

The above propagator function assumes a square input, i.e. $N_x = N_y$. Since the detector used in the experimental set-up presented here is circular, this is in general the case for experimentally obtained holograms. To allow for non-square input, the propagator has to be slightly modified:

```
1   def Propagatornonsquare(ar0, ar1, distance, holin, wavelength):
2       delta0 = 1/ar0
3       delta1 = 1/ar1
4       (Nx,Ny)=holin.shape[:2]
5       uv=[((ii-Nx/2)*delta0)**2+((jj-Ny/2)*delta1)**2 for ii in range(Nx) for jj in
             range(Ny)]
6       uv=np.array(uv, dtype=complex)
7       p=np.empty((Nx,Ny), dtype=complex)
8       p= np.exp(1j*np.pi*wavelength*distance*uv)
9       p=np.reshape(p, (Nx, Ny))
10      return p
```

By taking into account the different physical sizes of the object's dimensions (`ar0` and `ar1`), one obtains different Fourier pixel sizes for the different dimensions that need to be multiplied with the respective pixel numbers (line 5).

```
1  U_t=IFT2D(FT2D(t)*np.conjugate(Propagator(area, z0, t, Lambda)))
2  hologram_t=abs(U_t)**2
```

To account for the change in propagation direction, the complex conjugate of the propagator function defined above is employed. Since the first line yields a complex-valued array, the modulus squared has to be taken in the second line to create the hologram.

The reconstruction of the hologram takes a similar form with the propagator as defined above:

```
1  U_rec=IFT2D(FT2D(hologram_t)*Propagator(area, z0, hologram_t, Lambda))
2  arec=abs(U_rec)
```

In the second line, the absolute value of the complex-valued array U_{rec} has been taken to extract the reconstructed amplitude `arec`.

One of the input parameters of the above code is the source-to-sample distance `z0`. In the case of simulated examples, the in-focus source-to-sample distance is always known, since it is used in the simulation step. For experimentally obtained holograms, however, the in-focus distance z_0 has to be reconstructed alongside the object. Since the physical size of the hologram h, which is given by the size of the detector, and the sample-to-detector distance z are known, `z0` can be determined via the focus plane. This can be implemented by reconstructing the hologram for a range of source-to-sample distances from `z0_start` to `z0_end` with a distance `z0_step` between the reconstruction planes:

```
1  S = round((z0_end - z0_start)/z0_step)
2  for x in range(0,S):
3      z0=z0_start + x*z0_step
4      area = z0*h/(z+z0)
5      arec=abs(IFT2D(FT2D(hologram_t)*Propagator(area, z0, hologram_t, Lambda)))
```

The object size, which is an input parameter of the propagator function, can be calculated via the magnification for each source-to-sample distance (line 4).

The resulting stack of reconstruction planes can then be inspected by eye to find the in-focus image and the corresponding source-to-sample distance `z0`. In in-focus images, the reconstructed objects appear with sharp edges. Alternatively, focus-finding algorithms can be employed to determine the focus [80, 81]. The need to reconstruct a stack of focus planes for each image significantly increases the number of images to be processed in the reconstruction of an experimental data set. This warrants the use of powerful computing tools such as a parallelized architecture like a cluster for the reconstruction of experimental data since this allows the simultaneous reconstruction of a large number of images and thus considerably decreases the time required to obtain the desired reconstructions.

The parameters `z0`, `area`, and `Lambda` are interrelated via the propagator function, hence there are value regimes for which those parameters are optimized, which results in holograms with high-order interference fringes. Since the wavelength is fixed by the experimentally measured energy value, this optimization mostly concerns the source-to-sample distance. In order to obtain a high number of interference fringes in the hologram, which encode the desired information, the source-to-sample distance has to be chosen such that empty spaces around an object are optimally filled with fringes [74]. From this condition, suitable values for `z0` can be derived depending on the area of empty space, the maximal attainable resolution and the wavelength [74].

Another consideration regarding parameter optimization concerns sampling, i.e. the number of pixels in a hologram of a given physical size. For a correct sampling of the propagator function $S^*(u, v)$ with $N \times N$ pixels, the argument of the exponential $\exp\left(i\pi\lambda z(u^2 + v^2)\right)$ must be reducible to $\frac{i\pi}{N}(m^2 + n^2)$ [73]. Since $u = m\Delta_F$ and $v = n\Delta_F$, with $\Delta_F = \frac{1}{S_{object}}$, this yields the condition [73]

$$\frac{S_{object}^2}{\lambda z} \leq N, \tag{1.77}$$

which relates the experimental parameters to the number of pixels. A small number of pixels thus leads to a decreased resolution.

When reconstructing experimental holograms, further steps may have to be taken to avoid artefacts due to the imaging conditions. One source of artefacts that result in fringe patterns in the reconstructed image are sharp edges, i.e. localised large changes in contrast. Since the detector utilized in the set-up presented here (see section 1.2.1) is a phosphor screen, digitizing the holograms involves photographing the interference pattern formed on the circular screen with a digital camera. This means that the rim of the detector creates a sharp edge in the digital holograms.

To avoid artefacts in the reconstruction arising from these sharp edges, an apodization filter is used. This filter works as a mask multiplied with the hologram [73]: inside a radius η, the multiplication factor is 1, leaving the values of the hologram unchanged. This is followed by a ring of width ω in which the multiplication is proportional to a squared cosine function which results in a smooth fall-off to zero. Outside the radius $\eta + \omega$, the multiplication factor is zero, i.e.

$$C(\rho) = \begin{cases} 1, & 0 \leq \rho \leq \eta \\ \cos^2\left(\frac{\pi}{2\omega}(\rho - \eta)\right), & \eta < \rho < \eta + \omega \\ 0, & \rho \geq \eta + \omega \end{cases} \tag{1.78}$$

Numerically, this can be implemented as follows:

```
def Cosapo(holin, omega, eta):
    (Nx,Ny)=holin.shape[:2]
    c=np.empty((Nx,Ny), dtype=complex)
    for ii in range(Nx):
        for jj in range(Ny):
            if (0<=np.sqrt((ii-Nx/2-1)**2+(jj-Ny/2-1)**2)
                and np.sqrt((ii-Nx/2-1)**2+(jj-Ny/2-1)**2) <=eta):
```

```
8                       d=1
9                   elif (eta<np.sqrt((ii-Nx/2-1)**2+(jj-Ny/2-1)**2)
10                       and np.sqrt((ii-Nx/2-1)**2+(jj-Ny/2-1)**2)<eta+omega):
11                   d=np.cos((np.pi/(2*omega))*(np.sqrt((ii-Nx/2-1)**2
12                       +(jj-Ny/2-1)**2)-eta))**2
13                   else: d=0
14                   c[ii][jj]=d
15       return c*holin
```

To demonstrate that the reconstruction algorithm presented in this section works on both simulated and experimental data, Fig. 1.5 shows amplitude reconstructions of a hologram obtained by applying the simulation algorithm to a disk of amplitude $a = 0.6$ and of an experimentally acquired hologram of a monoclonal antibody imaged by low energy electrons (see Chapter 2).
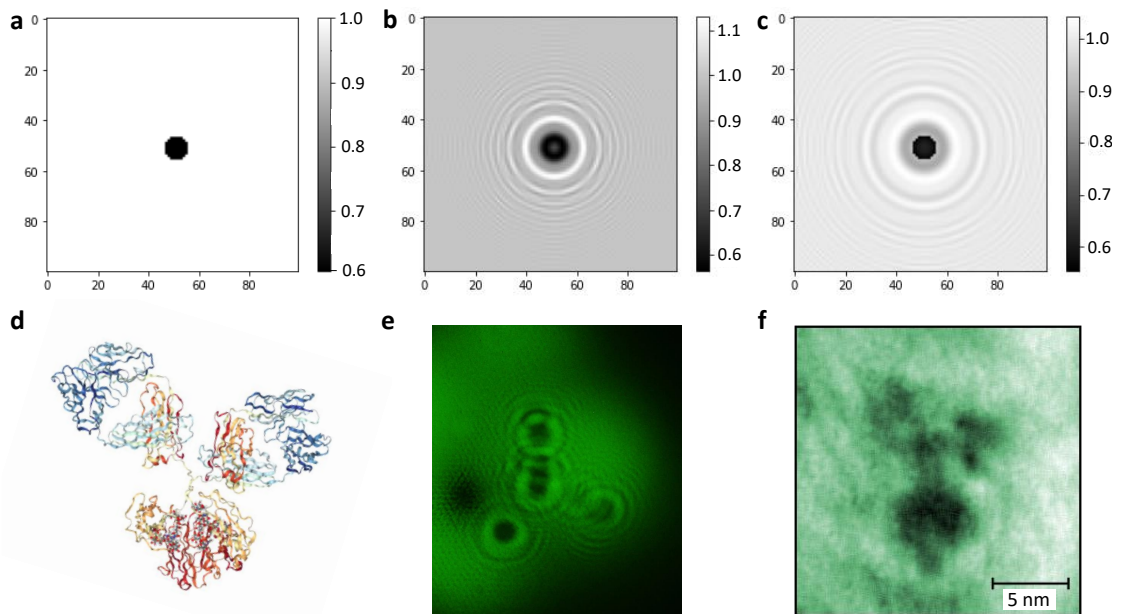
Figure 1.5: **Amplitude reconstruction of a simulated and an experimentally acquired hologram: a** Disk with amplitude $a = 0.6$ that serves as input for the hologram simulation. **b** Hologram simulated from the disk shown in **a** using the convolution approach presented in this section. **c** Amplitude reconstruction obtained with the algorithm described in this section from the hologram simulated in **b**. The amplitude of the object itself is reconstructed to a high degree of accuracy, the fringes in the background can be attributed to the contribution of the twin image. **d** Molecular model of an antibody molecule (PDB: 1IGT [82]). **e** Experimentally acquired hologram of a monoclonal antibody molecule (Herceptin). **f** Amplitude reconstruction of the hologram in **e** at source-to-sample distance $z_0 = 310$ nm. The three subunits and the hinge region connecting them are clearly identifiable in the reconstructed image.

In the simulated example (Fig. 1.5a-c), the amplitude of the object itself is reconstructed to a high degree of accuracy, but the background amplitude is not uniform. It exhibits modulations in the form of fringes that can be explained by the out-of-focus contribution of the twin image in the object plane. In a single-step reconstruction, as presented in this section, this contribution cannot be avoided in an in-line geometry. An iterative

treatment that can remove these contributions is discussed in Chapter 3.

In the case of the reconstruction obtained from the experimentally measured hologram (Fig. 1.5e), the transmission function of the object is unknown, hence it is harder to estimate the accuracy of the reconstructed values. The size and shape of the reconstructed molecule (Fig. 1.5f), however, match those known from crystallographic data [82] (Fig. 1.5d), and many structural features are identifiable. In general, the reconstructed amplitude can be associated with absorption properties and inelastic scattering interactions, as briefly discussed in the introduction to this chapter. The information about the object that can be extracted from amplitude images will be discussed in more detail in Chapters 3 and 4 in relation to the phase information encoded in the hologram. In Chapter 2, the amplitude reconstruction algorithm presented in this section will be utilized for the single molecule imaging of highly flexible proteins, thus demonstrating its value in scientific applications.

1.1.4 Coherence

Coherence is a measure for the ability of two waves to interfere. Since the interference of the scattered wave with the reference wave is at the basis of holography, a coherent source is an essential prerequisite for holographic imaging. In the discussion carried out in the previous sections, a point source was assumed, which yields a perfectly coherent monochromatic reference wave. Since a perfect point source cannot physically exist, the effect of a partial loss of coherence due to an extended source on holographic imaging has to be considered.

Coherence implies that both the waveform and the frequency of the two wave fields in question are the same, these two components are often called spatial and temporal coherence. While full coherence requires both spatial and temporal coherence, spatial coherence can exist without temporal coherence and vice versa.

Temporal coherence can be viewed as a measure of the monochromaticity of the source. While two monochromatic waves with a given phase difference will retain that phase difference over time, this is not the case any more if at least one of them is not perfectly monochromatic, i.e. if it has a frequency spread Δf. Then, the waves lose their correlation, i.e. become incoherent, on the time scale $\tau_c = \frac{1}{\Delta f}$ [83, 84]. τ_c is called the coherence time. Temporal coherence can also be characterized by a spatial length, called the longitudinal coherence length L_l since it characterizes the coherence along the propagation direction, which is the length it takes for two waves with a small difference in wavelength to interfere destructively. It can be shown to depend both on the wavelength λ and the spread in wavelength $\Delta\lambda$ [85]:

$$L_l \propto \frac{\lambda^2}{2\Delta\lambda} \tag{1.79}$$

For a known energy spread of the source, L_l can hence be directly calculated.

Spatial coherence, on the other hand, describes the coherence properties in the direction transverse to the wave propagation. The corresponding coherence length is thus referred to as the transverse coherence length L_t. A limited spatial coherence can directly
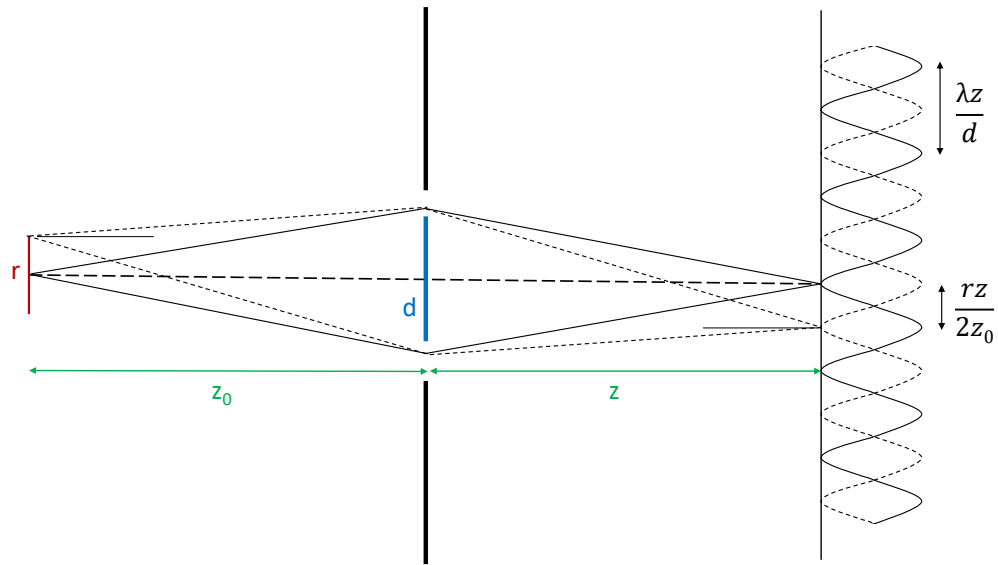
Figure 1.6: **Derivation of the transverse coherence length:** Interference patterns produced by a double slit set-up with slit distance d with an extended source of size r. The transverse coherence length L_t is defined as the slit distance for which the waves originating from the centre and the edge of the source interfere destructively.

be related to the source being physically extended rather than a perfect point. That an extended source affects the resulting interference pattern can easily be seen when considering a double slit experiment to generate an interference pattern as sketched in Fig. 1.6. Considering waves originating from different points of a source with extension r, they will create the same interference patterns (i. e. the spacing between maxima will be identical), but the diffraction angles will differ, which leads to a shift of the interference patterns with respect to one another, which in turn reduces the contrast of the interference pattern. This can be elucidated quantitatively by examining a wave emitted from the centre of the extended source and comparing it to a wave emitted from the upper edge of the source. Maxima of the interference pattern generated by a wave originating from the centre of the source emerge at diffraction angles $\sin\theta = n\frac{\lambda}{d}$, $n \in \mathbb{Z}$, where d is the separation between the two slits. The minima emerge at the corresponding angles with $n = \pm\frac{1}{2}, \pm\frac{3}{2}, \dots$. For small angles, when the approximation $\sin\theta = \tan\theta$ can be made, this yields a fringe spacing of $\frac{\lambda z}{d}$. The maxima generated by the wave originating from the edge of the source are shifted by an angle $\beta = \frac{r}{2z_0}$, i. e. they appear at angles $n\frac{\lambda}{d} + \frac{r}{2z_0}$ and are spaced $\frac{rz}{2z_0}$ apart, which can be derived by again assuming small angles. The transverse coherence length is then defined as the slit separation d for which the first maximum of the wave from the edge coincides with the first minimum of

the central wave. Hence

$$\frac{1}{2}\frac{\lambda z}{d} \overset{!}{=} \frac{rz}{2z_0}.$$ (1.80)

$$\Leftrightarrow d = \frac{\lambda z_0}{r}$$ (1.81)

$$\Leftrightarrow L_t = \frac{\lambda z_0}{r}$$ (1.82)

Equation (1.82) directly relates the transverse coherence length to the source size r. If the transverse coherence length can be determined experimentally, e. g. by measuring the extent of the fringe pattern in the detector plane, this relation can serve as an estimate of the effective source size r_{eff}. A similar relation can be derived via the van Cittert-Zernike theorem, which yields

$$r_{\text{eff}} = \frac{\lambda z_0}{\pi L_t}$$ (1.83)

for an incoherent Gaussian source [83, 86].

In the case of spherical illumination, the geometrical magnification has to be taken into account, hence the distance z_0 has to be adjusted to match the distance to the plane where the transverse coherence length is measured. Thus, if the transverse coherence length is measured in the detector plane, z_0 has to be replaced by the source-to-detector distance. Equation (1.83) is often used to estimate the effective radius of a source in the form of a sharp tip [87, 88, 89].

For a (partially) coherent source, the effective source size is smaller than the geometric size of the source [87, 89]. The effective source, also often called the virtual source, is the area in which the trajectories of emitted electrons intersect when backpropagating them into the tip [87, 90]. Since the algorithm presented in the previous section assumes a point source, the source-to-sample distance provided by the algorithm while focusing the image is the distance to the virtual source, not to the geometrical apex of the tip.

Thus, while the preparation of emitters in the form of sharp tips (see section 1.2.2) yields geometric tip radii in the range of several nanometers (see Fig. 1.9), a high degree of coherence of the source can decrease the radius of the virtual source to values in the low angstrom range [91, 89, 87], which is particularly relevant since the effective source size is often considered to be an estimate of the attainable resolution [88].

Emitter preparation is hence a crucial part of holographic imaging, the general method is described in section 1.2.2. Since there is evidence that cooling the tip to liquid nitrogen temperatures [89] increases the coherence of the source due to the increased inelastic mean free path in the tungsten tip at low temperatures, which allows a larger number of electrons to interfere with one another, holographic imaging at low temperatures appears to be a promising path to increase stability and resolution.

1.1.5   Resolution

The question of attainable resolution in in-line holography with low energy electrons is not easy to answer. In general, the system is purely diffraction-limited, i. e. the size d of

the smallest resolvable feature is given by

$$d = \frac{\lambda}{2n \sin \theta} = \frac{\lambda}{2\text{NA}}, \tag{1.84}$$

where θ is the half-angle describing the divergence of the beam and n is the refractive index of the medium. In vacuum, which is needed for low energy electron holography, $n = 1$. $n \sin \theta$ is also referred to as the numerical aperture NA of the system. This limit is known as the *Abbe diffraction limit.*

For a typical electron beam energy of $E = 100\,\text{eV}$, corresponding to a wavelength of $\lambda = \hbar/\sqrt{2m_e E} = 1.23\,\text{Å}$ and a half-opening angle of $18°$, which corresponds to the maximum opening angle of the system presented in section 1.2, this yields a resolution limit of $1.99\,\text{Å}$.

Depending on the exact criterion utilized to classify a feature as resolved and on whether the derivation is carried out in real space or Fourier space, different prefactors can occur, but all expressions for the resolution limit share the proportionality to $\frac{\lambda}{\text{NA}}$ [92]. The coherence of the source is of relevance for the determination of the theoretical resolution limit, too. To account for that, the prefactors in eq. (1.84) have to be adjusted, while the aforementioned proportionality remains [92].

In practice, however, there are many factors that can affect resolution, such as mechanical vibrations, instabilities of the tip, limited coherence due to an extended source, which decreases the quality of the fringe pattern, and the interaction with the object. Since no lenses are present in an in-line geometry, lens aberrations, which determine the attainable resolution in many other imaging systems, do not play a role here.

Furthermore, due to the numerical reconstruction, sampling also plays a role in determining the possible resolution. According to the Shannon sampling theorem [93], the correct representation of a periodic system requires a sampling of at least two pixels per period. Thus, the smallest resolvable fringe needs to have a width of two pixels. With a pixel size Δ_O in the object plane, this yields a resolution limit of $d_S = 2\Delta_O$. Since d_S can be decreased by increasing the number of pixels, thereby decreasing Δ_O, $d_S < d$ is true in most cases, especially for high-magnification images. At low magnification, d_S will at some point exceed the Abbe limit d and become the limiting factor for the achievable resolution of the system.

While eq. (1.84) implies that the theoretical resolution limit can be lowered by employing shorter wavelengths (i. e. higher energies) and larger numerical apertures (larger opening angles), this course of action is restricted by practical considerations. When measuring in an in-line geometry at high magnification, i. e. at small source-to-sample distances z_0, the sustainable voltages are limited to the range of approximately $150\,\text{V}$ and below. Applying higher voltages at these distances would require currents in a range that can induce damage to the sample (see section 1.2.2 for details). Hence wavelengths below $1\,\text{Å}$, which is the wavelength corresponding to an electron energy of $150\,\text{eV}$, are impractical to achieve in high magnification imaging. To increase the opening angle, either the sample-to-detector distance can be decreased or the detector size can be increased. The sample-to-detector distance could be decreased without significant changes to the experimental set-up, however, it would also lead to a decrease in the possible magnification if the detector size remains the same. A change in detector size would require changes to

the vacuum chamber in which the holography microscope is set up (see section 1.2.1), and is hence difficult to implement. Furthermore, increasing the opening angle will require modifications in the reconstruction algorithm presented in section 1.1.2 and 1.1.3, since the paraxial approximation does not hold for large angles. Thus, one would either need to evaluate the diffraction integral directly or implement coordinate transforms to project the spherical wave front at the detector that cannot be approximated by a flat screen any more, to the detector plane [73].

Due to imperfect experimental conditions, it is unlikely that a resolution at the theoretical limit is achieved, hence other possibilities to determine the resolution of a hologram, and, most importantly, its reconstruction, have to be considered. In imaging geometries using a point-like source, like the one employed here, the size of the virtual source is often used as an estimate of the attainable resolution [88], which implies that the resolution critically depends on the coherence properties of the source. The virtual source size can be calculated from the transverse coherence length (see section 1.1.4), which can be approximated by the spatial extent of the fringe pattern observed on the detector. Thus, in order to improve resolution, enhancing the coherence of the source is essential. This could be done by cooling tip and sample, which should increase the coherence of the tip due to an increased electron mean free path within the tip [89] and also reduce molecular motion and vibrations of the graphene substrate.

Another possibility to estimate resolution is to determine the size of the smallest resolved feature in the reconstructed image. In the case of the antibody molecule shown in Fig. 1.5f, such a feature would be the distinguishable peptide chains of the hinge region, whose width in this image is approximately 6 Å.

Since the size of the smallest resolvable feature in a given image ultimately depends on many factors, such as tip stability, coherence, electron energy and vibrations, the resolution can differ significantly between images, even between consecutive images of the same molecule taken within a short time span. Furthermore, while small, isolated features, such as the hinge region in Fig. 1.5f can be resolved, features of the same size within the bulk of a molecule, e.g. within a subunit of the antibody in Fig. 1.5f, often cannot be resolved because they are too close to other structural features. Thus, the resolution estimated in that way has to be understood as a local measure, not as a global property of the image.

An additional option for estimating the resolution of a reconstructed image is to examine its Fourier spectrum. An estimate of the full width half maximum of a Gaussian approximation to the point spread function can be derived from a radially averaged logarithmic intensity plot in Fourier space [94], which can be used as a measure of resolution. Since this procedure involves the Fourier transform of the whole image, it provides a global estimate of the resolution of the image that is not tied to a specific spatial feature [94].

So far, the discussion has been restricted to lateral resolution within the object plane. A physical object, however, is in general three-dimensional, hence the resolution along the optical axis is another important quantity. Ordinarily, the size of laterally resolvable features is smaller than that of axially resolvable features, which is also reflected in the expression of the axial resolution limit d_{ax} which is inversely proportional to NA^2 rather

than to NA [92, 95]:

$$d_{ax} \propto \frac{\lambda}{\mathrm{NA}^2}. \tag{1.85}$$

Again, different prefactors can occur depending on the specific derivation of the expression. Reconstruction along the axial direction is further exacerbated by the fact that when looking at a given reconstruction plane, not only the part of the object in focus in that plane will have contributed to the reconstructed signal, but also parts of the object located in different planes along the optical axis, which has already been shown in the discussion of the twin image (see section 1.1.1). Separating the contributions from the different planes, especially when they are in significant lateral overlap and the planes are in close proximity, as is generally the case for three-dimensional objects such as proteins, is very complex and requires the application of additional methods as discussed in Chapter 5.

In general, resolution is determined by the experimental conditions, in some cases, however, the processing of the recorded hologram can enhance resolution slightly, e. g. by removing high frequency noise from the hologram. While this can also lead to a blurring of the smallest fringes, it in general makes the visible fringes more pronounced, and thus can bring out some features more clearly in the reconstruction. This becomes apparent when comparing the amplitude reconstructions of a single antibody shown in Fig. 1.7. While Fig. 1.7a shows the reconstruction from the unprocessed hologram, Fig.1.7b is the result of the reconstruction of the same hologram with the high frequency noise removed in Fourier space. The image processing steps will be discussed in section 4.1. Some of the features appear more pronounced, and the smallest resolved features in the hinge
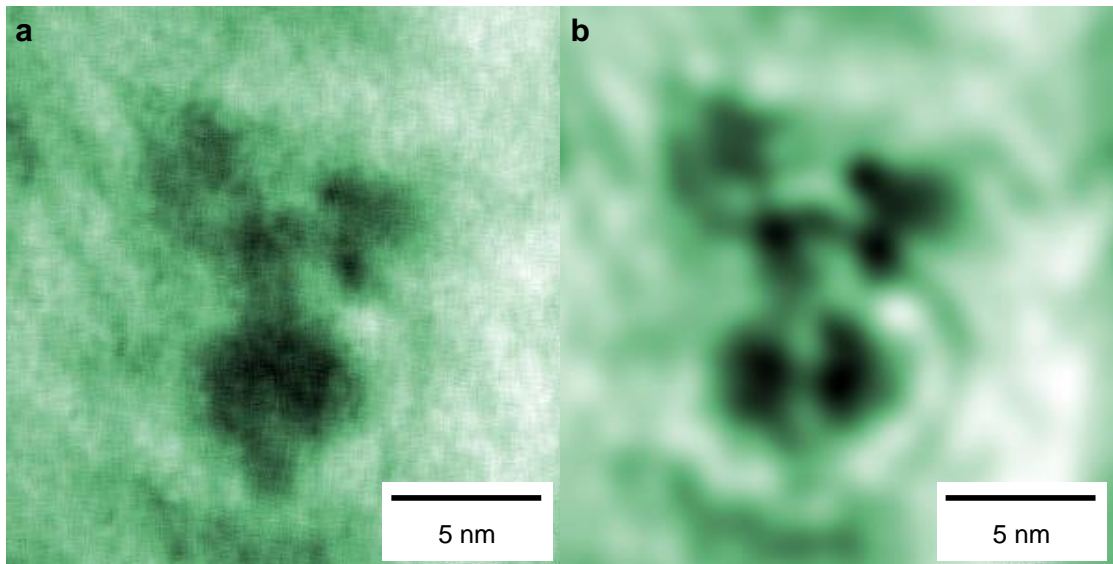

Figure 1.7: **Comparison of reconstructions from processed and unprocessed holograms of a single antibody: a** Reconstruction of a single antibody from the unprocessed hologram. **b** Reconstruction of the same antibody molecule from a hologram with high frequency noise removed in Fourier space.

region near the Fab and the Fc subunits are measured to have a size of 5 Å.

Further possibilities for image enhancement after measuring are compensation of noise by deblurring with a Gaussian [93] and extrapolation of the fringe pattern [96].

1.2 Experimental set-up

This section focuses on the experimental requirements and implementation of both LEEH imaging and sample preparation. The first part provides an overview of the design of the LEEH microscope and the imaging conditions, followed by a discussion of sharp tips as coherent low energy electron sources and their preparation. The second part addresses sample preparation, discussing the steps necessary for obtaining specimen suitable for LEEH imaging. This involves both the preparation of the single layer graphene (SLG) substrate and the deposition of the molecules on the substrate by native electrospray ion beam deposition (ES-IBD).

The development and construction of the experimental LEEH set-up used to measure the experimentally acquired holograms discussed in this thesis is described in detail in Sven Szilagyi's doctoral thesis "Low-energy Electron Holography Microscope for Imaging of Single Molecules".

1.2.1 Low energy electron holography

LEEH imaging needs to take place in an ultraclean environment, because any additional scatterers, such as impurities on the sample, would contribute to the hologram and hence complicate the discrimination between the signal from the system of interest and from such impurities. This, along with the increasing instability of the electron sources used in LEEH at higher pressures due to adsorbates, requires ultra high vacuum (UHV) imaging conditions [24].

Our LEEH microscope is integrated into a vacuum chamber with a base pressure in the low 10^{-10} mbar range. Next to the microscopy chamber, where LEEH imaging takes place, our set-up is also equipped with a preparation chamber in which the tips used as electron emitters can be characterized and prepared (see section 1.2.2).

The microscope itself consists of three parts: the electron source, the sample and the detector, which are arranged in a vertical in-line geometry (see Fig. 1.8a-b) that allows for lensless holographic imaging [27].

As electron source, a sharp tungsten tip is used (see section 1.2.2), which is mounted on a custom-build tip holder, allowing for the application of a voltage to the tip during imaging as well as a current for annealing during tip preparation. For imaging, the tip holder is placed on an xyz piezo motor stack to enable relative motion between tip and sample. This relative movement is critical for imaging different parts of the sample and for tuning the magnification by adjusting the source-to-sample distance. While the source-to-sample distance is variable, ranging from mm to nm distances, the sample-to-detector distance is kept constant at 11.5 cm. The resulting magnification is purely geometric and yields magnification factors of up to 10^7. The extended range of magnifications allows the acquisition of both survey images (Fig. 1.8c) and holograms of individual molecules (Fig. 1.8d). The source-to-sample distances for high magnification imaging are typically in the range of $200 - 500$ nm, whereas they are in the micron range
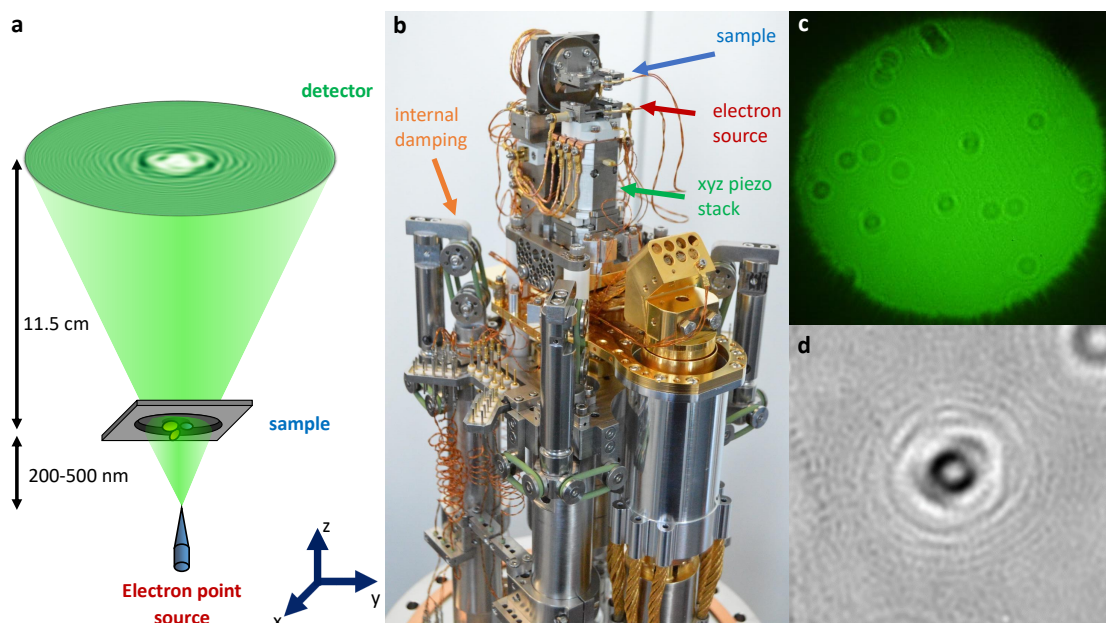
Figure 1.8: **LEEH set-up and imaging: a** Schematics of in-line holography geometry **b** LEEH microscope head **c** survey hologram **d** high magnification hologram.

in the case of survey images. Since the source-to-sample distance cannot directly be measured experimentally in this set-up, it has to be determined numerically by finding the source-to-sample distance for which the reconstructed object appears in focus (see section 1.1.3).

For imaging, a negative voltage is applied to the tip to generate electrons by field emission. At high magnifications suitable for imaging individual molecules, the voltage is in the range between -50 and $-150\,$V, corresponding to electron wavelengths between $1.7\,$Å and $1\,$Å. At low magnification, which allows a view of the whole sample (see section 1.2.3 and Fig. 1.10c), higher voltages in the range of -350 to $-750\,$V are typically required.

The wave scattered by the investigated molecule, and the reference wave, which did not interact with the molecule, overlap in the region behind the sample. The resulting interference pattern (the hologram) is recorded on the fluorescent screen of a microchannel plate detector. The holograms are digitalized by photographing the pattern displayed on the screen with a digital camera. The final image is obtained by numerical reconstruction of the digital images as described in section 1.1.

In order to obtain high resolution reconstructions, high-order interference fringes must be visible in the holograms since these high-order interference fringes encode small spatial features. To achieve this, it is crucial to isolate the microscope from vibrations as the interference fringes appear closer together the higher their order. Since high-order interference fringes encode short length scale spatial information, even small vibrations can lead to a blurring of the holograms that negatively affects the image resolution. To eliminate sources of noise that could affect the quality of the holograms, the chamber was mounted on a stiff and heavy frame attached to a concrete block on air springs and set up in an electromagnetically and acoustically shielded box [97]. To avoid noise induced by the experimenter, measurements are conducted remotely, i. e. from outside the box.

In addition to these vibration isolation measures aimed at shielding the whole system from noise, the microscope head is placed on a stage that is decoupled from the chamber via a Viton$^{\circledR}$ O-ring internal damping system (Fig. 1.8b). This internal damping system significantly reduces high-frequency vibrations and even allows for certain experiments, such as establishing the cleanliness of the substrate prior to sample preparation (see section 1.2.3), to be conducted while the turbo pumps are running. High resolution images can, however, only be acquired with the pumps turned off, in these situations the pressure is kept in the low 10^{-10} mbar range by an ion getter pump.

1.2.2   Tip preparation

Electron holography requires a coherent electron source to create stable interference patterns (see section 1.1). The coherence of an electron source is inversely proportional to the source size, i.e. to the volume from which electrons are emitted [98]. Thus, a small source size – as close to the ideal of a point source as possible – is desirable. Such point-like electron sources, which can be considered to emit spherical electron waves, can be created by field emission from sharp metal tips with only a few atoms at the apex [29]. Field emission from metals into vacuum is a quantum mechanical effect in which the potential barrier defined by the work function of the material is overcome by electron tunnelling facilitated by the presence of an electric field. The application of an electric field modifies the shape of the potential barrier, resulting in a rounded triangular barrier [99, 100, 101] that decreases in height with increasing field strength. The specific tunnelling behaviour through a triangular barrier, as is the case in these geometries, is called Fowler-Nordheim tunnelling. It can be described by Fowler-Nordheim type equations that relate the emission current from the tip to the applied electric field [99]. Sharp tip geometries are efficient field emission sources since the electric filed at a sharp tip is enhanced due to an increased density of field lines. The electric field E at a sharp tip can be described as $E = k\frac{V}{d}$, where V is the applied voltage, d is the source-to-sample distance, and k is an enhancement factor that describes the local enhancement of the field at the tip and depends both on the source-to-sample distance and the tip radius [102]. This implies that the extraction voltage decreases with decreasing source-to-sample distance when the emission current is kept constant. LEEH measurements can be carried out both in constant current and in constant voltage mode; in most cases, constant current mode is preferable since this ensures a consistent illumination throughout the measurement.

Due to field enhancement, sharp tips allow low extraction voltages [53] in the range desired in LEEH. Typically, the voltages applied to the tip are in the range of -50 to -200 V at high magnification source-to-sample distances (ca. $200 - 500$ nm) and emission currents in the range of $10 - 50$ nA, producing electrons with energies between 50 and 200 eV, which are low enough to avoid radiation damage to the sample. While higher voltages produce electrons of shorter wavelength, which is in principle desirable in terms of resolution, higher currents are usually the consequence of operating at such voltages. High currents, however, can decrease the stability of the tip and can lead to excessive brightness and even to sparking, which can severely damage both tip and sample. Electron energies between 100 and 150 eV at source-to-sample distances suitable for single molecules imaging have proven to produce the highest quality holograms, since
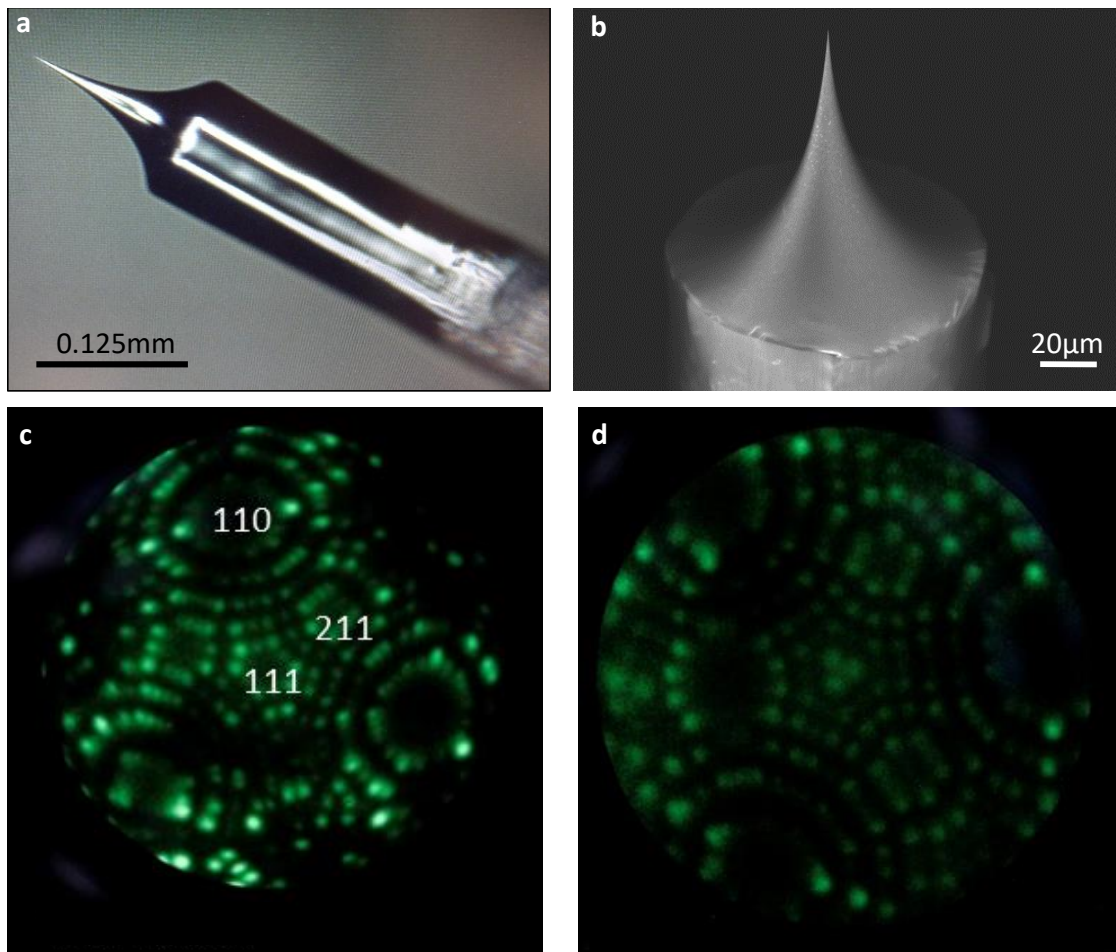
Figure 1.9: **Tip preparation and characterization: a** Light microscope image of a tungsten tip after electrochemical etching. The symmetry of the tip is vital for obtaining stable and coherent field emission. **b** Scanning electron microscopy (SEM) image of a tungsten tip. **c** Field ion microscopy (FIM) image of a tungsten (111) tip imaged at 18 kV. The orientation of several of the facets can be identified as indicated in the image. The geometrical tip radius was measured to be approx. 9 nm. **d** FIM image of a different tip with a clearly recognizable three-atomic apex with a measured geometrical tip radius of 10.5 nm.

at these energies, source size and wavelength are optimally balanced [103].

Field emission is in principle possible for all metals, tungsten, however, is the material of choice for electron sources for low-energy electron holography since it can be fabricated in ultrasharp geometries [29], has a high melting temperature, which makes it possible to clean the tips by annealing without blunting them, and is in general highly stable. Polycrystalline tungsten tips can in principle be used as electron sources for low-energy electron holography since their apex will in general be smaller than the crystallite grain size the wire is composed of, hence the emission area itself will be monocrystalline, although the orientation of the plane at the apex is not a priori predictable [104]. Due to their more consistent field emission characteristics, however, it is usually preferable to fabricate LEEH tips from single crystal tungsten (111) wire, which always yields a (111)

orientation at the apex, and results in smoother and more symmetric tips [105] because the electrochemical etching process selectively etches different crystal orientations with different etching rates [106].

Sharp tip geometries, as required for a LEEH electron source, can be prepared by electrochemical etching [106] (see Fig. 1.9a, b). For this, a piece of tungsten wire (diameter 0.125 mm, length 2-5 mm) is spot-welded to a tantalum arc (diameter 0.25 mm) and subsequently immersed into a 20 wt% NaOH solution while a voltage is applied between the wire and a counter electrode. The etching process leads to the formation of a neck in the wire until enough tungsten has been etched away for the lower part of the wire to break off [107]. At this point, due to the sudden change in current, the voltage is cut off, which stops the etching process. After the electrochemical etching process, the tip is cleaned in water to remove salt residues and is immediately brought into vacuum to avoid oxidation and the adsorption of impurities. The field emission of the tip is tested in ultra high vacuum. In order to improve the performance of the emitter, additional UHV preparation steps can be employed. Tip preparation in UHV includes self-sputtering by field emission in vacuum and sputtering with Neon to sharpen the tip, as well as annealing to remove impurity adsorbates and tungsten oxide.

Since the tip geometry is essential for the creation of coherent LEEH sources, it can be useful to characterize the tip by imaging it on the atomic level. This can be done with a field ion microscope (FIM) [108, 109]. In field ion microscopy, the chamber is filled with helium (partial pressure ca. 10^{-4} mbar) and a high positive voltage (in the range of $10 - 30$ keV) is applied to the tip. The high electric field at the tip ionizes the helium atoms close to the tip. The positively charged gas ions are repelled from the tip and move along the field lines to the detector, thereby creating an atomically resolved image of the tip (see Fig. 1.9c, d). In addition to that, the tip can be sputtered with nitrogen during FIM imaging, which both sharpens and stabilises the tip [91], and functionalised by picking up graphene flakes or carbon nanotubes.

1.2.3   Sample preparation

In general, molecular imaging with LEEH is not tied to a specific method of sample preparation. However, several conditions have to be fulfilled in order to be able to acquire high quality holograms.

First of all, a suitable substrate is crucial for successful imaging with LEEH. LEEH requires an atomically clean, low-energy electron-transparent substrate in order to produce a stable and undisturbed reference wave. An ideal substrate for LEEH would only induce a global phase shift which does not affect the interference pattern recorded on the detector. Furthermore, the substrate needs to be conductive to allow for distortion-free imaging by providing an equipotential surface at the sample and by avoiding charging effects. Additionally, the substrate material should weakly interact with the molecules to be studied in order to minimize structural changes in the molecules induced by molecule-substrate interaction. Single layer graphene (SLG) possesses all the aforementioned characteristics, i. e., it is transparent to low-energy electrons, can be prepared in an ultraclean fashion, is conductive and interacts with biomolecules via weak van der Waals forces [110], hence it is an ideal substrate for imaging proteins with LEEH.

In addition to a substrate, a method of depositing the proteins onto the substrate is

required to produce a sample for LEEH imaging. Dropcasting might first come to mind as a method of transferring molecules directly from solution onto a substrate. LEEH, however, requires an ultraclean sample to avoid disturbances of the reference wave. Solution residues on the sample could distort the reference wave significantly as they could lead to a non-uniform background and introduce additional scatterers that can interfere with the signal from the molecule. Depending on the solvent, solution residues can be eliminated from the graphene surface by heating the sample after dropcasting. For aqueous solutions that are used to maintain a native state of the proteins, however, temperatures of at least 100°C would be necessary to ensure that the purity of the sample is sufficient for LEEH imaging, which could in turn damage the proteins. If the molecules can withstand the temperatures needed for evaporating the remaining solvent, dropcasting could be used as a method of sample preparation for LEEH. While we have successfully prepared carbon nanotube samples for LEEH by dropcasting, these results will not be discussed here. The data presented in this thesis has solely been obtained from samples prepared using native electrospray ion beam deposition (native ES-IBD). Since native ES-IBD can both retain a native state of the sample, and, by mass-selective deposition that eliminates possible contaminants, produce clean samples in UHV [33], it is the method of choice for LEEH sample preparation [24].

In the following sections, the steps required to prepare both graphene substrates and protein samples are described in detail.

Single layer graphene substrate preparation

To fabricate ultraclean single layer graphene substrates for our LEEH experiments, the protocol described in [111] is followed: flakes of PMMA-coated single layer graphene on a copper substrate are suspended on a 0.33 mol/l ammonium persulfate solution to etch away the copper. After the etching, the flakes are transferred into clean water several times to eliminate residues. Following the cleaning procedure, the single layer graphene is fished using TEM grids. These grids consist of a silicon nitride membrane (thickness 200 nm) with four arrays of holes (hole diameter 500 nm, pitch distance between holes $2\,\mu$m), see Fig. 1.10a, b. Before the graphene is fished on the grids to cover the holes and thus create a free-standing single layer graphene substrate for protein deposition (Fig. 1.10), the grids are plasma cleaned and sputter coated with 5 nm of platinum and 10 nm of chromium on both sides to make the grids conductive, which is necessary to avoid charging effects during imaging. The platinum is additionally used as a catalyst to remove the PMMA layer in a stepwise heating procedure. Once the graphene has been held at 300°C for 45 minutes, it is immediately transferred into vacuum to avoid contamination from the air.

The cleanliness of the graphene substrate is ensured by a LEEH measurement before molecules are deposited. Characterization of specific holes before the deposition of proteins also facilitates returning to the same holes when imaging proteins to ensure that the imaged molecules stem from the deposition.

In general, this procedure yields holes covered by single layer graphene, but both empty holes (Fig. 1.10c) and holes partially covered by multiple layers of graphene can occur. Multi-layer graphene results in a much darker contrast than single layer graphene, which in most cases makes it too dark for successful imaging.
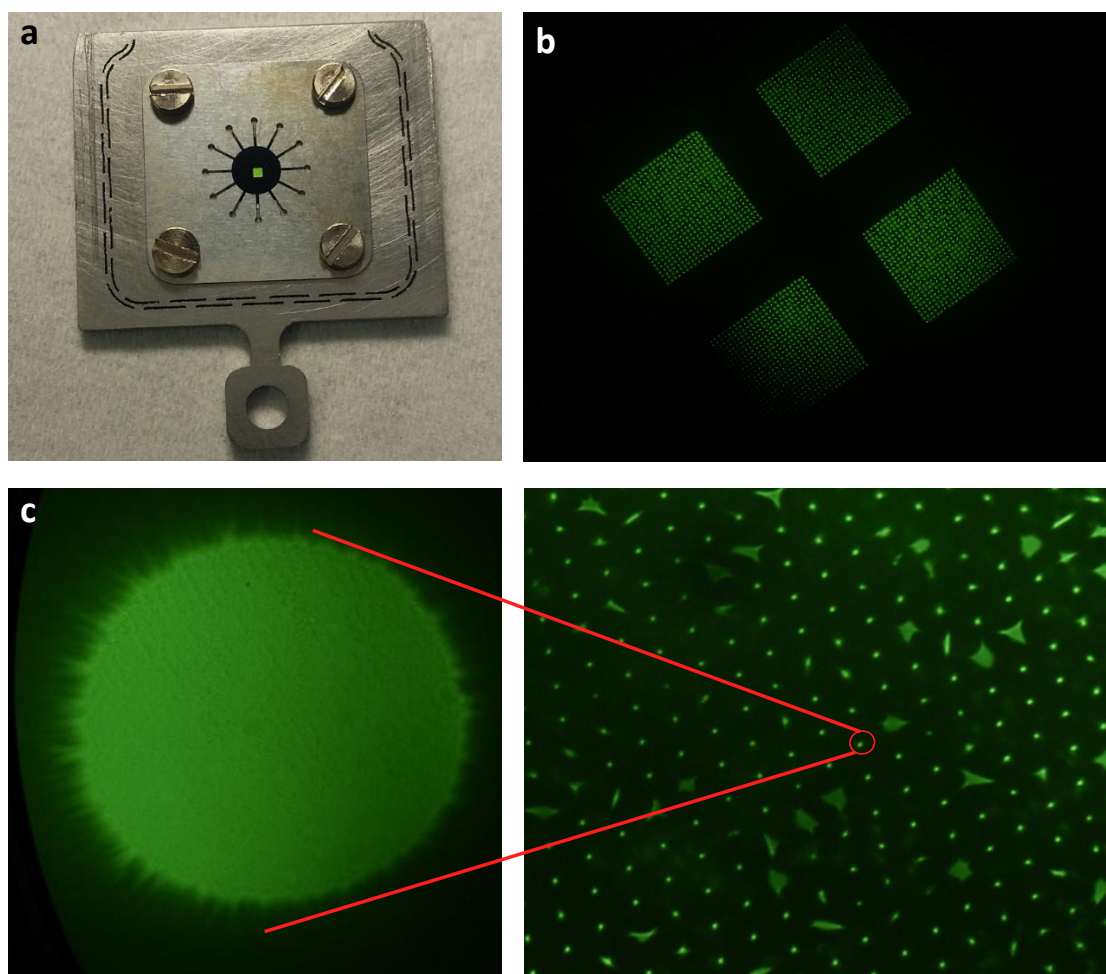
Figure 1.10: **Single layer graphene-covered TEM grids: a** Image of the sample holder with the graphene-covered TEM grid in the centre. **b** Low magnification image of the sample showing the holes in the TEM grid that are arranged in four arrays consisting of 25×25 holes each. Each of the holes has a diameter of $500\,\mathrm{nm}$, the pitch distance between two holes in $2\,\mu\mathrm{m}$. A sufficient pitch distance is necessary to avoid imaging artefacts due to contributions of neighbouring holes. This geometry facilitates finding and coming back to certain holes, which allows a comparison of the holes before and after the deposition of molecules. **c** Free-standing graphene sample at high magnification (left) and medium magnification (right). In the medium magnification image, holes covered with single layer graphene appear small and circular, whereas empty holes result in larger, angular shapes, which are due to charging and lensing effects. The high magnification image shows a perfectly clean SLG substrate without any contaminants or defects.

The graphene grids are transferred between the holography microscope, where characterization and imaging happens, and the native ES-IBD deposition chamber, in which the protein sample is prepared, utilizing a vacuum suitcase (base pressure $1 * 10^{-10}\,\mathrm{mbar}$) to avoid contamination both before and after protein deposition. Thus, once the samples are brought into vacuum, they are constantly kept at low $10^{-10}\,\mathrm{mbar}$ pressure.

Protein preparation

The most important prerequisite for protein sample preparation by native ES-IBD are solution conditions that support native protein conformations [112, 113], which will be referred to as native solution conditions in the following. Native solution conditions require a spray solution with stable pH, to achieve this, aqueous solutions like ammonium acetate (200 mM) are used, which is very common in native electrospray ionization techniques [112].
Depending on the proteins and their properties, they can either be stored as lyophilised powders or in a buffer solution (tris buffered saline solution, pH=7.6). In the former case, the proteins are directly desolved in 200 mM ammonium acetate, this solution is then purified via a biospin process. In the latter case, the buffer solution has to be exchanged with ammonium acetate, which is also done via a biospin process. For the biospinning, size exclusion columns are used in a centrifuge that is cooled to 4°C [114, 115]. The resulting solution is then diluted to concentrations of $0.5 - 1\,\mathrm{mg/ml}$ and kept on ice until it is filled into the nanospray emitter.

Native electrospray ion beam deposition

Electrospray ionization (ESI) is a versatile tool for transferring a broad range of molecules, including large and complex biomolecules such as proteins, into the gas phase by soft ionization and is hence widely used in mass spectrometry applications [116, 117, 118, 119, 120, 121, 122].
The transition from solution to the gas phase is achieved by applying a high voltage to a sharp emitter filled with solution. The strong electric field at the apex of the emitter results in a deformation of the liquid into a Taylor cone [123], which emits a jet of charged droplets once a certain threshold voltage is reached [124]. While the droplets move towards the counter electrode at the vacuum interface, the solvent evaporates, causing the droplets to become unstable due to the increase of charge per droplet volume. The repulsion of the charges within the droplets results in Coulomb fission, which leads to the disintegration of the larger droplets into several smaller ones. This process iterates as a result of continued solvent evaporation until only desolvated molecular ions remain. The specific mechanism governing the final desolvation step and thus the transition into the gas phase can follow one of two main pathways: the *charge residue model* and the *ion evaporation model* [125, 126, 127, 128, 129, 130, 131]. The *charge residue model* proposes that the fission process due to solvent evaporation continues until a droplet only contains a single molecule. Upon the evaporation of the remaining solvent, the charge in the droplet is transferred to the molecule, producing molecular gas phase ions. In contrast, the *ion evaporation model* suggests that molecular ions can be ejected from droplets whose radius is small enough to lead to electric fields at the surface strong enough to ionize the molecules in the droplet. By which process the transition into the gas phase takes place may also depend on the size of the molecules [132]. Both models allow the molecular ions to be multiply charged, and large molecules, such as proteins, have been shown to carry high amounts of charge [132, 133, 134].
The molecular ions are subsequently transferred into vacuum by an ion funnel that creates an ion beam of defined kinetic energy that can be steered towards mass spectrometry

and deposition by ion optics.

In principle, electrospray ionization is possible both in positive and negative mode, i. e. with positive and negative voltages applied to the emitter, respectively. This, in turn, creates positively and negatively charged molecular ions. Which mode is used depends strongly on the properties of the molecules, especially their functional groups, and the solvents used. Proteins are usually sprayed in positive mode [112, 135, 136, 134, 137], whereas other types of molecules, such as saccharides, work best in negative mode [138, 139, 140].

As a sample preparation tool, electrospray ionization has the advantage that it can directly be combined with preparative mass spectrometry techniques and ion mobility measurements, which does not only allow a clean sample preparation in UHV, but also the mass selection of specific charge states and hence the control of molecular properties such as the protein fold [141, 137] (see Fig. 1.11).

In our electrospray ion beam deposition setup [142], a quadrupole mass filter and a time-of-flight mass spectrometer are used to characterize and mass-select molecules with the desired m/z ratio, which are subsequently deposited onto a substrate in UHV ($p \approx 10^{-10}$ mbar). This yields samples with the high chemical purity necessary for LEEH imaging, since additional scattering due to impurities is eliminated.

The landing energy of the molecular ions on the substrate can be tuned via a voltage applied to the sample, which permits the exploration of both soft-landing and reactive landing regimes [143, 142, 144, 24, 138, 145, 146, 147, 148, 149]. The amount of deposited charge, and thereby, since the charge states are known from mass spectrometry, the amount of deposited molecules, is determined via a current measurement on the sample, which allows the control of the coverage of the sample.

When using ES-IBD to prepare samples for studying proteins in a native state, the experimental conditions have to be carefully chosen in order to retain a *native-like* state of the protein at every point of the process. The term *native-like*, which is to be used in the following, indicates that while the proteins' primary, secondary and overall tertiary structure remains intact, changes in environment during the sample preparation process can induce deviations from the native protein structure in solution [150], for example side chain collapse [151] and subunit rearrangement [152].

Conditions appropriate for the retention of native-like conformations include the usage of low ionization voltages ($1.0 - 1.5$ kV) applied to sharp metal-coated glass emitters to avoid damaging or unfolding of the molecules during the ionization process, temperatures at the air-vacuum interface of approx. 70°C to assist solvent evaporation without damaging the proteins, and small voltage offsets in the higher pressure parts of the setup to avoid activation due to background gas collisions. This ensures gas phase protein structures with intact stoichiometry and minimal structural deviation from the native solution structures [112]. Furthermore, a retarding voltage applied to the sample yields soft-landing conditions with low kinetic energies (≤ 5 eV per charge) upon impact to exclude damage to the proteins during the landing process.
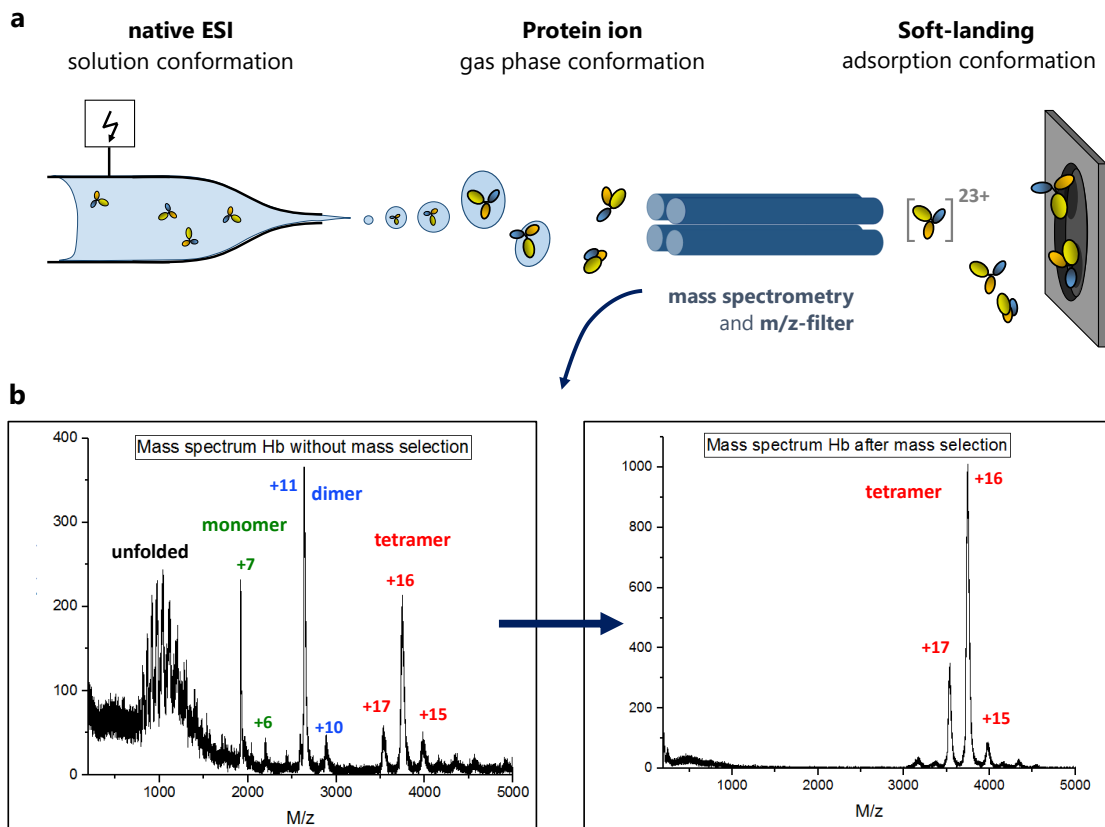
Figure 1.11: **a Native ES-IBD workflow:** Starting from native solution conditions, in which proteins can retain their native tertiary and quarternary structure, the application of a voltage (approximately $1000 - 1500\,\mathrm{V}$) to a thin emitter results in a jet of charged droplets. Desolvation produces molecular gas phase ions that are mass filtered by two quadrupoles after characterization by mass spectrometry with a time-of-flight mass spectrometer and deposited on the surface using soft-landing conditions (kinetic energy upon landing $\leq 5\,\mathrm{eV}$ per charge) regulated by the application of a retarding voltage to the sample. During the ES-IBD process, the environment of the molecules changes several times, from solution via the gas phase to the surface, which can affect the molecules' conformation, as discussed in detail in Chapter 2.

b Mass spectrometry: Mass spectrum of a native hemoglobin (Hb) beam before (left) and after (right) mass filtering. Hemoglobin is a protein complex with a total mass of $64\,\mathrm{kDa}$, consisting of four subunits with a mass of $16\,\mathrm{kDa}$ each. Before mass selection, a large number of peaks is visible in the mass spectrum (left) that can be categorised into four classes by fitting the observed mass-to-charge values to charge states of either the intact protein or its subunits. The peaks in the high mass range (approx. 3500-4000 m/z, charge states +15 to +17, marked in red) correspond to charge states of the intact hemoglobin molecule, i. e. to the tetramer composed of four subunits. The peaks between 2500 and 3000 m/z can be identified as two different charge states (+10, +11, marked in blue) of the dimer, i. e. the partially dissociated molecule consisting of only two subunits, while the peaks around 2000 m/z correspond to the charge states +6 and +7 (labelled in green) of the momomer, i. e. to a single subunit. The broad peak distribution centred around 1000 m/z can be associated with unfolded molecular conformations. After mass filtering (right), only the charge states corresponding to the tetramer remain. This allows for a selective deposition of specific charge states associated with the molecular species of interest.

2 Imaging of Conformational Variability of Single Antibody Molecules by LEEH

Imaging on the single molecule level is of particular importance when studying molecules with a high degree of structural flexibility. Biologically, this is of relevance since conformational changes are often associated with the molecules' functionality [1, 2, 3, 4, 5, 153]. Flexible molecules are often hard to crystallize [154, 155, 156] and to image with techniques relying on averaging over a large number of molecules. In contrast, single-molecule techniques can explore the full conformational space of the molecule in the imaging environment [157, 158]. While the conformational space on the surface, which can be examined by LEEH imaging, is not a one-to-one map of the conformational space of the molecule in solution, studying surface conformations can still provide useful insights into the general flexibility of the molecule and into the range of stable conformers.

While LEEH has been shown to be capable of single-molecule imaging of small globular proteins [24] and suspended macromolecules [159, 160, 161], highly flexible molecules have so far not been successfully imaged.

Probing LEEH's capability of imaging conformational variability on the single-molecule level requires a test system. Such a system should both exhibit a high degree of structural flexibility on the large scale – yielding the possibility of a change to the overall shape of the molecule that is detectable with the current LEEH resolution of approx. 1 nm – and additionally have characteristic structural features, such as well-distinguishable subunits, to facilitate the identification and distinction of different conformations.

An molecular system that meets these criteria are Immunoglobulin G (IgG) antibodies. IgG antibodies are glycoproteins that consist of four peptide chains (two heavy chains and two light chains), which interconnect to form three distinct subunits. Each subunit is composed of two peptide chains forming a loop around a cavity. Two of the subunits, the antigen-binding Fab subunits, which consist of a light chain and part of a heavy chain, respectively, are identical, while the third subunit, the Fc subunit, consists of two heavy chains and mediates the antibody's biological function. The hinge region connecting the subunits is highly flexible, which allows for each of the subunits to reorient independently. This high degree of flexibility is vital for antibodies to fulfil their biological role [162, 157] and is reflected by a multitude of possible antibody conformations. This

flexibility and the related variety in conformation pose a substantial challenge for the imaging of antibodies by techniques relying on averaging over many molecules [23, 158]. Additionally, IgG antibodies are a well-known system in native electrospray ionization mass spectrometry [141, 136, 152]. In the light of these considerations, IgG antibodies are ideal candidates for exploring the potential of LEEH in imaging protein flexibility on the level of individual molecules.

Furthermore, antibodies, being at the centre of the human immune response, are of immense importance in biology and medicine, and are specifically studied with the aim of developing therapeutics [163, 164, 165]. In this respect, the antibodies' binding behaviour and interaction with other molecules is of particular interest [166, 167]. Successful imaging of antibodies on the single-molecule level could be a first step towards imaging binding behaviour and structural changes induced by interactions with protein antigens, which could help answer questions of direct medical relevance.

Among the large number of IgG antibodies, Herceptin (Trastuzumab), a monoclonal antibody that binds to the HER2-receptor and is used to treat HER2-positive forms of breast and stomach cancer, was chosen as a test system for LEEH since it has been studied extensively by native electrospray ionization mass spectrometry [136, 168, 152, 169].

This chapter focuses on the analysis of Herceptin molecules, deposited on single layer graphene in 10 separate ES-IBD experiments (see section 1.2.3). LEEH imaging revealed a multitude of conformations that can be classified into two categories: structures with distinguishable antibody subunits that can be mapped to the crystallographic model (22% of the observed molecules at a landing energy of 5 eV per charge) and compact conformations with no discernible substructure (78% of the observed molecules at a landing energy of 5 eV per charge). In order to understand the origin of these different classes of antibody conformations on the surface and their relation to the antibody's native solution structure, the influence of the whole sample preparation process of ES-IBD has to be taken into account.

The chapter will thus be split into two parts: The first part of this chapter focuses on LEEH's ability of imaging highly flexible molecules by studying the first class of conformations, which features distinguishable subunits, showing that the imaging of variability on different structural levels – the overall molecular shape and the subunits – is possible in single-molecule LEEH experiments.

The second part of the chapter examines the influence of the ES-IBD process on the protein conformation during the different stages of LEEH sample preparation and the resulting relationship of the surface conformations probed by LEEH to the native conformations in solution that is of biological interest.

The results discussed in this chapter have been published in [170].

2.1 Imaging antibody conformations with LEEH

The class of antibody conformations with distinguishable subunits is highly diverse, which makes it ideal for studying the structural variability both within the proteins and in relation to the graphene substrate. 10% of the imaged structures that fall into this
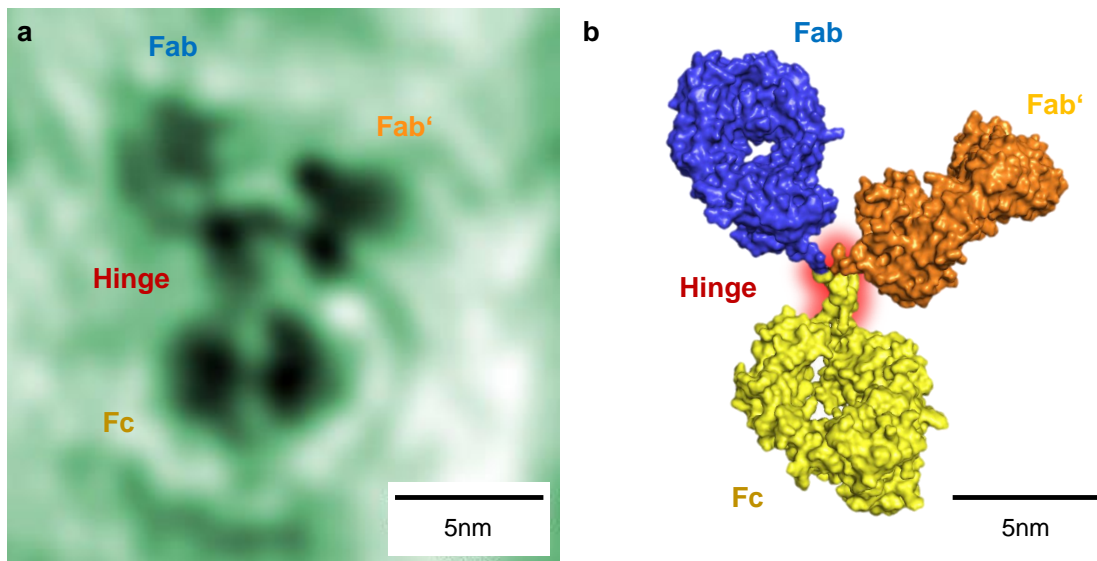
Figure 2.1: **Y-shaped conformation: a** High-resolution amplitude reconstruction of an individual antibody molecule. The size of the smallest resolved features is 5Å. The three subunits and the hinge region are clearly discernible. The hinge region consists of two strands stemming from the lower subunit, identifying said subunit as the Fc subunit. The two Fab subunits (top) exhibit different orientations, on the top left subunit, the cavity is visible (bright spot at the centre of the subunit). **b** Molecular model of the antibody conformation observed in **a**, generated from the 1IGT [82] crystallographic structure by bond rotation in the hinge region. The model and the experimental structure are in close agreement, the slight deviations can likely be attributed to additional degrees of freedom in the hinge region that are not included in the model.

class (2% of the total amount of imaged molecules) exhibit three clearly distinguishable subunits and can hence be readily recognized as the characteristic Y-shaped antibody structure known from crystallographic models [82, 171]. Fig. 2.1a shows a high resolution amplitude reconstruction of a single antibody molecule in a Y-shaped conformation with three distinguishable subunits. The three subunits are interconnected by a well-resolved hinge region consisting of two discernible peptide chain segments originating from the lower subunit. This characteristic of the hinge region identifies the lower subunit as the Fc subunit. The subunit's dimensions, which are slightly larger than those of the other two subunits, further support this classification. The other two subunits, depicted by the areas of high contrast in the upper part of Fig. 2.1a, can consequently be recognized as the Fab subunits. The image yields further structural details beyond the identification of the subunits: within the Fc subunit, two distinct dark regions are resolved, which can be interpreted as the two heavy chains composing the subunit. The area of low contrast that separates the peptide chains depicts the hydrophobic pocket hosting the glycan chains. Both the overall size of the molecule and the size of the individual subunits are in agreement with the expected molecular dimensions based on X-ray crystallography models of IgG antibodies [82, 171].

As described in section 1.1, the numerical reconstruction of LEEH holograms yields two-dimensional (2D) amplitude images that map the object's absorption. Given the imaging geometry, these 2D images can be understood as projections of the molecules
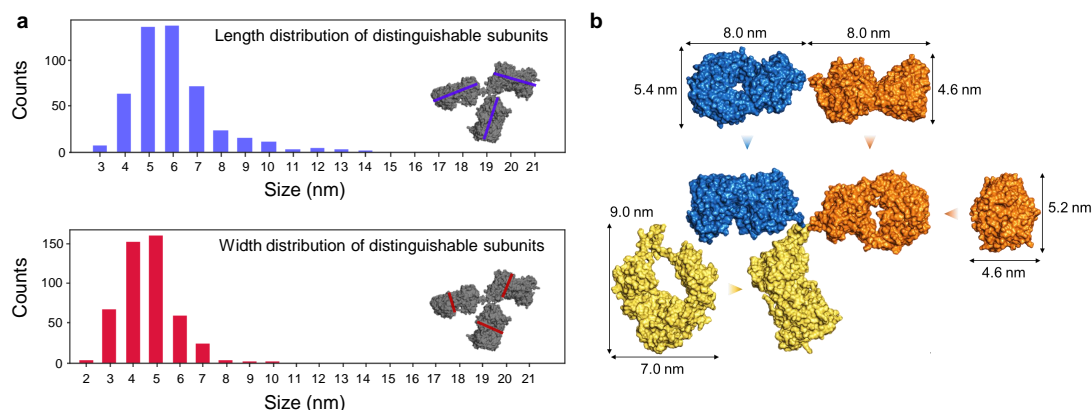
Figure 2.2: **Subunit size and shape: a** Measured size distributions of individual distinguishable subunits. Length (top) and width (bottom), longer and shorter subunit dimension, respectively, were determined at 90° angles as indicated by the blue and red lines in the insets of **a**. **b** Projections of the 1IGT crystallographic structure and its individual subunits. The directions of projection are marked by the arrows. The projection directions were selected to present maximal differences in orientation-dependent subunit shape and size. Mapping the flexibility of the hinge region, these projections provide an indication of the range of shapes and sizes the antibody subunits can assume on the surface. The experimentally measured size distribution in **a** matches the size range obtained in **b**, suggesting that the antibody subunits remain intact upon landing on the surface.

along the optical axis. This is crucial for interpreting the reconstructed images since it implies that different protein conformations, including conformations with three-dimensional components, will result in different 2D shapes. This is illustrated by the two Fab subunits in Fig. 2.1a. While they are chemically and structurally identical, they present different molecular shapes: the Fab subunit on the right appears as a compact shape, whereas the Fab subunit on the left exhibits a torus-like shape with a low-contrast area at its centre. This region of low contrast can be interpreted as the characteristic substructure of the subunit, the cavity between the two chains constituting the subunit. The diversity in subunit appearance can be explained by the subunits' orientation with respect to the graphene surface. The intrinsic flexibility of the hinge region allows for the independent rotation and reorientation of the subunits within a single molecule, which yields a large number of subunit configurations that manifest in significantly different subunit appearances in the reconstructed images. Thus, shape and size, and especially the visibility of the cavity, can vary significantly with subunit orientation. The range of possible subunit geometries and the corresponding sizes is illustrated in Fig. 2.2b, which shows projections of the subunits of the crystallographic 1IGT model in directions that display maximal differences in size and shape.

The shapes presented by the experimentally observed molecules can be replicated by rotating the subunits around the hinge region, thus changing their orientation with respect to the surface, while assuming a flexible hinge region and rigid subunits. The best-fitting structure for the molecule presented in Fig. 2.1a is shown in Fig. 2.1b. The structural variability of antibody molecules is not only manifest within a single molecule, but is further elucidated when comparing different examples of experimentally observed Y-shaped antibody molecules, as depicted in Fig. 2.1a and Fig. 2.3a-c. While they display

a large variety of subunit shapes and sizes, there are also clear variations in the overall shape and size of the molecules. These variations can be explained by the fact that the molecules are three-dimensional entities, hence subunits cannot only reorient by rotation within the substrate plane, but also rotate out of this plane. Such out-of-plane rotations decrease the apparent distances between subunits in the 2D reconstruction. Despite these variations in overall shape and subunit appearance in the antibodies imaged by LEEH, the statistical distribution of the sizes of individual subunits (Fig. 2.2a) is in agreement with the expected size distribution for the subunits of the crystallographic model [82] under different orientations (indicated in Fig. 2.2b). This congruence of measured and expected size distributions on the level of the individual antibody subunits, along with the appearance of characteristic structural features such as the cavity, suggests that the individual antibody subunits, that are in itself more rigid than the full antibody structure, retain their conformation even when the overall shape of the antibody changes.

The plethora of experimentally observed structures can be mapped to the crystallographic structure by modelling the inherent flexibility of the hinge region by rotating the φ and ψ angles of the Gly 236 residues [82] of both the heavy chains, see Fig. 2.4. These angles were chosen because of the location of the residues in the middle of the hinge region between the Fc subunit and the respective Fab subunit. Moreover, changes in the φ and ψ angles of the Gly 236 residues do not alter the disulfide bonds between the two heavy chains that stabilize the hinge region. By sampling all possible permutations of the φ and ψ angles for stepwise rotation by 60° (0°, 60°, 120°, 180°, 240°, 300°, with 0° being the angle observed in the original PDB model), $6^4 = 1296$ different configurations are generated.

Among those, however, a significant portion of structures features overlapping backbones in the subunits which eliminates these configurations from the set of possible models due to steric hindrance. The remaining 400 configurations that exhibit no relevant steric hindrance were considered for the interpretation of the experimental data.

Many projections, corresponding to different imaging directions, were created for each structure in this subset. Suitable matches between these projections and the experimentally observed molecules are found by a visual comparison (see Fig. 2.3). In most cases, several of the generated projections fit the structure presented by the experimental data (see Fig. 2.5). The remaining differences can be explained by additional degrees of freedom of the hinge region that are not incorporated in the model used to generate the projections, as well as surface interactions.

While part of the model-generated molecular configurations result in Y-shaped projections with three clearly distinguishable subunits, most of the structures produced by the model feature a degree of overlap of the antibody subunits resulting in projections in which not all three subunits are separable. This is mirrored by the experimental data, in which only 2% of the total number of molecules feature three distinct subunits, whereas 20% of the molecules display only two distinguishable subunits. To explain the appearance of the experimentally observed structures that match the projections with two distinguishable subunits, the relationship between the antibodies' adsorption geometry on the graphene and the imaging geometry of the LEEH setup has to be considered. Since the LEEH imaging process leads to a 2D image of a three-dimensional object, a Y-shaped antibody structure with three distinct subunits can only be experimentally observed when the
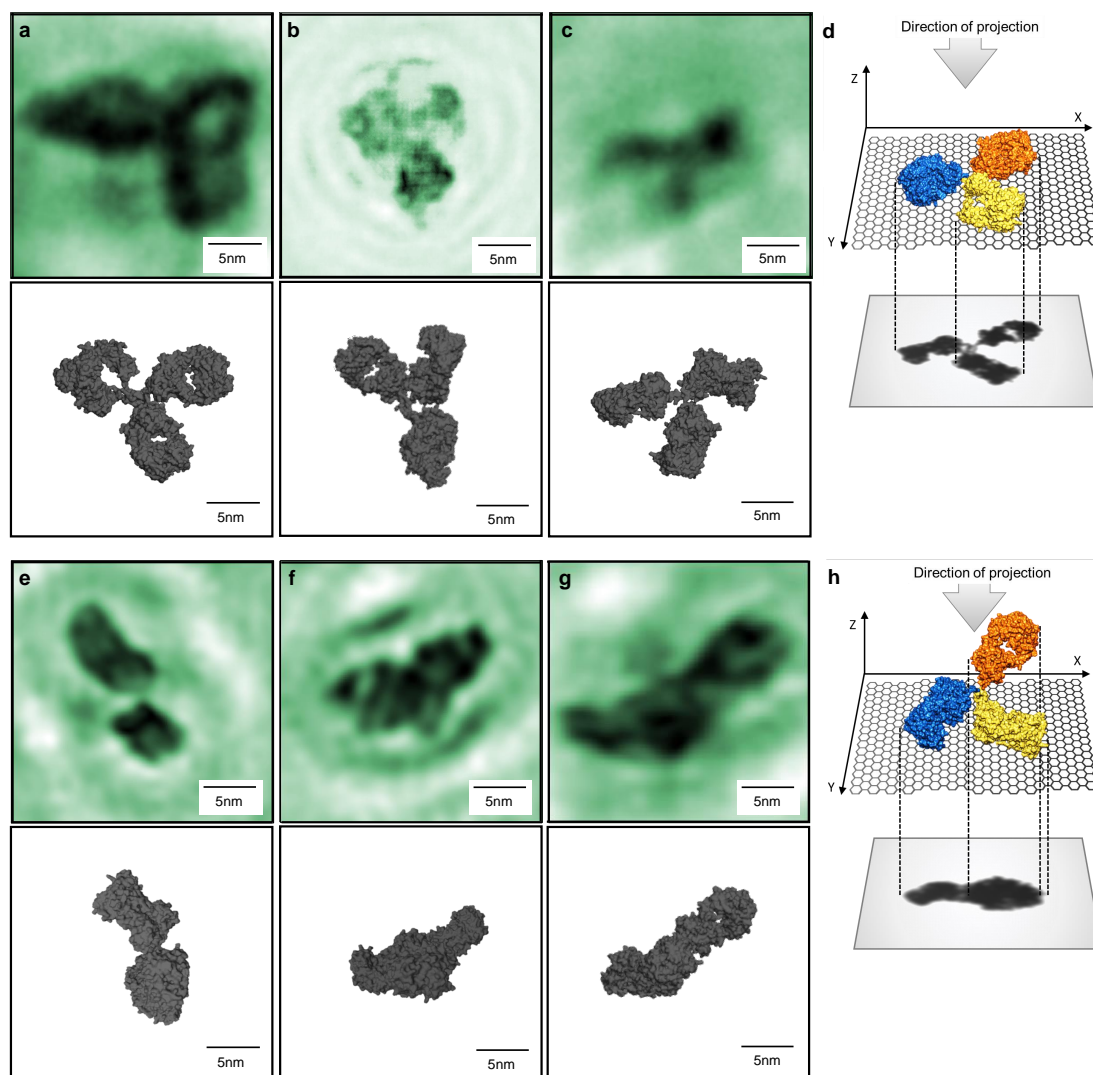
Figure 2.3: **Diversity in antibody adsorption geometry: a-c** Amplitude reconstructions of individual antibody molecules adsorbed in a flat geometry, as depicted schematically in **d**, leading to a Y-shaped appearance of the molecules, and corresponding model projections. The three subunits are clearly distinguishable in all examples, but the visibility of the hinge region and the cavities in the subunits varies: **a** hinge region not visible, cavities visible on all three subunits **b** hinge region partially visible, cavity visible on one subunit (top left) **c** hinge region partially visible, no cavities discernible.

e-g Amplitude reconstructions of individual antibody molecules in a vertical adsorption geometry as schematically depicted in **h** and matching projections obtained from the PDB structure by rotation of the subunits. Because of the ambiguity induced by the obscured third subunit in the projection, it is harder to assign molecular models to conformations featuring a vertical adsorption geometry. All three molecules display two clearly distinguishable subunits, in **f**, the third subunit is partially visible. In **e** and **g**, one of the subunits appears larger, respectively, indicating a possible contribution of the third subunit located above the other subunits as sketched in **h**. All six images correspond to a field of view of $25 \times 25\,\text{nm}^2$.

Figure 2.4: **Modelling antibody flexibility I: Theoretical model a** Ball and stick representation of the 1IGT X-ray crystallographic structure with the two Gly-236 residues highlighted in green. **b** A close-up view of the hinge region around the two Gly-236 residues (yellow rectangle in **a**). The pale blue and orange arrows represent the rotations corresponding to a change of the φ and ψ angles, respectively; the rotation axes are depicted as dashed lines. The color scheme used in both panels is: carbon in light gray, oxygen in red, nitrogen in blue and sulfur in yellow.
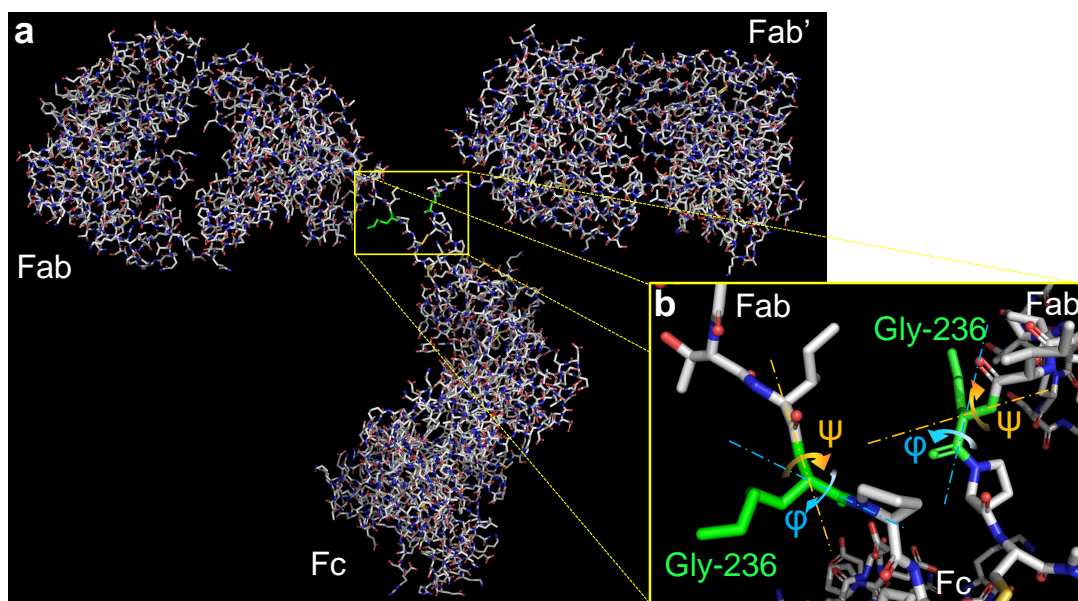
antibody molecule is absorbed on the graphene in a flat geometry, with all three subunits in the graphene plane. The antibody molecules' structural flexibility, however, extends into the third dimension. Thus, once one subunit is rotated out of the substrate plane, which, according to the model presented above, is a frequent occurrence, the subunits will overlap in the resulting projection. Hence, the structures with only two distinguishable subunits can be interpreted as antibody conformations that are adsorbed to the graphene substrate in vertical geometries, as schematically depicted in Fig. 2.3h.

In a vertical geometry, only one or two of the antibody subunits are interacting with the graphene substrate, leading to a partial (Fig. 2.3f) or complete (Fig. 2.3g) eclipse of the third subunit in a projection along the optical axis. In some cases, the third subunit is partially visible (Fig. 2.3f), in others, one of the two subunits appears larger than the other, which indicates a partial overlap of the subunits (Fig. 2.3e). In general, the subunits observed in these conformations match the intact subunits of the crystallographic model in size and shape (see Fig. 2.2), supporting the interpretation that the subunits themselves remain rigid while the antibody molecule as a whole changes its conformation. The existence of vertical antibody adsorption geometries has also been confirmed by other imaging techniques, e. g. in a combined atomic force microscopy and molecular dynamics investigation, both in water and in air [172].

As in the case of Y-shaped structures, the flexibility of the molecule results in vertical conformations that are diverse in appearance, especially regarding the separability of the subunits and the visibility of the hinge region, see Fig. 2.3e-g and Fig. 2.6. While
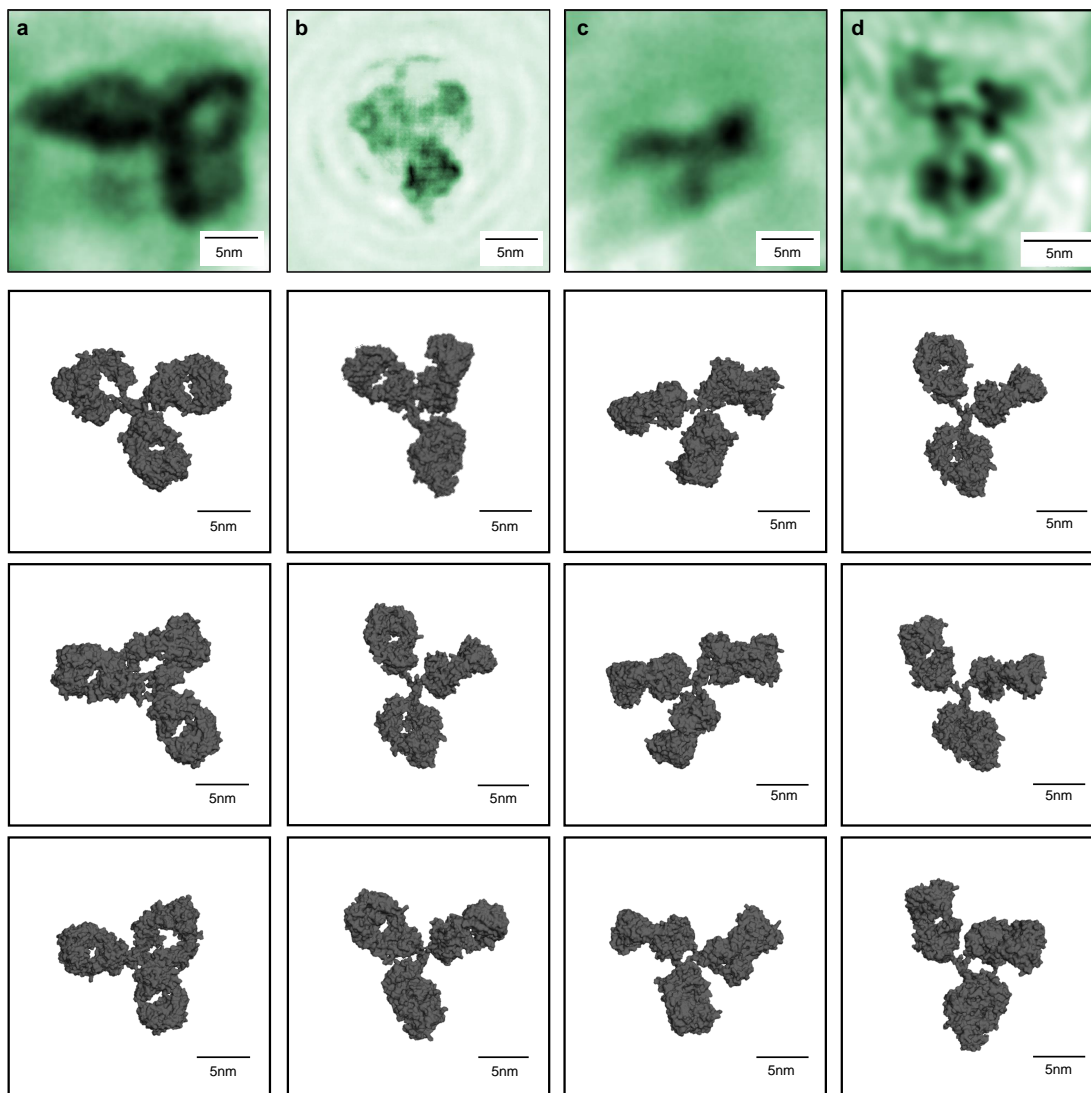
Figure 2.5: **Modelling antibody flexibility II:** Comparison of experimental data with several conformations obtained from the model described in Fig. 2.4. For each experimentally obtained image (**a** - **d**), several model conformations can be found that fit the experimental conformation in both size and shape. Despite the fact that the model only introduces modifications at one specific place in the hinge region, most experimentally observed features can be reproduced to a high degree of accuracy. The remaining discrepancies between the structures derived from the model and the imaged antibody structures can be attributed to degrees of freedom of the hinge region not included in the model and to interactions between the molecules and the graphene surface.

antibodies in flat and vertical adsorption geometries, i.e. structures with three and two distinguishable subunits, respectively, might at a first glance appear as two different classes of conformations, this analysis shows that they are to be classified as the same

Figure 2.6: **Diversity of vertical adsorption geometries:** The experimental data yields a range of conformations that can be classified as vertical geometries with two distinguishable subunits. As shown in this figure, this includes conformations with a clearly discernible hinge region (**a**, **b**), regions of lower contrast between the subunits (**c**), touching subunits (**d**, **e**) and subunits blending into each other (**f**).

type of conformation – a conformation with an extended hinge region that can be mapped to the crystallographic structure – and only differ in their adsorption geometry.

In conclusion, LEEH is capable of imaging the structural variability of antibody molecules in extended conformations, both of the molecular structure as a whole and on the subunit level. Furthermore, it cannot only elucidate variability within the substrate plane, as exemplified by the observation of the cavities as substructure of the subunits, but can partially map the flexibility towards the third dimension by distinguishing adsorption geometries.

2.2 The influence of the ES-IBD process on the conformations observed on the surface

While the structures with two distinguishable subunits are congruent with out-of-plane antibody geometries, the compact structures that make up the majority of observed molecules in the LEEH experiments do not match the projections obtained from the modified crystallographic model, neither in flat nor vertical geometry, i. e., they cannot be mapped to structures with an extended hinge region. The dimensions of the compact
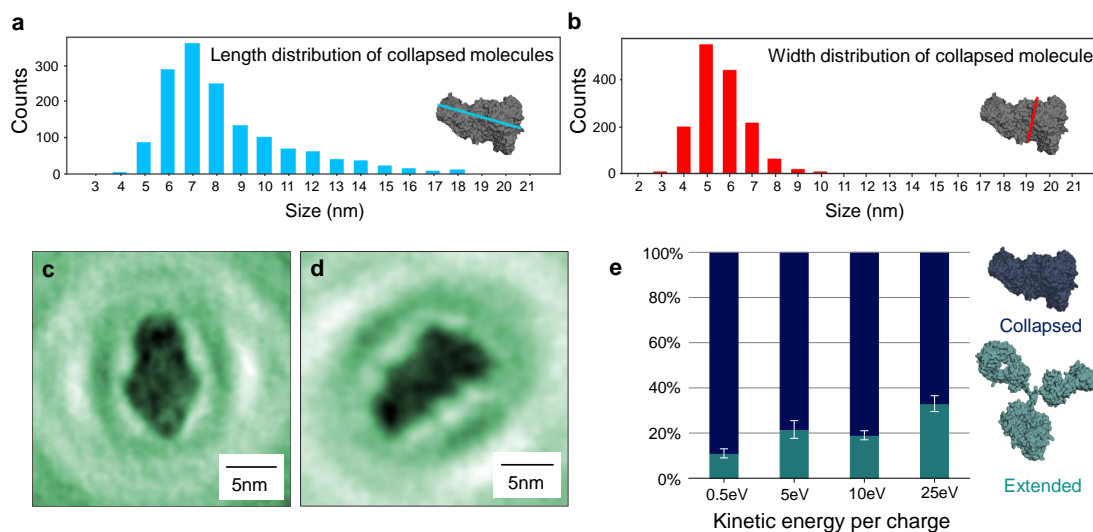
Figure 2.7: **Gas phase-related conformations: a** Length and **b** width distributions of Herceptin molecules in collapsed conformations. Length and width were measured at 90° angles as indicated by the blue and red lines in the insets of **a** and **b**. Comparison with the size distributions in Fig. 2.2a shows that the collapsed antibody conformations have larger dimensions than individual antibody subunits. **c-d** Amplitude reconstructions of antibodies in a collapsed conformation. **e** Percentage of extended and collapsed conformations for different landing energies. The amount of collapsed structures increases at lower landing energies, suggesting a close relation between the collapsed structure and the gas-phase conformation. For this analysis, a large number of molecules was evaluated at each landing energy: 378 molecules at 0.5 eV, 1259 molecules at 5 eV, 382 molecules at 10 eV and 239 molecules at 25 eV.

structures (Fig. 2.7a-b) are larger when compared to the length and width distributions of the individual subunits (Fig. 2.2a), hence these structures cannot be individual subunits. Thus, the compact conformations can be interpreted as intact antibodies with all three subunits in significant overlap, which results in the lack of observable substructure.

In order to expound the origin of the different classes of surface conformations – extended conformations with distinguishable subunits and compact conformations without discernible substructure – the impact of the ES-IBD process on the protein conformation has to be taken into consideration. The conformation of a protein could change during several steps within the ES-IBD sample preparation process: during ionization, gas-phase flight, collision with the graphene surface, and due to adsorption interaction. The transitional steps, i. e. the passage from the liquid phase into the gas phase and the landing process, are of particular importance to this investigation, since during these two processes, changes in the molecules' environment occur, which in turn potentially alter the proteins' conformation space. The most dramatic change in environment takes place during the final desolvation step upon transition into the gas phase. The loss of water has a direct effect on the hydrophobic bonding and the electrostatic interactions within the molecule which can affect the protein's stability [150]. Mass spectrometry and ion mobility measurements indicate that antibodies do not unfold or fragment when transitioning into the gas phase since the peaks in the mass spectra (see Fig. 2.8) match those of the folded protein and collisional cross section measurements are in the range
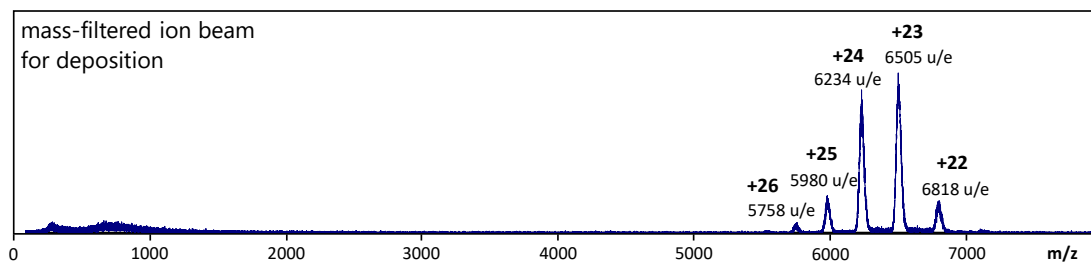
Figure 2.8: **Native mass spectrum of antibodies:** Before deposition on the single layer graphene substrate, the Herceptin molecules (molar mass 149 kDa) are characterized and mass-selected in a linear time-of-flight mass spectrometer. The native solution conditions and gentle ionization parameters ensure that the antibody molecules remain in a folded, native-like state. Each peak corresponds to a low charge state ($z = +22$ to $+26$).

expected for native-like protein conformations [136, 168, 152].

To ensure that the conformations observed on the surface originate from native solution conformations, a mass spectrum of denatured antibodies was obtained by modifying the solution conditions. This leads to significant changes in the mass spectrum (Fig. 2.9a) and an altered size distribution of the deposited molecules (Fig. 2.9b-d), confirming that the surface conformations observed after native ES-IBD sample preparation are indeed related to the native solution structure.

However, when comparing antibody collision cross sections that were experimentally obtained by ion mobility measurements with the corresponding cross sections calculated from the crystallographic model, the experimental cross sections are up to 30% smaller than the theoretical values [152]. This suggests that the tertiary structure of proteins can collapse onto itself upon transition into the gas phase [152, 173, 174, 175]. This conclusion is supported by molecular dynamics simulations of the transition into the gas phase that yields collapsed, compact gas-phase structures [152]. While compaction in the gas phase can occur for a large range of molecules, it is specifically relevant for non-globular proteins and proteins featuring flexible, elongated parts [173, 176, 152]. The highly flexible hinge region of the antibodies is primarily responsible for the collapse of antibodies in the gas phase since it allows the subunits to come into close contact [152, 173]. The compact surface conformations that do not feature distinguishable subunits can thus be interpreted as chemically intact antibodies that have collapsed in the gas phase and retained this compact conformation upon landing on the graphene surface. The comparison of the size distribution of the experimentally observed compact structures to the sizes derived from projections of simulated collapsed antibody structures provided by Hansen et al. [152] further underpins this interpretation. The simulated models generate a length range of $6 - 12$ nm and a width range of $5 - 9$ nm, which is in good agreement with the size distributions found experimentally (Fig. 2.7a-b).

In the LEEH experiments, however, a significant portion of the molecules presents extended conformations, as discussed in section 2.1, despite the strong indication that antibody molecules collapse in the gas phase. The presence of extended conformations on the graphene surface can be explained in two ways: on the one hand, the extended structures could already be present in the gas phase and retained upon landing on the surface. On the other hand, these structures could be a product of the landing process.
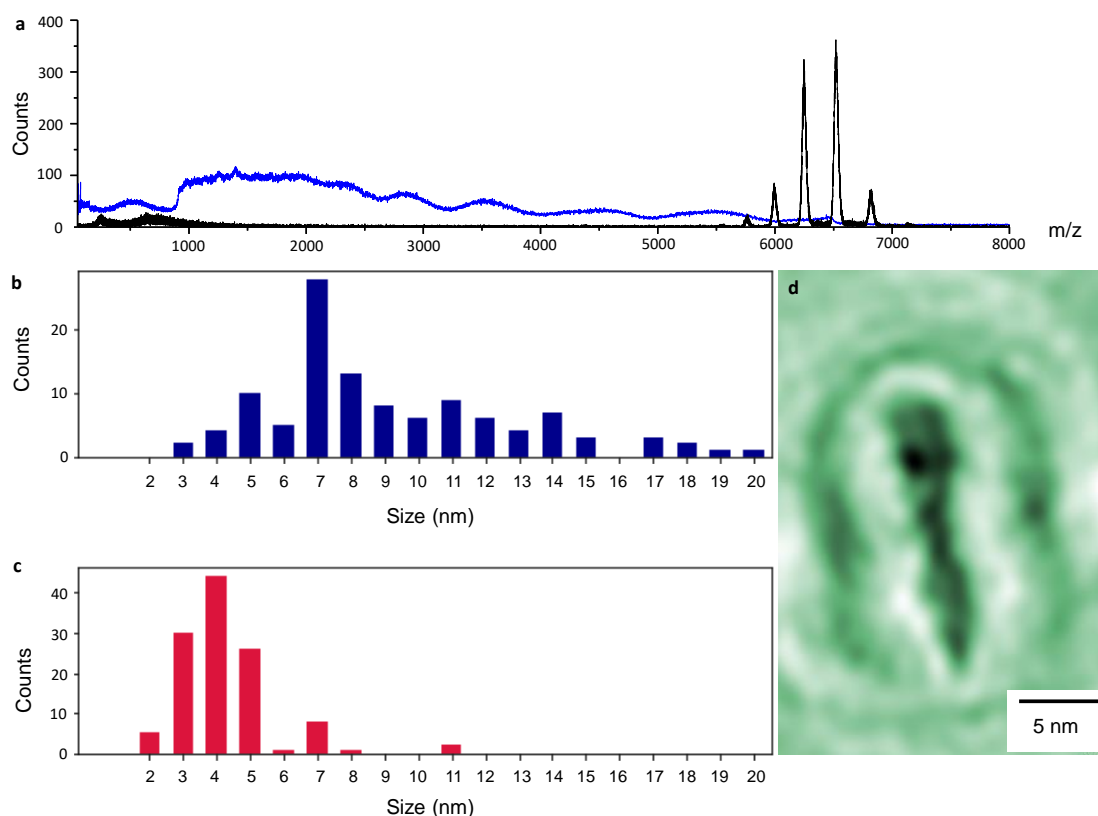
Figure 2.9: **Denaturing of antibodies: a** Mass spectra of denatured antibodies (blue) and native antibodies (black). Denaturing of the molecules in solution was achieved by adding both organic solvent and formic acid to the solution. While the native spectrum exhibits clear peaks in the high mass-to-charge range, corresponding to intact, folded molecules, the denatured spectrum shows a broad distribution in the lower mass-to-charge range, corresponding to overlapping, unresolved high charge states. Due to the limited resolution of the linear time-of-flight mass spectrometer employed to obtain these spectra, the peaks corresponding to high charge states that should appear close to each other cannot be resolved. This is exacerbated by the heterogeneity introduced into the molecular sample through denaturing, which creates a multitude of peaks in a given mass-to-charge range. **b** Length distribution of denatured antibodies imaged by LEEH. **c** Width distribution of denatured antibodies imaged by LEEH. In comparison with the size distributions for the collapsed conformations shown in Fig. 2.7a-b, the denatured molecules have a broader length distribution and a width distribution that is shifted to lower values. The heterogeneity indicated by the broad intensity distribution in the denatured mass spectrum is also reflected in these size distributions, especially in the broadened length distribution. The shift to smaller width values illustrates the loss of folding due to the denaturing of the molecules, resulting in chain-like conformations. **d** Reconstruction of a denatured antibody molecule measured by LEEH presenting in a conformation that is longer and narrower than the structures observed after deposition from native solution conditions. The heterogeneity exhibited both on the level of the mass spectrum and of the size distributions is also represented by the diversity of the appearance of the molecules observed on the surface. This implies that molecules in various states of folding, from completely unfolded to partially folded, are present in the denatured molecular ion beam.

The results from both ion mobility measurements and the molecular dynamics simulations [152] indicate that the existence of extended conformations in the gas phase is improbable,

hence it seems that the former hypothesis could not generate the observed percentage of extended structures on the surface. The absence of extended conformations from the gas phase is further reflected by the fact that the collision cross-sections that can be calculated from the models assigned to the observed Y-shaped conformations (Fig. 2.3) with the software tool IMPACT [177] are larger than the collision cross-sections obtained from ion mobility measurements ($9700 \pm 100 \, \text{Å}^2$ for the model vs. $6827 \pm 81 \, \text{Å}^2$ measured experimentally), but are close to the collision cross-sections calculated from the initial crystallographic model (e.g. $9780 \pm 30 \, \text{Å}^2$ for 1IGT [82]).

Assuming that the antibody gas-phase conformation indeed takes the form of a collapsed structure, the landing process has to allow for the re-extension of the hinge region upon impact on the surface. In general, the deposition process again involves a change of the molecule's conformation space since the landing can be described as a transition from a gas phase conformation to an adsorption conformation [143]. This transition is characterised by the conversion of translational kinetic energy into vibrational modes of both the molecule and the graphene substrate [149, 138, 148], hence kinetic energy transfer is a defining factor regarding possible surface conformations. Charge transfer and adsorption interaction also affect the surface conformation space. Since thermal energy can be excluded as a relevant influence on the surface conformations given that neither molecular motion nor diffusion are observed during the LEEH measurement, the presence of extended antibody conformations on the surface can be attributed to the dissipation of kinetic energy during the landing process or the adsorption interaction, which allows the recovery of extended conformations from collapsed structures.

Whether a molecule changes from a collapsed to an extended structure during landing depends both on the orientation of the molecule upon impact and on its kinetic energy [138, 148]. Before they reach the surface, the molecules are slowed down by the application of a retarding voltage to the graphene substrate. The default retardation to create soft landing conditions is such that the maximal landing energy is $5 \, \text{eV}$ per charge, which yields a kinetic energy at impact of around 110 - 130 eV per molecule for charge states between $+22$ and $+26$ as observed in the mass spectrum of native antibodies (see Fig. 2.8). This translates to approximately $10 \, \text{meV}$ per atom, which is less than the thermal energy at room temperature ($k_B T = 25$ meV) and hence far from a reactive collision regime [148] that could induce the breaking of bonds and result in damage to the molecules. Since the kinetic energy is mostly transferred into soft vibrational modes of the molecule, conformational changes coupled to these soft modes can occur even at low collision energy, while the primary and secondary structure remain intact. In the case of antibodies, these soft modes can lead to modifications of the hinge region. As long as the landing energy is not high enough to reach the reactive regime, the energy transfer during the landing process favours extended surface conformations with increasing deposition energy [138]. Such a dependence of the conformation of the antibody molecules on deposition energy can also be observed experimentally. When examining a range of landing energies between 0.5 and $25 \, \text{eV}$ per charge, the observed percentage of extended conformations on the surface increases from $11 \pm 2\%$ at the lowest landing energy of $0.5 \, \text{eV}$ per charge to $22 \pm 4\%$ at a landing energy of $5 - 10 \, \text{eV}$ per charge and to $33 \pm 3.5\%$ at $25 \, \text{eV}$ per charge (Fig. 2.7e). While the percentages measured at $5 \, \text{eV}$ and $10 \, \text{eV}$ are in the same range within the error ($22 \pm 4\%$ at $5 \, \text{eV}$ and $18 \pm 2\%$ at $10 \, \text{eV}$), this constitutes a clear trend relating the amount of extended structures on the surface

to the landing energy. The relatively large error bars are mostly due to the ambiguity during the manual classification of the imaged structures into extended and collapsed configurations. Next to the error calculated from the number of ambiguous molecules in the sample, the standard error $SE(\hat{p}) = \sqrt{Var(\hat{p})}$ was also calculated for each set of landing energies under the assumption of a binomial distribution of the extended and collapsed populations, respectively, the larger one of the two errors was used as the error bar in Fig. 2.7e. The fact that almost all molecules at a landing energy of $0.5\,\mathrm{eV}$ present compact conformations further underpins the conclusion that these compact structures are directly related to the gas-phase structure of the molecules since the dissipated energy per atom ($\approx 1\,\mathrm{meV}$ per atom at $0.5\,\mathrm{eV}$ per charge) is too low to induce conformational changes in most cases.

When increasing the landing energies even further, to $50\,\mathrm{eV}$ per charge ($80 - 100\,\mathrm{meV}$ per atom), the reactive regime is reached. While both extended ($25 \pm 4\%$) and collapsed ($47 \pm 4\%$) conformations can still be observed on the surface, a new class of conformations appears: approx. 20% of the structures are elongated structures without clear substructure with lengths between 20 and $40\,\mathrm{nm}$ and widths between 4 and $7\,\mathrm{nm}$. These structures could be the result of a partial unfolding of the molecules during the landing process.

The collapsed conformations that are still observed at high landing energies indicate that not only the landing energy itself, but also other factors such as the orientation of the molecule with respect to the surface upon landing [148] play an important role in determining whether the energy transfer can induce conformational changes. The molecular orientation upon impact could also affect the molecules' adsorption geometry.

This analysis shows that the conformations imaged on the surface are formed during several processes within the ES-IBD sample preparation. While the extended structures recovered during the landing process can be mapped to the crystallographic model and thus to the solution structure, the compact conformations are related to the molecules' gas-phase conformations. To some extent, the ratio of extended and collapsed conformations can be steered by tuning the parameters of the ES-IBD process, primarily the landing energy.

LEEH imaging hence allows the examination of both conformation spaces and hence could be used both for answering questions pertaining to the solution structure of flexible proteins as well as to applications of structural biology in the gas phase.

2.3   Outlook

The results presented in this chapter show that LEEH can indeed be used as a tool for single-molecule imaging of highly flexible proteins such as IgG antibodies at subnanometer resolution. While antibodies by themselves are systems of biological relevance, their interaction with other molecules is of particular interest in the context of medical research, hence the demonstration of the imaging of structural variability on the singlemolecule level could be a first step towards targeting open questions of biological and medical relevance with LEEH.

A question of interest in the development of therapeutics for inflammatory bowel diseases is the interaction of the IgG antibody Ontamalimab with MAdCAM-1 molecules [178, 179]. MAdCAM-1 (mucosal vascular addressin cell adhesion molecule 1) is a protein that plays a role in the inflammatory response by bringing leukocytes to the site of inflammation. An overexpression of MAdCAM-1 is associated with inflammatory bowel diseases [178], hence a possible therapeutic approach is to inhibit the interaction of MAdCAM-1 with the leukocytes by targeting it with the monoclonal antibody Ontamalimab.

Preliminary LEEH results show that Ontamalimab can be imaged by LEEH and exhibits very similar behaviour and statistics as Herceptin. However, it has so far been difficult to retain a bound state between Ontamalimab and MAdCAM-1 that can be observed in the mass spectrum. The mass spectra featured both peaks in the high m/z range that could be associated with the intact antibodies and a less defined, broadened distribution at lower m/z values, indicating that the bound state has not been preserved in the gas phase. This could be due to the bound state already being unstable in solution, which in turn could be related to the choice of buffer, since the binding is facilitated by the presence of salts, while in native ES-IBD, aqueous solutions without salts are preferable. These results show that the challenges regarding binding experiments with LEEH are not only limited to the imaging process and analysis, but also pertain to the sample preparation. In order to transfer a bound complex to the graphene surface intactly to enable imaging with LEEH, the conditions have to be optimized at each step of the native ES-IBD process. Still, it is likely that once a sample is successfully prepared, LEEH can contribute towards elucidating antibody binding behaviour.

3 Phase reconstruction: Theory

The exit wave produced by the scattering of an incident wave by an object, such as a protein, can in general be described as a complex-valued function, i.e. the scattering process does not only change the amplitude of the incoming wave, but also its phase. Detectors, however, can only record intensity distributions, hence the phase of the complex wave field in the detector plane cannot be directly measured. The loss of information resulting from the lack of direct phase detection is often referred to as the *phase problem* [180, 181, 182]. This is especially relevant in the context of diffraction methods such as crystallography, where knowledge of the phases is needed for the calculation of the electron density [180, 181].

A widely used approach to phase retrieval are iterative methods. Iterative methods start from either a random initial phase distribution or a suitable initial guess for the phases and use this, along with the measured intensity data, to calculate the missing phases by iteratively applying constraints to both amplitude and phase. Enforcing suitable constraints changes the reconstructed amplitudes and phases, which consequently leads to a better approximation of the object's properties in each consecutive iteration. After a number of iterations, the calculated phases converge to the actual phase shift distribution induced by the object.

The phase problem in holography is much less severe than in crystallography or other diffraction-based methods. While diffraction patterns only record intensities [38], the presence of the reference wave in holography yields relative phase information stored in the interference pattern recorded by the detector, i.e. the hologram. Thus, holography is part of a class of methods employed to measure phase information [183, 184]. However, the phase problem itself, i.e. the loss of absolute phase at the detector, persists in holography and leads to an ambiguity that impedes full phase reconstruction in in-line holography.

This ambiguity takes the form of the twin image, the complex conjugate of the object, which induces the same phase shifts as the object itself, but with opposite sign, and hence also solves the reconstruction integral when the absolute phases at the detector are not known.

In an in-line geometry, the twin image appears in a reconstruction plane that is mirror-symmetric to the object plane with respect to the source and hence creates an out-of-focus contribution to the object plane (see section 1.1.1). This ambiguity can be avoided by
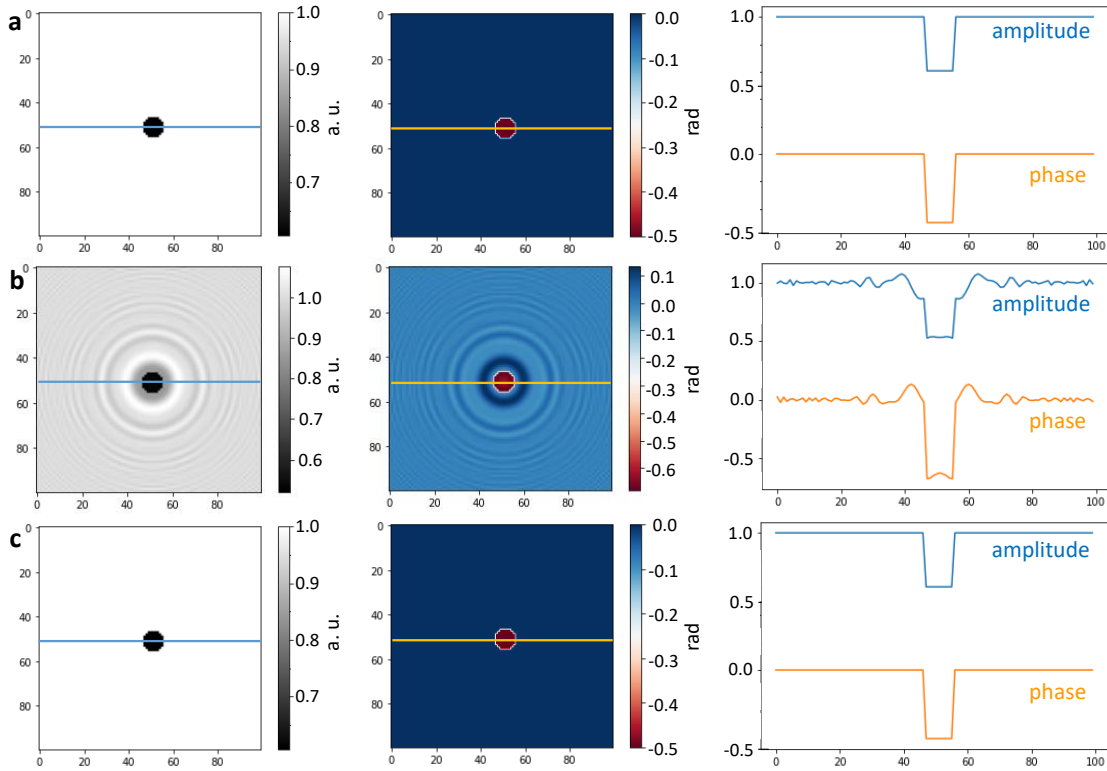
Figure 3.1: **Comparison of the reconstruction from the hologram and the complex wave field: a** Object used as input with absorption $\alpha = 0.5$ (i.e. amplitude $a = e^{-\alpha} = 0.61$ (arbitrary units)) and phase $\varphi = -0.5$ rad. The three panels show the amplitude, the phase and the line profile of each (amplitude in blue and phase in orange) through the centre of the image. **b** Amplitude and phase along with the respective line profiles of the object reconstructed from the hologram simulated from the object shown in **a**. Since the hologram is the modulus squared of the complex wave front impinging on the detector, the absolute phases in the detector plane are unknown. While the obtained reconstructions are close to the input, they suffer from imperfections. **c** Amplitude and phase along with the respective line profiles of the object reconstructed from the complex wave field in the detector plane simulated from the object shown in **a**. Knowledge of both amplitude and phase in the detector plane results in the perfect reconstruction of the input amplitude and phase in the object plane.

performing holography in an off-axis rather than an in-line set-up since the object and the twin image can be separated in Fourier space [54, 55, 49] when they do not share the same optical axis. Hence, using an off-axis geometry facilitates phase reconstruction in holography; to experimentally implement such a geometry, however, optical elements are required, which, for the case of electrons, suffer from severe aberrations in the low-energy range. Low-energy electron holography imaging is thus tied to an in-line geometry and an accurate phase reconstruction warrants the elimination of the twin image contribution in the object plane.

The effect of the presence of the twin image on the reconstruction is clearly visible when comparing the reconstruction of a simulated object from the complex wave field at the detector and from the hologram (the modulus squared of the complex wave field). While the reconstruction from the complex wave field leads to a perfect reconstruction of

amplitude and phase (see Fig. 3.1c), this is not the case for the reconstruction from the hologram (Fig. 3.1b). A perfect reconstruction of the object hence requires the knowledge of both the amplitude and phase distributions in the detector plane.

In order to understand the significance of the information encoded in the phase, the interaction of the incident beam with the object has to be considered. Scattering within an object can be due to both elastic and inelastic processes. In an elastic scattering process, the energy of the scattered electron remains constant, but its momentum changes. From a particle point of view, this is associated with a change of direction, which translates to a path difference and hence to a phase shift in the wave picture. Elastically scattered electrons retain their coherence with the reference wave and hence contribute to the interference pattern. Inelastic scattering events can have more diverse results; their defining characteristic is a transfer of energy. On the one hand, these processes can result in a reduction of beam intensity by preventing electrons from reaching the detector by absorption or high-angle scattering. On the other hand, inelastic scattering can, while inducing a change in energy, lead to virtually unchanged directions [185]. In general, inelastic scattering leads to loss or reduction of coherence [49, 185, 186]. If partial coherence can be retained, the inelastically scattered electrons can contribute to the hologram. The (partial) loss of coherence leads to a reduction in the contrast of the hologram [49], which, at least in the high-energy regime, can be directly related to the thickness of the object.

Phase imaging is of particular importance in situations of low amplitude contrast, i. e. for objects that do not absorb the incident radiation, e. g. transparent objects in light microscopy [187, 188] or weakly interacting objects in high-energy electron imaging. Still, the phase can also yield additional insights when amplitude contrast is present, as is the case for LEEH. On the one hand, phase imaging can increase spatial resolution [189], on the other hand, it can map the local electric potential [49, 190] as well as time-independent magnetic fields [49, 191, 190].

In in-line holography, the correct reconstruction of the object's phase is contingent on the elimination of the twin image from the reconstruction process. This can be achieved with the help of iterative methods such as the ones used to overcome the phase problem in crystallography.

This chapter is split into two parts: in the first part, iterative phase reconstruction algorithms in general and their application to phase retrieval in holography are discussed. The algorithm used for phase reconstruction in LEEH is described in detail and its performance is characterized with simulated examples. The second part is dedicated to the simulation of effects that likely play a role in the holographic imaging of complex structures, such as proteins, in order to provide a framework for interpreting phase data reconstructed from experimentally obtained holograms of such systems.

3.1 Iterative phase reconstruction

In 1972, Gerchberg and Saxton proposed an iterative algorithm [183] that uses two sets of intensity measurements in two planes (the image plane and the diffraction plane) to calculate the missing phase information starting from a random initial phase input.

The intensity measurements in these two planes are connected via a Fourier transform relation, hence one can propagate between the two planes by Fourier transforming the complex wave field in each plane and then constrain the amplitudes to the measured amplitude data in each plane. Both constraints are implemented once per iteration.

In holography, one could realize a similar reconstruction scheme by measuring two holograms at different source-to-sample distances. Propagating back and forth between these two holograms and applying the respective amplitude constraints would then yield the correct complex wave fields in the hologram planes from which the object could then be reconstructed [192].

The general idea of the Gerchberg-Saxton algorithm, i. e. the propagation of a wave field between two planes while enforcing constraints in each plane, can be applied to a multitude of imaging situations, including those with a single data set, e. g. a single hologram. In those cases, the constraint applied in the image plane cannot come from measured data, but has to be something else that is known about the object or the imaging system.

In general, this variant of the Gerchberg-Saxton algorithm can be delineated in the following way [193, 192] (see also Fig 3.2): the object is described as a complex-valued transmission function of the form

$$t = e^{-\alpha(r)}e^{i\varphi(r)}, \tag{3.1}$$

where $\alpha(r)$ and $\varphi(r)$ are the absorption and phase distribution of the object, respectively, and $r = (x, y)$ are the spatial coordinates in the object plane. The object's amplitude is related to the absorption as follows:

$$a(r) = e^{-\alpha(r)} \tag{3.2}$$

This representation allows us to easily separate amplitude and phase contributions and to apply constraints to each of them. In the following, phase values are given in rad, and amplitude and absorption values in arbitrary units (a. u.).

Starting from the input, which consists of the measured amplitudes in the hologram plane $A = \sqrt{I_{\text{measured}}} = \sqrt{H}$, i. e. the square root of the measured intensity distribution (the hologram H), $I_{\text{measured}} = H$, and an initial phase distribution Φ_{init}, the corresponding complex wave field at the detector is formed, $U = Ae^{i\Phi_{\text{init}}}$. This complex wave field is then propagated from the hologram plane to the object plane which yields the first estimate of the transmission function of the object, $t = ae^{i\varphi}$, with amplitude a and phase φ. If the initial guess for the phase is zero, this step results in the same reconstruction as the basic reconstruction algorithm discussed in section 1.1. Following that, the constraints in the object plane are applied, leading to a modified transmission function $t' = a'e^{i\varphi'}$. Different possible constraints in the object plane are discussed in section 3.1.1. This new transmission function is subsequently propagated to the hologram plane, yielding a new complex wave field $U' = A'e^{i\Phi'}$. The constraint in the hologram plane is applied, i. e. the calculated amplitudes A' are replaced by the measured amplitudes $A = \sqrt{H}$. The complex wave field then has the form $U = Ae^{i\Phi'}$. This is the input for the second iteration of the algorithm.

As stated above, this iterative scheme is very general and works with a variety of different constraints and imaging strategies. In the case of inline holography, the wave field propagation is carried out with Fresnel-Kirchhoff propagators, which are used to simulate

and reconstruct holograms in non-iterative reconstruction schemes as well (see section 1.1.2).
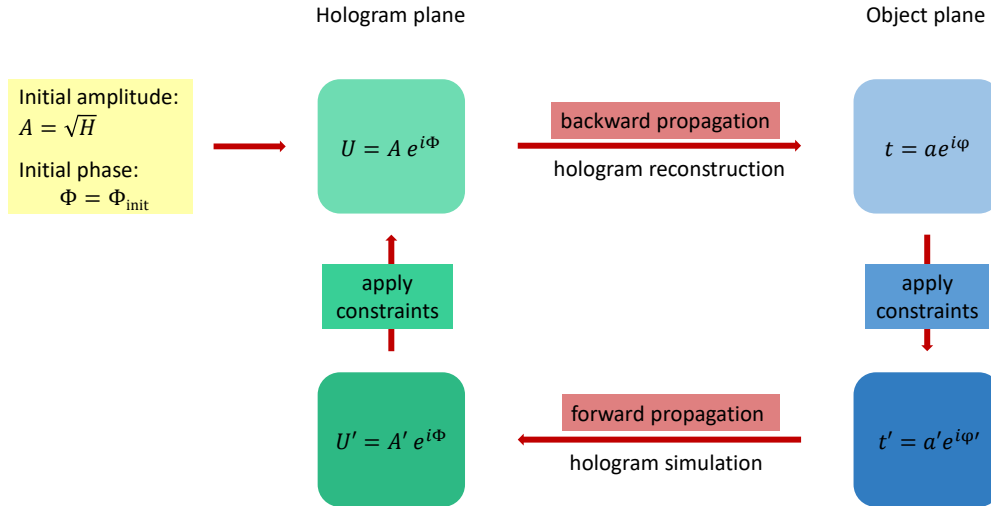
Figure 3.2: **Iterative reconstruction algorithm:** Starting from the measured amplitudes A and an initial guess for the phase Φ_{init}, the complex wave field U in the detector plane is calculated. This wave field is propagated backwards to the object plane using the propagator employed for the reconstruction of holograms recorded by spherical waves (see section 1.1.2), which results in a transmission function t. As a next step, suitable constraints are enforced, yielding a modified transmission function t'. From t', a complex wave field U' in the hologram plane is simulated by forward propagation. Subsequently, the constraints in the hologram plane are applied, which completes one iteration of the algorithm. The resulting complex wave field serves as input for the next iteration.

3.1.1   Constraints

The choice of constraints depends on the imaging situation and on what is known about the object. A type of constraint often used in iterative phase retrieval processes are support constraints [194, 195, 196, 197]. A support constraint is a mask that determines which parts of the object plane are considered background. In these regions, the complex wave field is set to zero, thereby specifying the area in which the object can be located. Support constraints work well in general, but too loose supports can be inefficient, while to tight supports can bias the outcome of the reconstruction. The constraints applied in the object plane do not generally have to involve considerations regarding the object's support, any other properties of the object that are known a priori (such as certain phase properties, symmetries etc.) can be used to formulate constraints.

Since we in general do not have a priori knowledge of the objects we study in LEEH, we choose a very general constraint, as suggested in [193]: the requirement that energy needs to be conserved during the imaging process. Energy conservation necessitates

that an interaction with the object will in general lead to reduced transmission (or in the case of a pure phase objects to unit transmission), but can never lead to increased transmission. Thus, the amplitude of the exit wave must be lower or equal to the amplitude of the incident wave. Given the relation between amplitude and absorption (eq. 3.2), this translates to a constraint on the object's absorption distribution: it has to be non-negative, $\alpha \geq 0$. Negative absorption values occurring in a transmission function can be attributed to the interference of the twin image with the reference wave and can hence be set to zero. Alternatively, the same constraint can be directly applied to the amplitude by setting amplitudes larger than 1 (corresponding to increased transmission and hence negative absorption) to 1. For these constraints to be effective, the hologram needs to be normalized by division by the background intensity. If the background intensity cannot be measured separately, this can be achieved by division by the mean of the hologram.

To obtain a perfect reconstruction of both amplitude and phase, it turns out that requiring the absorption to be non-negative is not sufficient (see Fig. 3.3a, c), and that the phase needs to be constrained in each iteration along with the amplitude. The constraint on the phase employed here involves setting the phase to zero at points of negative absorption. In a simulated example, pixels with negative absorption values in general only occur in the background since physical objects with non-negative absorption are assumed as simulation input. Negative absorption values are due to the contribution from the twin image, which most prominently appear in the form of fringes in the reconstruction of both absorption and phase. In the background, the correct phase should be the phase of the reference wave in the object plane, thus relative to the reference wave, the background phase outside of the object should remain unchanged and hence take the value zero in the reconstruction. Hence, setting the phase values of pixels with negative absorption values to zero is a phase constraint that will yield a better approximation of object and background.

Without this additional constraint on the phase, both amplitude and phase can be retrieved with much higher accuracy than with a one-step reconstruction, however, the process falls short of perfectly reconstructing both the phase and the amplitude input (see Fig. 3.3a, c). This appears to be due to the fact that the contributions originating from the twin image do not only result in pixels of negative absorption in the background, but also in pixels of incorrect positive absorption values (the correct background absorption should be zero), which can be seen in the cross sections shown in Fig. 3.3e, f. Incorrect positive absorption values cannot be corrected by the non-negativity constraint applied on the absorption level. If the phase values are left unchanged during the iterative process, positive absorption values will again be induced in the reconstruction of the next iteration, and the algorithm will ultimately converge to imperfectly reconstructed amplitude and phase images. Adding the constraint on the phase level fixes this problem since correcting the phases for pixel of negative absorption affects the phase estimate for the neighbouring pixels in the next iteration. This process results in a step-wise correction of both absorption and phase values over the whole of the image (Fig. 3.3e, f), and thereby yields perfect reconstructions of both amplitude and phase (Fig. 3.3b, d). Employing this phase constraint is especially important when dealing with objects of high absorption (see Fig. 3.3c-d), in this case, the algorithm with constraints for both absorption and phase still succeeds in perfectly reconstructing amplitude and phase,
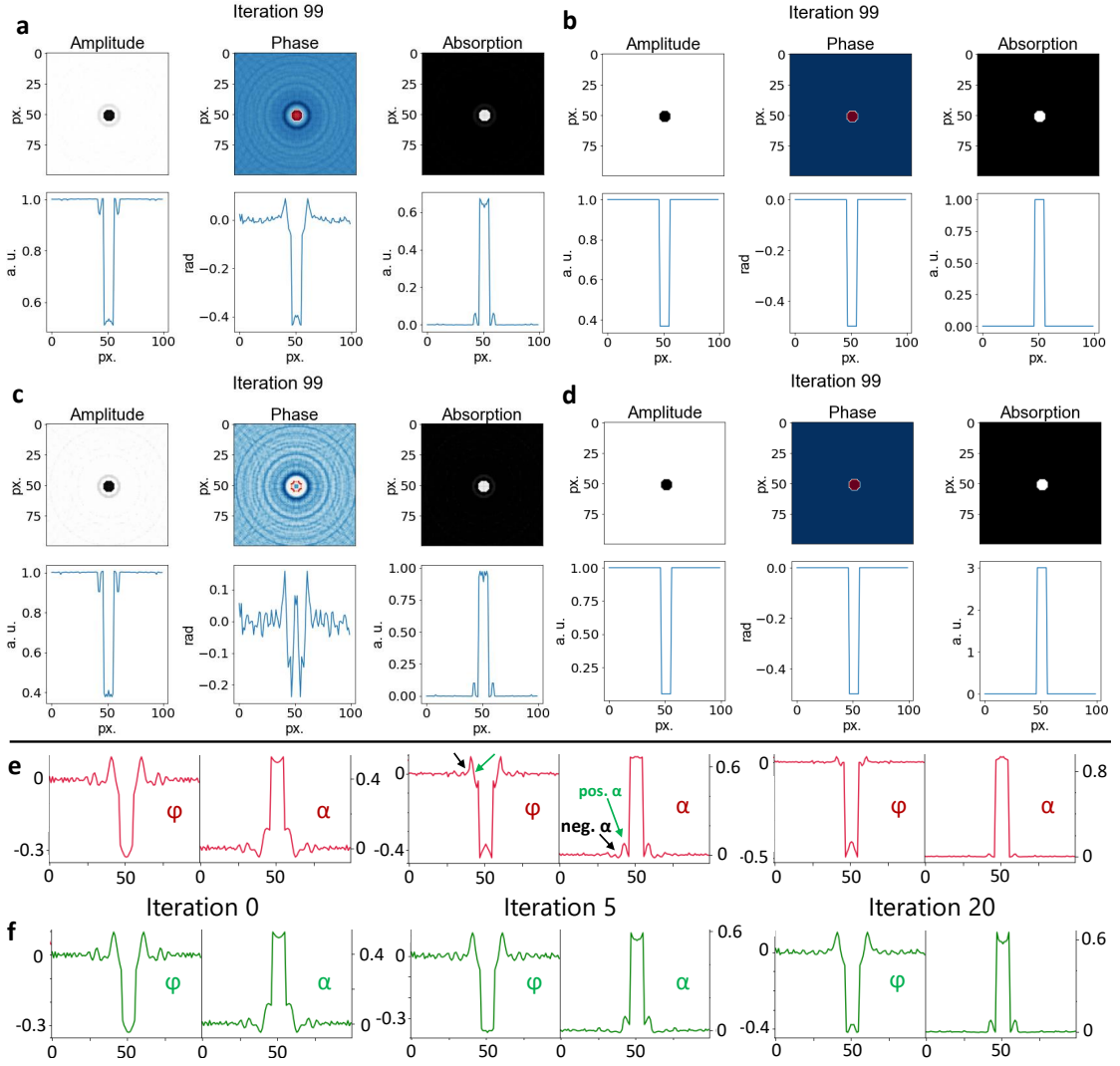
Figure 3.3: **Significance of constraining both phase and absorption: a** Iterative reconstruction of amplitude, phase and absorption of a disk of absorption $\alpha = 1$ and phase $\varphi = -0.5\,\text{rad}$ after 100 iterations. The constraint employed sets negative absorption values to zero, but leaves the phase unchanged. Neither amplitude nor phase are perfectly recovered. The imperfect reconstruction is not due to an insufficient number of iterations since the reconstruction converges to the values shown here after less than 50 iterations. **b** Iterative reconstruction after 100 iterations of the same object as used in **a**, with the additional constraint that the phase value at pixels with negative absorption values is also set to zero. The application of the constraints to both absorption and phase yields a perfect reconstruction of both amplitude and phase input after approximately 50 iterations. **c** Iterative reconstruction of amplitude, phase and absorption of a disk of absorption $\alpha = 3$ and phase $\varphi = -0.5\,\text{rad}$ after 100 iterations. The constraint employed is the same as in **a**, i.e. only the absorption is constrained. Compared to **a**, especially the phase is much harder to recover for the strongly absorbing object. **d** Iterative reconstruction after 100 iterations of the same object as in **c**, with constraints on both absorption and phase as described in **b**. Despite the strong absorption of the object, both amplitude and phase are recovered perfectly. **e** Series of line profiles of phase (φ, left) and absorption (α, right) for several iteration steps (for iteration 0, 5, and 20) of the iterative process with constraints applied to both absorption and phase as in **b** and **d**. The progression towards the correct values happens for both the pixel values of negative (indicated by black arrows in **e**) and for those of non-negative absorption values (indicated by green arrows in **e**). The latter is not the case if only the absorption constraint is used, as can be seen from the corresponding series of line profiles in **f**.

67

while the algorithm featuring only the absorption constraint yields a phase reconstruction significantly different from the phase input. This is particularly relevant when dealing with experimental data (see Chapter 4) since high absorption values frequently occur in this context.

If negative absorption would also occur within the object, potentially due to charging effects, this additional phase constraint could lead to artefacts in the reconstruction. However, assuming that the phase value corresponding to a pixel with negative absorption is incorrect, changing it to zero should not a priori induce more artefacts than keeping it at its initial value. An option to avoid this, however, would be to combine the constraints discussed here with a support constraint which allows the phase constraint only to be applied to pixels of negative absorption outside the molecule.

3.1.2   Iterative phase retrieval algorithm

With the constraints discussed in the previous section, the algorithm described here takes the following form:

1. Form an initial complex wave field U in the hologram plane from the measured amplitudes and an initial phase distribution.

2. Propagate the complex wave field to the object plane using a Fresnel-Kirchhoff propagator in a convolution approach.

3. Apply the constraints for both absorption and phase to the transmission function t that resulted from the previous step, i.e. set both absorption and phase of t to zero for pixels with negative absorption values. This yields the modified transmission function t'.

4. Propagate the modified transmission function t' to the hologram plane.

5. Apply the amplitude constraint: replace the calculated amplitude distribution by the measured amplitude distribution.

The code corresponding to these steps can be implemented as follows:

Step 1: Load and normalize hologram, take the square root, which yields the measured amplitudes, and create an initial phase distribution. Form the initial complex wave field in the detector plane from the amplitude and phase distributions.

```
1  hologram=np.load(filename)  #load hologram
2  hologram=hologram/np.mean(hologram)  #normalize hologram
3  hologram_sqrt=np.sqrt(hologram)  #measured amplitude
4  phase=np.random.rand(nx,ny)  #initialise with random phase distribution
5  hologram_field=hologram_sqrt*np.exp(1j*phase)  #form complex wave field
```

Step 2: Propagate the complex wave field `hologram_field` to the object plane using a Fresnel-Kirchhoff propagator in a convolution approach to calculate the transmission function `t`. `area` is the object size, `z0` is the source-to-sample distance, and `Lambda` is

the electron wavelength. The Fresnel-Kirchhoff propagator `Propagator` is defined as in section 1.1.3.

```
1  t=IFT2D(FT2D(hologram_field)*Propagator(area, z0, hologram_field, Lambda))
```

Step 3: Apply the constraints in the object plane.

```
1   #calculate amplitude and absorption from transmission function
2   amplitude=abs(t)
3   absorption=-np.log(amplitude)
4   #calculate phase from transmission function
5   Phase=np.angle(t)
6   #apply constraints
7   for i in range (0, nx):
8           for j in range (0, ny):
9               if absorption[i][j]<0:
10                  absorption[i][j]=0
11                  Phase[i][j]=0
12  #calculate modified transmission function
13  new_amplitude=np.exp(-absorption)
14  t_new=new_amplitude*np.exp(1j*Phase)
```

Step 4: Propagate the modified transmission function `t_new` to the hologram plane.

```
1   hologram_field_new=IFT2D(FT2D(t_new)*np.conjugate(Propagator(area, z0, t_new,
      Lambda)))
```

Step 5: Apply the constraints in the hologram plane.

```
1   #calculate updated wave field in hologram plane
2   phase=np.angle(hologram_field_new)
3   hologram_field=hologram_sqrt*np.exp(1j*phase)
```

In step 1 of the algorithm outlined here, the phase has been initialised with an array of uniformly distributed random numbers in the range $[0, 1)$, implemented by the NumPy function `rand()`. In order to allow for both positive and negative initial phase values, the NumPy function `randn()` can alternatively be used, which yields an array of normally distributed random numbers with mean 0 and variance 1. Instead of using a random initial phase distribution, different initial guesses can also be used and can lead to a faster convergence of the algorithm, especially when the initial guess reflects known properties of the object's phase distribution. Since the phase of the background should not change with respect to the phase of the reference wave during holographic imaging, the background phase can be set to a constant value. Considering that the background pixels constitute the majority of the image, using an array of zeros as the initial guess for the phase distribution (NumPy function `zeros()`), is thus a viable guess.

Fig. 3.4 shows a comparison of the iterative phase reconstruction of a disk of absorption $\alpha = 0.5$ and phase $\varphi = -0.5\,\text{rad}$ for different initial phase inputs, generated by a uniform distribution of random numbers using the function `rand()` (Fig. 3.4a), a normal distribution of random numbers created by the function `randn()` (Fig. 3.4b) and an array of zeros (Fig. 3.4c). While the results after the first iteration (Fig. 3.4a-c, left panels, labelled Iteration 0 since Python uses zero-based numbering, which results in the first element of a sequence being labelled with the index 0, consequently the result after 100 iterations is labelled Iteration 99) differ strongly, in all three cases, the input values are retrieved in less than 100 iterations: in approximately 55 iterations for the uniformly
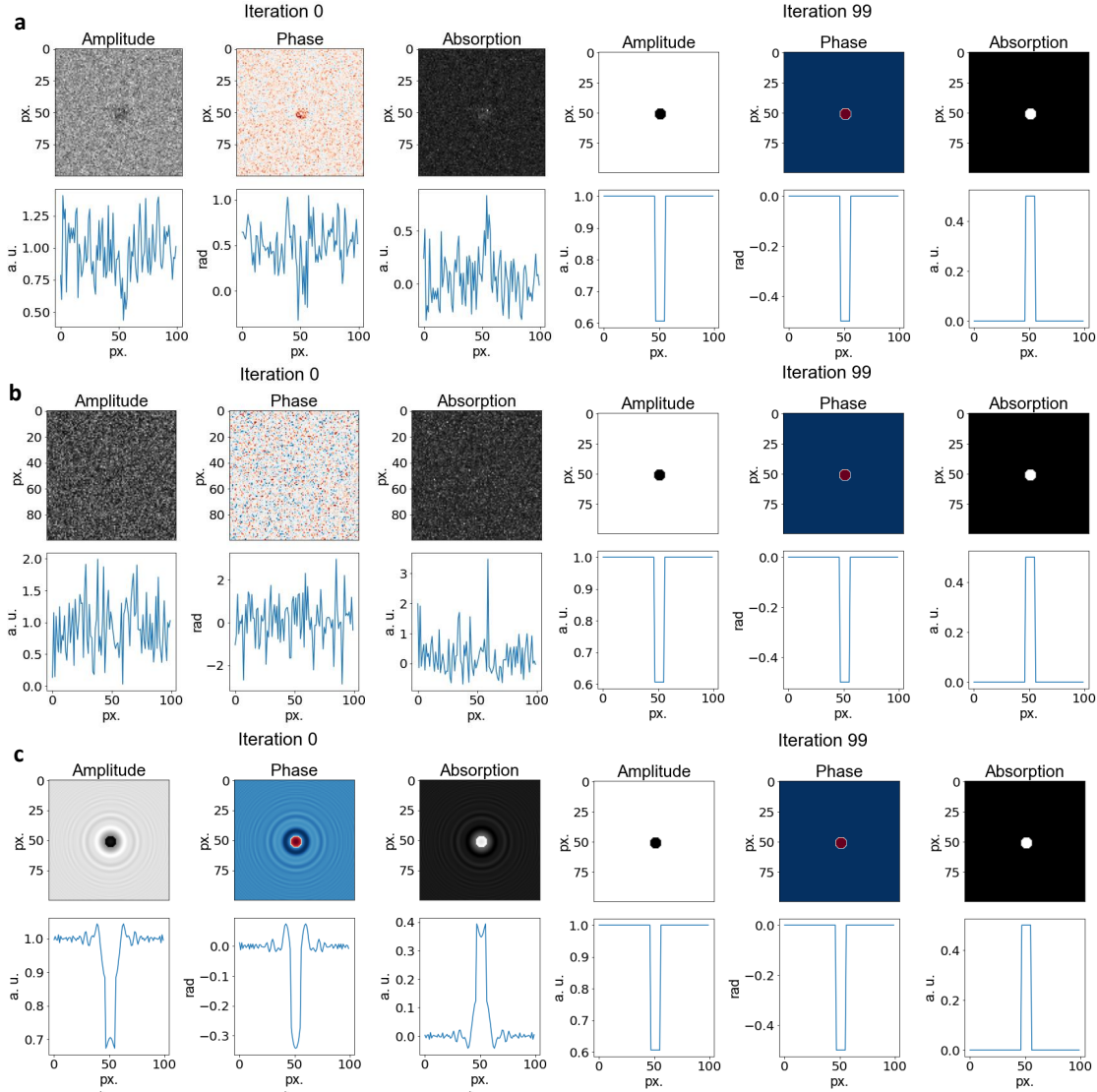
Figure 3.4: **Independence of the reconstruction of the initial phase distribution:**
a Iterative reconstruction of a disk with absorption $\alpha = 0.5$ and phase $\varphi = -0.5\,\text{rad}$ with a uniform distribution of random numbers generated by the NumPy function `rand()` as initial phase distribution. The input is perfectly reconstructed after approximately 55 iterations. The error between input and reconstruction after 100 iterations is $E_a = 3.5 \times 10^{-18}$ for the amplitude and $E_\varphi = 1.6 \times 10^{-16}$ for the phase. **b** Iterative reconstruction of a disk with absorption $\alpha = 0.5$ and phase $\varphi = -0.5\,\text{rad}$ with a normal distribution of random numbers generated by the NumPy function `randn()` as initial phase distribution. The input is perfectly reconstructed after approximately 75 iterations. The error between input and reconstruction after 100 iterations is $E_a = 3.6 \times 10^{-15}$ for the amplitude and $E_\varphi = 2.4 \times 10^{-15}$ for the phase. **c** Iterative reconstruction of a disk with absorption $\alpha = 0.5$ and phase $\varphi = -0.5\,\text{rad}$ with an array of zeros as initial phase distribution. The input is perfectly reconstructed after approximately 35 iterations. The error between input and reconstruction after 100 iterations is $E_a = 1.4 \times 10^{-21}$ for the amplitude and $E_\varphi = 5.1 \times 10^{-22}$ for the phase.

In all reconstructions shown in this figure, both the absorption and the phase are constrained in the object plane. While the reconstruction after one iteration differs significantly depending on the initial input, all examples converge to the input object in less than 100 iterations.

distributed random input, in approximately 75 iterations for the normally distributed random input, and in approximately 35 iterations for the zero-phase input.

For simulated input, the quality of the reconstruction can be evaluated by calculating the error between the input amplitude and phase and the reconstructed amplitude and phase. This can be done by calculating and summing the squared pixel-by-pixel differences, yielding the following error measures [193]:

$$E_a = \frac{1}{N^2} \sum_{j,k=1}^{N} |a_{in} - a_{rec}|^2 \tag{3.3}$$

$$E_\varphi = \frac{1}{N^2} \sum_{j,k=1}^{N} ||\varphi_{in}| - |\varphi_{rec}||^2, \tag{3.4}$$

where a_{in} and φ_{in} are the input amplitude and phase, a_{rec} and φ_{rec} are the reconstructed amplitude and phase, and N is the number of pixels of the object in each dimension. The summation is carried out over all pixels (j, k). The additional absolute values in the calculation of the phase error occur because the phase values can be negative.

The errors calculated for the three cases shown in Fig. 3.4 are $E_a = 3.5 \times 10^{-18}$ and $E_\varphi = 1.6 \times 10^{-16}$ for the reconstruction with the uniformly distributed random phase input as shown in Fig. 3.4a, $E_a = 3.6 \times 10^{-15}$ and $E_\varphi = 2.4 \times 10^{-15}$ for the case of the normally distributed random phase input (Fig. 3.4b), and $E_a = 1.4 \times 10^{-21}$ and $E_\varphi = 5.1 \times 10^{-22}$ for the initial phase input in the form of an array of zeros (Fig. 3.4c). While the errors are not exactly the same, they are in a similar range, this error range corresponds to a very accurate reconstruction of the input as corroborated by the plots and cross sections in Fig. 3.4.

As demonstrated by Fig. 3.4, the algorithm described here, which constrains both absorption and phase in the object plane, can reconstruct the input object independently of the initial phase distribution. In contrast, an algorithm that does not constrain the phase on the object level, but only employs the non-negativity constraint for the object's absorption, exhibits a significantly different performance depending on the initial phase input (see Fig. 3.5). If the initial guess for the phase is a constant distribution, such as an array of zeros or ones, the object is recovered to a high degree of accuracy (Fig. 3.5c), although not perfectly, as discussed in Fig. 3.3. The corresponding errors are $E_a = 2.5 \times 10^{-12}$ and $E_\varphi = 1.7 \times 10^{-9}$. While using a uniformly distributed random input yields a noisy reconstruction of the object (Fig. 3.5a, $E_a = 5.2 \times 10^{-9}$ and $E_\varphi = 2.2 \times 10^{-5}$), a normally distributed random input does not result in a recognisable reconstruction of the object (Fig. 3.5b, $E_a = 7.1 \times 10^{-8}$ and $E_\varphi = 3.1 \times 10^{-5}$). These errors are several orders of magnitude larger than the errors in the iterative reconstructions calculated by applying constraints to both amplitude and phase.

This further indicates that the phase constraint should be enforced to obtain reliable reconstructions. In the following, "iteratively reconstructed" will be taken to mean iteratively reconstructed with the presented algorithm employing both the amplitude constraint and the phase constraint.

In summary, the constraint requiring non-negative absorption was shown to be effective in reconstructing simulated holograms, especially when combined with an additional phase constraint enforced on the pixels targeted by the absorption constraint. The iterative

algorithm employing both constraints uniquely reconstructs the input, independently of the initial guess used for the phase distribution.
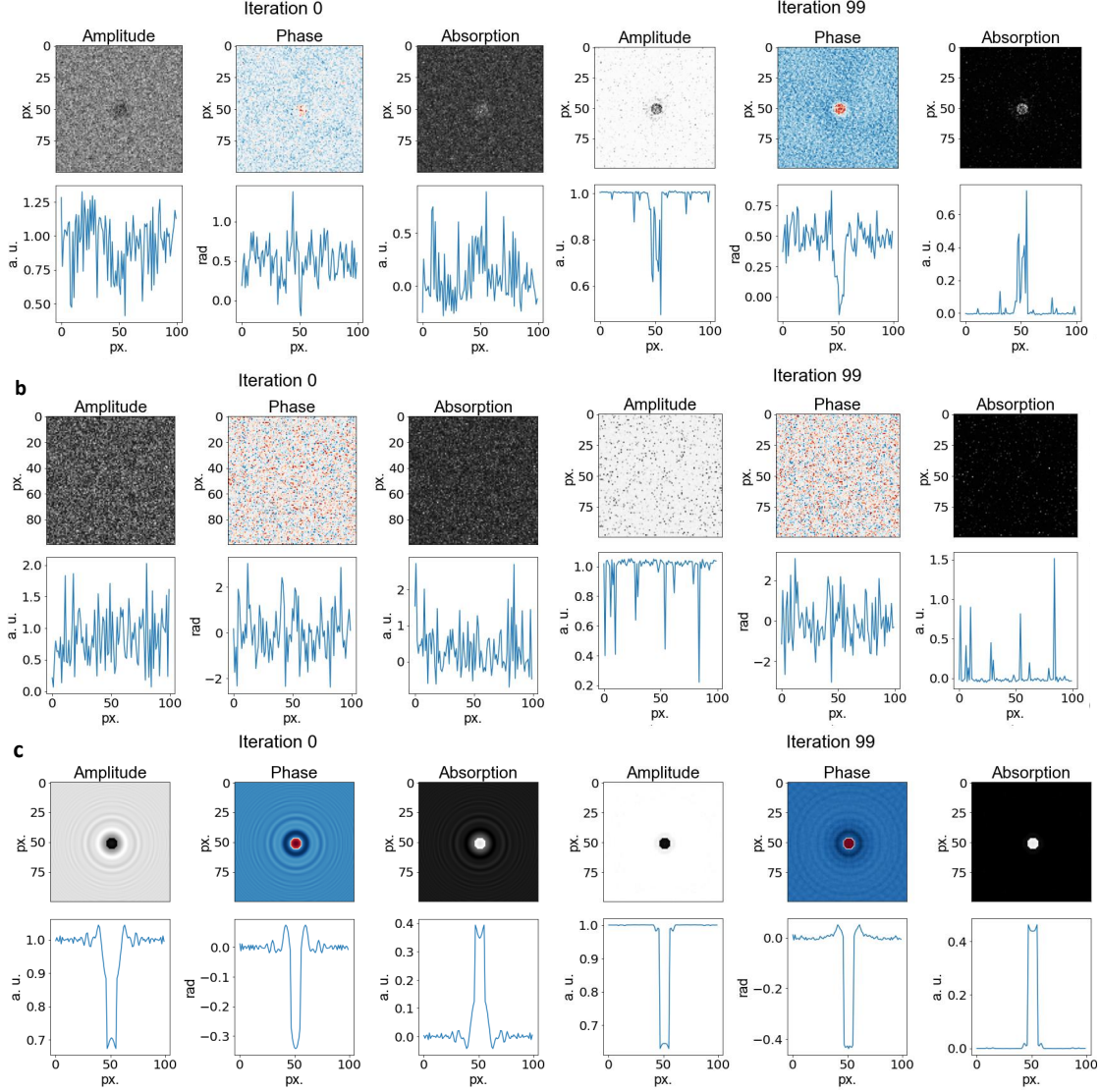
Figure 3.5: **Dependence of reconstruction success on initial phase distribution for an algorithm without a phase constraint in the object plane: a** Iterative reconstruction without phase constraint in the object plane of a disk with absorption $\alpha = 0.5$ and phase $\varphi = -0.5\,\text{rad}$ with a uniform distribution of random numbers generated by the NumPy function `rand()` as initial phase distribution. While some of the properties of the object can be recognised in the reconstruction, it is very noisy. **b** Iterative reconstruction without phase constraint in the object plane of a disk with absorption $\alpha = 0.5$ and phase $\varphi = -0.5\,\text{rad}$ with a normal distribution of random numbers created by the function `randn()` as initial phase distribution. The object cannot be recovered at all. **c** Iterative reconstruction without phase constraint in the object plane of a disk with absorption $\alpha = 0.5$ and phase $\varphi = -0.5\,\text{rad}$ with an array of zeros as initial phase distribution. The object is not reconstructed perfectly, but with a high degree of accuracy. The error between input and reconstruction after 100 iterations is $E_a = 2.5 \times 10^{-12}$ for the amplitude and $E_\varphi = 1.7 \times 10^{-9}$ for the phase.

3.2 Characterization of the iterative phase reconstruction algorithm

While the previous section presented an iterative phase reconstruction algorithm based on the elimination of twin image contributions from the reconstruction and showed that it accurately retrieves both the amplitude and phase input of a simulated object, this section focuses on the characterization of the algorithm. For this, the algorithm is applied to a variety of simulated objects, for example pure phase objects, which change the phase of the incident wave, but not its amplitude, and strongly absorbing objects, and the algorithm's performance is evaluated. The influence of factors such as the ratio of the object size to the size of the illuminated area, the amount of fringes in the hologram, the physical size of the object, and noise will be discussed. This is of relevance for identifying possible sources of artefacts or poor performance of the algorithm, which is important for the application of the algorithm to experimental data, for which an exact input is not known.

Before turning to the limiting cases of pure phase objects and purely absorbing objects, the general case of an object that induces a change in both the amplitude and the phase of the reference wave shall be examined in more detail. The transmission function of such an object can be written as

$$t(x,y) = e^{-\alpha(x,y)} e^{i\varphi(x,y)}, \tag{3.5}$$

where both α and φ are non-zero. While φ can assume both positive and negative values, α is always positive for physical objects.

Figures 3.6 and 3.7 show the iterative reconstruction after 100 iterations of objects that exhibit both absorbing and phase-shifting properties of various degrees. In all cases, the reconstructed absorption, amplitude and phase values match the input values. Fig. 3.6 shows the reconstruction of a disk of radius 5 pixels with uniform amplitude and phase distributions for the input amplitudes $\alpha = 0.1$, $\alpha = 0.5$, $\alpha = 1$ and $\alpha = 3$, spanning a range from almost full transmission to strong absorption ($a = 0.9, 0.61, 0.37, 0.05$), for input phases of both $\varphi = 1 \, \text{rad}$ and $\varphi = -1 \, \text{rad}$. The correct retrieval of the input for all absorption values considered shows that the algorithm performs well for simulated objects that are both absorbing and phase-shifting over a large range of absorption values. This implies that the code is applicable to a wide variety of physical objects, from weakly absorbing to strongly absorbing. The extreme cases of zero absorption (pure phase objects) and strong absorption will be discussed in the following sections.

To show that the algorithm can handle not only a large range of absorption values, but also a large range of phase values, Fig. 3.7 shows iteratively reconstructed disks with uniform absorption $\alpha = 1$ and uniform phase shifts $\varphi = \pm 0.1$, $\varphi = \pm 0.5$, $\varphi = \pm 3$ and $\varphi = \pm 4 \, \text{rad}$. The limiting case $\varphi = 0 \, \text{rad}$, corresponding to a purely absorbing object, will be considered in section 3.2.1. Again, all input values are correctly reconstructed, thus confirming the adequacy of the algorithm regarding the reconstruction of phases in the whole domain $[-\pi, \pi] \, \text{rad}$.

Since the phase is a 2π-periodic function defined on the domain $[-\pi, \pi] \, \text{rad}$, the phase reconstructed from input phases larger than $\pi \, \text{rad}$ or smaller than $-\pi \, \text{rad}$ is mapped
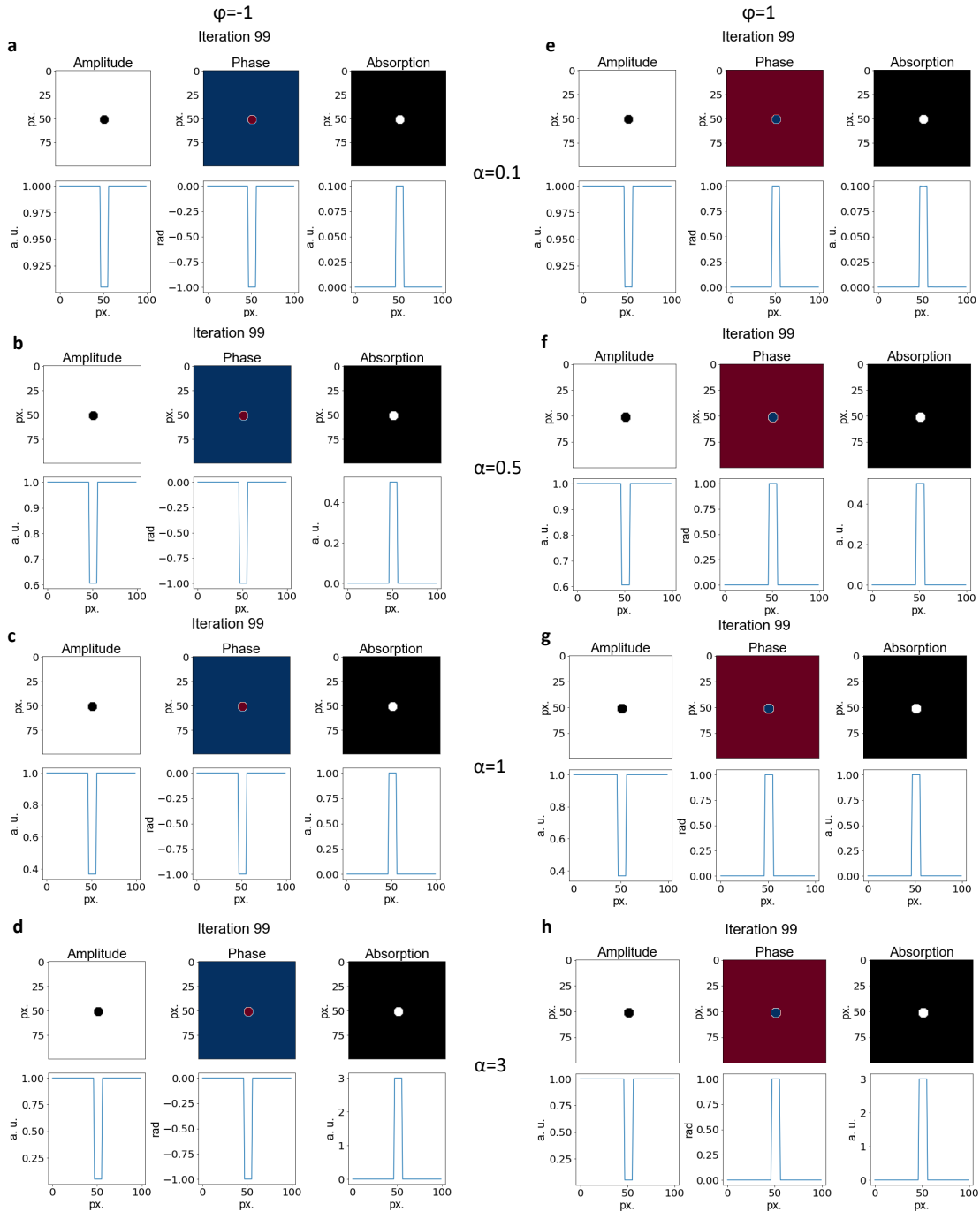
Figure 3.6: **Iterative reconstruction for different input amplitudes:** Iterative reconstruction of a disk of radius 5 pixels after 100 iterations for different input amplitudes, $\alpha = 0.1$ (**a**, **e**), $\alpha = 0.5$ (**b**, **f**), $\alpha = 1$ (**c**, **g**) and $\alpha = 3$ (**d**, **h**). For each amplitude, the object has been reconstructed for input phase values of -1 rad (**a-d**) and $+1$ rad (**e-h**). For all combinations, the input values of absorption, amplitude and phase are correctly reconstructed.

onto the interval $[-\pi, \pi]$ rad, as shown in Fig. 3.7d and h. The input phase -4 rad is reconstructed as approximately $+2$ rad (Fig. 3.7d), whereas the input phase $+4$ rad

results in a reconstructed phase of approximately $-2\,$rad.

Figure 3.7: **Iterative reconstruction for different phase inputs:** Iterative reconstruction of a disk of radius 5 pixels and amplitude $\alpha = 1$ after 100 iterations for different input phases, $\varphi = \pm 0.1$ (**a**, **e**), $\varphi = \pm 0,5$ (**b**, **f**), $\varphi = \pm 3$ (**c**, **g**) and $\varphi = \pm 4\,$rad (**d**, **h**). The reconstructions for negative (positive) φ are shown in the panels on the left (right). For all combinations, the input values of absorption, amplitude and phase are correctly reconstructed. Because of the 2π-periodicity of the phase, phase values $|\varphi| > \pi\,$rad are mapped onto the interval $[-\pi, \pi]\,$rad, as exemplified in **d** and **h**.

So far, all examples consisted of an object with a uniform phase and absorption distribution over the whole extent of the object. Fig. 3.8 shows examples, which demonstrate that the algorithm can also adequately reconstruct more complex objects with non-uniform amplitude and phase distributions. The examples shown in Fig. 3.8 include a disk of uniform amplitude with a step change in phase from -1 to $+1\,\mathrm{rad}$ (Fig. 3.8a), a disk with a step change in phase from -1 to $+1\,\mathrm{rad}$ and a step change in absorption from 1 to 0.5 (Fig. 3.8b), a Gaussian object with continually changing amplitude and phase (Fig. 3.8c), and two overlapping disks, each with absorption 0.5 and phase $+0.5\,\mathrm{rad}$ (Fig. 3.8d).
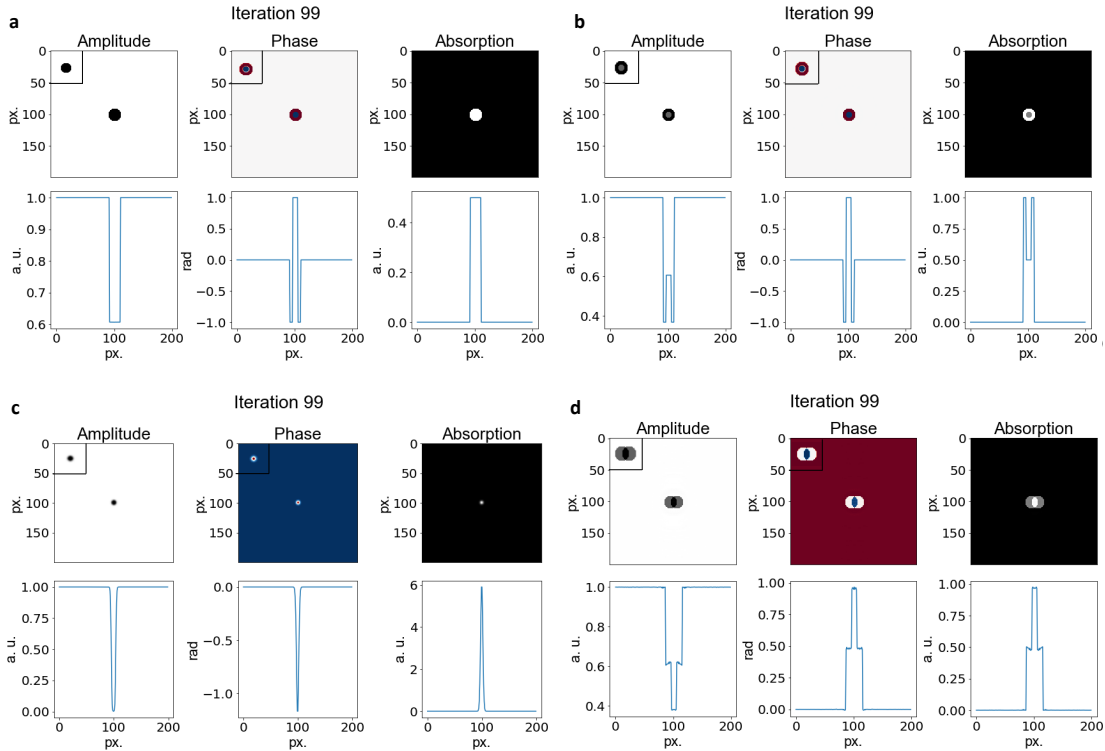
Figure 3.8: **Iterative reconstruction of more complex inputs: a** Iterative reconstruction after 100 iterations of an object with a uniform amplitude of $\alpha = 0.5$ and a step change in phase from $\varphi = -1$ to $\varphi = 1\,\mathrm{rad}$. **b** Iterative reconstruction after 100 iterations of an object with a step change in absorption from 1 to 0.5 and a step change in phase from $\varphi = -1$ to $\varphi = 1\,\mathrm{rad}$. **c** Iterative reconstruction after 100 iterations of a Gaussian object with continually changing absorption from $\alpha = 0$ to $\alpha = 6$ and continually changing phase from $\varphi = 0$ to approximately $\varphi = -1\,\mathrm{rad}$. **d** Iterative reconstruction after 100 iterations of an object composed of two overlapping disks, each of absorption $\alpha = 0.5$ and phase $\varphi = 0.5\,\mathrm{rad}$. The input is shown in the insets of each panel.

3.2.1   Purely absorbing objects and pure phase objects

At the electron energies employed in low-energy electron holography, most physical objects will change both the amplitude and the phase distribution of the incident reference wave. To characterize the algorithm, however, it is relevant to also consider the limiting cases of purely absorbing objects that do not induce a phase shift by interacting with the reference wave, and pure phase objects that only induce a phase shift, but leave the amplitude unaltered.

Purely absorbing objects

Purely absorbing objects are characterized by a transmission function with a non-zero positive absorption distribution $\alpha(x, y)$ and zero phase shift $\varphi(x, y) = 0$, i. e.

$$t_{\text{absorbing}}(x, y) = e^{-\alpha(x,y)} e^{i\varphi(x,y)} = e^{-\alpha(x,y)}. \tag{3.6}$$

Fig. 3.9 shows the result of the iterative reconstruction of a disk with absorption $\alpha = 1$ and phase $\varphi = 0$. The reconstructions of the amplitude and absorption converge to the input values in less than 100 iterations, i. e. on a similar scale as the example of an object with both non-zero absorption and non-zero phase as shown in Fig. 3.3. After 100 iterations, the remaining deviations between the reconstructed phase distribution and the input phase distribution are in the order of 10^{-5} (Fig. 3.9a). The remaining deviation decreases with an increasing number of iterations, to 10^{-9} after 200 iterations, 10^{-12} after 300 iterations, and 10^{-15} after 500 iterations (Fig. 3.9b). Iteration numbers larger than 500 do not significantly improve the result. This is likely the case since the reconstructed phase after 500 iterations (Fig. 3.9e) is very close to the phase distribution reconstructed in a single-step reconstruction with the complex wave field in the detector plane as input (Fig. 3.9d). Since even the reconstruction from the known phases shown in (Fig. 3.9d) yields small deviations from the input, this implies that an iterative approach will come up against the same boundaries.

The error calculated with equations (3.3) and (3.4) are $E_a = 7.6 \times 10^{-16}$ and $E_\varphi = 6.1 \times 10^{-16}$ after 100 iterations, and $E_a = 2.8 \times 10^{-34}$ and $E_\varphi = 5.8 \times 10^{-34}$ after 500 iterations.

The same results hold for purely absorbing objects of different absorption values, both for low absorption ($\alpha = 0.5$) and high absorption values ($\alpha = 3$ and $\alpha = 6$). In all cases, the phase reconstruction converges to the reconstruction from the complex wave field after approximately 500 iterations. While the absolute value of the deviations is higher in both the iteratively reconstructed phase and the phase reconstructed from the complex wave field for strongly absorbing objects, for weakly absorbing objects, the number of iterations needed to converge to the phase reconstructed from the complex wave field is slightly larger than for strongly absorbing objects.
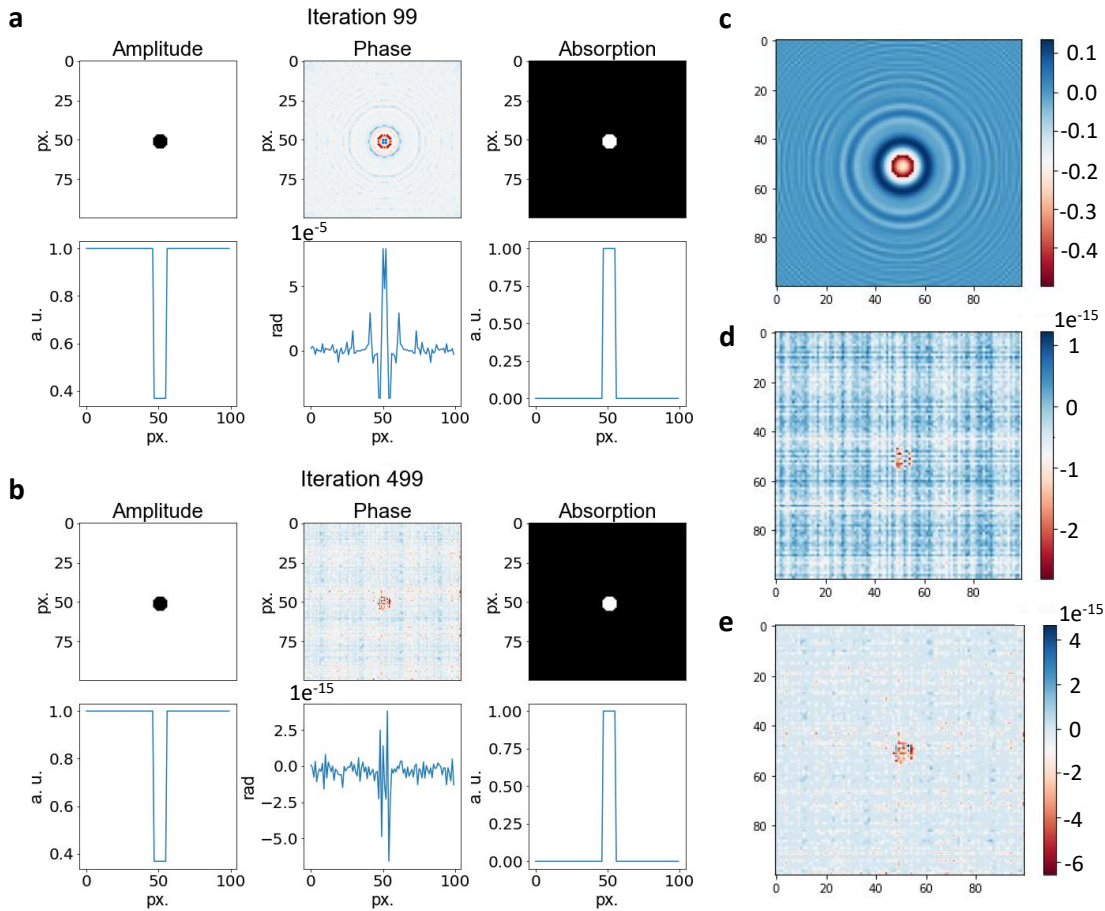
Figure 3.9: **Purely absorbing objects: a** Iterative reconstruction of a disk with absorption $\alpha = 1$ and phase $\varphi = 0$ after 100 iterations. The input values of the amplitude and absorption are recovered, small deviations from the input phase remain. **b** Iterative reconstruction of a disk with absorption $\alpha = 1$ and phase $\varphi = 0$ after 500 iterations. The deviations from the input phase have been decreased by several orders of magnitude, the remaining deviations are close to the ones occurring in a one-step reconstruction of the complex-valued wave field in the detector plane as shown in **d**. **c** Phase distribution reconstructed with a one-step algorithm as presented in Chapter 1.1 from the same input hologram as used in **a** and **b**. It deviates significantly from the input phase distribution, which is equal to zero. **d** One-step phase reconstruction starting from the complex-valued wave field at the detector simulated from the same input as **a-c**. Knowledge of the phases in the detector plane yields a near-perfect reconstruction of the input, only a very small error remains. **e** Phase distribution as reconstructed in **b**. The small deviations from the input phase match those of the reconstruction shown in **d**, both are in the range of 10^{-15}.

Pure phase objects

A pure phase object is defined as an object of non-zero phase shift distribution and zero absorption, i. e. the interaction of such an object with the reference wave only changes the phase, not the amplitude. A pure phase object would hence be fully transparent.

The transmission function takes the form

$$t_{\text{phase}}(x, y) = e^{i\varphi(x,y)}. \tag{3.7}$$

As demonstrated by Fig. 3.10, which depicts iteratively reconstructed phase objects ($\alpha = 0$) for a variety of phase values, $\varphi = \pm 0.5, \pm 1, \pm 3, \pm 4$ rad, the correct iterative reconstruction of pure phase objects is more complicated than that of purely absorbing objects, especially in certain phase regimes.

Fig. 3.10 shows that pure phase objects with negative phase values between 0 and $-\pi$ rad (Fig. 3.10a-c) can be reconstructed with a high degree of accuracy. While the phase is reconstructed almost perfectly, a small error in amplitude and absorption remains. The convergence of the algorithm, however, is not as stable for these types of objects as it is in the case of objects with both absorbing and phase-shifting properties or in the case of purely absorbing objects, since for some phase values, additional iterations will lead to a jump from a highly accurate reconstruction to one quite far from the input values. Additionally, for some cases, even though the reconstructed values are highly accurate, some fluctuations around the correct values remain (e.g. Fig. 3.10h).

For pure phase objects with positive phases, an accurate reconstruction of the input is harder to obtain, especially for small positive phase values (Fig. 3.10e-f). In these cases, the phase reconstruction is noisy and far from the input values, and the remaining deviations from zero in the absorption are large, independently of the number of iterations. While the object starts to become recognizable in the phase reconstruction for phase values above approximately $+\frac{\pi}{2}$ rad, although the exact phase values are not reached and the errors in the absorption remain large, the correct phase values can be stably retrieved for phase values above approximately $+2.2$ rad, see Fig. 3.10g.

Reconstructing pure phase objects with phase shifts $|\varphi| > \pi$ rad yields the same result as the corresponding phase shift values in the domain $[-\pi, \pi]$ rad (Fig. 3.10d, h). The phase values ± 5 rad approximately correspond to ∓ 1 rad, as can be seen when comparing Fig. 3.10d, h to Fig. 3.10b, f. Thus, reconstructing pure phase objects correctly is especially difficult for small, positive phase shifts.

Adding a small absorption value can remedy the situation, an absorption contribution of $\alpha = 0.1$, corresponding to 90% transmission, is enough to yield a perfect phase reconstruction independently of the input phase. The accurate reconstruction of strong phase shifting objects ($\varphi = 4$ rad) for objects with absorption $\alpha = 0.1$ has also been reported in [198].

Physically, the reason for the problems occurring in the reconstruction of pure phase objects with small positive phase shifts could be that in the absence of inelastic interaction, only negative phase shifts would be expected to occur. Positive phase shifts occur when the scattering potential is attractive, while negative phase shifts are the result of repulsive potentials [199]. Assuming that most of the elastic interaction in low-energy electron scattering is due to the interaction of the incident electron beam with the electron density of the object, the scattering is dominated by a repulsive potential yielding negative phase shifts. Since the simulated object is thin, multiple scattering, which could lead to an accumulation of phase, can be neglected, hence if a pure phase object would physically exist, it should induce negative phase shifts.

Despite the difference in energy, on an interaction level, this is similar to the X-ray regime, in which negative phase shifts can be assumed for thin objects [200, 201].
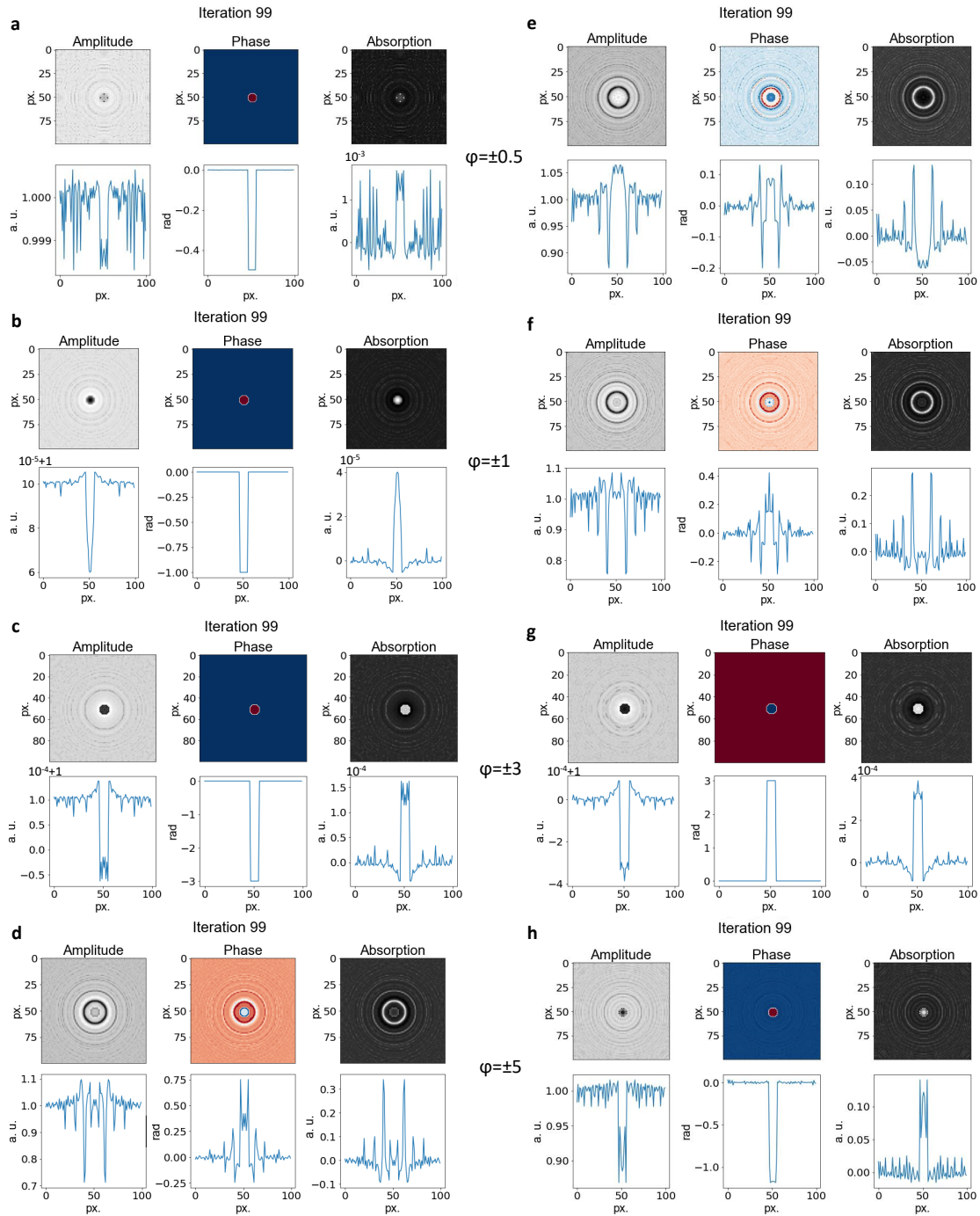
Figure 3.10: **Pure phase objects:** Iterative reconstructions after 100 iterations of disks of absorption $\alpha = 0$ for different phase values: **a** $\varphi = -0.5\,\mathrm{rad}$ **b** $\varphi = -1\,\mathrm{rad}$ **c** $\varphi = -3\,\mathrm{rad}$ **d** $\varphi = -5\,\mathrm{rad}$ **e** $\varphi = +0.5\,\mathrm{rad}$ **f** $\varphi = +1\,\mathrm{rad}$ **g** $\varphi = +3\,\mathrm{rad}$ **h** $\varphi = +5\,\mathrm{rad}$. While the input values are reconstructed with a high degree of accuracy for negative phase shifts in the range $[-\pi, 0]\,\mathrm{rad}$ (**a-d**) as well as for positive phase shifts close to $+\pi\,\mathrm{rad}$ (**g**), the input values cannot be retrieved for low positive input phases (**e-f**). Phase shifts larger than $\pm\pi\,\mathrm{rad}$ are mapped to the interval $[-\pi, \pi]\,\mathrm{rad}$ (**d, h**).

On the level of the algorithm, the difficulties arising during reconstruction could be explained by the fact that such an "unphysical" object would produce a hologram with properties that differ from the properties of holograms for which a reconstruction of the input can be achieved. Comparing the hologram in Fig. 3.11a, which is simulated with parameters in the regime for which reconstruction has been shown to be problematic in Fig. 3.10, to the holograms in Fig. 3.11b-f, from which the input parameters can be retrieved accurately, clear differences are observable, most obviously in the hologram contrast. Holograms for which the correct amplitude and phase values can be reconstructed, feature maximum intensity values only slightly above 1 (1 is the background intensity) as well as values significantly below 1, both for the case of pure phase objects (Fig. 3.11d) and for the case of objects with both absorbing and phase-shifting characteristics (Fig. 3.11e-f). For holograms simulated from a pure phase object in the low positive phase regime, the
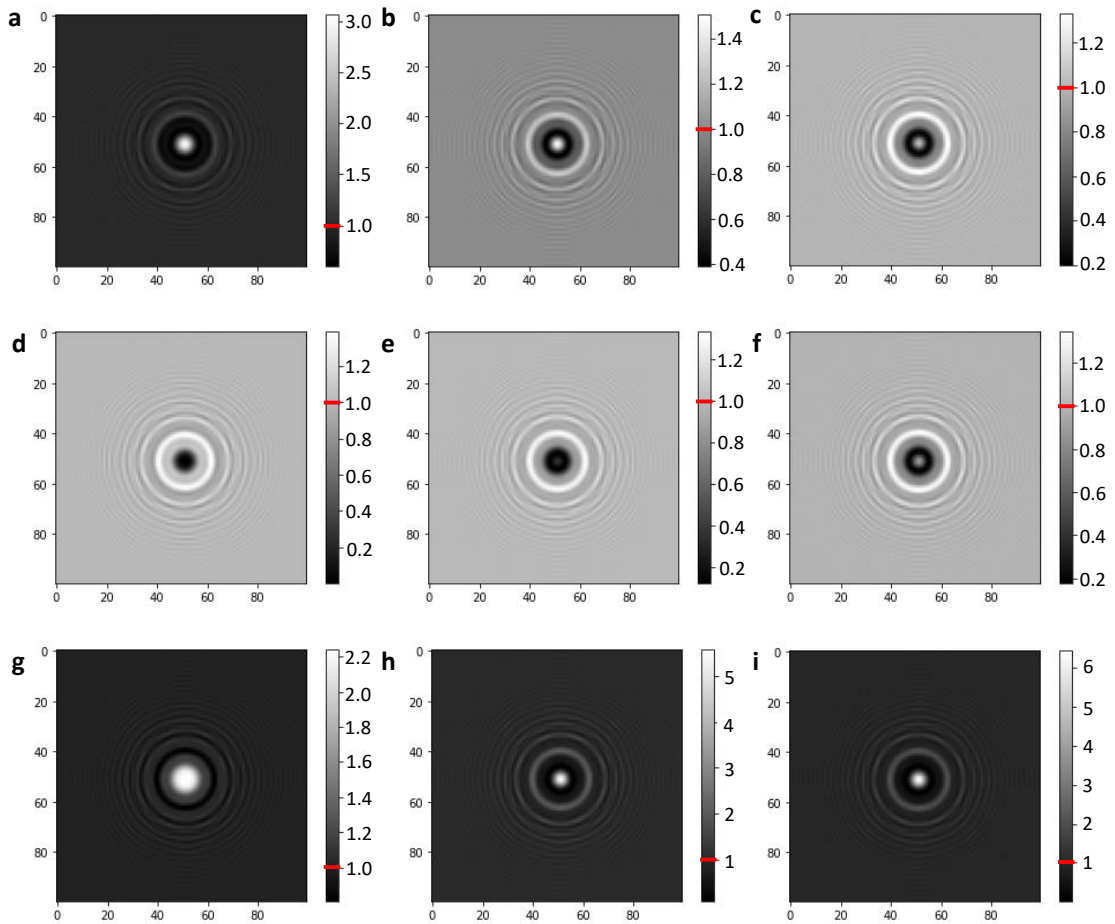
Figure 3.11: **Hologram contrast: a - f** Holograms simulated from disks of **a** absorption $\alpha = 0$ and phase $\varphi = +1\,\mathrm{rad}$, **b** absorption $\alpha = 1$ and phase $\varphi = +1\,\mathrm{rad}$, **c** absorption $\alpha = 3$ and phase $\varphi = +1\,\mathrm{rad}$, **d** absorption $\alpha = 0$ and phase $\varphi = -1\,\mathrm{rad}$, **e** absorption $\alpha = 1$ and phase $\varphi = -1\,\mathrm{rad}$, **f** absorption $\alpha = 3$ and phase $\varphi = -1\,\mathrm{rad}$. The contrast in the hologram in **a** is reversed with respect to **b-f**. **g - i** Holograms simulated from disks of negative absorption $\alpha = -0.5$ for phase $\varphi = 0\,\mathrm{rad}$ (**g**), $\varphi = -3\,\mathrm{rad}$ (**h**) and $\varphi = 3\,\mathrm{rad}$ (**i**). The background intensity is marked in red on the colorbars.

contrast is reversed. While the background still has intensity 1, the minimum values are slightly below 1, whereas the maximum values in the hologram significantly exceed the background intensity (Fig. 3.11a). Since these differences appear already at the level of hologram simulation, the reconstruction behaviour of pure phase objects in the low positive phase regime is not a feature specific to the iterative reconstruction, but of the basic hologram simulation and reconstruction.

The high maximum values in Fig. 3.11a stem from the addition of two waves of unit amplitude in regions of constructive interference. For non-zero absorption, the amplitude of the object wave is less than 1, hence the maximum values in the hologram also decrease. The hologram contrast for absorbing objects with positive phase shifts (Fig. 3.11b-c) approximates that of absorbing objects with negative phase shifts (Fig. 3.11e-f), for large values of absorption, the two holograms simulated from objects of opposite phase have a very similar contrast (Fig. 3.11c, f). The hologram in Fig. 3.11a shows significant similarities to the holograms simulated from objects with negative absorption, as depicted in Fig. 3.11g-i. Such an object with negative absorption is unphysical since it would correspond to an object that, instead of absorbing part of the incident beam, would emit additional electrons in the forward direction, thereby violating the conservation of energy. Adding a small absorption remedies the problem since absorbing characteristics of the object reduce the amplitude of the constructively interfering waves, which in turn reduces the maximum values in the holograms. Physically, adding absorption adds inelastic interaction, which can yield large phase shifts as well as large scattering angles, both of which can result in positive phase shifts.

Since in the low-energy electron regime, proteins, which constitute the main class of molecules studied in this work, are far from the pure phase object regime, the artefacts described in this section should not pose a problem on the practical level when dealing with experimentally measured holograms.

3.2.2 Strongly absorbing objects

While the previous section examined the algorithm's performance regarding objects with very low absorption values, this section will focus on strongly absorbing objects which only very weakly transmit.

As shown in Fig. 3.12, amplitude, absorption and phase can be reconstructed correctly even in regions of high absorption, i.e. of low amplitude. Up to an absorption of approximately $\alpha = 5$, which corresponds to an amplitude of $a = 6.7 \times 10^{-3}$ (transmittance 0.67%), the reconstruction is artefact-free (Fig. 3.12a-b). For absorption values between $\alpha = 6$ and $\alpha = 9$, corresponding to amplitudes between $a = 2.5 \times 10^{-3}$ (transmittance 0.25%) and $a = 1.2 \times 10^{-4}$ (transmittance 0.012%), small errors occur in the reconstruction of both amplitude and phase (Fig. 3.12c-d, shown for $\alpha = 8$, i.e. $a = 3.3 \times 10^{-4}$, transmittance 0.033%). For absorption values $\alpha \geq 10$ ($a \leq 5 \times 10^{-5}$, transmittance $\leq 0.005\%$), as shown for the example of $\alpha = 12$ ($a = 6.1 \times 10^{-6}$, transmittance 0.00061%) in Fig. 3.12e-f, the errors in the phase become large, and the reconstructions for the different input phases in the strongly absorbing part are not distinguishable any more. Mathematically, the reason for this artefact is that the amplitude of the wave is so close to zero that the propagation equations are fulfilled for an arbitrary phase. Thus, for strongly absorbing objects, artefacts and errors can occur in the phase reconstruction.

This point will be examined further on the level of reconstructions of experimentally acquired holograms in Chapter 4.
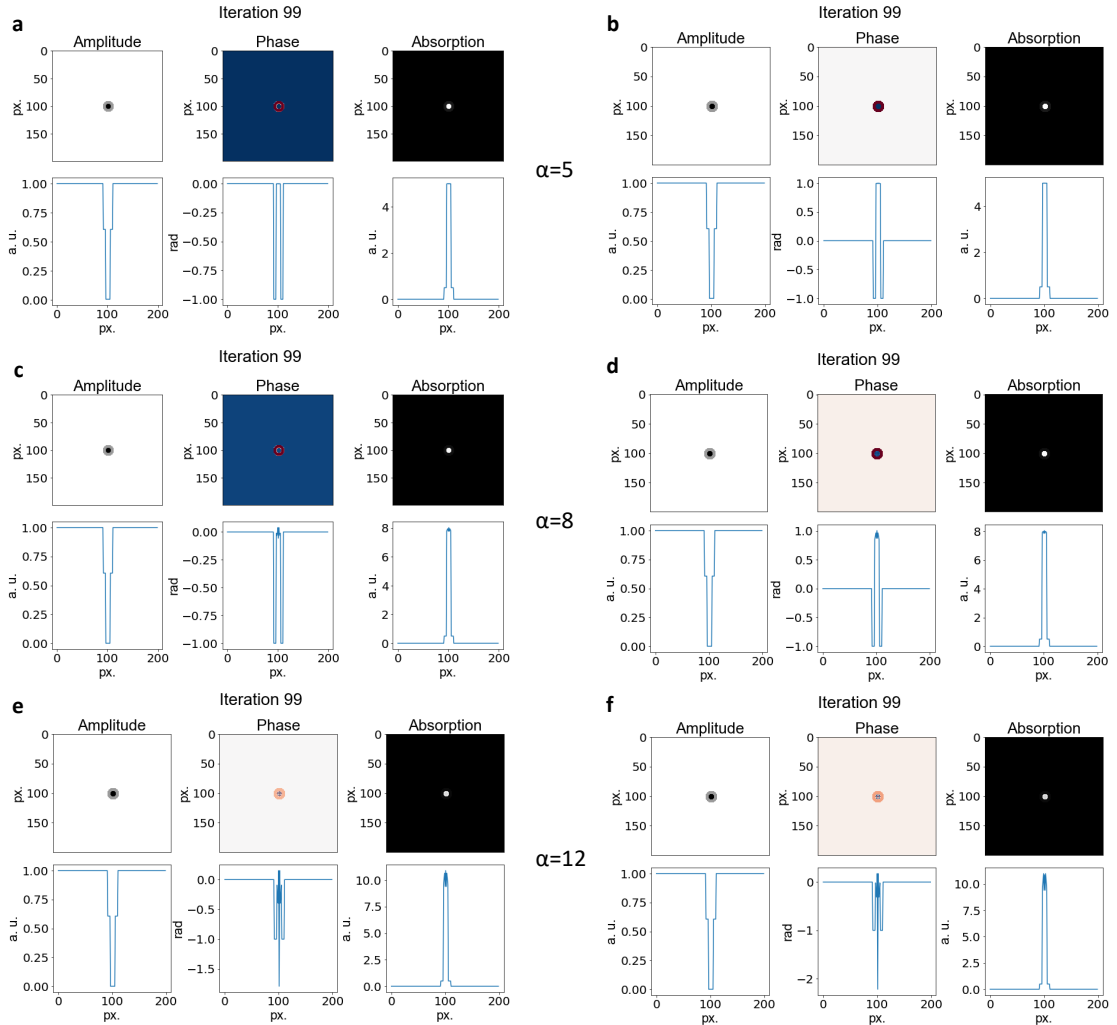

Figure 3.12: **Strongly absorbing objects:** Iterative reconstruction after 100 iterations of objects with a strongly absorbing component. The absorption of the outer ring is $\alpha = 0.5$ for all examples, the phase shift of the outer ring is $\varphi = -1$ rad. The absorption of the strongly absorbing inner disk varies from $\alpha = 5$ (**a**, **b**) to $\alpha = 8$ (**c**, **d**) and $\alpha = 12$ (**e**, **f**), the phase of the strongly absorbing part is either $\varphi = 0$ (**a**, **c**, **e**) or $\varphi = +1$ rad (**b**, **d**, **f**). For $\alpha = 5$, the reconstruction is artefact-free, for $\alpha = 8$, small errors occur, whereas for $\alpha = 12$, the amplitude is too low for the phase to be correctly reconstructed.

3.2.3 Noise

While simulated holograms are in general noise-free, experimental holograms feature a certain degree of noise due to the experimental conditions. Hence, the robustness of the algorithm's performance regarding noisy holograms needs to be examined.

This can be done by adding noise with a Gaussian distribution with mean 0 and different

standard deviations σ to the hologram. Since the square root of the hologram has to be taken to create the initial input for the iterative reconstruction and to fulfil the amplitude constraint, the noise is added directly to the square root of the hologram to avoid artefacts when calculating the square root.

Fig. 3.13a shows that the algorithm retrieves an almost perfect reconstruction of the input values ($\alpha = 1$, $\varphi = -1\,\text{rad}$) for holograms with a low noise level ($\sigma = 0.001$). As shown in Fig. 3.13b, the algorithm can still handle a noise level of $\sigma = 0.01$, the resulting deviations from the input object remain small. Increasing the noise level to $\sigma = 0.1$ still yields a recognizable object, but the deviations from the input increase significantly (Fig. 3.13c). An additional increase of the noise level to $\sigma = 0.2$ results in a reconstruction in which the object is barely discernible (Fig. 3.13d).
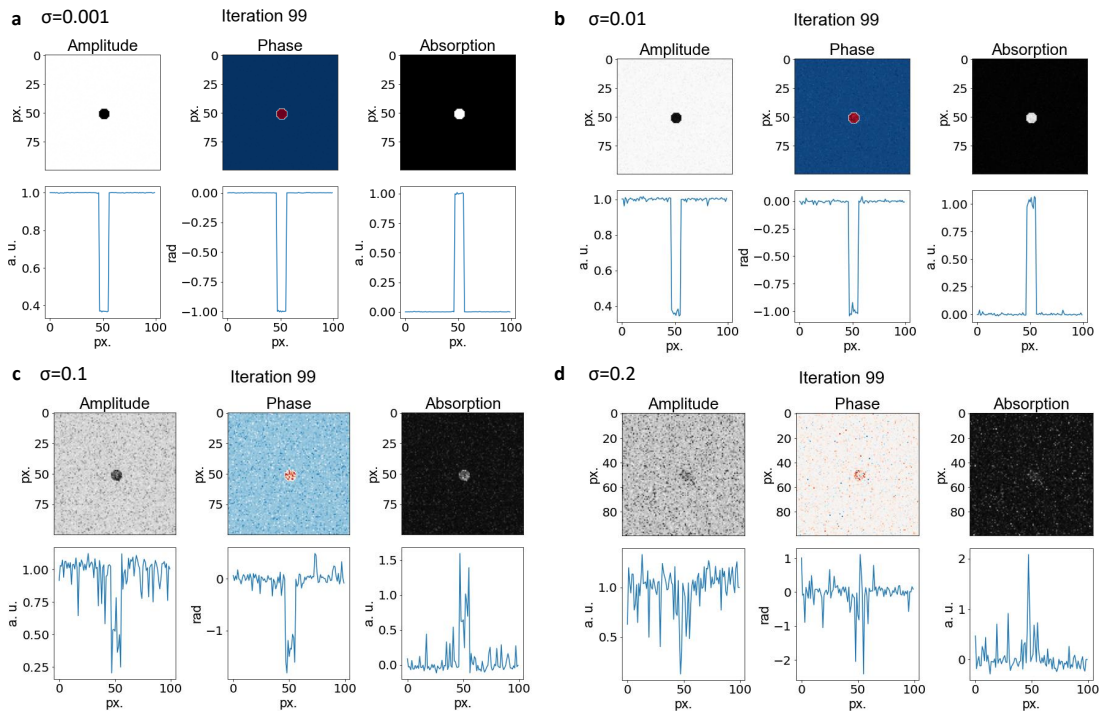
Figure 3.13: **Effect of noise on the reconstruction:** Iterative reconstruction after 100 iterations of a disk of $\alpha = 1$ and $\varphi = -1\,\text{rad}$ with Gaussian noise with mean 0 and standard deviation σ added to the square root of the hologram. The quality of the reconstruction depends strongly on the noise level. **a** $\sigma = 0.001$ **b** $\sigma = 0.01$ **c** $\sigma = 0.1$ **d** $\sigma = 0.2$.

Experimentally, random noise does occur in LEEH measurements, but in general, the signal-to-noise ratio of the images is high compared to other methods, due to the high contrast inherent to this technique even at the single-molecule level. The signal-to-noise ratio and thus the noise level in the image can be further improved by optimizing the shutter times used for recording the experimental holograms.

A further effect of noise can be the blurring of holograms, for example due to mechanical vibrations in the experimental set-up. Since the blurring of the hologram leads to a reduced number of resolvable fringes, this affects the quality of the reconstruction as well

as the resolution.

This is not only relevant because experimental conditions might result in a blurring of the holograms, but also since a successful phase reconstruction of experimental data (see Chapter 4) requires preprocessing of the images which involves both cropping and smoothing the image.
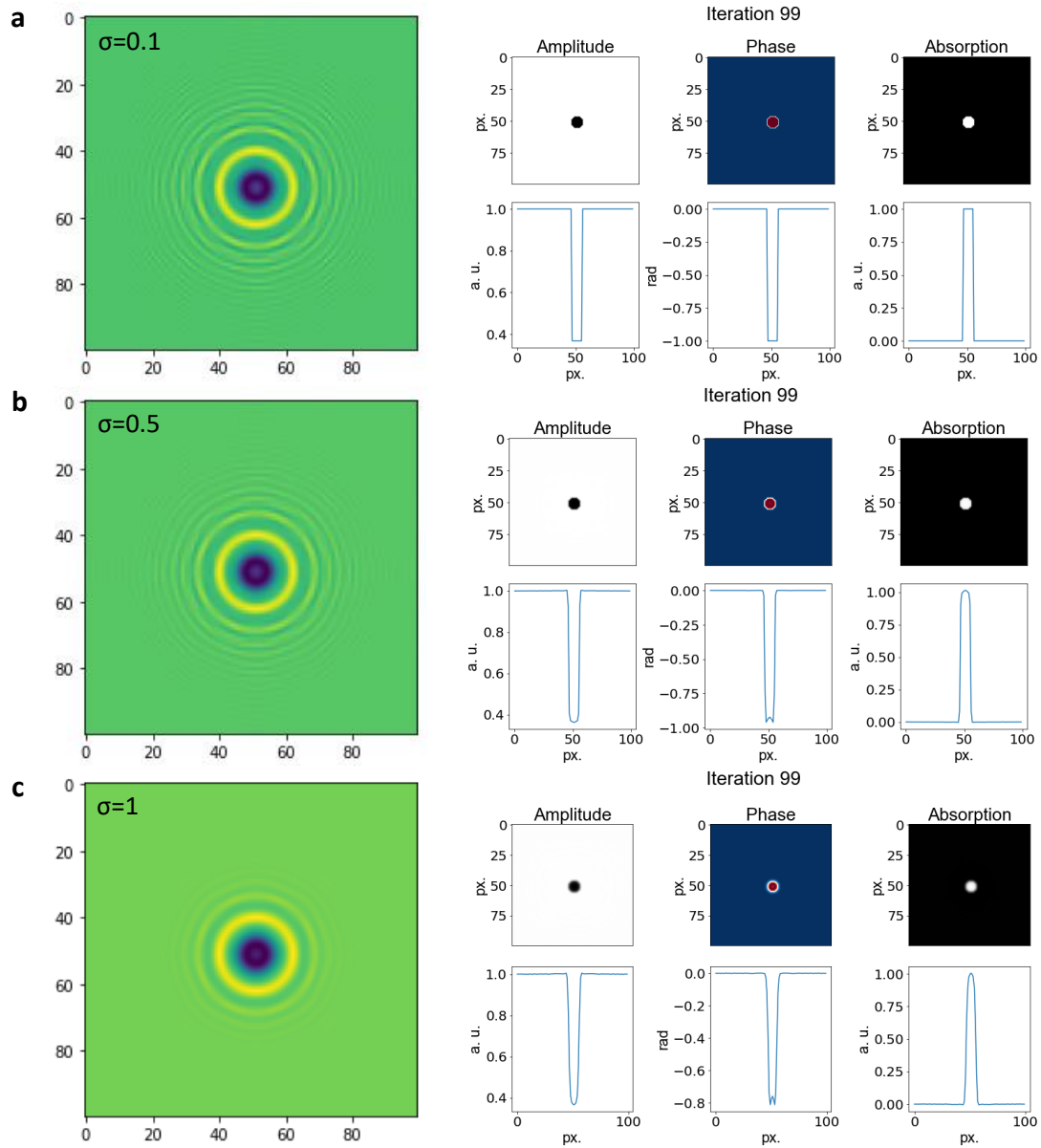
Figure 3.14: **Effect of blurring on the reconstruction:** Hologram of a disk of $\alpha = 1$ and $\varphi = -1$ rad blurred with a Gaussian filter of kernel size 5×5 pixel and corresponding iterative reconstruction after 100 iterations. **a** Slight blurring, $\sigma = 0.1$ **b** Moderate blurring, $\sigma = 0.5$ **c** Strong blurring, $\sigma = 1$.

Fig. 3.14 shows holograms blurred with a Gaussian filter with kernel size 5×5 pixel

and the corresponding reconstructions. Slight blurring still allows for an accurate reconstruction (Fig. 3.14a, $\sigma = 0.1$). Moderate blurring (Fig. 3.14b, $\sigma = 0.5$) yields reconstructions with values close to the input values, but the sharpness of edges is lost. A further increase of the blurring in the hologram (Fig. 3.14, $\sigma = 1$) results in inaccurately reconstructed phase values and additional blurring of the reconstructed edges.

Cutting off part of the fringes of the hologram while leaving the central part intact has a similar effect as the blurring of the hologram since both types of treatment result in the loss of information contained in the outer fringes.

The result of cutting off the outer interference fringes with an apodization filter as described in section 1.1.3 with different inner radii η and a sharp cutoff ($\omega = 0$) is shown in Fig. 3.15. The parts of the image outside the radius η are set to zero. The images depicted in Fig. 3.15 have a size of 100×100 pixels.

A cutoff radius of $\eta = 50$ pixels (Fig. 3.15 a), which only cuts off the fringes in the corners of the image, still yields an accurate reconstruction of the object, a disk of absorption $\alpha = 1$ and phase $\varphi = -1$ rad, with only slight deviations from the input values. The deviations from the input increase with decreasing cutoff radius ($\eta = 30$ pixels in Fig. 3.15b and $\eta = 20$ pixels in Fig. 3.15 c), but the size of the object is still accurately reconstructed. While the noise-like deviations from the input values are stronger when cutting off the fringes than when blurring the entire hologram (Fig. 3.14), the sharpness of the edges is better recovered in the former case.

When decreasing the cutoff radius such that only the innermost part of the hologram remains ($\eta = 10$ pixels, Fig. 3.15d), the object cannot be accurately reconstructed any more.

Since both moderate blurring and the cutting of fringes with large cutoff radius still yield reconstructions that match the input closely, moderate hologram processing before reconstruction should not hinder an accurate retrieval of the properties of the object.

3.2.4   Dependence of the algorithm performance on object size

As already predicted by Gabor [38], the ratio of the object size to the illuminated area is a parameter of relevance for a correct reconstruction of the object. Starting from a consideration of the contribution of the error terms due to the twin image to the reconstructed object, Gabor arrives at the conclusion that in order to obtain sufficient background uniformity, the object should only take up 1% of the illuminated area, i. e. the object-to-area ratio should not exceed 0.01 [38].

Gabor derived this criterion in the framework a non-iterative reconstruction for both absorbing objects and pure phase objects [38], hence, since the iterative reconstruction discussed here is based on the removal of the twin image contributions, reconstructing iteratively could lead to a relaxation of the criterion by eliminating error terms. While, as expected, an iterative reconstruction yields results closer to the input values, especially regarding phase, inaccuracies in both phase and amplitude reconstruction remain for objects whose size-to-area ratio exceeds the Gabor criterion, as shown in Figures 3.16 and 3.17. Similar conclusions have been drawn in [198].
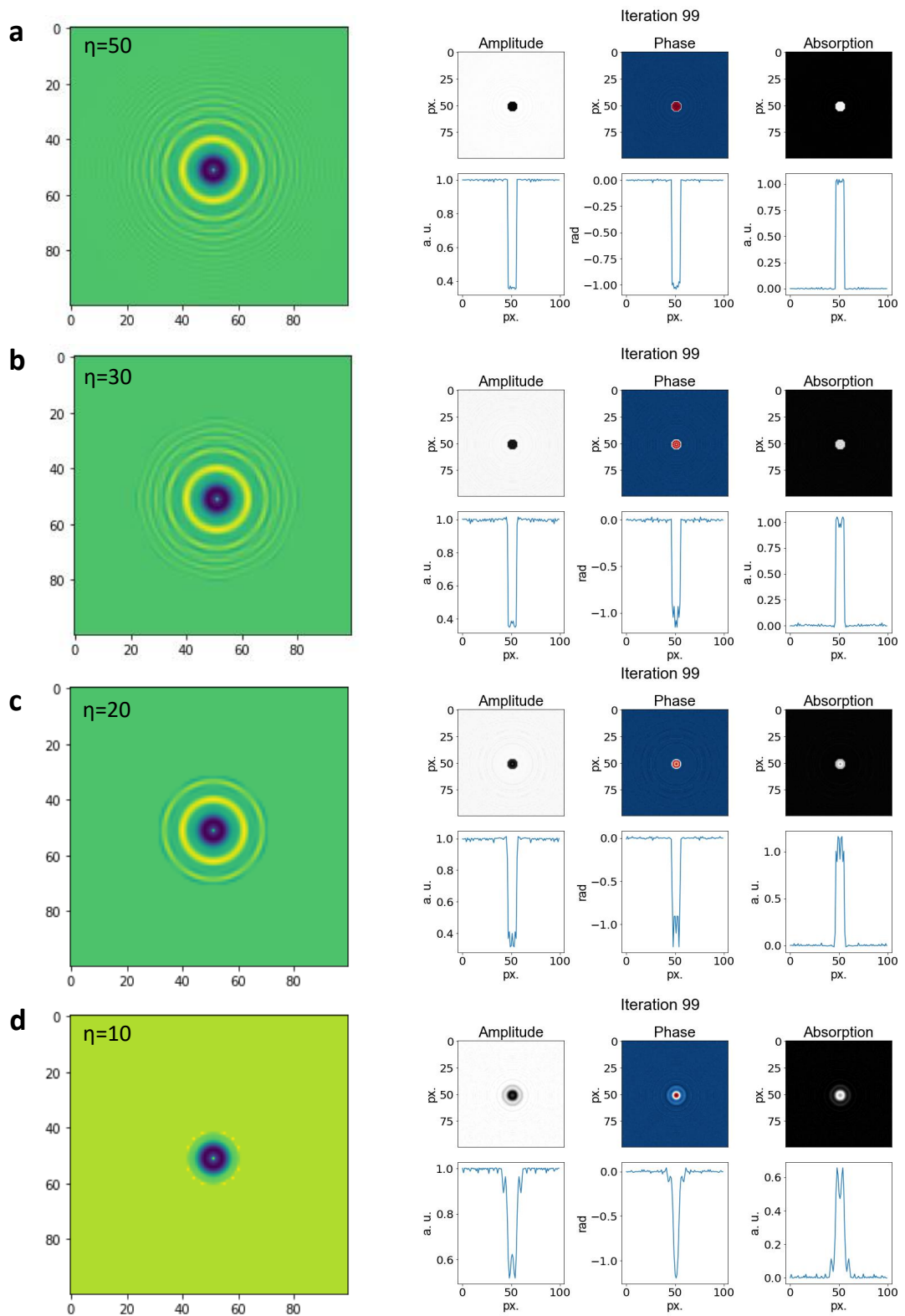
Figure 3.15: **Effect of the cutting of fringes:** Hologram of a disk of $\alpha = 1$ and $\varphi = -1\,\mathrm{rad}$ with part of the fringes cut off by an apodization filter. The cutoff radii are **a** 50 pixels **b** 30 pixels **c** 20 pixels and **d** 10 pixels.

So far, disks with radius 5 pixels for image sizes of 100×100 pixels, corresponding to a size-to-area ratio of 7.85×10^{-3} have been used to characterize the algorithm. As shown in Fig. 3.6 and Fig. 3.7, accurate reconstructions of the object can be obtained for different values of absorption and phase for this size-to-area ratio. Analogously, in the case of larger objects of radius 10 pixels for image sizes of 200×200 pixels, which yields the same size-to-area ratio, the objects are correctly reconstructed (see Fig. 3.8a-c, Fig. 3.12). The slight deviations from the input values in Fig. 3.8d could be attributed to a size-to-area ratio slightly above the limit set by the Gabor criterion.
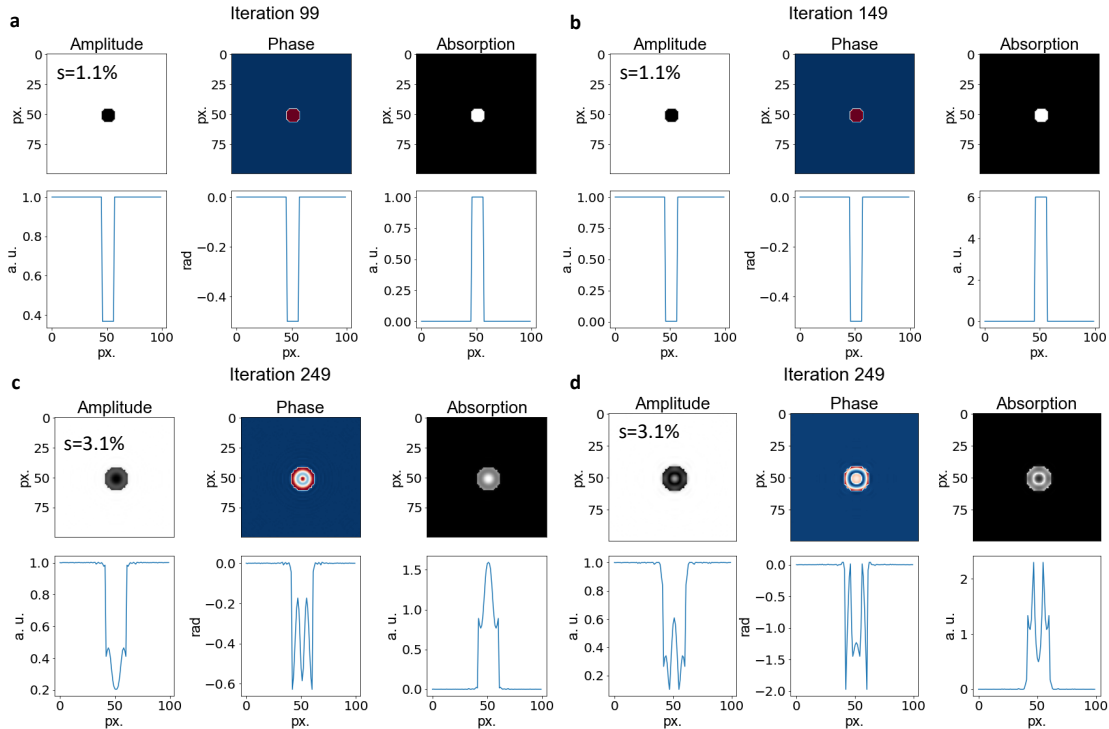

Figure 3.16: **Effect of size-to-area ratio: a** Iterative reconstruction of a disk of radius $r = 6$ pixels (size-to-area ratio s=1.1%), absorption $\alpha = 1$ and phase $\varphi = -0.5\,\mathrm{rad}$ after 100 iterations. **b** Iterative reconstruction of a disk of radius $r = 6$ pixels (size-to-area ratio s=1.1%), absorption $\alpha = 6$ and phase $\varphi = -0.5\,\mathrm{rad}$ after 150 iterations. **c** Iterative reconstruction of a disk of radius $r = 10$ pixels (size-to-area ratio s=3.1%), absorption $\alpha = 1$ and phase $\varphi = -0.5\,\mathrm{rad}$ after 250 iterations. **d** Iterative reconstruction of a disk of radius $r = 10$ pixels (size-to-area ratio s=3.1%), absorption $\alpha = 6$ and phase $\varphi = -0.5\,\mathrm{rad}$ after 250 iterations. The image size in all cases is 100×100 pixels.

To examine the influence of the size-to-area ratio, Fig. 3.16 shows the reconstruction of disks of radius 6 pixels (Fig. 3.16a, b) and of radius 10 pixels (Fig. 3.16c, d) for an image size of 100×100 pixels and absorption values of $\alpha = 1$ and $\alpha = 6$, with a phase shift of $\varphi = -0.5\,\mathrm{rad}$ in all cases. For $r = 6$ pixels, the size-to-area ratio is 0.011, i. e. slightly above the criterion defined by Gabor, while the size-to-area ratio corresponding to $r = 10$ pixels is 0.031 and hence considerably above Gabor's criterion. For $r = 6$ pixels, despite the size-to-area ratio slightly exceeding 0.01, the object can be correctly reconstructed for both absorption values shown, although in the case of higher absorption, the full

convergence requires a higher number of iterations (150 iterations at $\alpha = 6$ vs. less than 100 iterations for $\alpha = 1$). A similar behaviour is observable for radii of 7 and 8 pixels (size-to-area ratios of 0.015 and 0.020, respectively). In this range, for some cases, the object can be correctly reconstructed in less than 100 iterations, while for other values of absorption and phase, a larger number of iterations is necessary, or no convergence to the input is obtained at all. Hence, for size-to-area ratios slightly exceeding 0.01, the convergence properties strongly depend on absorption and phase values.

For size-to-area ratios above 2%, the object is not reconstructed accurately any more, not even for a higher number of iterations, as demonstrated in Fig. 3.16c, d for the case $r = 10$ pixels. The reconstructed phase changes rapidly between several values, and the amplitude continuously decreases towards the centre of the disk. For high absorption values (Fig. 3.16d), this trend likely continues towards negative amplitude values, which then appear in the reconstruction as higher amplitudes, i.e. a bright spot at the centre of the disk, because the absolute value of the transmission function is taken to calculate the amplitude.
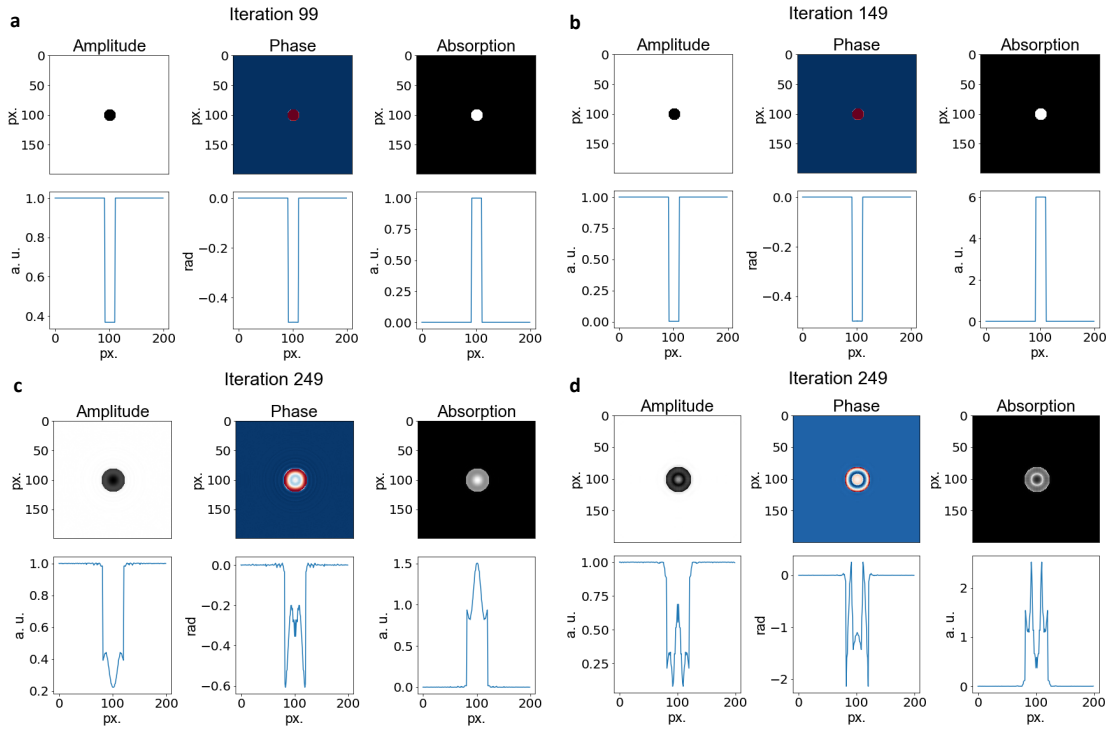
Figure 3.17: **Effect of size-to-area ratio II: a** Iterative reconstruction of a disk of radius $r = 10$ pixels, absorption $\alpha = 1$ and phase $\varphi = -0.5\,\mathrm{rad}$ after 100 iterations. **b** Iterative reconstruction of a disk of radius $r = 10$ pixels, absorption $\alpha = 6$ and phase $\varphi = -0.5\,\mathrm{rad}$ after 150 iterations. **c** Iterative reconstruction of a disk of radius $r = 20$ pixels, absorption $\alpha = 1$ and phase $\varphi = -0.5\,\mathrm{rad}$ after 250 iterations. **d** Iterative reconstruction of a disk of radius $r = 20$ pixels, absorption $\alpha = 6$ and phase $\varphi = -0.5\,\mathrm{rad}$ after 250 iterations. The image size in all cases is 200×200 pixels.

To show that this dependence of the size-to-area ratio is independent of the image size, the analogue to Fig. 3.16 is shown in Fig. 3.17 for an image size of 200×200 pixels and

object radii of $r = 10$ (size-to-area ratio 7.85×10^{-3}) and $r = 20$ (size-to-area ratio 0.031). The same behaviour as in the case of an image size of 100×100 pixels is observed. Since the derivation of the admissible size-to-area ratio is based on the requirement of a certain degree of uniformity of the background, this emphasises the importance of a clean substrate and of a stable illumination for the successful experimental implementation of holography.

Next to the size-to-area ratio, there is another size-related parameter whose value determines whether the object can be correctly reconstructed. This parameter is related to the physical size of the object S_O, i.e. in the case of simulated objects as presented here, to the physical size assigned to the input image. The inverse of that parameter appears in the propagator function as the pixel size in Fourier space $\Delta_O = \frac{1}{S_O}$ and is crucial for correct sampling when transforming between real space and Fourier space. While the physical object size in itself is not theoretically limited, its interplay with the other parameters occurring in the argument of the propagator function, namely wavelength and source-to-sample distance, determines the imaging regime and hence the amount of fringes in the hologram, which in turn affects the performance of the algorithm and the quality of the reconstruction.

The prefactor f of the pixel numbers in Fourier space in the argument of the propagator function takes the form

$$f = \lambda z \Delta_O^2 = \frac{\lambda z}{S_O^2}, \tag{3.8}$$

where λ is the electron wavelength, z is the source-to-sample distance and S_O is the physical object size, corresponding to the length of one dimension of the image, hence S_O^2 is the image area.

This factor is closely related to a quantity that is often used to characterize imaging systems, the Fresnel number N_f. N_f is a measure of the strength of diffraction at an aperture and can thus be used to define different imaging regimes, specifically near-field imaging (such as holography) and far-field imaging (such as diffraction imaging). The Fresnel number is defined as [194, 198]

$$N_f = \frac{(\text{maximum object size})^2}{\lambda Z}. \tag{3.9}$$

In the general definition, the sample-to-detector distance Z appears instead of the source-to-sample distance z, since in many imaging systems, especially when imaging with incident plane waves, this is the relevant parameter to describe the system. In a holography set-up with plane waves, for example, Z rather than z appears in the propagator (see section 3.3.1 for the derivation of the plane wave propagator), hence the Fresnel number is proportional to the inverse of the prefactor occurring in the propagator function. Since in literature, the Fresnel number values for which an accurate holographic reconstruction can be obtained have been determined for incident plane waves [194, 198], the values are not directly comparable to the ones calculated from the parameters used in Fig. 3.18, but multiplication with the magnification factor $\frac{Z}{z}$ results in values in the same range.
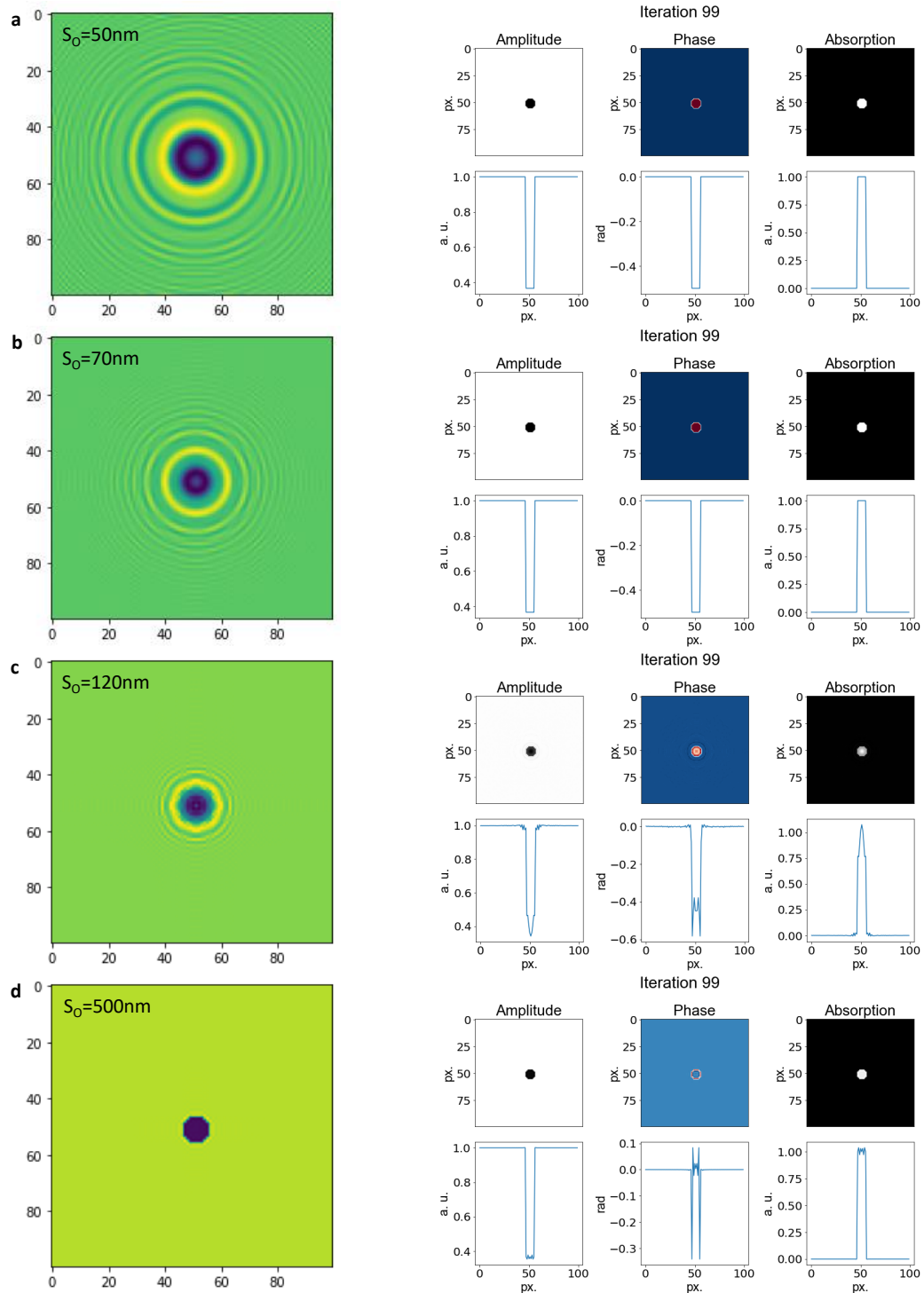
Figure 3.18: **Effect of physical object size:** Hologram and iterative reconstruction of a disk of radius $r = 5$ pixels, absorption $\alpha = 1$ and phase $\varphi = -0.5$ rad after 100 iterations for different physical object sizes S_O. **a** $S_O = 50$ nm **b** $S_O = 70$ nm **c** $S_O = 120$ nm **d** $S_O = 500$ nm.

The change in the imaging regime resulting from changing the physical object size while keeping λ and z as well as the pixel number N of the image constant can directly be observed on the hologram level, as shown in Fig. 3.18. Fig. 3.18 depicts holograms simulated from a disk of radius $r = 5$ pixels, absorption $\alpha = 1$ and phase $\varphi = -0.5\,\text{rad}$, along with their iterative reconstructions after 100 iterations, for different physical image sizes ranging from $S_O = 50\,\text{nm}$ to $S_O = 500\,\text{nm}$. The physical image size $S_O = 70\,\text{nm}$, as shown in Fig. 3.18b, has been used as the default in all simulations in this chapter. Experimentally, a change in physical image size and thus in imaging regime can be achieved by changing the sample-to-detector distance.

The effect of the increased physical object size for a given set of parameters λ, z, N is observable in a decrease in the number and extent of fringes due to the loss of high frequencies in Fourier space. If the other parameters are kept constant, the amount of visible fringes decreases with increasing object size until no fringes are visible at $S_O = 500\,\text{nm}$, which means that the imaging does not happen in the holographic regime. This directly implies that the object cannot be reconstructed in the way presented here, which is apparent in the corresponding reconstruction: the amplitude is still reconstructed fairly accurately in this projection-like regime, but the phase is not. Thus, if objects of large physical size need to be reconstructed, the other parameters have to be adjusted accordingly.

3.3 Towards phase interpretation

When considering the holographic imaging of real, physical objects such as proteins, it is obvious that such an object would only in very special cases have uniform amplitude and phase distributions as the ones used in most of the simulated examples in this chapter so far. To incorporate effects that likely occur in real samples, simulation methods for modelling some of these aspects have to be developed, which could yield insights into the interpretation of experimental phase data.

Non-uniformity in phase and amplitude, both in the lateral and axial dimensions, can be the result of many different contributions stemming from the interaction of the electrons with the sample, such as variations in sample thickness, which could lead to an accumulation of phase and amplitude due to multiple scattering processes, and locally different scattering strengths, e. g. due to local electric potentials and charges.

To model these effects, in the following sections a multislice algorithm to model multiple scattering effects and an algorithm based on element-dependent scattering are discussed.

3.3.1 Multislice algorithm

Proteins, the main experimental focus of this thesis, are three-dimensional objects with a molecular thickness in the range of several nanometers. Since the mean free path in proteins for low-energy electrons is in a similar range as the molecular thickness [57], multiple scattering is likely to occur in LEEH imaging of proteins. To simulate multiple scattering processes, multislice algorithms can be applied [202, 203, 204].

A multislice approach to scattering in order to incorporate the effects of multiple scattering events has first been proposed by Cowley and Moodie in 1957 [202] as a method of

simulating diffraction patterns in electron crystallography. The general idea is based on the division of the object into slices, with each slice being assigned a transmission function t_i, where i denotes the slice number [205]. After the interaction of the incoming wave with the first slice slice (i. e. multiplication of the incident wave with the slice's transmission function), the resulting exit wave is freely propagated to the next slice with a Fresnel-Kirchhoff propagation function as employed in the single scattering algorithms described in the previous sections. The exit wave produced by the interaction with the last slice is then propagated to the detector, where the hologram is recorded. This exit wave contains the information from subsequent interaction with all object slices.

Since no assumptions regarding periodicity are made, the method is applicable to a wide range of objects.

In the high-energy regime, for which the method was originally developed, the phase shift and hence the transmission function can directly be related to the integrated potential $V(\rho, z)$ over the slice thickness z of the molecule [206, 205]; by including an imaginary part in the potential, absorbing properties can be added, and the transmission function of a slice i can be written as $t_i = \exp(\int_{z_i}^{z_i+1} V_i(\rho) \, \mathrm{d}z)$ [205].

For low-energy electrons, this relation does not hold, since further effects need to be included in the description of the interaction. However, the general idea of partitioning an object into slices with which the electron beam interacts successively, is still applicable, even if the exact interaction is not modelled. Since the method is compatible with the transmission function approach utilized in both the single-step and the iterative algorithms described above, objects consisting of multiple slices can easily be incorporated into the general framework of the simulation algorithm presented so far. The object can either be build from a series of known two-dimensional transmission functions, for example if the object consists of several separate overlapping or non-overlapping parts (Fig. 3.21 – Fig. 3.23), or, if the initial object structure is three-dimensional, for example in the case of a molecular model (Fig. 3.24), projections of volume segments of the object onto slices can be used as input for a multislice hologram simulation.

The general form of the multislice algorithm is depicted in Fig. 3.19. The object is divided into n slices with a distance d_i in between the slices. The slices can be chosen to be equidistant, but do not need to be.

An incoming wave ψ interacts with the first slice at distance z_1, characterized by the transmission function t_1, which results in the exit wave $\psi_1(z_1) = \psi(z_1)t_1$. Subsequently, ψ_1 is propagated a distance d_1 to the next slice at z_2 with transmission function t_2, the interaction yields a new exit wave $\psi_2(z_2) = \psi_1(z_2)t_2$. This process is repeated until the last slice at distance z_n is reached. The interaction in that plane produces the final exit wave $\psi_n(z_n) = \psi_{n-1}(z_n)t_n$, which is then propagated to the detector at distance z, where the hologram $H = |\psi_n(z)|^2$ is calculated.

If the incident wave is a plane wave, the multislice method is straightforward to incorporate [192, 207] since the propagator used to propagate between the sample slices is the same type of propagator that is used for propagation to the detector. In the case of illumination with spherical waves, as experimentally employed in LEEH, the geometrical magnification needs to be taken into account. Because the propagation distance between two object slices is much smaller than the source-to-sample distance, using the spherical propagator function for propagation within the object would lead to a demagnification of the slice
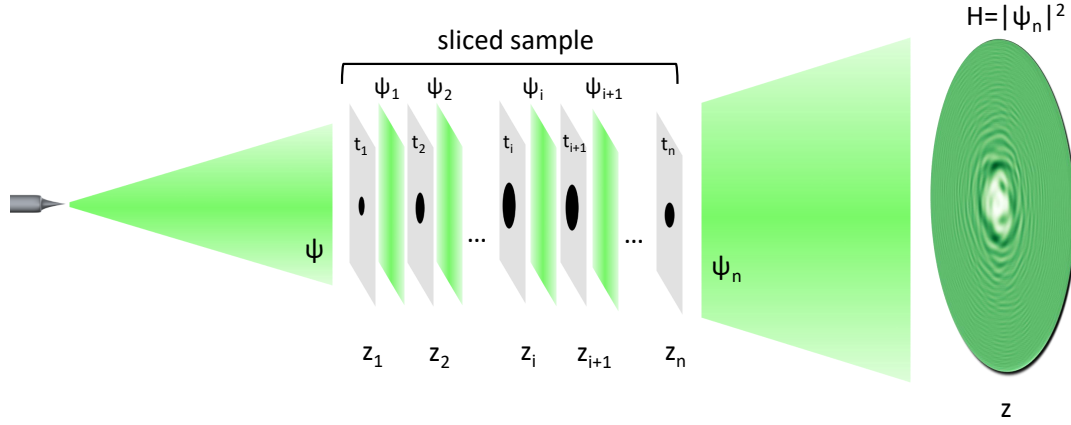
Figure 3.19: **Multislice algorithm:** Schematic depiction of the multislice algorithm. The incident wave is propagated to the first object slice of transmission function t_1, the interaction results in an exit wave ψ_1 that is subsequently propagated to the next slice, where it interacts with t_2, etc. Since the incident wave is a spherical wave, the final propagation step to the detector is calculated using a spherical Fresnel-Kirchhoff propagator, whereas the propagation over the short distances within the object is carried out with a plane wave propagator.

sizes. Thus, since the object's thickness is small in comparison to the source-to-detector distance, one can assume that no significant radial spreading of the wave front occurs between slices, and that the spherical wave can be approximated by a plane wave for propagation distances on the scale of the slice separation. This implies that the propagation within the object can be described by plane wave propagation, which does not involve magnification.

In order to amend the code accordingly, a propagator function for plane wave propagation needs to be derived. Since this propagator function will be applied to very small propagation distances z, the derivation must be in a regime that is valid for small z and hence cannot rely on approximations regarding large propagation distances. A method that fulfils this requirement is the angular spectrum method [73], which is valid for both small and large z.

The angular spectrum method is based on the description of the Fourier transform of a monochromatic, complex wave field as a sum of plane waves travelling in different directions, which yields different angular components in the argument of the exponential describing the plane wave [208, 209].

To derive the angular spectrum propagator, the propagation of a monochromatic complex wave field $U(x, y, 0)$ in the plane $z = 0$ to the corresponding wave field $U(x, y, z)$ in the parallel plane $z = z$ shall be considered [209]. The complex wave field $U(x, y, 0)$ can be written as an inverse Fourier transform of a function $A(u, v, 0)$,

$$U(x, y, 0) = \iint_{-\infty}^{\infty} A(u, v, 0) e^{2\pi i (ux + vy)} \mathrm{d}u \mathrm{d}v, \tag{3.10}$$

where u and v are coordinates in Fourier space. The wave vector $\mathbf{k} = (k_x, k_y, k_z)$ depends

on the directional angles and can be written as

$$\mathbf{k} = \frac{2\pi}{\lambda}(\cos\phi\sin\theta, \sin\phi\sin\theta, \cos\theta) = \frac{2\pi}{\lambda}(\alpha, \beta, \gamma), \tag{3.11}$$

with the polar angle θ and the azimuthal angle ϕ. This implies the following relation for the direction cosines α, β, γ (see Fig. 3.20):

$$\alpha^2 + \beta^2 + \gamma^2 = 1 \quad \Rightarrow \quad \gamma = \sqrt{1 - \alpha^2 - \beta^2} \tag{3.12}$$
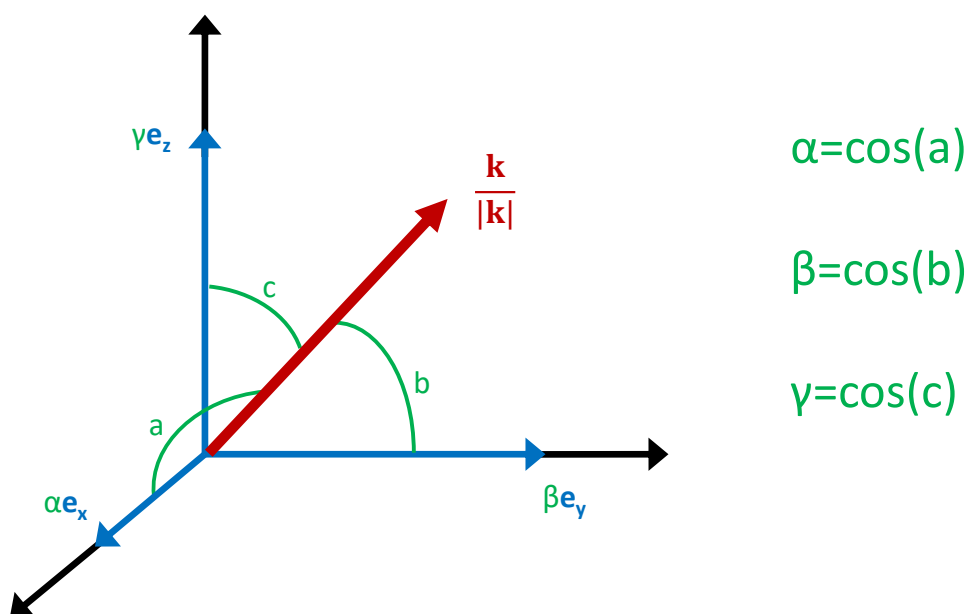

Figure 3.20: **Direction cosines**

With this notation, the angular spectrum $A\left(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda}, 0\right)$ of the wave field $U(x, y, 0)$ takes the form

$$A\left(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda}, 0\right) = \iint_{-\infty}^{\infty} U(x, y, 0)e^{-2\pi i(\frac{\alpha}{\lambda}x + \frac{\beta}{\lambda}y)}\mathrm{d}x\mathrm{d}y. \tag{3.13}$$

Analogously, the angular spectrum of a wave field $U(x, y, z)$ at a plane z is given by

$$A\left(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda}, z\right) = \iint_{-\infty}^{\infty} U(x, y, z)e^{-2\pi i(\frac{\alpha}{\lambda}x + \frac{\beta}{\lambda}y)}\mathrm{d}x\mathrm{d}y. \tag{3.14}$$

Thus, to determine the function that propagates the wave field from plane $z = 0$ to $z = z$, the relationship between the angular spectra $A\left(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda}, 0\right)$ and $A\left(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda}, z\right)$ has to be found. For this, the wave field $U(x, y, z)$ can be rewritten in terms of its angular spectrum,

$$U(x, y, z) = \iint_{-\infty}^{\infty} A\left(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda}, z\right) e^{2\pi i(\frac{\alpha}{\lambda}x + \frac{\beta}{\lambda}y)}\mathrm{d}\frac{\alpha}{\lambda}\mathrm{d}\frac{\beta}{\lambda}. \tag{3.15}$$

$U(x, y, z)$ also satisfies the Helmholtz equation (see section 1.1), i.e.

$$\nabla^2 U + k^2 U = 0, \tag{3.16}$$

which, when applied to the expression in eq. (3.15), yields

$$\frac{\mathrm{d}^2}{\mathrm{d}z^2} A\left(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda}, z\right) + \left(k^2 - \frac{4\pi^2}{\lambda^2}\left(\alpha^2 + \beta^2\right)\right) A\left(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda}, z\right) = 0. \tag{3.17}$$

With

$$(k_x, k_y, k_z) = \frac{2\pi}{\lambda}(\alpha, \beta, \gamma), \tag{3.18}$$

this results in

$$\frac{\mathrm{d}^2}{\mathrm{d}z^2} A\left(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda}, z\right) + \left(\frac{2\pi}{\lambda}\right)^2 (1 - \alpha^2 - \beta^2) A\left(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda}, z\right) = 0. \tag{3.19}$$

This is a well-known differential equation which is solved by the exponential function

$$A\left(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda}, z\right) = A\left(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda}, 0\right) e^{i\frac{2\pi}{\lambda}\sqrt{1-\alpha^2-\beta^2}z}. \tag{3.20}$$

Hence, the angular spectrum propagator has the form

$$e^{i\frac{2\pi}{\lambda}\sqrt{1-\alpha^2-\beta^2}z}, \tag{3.21}$$

and the propagated wave field $U(x, y, z)$ can be expressed as

$$U(x, y, z) = \iint_{-\infty}^{\infty} A\left(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda}, 0\right) e^{i\frac{2\pi}{\lambda}\sqrt{1-\alpha^2-\beta^2}z} e^{2\pi i(\frac{\alpha}{\lambda}x + \frac{\beta}{\lambda}y)} \mathrm{d}\frac{\alpha}{\lambda} \mathrm{d}\frac{\beta}{\lambda} \tag{3.22}$$

$$= \mathcal{F}^{-1}\left(\mathcal{F}(U(x, y, 0)) e^{i\frac{2\pi}{\lambda}\sqrt{1-\alpha^2-\beta^2}z}\right). \tag{3.23}$$

Applied to the propagation from the sample plane (transmission function $t(x, y)$) to the detector plane in holography, this yields

$$U_{det}(X, Y) = \mathcal{F}^{-1}\left(\mathcal{F}(t(x, y)) e^{i\frac{2\pi}{\lambda}\sqrt{1-\alpha^2-\beta^2}z}\right). \tag{3.24}$$

Numerically, the angular spectrum propagator can be implemented as follows:

```
def PropagatorAngularSpec(ar, distance, wavelength, holin):
    delta0 = 1/ar
    (Nx,Ny)=holin.shape[:2]
    p=np.zeros((Nx,Ny), dtype=complex)
    for ii in range(Nx):
        for jj in range(Ny):
            alpha=(ii-Nx/2)*wavelength*delta0
            beta=(jj-Nx/2)*wavelength*delta0
            if alpha**2+beta**2<=1:
                p[ii][jj]=np.exp(1j*2*np.pi*distance*
                            np.sqrt(1-alpha**2-beta**2)/wavelength)
    return p
```

As before, `delta0` is the pixel size in Fourier space, calculated from the real space object size `ar`. In line 7 and 8, α and β are sampled in Fourier space. `distance` denotes the propagation distance.

With this definition, the hologram simulation and reconstruction for incident plane waves can be implemented with the exact same code as for spherical waves if the spherical propagator is replaced by the angular spectrum propagator:

```
1   Udet=IFT2D(FT2D(t)*PropagatorAngularSpec(area, z, Lambda, t))
```

As an example, the following code simulates an object consisting of three slices with transmission functions t_0, t_1 and t_2, respectively, at distances z_0, z_1 and z_2 from the source. One can assume that, since the object is small, the incident spherical wave can be approximated by a plane wave over the extent of the object. This implies that the exit wave after the interaction with the first slice can be represented by the corresponding transmission function t_0 since the remaining factors are constant and can hence be dropped. Thus, as a first step, the slice characterized by the transmission function t_0 is propagated forward to the next slice with the help of the angular spectrum propagator, which yields the incident wave for the z_1-plane:

```
1   U1_in=IFT2D(FT2D(t_0)*PropagatorAngularSpec(area, z1-z0, Lambda, t_0))
```

The subsequent interaction with the transmission function t_1 results in the exit wave of the z_1-plane

```
1   U1_exit=U1_in*t_1
```

These steps are then repeated, first `U1_exit` is propagated to the plane z_2, then it interacts with t_2 to create the exit wave `U2_exit`:

```
1   U2_in=IFT2D(FT2D(U1_exit)*PropagatorAngularSpec(area, z2-z1, Lambda, U1_exit))
2   U2_exit=U2_in*t_2
```

In the case of three slices considered here, `U2_exit` is the final exit wave which then has to be propagated to the detector. For this, the spherical wave propagator is used:

```
1   U_det=IFT2D(FT2D(U2_exit)*np.conjugate(Propagator(area, z2, Lambda, U2_exit)))
```

Fig. 3.21a-d shows an example of non-overlapping disks in two different focal planes, reconstructed from a hologram simulated with the multislice algorithm presented above. The reconstruction has been carried out in the two focus planes using the iterative reconstruction algorithm. The objects appear in the correct plane, respectively, while the respective out-of-focus reconstruction of the other disk is also visible, albeit with less sharp edges. The small imperfections in the reconstruction of the in-focus disk are due to the contributions from the other slice. The imperfections in the reconstruction decrease with smaller slice separation (Fig. 3.21a-d).

Fig. 3.21 also shows that the reconstructed amplitude and phase values of the out-of-focus objects in a given slice are still close to the correct values for a slice separation range of both 5 nm (Fig. 3.21a-b) and 10 nm (Fig. 3.21c-d), i.e. over the thickness of a single protein. This is of relevance since the exact source-to-sample distance and thus the correct reconstruction distance is in general not known in experimental imaging situations
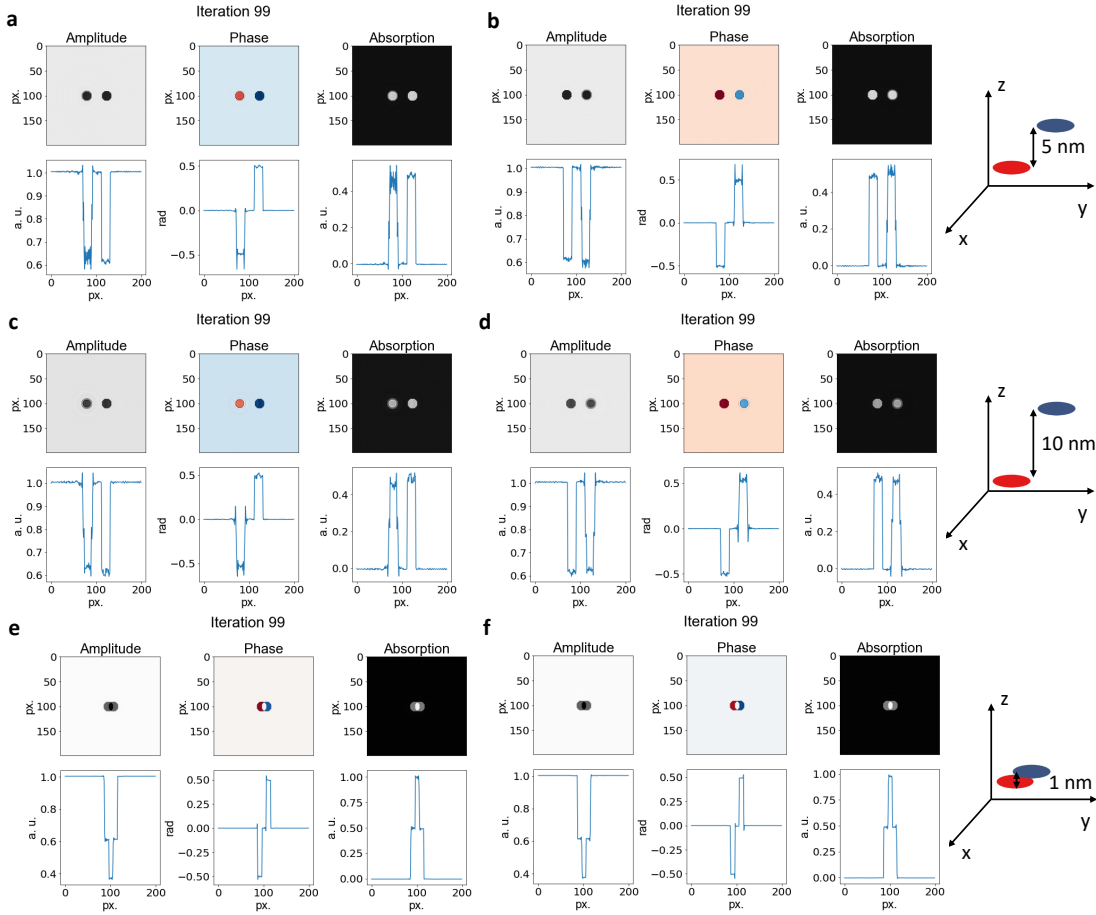
Figure 3.21: **Multislice algorithm: Non-overlapping and partially overlapping objects a**, **b** Iterative reconstruction a hologram simulated with the multislice algorithm of two non-overlapping disks separated along the optical axis by 5 nm (shown on the right) in their respective focus planes. Both disks have an absorption of $\alpha = 0.5$ and a phase shift of $\varphi = 0.5$ and $\varphi = -0.5$, respectively. In **a**, the disk on the right is in focus, in **b**, the disk on the left is in focus. **c**, **d** Iterative reconstruction of a hologram generated by the multislice algorithm of the two disks described in **a** and **b**, with an increased slice separation of 10 nm. In **c**, the focus plane of the right disk is depicted, while **d** shows the focus plane of the left disk. **e**, **f** Iterative reconstruction in the respective focus planes of a multislice-generated hologram of two overlapping disks with the same specifications as in **a**. The slice separation is 1 nm. The respective object geometries are sketched on the right.

and has to be retrieved by determining the focal plane of the object (see Chapter 1). While the example of two non-overlapping disks demonstrates that the algorithm can reconstruct the object in the correct z-plane and that the correct values of amplitude and phase can be retrieved in the focus plane, it is of greater interest to apply the multislice algorithm to objects that have (partially) overlapping parts in different slices, which could be a model for an object with a spatially varying thickness. Fig. 3.21e-f demonstrates that both phase and absorption values of the slices' transmission function are added in the multislice process. In the example presented in the figure, this results in higher absorption and zero phase shift in the overlap region. Since multiple scattering

becomes relevant in the low-energy electron imaging regime even for thin samples, the accumulation of amplitude and phase by subsequent interaction with different parts of the object is an important process to consider when interpreting experimental phase data.
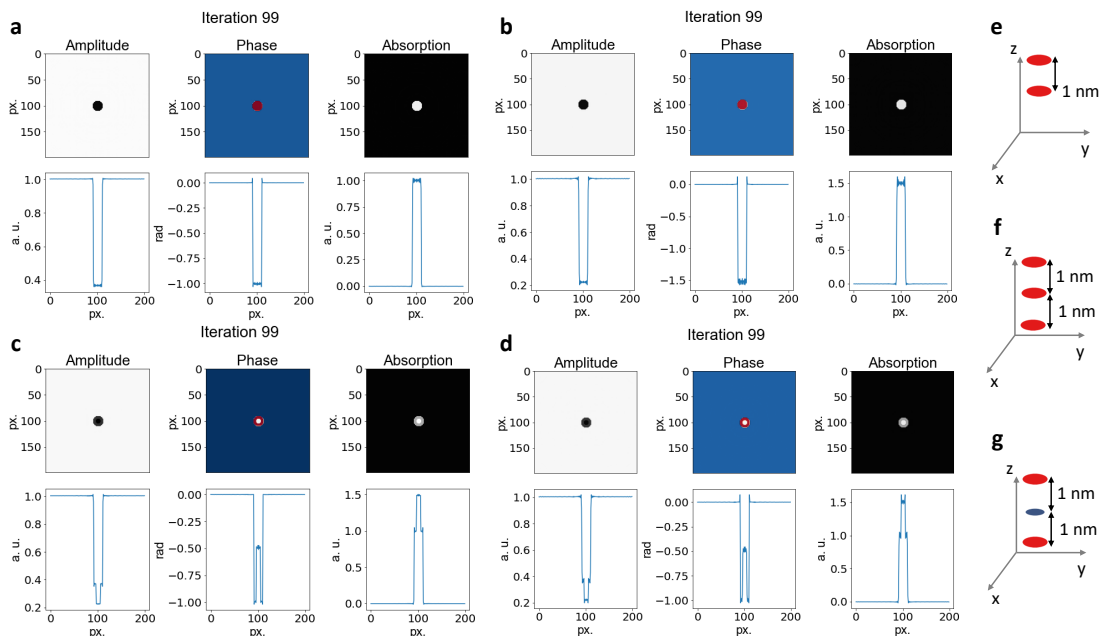

Figure 3.22: **Multislice algorithm: overlapping objects. a** Iterative reconstruction of a hologram simulated with the multislice algorithm from two identical disks with absorption $\alpha = 0.5$ and phase $\varphi = -0.5\,\mathrm{rad}$ separated along the optical axis by $1\,\mathrm{nm}$; the geometry is sketched in **e**. The amplitude, absorption and phase values of the slices are accumulated. The reconstruction is shown in the focal plane of the top disk. **b** Iterative reconstruction of a hologram simulated with the multislice algorithm from three identical disks with absorption $\alpha = 0.5$ and phase $\varphi = -0.5\,\mathrm{rad}$, separated along the optical axis by $1\,\mathrm{nm}$; the corresponding geometry is sketched in **f**. The reconstruction is shown in the focal plane of the top disk. **c**, **d** Iterative reconstruction of a hologram generated by the multislice algorithm from an object consisting of three disks of different radii, as depicted in **g**. **c** and **d** show the reconstructions in the focal planes of the middle and the top disk, respectively. The slice separation is $1\,\mathrm{nm}$. The outer two slices are identical disks of radius 10 pixels and absorption $\alpha = 0.5$ and phase $\varphi = -0.5\,\mathrm{rad}$, the middle disk has a smaller radius of 5 pixels with absorption $\alpha = 0.5$ and phase $\varphi = +0.5\,\mathrm{rad}$.

In figures 3.22 and 3.23, the case of fully overlapping objects in different slices is examined. Overlapping identical disks, as shown in Fig. 3.22a-b yield the same result as a single disk with the accumulated absorption and phase values, i.e. a hologram simulated from two fully overlapping disks (Fig. 3.22e), with absorption $\alpha = 0.5$ and phase $\varphi = -0.5\,\mathrm{rad}$ each, yields a reconstructed object of absorption $\alpha = 1$ and phase $\varphi = -1\,\mathrm{rad}$ (Fig. 3.22a). The contributions of the two slices are not well-separable in the axial direction, i.e. the two focal planes are hard to distinguish and, as a result, the reconstructions are slightly noisier than what would be expected from a single-slice object. For fully overlapping, non-identical objects in different planes, the amplitudes and phases also add up as expected, but again, the different reconstruction planes cannot easily be distinguished,

as shown in Fig. 3.22c-d, which shows the reconstruction of an object built up of 3 slices with a slice separation of $1\,\mathrm{nm}$ as depicted in Fig. 3.22g. While the outer two slices are identical disks of radius 10 pixels and absorption $\alpha = 0.5$ and phase $\varphi = -0.5\,\mathrm{rad}$, the middle disk has a smaller radius of 5 pixels with absorption $\alpha = 0.5$ and phase $\varphi = +0.5\,\mathrm{rad}$.
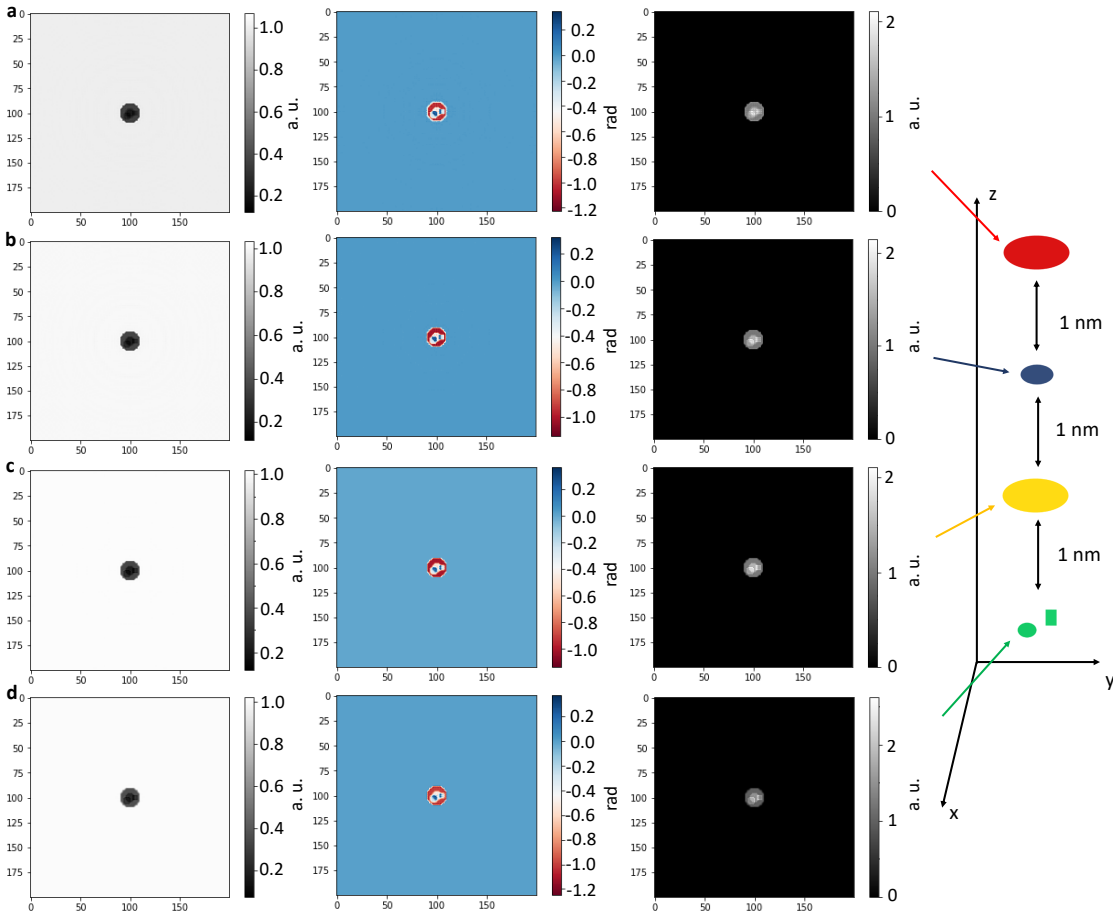

Figure 3.23: **Multislice algorithm: overlapping objects II** Iterative reconstruction of an object consisting of four different slices in the respective focal planes. The object geometry is sketched on the right, shapes of the same colour are located in the same z-plane. In all planes, all contributions are visible, but the sharpness of the edges allows the identification of the in-focus parts: **a** a disk of radius 10 pixels and absorption $\alpha = 0.5$ and phase $\varphi = -0.5\,\mathrm{rad}$ is in focus (red) **b** a disk of radius 5 pixels, absorption $\alpha = 0.5$ and phase $\varphi = +0.5\,\mathrm{rad}$ is in focus (blue) **c** a disk of radius 10 pixels and absorption $\alpha = 0.5$ and phase $\varphi = -0.5\,\mathrm{rad}$ is in focus (yellow) **d** a smaller disk (radius 4 pixels) and a square, whose centres are shifted away from the slice centre (green), both with absorption $\alpha = 0.3$ and phase $\varphi = 0.7\,\mathrm{rad}$, are in focus. The two identical disks in **a** and **c** are hard to distinguish via the sharpness of the edges.

To examine how well the different focus planes in an object with laterally overlapping parts in different z-planes can still be distinguished, Fig. 3.23 shows the reconstructed amplitude, phase and absorption of an object consisting of four different slices with overlapping content. The contributions of the different slices are a disk of radius 10

pixels and absorption $\alpha = 0.5$ and phase $\varphi = -0.5 \, \text{rad}$, followed by a disk of radius 5 pixels, absorption $\alpha = 0.5$ and phase $\varphi = +0.5 \, \text{rad}$, followed by a disk identical to the first one. The disks in all three slices are centred around the centre of the slice. The last slice is characterized by two smaller objects, a disk and a square, whose centres are shifted away from the slice centre, both with absorption $\alpha = 0.3$ and phase $\varphi = 0.7 \, \text{rad}$. For all reconstructed focal distances, all parts of the object are visible, the sharpness of the edges in amplitude and absorption differs, however, and thus indicates which part of the object actually is in focus. In the phase reconstruction, the planes are hard to distinguish. Additionally, the two identical overlapping objects (i.e. the two disks of radius 10 pixels in focus in Fig. 3.23a and c) are hard to distinguish via the sharpness of edges as already demonstrated in Fig. 3.22a-b.

To apply the multislice algorithm to an object close to an experimental structure of a protein, which will be discussed in the next chapter, a molecular model of the chaperone protein GroEL is examined in Fig. 3.24. Starting from the pdb structure 5W0S [210], the model can be divided into slices whose separation is calculated from the molecular thickness and the chosen number of slices. The atoms in one slice are then projected into one plane, these planes can subsequently be used as input for the multislice algorithm. Fig. 3.24 compares the reconstruction from a hologram with all atoms projected into a single plane (Fig. 3.24b) and that from a hologram generated by the multislice algorithm from 5 slices with a slice separation of $3 \, \text{nm}$ (Fig. 3.24c). The 7-fold symmetry of the molecule is present on the hologram level in both the single slice and the multislice hologram. The iterative reconstruction obtained from the single-slice hologram accurately retrieves the density variations mapped by the projection (Fig. 3.24d). The iterative reconstruction calculated from the multislice hologram shows lower contrast than the single-slice hologram reconstruction, but the three-dimensional information imprinted upon the hologram by the multislice simulation can be partially retrieved, as can be seen when comparing the iterative amplitude (Fig. 3.24f) and phase reconstructions (Fig. 3.24g) in the respective z-planes to the slice projections in Fig. 3.24e. While the individual slices are not fully recovered, structural details pertaining to the respective slices are discernible. The fuzzier appearance of the reconstructions of the multislice hologram as compared to the single-slice hologram can be attributed to the out-of-focus contributions from the other slices.

Thus, multislice simulations are a way of incorporating three-dimensional information into the hologram simulation process. Since the assumption that an object like a protein can be described by a two-dimensional transmission function is quickly stretched to its limits in an experimental context due to the presence of multiple scattering events, multislice simulations can help creating more accurate models for protein holograms that could be used for comparison with the experimental data.

Fig. 3.22 - Fig. 3.24 demonstrate that, as in the case of non-overlapping objects (Fig. 3.21), out-of-focus phase and amplitude data also contribute to the reconstruction plane for fully overlapping objects. While this hinders an accurate axial resolution, since it is more difficult to ascertain which part of the objects belongs to which focal plane in cases of significant lateral overlap, experimentally, this can be beneficial since the information is not lost in cases in which the source-to-sample distance cannot be exactly determined. Although the quantitative values will lose accuracy the further the distance from the focal
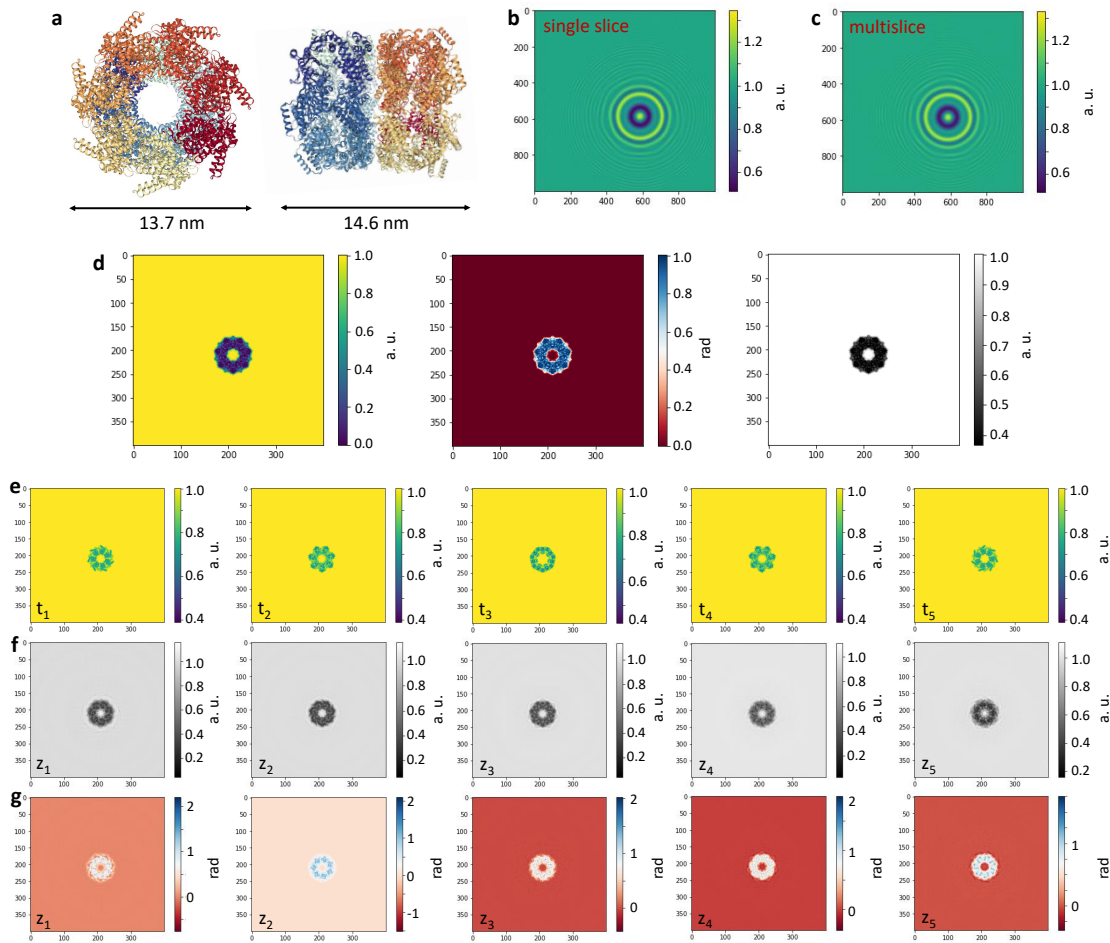
Figure 3.24: **Multislice algorithm: protein example a** PDB structure 5W0S of GroEL [210] both in top and side view. The projections used in **d** and **e** show the top view. **b** Hologram simulated from a single-slice transmission function of the PDB structure 5W0S of the protein GroEL [210], for which all the atoms in the PDB model have been projected into one plane (shown in **d**). **c** Hologram simulated with the multislice algorithm from projections of the molecular model into 5 slices (shown in **e**) with a slice separation of 3 nm to match the molecular thickness. **d** Single-slice projection of the PDB model 5W0S used as input for the single-slice hologram in **b** and the iterative phase and amplitude reconstruction after 100 iterations of the single-slice hologram. **e** Projections of the PDB model 5W0S into 5 slices which are used as input for the multislice algorithm. **f** Iterative amplitude reconstructions of the multislice hologram after 100 iterations in the respective z-planes. **g** Iterative phase reconstructions of the multislice hologram after 100 iterations in the respective z-planes.

plane, the qualitative features are retained over an axial range of several nanometers. Because of these additional contributions, however, the reconstruction algorithm employed here is not perfectly suited to deal with objects described by three-dimensional transmission functions. Ideally, such holograms would be reconstructed using an algorithm that takes the inherently three-dimensional nature of the objects into account. Possible approaches to this include multislice iterative phase retrieval [207] and deconvolution approaches [211, 212], the latter will be discussed in Chapter 5.

In general, the artefacts that can occur in the iterative phase reconstruction algorithm for a single slice hologram can also occur during the iterative reconstruction of multislice holograms since the reconstruction part is identical. Since the absorption values of the transmission functions of each individual slice are added up, it is easy to create strongly absorbing objects, hence in simulations, the absorption values need to be chosen carefully.

The multislice algorithm presented in this section can be considered as a model that incorporates multiple scattering processes and information about an object's three-dimensional structure into the hologram simulation. There is one important difference, however, to experimental multiple scattering processes: in the experiment, many multiple scattering events will be inelastic, which in most cases means that they lose coherence with respect to the reference wave. This implies that their contribution to the hologram is not to the fringe pattern, but only towards the absorptive properties. The algorithm discussed here, however, does not include inelastic effects, hence it implements elastic multiple scattering processes.

3.3.2 Scattering algorithm

So far, the absorption and phase shift values in the transmission function have been chosen such that they are suitable for test purposes, and the interaction of the incident wave and the object has been described classically. Even in the case shown in Fig. 3.24, where different absorption and phase shift values could in principle be assigned to atoms of different elements, the assigned values are not based on the actual interaction of the electrons with the atoms of the protein.

The scattering amplitudes and phase shifts of electrons interacting with atoms can be calculated using partial waves [213, 214, 103], which allows a fully quantum mechanical description of the interaction. With this approach, rather than using a transmission function defined by a two-dimensional array with absorption and phase values assigned to each pixel, the object can be built up atom by atom. The reference wave then interacts with each atom individually, the scattered waves resulting from each interaction are propagated to the screen and superimposed with the unaltered reference wave, which creates a hologram.

In general, the time evolution of the wave function of a particle moving in a potential, e. g. an electron in an atomic potential, can be described by stationary states of well-defined energy E

$$\Psi(\mathbf{r}, t) = \psi(\mathbf{r})e^{-i\frac{E}{\hbar}t}, \tag{3.25}$$

where $\psi(\mathbf{r})$ are solutions of the time-independent Schrödinger equation, i. e. energy eigenstates of the corresponding Hamiltonian [215].

Scattering processes can be delineated in terms of stationary scattering states that are examined at temporal and spatial coordinates at which the particle can be considered as asymptotically free. For large negative values of t, i. e. before the interaction with the potential, the free particle can be represented as a plane wave e^{ikz}, whereas, at large positive times t, when the particle has moved away from the potential after interacting, the particle's wave function can be described by a superposition of a transmitted and a

scattered wave

$$\psi_k(\mathbf{r}) \overset{r \to \infty}{\sim} e^{ikz} + f(\theta, \phi)\frac{e^{ikr}}{r}, \tag{3.26}$$

where $f(\theta, \phi)$ is the scattering amplitude that depends on the form of the potential [215, 216].

In the case of a central potential, the orbital angular momentum $\mathbf{L}$ is conserved. Hence, the Hamiltonian of a particle moving in a central potential has eigenstates with well-defined angular momentum of the form

$$\psi_{k,l,m}(\mathbf{r}) = \sqrt{\frac{2k^2}{\pi}} j_l(kr) Y_l^m(\theta, \phi), \tag{3.27}$$

whose angular dependence is given by the spherical harmonics $Y_l^m(\theta, \phi)$ and whose radial component is given by the spherical Bessel function $j_l(r)$ [215, 217]. These eigenstates can be used as a basis in which to expand the incoming wave and thus to describe the scattering interaction via the partial wave amplitude $f(\theta)$. Since the interaction with the potential induces an additional, angular momentum-dependent phase shift δ_l, the scattering amplitude and scattering cross section can be expressed as a function of these phase shifts in a partial wave analysis [215].

An expression for the complex-valued partial wave amplitude $f(\theta)$ can be derived by considering the scattering of an incoming plane wave e^{ikz}, with wave fronts perpendicular to the z-axis, from a spherically symmetric potential $V(\mathbf{r}) = V(r)$.

Expanding the incoming plane wave in terms of the angular momentum eigenstates and applying the spherical harmonic addition theorem [215, 217] yields

$$e^{ikz} = e^{ikr\cos\theta} = \sum_{l=0}^{\infty} i^l \sqrt{4\pi(2l+1)} j_l(kr) Y_l^0(\theta) \tag{3.28}$$

$$= \sum_{l=0}^{\infty} i^l (2l+1) j_l(kr) P_l(\cos\theta). \tag{3.29}$$

Since e^{ikz} does not depend on ϕ, it is composed of angular momentum eigenstates with $m = 0$. P_l is a Legendre polynomial of degree l, which is related to $Y_l^0(\theta)$ by

$$Y_l^0(\theta) = \sqrt{\frac{2l+1}{4\pi}} P_l(\cos\theta). \tag{3.30}$$

Since the sum occurring in the expansion in eq. (3.29) includes all angular momenta up to infinity, this does on first glance appear computationally costly. It is, however, often sufficient to only include a small number of l up to a maximum angular momentum l_{max} in the sum, especially at low energies. This can be heuristically understood in analogy to the impact parameter b in classical scattering [217]. The angular momentum of an incoming particle at a distance b from the axis on which the scattering target is placed can be approximated by

$$\hbar l \cong \hbar k b. \tag{3.31}$$

Hence, for a potential of finite range r_0, particles with large angular momenta associated with $b > r_0$ do not interact with the potential (in the classical analogue, the particle would miss the target). Hence a maximum angular momentum that contributes to the scattering can be defined:

$$l_{max} = kb_{max} \cong kr_0. \tag{3.32}$$

A rigorous derivation based on the properties of the angular momentum eigenstates yields the quantum mechanical impact parameter analogue $b_l = \frac{\sqrt{l(l+1)}}{k}$ [215].
Considering the asymptotic case $r \to \infty$, for which the scattering states take the form discussed above, eq. (3.29) can be simplified by using the approximate form of the spherical Bessel function for large r:

$$j_l(kr) \overset{r\to\infty}{\Rightarrow} \frac{\sin(kr - l\frac{\pi}{2})}{kr}. \tag{3.33}$$

The asymptotic case is valid when the distances between the scatterer and the points where the wave is evaluated (the screen and the emitter) are much larger than the range of the potential. For the scattering at an atomic potential, this is the case in an in-line holography set-up.
Inserting this into eq. (3.29) and expressing the sine function in terms of exponentials by Euler's formula results in

$$e^{ikz} \overset{r\to\infty}{\Rightarrow} \frac{1}{2ik} \sum_{l=0}^{\infty} i^l (2l+1) \left[\frac{e^{i(kr-l\frac{\pi}{2})}}{r} - \frac{e^{-i(kr-l\frac{\pi}{2})}}{r} \right] P_l(\cos\theta) \tag{3.34}$$

$$= \frac{1}{2ik} \sum_{l=0}^{\infty} (2l+1) \left[\frac{e^{ikr}}{r} - \frac{e^{-i(kr-l\pi)}}{r} \right] P_l(\cos\theta), \tag{3.35}$$

where $i^l = e^{i\frac{\pi}{2}l}$ has been used in the last step.
With this, the asymptotic form of the full wave function (eq. (3.26)) becomes

$$\psi_k(\mathbf{r}) \overset{r\to\infty}{\Rightarrow} \frac{1}{2ik} \sum_{l=0}^{\infty} (2l+1) \left[\frac{e^{ikr}}{r} - \frac{e^{-i(kr-l\pi)}}{r} \right] P_l(\cos\theta) + f(\theta)\frac{e^{ikr}}{r} \tag{3.36}$$

$$= -\frac{1}{2ik} \sum_{l=0}^{\infty} (2l+1) \frac{e^{-i(kr-l\pi)}}{r} P_l(\cos\theta) + \frac{e^{ikr}}{r} \left(f(\theta) + \frac{1}{2ik} \sum_{l=0}^{\infty} (2l+1) P_l(\cos\theta) \right) \tag{3.37}$$

In the last step, all terms proportional to $\frac{e^{ikr}}{r}$ have been grouped together.
The wave function likely changes significantly during the interaction. Still, the radial part at large distances $r \to \infty$ can be approximated by a phase-shifted free-particle wave function, again making use of the properties of the spherical Bessel function:

$$R_l(r) = \frac{U_l(r)}{r} \overset{r\to\infty}{\Rightarrow} A_l \frac{\sin(kr - l\frac{\pi}{2} + \delta_l(k))}{kr}, \tag{3.38}$$

105

where $\delta_l(k)$ is the phase shift induced by the potential. Thus, the wave function $\psi_k(\mathbf{r})$ can be expressed in terms of the radial basis as follows:

$$\psi_k(\mathbf{r}) \overset{r \to \infty}{\to} \sum_{l=0}^{\infty} A_l \frac{\sin(kr - l\frac{\pi}{2} + \delta_l(k))}{kr} P_l(\cos\theta) \tag{3.39}$$

$$= \frac{1}{2ik} \sum_{l=0}^{\infty} A_l \left[\frac{e^{i(kr - l\frac{\pi}{2} + \delta_l)}}{r} - \frac{e^{-i(kr - l\frac{\pi}{2} + \delta_l)}}{r} \right] P_l(\cos\theta). \tag{3.40}$$

The coefficients A_l can now be determined by comparison to eq. (3.37). Since the term associated with the incoming plane wave remains unaltered by the scattering process, the coefficients of the terms proportional to e^{-ikr} have to match in equations (3.37) and (3.40). This requirement directly yields an expression for A_l:

$$A_l = (2l + 1)e^{il\frac{\pi}{2}} e^{i\delta_l} \tag{3.41}$$

Inserting this into eq. (3.40) and comparing the remaining terms in eq. (3.40) and eq. (3.37) yields the partial wave amplitude $f(\theta)$:

$$f(\theta) + \frac{1}{2ik} \sum_{l=0}^{\infty} (2l + 1) P_l(\cos\theta) = \frac{1}{2ik} \sum_{l=0}^{\infty} (2l + 1)e^{2i\delta_l} P_l(\cos\theta) \tag{3.42}$$

$$\Rightarrow f(\theta) = \frac{1}{2ik} \sum_{l=0}^{\infty} (2l + 1) \left(e^{2i\delta_l} - 1 \right) P_l(\cos\theta) \tag{3.43}$$

$$= \frac{1}{k} \sum_{l=0}^{\infty} (2l + 1)e^{i\delta_l} \sin(\delta_l) P_l(\cos\theta). \tag{3.44}$$

In the last step, the exponential term has been rewritten as

$$e^{2i\delta_l} - 1 = e^{i\delta_l}(e^{i\delta_l} - e^{-i\delta_l}) = e^{i\delta_l} 2i \sin(\delta_l). \tag{3.45}$$

From $f(\theta)$, expressions for the total scattering cross section σ and the differential cross section $\frac{d\sigma}{d\Omega}$ can be obtained:

$$\frac{d\sigma}{d\Omega} = |f(\theta)|^2 \tag{3.46}$$

$$\Rightarrow \sigma = \int \frac{d\sigma}{d\Omega} d\Omega = \int_0^{2\pi} \int_0^{\pi} |f(\theta)|^2 \sin\theta d\theta d\phi \tag{3.47}$$

$$= \frac{2\pi}{k^2} \sum_{l=0}^{\infty} \sum_{l'=0}^{\infty} (2l + 1)(2l' + 1)e^{i(\delta_l - \delta_{l'})} \sin(\delta_l) \sin(\delta_{l'}) \int_0^{\pi} P_l(\cos\theta) P_{l'}(\cos\theta) \sin\theta d\theta \tag{3.48}$$

With the orthogonality relation for Legendre polynomials [217],

$$\int_0^{\pi} P_l(\cos\theta) P_{l'}(\cos\theta) \sin\theta d\theta = \frac{2}{2l + 1} \delta_{ll'}, \tag{3.49}$$

this results in

$$\sigma = \frac{4\pi}{k^2} \sum_{l=0}^{\infty} (2l + 1) \sin^2(\delta_l). \tag{3.50}$$

In the more general case of particles with spin, two scattering amplitudes have to be considered: the direct scattering amplitude $f(\theta)$ and the spin-flip scattering amplitude $g(\theta)$, which are of the form [218, 214, 213]

$$f(\theta) = \frac{1}{2ik} \sum_{l=0}^{\infty} \left[(l+1) \left(e^{2i\delta_l^+} - 1 \right) + l \left(e^{2i\delta_l^-} - 1 \right) \right] P_l(\cos\theta) \qquad (3.51)$$

and

$$g(\theta) = \frac{1}{2ik} \sum_{l=0}^{\infty} \left(e^{2i\delta_l^-} - e^{2i\delta_l^+} \right) P_l^1(\cos\theta), \qquad (3.52)$$

where $P_l(\cos\theta)$ are the Legendre polynomials and $P_l^1(\cos\theta)$ are the associated Legendre polynomials.

For $\delta_l^+ = \delta_l^-$, $g(\theta) = 0$ and $f(\theta)$ reduces to the expression derived in eq. (3.44).

The NIST Electron Elastic-Scattering Cross-Section Database [218] provides the phase shifts δ_l^+ and δ_l^- and the associated cross sections for the elastic interaction of electrons with energies between 50 eV and 300 keV and target atoms of atomic number Z, calculated by a Dirac partial wave analysis. The potential used to model the interaction includes the following terms [213]:

$$V(r) = V_{st}(r) + V_{ex}(r) + V_{cp}(r) - iW_{abs}(r). \qquad (3.53)$$

$V_{st}(r)$ describes the electrostatic potential which is dominated by Coulomb interaction. $V_{ex}(r)$ is the exchange potential representing the interaction of identical particles. The correlation-polarization term $V_{cp}(r)$ characterizes the polarization of the target charge distribution due to the electric field of the electrons and the resulting effect of the induced dipole on the incoming electrons. $W_{abs}(r)$ describes the absorptive potential that accounts for a partial loss of elastic electron flux due to inelastic effects. The last three terms are especially relevant in the low-energy regime.

Since the phase shift for individual elements are known, holograms of objects built up from individual atoms can be simulated based on the scattering amplitude $f(\theta)$. The wave function at the detector then takes the form [103]

$$U_{\text{det}}(\mathbf{R}) = A_{\text{ref}} \frac{e^{ikR}}{R} + f(\theta) \sum_{j=1}^{N} \frac{e^{ikr_j}}{r_j} \frac{e^{ik|\mathbf{r}_j - \mathbf{R}|}}{|\mathbf{r}_j - \mathbf{R}|}, \qquad (3.54)$$

where $\mathbf{R}$ is a position on the screen (relative to the emitter position), $\mathbf{r}_j$ is the position of the jth atom in the sample, N is the number of atoms in the sample, and A_{ref} is the amplitude of the reference wave.

Numerically, this can be implemented with the help of the SciPy function `legendre` as a representation of the Legendre polynomials. For an atom `s` within the molecule and an incoming electron with wave vector `k`, the scattering amplitude is calculated as follows:

```
1  table = atomset.index(atom[s]) #select phase shift table for element
2      lgd=np.zeros((pixel*pixel)) # initialise Legendre polynomials
3      f = lgd # initialise scattering amplitude
4      # Calculate scattering amplitude
```

```
5       for l in range(0, lmax):
6           Pn=legendre(l)
7           lgd=Pn(cosTheta)
8           f=(2*l+1)*np.sin(phase[table,l,1])*np.exp(1j*phase[table,l,1])*lgd/k + f
```

In line 1, the correct phase shift table for the atomic species of the atom `s` is selected, in the following two lines, the arrays for the Legendre polynomials `lgd` and the scattering amplitude `f` are initialised. The for-loop then calculates the scattering amplitude according to eq. (3.44), for angular momenta `l` from 0 to a maximum angular momentum `lmax`. While theoretically, any value for `lmax` could be chosen, it is practically limited by the availability of the corresponding phase shifts in the NIST data base. For light elements such as carbon and nitrogen, phase shifts at low energies of 100 eV are available up to $l = 25$. `phase[table,l,1]` denotes the phase shift corresponding to the correct atomic number (as given by `table`) and the angular momentum `l`. Since the NIST phase shifts tables include both δ_l^+ and δ_l^-, the last argument of `phase` selects which phase shift to use, 1 stands for δ_l^+, -1 for δ_l^-. Since δ_l^+ and δ_l^- are almost identical for most l (except at $l = 0$, where δ_l^- is often 0) at the energies considered here, both phase shifts could be chosen when implementing eq. (3.44), because of the discrepancy at $l = 0$, δ_l^+ has been chosen here.

Alternatively, eq. (3.51) with both δ_l^+ and δ_l^- can be implemented as

```
1   table = atomset.index(atom[s]) #select phase shift table for element
2       lgd=np.zeros((pixel*pixel)) # initialise Legendre polynomials
3       f = lgd # initialise scattering amplitude
4       # Calculate scattering amplitude
5       for l in range(0, lmax):
6           Pn=legendre(l)
7           lgd=Pn(cosTheta)
8           f=1/(2*1j*k)*(((l+1)*(np.exp(2*1j*phase[table,l,1])-1)
9               +l*(np.exp(2*1j*phase[table,l,-1])-1))*lgd) + f
```

Because the differences between δ_l^+ and δ_l^- are mostly small, the scattering amplitudes also yield very similar values.

In both cases, the argument of the Legendre polynomials `cosTheta` has to be calculated point by point for each scatterer. For a given scatterer and a given point on the screen, the scattering angle θ is determined by the triangle spanned by the scatterer, the point on the screen, and the direct projection of the scatterer to the screen. $\cos\theta$ can then be calculated by the law of cosines as

```
1   cosTheta=(directlength**2+rss**2-screenlength**2)/(2*directlength*rss)
```

where `rss` is the distance between the scatterer and the point on the screen, `directlength` is the distance of the scatterer to its direct projection on the screen, and `screenlength` is the distance between the point on the screen defined by `rss` and the point on the screen corresponding to the direct projection of the scatterer.

Since `cosTheta` has to be calculated for each point on the screen for a given scatterer `s`, the distances are most efficiently calculated by expressing the distance vectors in matrix form and exploiting pre-compiled `numpy.array` functions to achieve acceptable running times.

To check that the scattering amplitude is calculated correctly, the differential cross section can be calculated from $f(\theta)$ and compared to the cross sections provided by the NIST

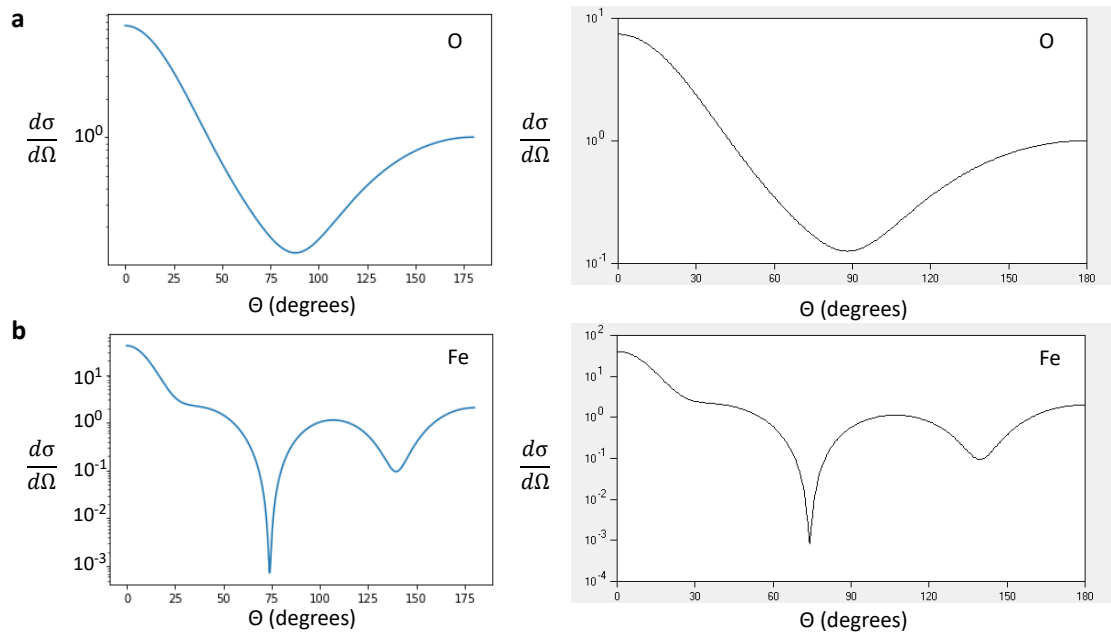data base. The cross sections match precisely, as shown in Fig. 3.25 for the examples of oxygen and iron.

Figure 3.25: **Differential scattering cross sections: a** Differential scattering cross section for oxygen (O) at an electron energy of 100 eV, calculated with the code presented above (left) and screenshot of the corresponding values taken from the NIST database [218] (right). **b** Differential scattering cross section for iron (Fe) at an electron energy of 100 eV, calculated with the code presented above (left) and screenshot of the corresponding values taken from the NIST database [218] (right).

Once the scattering amplitude $f(\theta)$ is calculated, the complex wave field at the detector can be simulated (Fig. 3.26a) by implementing eq. (3.54) as

```
1  Udet = p+ref
```

with the scattered wave

```
1  p = (a0/rse) * np.exp(1j*(k*rse+phi0))*(np.exp(1j*k*rss )/rss)*f + p
```

and the reference wave

```
1  ref = (a0/r) * np.exp(1j*(k*r + phi0))
```

where `a0` and `phi0` are the initial amplitude and phase of the reference wave, `rse` is the distance between a given scatterer and the emitter and `rss` is the distance between the scatterer and a given point of the screen, as above.

Since the amplitude of the reference wave is known, the hologram can be normalized by division by the amplitude of the reference wave (Fig. 3.26b):

```
1  holoabs=abs(holo)
2  absref=abs(ref)
3  holobackdiv=(holoabs**2)/(absref**2)
```
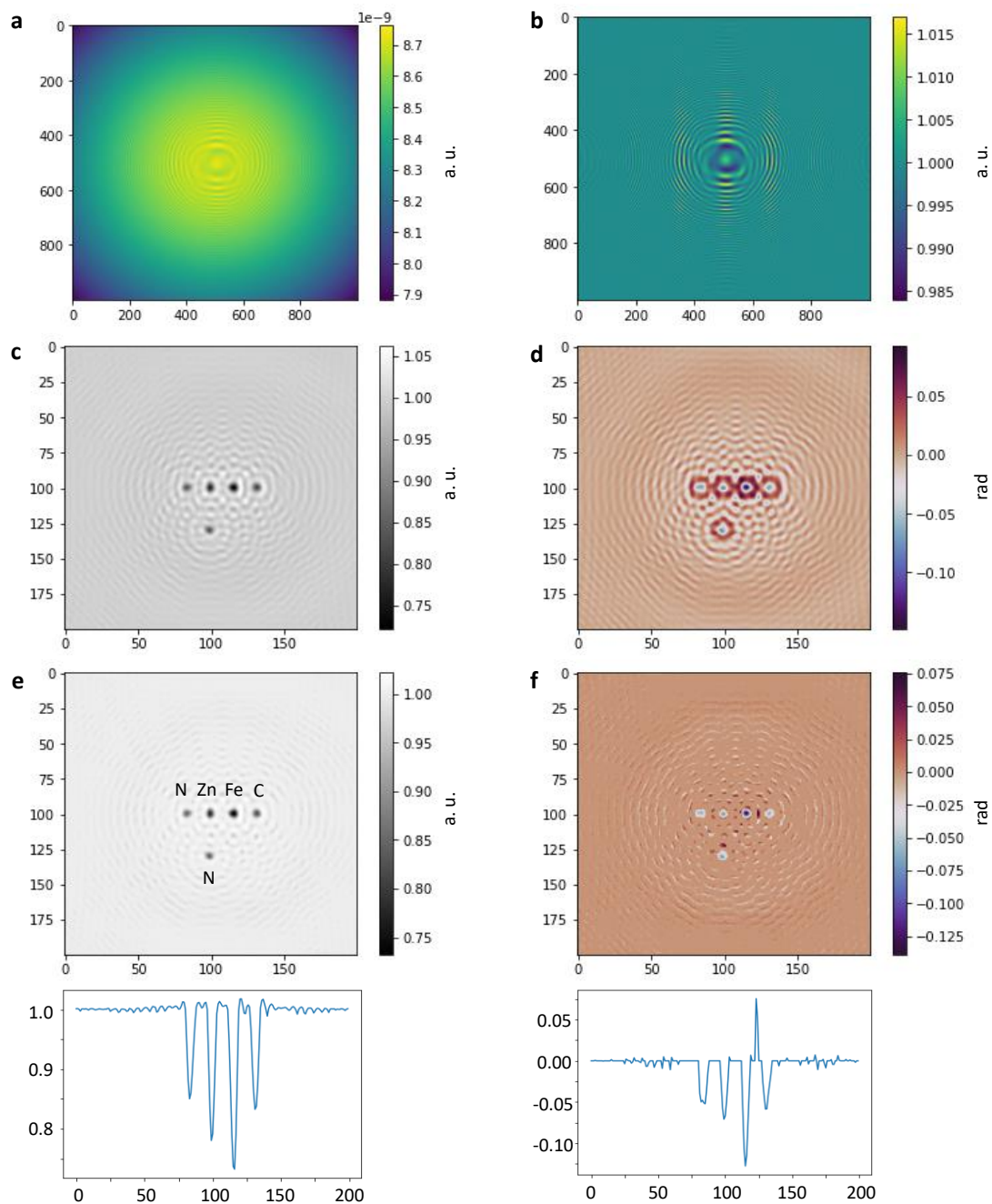
109

Figure 3.26: **Scattering simulation for individual atoms: a** Amplitude of the complex wave field at the detector and **b** normalized hologram, simulated with the scattering algorithm from the object reconstructed in **c-f**. **c** One step amplitude reconstruction and **d** one-step phase reconstruction at $z_0 = 96\,\text{nm}$ of the hologram shown in **b**. Despite $z_0 = 100\,\text{nm}$ being the simulated source-to-sample distance, the atoms appear in focus at $z_0 = 96\,\text{nm}$. The different atomic species (labelled in **e**) appear at different contrast in both amplitude and phase. **e-f** Iterative amplitude and phase reconstructions after 50 iterations of the hologram in **b** and corresponding cross sections through the centre of the image.

The hologram can then be reconstructed by employing the code presented in Chapter 1. The reconstruction is in principle successful, since even the one-step algorithm yields reconstructions in which different atomic species can be distinguished in both amplitude and phase (see Fig. 3.26c-d). As expected, atoms with high atomic number scatter more strongly than elements with low atomic number, as exemplified in Fig. 3.26c-d, where the metal atoms Fe and Zn appear darker than the C and N atoms in both amplitude and phase.

The iterative reconstruction algorithm can also be applied, while it does reduce the fringes, it does not remove them completely as in the case of the holograms simulated with the convolution-based algorithm (Fig. 3.26e-f). The reason for this could be the conceptual difference between the two algorithms used for simulation (via scattering) and reconstruction (via convolution). As a result of that, the constraints applied during the iterative reconstruction are likely not ideal. This conceptual difference also complicates the quantitative evaluation of the reconstruction: to compare the values reconstructed by the convolution approach to the scattering amplitudes, one would need to calculate a transmission function from the scattering amplitudes directly, i.e. not via simulation and reconstruction of a hologram, which is not possible since the form of the wave function used in the scattering approach is only valid for large distances from the object. Additionally, in the convolution approach, a constant factor such as the initial reference wave amplitude a0, does not qualitatively change the reconstruction. The relative pixel values will thus remain the same even if the absolute values of the maximum and minimum pixel intensities change. Hence, the hologram can be normalized without a qualitative change in the reconstruction, which is of particular relevance for the iterative reconstruction algorithm since some of the constraints are only applicable for certain value ranges of the hologram.
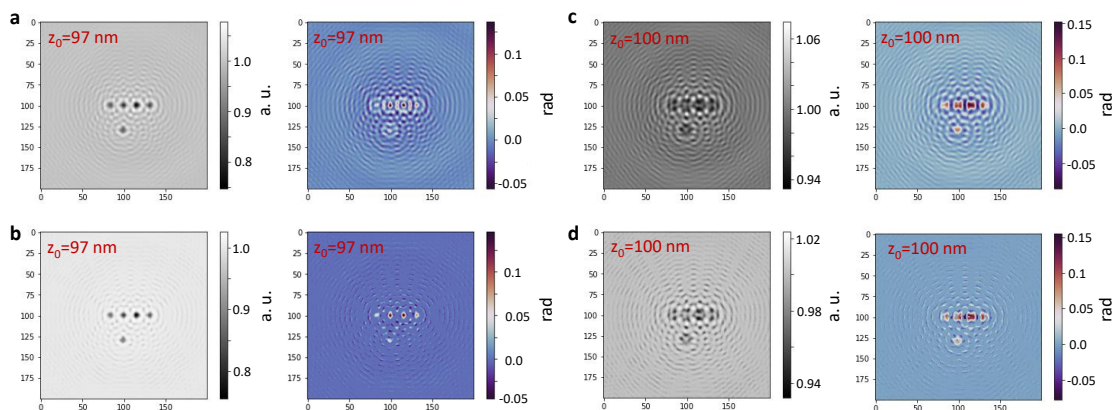

Figure 3.27: **Scattering algorithm: focus a** One-step amplitude and phase reconstruction of the hologram shown in Fig. 3.26b at $z_0 = 97\,\mathrm{nm}$. **b** Iterative amplitude and phase reconstruction after 50 iterations of the hologram shown in Fig. 3.26b at $z_0 = 97\,\mathrm{nm}$. While the relative phases are similar, the sign of the phases is reversed in comparison to amplitude and phase reconstruction of the hologram shown in Fig. 3.26d, f. **c** One-step amplitude and phase reconstruction of the hologram shown in Fig. 3.26b at $z_0 = 100\,\mathrm{nm}$. **d** Iterative amplitude and phase reconstructions after 50 iterations of the hologram shown in Fig. 3.26b at $z_0 = 100\,\mathrm{nm}$. While this should be the focal plane corresponding to the distance between source and sample plane, the reconstruction appears out of focus.

Additionally, the focusing of the reconstruction is met with some difficulties. Since the source-to-sample distance is in principle known (100 nm in the images presented here), the focused image of an object consisting of atoms in one plane would be expected to appear at the simulated source-to-sample distance. When applying the convolution approach as described in Chapter 1 to the holograms simulated with the scattering algorithm, however, the atoms appear in focus (in the sense in which a transmission function object would be focused) at source-to-sample distances that differ from the expected value by several nanometers. Fig. 3.27 shows the reconstruction of the same object shown at a reconstruction distance of 96 nm in Fig. 3.26, an array of five atoms of four different atomic species (N, Zn, Fe and C), for the reconstruction distances 97 nm and 100 nm. While $z_0 = 100$ nm would be the expected focus distance, the object appears in focus at $z_0 = 96/97$ nm (Fig. 3.26c-f, Fig. 3.27a-b). It is, however, not clear which of the two distances should be considered as the focal plane. The amplitude reconstructions at both distances are similar, but while the relative reconstructed phases are also similar for the two distances, the absolute phases are not: at 96 nm, the reconstructed phases are negative, at 97 nm, they are positive (Fig. 3.26d,f, Fig. 3.27a-b).

This indicates that a different reconstruction procedure might be necessary for reconstructing holograms simulated with a scattering algorithm. The ambiguity in the focus of the reconstruction, specifically, could be due to the fact that the paraxial approximation cannot be applied in the scattering case, i.e. that the wave front at the detector is too curved to be approximated by a flat screen.

To test this hypothesis, the Fresnel-Kirchhoff integral (eq. (1.58)) has been evaluated directly:

```
1  for u in range (0,Dim):
2      pix=np.transpose(np.repeat(np.reshape(recPlane[u,:],(3,1)),
3                              ScreenPixel*ScreenPixel, axis=1))
4      rpsi=(np.sum(((pix-emitter)-(screen-emitter))**2,1))**0.5
5      psi[u]=1j/wavelen*np.sum(h*ref*np.exp(-1j*k*rpsi)/rpsi)
```

where `pix` is a pixel in the object plane, `h` is the normalized hologram and `ref` is the reference wave.

The results of this reconstruction are shown in Fig. 3.28a-b. The focus of the reconstructed object is indeed at 100 nm, however, a strong background pattern leads to errors in the reconstruction, especially in the phase reconstruction (Fig. 3.28b). The phase background can in part be removed by unwrapping the phase with the `skimage.restoration` function `unwrap_phase` (Fig. 3.28c).

Shifting the values of the normalized hologram by subtracting 1 (i.e. `h=holobackdiv-1` instead of `h=holobackdiv`), a normalization for example employed in references [73, 219], removes the pattern in the amplitude reconstruction while also inverting the contrast (Fig. 3.28d). The phase reconstruction, however, deteriorates when using holograms with the shifted values (Fig. 3.28e-f). Additionally, the direct reconstruction algorithm is computationally very costly, the reconstruction of a small image of 200×200 pixels takes approximately 600 s, while the convolution-based reconstruction of an image of the same size takes 0.8 s.
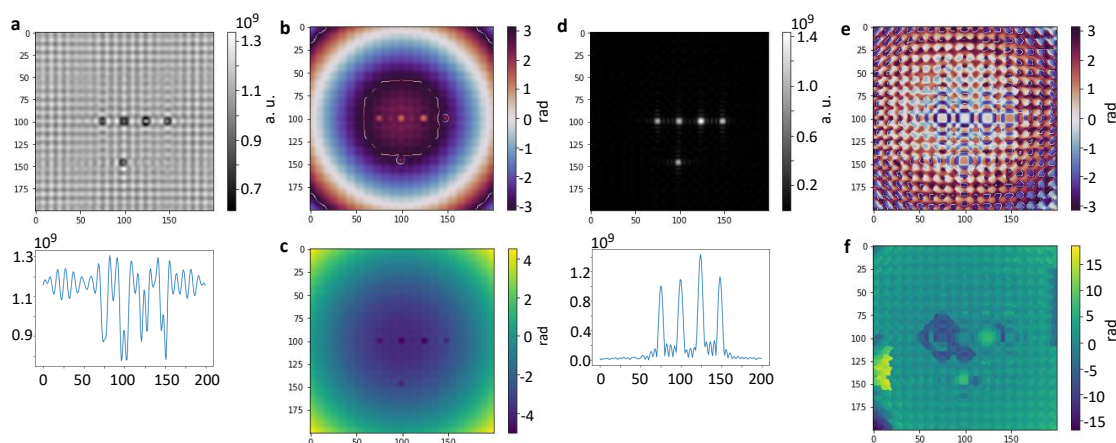
Figure 3.28: **Direct integral reconstruction a** Direct integral amplitude reconstruction of the hologram shown in Fig. 3.26b at $z_0 = 100$ nm and corresponding cross section through the centre of the image. **b** Direct integral phase reconstruction of the hologram shown in Fig. 3.26b at $z_0 = 100$ nm. The slowly varying background contribution can be partially removed by phase unwrapping as shown in **c**. **d** Direct integral amplitude reconstruction and corresponding cross section through the centre of the image of the shifted hologram (holobackdiv-1) at $z_0 = 100$ nm. **e-f** Direct integral phase reconstruction of the shifted hologram at $z_0 = 100$ nm, the directly reconstructed phase is shown in **e**, the unwrapped phase in **f**.

Despite these ambiguities in the reconstruction, the code can be used to simulate the interaction of the electron beam with molecules such as porphyrins or phtalocyanines with metal centres that result in locally stronger scattering. The example of platinum phtalocyanine (PtPc) is shown in Fig. 3.29, the stronger scattering by the metal centre is evident in both amplitude and phase.
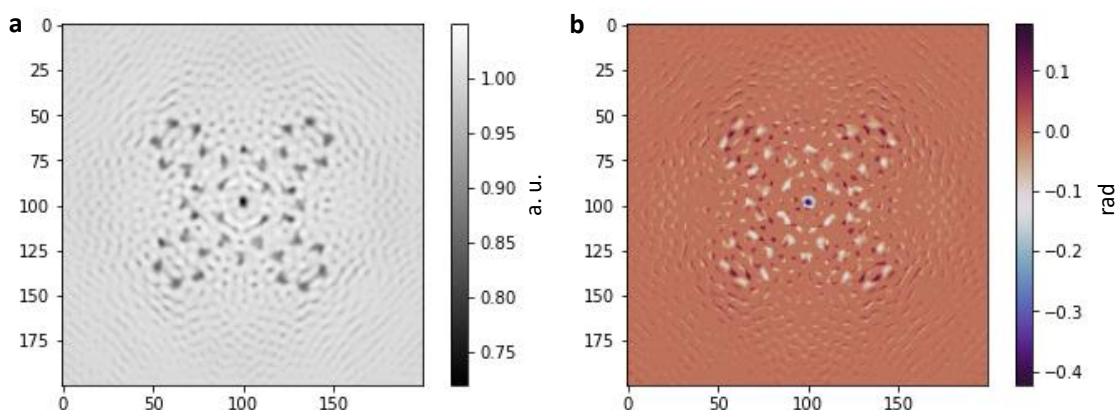
Figure 3.29: **Scattering algorithm: molecular example a** Iterative amplitude reconstruction after 50 iterations of a molecular model of PtPc at $z_0 = 96$ nm. **b** Iterative phase reconstruction after 50 iterations of a molecular model of PtPc at $z_0 = 96$ nm.

This shows that the algorithm can be used to model the effect of locally different scattering strengths and could be applied to produce simulations to compare to experimental data

of molecules such as PtPc. Such a comparison would not only provide insights relevant to understanding the origin of certain features in the phase reconstruction, but is also of relevance to the imaging of biological systems, since similar structures to the one simulated in Fig. 3.29 occur within biomolecules, for example in the form of heme groups. So far, however, the experimental results for these types of molecules have been ambiguous, mostly due to difficulties regarding sample preparation and insufficient resolution.

While it is unlikely that, at the current resolution, effects of a local change in scattering strength due to a single metal atom could be detected in large three-dimensional molecules such as proteins, the results of this section could be extrapolated to the imaging of local changes of potential on a scale larger than an individual atom.

In summary, this algorithm can be considered as an approximate model for the interaction of the low-energy electrons employed in LEEH and can be used for a qualitative comparison to experimental data.
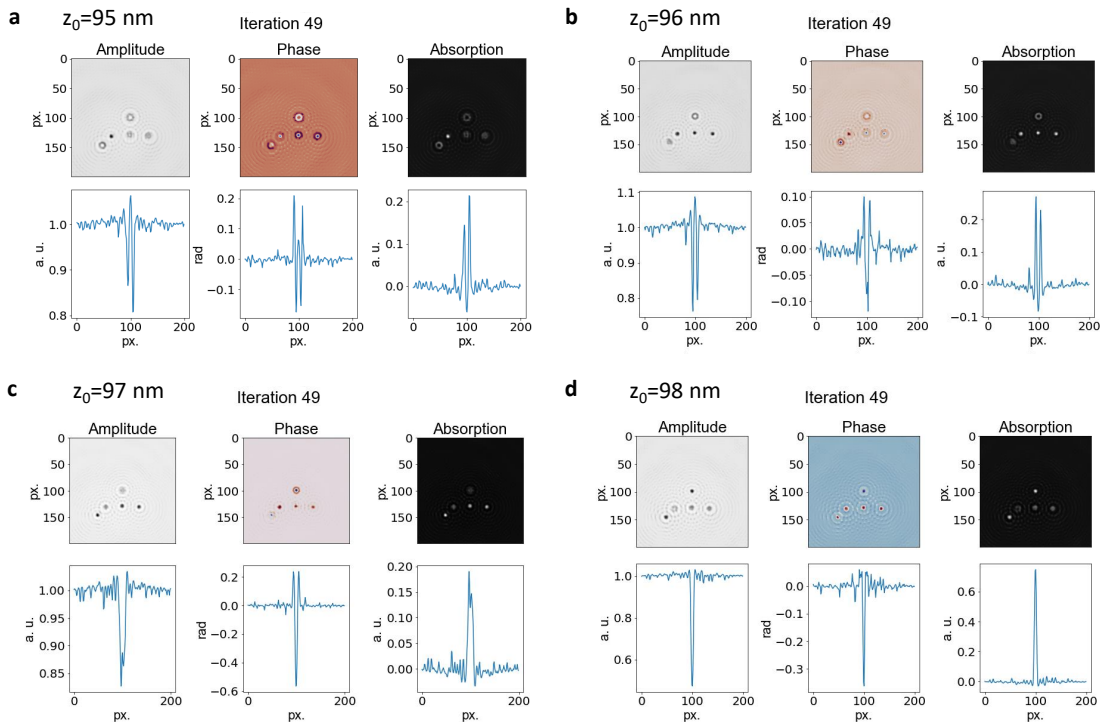

Figure 3.30: **Scattering algorithm: 3D input** Iterative amplitude, phase and absorption reconstruction after 50 iterations of an object consisting of 5 Tb atoms distributed in 4 different planes along the z-axis. Atom 1 is the closest to the emitter, atoms 2 and 3 are located in the same plane, 1nm further towards the screen than atom 1, followed by atom 4 and 5, whose z-planes are located 2 nm and 3 nm towards the detector with respect to atom 1. The reconstructions were obtained for **a** $z_0 = 95$ nm, **b** $z_0 = 96$ nm **c** $z_0 = 97$ nm, **d** $z_0 = 98$ nm. As expected, the atoms come into focus in the correct order. The cross sections are taken through the image centres and show atom 5 transitioning from being out of focus in **a** to being in focus in **d**.

While the simulation of flat molecules at atomic resolution has been demonstrated using PtPc in Fig. 3.29, the algorithm can also handle three-dimensional data, as shown by a simple three-dimensional object consisting of 5 atoms in four different planes in Fig. 3.30. This algorithm, however, only allows the modelling of single scattering events since the wave resulting from each scattering interaction is directly propagated to the detector.

Furthermore, while an inelastic term has been added to the potential used to calculate the phase shifts in the NIST data base [218, 213], the interaction modelled here is still primarily elastic, specifically, the scattering amplitude $f(\theta)$ has been derived for the case of elastic scattering. In principle, inelastic contributions can be included in the scattering amplitude. In the expression of the scattering amplitude given in eq. (3.43), the term $e^{2i\delta_l}$ can be associated with the scattering matrix for angular momentum l, S_l. In the case of elastic scattering, no flux is lost, hence $|S_l| = 1$. If absorption and inelastic effects are included, the number of elastically scattered particles is reduced, hence $|S_l| \leq 1$. Thus, S_l can be written as [217]

$$S_l = \eta_l e^{2i\delta_l}, \tag{3.55}$$

with $0 \leq \eta \leq 1$. With this, the scattering amplitude $f(\theta)$ takes the form

$$f(\theta) = \frac{1}{2ik} \sum_{l=0}^{\infty} (2l+1)\, (S_l - 1)\, P_l(\cos\theta) \tag{3.56}$$

$$= \frac{1}{2ik} \sum_{l=0}^{\infty} (2l+1) \left(\eta_l e^{2i\delta_l} - 1 \right) P_l(\cos\theta) \tag{3.57}$$

$$= \frac{1}{2ik} \sum_{l=0}^{\infty} (2l+1)\, (\eta_l(\cos(2\delta_l) + i\sin(2\delta_l)) - 1)\, P_l(\cos\theta) \tag{3.58}$$

$$= \frac{1}{2k} \sum_{l=0}^{\infty} (2l+1)\, (\eta_l \sin(2\delta_l) + i(1 - \eta_l \cos(2\delta_l)))\, P_l(\cos\theta). \tag{3.59}$$

The corresponding total elastic cross section is

$$\sigma_{el} = \frac{\pi}{k^2} \sum_{l=0}^{\infty} (2l+1) \left(1 + \eta_l^2 - 2\eta_l \cos(2\delta_l) \right). \tag{3.60}$$

Thus, even in the case of full absorption, $\eta_l = 0$, an elastic scattering contribution remains, which is of interest especially when considering complex and three-dimensional molecules like proteins, which could behave like strongly absorbing objects in the context of LEEH. Adding absorptive properties to the scattering cross section would be relevant in a simulation of an entire protein consisting of thousands of atoms; while this is in principle possible, it is computationally very costly, since the interaction with each scatterer has to be calculated separately.

4 Phase reconstruction: Application to experimental data

The interaction of low-energy electrons with biological matter, such as proteins, will in general result in changes to both the amplitude and the phase of the incoming electron wave. The phase maps the elastic part of the scattering interaction, whereas the changes in amplitude are related to the inelastic scattering events [62, 49, 220]. Thus, while amplitude images retrieved from experimental holograms can yield a variety of insights regarding the structure and conformational variability of proteins, as discussed in Chapter 2, the application of the iterative phase retrieval algorithm presented in Chapter 3 to experimental data is of great interest as this could reveal additional information about the imaged molecules.

So far, it has been demonstrated that iterative phase retrieval algorithms like the one discussed in the previous chapter can be applied to experimentally acquired optical holograms [193] and to low-energy electron holograms of charged adsorbates [221]. This chapter focuses on the phase reconstruction of low-energy electron holograms of proteins. It will be shown that the phase retrieval algorithm described in Chapter 3 can be employed to iteratively calculate the amplitude and phase distributions of a wide range of LEEH-imaged proteins with different sizes and characteristics. A selection of phase reconstructions of different proteins, ranging from small proteins (myoglobin and hemoglobin) to large proteins (β-Galactosidase) is shown in Fig. 4.1. All phase reconstructions in Fig. 4.1 feature a uniform background, indicating that the iterative phase retrieval algorithm has succeeded in removing the background modulations due to the contributions of the twin image. Moreover, the phase images in Fig. 4.1 show distinct features of strong contrast that cannot immediately be interpreted. In order to evaluate the phase reconstructions, the phase information needs to be carefully compared to amplitude data as well as additional information that can, for example, be inferred from the holograms themselves, or by comparison with molecular models of the proteins derived from other imaging methods, such as X-ray crystallography or cryo-EM.

Before turning to the interpretation of the retrieved phase distributions, the image processing workflow employed for the phase reconstruction of experimentally acquired holograms is discussed. Following that, possible interpretations of the phase images and their features are proposed, building on the concepts of molecular thickness and density as well as variations in the scattering potential that were explored in Chapter
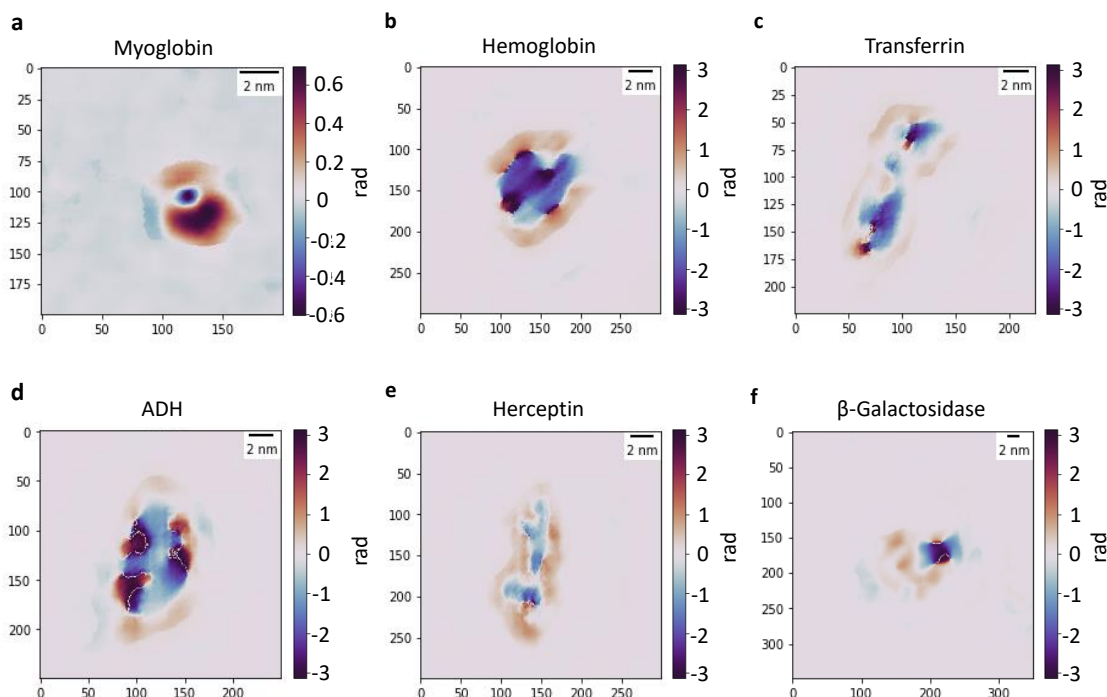
Figure 4.1: **Phase reconstruction of proteins: a** Myoglobin (mass: 17 kDa) **b** Hemoglobin (mass: 64 kDa) **c** Transferrin (mass: 80 kDa) **d** ADH (mass: 150 kDa) **e** Herceptin (mass: 150 kDa) **f** β-Galactosidase (mass: 465 kDa).

3. For this, amplitude and phase data of a variety of proteins will be evaluated. The examined proteins have a large size range, from cytochrome C (cyt C, 12 kDa) to GroEL (812 kDa). Some of the proteins have characteristic structural features, while others have a globular shape. In some cases, metal atoms occur as part of the protein structure, both in the form of individual ions (e. g. ADH, transferrin) and embedded in a porphyrin cofactor known as a heme group (e. g. cyt C, myoglobin, hemoglobin). All these types of characteristics will be taken into account in the interpretation of the phase. Additionally, phase features induced by charging effects will be examined.

4.1 Image processing for phase reconstruction

The one-step amplitude reconstruction routine, which has been described in Chapter 1 and applied to protein data in Chapter 2, can be directly applied to the photographs of the holograms featuring the full detector area; while the sharp edge of the detector can cause edge artefacts, this can be remedied by applying an apodization filter to the hologram before reconstruction (see section 1.1.3). In the case of phase reconstruction, however, the sharp detector edge induces a large phase shift that dominates the reconstructed image, hence an efficient and high-contrast reconstruction of the object phase requires the hologram to be cropped, featuring all observable fringes, but as little of the detector edge (or other high-contrast features, such as impurities) as possible. While it is helpful
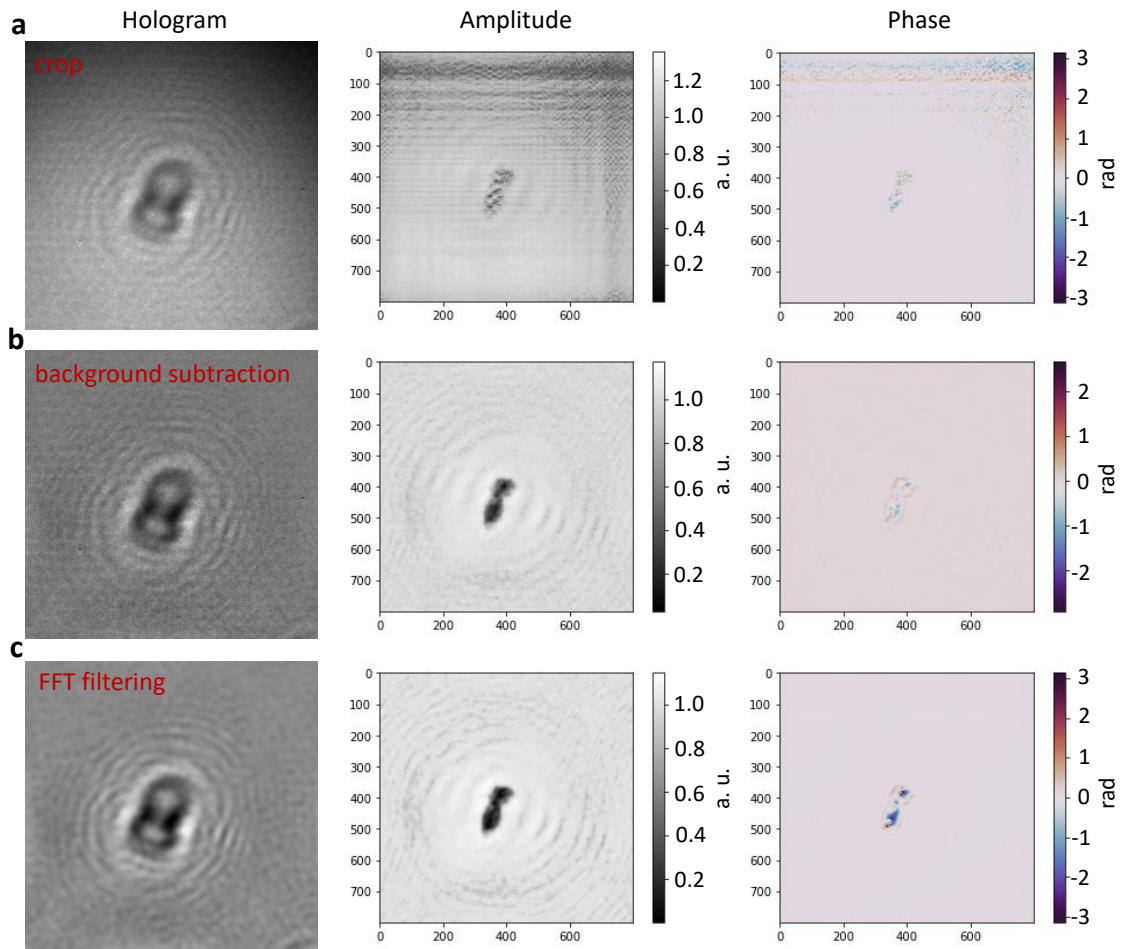
Figure 4.2: **Hologram processing for phase reconstruction: a** Hologram, iterative amplitude reconstruction after 100 iterations and iterative phase reconstruction after 100 iterations of a LEEH hologram of a transferrin molecule. The hologram has been cropped, but no further image processing steps have been performed. The non-uniform background results in severe edge artefacts. **b** Hologram, iterative amplitude reconstruction after 100 iterations and iterative phase reconstruction after 100 iterations of the hologram in **a** after an additional polynomial background subtraction with a polynomial of order 5 has been performed. This processing step leads to a removal of the artefacts in the reconstruction. **c** Hologram, iterative amplitude reconstruction after 100 iterations and iterative phase reconstruction after 100 iterations of the hologram in **b** with high-frequency noise removed in Fourier space by the procedure depicted in Fig. 4.3.

to crop additional molecules and other features, this should not be done at the expense of cutting off fringes since this would affect the resolution as well as the visibility of some structural features of the reconstructed images. To achieve this, and in order not to cut off fringes when applying apodization filters to the holograms, the hologram of the molecule should ideally be located at the centre of the cropped image.

Simply cropping the hologram of the molecule to be reconstructed, however, is not sufficient for achieving an artefact-free reconstruction. For this, an additional, important step is the subtraction of a polynomial background in order to obtain a uniform background intensity over the whole of the hologram. Without background subtraction,

the differences in background intensity can lead to severe artefacts in the reconstruction, which impede a high contrast reconstruction of the phase, as demonstrated in Fig. 4.2a. The cropping of the holograms as well as the polynomial background subtraction can be performed by a variety of programs, the holograms used for the reconstructions shown here have been processed using the program Gwyddion [222]. In general, the order of the polynomials used for the background removal should be chosen such that the background intensities are uniform, while avoiding a change in intensity in the area of the molecule. As a rule of thumb, polynomials of order 3 to 5 seem to fit this requirement for LEEH holograms measured with our set-up. As demonstrated in Fig. 4.2b, the background subtraction step removes the artefacts present in Fig. 4.2a.

The quality of the reconstruction can be increased by removing high-frequency noise in the hologram, which results in more pronounced fringes, as shown in Fig. 4.2c. This leads to a smoother amplitude reconstruction and a higher contrast in amplitude and phase due to the lower noise level and higher hologram contrast. One possibility to remove high-frequency noise is to cut off high frequencies in Fourier space. Noise removal in Fourier space has the advantage that periodic signals, such as the one created by the pixels of the detector, can simultaneously be eliminated.

The noise removal process used in this chapter is depicted in Fig. 4.3. As a first step, the 2D Fourier transform of the background-subtracted hologram (Fig. 4.2b) is calculated (Fig. 4.3a) using the *2D FFT filtering* function in Gwyddion. With the same Gwyddion tool, a Fourier space mask can be created to determine which Fourier components should be retained. For this, at first the part of Fourier space to be retained is selected by the function *Add an ellipse to the Fourier mask* (Fig. 4.3b). Subsequently, periodic signal contributions, which appear in Fourier space in the form of diffraction spots (marked by red arrows in Fig. 4.3), can be removed by using the *Subtract an ellipse from the FFT mask* (Fig. 4.3c). Fig. 4.3d and e demonstrate the effect of the Fourier filtering step: compared to the unfiltered hologram (Fig. 4.3d), the fringes are more pronounced and high-frequency noise is reduced in the filtered hologram (Fig. 4.3e).

While FFT filtering usually improves quality of the reconstruction and can accentuate details that are not discernible in the reconstruction from the unfiltered hologram, the smoothing that results from the FFT filtering step can also blur small spatial features if the cutoff frequency is chosen too low, hence a comparison of the reconstructions obtained from filtered and unfiltered holograms, respectively, can help elucidating further structural detail.

The processed hologram is used as the input for the iterative phase reconstruction routine (see section 3.1.2). In order to avoid artefacts, it is important to save the filtered hologram as a 16-bit greyscale image that does not include any additional features such as scale bars. To deal with remaining edge artefacts, an apodization filter in which the values outside the masked region are set to the mean value of the hologram can be applied to the hologram, which further contributes to a uniform background intensity at the hologram edges. If only small deviations from a uniform background are present at the edges of the image, applying an apodization filter is not necessary, but for holograms with features at the edge, this is an efficient method for suppressing artefacts. It is, however, important to choose the inner radius of the apodization filter mask such that no fringes are obscured since this can significantly affect the reconstructed image due to the loss of information contained in the outer fringes.
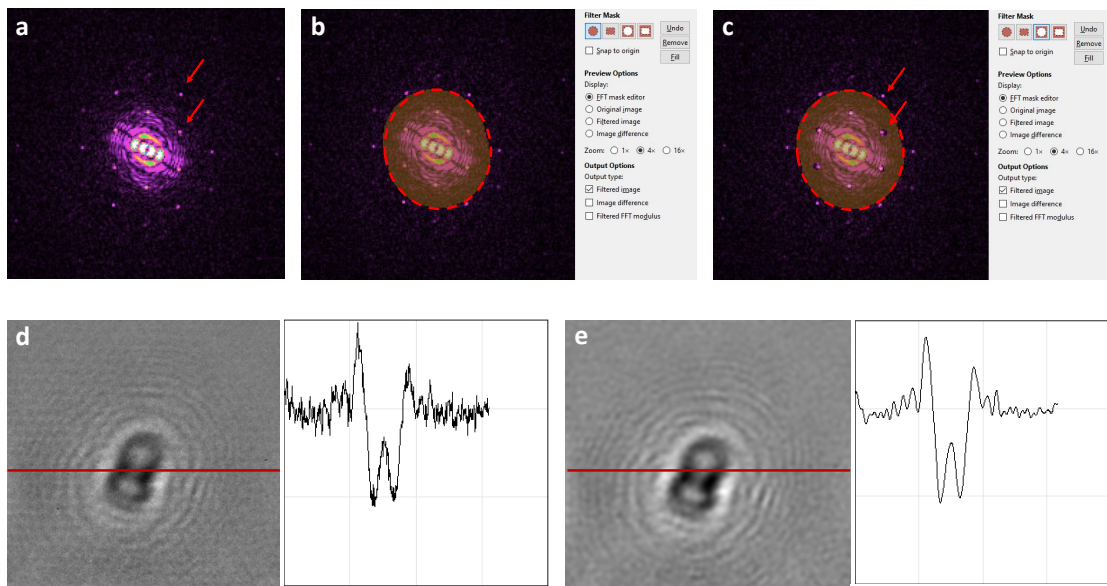
Figure 4.3: **Hologram FFT filtering: a** 2D FFT of the hologram in **d**, calculated with the Gwyddion [222] function 2FFT filtering. **b** Creation of the FFT filter mask, which selects the Fourier components of the image to be kept. **c** Periodic signal contributions can be removed from the Fourier filter mask. **d** Hologram before Fourier filtering and the corresponding cross section along the red line. **e** Hologram after Fourier filtering and the corresponding cross section along the red line. Fourier filtering reduces the noise in the hologram and increases the discriminability of the fringes.

As a last step before the iterative reconstruction process can be initialized, the hologram has to be divided by its mean so that the range of intensity values in the hologram matches the value range that the constraints in the iterative reconstruction routine apply to. Physically, this step corresponds to a hologram normalization by division with the background intensity, i.e. the intensity of the reference wave. Since an empty beam measurement is not feasible experimentally, as the performance and brightness of the electron source can change significantly between measurements, this normalization process is approximated by a division by the mean value of the image via the NumPy function `np.mean`.

In summary, the following image processing steps are taken on the hologram level to generate suitable input for the iterative phase retrieval algorithm from experimentally acquired holograms: Firstly, the hologram of the molecules has to be cropped such that all of the resolvable fringes are retained while as many additional features (such as the detector edge or other molecules) as possible are removed. Ideally, the hologram is centred in the cropped image. As a second step, a polynomial background subtraction is performed, with typical polynomial orders in the range of 3 to 5. Subsequently, a Fourier filter is applied to remove high-frequency noise and to enhance the contrast of the fringes. An apodization filter can be applied to suppress remaining edge artefacts. In a final step, the hologram is normalized by division by its mean. A hologram prepared by this routine can then be reconstructed by the algorithm presented in Chapter 3.

4.2 Test of algorithm performance for experimental holograms

The analysis and characterization of the algorithm in Chapter 3 has shown that the algorithm can retrieve the correct phases independently of the input phase and that artefacts due to the algorithm are limited to specific cases. This characterization carries over to the reconstruction of experimental holograms. In the case of experimental holograms, however, the amplitude and phase distributions of the object are unknown, i. e., no direct comparison is possible to ascertain the correctness of the reconstruction. Moreover, experimental holograms differ from simulated holograms in features like noise
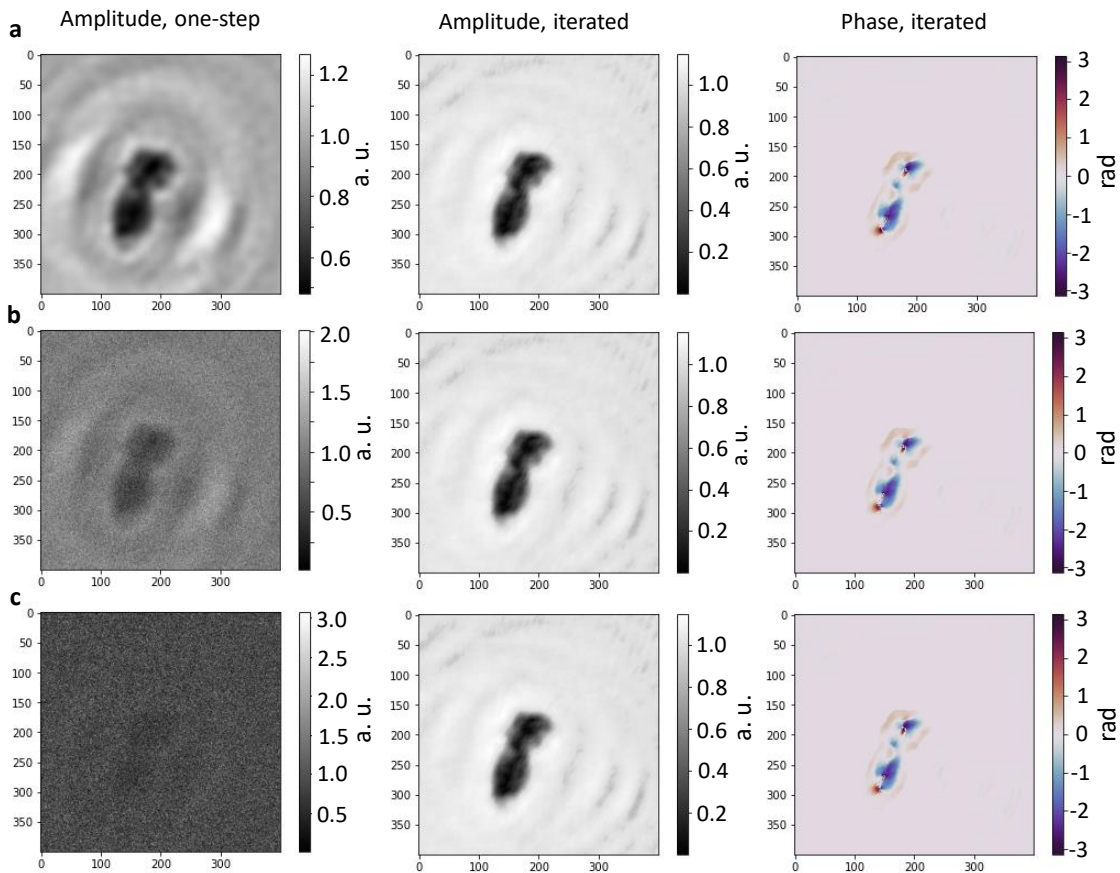

Figure 4.4: **Independence of input phase, experimental holograms: a** One-step amplitude reconstruction, iterative amplitude reconstruction and iterative phase reconstruction of an experimentally acquired hologram of a transferrin molecule with an initial phase distribution given by an array of zeros. The hologram has been processed as described in section 4.1 before reconstruction. **b** One-step amplitude reconstruction, iterative amplitude reconstruction and iterative phase reconstruction of the same hologram as in **a** with an initial phase distribution given by an array of uniformly distributed random numbers created by the NumPy function `np.random.rand`. **c** One-step amplitude reconstruction, iterative amplitude reconstruction and iterative phase reconstruction of the same hologram as in **a** with an initial phase distribution given by an array of normally distributed random numbers created by the NumPy function `np.random.randn`.

level and background uniformity, hence it is worth ensuring that the algorithm retains
these characteristics when applied to experimental data.

To ascertain that experimental holograms, as their simulated counterparts, yield the
same iterative reconstruction independent of the initial input phase, an experimentally
acquired hologram of a transferrin molecule was processed as described in section 4.1 and
reconstructed with three different initial phase distributions: an array of zeros (Fig. 4.4a),
an array of uniformly distributed random numbers (Fig. 4.4b) and an array of normally
distributed random numbers (Fig. 4.4c). Fig. 4.4 demonstrates that while the one-step
reconstructions for these three cases differ strongly, the iterative reconstructions are
identical (shown here for 100 iterations). Since in all three cases, the algorithm converges
to the same reconstructed amplitude and phase images, using an initial input phase array
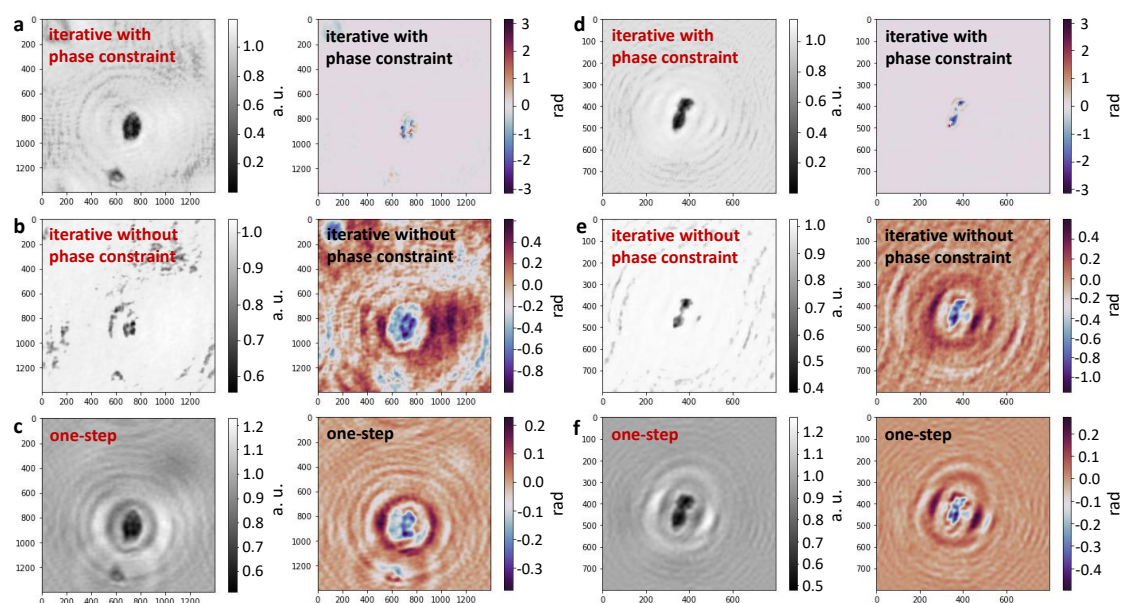of zeros as default for the iterative reconstruction routine is justified.

Figure 4.5: **Relevance of phase constraint, experimental holograms: a** Iterative amplitude
and phase reconstruction after 100 iterations of an experimentally acquired hologram of a
hemoglobin molecule. Both amplitude and phase have been constrained in every iteration step
as described in Chapter 3. **b** Iterative amplitude and phase reconstruction after 100 iterations
of the same hologram as in **a**. Only the amplitude has been constrained during the iterative
process. **c** One-step amplitude and phase reconstruction of the same hologram as in **a**. **d** Iterative
amplitude and phase reconstruction after 100 iterations of an experimentally acquired hologram
of a transferrin molecule. Both amplitude and phase have been constrained in every iteration step
as described in Chapter 3. **e** Iterative amplitude and phase reconstruction after 100 iterations of
the same hologram as in **d**. Only the amplitude has been constrained during the iterative process.
f One-step amplitude and phase reconstruction of the same hologram as in **d**.

The analysis of the constraints enforced during the iterative reconstruction process (section
3.1.1) demonstrated that constraining the phase along with the amplitude significantly
improves the quality of the iterative reconstruction for simulated holograms. Because of
the importance of this step, it is relevant to test the behaviour of experimental holograms
under these constraints. Fig. 4.5 explores this with two examples of experimentally

acquired protein holograms, one of a hemoglobin molecule (Fig. 4.5a-c) and one of a transferrin molecule (Fig. 4.5d-f). Fig. 4.5 shows that the reconstructions performed with the phase constraint applied (Fig. 4.5a, d) differ strongly from the reconstructions retrieved without enforcing the phase constraint (Fig. 4.5b, e). Specifically, without the phase constraint, the phase reconstructions remain close to the one-step phase reconstructions (Fig. 4.5c, f).

Analogously to the simulated case, the algorithm without the phase constraint is much less robust towards random initial phase inputs than the algorithm with the phase constraint when applied to experimental data. As in the simulated case, the reconstructions differ significantly depending on the initial phase input, and the noise induced via a random initial phase input cannot be removed by the iterative steps, as opposed to the case of a reconstruction with the phase constraint (Fig. 4.4).

In general, it thus appears to be beneficial to employ the constraints on both absorption and phase in the case of experimental holograms as well, as expected from the results on simulated data presented in the previous chapter. As mentioned in Chapter 3, if negative absorption values occur within the molecule, applying the phase constraint could in principle lead to pixel-wise artefacts. Indeed, this type of artefact occasionally occurs in the reconstruction of experimental holograms, but is usually not very severe. A more detailed discussion of this, alongside other artefacts observed in the iterative reconstruction of experimentally acquired holograms, is carried out in section 4.6.

4.3   Interpretation of experimental phase data

The one-step reconstruction introduced in Chapter 1 and the iterative reconstruction routine presented in Chapter 3 yield reconstructions of the one-step amplitude, the iterated amplitude and the iterated phase; all three contain relevant information about the imaged molecules.

In general, the contours of the molecules are more pronounced in the amplitude reconstructions than in the phase reconstructions (see for example Fig. 4.6 and Fig. 4.12). Specifically, phase reconstructions often feature low-contrast envelopes that trace the outline of the molecule and usually have a phase contribution of opposite sign as compared to the phase shift induced by the inner part of the molecule (see e. g. Fig.4.1b-e, Fig. 4.6, Fig. 4.12c). Because of the less clear contours in the phase reconstructions, it is often difficult to determine the source-to-sample distance z_0 for which a molecule appears in focus from the phase images. Hence, the amplitude reconstructions should be used to identify the appropriate source-to-sample distances.

While phase reconstructions may not feature clear contours, they can help identifying structural features within the molecules that are not discernible in the amplitude reconstruction. Fig. 4.6 shows two examples where the substructure of the imaged object is much more clearly recognizable in the phase reconstruction than in the amplitude reconstruction. Fig. 4.6a shows a hemoglobin molecule, a protein consisting of four subunits. The subunits are not distinguishable in either of the amplitude reconstructions, but they are discernible in the phase reconstruction. Similarly, in Fig. 4.6b, which depicts the reconstruction of a dimer consisting of two cytochrome C molecules, the two molecules are clearly distinguishable in the phase, but not in the amplitude.
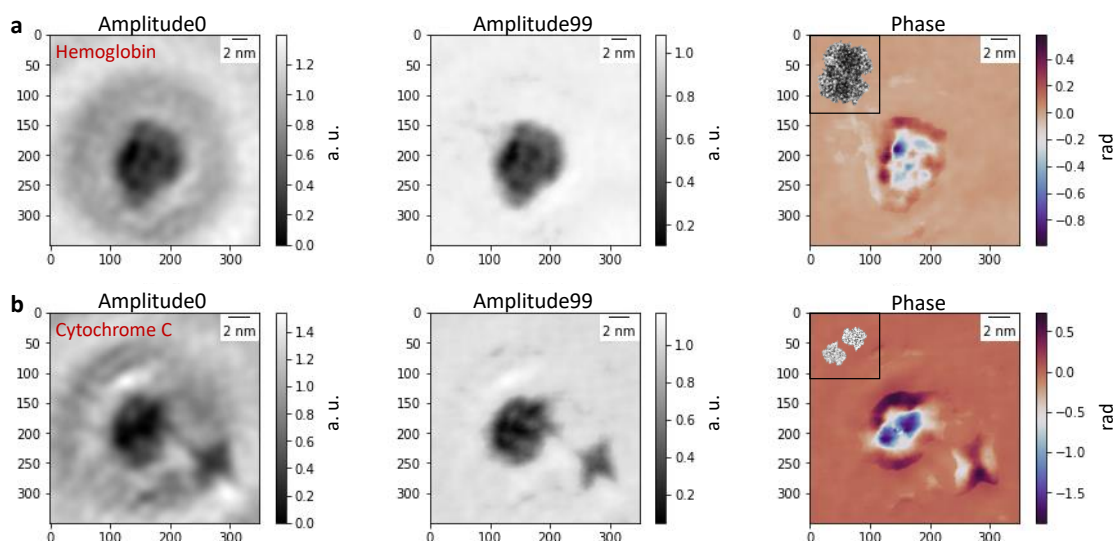
Figure 4.6: **Phase features: subunit identification. a** One-step amplitude reconstruction, iterated amplitude reconstruction after 100 iterations and phase reconstruction after 100 iterations of a single hemoglobin molecule. The phase reconstruction reveals the substructure of the molecule. The inset in the phase reconstruction shows a model of a hemoglobin molecule (PDB: 1FSX [223]). **b** One-step amplitude reconstruction, iterated amplitude reconstruction after 100 iterations and phase reconstruction after 100 iterations of two cytochrome C molecules. The molecules are more clearly distinguishable in the phase reconstruction than in the amplitude reconstruction. The feature in the bottom right corner is an additional charge, possibly due to a defect in the graphene. The inset in the phase reconstruction shows a model of two cytochrome C molecules (PDB: 1HRC [224]).

As shown in Fig. 4.1, a variety of different phase features, from large regions of constant phase to small localised areas of rapidly changing phase, can occur in the phase reconstruction of a protein. In order to be able to interpret these features, one important aspect is to ascertain whether these features are robust with respect to a change in reconstruction distance, corresponding to reconstructions in different focal planes. The results from the simulations and reconstructions of three-dimensional objects performed in section 3.3.1 suggest that a phase feature should be observable over a z-range spanning tens of nanometers around the focal plane. This could imply that phase features that remain similar over the range of in-focus distances determined from the amplitude reconstruction are related to structural features of the molecule.

Fig. 4.7 depicts the phase reconstruction of an ADH molecule at four different source-to-sample distances. The amplitude reconstruction yields a focus distance of around $z_0 = 390 - 400$ nm. The four prominent phase features, appearing in the reconstructions as localised red areas, are present over the range of focal distances shown in the figure, indicating that those are stable phase features. The symmetry present in both amplitude and phase reconstructions suggests that these phase features are related to the molecule's four-subunit structure. Additionally, the behaviour of phase features over different focal distances is of relevance since a rapid change in the phase with z_0 could be indicative of artefacts. Especially around the edge of a molecule, the phase can change significantly for different reconstruction distances.
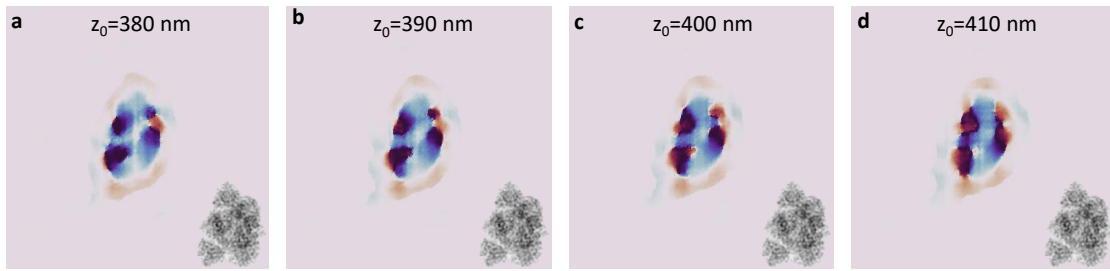
Figure 4.7: **Phase features: z-focus. a-d**: Phase reconstruction of a single ADH molecule for different focal distances from $z_0 = 380\,\mathrm{nm}$ to $z_0 = 410\,\mathrm{nm}$. Four distinctive phase features are present over the whole z_0-range depicted here and change very little between the two in-focus distances ($z_0 = 390 - 400\,\mathrm{nm}$). The insets show a model of an ADH molecule (PDB: 7KCQ [225]).

Before turning towards exploring different interpretations of the phase features, such as the molecular density and thickness (section 4.4) as well as changes in the scattering strength due to the presence of metal atoms (section 4.5) and charges (section 4.5.4), a few general aspects of phase features are discussed in the following.

Unlike the amplitude, the phase can assume both positive and negative values, hence features that involve a change in the sign of the phase are especially prominent. On the one hand, this can help identifying structural features that are hard to discern in the amplitude reconstructions, on the other hand, quantitatively small changes of phase that result in a change of sign could be interpreted as more significant than they actually are. The phase shift induced by a molecule and reconstructed by an iterative reconstruction routine cannot only take both positive and negative values, it is also cyclic, i.e. the reconstructed phase is mapped on the interval $[-\pi,\,\pi]\,\mathrm{rad}$, thus a phase value slightly above $+\pi\,\mathrm{rad}$ will be represented as a negative value close to $-\pi\,\mathrm{rad}$. To account for this, the phase reconstructions are depicted using a cyclic colormap, hence values of high positive and negative phase appear in the same colour (dark purple in the colormap used here). While the use of a cyclic colormap avoids some of the ambiguities related to the $\pi/-\pi$-identity, a full quantitative phase analysis is not possible since the projection on $[-\pi,\,\pi]\,\mathrm{rad}$ does not allow a distinction of phase shift values outside this interval and the corresponding value within the interval. Preliminary tests with phase unwrapping algorithms were unsuccessful, most likely because these algorithms are developed for data with strong background phase modulations due to continuous phase wrapping, a type of feature that does not appear in the reconstructed phase data discussed here.

In our data, areas of localised phase change often seem to be correlated with a change in sign and thus the appearance of high phase values, hence it is worth examining the behaviour of the phase values in these regions in more detail. Fig. 4.8 shows the phase reconstruction of an ADH molecule featuring four regions of localised phase change. By taking the cross section of the image at four different pixel values (indicated by the black lines, the respective pixel value is given as a number at the right side of each image), chosen such that each of the four distinct phase features is crossed, the progression of the phase values across the localised areas of phase change can be studied.
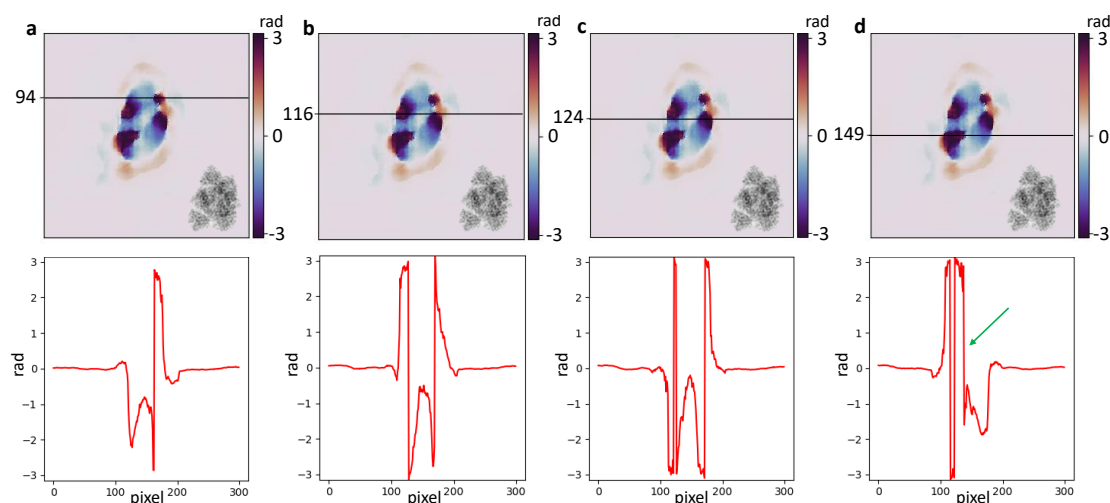
Figure 4.8: **Phase features: areas of localised phase change. a-d**: Phase reconstruction of an ADH molecule and corresponding line profiles along the lines indicated in black, crossing each of the four phase features, respectively. While the phase appears to exhibit discontinuous behaviour, the majority of the features in the line profiles can be explained by continuous changes in phase. The insets show a projection of an ADH molecule (PDB: 7KCQ [225]) matching the orientation of the molecule shown here.

Two types of behaviour that lead to a change in sign can be identified: in the first type of progression, the phase starts out as positive (negative), increases (decreases) towards π $(-\pi)$ rad, then crosses over the $\pi/-\pi$ threshold and decreases (increases) towards zero, as in the left peak in Fig. 4.8b (positive to negative) and in the right peak in Fig. 4.8c (negative to positive). Such behaviour could be interpreted as a continuous increase or decrease of phase shift values that at some point crosses the $\pi/-\pi$ threshold. The peak in Fig. 4.8a and the right peak in Fig. 4.8b exhibit a similar behaviour, although the progression towards the $\pi/-\pi$ threshold happens more rapidly. The second type of behaviour is represented by the peaks on the left of Fig. 4.8c and d. Once the $\pi/-\pi$ threshold is reached, the phase briefly crosses it and thus changes sign, but then crosses back immediately. This can be interpreted as an increasing (decreasing) phase shift that increases (decreases) slightly above π rad (below $-\pi$ rad) and then decreases (increases) again. In both types of phase behaviour, multiple rapid crossings of the $\pi/-\pi$ threshold are possible.

While, especially in the line profiles, it may look as if the phase is jumping when the $\pi/-\pi$ threshold crossings happen, it seems that the behaviour described above can be explained by continuous changes of phase. This indicates that the majority of the observed local phase changes can be attributed to a continuous phase behaviour, yet, the existence of discontinuous phase jumps cannot be fully ruled out. There is evidence that rapid changes in phase can happen without the involvement of a $\pi/-\pi$ threshold crossing, see for example the phase change in Fig. 4.8d from approximately $+3$ rad to -1 rad at $x \approx 130$ pixels (indicated by the green arrow in Fig. 4.8d).

127

4.4 Molecular density

In high-energy electron imaging, for example in TEM, the phase shift induced by the imaged object is directly related to the mean inner potential of the object, which is defined as $V_{\mathrm{mean}} = \frac{1}{\Omega} \int_\Omega V(\mathbf{r}) \mathrm{d}^3 \mathbf{r}$, where $V(\mathbf{r})$ is the electric potential of the object [56, 49]. If the scattering potential of all atoms in the object is similar, this implies a direct connection between the induced phase shift and the object's molecular density, or in case of a uniform density, the molecular thickness. In imaging processes that use low-energy electrons, however, additional contributions to the potential can arise, which make the interpretation of the phase data less straightforward. Specifically, contributions to the potential due to exchange interaction [56, 213] and polarization effects [213] become relevant at low energies. Additionally, inelastic scattering and absorption effects are not negligible in this energy range [56, 213]. Thus, molecular density and the thickness of the molecule should, at least to some extent, be mapped by both the amplitude and phase distribution of the object. To interpret the data retrieved from experimental holograms, it is hence of interest to study both the amplitude and phase reconstructions, as well as their interrelation.

4.4.1 Absorption

Since strongly absorbing objects can induce artefacts in the phase reconstruction, as discussed for simulated objects in section 3.2.2, it is important to check whether the molecules studied here, specifically the larger proteins, would fall into the category of strong absorbers. The two largest proteins imaged by LEEH on our set-up are β-Galactosidase (465 kDa, ca. 18 nm × 14 nm × 8.7 nm [226]) and GroEL ($\sim$ 800 kDa, diameter: 13.5 nm, height: 14.5 nm [227]). One can calculate the transmission through such an object with the help of the Lambert-Beer law, which relates the ratio of transmitted to incident intensity $\frac{I}{I_0}$ to the ratio of the object's thickness d and the inelastic mean free path λ_i [220],

$$\ln\left(\frac{I}{I_0}\right) = -\frac{d}{\lambda_i}. \tag{4.1}$$

The inelastic mean free path of an electron depends on its energy E and can be calculated as follows [228]:

$$\lambda_i = \frac{A}{E^2} + B\sqrt{E}, \tag{4.2}$$

where A and B are material-dependent constants. According to ref. [228], appropriate values for organic compounds are $A = 31$ and $B = 0.087$, the corresponding plot of the inelastic mean free path as a function of energy is depicted in Fig. 4.9a. For an electron energy of 100 eV, a typical energy value used for acquiring experimental holograms by LEEH, this yields an inelastic mean free path of slightly below 1 nm. The values for the inelastic mean free path calculated here match the values reported elsewhere in literature [57].

With this, the transmission through a protein of known thickness can be estimated according to eq. 4.1, as shown in Fig. 4.9b for electron energies in the range of $50 - 250$ eV.
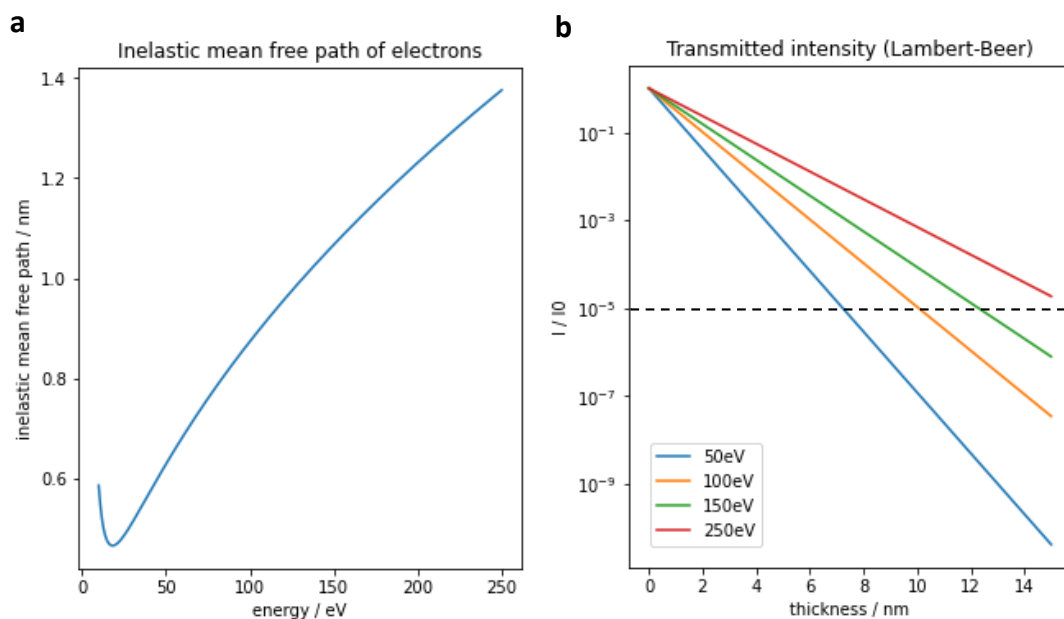
Figure 4.9: **Inelastic mean free path and Lambert-Beer law: a** Energy-dependent inelastic mean free path in nm for electrons in organic matter, calculated according to eq. (4.2) [228]. **b** Lambert-Beer law: transmission of electrons of energies in the range of $50 - 250\,\text{eV}$ through an organic component as a function of the object's thickness (eq. (4.1)). The dashed black line marks the transmission value below which artefacts occurred in the case of simulated objects.

Section 3.2.2 showed that for transmission values below 10^{-5}, artefacts can occur in the phase reconstruction of simulated objects. According to Fig. 4.9b, large proteins of the dimensions of GroEL and β-Galactosidase should hence be in the strongly absorbing regime. Despite that, the experimental evidence suggests that a reconstruction free of the type of artefacts observed for the simulated strongly absorbing objects is possible for these molecules. On the one hand, LEEH imaging of large proteins produces high contrast holograms that present the same symmetry as holograms simulated from the molecular structure (Fig. 4.10), indicating that the scattered amplitude is large enough to generate sufficient contrast in the holograms. This does not necessarily imply that there is a strong transmitted part of the wave, since diffraction at the edge of the molecule alone could result in an interference pattern of similar symmetry, but hints at the fact that even if the object is strongly absorbing, the hologram and thus the reconstruction contains relevant structural information. Even in the case of zero transmission, the reconstruction of such a hologram should yield the overall shape of the molecule, although inner structure would likely be lost. On the other hand, both the amplitude and phase reconstructions feature inner structure and intensity modulation, as shown in the phase reconstructions of a β-Galactosidase molecule and a GroEL molecule presented in Fig. 4.10e and j, respectively, suggesting that there is indeed a significant transmitted part. Additionally, the phase reconstructions obtained from experimental holograms do not suffer from noise-like artefacts as the phase reconstructions of the strongly absorbing simulated objects (Fig. 3.12e, f). In general, this suggests that the transmitted intensity is higher than expected from
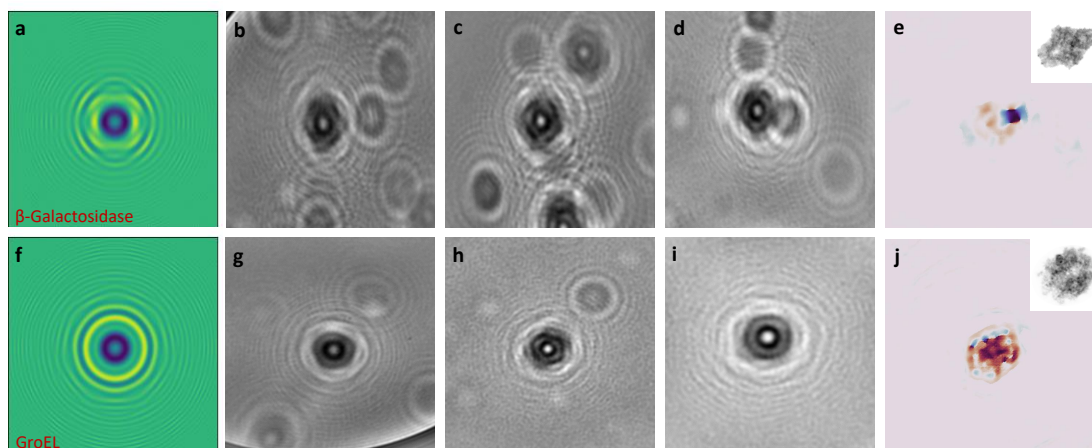
Figure 4.10: **Simulated and experimental holograms of large proteins: a** Hologram of a β-Galactosidase molecule simulated from the PDB structure 6CVM [229]. All atoms have been projected into one plane to create a transmission function. **b-d** Examples of experimentally obtained holograms of β-Galactosidase molecules. The symmetry of the experimental holograms matches the one observable in the simulated hologram. **e** Phase reconstruction of a β-Galactosidase molecule with a projection of the molecular model in the inset. Some inner substructure of the molecule is resolved. The blue feature on the right is due to a charge. Since the presence of the charge partially obscures the symmetry in the hologram, the hologram corresponding to the reconstruction in **e** is not presented here. **f** Hologram of a GroEL molecule simulated from the PDB structure 5W0S [210]. **g-i** Examples of experimentally obtained holograms of GroEL molecules. The symmetry of the experimental holograms is similar to the one observable in the simulated hologram. **j** Phase reconstruction of a GroEL molecule from the hologram in **h** with a projection of the molecular model in the inset. Some inner substructure of the molecule is resolved.

the Lambert-Beer plot in Fig. 4.9b. This could be due to an inexact assumption of the values of A and B used for the calculation, since the generic values for organic compounds found by low-energy electron diffraction might have to be adjusted for large proteins, or due to the fact that objects such as large proteins do not feature a uniform density distribution over the whole molecular thickness. Parts of lower density would decrease the effective thickness of the protein that is used as a parameter in the Lambert-Beer calculation, thus leading to a higher transmission than calculated in Fig. 4.9b for these types of objects.

4.4.2 Relation between amplitude and phase reconstruction

While the experimental evidence shows that large proteins can be imaged by LEEH in an artefact-free manner and that the reconstructions can reveal substructure, it is still possible that areas of high absorption, and thus of low amplitude, can result in artefacts in the phase. Hence, it is important to illuminate the interrelation between amplitude and phase data, specifically the correlation between low amplitude and distinctive phase features such as $\pi/-\pi$ threshold crossings, which could indicate the presence of artefacts. The relation between amplitude and phase is examined with the help of several visualiza-

tion tools presented in Fig. 4.11, applied to the example of three different molecules: a transferrin molecule in Fig. 4.11a-h, an antibody molecule (Herceptin) in Fig. 4.11i-p, and a hemoglobin molecule in Fig. 4.11q-x.

Since it is of interest to compare both the iterated and the non-iterated amplitude data to the phase data, panels a, i and q depict the non-iterated amplitude reconstruction (labelled Amp0), the iterated amplitude reconstruction after 100 iterations (labelled Amp99) and the corresponding phase reconstruction after 100 iterations for the respective examples. A direct comparison of Amp0 and Amp99 shows that the iterative phase retrieval algorithm effectively reduces the contributions from the twin image since the background modulations are less pronounced in the iterated amplitude reconstruction. Specifically, the constraint on the absorption leads to the removal of very high background intensity values, which yields an overall background intensity of the iterated amplitude image close to 1.

The effect of the twin image removal on the value distribution in object and background can also be observed in amplitude-vs.-phase plots. In panels b, c, j, k, r and s, the phase values are plotted against the amplitude values for the non-iterated amplitude (Amp0, panels b, j, r) and the iterated amplitude (Amp99, panels c, k, s) for each pixel. The colour of each point in the scatter plot represents the magnitude of the phase gradient at this pixel. This type of plot directly shows the distribution of the phase and amplitude values. The comparison of the plots for iterated and non-iterated amplitude suggests that the value distribution in the case of the iterated amplitude is more levelled, specifically, the background amplitude value distribution is much narrower in the iterated reconstructions. Regions of low non-iterated amplitude correlate strongly with regions of high phase, whereas low iterated amplitude values appear to coincide with a larger range of phase values. In both cases, the phase gradient is highest for regions of high phase, matching the observations gleaned from the phase gradient plots (Fig. 4.11d, l, t).

Panels d, l and t show the magnitude of the gradient of the phase reconstructions in a, i and q, respectively, calculated as the square root of the sum of the squared gradients along the x- and y-directions. This is of relevance, since some of the distinct phase features are associated with rapid changes in phase and thus with a large phase gradient. The regions of high phase gradient in Fig. 4.11d, l, t correlate with regions of high phase close to π and $-\pi$ rad in Fig. 4.11a, i, q. This is not surprising since for these phase values, even small changes in phase can result in a crossing of the $\pi/-\pi$ threshold, which appears as a sudden jump from π to $-\pi$ rad (see Fig. 4.8) and thus corresponds to high phase gradients.

While the amplitude-vs.-phase plots reveal the general distribution and correlation of amplitude and phase values, the spatial component, i.e. the location of each pixel within the molecule, is lost, hence it is of interest to additionally create correlation plots that retain the spatial component while focusing on specific phase and amplitude value ranges. Panels e-h, m-p, and u-x, respectively, depict two different types of correlation plots that are based on spatial information. In the first type of plot, located in the respective upper rows, areas of low amplitude values smaller than 0.05 (red) are plotted on top of areas of high phase (yellow) with absolute phase values above 3, i.e. close to π rad, for the non-iterated amplitude (Amp0, panels e, m, u) and the iterated amplitude (Amp99, panels f, n, v), respectively. Pixels for which both conditions hold, i.e. amplitude < 0.05 and $|\text{phase}| > 3$ rad, are depicted in black. The second type of correlation plot shown here
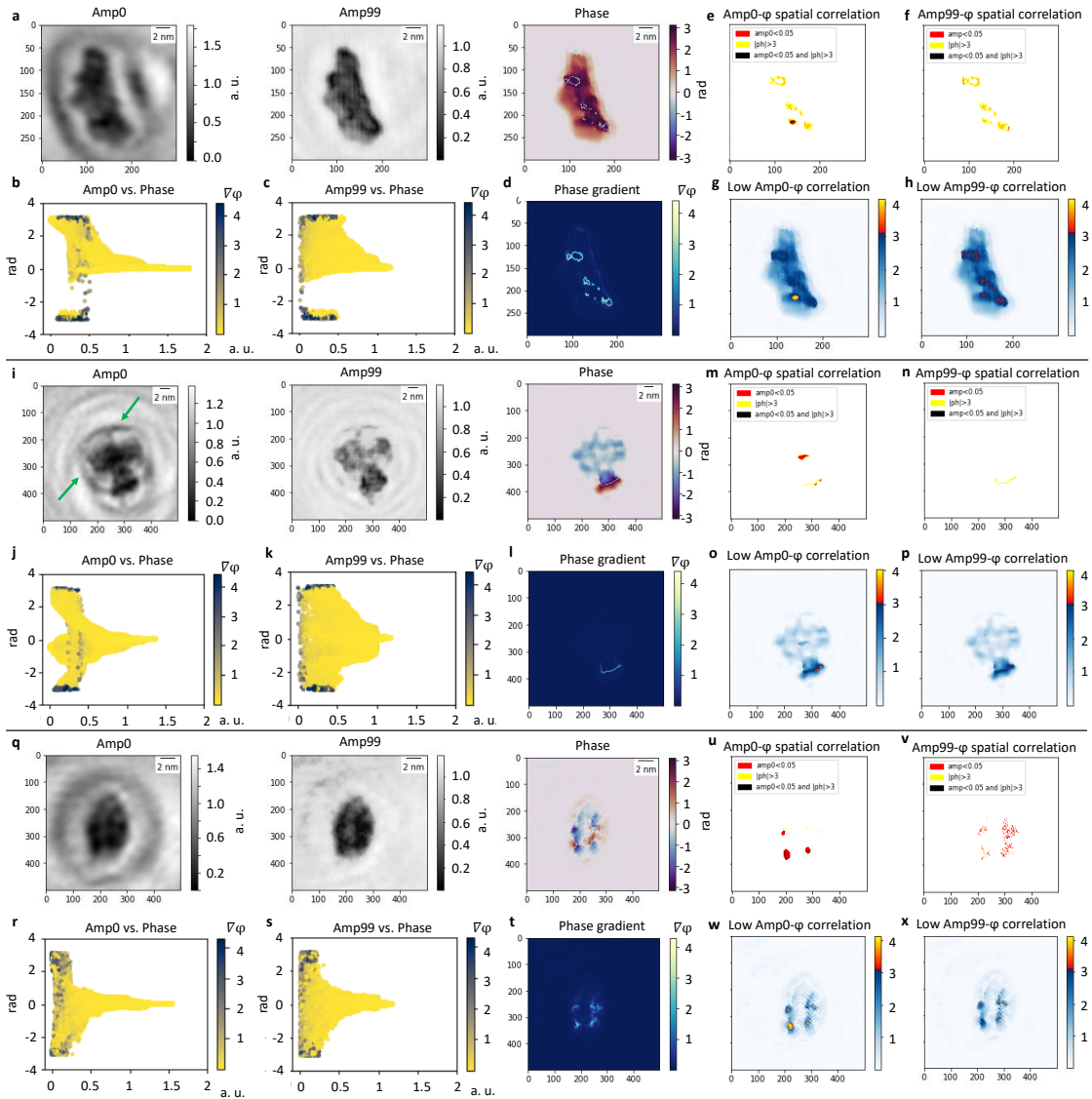
Figure 4.11: **Amplitude vs. phase: a** Left to right: one-step amplitude reconstruction (Amp0), iterative amplitude reconstruction after 100 iterations (Amp99) and phase reconstruction after 100 iterations of a transferrin molecule. **b** Scatter plot of the phase vs. Amp0. **c** Scatter plot of phase vs. Amp99. **d** Magnitude of the gradient of the phase image in **a**. **e** Correlation plot of areas of low non-iterated amplitude (Amp0 < 0.05) and high phase (|phase| > 3). **f** Correlation plot of areas of low iterated amplitude (Amp99 < 0.05) and high phase (|phase| > 3). **g** Correlation plot of the full range of phase values and areas of low non-iterated amplitude (Amp0 < 0.05). **h** Correlation plot of the full range of phase values and areas of low iterated amplitude (Amp99 < 0.05). **i** Left to right: Amp0, Amp99 and phase of a Herceptin molecule. **j** Scatter plot of the phase vs. Amp0. **k** Scatter plot of the phase vs. Amp99. **l** Magnitude of the gradient of the phase image in **i**. **m** Correlation plot of areas of low Amp0 and high phase. **n** Correlation plot of areas of low Amp99 and high phase. **o** Correlation plot of the full range of phase values and areas of low Amp0. **p** Correlation plot of the full range of phase values and areas of low Amp99. **q** Left to right: Amp0, Amp99 and phase of a hemoglobin molecule. **r** Scatter plot of the phase vs. Amp0. **s** Scatter plot of the phase vs. Amp99. **t** Magnitude of the gradient of the phase image in **q**. **u** Correlation plot of areas of low Amp0 and high phase. **v** Correlation plot of areas of low Amp99 and high phase. **w** Correlation plot of the full range of phase values and areas of low Amp0. **x** Correlation plot of the full range of phase values and areas of low Amp99.

features the superposition of the full range of the absolute phase values, which can take values between 0 and π rad and the areas of low amplitude for which amplitude < 0.05 holds. This is achieved by adding the value 1 to the phase value of every pixel that satisfies the low amplitude condition. The colormap used for these plots has been adjusted such that values above 3, i.e. pixels for which the phase is high and the amplitude is low, are easily visible by a change in colour. The areas coloured in orange to yellow hence correspond to the areas where absolute phase values above 2 rad coincide with low amplitude values.

These spatial correlation plots are especially interesting regarding the question of whether areas of low amplitude can result in artefacts in the phase, as indicated by some of the simulated examples discussed in section 3.2.2. To ascertain that localised phase signals can be attributed to properties of the molecule rather than to low-amplitude-related artefacts, we need to check that there is no pixel-wise correlation between areas of low amplitude and localised phase features such as $\pi/-\pi$ threshold crossings. The correlation plots in Fig. 4.11 show that for all three molecules, areas of low amplitude and high phase are correlated, i.e. they occur in the same parts of the molecule, but they are not pixel-wise identical, which suggests that the localised phase signals, as for example observable in Fig. 4.11a, are in general not artefacts due to pixels with low amplitude values.

Overall, the non-iterated amplitude (Amp0) appears to feature more and larger areas of low amplitude values, hence the correlation between low amplitude and high phase is stronger for the non-iterated amplitude than for the iterated amplitude. However, not all areas of low amplitude correspond to regions of high phase and vice versa. Fig. 4.11f features extensive areas of high phase, but almost none of low amplitude, the same is true for Fig. 4.11e, the corresponding correlation plot for the non-iterated amplitude, where one area of low amplitude is present and correlates with high phase values, but the vast majority of high phase regions are still independent of low amplitude values. In Fig. 4.11m and Fig. 4.11u and v, low amplitude areas exist that are not correlated with high phase regions. In Fig. 4.11m, two areas of low amplitude are present, but only one is correlated with a high phase region, the region at the top of the molecule, while appearing dark in Amp0 in Fig. 4.11i, does not show up as a feature in the corresponding Amp99 and phase reconstructions. The region of high phase in Fig. 4.11i, corresponding to the lower dark area in the Amp0 reconstruction, is likely not the result of a structural feature of the molecule, but is an artefact stemming from the fringes created by the upper part of the hologram, which modulate the intensity in the part of the hologram corresponding to the lower subunit. This contribution is visible both on the hologram level and in the non-iterative reconstruction. In the latter, it appears as a dark ring around the upper part of the molecule (marked by green arrows) that intersects the lower subunit and thus leads to a reduced amplitude value in the intersection area. While this feature was removed from the amplitude during the iterative reconstruction, it was retained in the phase reconstruction. This shows that, regarding the interpretation of the structural features revealed in the amplitude and phase reconstructions, all three types of reconstruction data, the non-iterated amplitude, the iterated amplitude and the phase, contain valuable information and that all three, together, should be used to gain an understanding of what can be concluded from the data.

The comparison of all three types of reconstruction also yields insights into the ways in

which the iterative reconstruction can improve the retrieved data as well as into which kind of artefacts can occur. The latter is discussed in more detail in section 4.6.

In summary, regions of low amplitude and high phase are often correlated, but can also occur independently of one another. This implies that localised areas of high phase, often associated with a change in sign of the phase, are not artefacts due to low amplitude, but can likely be attributed to continuous phase changes crossing the $\pi/-\pi$ threshold.

4.4.3 Density projections

The Lambert-Beer calculation shown in Fig. 4.9 and the results of the multislice simulations performed in section 3.3.1 suggest that molecular density and molecular thickness should affect both the amplitude and the phase reconstructions. A comparison of the reconstructions obtained from experimental holograms of proteins with molecular models of the imaged molecules can provide insights into both structural details and into the connection between molecular density and reconstructed phase. In order to explore this, two different types of projections of the molecular models are used in the following discussion. Both types of projections start from a PDB model of the imaged molecule. For the first type of projection, referred to as *molecular projection* in the following and depicted in a greyscale colour scale, all atoms of the PDB model are projected into one plane with each atom represented by a small disk with adjustable radius and contrast. Changing the radius represents adjusting the resolution, while the atom contrast, i. e. the greyscale colour of a single atom, determines the scattering amplitude. The changes in colour over the projection hence are a measure of density, since regions with larger numbers of atoms appear darker. For large molecules with a high number of atoms, it is thus important to set a low atom contrast in order to observe the molecule's inner substructure given by the density variations. This is specifically relevant for determining which molecular projection fits the orientation of the experimentally imaged molecule. To facilitate the comparison to different molecular projections, a script has been set up that can be directly run in PyMOL [231]. As a molecular visualization tool, PyMOL allows the rotation of the molecular structure, and promising orientations that could fit the reconstructions of the experimental data can then directly be turned into molecular projections.

Fig. 4.12 shows examples of the amplitude and phase reconstructions of several experimentally acquired protein holograms and a matching molecular projection obtained from the respective PDB structure. Firstly, this comparison shows that molecular orientations can be found that fit the experimental data very well, both in size and overall shape. In many cases, additional substructure is visible, both in the reconstructions and in the molecular projections, indicating that the molecules remain intact during ES-IBD sample preparation and LEEH imaging.

Fig. 4.12a and b depict two β-Galactosidase molecules in different orientations. In both cases, the projections match the reconstructions in shape and size. In Fig. 4.12b, the substructure of the molecule is well-resolved and fits the substructure visible in the molecular projection. These observations are in accordance with the results of an ES-IBD/cryo-EM study of β-Galactosidase molecules [232]. The dark spot in the non-iterated amplitude of Fig. 4.12b, corresponding to the blue spot in the phase reconstruction, can be attributed to a charge on the molecule.
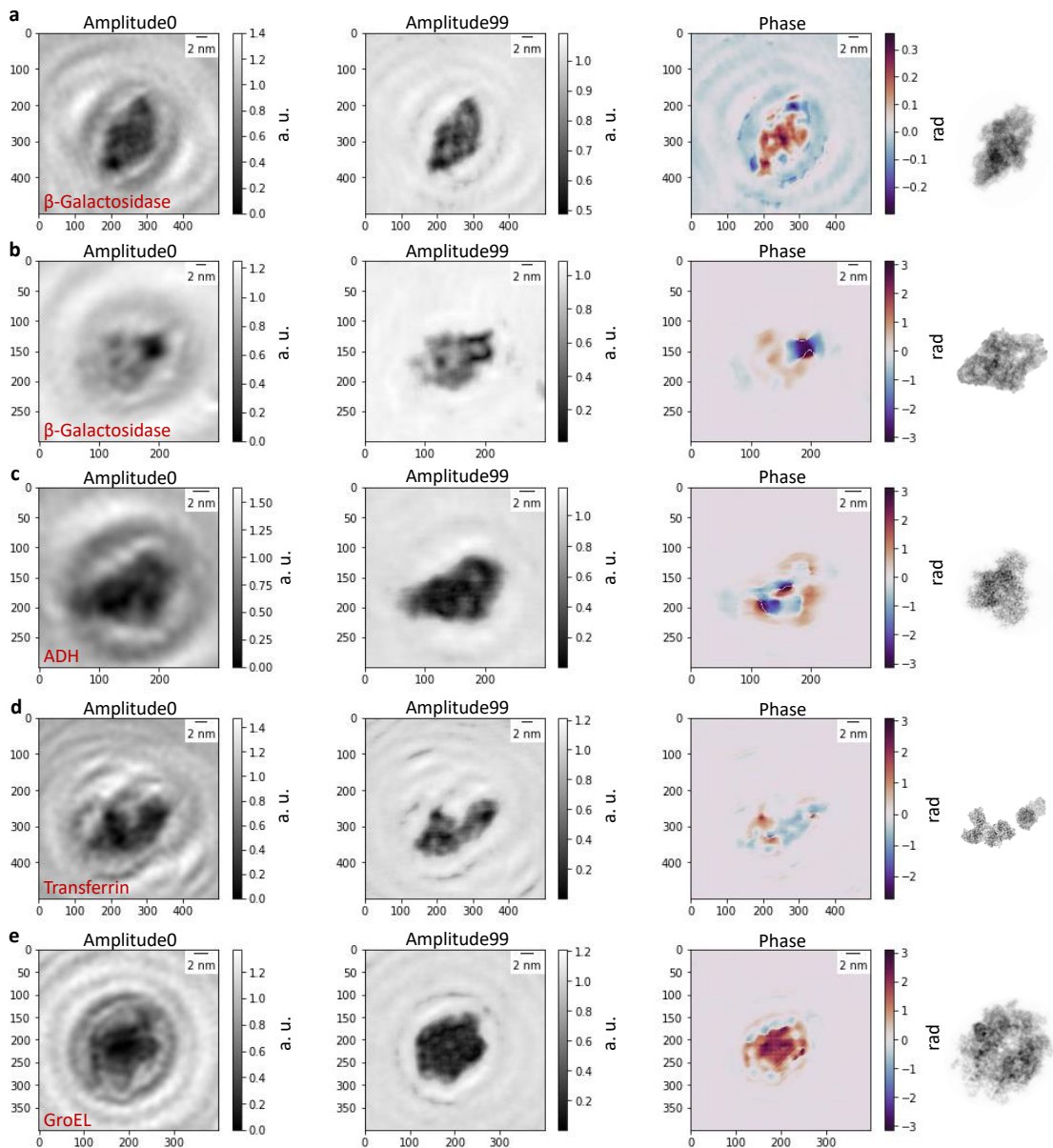
Figure 4.12: **Molecular projections: a** Left to right: one-step amplitude reconstruction (Amplitude0), iterated amplitude reconstruction after 100 iterations (Amplitude99), phase reconstruction after 100 iterations and corresponding molecular projection (PDB: 6CVM [229]) of a hologram of a β-Galactosidase molecule. **b** Left to right: Amplitude0, Amplitude99, phase reconstruction and corresponding molecular projection (PDB: 6CVM [229]) of a hologram of a different β-Galactosidase molecule. As seen in both the reconstructions and the projection, this molecule was captured in a different orientation than the one in **a**. **c** Left to right: Amplitude0, Amplitude99, phase reconstruction and corresponding molecular projection (PDB: 7KCQ [225]) of a hologram of an ADH molecule. **d** Left to right: Amplitude0, Amplitude99, phase reconstruction and corresponding molecular projection (PDB: 4X1B [230]) of a hologram of two transferrin molecules. By comparing the molecular projections to the reconstruction, it becomes clear that this is a dimer of two molecules in different orientations. **e** Left to right: Amplitude0, Amplitude99, phase reconstruction and corresponding molecular projection (PDB: 5W0S [210]) of a hologram of a GroEL molecule.

135

The molecular projection of the ADH molecule in Fig. 4.12c does not only match the triangular shape and size of the molecule, additionally, the darker regions in the reconstructions correspond to the area of higher density in the molecular projection.

Fig. 4.12d shows two transferrin molecules in different orientations, one in a streched-out and one in a compact shape, as shown by the molecular projections on the right. In terms of overall shape, these reconstructions could have been interpreted as a single molecule, but the comparison with the projections clearly identifies two molecules. The conformation of the molecule on the left is especially interesting, since it corresponds to an open configuration of the left lobe of the molecule. The subunits of transferrin, an iron-transporting protein, assume different conformations depending on whether an iron atom is bound or not. The observation of this open configuration thus indicates that no iron atom is present and demonstrates that LEEH imaging can not only distinguish different protein orientations, but also different conformations, if the change in conformation results in a structural change on a length scale that is resolvable by LEEH.

In the case of GroEL, depicted in Fig. 4.12e, which is the largest protein analysed with our ES-IBD/LEEH set-up, finding the right molecular projections has proven to be more difficult than for other proteins. In many cases, the reconstructions of the experimental hologram appear smaller in size than the projections obtained from the molecular model. While cryo-EM studies of ES-IBD-deposited GroEL [232] show that GroEL can remain intact during the gas phase flight, ion mobility studies suggest that a partial collapse of the molecular structure in the gas phase could occur [233], similarly to the gas phase collapse described for antibody molecules in Chapter 2. The latter phenomenon could explain the size differences observed between the LEEH reconstructions and the molecular model.

Next to the molecular projections, a second type of projection, that will be referred to as *density projection* and depicted in a green colour scale, is used in the following. Density projections are generated in a similar way as the molecular projections, by finding the correct orientation in PyMOL and subsequently creating a projection. While the molecular projection treats each atom as a small disk, the density projection creates a grid of adjustable resolution and calculates a histogram assigning density values according to how many atoms contribute to each histogram cell. This has the advantage that the resulting projections can be quantitatively compared to one another by comparing the calculated density values.

A comparison of density values and phase shift values is of interest for exploring the connection between molecular density and phase shift. Since the reconstructed phase shift values are projected onto the interval $[-\pi, \pi]$ rad, a direct quantitative comparison is not possible. While higher density, in general, leads to higher phase shifts, the cyclic nature of the phase shifts could result in a large absolute phase shift appearing as a small phase shift value. Since the experimental evidence shows that both smaller and larger proteins exhibit phase shifts covering the whole $[-\pi, \pi]$ rad interval, it is likely that the total phase shift induced by a larger molecule is in the range of several multiples of π rad. Still, the comparison of proteins in different orientations that correspond to different densities or molecular thicknesses can yield valuable insights into the connection between phase and molecular density.

Since other factors likely play a role in this, as mentioned at the beginning of section

4.4, it is best to use molecules of the same sample for this kind of comparison, ideally molecules that are located in close proximity to one another since this will ensure identical imaging conditions.
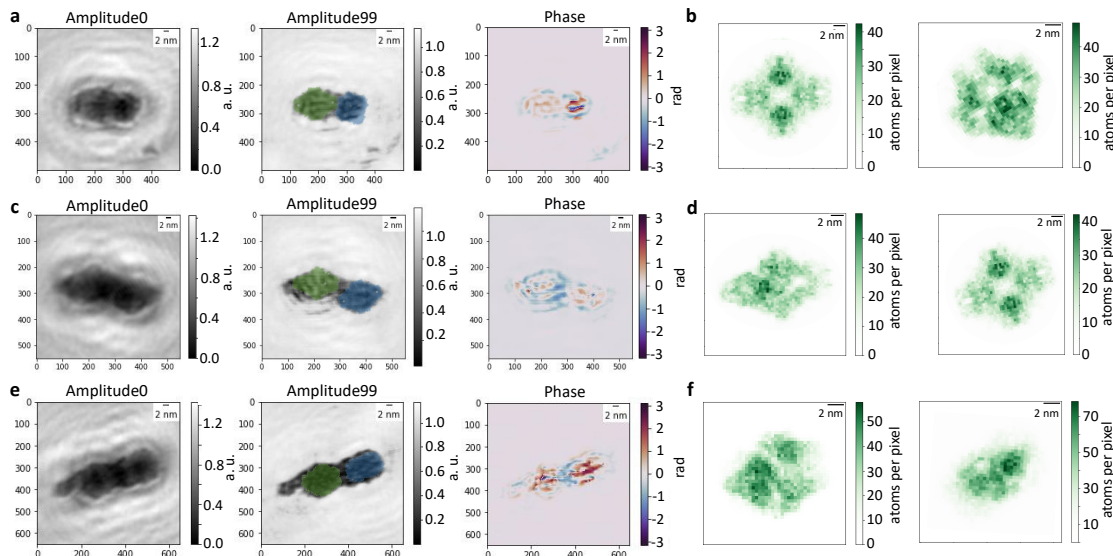

Figure 4.13: **Direct comparison of molecular thickness: dimers of β-Galactosidase: a** Dimer of two β-Galactosidase molecules in two different orientations. The molecular projections (PDB: 6CVM [229]) semi-transparently overlay the iterated amplitude reconstruction and have been coloured green and blue to facilitate the distinction of the two molecules. **b** Density projections for each molecule, the left (right) molecule is depicted on the left (right). **c** Dimer of two β-Galactosidase molecules in two very similar orientations. The molecular projections (PDB: 6CVM [229]) semi-transparently overlay the iterated amplitude reconstruction and have been coloured green and blue to facilitate the distinction of the two molecules. **d** Density projections for each molecule, the left (right) molecule is depicted on the left (right). **c** Dimer of two β-Galactosidase molecules in similar orientations. The molecular projections (PDB: 6CVM [229]) semi-transparently overlay the iterated amplitude reconstruction and have been coloured green and blue to facilitate the distinction of the two molecules. **f** Density projections for each molecule, the left (right) molecule is depicted on the left (right). While the general orientation of the two molecules appears similar, the maximum density is larger for the molecule on the right.

Fig. 4.13 shows three examples of β-Galactosidase dimers. In Fig. 4.13a, c and e, the non-iterated amplitude (Amplitude0), the iterated amplitude (Amplitude99) and the phase reconstructions are shown. Fig. 4.13b, d and f depict the density projections corresponding to the two molecules in each image, respectively. Semi-transparent molecular projections, coloured in green and blue to make them easy to distinguish, overlay the iterated amplitude reconstructions. In Fig. 4.13a, the dimer consists of two molecules in different orientations. The molecule on the left has a lower maximum density value of approximately 40 atoms per pixel and features large areas of low density, while the molecule on the right has a higher overall density as well as a higher maximum density value of approximately 50 atoms per pixel (Fig. 4.13b). The pixel size refers to the size of the cells of the histogram that is used to create the density projections, in the examples presented here, the side length of the cells was chosen to be 4Å. The difference in density is clearly reflected in

the phase reconstruction: the phase shift in the area of the lower-density molecule on the left is much lower than the phase shift induced by the molecule on the right.

The two molecules in Fig. 4.13c are in similar orientations that result in similar density values and distributions (Fig. 4.13d). The phase reconstruction reproduces this, the phase shifts induced by both molecules are almost identical. Fig. 4.13e features two molecules of similar shape, however, the density projections reveal a locally increased density in the molecule on the right (maximum density value of 70 atoms per pixel vs. 50 atoms per pixel in the molecule on the left, see Fig. 4.13f). Although the phase contrast is not as obvious as in Fig. 4.13a, this difference in density correlates with higher phase values in the molecule on the right.

This suggests that the phase shift induced by a molecule is indeed related to the molecular density. To some extent, a similar relationship can also be observed in the amplitude reconstruction, where areas of high density appear to be correlated with areas of lower amplitude. The contrast in the phase images, however, is higher, which facilitates the distinction of areas of different density.

A similar observation can be made when comparing transferrin molecules in different orientations. As shown in Fig. 4.14, compact orientations, corresponding to higher molecular density (Fig. 4.14a, c, maximum density values: 20-25 atoms per pixel, see Fig. 4.14c, f), exhibit a higher overall phase shift over the whole extent of the molecule than less compact orientations (Fig. 4.14g, j, maximum density values: 14-16 atoms per pixel, see Fig. 4.14i, l). While the compact orientations exhibit phase shift values close to π rad over almost the whole area of the molecule, the phase shift values reconstructed in the case of the more extended molecular conformations are lower and only reach values close to π rad in the regions in which the sign of the phase shift changes. Hence, despite not knowing how these phase values quantitatively compare, since we can only determine them modulo 2π, the qualitative difference in the phase reconstructions supports the conclusion that molecular density and induced phase shift are directly related.

In order to pursue this analysis quantitatively, a test system with a defined molecular thickness and density would be required. An ideal test sample would be a system of uniform density that can be produced with varying molecular thicknesses. A possibility for creating such a system could be DNA origami since with this technique, structures with well-defined properties can be tailor-made to fit the requirements of the experiment [235]. Since it is possible to create flat patches of DNA-origami of uniform thickness and defined shape, stacking several of those patches could yield a test system that allows for a quantification of the relation between molecular density and phase shift if they can be prepared on graphene in a way suitable for LEEH imaging.

In summary, comparing both the molecular projections and the density projections to the reconstructed data yields several important insights: Firstly, we can find molecular orientations for which size, shape and, if resolved, inner structure fit the experimentally acquired data, indicating that the molecules are intact and that LEEH imaging can distinguish different molecular orientations on the single-molecule level. Furthermore, important structural information is contained in both amplitude and phase reconstruction, hence both are important to consider when interpreting the experimental data. Secondly, the comparison of density projections to the data implies a connection between molecular density and induced phase shift.
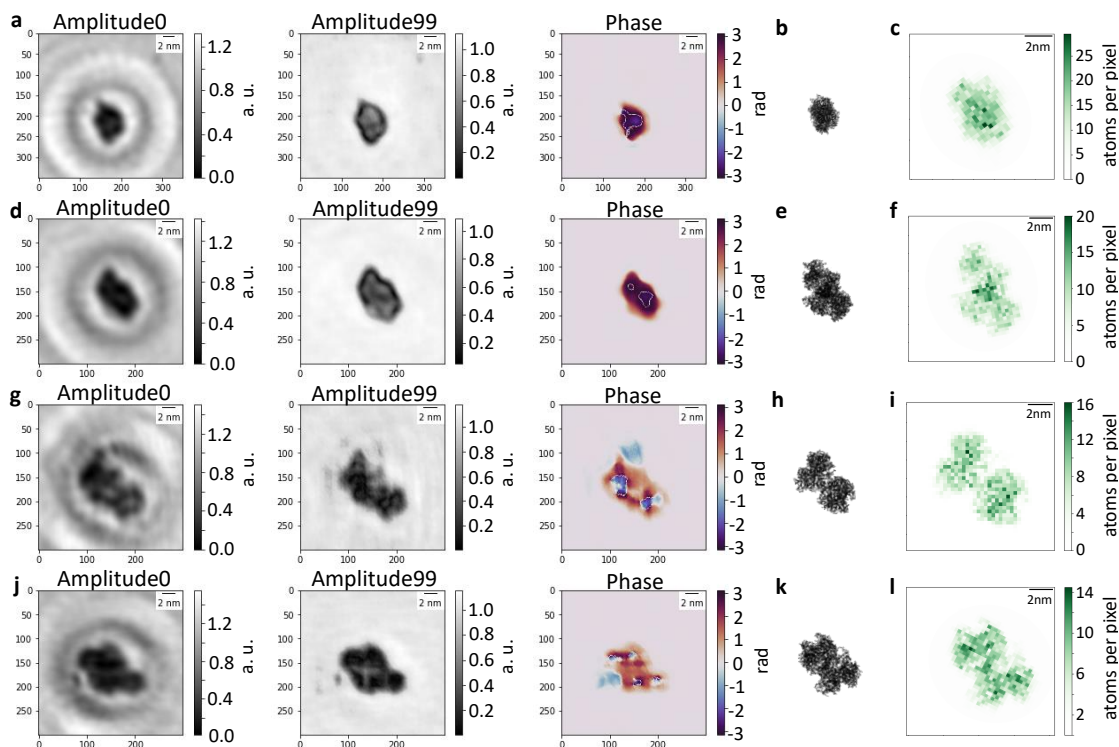
Figure 4.14: **Direct comparison of molecular thickness: Transferrin: a-f** One-step amplitude reconstruction (Amplitude0), iterated amplitude reconstruction after 100 iterations (Amplitude99), phase reconstruction after 100 iterations (**a, d**), molecular projections of each molecule (**b, e**) and corresponding density projections (**c, f**, PDB: 1JNF [234]) of holograms of two different transferrin molecules in a compact conformation associated with higher molecular density. The size of the molecular projection has been adjusted to the scale of the reconstructions, while the density projections are depicted at a slightly larger size to facilitate the identification of density variations. **g-l** One-step amplitude reconstruction (Amplitude0), iterated amplitude reconstruction after 100 iterations (Amplitude99), phase reconstruction after 100 iterations (**g, j**), molecular projections of each molecule (**h, k**) and corresponding density projection (PDB: 4X1B [230] (**i**) and 1JNF [234] (**l**)) of holograms of two different transferrin molecules in an extended conformation associated with lower molecular density. The size of the molecular projection has been adjusted to the scale of the reconstructions, while the density projections are depicted at a slightly larger size to facilitate the identification of density variations. There is a qualitative difference in the phase shift values between the high-density and the low-density molecules.

4.5 Local changes in scattering strength/potential

In section 3.3.2, the interaction between the electron beam and the molecule was modelled by a partial wave approach taking into account the element-dependent scattering strengths and phase shifts for each atom in the simulated molecule. These simulations demonstrate that local differences in potential or scattering strength should be observable, at least for the flat geometries explored in the simulations, given sufficient resolution. This implies that special characteristics of the molecule that could be associated with stronger scattering or local changes in potential, such as the presence of metal ions in a biomolecule that otherwise consists of light atoms, could lead to distinct phase signatures.

While proteins are mostly composed of carbon, nitrogen, oxygen and hydrogen atoms, many proteins contain ions and their relative metal-binding sites, which are often closely associated with the protein's biological function; as an example, the iron in a heme group plays a crucial role in both oxygen transport (hemoglobin) and oxygen storage (myoglobin).

Although it is unlikely that a contribution by a single metal is resolvable, especially in large proteins consisting of thousands of non-metal atoms, it is still of interest to explore the concept of phase features related to changes in scattering strength and in potential. Specifically, the metals occurring within proteins are ions with an associated charge bound to a ligand of opposite charge.

In the following sections, the phase reconstructions of proteins involving metals will be studied for the case of individual metal ions within a protein (section 4.5.1), paying specific attention to the behaviour of metalloproteins containing a heme group (section 4.5.2). Following that, the presence of metals in small molecules will be discussed via the examples of the heme group itself (hemin) and phosphotungstic acid (PTA) (section 4.5.3) and the influence of charges on the phase reconstruction will be explored (section 4.5.4).

4.5.1 Metalloproteins: Alcohol dehydrogenase and Transferrin

While a metal atom in principle leads to a stronger scattering contribution than an element with low atomic number, such as carbon or nitrogen (see section 3.3.2), the differences in phase shift observed in the simulations are likely not large enough to be observable against the contributions of the large number of non-metal atoms within a protein. The data resulting from the reconstruction of experimentally acquired holograms of Alcohol dehydrogenase (ADH) and transferrin molecules, both of which contain a small number of individual metal ions, substantiate this expectation.

Each of the four subunits of ADH contains two zinc ions, one of which plays a critical role in the catalytic process carried out by this enzyme [236]. Fig. 4.15 shows two examples of reconstructions of ADH molecules in different orientations (a, c), with the corresponding molecular projections depicted in b and d. The locations of the metal atoms are marked by orange dots surrounded by red circles for better visibility. In the case of the molecule shown in Fig. 4.15a, no correlation between the structural features in the reconstructed images and the location of the metal atoms is observable. The iterated amplitude reconstruction (Amplitude99) is in good agreement with the molecular projection, including the central area of lower density, but no additional features in the amplitude indicate a resolvable contribution from the metal ions. The phase reconstruction only shows one localised feature and is otherwise uniform. While the localised phase feature on the right does approximately coincide with the location of three of the zinc ions, no corresponding feature is visible for the identical arrangement of zinc ions on the left side of the projection. Hence it is unlikely that this phase feature can be attributed to the presence of the metal ions.

In the molecule depicted in Fig. 4.15c, a four-fold symmetry, corresponding to the four subunits of the molecule, is visible both in the amplitude and in the phase reconstruction. The phase reconstruction exhibits four localised features, one per subunit. The location of the metal ions partially overlaps with the localised phase features, especially on the

left side of the molecule.

Thus, although at least in the case of the molecule in Fig. 4.15c, the location of the metals correlates to some extent with the phase features, there is no one-to-one mapping that would clearly indicate a relationship between the phase features and the metal ions.
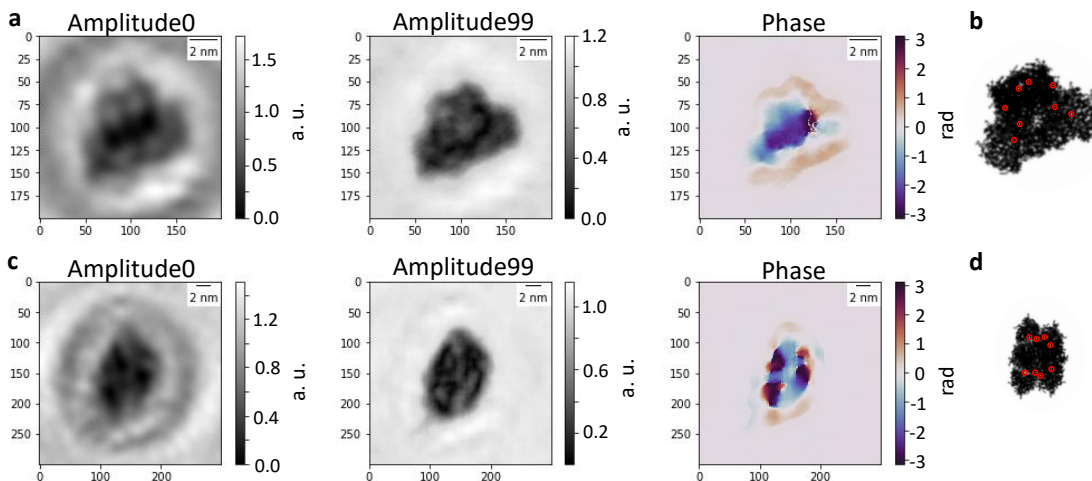

Figure 4.15: **ADH a** Left to right: one-step amplitude reconstruction, iterative amplitude reconstruction after 100 iterations and phase reconstruction after 100 iterations of an ADH molecule. **b** Molecular projection of an ADH molecule (PDB: 7KCQ [225]) in an orientation matching the molecule in the reconstructed images in **a**. The projection is scaled to the same scale as the reconstructed images. The 8 zinc ions are marked by orange dots, which are circled in red for better visibility. **c** Left to right: one-step amplitude reconstruction, iterative amplitude reconstruction after 100 iterations and phase reconstruction after 100 iterations of a different ADH molecule. **b** Molecular projection of an ADH molecule (PDB: 7KCQ [225]) corresponding to the orientation of the molecule in the reconstructed images in **c**. The projection is scaled to the same scale as the reconstructed images. The 8 zinc ions are marked by orange dots, which are circled in red for better visibility.

Another interesting molecule for studying whether the presence of single metal ions within a protein can be detected in the reconstructed phase is transferrin. Since its biological function is the transport of iron in the blood plasma, it can both bind and release iron atoms, and thus exist both in holo form, i.e. with iron atoms bound, and in apo form, i.e. without iron. Comparing these two different forms will hence provide insights into the question whether features occurring in the phase reconstruction can be attributed to the presence of individual metal ions.

Transferrin has two iron-binding sites, hence a transferrin molecule in holo form could have either one or two iron atoms bound. The binding process is related to a conformational change of the subunit from an open form (apo) to a closed form (holo) [237, 238, 239]. Thus, differences in the phase reconstructions between the two molecular forms could also be due to structural differences. The overall structure of the molecule remains similar, and in many molecular orientations, the molecular projections made from PDB structures in both holo and apo form are almost indistinguishable (see for example Fig. 4.16f, i and j). Hence, while open subunit conformations have been observed (Fig. 4.12d, left molecule) in LEEH imaging, these orientations appear to be rare and transferrin molecules in holo

and apo form are in most cases hard to distinguish in terms of molecular shape. Fig. 4.16 shows the reconstructions and corresponding molecular and density projections of five transferrin molecules, four in holo and one in apo form. While the molecules exhibit different features in the respective phase reconstructions, there are no features that would mark a clear distinction between the apo and holo forms.
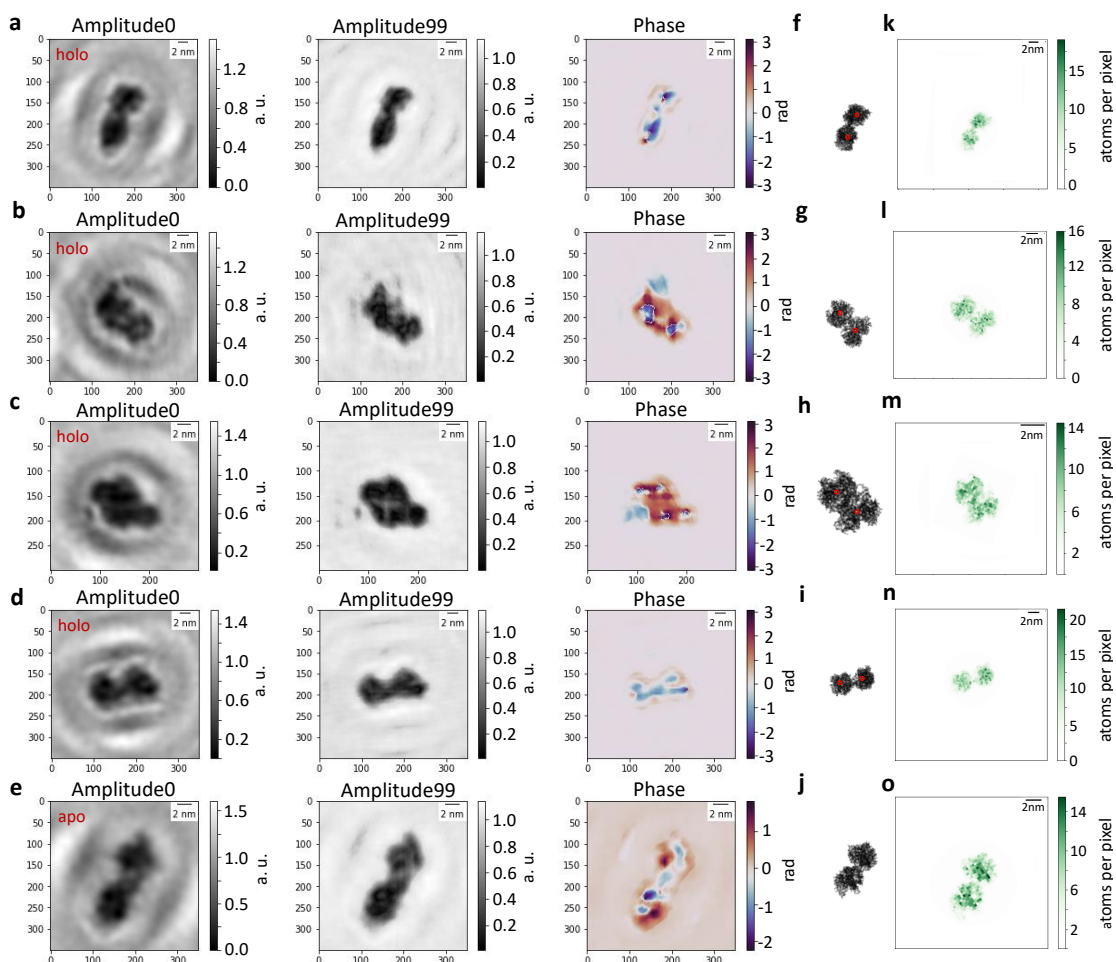

Figure 4.16: **Transferrin: holo vs. apo a-d** Left to right: one-step amplitude reconstruction (Amplitude0), iterative amplitude reconstruction (Amplitude99) after 100 iterations and phase reconstruction after 100 iterations of four different transferrin molecules in holo form. **e** Left to right: one-step amplitude reconstruction (Amplitude0), iterative amplitude reconstruction (Amplitude99) after 100 iterations and phase reconstruction after 100 iterations of a transferrin molecule in apo form. **f-i** Molecular projections of a transferrin molecule in holo form (PDB: 1JNF [234]) in orientations matching the orientations of the molecules shown in **a-d**, respectively. The projections are scaled to the respective scale bars of the corresponding reconstructed images. In each projection, the two iron atoms are marked by orange dots, which are circled in red for better visibility. **j** Molecular projection of a transferrin molecule in apo form (PDB: 1RYX [240]) in an orientation matching the orientation of the molecule shown in **e**. The projection is scaled to the scale bar of the corresponding reconstructed image. **k-n** Density projections of the respective models in **f-i**. **o** Density projection of the model in **j**.

The molecules shown in Fig. 4.16a, d (holo) and e (apo) have very similar conformations. Their density projections also closely resemble one another; despite the maximum density values differing slightly, the mean density over the extent of the molecule appears to be approximately the same in all three cases. The lack of differences between the density projections of transferrin molecules in holo and apo form is not surprising, since in a protein consisting of several thousands of atoms, the presence or absence of a single atom should not noticeably affect the density projection. Localised features of high phase also do not seem to be a distinguishing characteristic of the molecules in holo form, since the holo-form molecule in Fig. 4.16a exhibits them, while the holo-form molecule in Fig. 4.16d as well as the apo-form molecule in Fig. 4.16e do not. The localised phase features in Fig. 4.16a correlate with the location of the metal ions in the molecule in the sense that there is one feature per subunit, but since the reconstructed molecule appears slightly more extended than the corresponding projection, it is hard to determine whether there is an actual overlap of the phase feature locations and the iron binding sites. In general, for both molecules in holo and in apo form in orientations similar to the ones shown in Fig. 4.16f, i and j, the overall sizes of the reconstructed molecules appear slightly larger than the projected sizes, while the sizes of the individual subunits match in reconstruction and projection.

Regarding localised phase features, both Fig. 4.16b and c prominently feature areas in which the phase changes sign. In Fig. 4.16b, two such features occur; while they approximately match the location of the metal atoms, they extend over a significant part of each subunit, thus making it unlikely that these contributions are associated with a single metal atom. The phase reconstruction of the molecule in Fig. 4.16c exhibits four similar, albeit smaller, phase features. Given that a transferrin molecule can only bind two iron atoms, these features cannot be attributed to the presence of metal atoms. When comparing the location of the phase features to the projected molecular structure and density, one can see that these phase features are associated with larger structural components and the corresponding densities, which further supports the conclusion that phase and density are related.

In conclusion, there is no indication that the presence of single metal atoms produces an observable contribution to the phase reconstruction of a protein consisting of several thousands of atoms.

4.5.2   Heme-bearing metalloproteins

Metal ions represent some of the simplest cofactors for proteins, but more elaborate molecular structures featuring the presence of metals often appear as prosthetic and coenzyme groups. One of the most prominent example of such a structure is the heme group, a porphyrin-derivative with a central iron ion, that plays a crucial role in oxygen transport and storage. As opposed to the individual metals discussed in the previous section, the heme-bearing metalloproteins exhibit a conjugated π-system along with the presence of the iron. Additionally, the examples of proteins with heme groups discussed in this section are smaller than the proteins that are subject of the previous sections. The heme group itself has a diameter of approximately 1 nm [223], hence it takes up a significant amount of space within the protein for the examples discussed here (see e. g. Fig. 4.19d for the case of cytochrome C, which has a molecular diameter of

approximately 2.5 nm [224]). Thus, the heme group could contribute significant features to the phase reconstruction if it interacts differently with the electron beam than the rest of the protein.
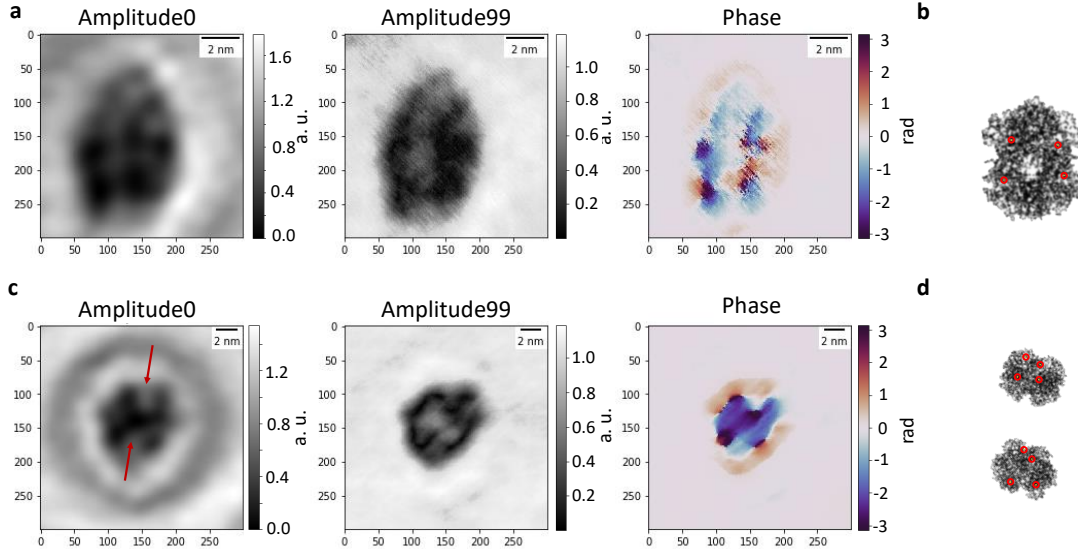

Figure 4.17: **Hemoglobin: a, c** Left to right: one-step amplitude reconstruction (Amplitude0), iterative amplitude reconstruction (Amplitude99) after 100 iterations and phase reconstruction after 100 iterations of two different hemoglobin molecules. **b, d** Molecular projections of a hemoglobin molecule (PDB: 1FSX [223]) matching the orientations in **a** and **c**, respectively. In each projection, the 4 iron ions in the heme groups are marked by orange dots, which are circled in red for better visibility.

Fig. 4.17a and c shows the amplitude and phase reconstructions of two hemoglobin molecules in different orientations and the corresponding molecular projections (Fig. 4.17b, d). Hemoglobin consists of four subunits, each of which contains a heme group. The locations of the iron ions in the heme groups have been marked by orange dots encircled in red in the molecular projections in Fig. 4.17b, d. In the case of the molecule shown in Fig. 4.17a, several structural features are discernible both in the amplitude and phase reconstructions, including the four subunits and the area of low density at the centre of the molecule. The four localised features visible in the phase reconstruction match the location of the metal ions, i.e. the heme groups, very well.

The case of the molecule in Fig. 4.17c is less clear. While the phase reconstruction also exhibits four localised features, the exact orientation of the molecule is hard to determine, especially since the iterated amplitude (Amplitude99) does not feature a well-resolved inner structure. Fig. 4.17d depicts two possible projections which fit the molecular structure, although neither perfectly reproduces the two narrow bright areas in between the subunits (indicated by red arrows) in the non-iterated amplitude (Amplitude0). The upper projection fits the structure in Amplitude0 slightly better, but the phase features do not match the locations of the heme groups. Although only marginally different in terms of overall shape, the lower projection features an arrangement of heme groups that coincides with the localised phase features. While orientations can be found in which the

locations of the heme groups fit the phase features, some ambiguity remains due to the globular nature of the protein.

The statistical evaluation of the sample is complicated by this ambiguity in orientation since cases like the molecule in Fig. 4.17a, for which the orientation can be determined with a high degree of accuracy, are rare. Furthermore, many of the iterated amplitude reconstructions suffer from a lack of inner structure as in Fig. 4.17c. As shall be discussed in section 4.6, this could affect the phase reconstruction, and introduces a further degree of uncertainty into the interpretation of the reconstructed images.

While hemoglobin is still a relatively large molecule (molar mass 64 kDa, molecular diameter approx. 5 nm [223]), the two other molecules to be discussed in this section are much smaller. The first one, myoglobin, an oxygen-storing protein, has a mass of 17 kDa and a diameter of approximately 3 nm [241] and thus roughly corresponds to one of the subunits of hemoglobin. A myoglobin molecule contains a single heme group.

Fig. 4.18a and c depict the amplitude and phase reconstructions of two different myoglobin molecules alongside possible molecular projections in Fig. 4.18b, d. The location of the
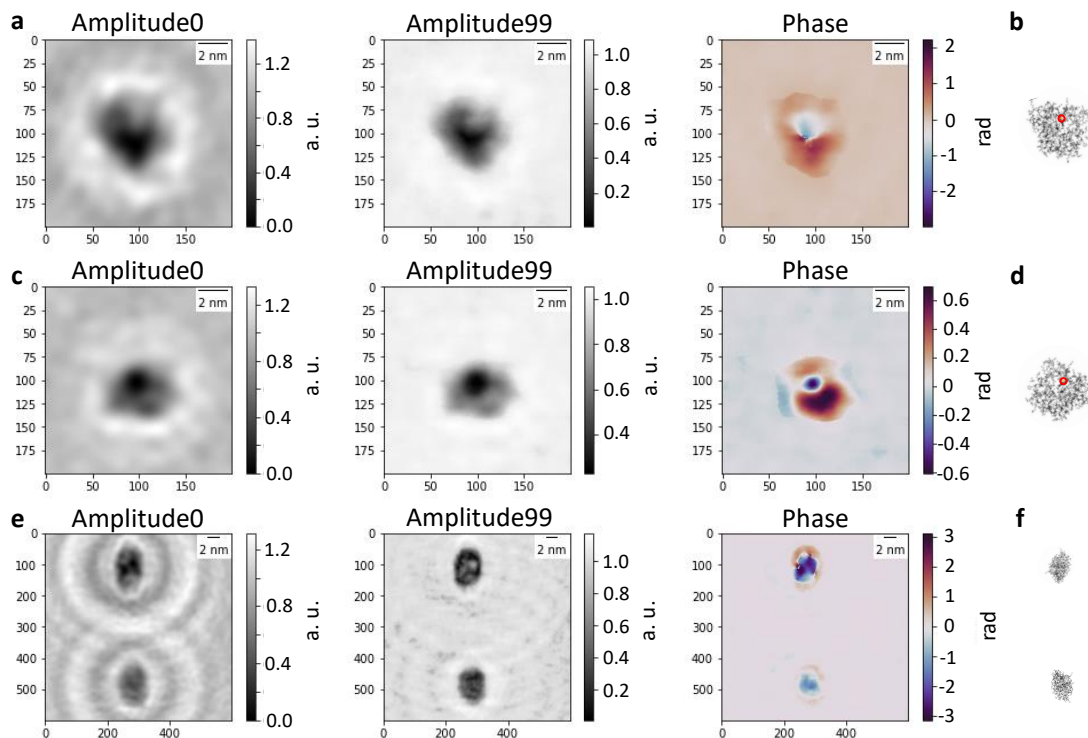

Figure 4.18: **Myoglobin: a, c** Left to right: one-step amplitude reconstruction (Amplitude0), iterative amplitude reconstruction (Amplitude99) after 100 iterations and phase reconstruction after 100 iterations of two different myoglobin molecules. **b, d** Molecular projections of a myoglobin molecule (PDB: 5ZZE [241]) matching the orientations in **a** and **c**, respectively. In each projection, the iron atom in the heme group is marked by an orange dot and circled in red for better visibility. **e** Left to right: one-step amplitude reconstruction (Amplitude0), iterative amplitude reconstruction (Amplitude99) after 100 iterations and phase reconstruction after 100 iterations of two apo-myoglobin molecules. **f** Molecular projections of a myoglobin molecule (PDB: 5ZZE [241]) matching the orientations in **e**.

iron ion in the heme group is again marked by an orange dot surrounded by a red circle. Both molecules exhibit a localised phase feature, which corresponds to a dark feature in the amplitude reconstructions, respectively. Orientations for which the location of the heme group in the molecular projection matches the localised phase feature can be found, but the protein is structurally highly unspecific, hence it is hard to find criteria for identifying the correct orientation.

Like transferrin, myoglobin can also exist in an apo form, i. e. without the heme group. Although being slightly more disordered, the overall structure of apomyoglobin remains similar to myoglobin [242]. Fig. 4.18e shows the amplitude and phase reconstructions of two apomyoglobin molecules. Molecular projections obtained from the same PDB structure as in Fig. 4.18b, d depict possible orientations. In contrast to the myoglobin molecules in Fig. 4.18a, c, no localised phase features occur within the molecules. This difference might point towards a contribution from the heme group that is visible in the phase reconstruction. However, because the molecules are small and appear to be mobile on the graphene surface during the landing process, many objects observed in LEEH on both the myoglobin and apomyoglobin samples are dimers or clusters, or suffer from artefacts such as charging, possibly due to the molecules adsorbing at defects on the SLG surface, which exacerbates the interpretation.
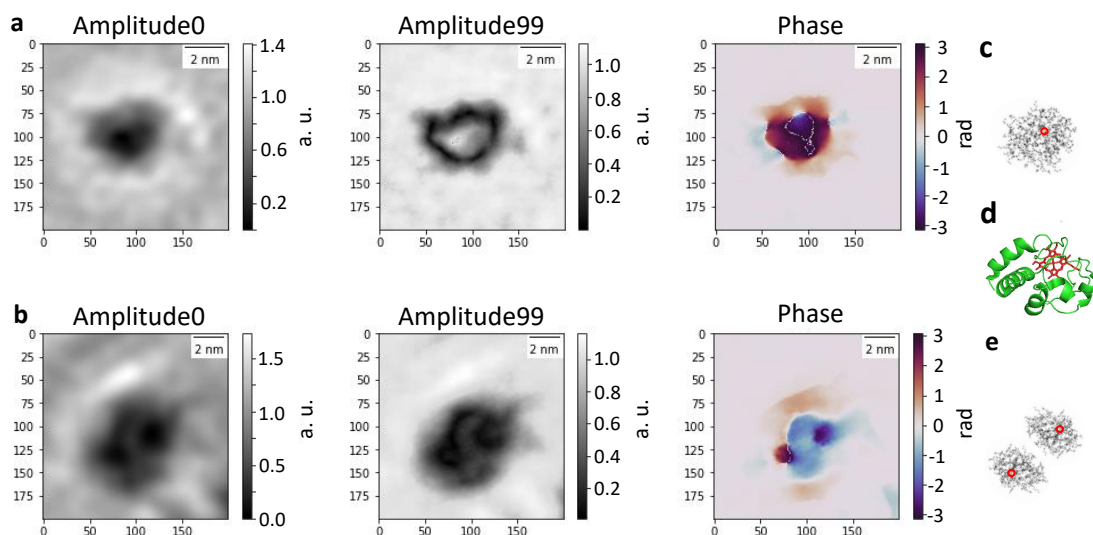

Figure 4.19: **Cytochrome C: a** Left to right: one-step amplitude reconstruction (Amplitude0), iterative amplitude reconstruction (Amplitude99) after 100 iterations and phase reconstruction after 100 iterations of a single cytochrome C molecule. **b** Left to right: one-step amplitude reconstruction (Amplitude0), iterative amplitude reconstruction (Amplitude99) after 100 iterations and phase reconstruction after 100 iterations of a dimer of cytochrome C molecules. **c** Molecular projection of a cytochrome C molecule (PDB: 1HRC [224]) matching the orientation in **a**. The location of the iron in the heme group is marked. **d** Ribbon-sytle molecular model of cytochrome C. The heme group is shown in red. **e** Molecular projections of two cytochrome C molecules (PDB: 1HRC [224]) matching the orientations in **b**. The location of the iron in the heme group is marked.

The situation is similar with respect to cytochrome C, which is even smaller than

myoglobin (12 kDa, molecular diameter 2.5 nm [224]) and also contains one heme group. An even higher percentage of the molecules appears in clusters.

Fig. 4.19 shows the amplitude and phase reconstructions of a single cytochrome C molecule (Fig. 4.19a) and of a dimer consisting of two cytochrome C molecules (Fig. 4.19b) alongside the corresponding projections with the location of the iron ion marked in orange and encircled in red (Fig. 4.19c, e). A ribbon molecular model [224] is depicted in Fig. 4.19d to illustrate the size of the heme group in relation to the size of the molecule. As in the case of myoglobin, orientations can be found for which the position of the heme group matches that of the phase feature. The iterated amplitude reconstruction of the single molecule, however, lacks inner structure, the relatively large phase feature in the corresponding phase reconstruction could be influenced by that.

In conclusion, there is some indication that the positions of localised phase features can be correlated with the positions of heme groups within the molecules discussed in this section. However, the globular nature of the proteins in question often does not allow a clear identification of the molecular orientation of the imaged molecules. In general, due to the high number of clusters on the sample and the artefacts occurring in the reconstructions, the statistical analysis of these samples is difficult to interpret.

4.5.3 Small metal-containing molecules

Possible contributions to the phase signal from metal ions within a protein are difficult to clearly identify due to the large number of non-metal atoms in proteins and the resulting three-dimensional nature of the molecules. A possibility to gain a better understanding of the effect that the presence of metals has on the phase reconstructions of experimental holograms, as well as of the interpretation of reconstructed phase features in general, could be to measure holograms of small and flat molecules involving a small number of metal ions. As shown in section 3.3.2, holograms of molecules like phtalocyanines and porphyrins with metal centres can be simulated on the atomic level using a scattering algorithm, thus providing a comparison to possible experimental data. Hence, a molecule of similar size and structure would work well as a test system.

However, sample preparation for LEEH imaging for these kinds of molecules has proven to be difficult. Ideally, a sample for a LEEH measurement would feature single porphyrin or phtalocyanine molecules on SLG that are neither themselves charged nor sit on the site of a charged defect.

As discussed in section 1.2.3, the protein samples used in the LEEH experiments presented in this thesis have been prepared by native ES-IBD. Ionizing porphyrins and phtalocyanines non-destructively using ESI is in general not practical because solvents suitable for these molecules are not compatible with ESI.

The only porphyrin system that we could deposit onto SLG via ES-IBD is hemin, which is a heme group with a chlorine atom attached to the iron core of the porphyrin structure. During the ESI process, the chlorine atom is removed, thus creating a heme cation that is then deposited on the surface. The molecule will be referred to as hemin in the following, even though the chlorine atom is not present in the molecules observed on the surface. Fig. 4.20 shows two examples of Hemin molecule clusters reconstructed from LEEH measurements. Since the molecules are mobile on the graphene upon landing, no isolated hemin molecules have been observed on the sample. In some cases, the number
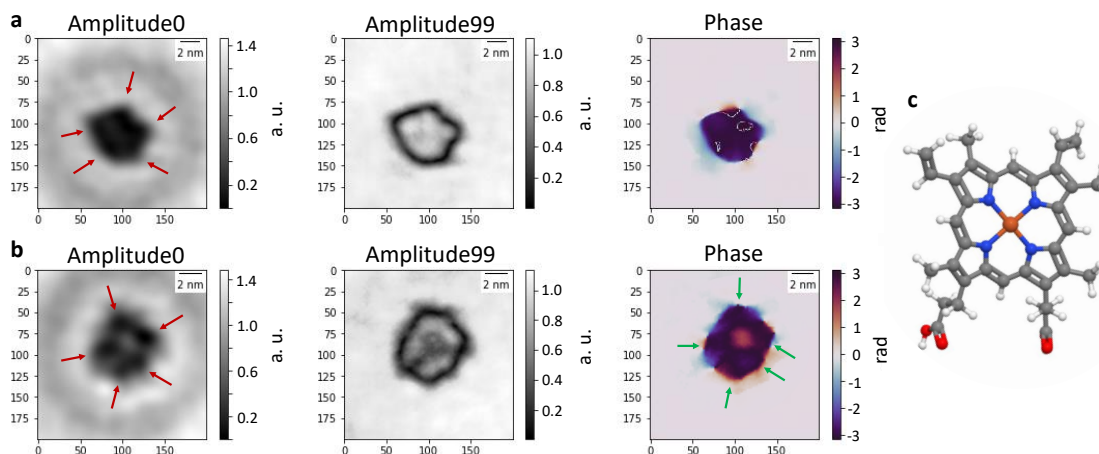
Figure 4.20: **Hemin: a, b** Left to right: one-step amplitude reconstruction (Amplitude0), iterative amplitude reconstruction (Amplitude99) after 100 iterations and phase reconstruction after 100 iterations of two clusters of 5 Hemin molecules, respectively. The positions of the individual molecules are indicated by the red arrows in the non-iterated amplitude reconstruction. **c** Molecular structure of an individual heme group.

of molecules within a cluster of hemin molecules can be determined. The positions of the individual molecules are marked by red arrows in the non-iterated amplitude (Amplitude0) reconstruction images in Fig. 4.20a and b. The molecules are especially well-distinguishable in the non-iterated amplitude reconstruction shown in Fig. 4.20b. In our data set, the hemin molecules all exhibit very high amplitude contrast, i. e. appear very dark, even though no strong absorptive behaviour is expected for small, flat molecules.

The phase reconstructions are harder to interpret. It could be argued that the positions of the molecules as identified in the Amplitude0 image in Fig. 4.20b coincide with extrema in the phase reconstruction, represented by a dark red colour in the corresponding phase reconstruction and marked by green arrows, although these phase features are not clearly localised. The phase reconstruction in Fig. 4.20a exhibits four features with clearer boundaries; however, despite the individual molecules not being as easily distinguishable as in Fig. 4.20b, the size of the cluster indicates that it likely consists of 5 molecules.

While there is some indication that an individual molecule can be associated with a phase feature within a cluster, the resolution is not sufficient to be able to attribute these features to the iron atom within the molecules.

As a further test system, we tried to evaporate platinum phtalocyanine (PtPc) on SLG. After evaporating for 90 minutes, molecules were visible on the surface, many of them were, however, charged, likely because the molecules are mobile on the SLG surface and thus settle on charged defect sites. While in general, it can be possible to reconstruct objects in the presence of charges (see section 4.5.4), this was not the case for this sample. The reconstruction process was severely hampered by the charges, hence it was not possible to determine the size of the clusters or infer further structural detail.

Given the aforementioned problems with sample preparation for porphyrin and phtalocyanine systems, another small molecule including metals, phosphotungstic acid (PTA),

for which sample preparation by ES-IBD had been demonstrated [243], was used as a further test system. While PTA is of a similar size as a porphyrin [243], instead of a single metal ion, PTA contains 12 tungsten atoms, which should result in strongly absorbing properties and thus, high contrast, despite the small size.

Fig. 4.21 shows two examples of PTA clusters observed on the surface, with 2-3 molecules in the cluster in Fig. 4.20a and at least 5 molecules in the cluster in Fig. 4.20b. While the molecules are at least to some extent distinguishable in the non-iterated amplitude, this does not translate to a corresponding phase feature in all cases. Amplitude and phase contrast appear to be overall similar, i.e. darker features in the amplitude are associated with darker features in the phase.
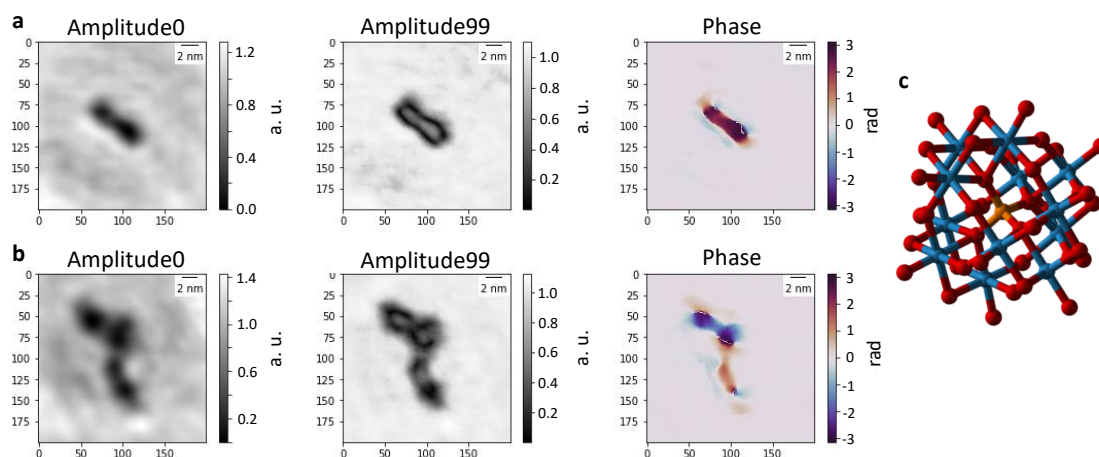

Figure 4.21: **PTA: a, b** Left to right: one-step amplitude reconstruction (Amplitude0), iterative amplitude reconstruction (Amplitude99) after 100 iterations and phase reconstruction after 100 iterations of two clusters of PTA molecules. The cluster in **a** likely consists of 2-3 molecules, the one in **b** of at least 5 molecules. **c** Molecular structure of a PTA molecule.

In both cases of the molecules presented here, the resolution is not sufficient to determine whether the phase contributions can be associated with the presence of individual metal atoms. Additionally, while the iterative phase reconstruction shows interesting features, the iterated amplitude reconstructions of all molecules analysed on these samples lacked inner structure to varying degrees. The origin of this behaviour is not fully clear and will be discussed in more detail in section 4.6.

Since sample preparation has been a major limiting factor in the study of small molecular systems, it seems that a necessary next step would be to find test systems that can be prepared either by ES-IBD, possibly by deposition on a cold substrate, or by dropcasting.

4.5.4 Charge

As mentioned in the previous section, charges on or close to the molecules can severely hinder the reconstruction process or alter the appearance of the molecule. This is, however, not always the case, and depends on the sign of the charge, the amount of charge and the location of the charges relative to the molecule.

In general, studying the effects of charge in the framework of phase reconstruction is of interest since the presence of a charge changes the electron trajectories, the resulting path difference then induces a phase shift [244]. Neglecting exchange interaction and polarization effects, the reconstructed phase shift distribution can be related to the projected electric potential of the charge [221, 49]. Gaining an understanding of how to interpret charges in the context of phase reconstruction could be a first step towards imaging local electric potentials within a molecule.

Fig. 4.22 shows examples of charges on or near molecules. While these charges hinder the reconstruction to various degrees, the charges converge in the same focal range as the molecules. If this is the case, a reconstruction of the molecule is often possible with only minor artefacts, if the charges do not fully converge, the artefacts are more severe and a high-quality reconstruction of the molecule is not possible.

Fig. 4.22a depicts a β-Galactosidase molecule with a charge, appearing as a black spot in the amplitude reconstruction on one end of the molecule, the corresponding hologram is shown in Fig. 4.22b. The position of the charge is marked by red arrows both in the non-iterated amplitude reconstruction and in the hologram. In the phase reconstruction, the charge clearly is the dominating feature, it appears as a blue spot. While the charge
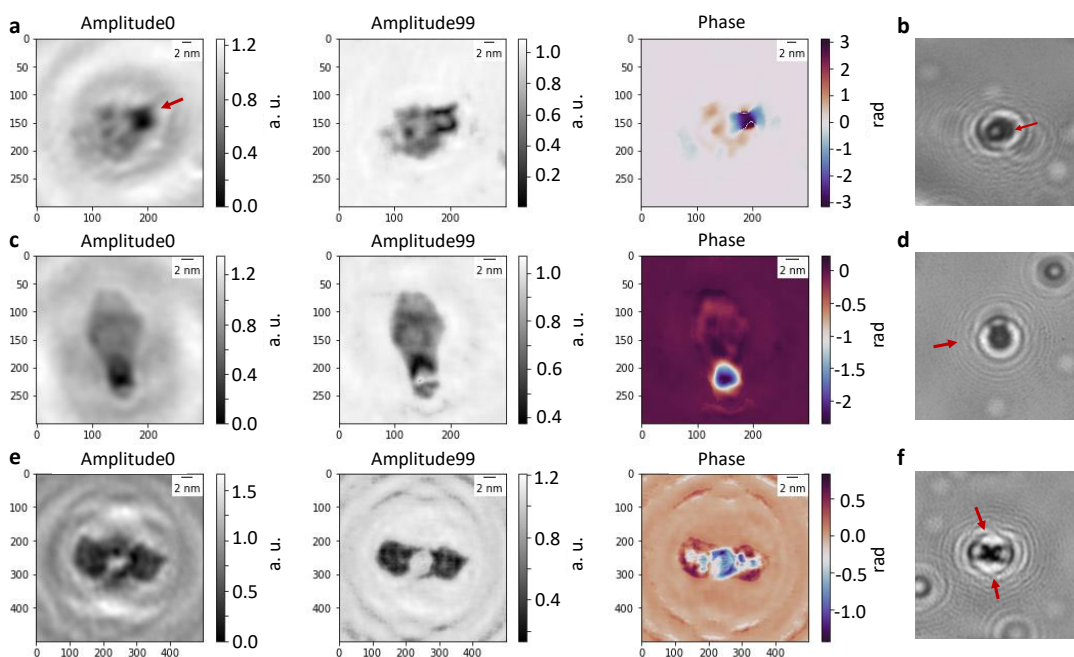

Figure 4.22: **Charge: a** Left to right: one-step amplitude reconstruction (Amplitude0), iterative amplitude reconstruction (Amplitude99) after 100 iterations and phase reconstruction after 100 iterations of a β-Galactosidase molecule with a charge on one end of the molecule. **b** Hologram corresponding to the reconstructions in **a**. The location of the charge has been marked with red arrows in both **a** and **b**. **c** Left to right: Amplitude0, Amplitude99 and phase reconstruction of a β-Galactosidase molecule with a charge on one end of the molecule. **d** Hologram corresponding to the reconstructions in **c**. The location of the charge has been marked with red arrows in both **c** and **d**. **e** Left to right: Amplitude0, Amplitude99 and phase reconstruction of an antibody molecule with a charged hinge region. **f** Hologram corresponding to the reconstructions in **e**. The location of the charges has been marked with red arrows in both **e** and **f**.

dominates both the amplitude and the phase reconstruction in terms of contrast, the inner structure of the molecule is still observable in both the amplitude and the phase reconstruction (see Fig. 4.12b for a comparison to the PDB structure). Without the charge, the contrast of the inner structure of the molecule would be higher, but the presence of the charge does not severely hinder the reconstruction.

In contrast to that, the charge in Fig. 4.22c clearly dominates the phase reconstruction. The phase shift due to the charge is much stronger than the phase shift due to the molecule, which results in the phase shift due to the molecule appearing at values similar to the background intensity. In the amplitude reconstructions, the charge is also the dominant feature, but the shape of the molecule is still reconstructed with sufficient contrast. The dominance of the charge in these images is already visible in the hologram (Fig. 4.22d), where the hologram of the molecule is almost fully obscured by the hologram of the charge. In comparison to the example in Fig. 4.22a, b, it seems likely that the amount of charge present in Fig. 4.22c, d is larger than in Fig. 4.22a, b.

The example in Fig. 4.22e, f shows two lobes of an antibody molecule with a charged hinge region in between. Given both the shape of the hologram (Fig. 4.22f) and the two features visible in the amplitude reconstructions in between the two subunits, it seems likely that there are two charged sites, as indicated by the red arrows. The phase reconstruction of this molecule-charge system is particularly interesting: while in the two examples in Fig. 4.22a and c, the feature produced by the charge in the amplitude reconstruction was enhanced in the phase reconstruction, but appeared in the same position and shape, in Fig. 4.22e the phase reconstruction shows a new feature occupying the whole of the space in between the subunits. This feature could be interpreted as the potential due to the charge distribution.

In summary, the phase reconstruction is sensitive to charges. Although this can in some cases lead to a decreased quality of the reconstruction, in other cases, additional information about the local charge distribution can be gleaned from the features in the phase reconstruction.

4.6 Artefacts

As discussed at different points throughout this chapter, many different factors can induce artefacts, such as sharp edges (section 4.1), modulations due to fringes of nearby molecules (section 4.4.2) and charges (section 4.5.4). While they usually lead to a deterioration of the quality of the reconstruction, these types of artefacts are easy to recognize and understand. In this section, two further types of artefacts will be discussed that have been observed in the reconstructions of experimental holograms, but not in the reconstructions of simulated holograms (see Chapter 3).

The first of the two, which shall in the following be referred to as *negative absorption spot* artefact, only occurs very rarely and is in most cases not severe. It is directly related to the phase constraint employed during the iterative reconstruction steps, and thus easy to understand. The second type of artefact, whose hallmark is a loss of inner structure in the iterated amplitude, occurs more frequently, at least for certain types of molecules, and will be called *hollow object* artefact in the following. Several examples of this type of

artefact can be seen in Figs. 4.17-4.21. Its origins are, however, not yet fully understood.

Negative absorption spot artefacts

As shown in Fig. 4.23a, b, the negative absorption spot artefact is an artefact that affects individual pixels. They occur as discontinuous black pixels in the iterated amplitude reconstruction and as discontinuous white pixels in the phase. Negative absorption spot artefacts can occur both in small numbers, i.e. only a few pixels are affected, and in large quantities as in Fig. 4.23a, b. Large numbers of negative absorption spot artefacts often occur in conjunction with the hollow object artefact.
The origin of the negative absorption spot artefact is the additional phase constraint that has been introduced in section 3.1.1 and which sets the phase value of every pixel with negative absorption values to zero in each step of the iteration. In general, negative absorption values mostly occur in the image background due to contributions by the twin image (see sections 1.1.1 and 3.1.1). Since the background phase shift should be zero, applying the constraint in these cases yields the correct phase value.
Negative absorption spot artefacts occur if pixels with negative absorption exist within the molecule. The phase values within the molecule are in general not zero, thus repeatedly setting the phase values of the negative absorption pixels to zero can result in pixel-wise artefacts. That this is indeed the origin of the artefacts can be tested by reconstructing the same hologram iteratively without applying the phase constraint. The result is shown in Fig. 4.23c, d. The negative absorption spot artefacts are removed, however, the phase reconstruction in Fig. 4.23d differs from the phase reconstruction in Fig. 4.23b as expected due to the discrepancy in the applied constraints. It has been shown for both simulated (section 3.1.1) and experimental examples (section 4.2) that enforcing the phase constraint leads to a higher-quality phase reconstruction. Thus, it is preferable to apply the constraint despite it potentially inducing negative absorption spot artefacts, especially since the artefacts rarely occur and if they do, then usually in conjunction with either charges or hollow object artefacts. Possible ways of avoiding negative absorption spot artefacts could be restricting the application of the phase constraint to the background, e.g. by an additional support constraint, or only applying the phase constraint for the first few iterations until the background phase has been stabilized.

Hollow object artefacts

The potentially most severe artefact so far encountered in the iterative reconstruction of experimentally acquired LEEH holograms are hollow object artefacts. These artefacts are characterized by a loss of inner structure in the amplitude reconstruction during the iterative process, resulting in a molecular shape with a thin, dark border and a bright, often completely structureless centre of amplitude 1, i.e. matching the background amplitude.
For certain classes of molecules, hollow object artefacts are a common occurrence: almost all analysed Hemin and PTA molecules featured hollow object artefacts (see Fig. 4.20, Fig. 4.21). Small proteins (hemoglobin, myoglobin, cytochrome C) also appear to be prone to hollow object artefacts, while this type of artefact is exceedingly rare in the
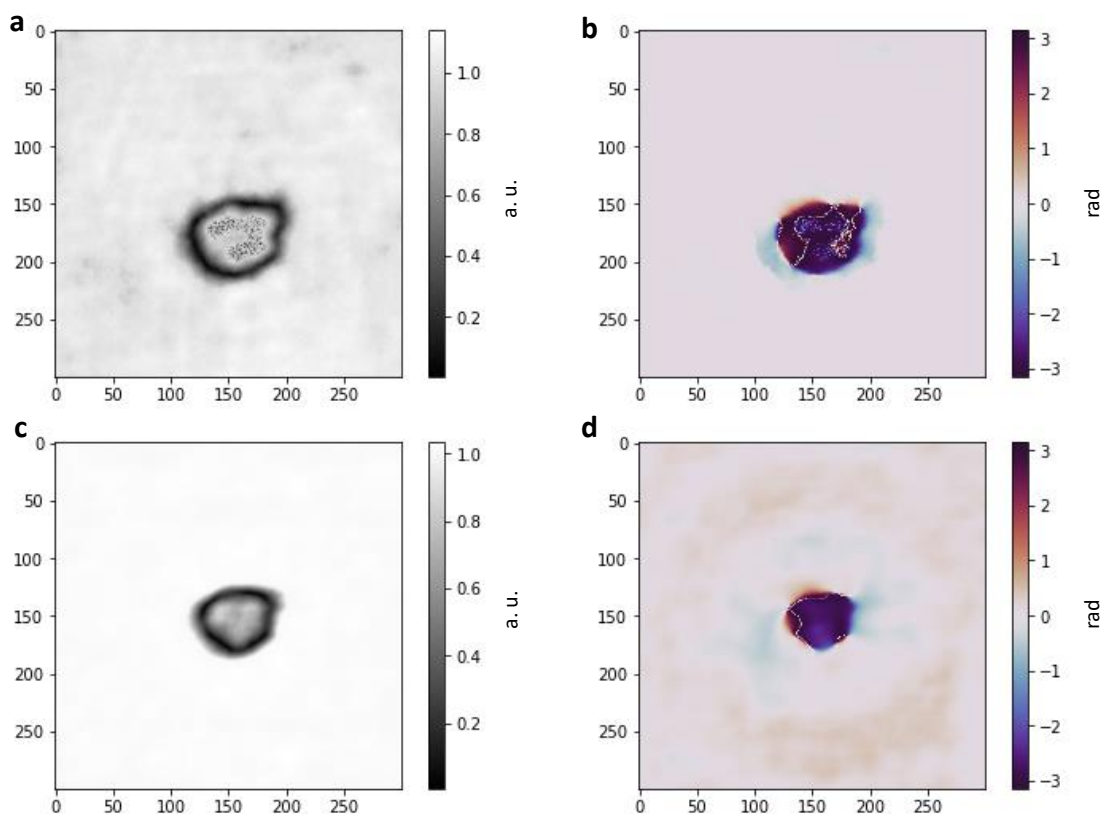
Figure 4.23: **Negative absorption spot artefact: a** Iterative amplitude reconstruction of a Hemin molecule after 100 iterations with both the absorption and the phase constraints applied. The black spots within the molecule correspond to the negative absorption spot artefacts. **b** Phase reconstruction of a Hemin molecule after 100 iterations with both the absorption and the phase constraints applied. The white spots within the molecule correspond to the negative absorption spot artefacts. **c** Iterative amplitude reconstruction of the same Hemin molecule after 100 iterations with only the absorption constraint applied. In this case, no negative absorption spot artefacts occur. **d** Phase reconstruction of the same Hemin molecule after 100 iterations with only the absorption constraint applied. The artefacts are removed, but the overall phase reconstruction also differs.

reconstructions of larger proteins.

Independently of the imaged molecule, hollow object artefacts are in most cases correlated with extended dark areas in the non-iterated amplitude, i. e. areas of low non-iterated amplitude with little to no amplitude variation. Dark areas are in general associated with strongly absorbing properties, however, strong absorption is unlikely to be the origin of these artefacts. If strong absorption would be the source of hollow object artefacts, one would expect these artefacts to mainly occur within large proteins. However, the opposite is the case, small proteins are disproportionately more affected by hollow object artefacts than large proteins, and the smallest molecules studied in this thesis, hemin and PTA, most consistently exhibit hollow object artefacts. While PTA, containing 12 tungsten atoms, could be a strong absorber, hemin should only very weakly absorb, given the discussion regarding electron penetration depth (section 4.4.1).

Despite the correlation between hollow object artefacts and very dark non-iterated amplitude appearances, there might not be a physical reason for this type of artefact, i. e. the artefact could be due to properties of the algorithm. The fact that the amplitude value in the inner part of the hollow objects matches the background amplitude suggests that this artefact is related to the amplitude constraint, which sets large amplitude values above 1 to 1. It is, however, unclear, how this fits with the dark amplitude appearances. Another possibility for explaining the hollow object artefact in terms of the algorithm is to compare it to an artefact we have encountered during the discussion of simulated objects that is at least similar in its general appearance. In section 3.2.4, the dependence of the quality of the reconstruction on object size was explored. If a certain ratio of (dark) object size to illuminated area was exceeded, artefacts resulting in an amplitude modulation occurred, which in many cases lead to a brighter appearance of the object's centre.

To check whether the hollow object artefacts are related to the area ratio between object size and illuminated area, the dependence of the occurrence of hollow object artefacts on the hologram size has to be tested. Fig. 4.24 explores this size-dependence for the example of a hemoglobin molecule (Fig. 4.24a, b) and a cluster of hemin molecules (Fig. 4.24c, d). In the case of hemoglobin, a larger hologram size makes a difference: the iterated amplitude in Fig. 4.24b is no longer hollow, however, as a result, the phase reconstruction changes drastically, too. In the phase reconstruction in Fig. 4.24a, the background is uniform, implying that contributions due to the twin image have been removed and the phase contrast in the centre of the molecule is higher than at the edges. In the phase reconstruction in Fig. 4.24b, on the other hand, twin image contributions in the form of fringes are retained in the background and the phase contrast of the molecule compared to the background is low. Additionally, the edge of the molecule features prominently in the phase in Fig. 4.24b, not unlike the edge of the molecule in the hollow object iterated amplitude reconstruction in Fig. 4.24a. In general, both the iterated amplitude and phase reconstructions shown in Fig. 4.24b are quite similar to the corresponding non-iterated amplitude and phase reconstructions. It thus seems that the iterated amplitude in Fig. 4.24b and the phase in Fig. 4.24a should represent the molecular structure more closely than their respective counterparts, the question is whether these two reconstructions can appear together as the result of an iterative reconstruction. Tests with different-sized holograms as well as changes in the other parameters such as the radius of the apodization filter have stably led to one of the two behaviours depicted in Fig. 4.24a and b, respectively. Either the iterated amplitude is hollow while the phase exhibits localised, high-contrast features, or the iterated amplitude is not hollow while the phase features are lost, but the non-hollow iterated amplitude and the high-contrast phase never occur together.

In contrast to that, changing the hologram size does not make any difference in the case of the hemin molecule cluster, the reconstructions look almost identical (Fig. 4.24c, d).

Since no feature of the algorithm could be clearly identified to be responsible for the hollow object artefacts, possible physical explanations need to be taken into account. The overall very dark appearance of the flat hemin molecules could be related to the molecule's electronic structure and its interaction with the graphene substrate, which could potentially stabilize a charge state and thereby affect the molecule's scattering
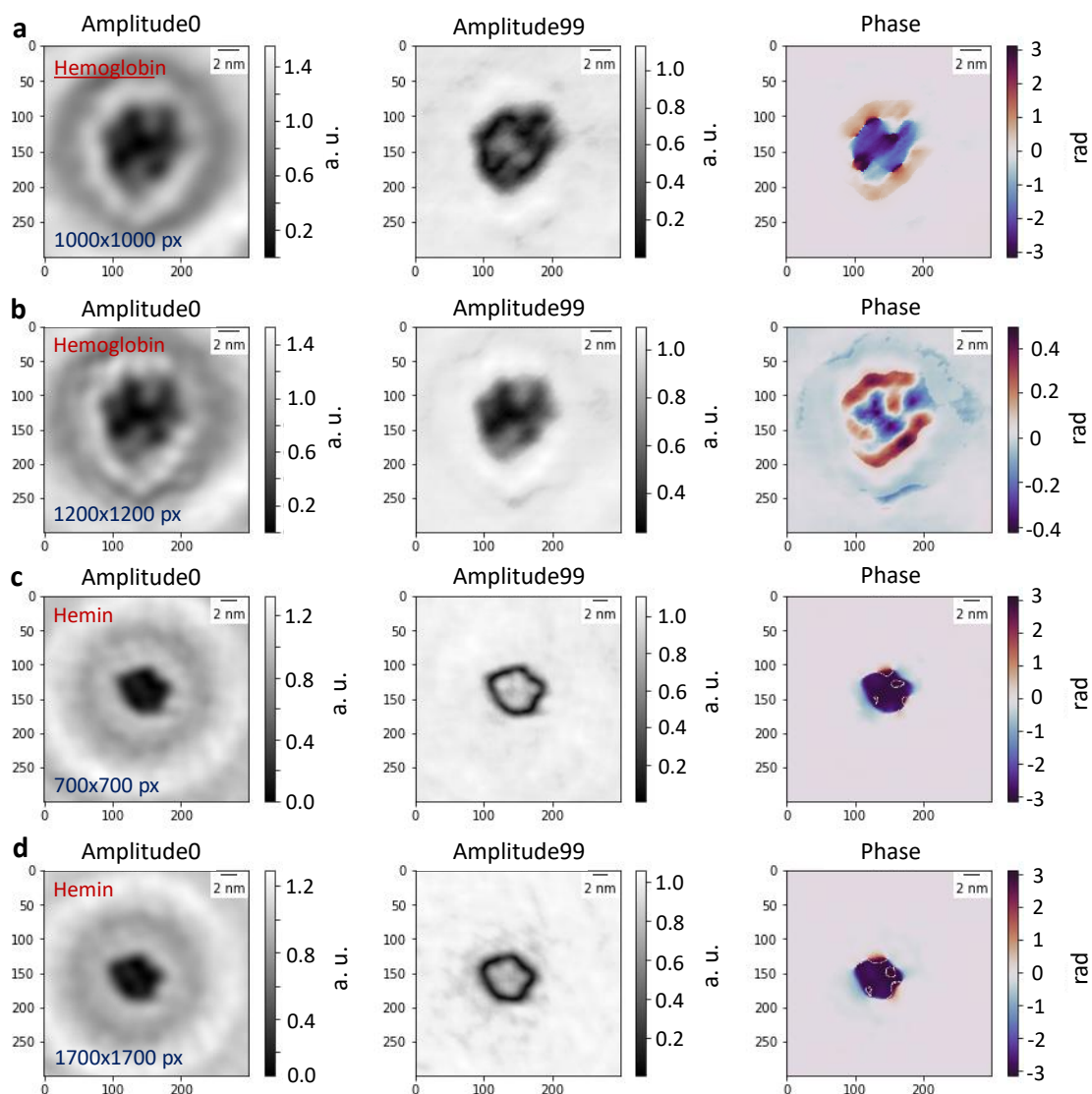
Figure 4.24: **Hollow object artefact: a** Left to right: one-step amplitude reconstruction (Amplitude0), iterative amplitude reconstruction (Amplitude99) after 100 iterations and phase reconstruction after 100 iterations of a hologram of a hemoglobin molecule cropped to a size of 1000×1000 pixels. **b** Left to right: Amplitude0, Amplitude99 and phase reconstruction after 100 iterations of the same hologram as in **a**, cropped to 1200×1200 pixels. **c** Left to right: Amplitude0, Amplitude99 and phase reconstruction after 100 iterations of a hologram of a cluster of Hemin molecules, cropped to 700×700 pixels. **d** Left to right: Amplitude0, Amplitude99 and phase reconstruction after 100 iterations of the same hologram as in **c**, cropped to 1700×1700 pixels.

properties. The presence of a high number of metal atoms within a small molecular volume in PTA could result both in strong scattering or in strong absorption. Since larger molecules only very rarely suffer from hollow object artefacts, it is unclear which physical properties of the smaller proteins could make them more prone to these types of artefact.

In summary, the origin of the hollow object artefact is still unclear. While in some examples, a size-dependence is observed, indicating a possible connection to artefacts observed in simulated examples, there appears to be no size-dependent behaviour in other examples. Furthermore, in the simulated case, both amplitude and phase featured similar size-related types of artefacts, whereas in the hollow object case, mainly the amplitude appears to be affected.

Given that mostly smaller molecules are affected by hollow object artefacts, the appearance of this type of artefact could be related to physical properties of these classes of molecules. However, it might also be possible that the imaging conditions play a role in the occurrence of such artefacts.

Overall, this discussion indicates that a further refinement of the phase retrieval algorithm as well as a better understanding of all the contributions to the amplitude and phase reconstructions are necessary for interpreting the experimental data.

5 Towards three-dimensional reconstruction

Large molecules, such as proteins, are inherently three-dimensional (3D) objects. The correct tertiary structure of a protein, which describes its three-dimensional folding structure, is often crucial for the protein to perform its biological function. Thus, many questions concerning protein structure that are relevant in a biological or biomedical context are related to 3D structural features. A full protein structure determination by LEEH would hence require a reconstruction of the imaged molecule in three dimensions, whereas the algorithms discussed in the previous sections have only provided reconstructions of two-dimensional (2D) planes.

In general, two different approaches towards retrieving 3D information in LEEH are possible: single-hologram and multi-hologram techniques. In single-hologram techniques, the input for the 3D reconstruction is an individual hologram, i. e. the same type of input as used for the reconstruction of two-dimensional images, as discussed in Chapters 1-4. In this case, the reconstruction needs to be improved in such a way that object parts located at different distances along the optical axis are assigned to the correct z-coordinate, while removing the respective out-of-focus contributions from each slice. Alternatively, one can draw upon multiple-hologram imaging of an individual molecule by recording a series of holograms that encode additional structural information, such as for example different orientations of the same molecule, by imaging it from different angles.

In the following, first steps towards a 3D reconstruction from LEEH holograms will be explored by discussing three-dimensional deconvolution, a single-hologram method, and tomography, which uses a series of holograms measured at different tilt angles of the samples as input. While the former method does not require changes to the experimental set-up and workflow, the latter can be implemented experimentally by mounting the sample holder on a rotation piezo, which allows the rotation of the sample around one of the axes in the same plane.

5.1   Three-dimensional deconvolution

In principle, a hologram contains three-dimensional information about the object from which it was generated. A three-dimensional volume can be reconstructed by stacking two-dimensional images, reconstructed at different distances along the optical axis. In simulated objects consisting of several laterally non-overlapping or partially overlapping two-dimensional objects located in different planes along the optical axis, the respective focal planes are distinguishable within such stacks even for small axial separations, as discussed in Fig. 3.21 in the context of multislice hologram simulation. However, as is apparent even in the case of objects with a large axial separation (Fig. 3.21c, d), the out-of-focus parts of the reconstructed wave field significantly contribute to each reconstruction plane.

This is exacerbated for objects with high degrees of lateral overlap and with small axial separation distances; in these cases, distinguishing the respective focal planes becomes more difficult (see Figures 3.22 - 3.24).

While the simulated examples of three-dimensional molecules consisted of object parts confined to two-dimensional slices that are in focus in exactly one plane, a protein has to be considered as a "continuous" object with slice separations on the order of atomic dimensions and a strong lateral overlap of the contributions from each slice. This is reflected by the fact that for reconstructions from experimental holograms, no singular focus plane can be determined, rather, the molecules appear in focus for a range of z_0-values. Thus, separating the contributions from the different parts of the molecule and assigning the correct value distributions to each slice is crucial for reconstructing the 3D properties of the object.

In order to do this, we need to know the impulse response of the LEEH imaging system, which is given by the point spread function (PSF). The point spread function describes the response of the imaging system to a point object, i.e. it maps how a point object is represented as a result of the imaging process. It is defined as the function $\mathrm{PSF}(\mathbf{r})$ that maps a points scatterer, described by a δ-distribution, to the corresponding distribution in the detector plane $U_p(\mathbf{r})$:

$$U_p(\mathbf{r}) = \delta(\mathbf{r}) * \mathrm{PSF}(\mathbf{r}) \equiv \mathrm{PSF}(\mathbf{r}), \qquad (5.1)$$

where * denotes a convolution. Since holography is a two-step process, consisting of hologram generation and reconstruction, the PSF in holography is given by the stack of reconstructed 2D images along the optical axis.

Fig. 5.1 shows the hologram of a point object generated in an in-line holography set-up by plane waves (Fig. 5.1a) and by spherical waves (Fig. 5.1d), along with the corresponding cuts through the reconstructed volume along the optical axis (z-axis), see Fig. 5.1b, e for a wider and c, f for a narrower z-range around the focus, respectively. While the profiles shown in Fig. 5.1b, e differ at z-values away from the focal plane, reflecting the different types of illumination, the z-range around the focus is similar in both cases (Fig. 5.1c, f). The profiles in Fig. 5.1b-c and e-f show that a point scatterer, while well-focused laterally, creates a significant contribution along a range of axial values that by far exceeds the width of the focal plane. The shape of the PSF thus directly limits the resolution in both lateral and axial directions [95].
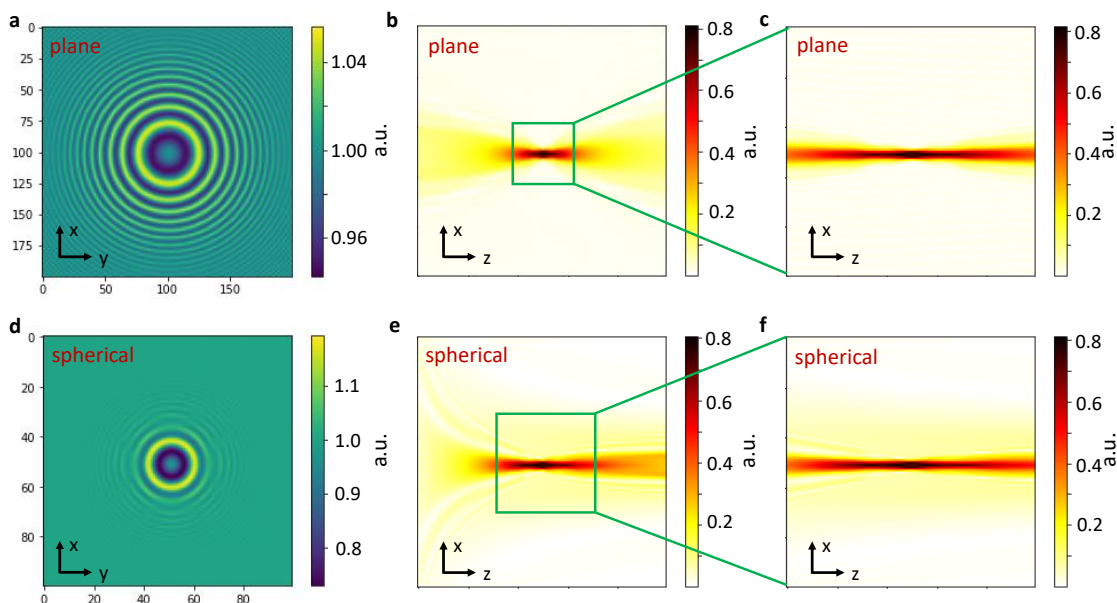
Figure 5.1: **Point spread function in LEEH: a** Hologram of a point scatterer simulated with incident plane waves using an angular spectrum propagator (see section 3.3.1). **b** Cut along the z-axis through a stack of 2D reconstructions obtained from the hologram in **a**. **c** Zoom into the inner region of the profile shown in **b**. **d** Hologram of a point scatterer simulated with incident spherical waves using a Fresnel-Kirchhoff propagator (see section 1.1.2). **e** Cut along the z-axis through a stack of 2D reconstructions obtained from the hologram in **d**. **f** Zoom into the inner region of the profile shown in **e**.

In analogy to the definition of the point spread function in eq. (5.1), the imaging of an object can also be described by a convolution with the point spread function of an imaging system. Assuming that an object $O(\mathbf{r})$ can be described as a sum of point scatterers (condition of linearity),

$$O(\mathbf{r}) = \int O(\mathbf{s})\delta(\mathbf{r} - \mathbf{s})\mathrm{d}\mathbf{s}, \qquad (5.2)$$

i.e. when multiple scattering can be neglected, and that each individual scatterer is treated in the same way during the imaging process (condition of shift invariance), the output of the imaging process can be described as [211, 245]

$$U_o(\mathbf{r}) = \int O(\mathbf{s})\mathrm{PSF}(\mathbf{r} - \mathbf{s})\mathrm{d}\mathbf{s}, \qquad (5.3)$$

which takes the form of a convolution

$$U_o(\mathbf{r}) = O(\mathbf{r}) * \mathrm{PSF}(\mathbf{r}) = O(\mathbf{r}) * U_p(\mathbf{r}). \qquad (5.4)$$

Thus, a deconvolution of the image output with the PSF yields the object distribution. Deconvolution is an established technique in optical microscopy [211, 246, 247]. In the following, three-dimensional deconvolution in an in-line holography set-up will be discussed based on the algorithms proposed in refs. [211, 212].

The most obvious way in which eq. (5.4) can be solved for the object distribution is to rewrite it as a product of Fourier transforms by applying the convolution theorem (see section 1.1.2):

$$\mathcal{F}(U_o(\mathbf{r})) = \mathcal{F}(O(\mathbf{r}))\mathcal{F}(U_p(\mathbf{r})), \tag{5.5}$$

where $\mathcal{F}$ denotes the three-dimensional Fourier transform.

This yields an expression for the three-dimensional object distribution:

$$O(\mathbf{r}) = \mathcal{F}^{-1}\left(\frac{\mathcal{F}(U_o(\mathbf{r}))}{\mathcal{F}(U_p(\mathbf{r}))}\right). \tag{5.6}$$

When inserting the appropriate expressions for the Fourier transforms of the complex wave fields $U_o(\mathbf{r})$ and $U_p(\mathbf{r})$ into eq. (5.6), however, the z-dependent components will cancel out in the direct division [211]. Thus eq. (5.6) cannot directly be used to retrieve the three-dimensional object distribution.

This problem can be avoided in two ways. If one considers the intensities $|U_o(\mathbf{r})|^2$ and $|U_p(\mathbf{r})|^2$ instead of the complex wave fields $U_o(\mathbf{r})$ and $U_p(\mathbf{r})$ as the measured signal, a direct division approach as in eq. (5.6) can be retained without loss of the z-dependent components. The object distribution is then calculated as

$$O(\mathbf{r}) = \mathcal{F}^{-1}\left(\frac{\mathcal{F}(|U_o(\mathbf{r})|^2)}{\mathcal{F}(|U_p(\mathbf{r})|^2) + \beta}\right), \tag{5.7}$$

where β is a small constant to avoid division by zero.

Because intensities rather than complex wave fields are used, however, the linearity condition, upon which the derivation of the convolution formulation was based, is only fulfilled for a sample consisting of isolated, point-like objects. In such cases, the intensity-based direct deconvolution approach is applicable in the context of in-line holography, as demonstrated in ref. [211] for incident plane waves.

When considering extended objects, eq. (5.4) needs to be solved iteratively. Since they operate on the level of complex wave fields, iterative deconvolution methods are applicable to a wide range of objects [212]. Because we are ultimately interested in applying deconvolution methods to holograms of extended molecules such as proteins, we will focus on the iterative deconvolution methods here.

Given that iterative deconvolution methods are widespread in optical microscopy, the iterative steps of one such method, the Gold algorithm [248], can be adapted for holography [211, 212]. The Gold algorithm consists of the following iterative steps:

The object distribution is initialized by the output distribution of the imaging process,

$$O^{(0)}(\mathbf{r}) = U_o(\mathbf{r}), \tag{5.8}$$

where $U_o(\mathbf{r})$ is the initially reconstructed stack of 2D images from the hologram. In the nth iteration step, $U_o^{(n)}(\mathbf{r})$ is calculated as the convolution of the nth object distribution $O^{(n)}(\mathbf{r})$ and the point spread function $U_p(\mathbf{r})$,

$$U_o^{(n)}(\mathbf{r}) = O^{(n)}(\mathbf{r}) * U_p(\mathbf{r}). \tag{5.9}$$

Subsequently, the updated object distribution $O^{(n+1)}(\mathbf{r})$ is calculated as

$$O^{(n+1)}(\mathbf{r}) = O^{(n)}(\mathbf{r}) \frac{U_o(\mathbf{r})}{U_o^{(n)}(\mathbf{r})}. \tag{5.10}$$

The algorithm can hence be summarized as

$$O^{(0)}(\mathbf{r}) = U_o(\mathbf{r}) \tag{5.11}$$

$$\text{(i)} \qquad U_o^{(n)}(\mathbf{r}) = O^{(n)}(\mathbf{r}) * U_p(\mathbf{r}) \tag{5.12}$$

$$\text{(ii)} \quad O^{(n+1)}(\mathbf{r}) = O^{(n)}(\mathbf{r}) \frac{U_o(\mathbf{r})}{U_o^{(n)}(\mathbf{r})} \tag{5.13}$$

$$\text{(iii)} \qquad n = n + 1 \tag{5.14}$$

The convolution in (i) can be evaluated via the convolution theorem. The division in (ii) can be calculated in the same way as the division step in the intensity-based deconvolution [211, 212]:

$$O^{(n+1)}(\mathbf{r}) = O^{(n)}(\mathbf{r}) \frac{U_o(\mathbf{r}) \left(U_o^{(n)}(\mathbf{r}) \right)^*}{|U_o^{(n)}(\mathbf{r})|^2 + \beta}, \tag{5.15}$$

where * denotes the complex conjugate and $\beta \ll |U_o^{(n)}(\mathbf{r})|^2$.

Additionally, to ensure that $U_o^{(n)}(\mathbf{r})$ has the same value range as $U_o(\mathbf{r})$, $U_o^{(n)}(\mathbf{r})$ has to be normalized after step (i) [212]:

$$U_o^{(n)}(\mathbf{r}) = \frac{\left(U_o^{(n)}(\mathbf{r}) - \min\left[U_o^{(n)}(\mathbf{r}) \right] \right) \left(\max\left[U_o(\mathbf{r}) \right] - \min\left[U_o(\mathbf{r}) \right] \right)}{\left(\max\left[U_o^{(n)}(\mathbf{r}) \right] - \min\left[U_o^{(n)}(\mathbf{r}) \right] \right)} + \min\left[U_o(\mathbf{r}) \right]$$

$$\tag{5.16}$$

In the following, the performance of the algorithm will be tested by applying it to a range of different objects. To facilitate a comparison to literature, the discussion will mostly focus on simulation and reconstruction with plane waves.

Fig. 5.2a shows an object built up from four point scatterers in different planes along the z-axis with a z-separation of $10\,\text{mm}$ between the focus planes and a distance of $70\,\text{mm}$ between the detector and the sample plane closest to the detector. This geometry reproduces the geometry of the examples of three-dimensional deconvolution discussed in ref. [211]. The point scatterers are centred at the pixels $(75, 75)$, $(75, 125)$, $(125, 75)$, and $(125, 125)$ in the $x - y$-plane; each point scatterer is simulated as a central pixel of intensity 1, surrounded by two rings of intensity 0.75 and 0.25, respectively. The hologram has been simulated from the four slices via the multislice algorithm presented in section 3.3.1 with plane waves of wavelength $500\,\text{nm}$.

Fig. 5.2b depicts the one-step 2D reconstructions from the hologram in the respective focus planes of the point scatterers, labelled by the respective distance to the detector. The $x - y$-location of the reconstructed objects is correctly retrieved and their shape matches that of the simulated point object. Each scatterer appears in focus in the correct
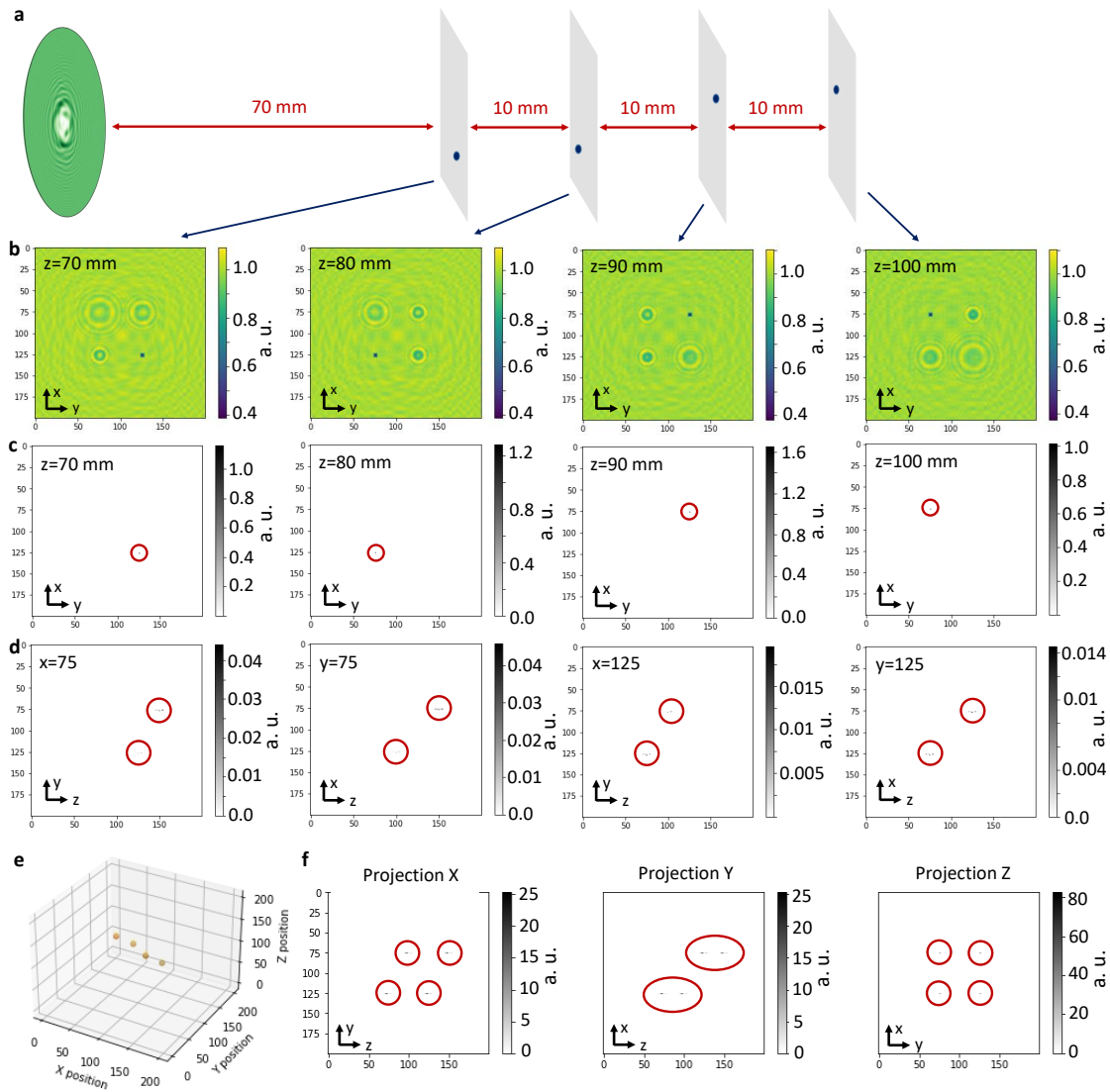
Figure 5.2: **Iterative deconvolution: point objects a** Sketch of the object, consisting of four point objects located in four different planes along the z-axis, separated by $10\,\mathrm{mm}$, respectively. **b** 2D reconstructions (U_o) in the respective z-planes. **c** Reconstructions ($O^{(n)}$) of the z-planes in which the objects are located after iterative deconvolution (25 iterations). **d** $x - z$ and $y - z$ cuts through the corresponding object planes. **e** 3D plot of the reconstructed object $O^{(n)}$ after deconvolution. **f** Projections along the x-, y- and z-axes of the 3D plot shown in **e**. For better visibility, the reconstructed signals are highlighted by red circles.

plane. However, out-of-focus contributions from the other point scatterers are present in each of the images. Fig. 5.2c shows the corresponding 2D slices through the focal planes of the point scatterers after 25 iterations of the three-dimensional deconvolution algorithm. The point scatterers still appear in the correct $x - y$-positions, but their shape differs from the shape reconstructed in Fig. 5.2b, they only appear as a single pixel. Contributions from the other scatterers have been removed by the deconvolution step. The effectiveness of the iterative deconvolution in removing out-of-focus contributions

along the z-axis is also demonstrated in Fig. 5.2d, which shows the $x - z$ and $y - z$-profiles through the x, $y = 75$ and x, $y = 125$ planes. While the point scatterers are not reduced to a single point in the axial direction, they only contribute to a few pixels around their z-location as opposed to the contribution over the full z-range shown in the PSF in Fig. 5.1b.

Fig. 5.2e shows a 3D representation of the reconstructed object $O^{(n)}$, the corresponding projections of the 3D plot along the x-, y, and z-axes are shown in Fig. 5.2f, summarizing the results of Fig. 5.2c-d.
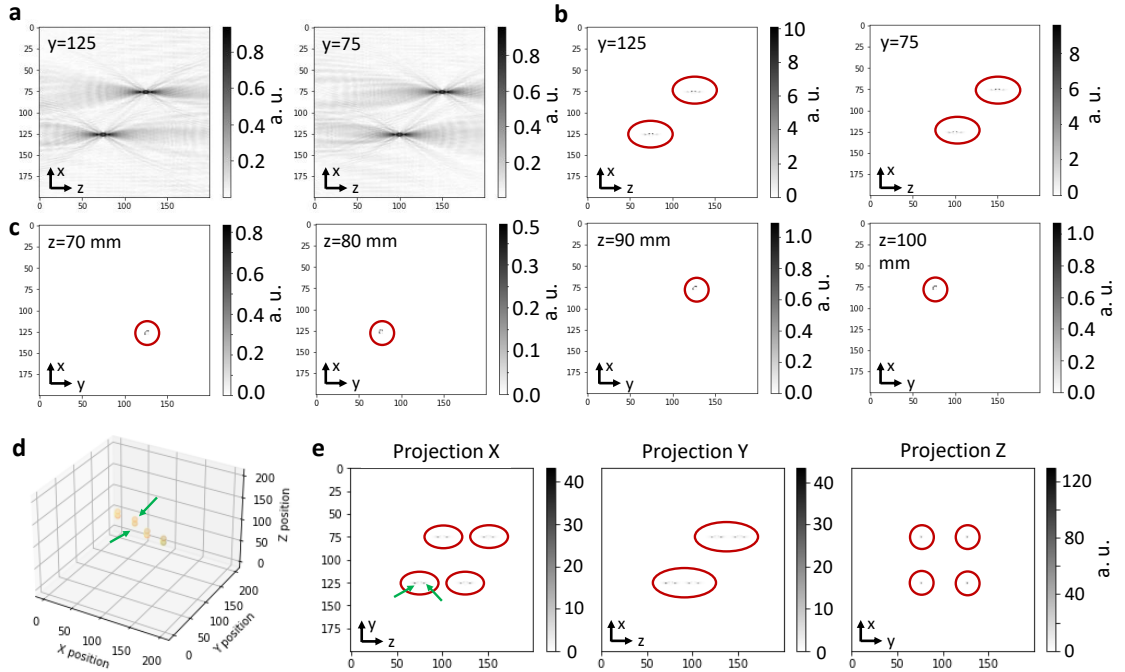

Figure 5.3: **Iterative deconvolution: small disks** Object of the same geometry as in Fig. 5.2 with small disks of radius 3 pixels instead of point objects. **a** $x - z$-cuts through the reconstructed object volume before deconvolution at $y = 125$ and $y = 75$ pixels. **b** $x - z$-cuts through the reconstructed object volume after deconvolution (25 iterations) at $y = 125$ and $y = 75$ pixels. **c** Reconstructions ($O^{(n)}$) of the z-planes in which the objects are located after iterative deconvolution (25 iterations). **d** 3D plot of the reconstructed object $O^{(n)}$ after deconvolution. **e** Projections along the x-, y- and z-axes of the 3D plot shown in **d**. For better visibility, the reconstructed signals are highlighted by red circles.

Fig. 5.3 shows the analogous object to Fig. 5.2 with small extended objects (disks of radius 3 pixels) instead of point objects. Fig. 5.3a and b depict the $x - z$-profiles through the objects before (Fig. 5.3a) and after (Fig. 5.3b) 25 iterations of the iterative deconvolution algorithm. The deconvolution step reduces the z-extension significantly. In the corresponding z-focus slices for each of the four objects (Fig. 5.3c), only the in-focus object can be seen, contributions from the other planes have been removed. The shape of the disks, however, is not fully recovered, the strongest part of the signal appears to come from the edge of the objects. In the 3D plot of the iterative reconstruction (Fig. 5.3d), each of the objects appears as a double sphere, which is also reflected in the

corresponding projections in Fig. 5.3e (marked by green arrows), the double signal could be associated with the objects' sharp edges as opposed to the smooth edges of the point objects discussed in Fig. 5.2.

This double signal appears to be a general feature of the reconstruction of extended objects with sharp edges by iterative three-dimensional deconvolution.

In order to discuss the case of extended objects in more detail, Fig. 5.4 shows the results of the reconstruction of a hologram of a single disk of radius 5 pixels, located in the plane $z = 80$ mm.

Fig. 5.4a and b show the $x - z$-profile through the object along the optical axis and the reconstruction in the focus plane, respectively. The reconstruction of the object plane contains some contributions from the twin image. Fig. 5.4c, d shows the corresponding images after 10 iterations of the deconvolution algorithm. The $x - z$-profile is reduced to its inner part. In the focal plane (Fig. 5.4d), the twin image contributions are eliminated, but the object shape is not recovered as well as in Fig. 5.4b. The strongest signal appears to come from the edge of the object. It is also interesting to note the difference in the amplitude value range between Fig. 5.4a, b and Fig. 5.4 c, d, respectively. The higher maximum amplitudes at the object edge in Fig. 5.4d, which might be due to an amplification of errors at the object edge during the iteration steps, could explain the loss of contrast in the inner part of the object; the amplitude values in the inner part of the object in Fig. 5.4d seem to match those in Fig. 5.4b.

The double-signal structure already observed in Fig. 5.3 reappears in Fig. 5.4e. The larger separation of the signals as compared to Fig. 5.3d supports the attribution of this phenomenon to the object's edges since the object in Fig. 5.4 has a larger radius than the objects in Fig. 5.3.

Fig. 5.4g-l shows that the situation can to some extent be remedied by the application of spatial filters that set values outside a masked region in the inner part of the 3D volume to zero. The filter is applied to the reconstructed object $O^{(n)}$ at the end of each iteration step. By choosing a tight mask along the axial direction, the double-signal structure can be partially suppressed (Fig. 5.4k, l). Strong contributions from the object's edge remain, however, as can be seen in the $x - z$-profile and the reconstruction of the focal plane in Fig. 5.4i and j, respectively.

Fig. 5.4 shows that the iterative deconvolution algorithm can in general handle extended objects and that edge artefacts can be to some degree eliminated with with spatial filters. In Fig. 5.5, a different type of artefact is discussed, namely a strong contribution along the optical axis. The object depicted in Fig. 5.5 is a ring with inner radius 10 pixels and outer radius 15 pixels, the $x - z$-profile and the focal plane of the reconstruction before deconvolution are shown in Fig. 5.5a, b. As demonstrated in Fig. 5.5d, the object itself is accurately recovered in its focal plane, although there again appear to be excessive contributions from the object's edge. In the $x - z$-profile of the object after deconvolution (Fig. 5.5c), however, contributions along the optical axis away from the object's focal plane dominate. This is also reflected in the 3D plot in Fig. 5.5e as well as in the corresponding projections in Fig. 5.5f, where the dominance of this artefact in the z-projection renders the ring itself almost invisible. Since these strong axial contributions do not coincide with the location of the object, they can be removed by applying a spatial filter, which sets the values away from the centre of the 3D volume to zero in each iteration step (Fig. 5.5g, h.). Alternatively, smoothing the reconstruction with a
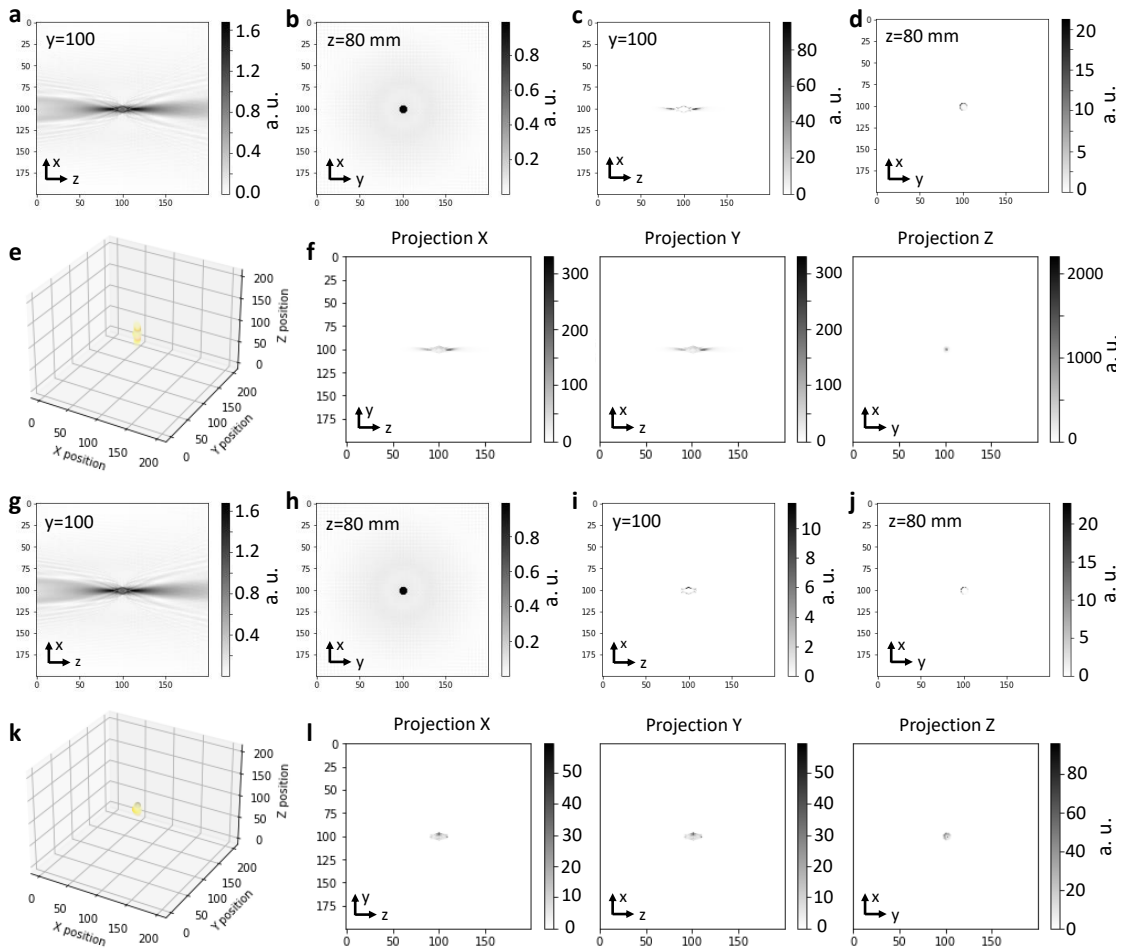
Figure 5.4: **Iterative deconvolution: extended objects** Reconstruction by iterative deconvolution of the hologram of a disk of radius 5 pixels located at $z = 80\,\text{mm}$ with (**g-l**) and without (**a-f**) the application of spatial filters. **a** $x - z$-cut through the reconstructed object volume before deconvolution at $y = 100\,\text{pixels}$. **b** 2D reconstruction (U_o) of the object in the plane corresponding to $z = 80\,\text{mm}$. **c** $x - z$-cut through the reconstructed object volume after deconvolution (10 iterations) at $y = 100\,\text{pixels}$. **d** Reconstruction of the object ($O^{(n)}$) after deconvolution (10 iterations) in the plane corresponding to $z = 80\,\text{mm}$. **e** 3D plot of the reconstructed object $O^{(n)}$ after deconvolution. **f** Projections along the x-, y- and z-axes of the 3D plot shown in **e**. **g** $x - z$-cut through the reconstructed object volume before deconvolution at $y = 100\,\text{pixels}$. **h** 2D reconstruction (U_o) of the object in the plane corresponding to $z = 80\,\text{mm}$. **i** $x - z$-cut through the reconstructed object volume after deconvolution (10 iterations) with the application of a spatial filter in each iteration, at $y = 100\,\text{pixels}$. **j** Reconstruction of the object ($O^{(n)}$) in the plane corresponding to $z = 80\,\text{mm}$ after deconvolution (10 iterations) and spatial filtering in each iteration. **k** 3D plot of the reconstructed object $O^{(n)}$ after deconvolution with spatial filtering. **l** Projections along the x-, y- and z-axes of the 3D plot shown in **k**.

Gaussian filter during the iterative process, as suggested in refs. [211, 212], could help suppress these types of artefacts.

Finally, while Fig. 5.4 and Fig. 5.5 involve laterally extended objects, it is of interest to study extended objects that include an extension along the axial direction. Fig. 5.6

shows two disks of radius 5 pixels located in the z-planes $z = 70\,\text{mm}$ and $z = 80\,\text{mm}$. Laterally, they are shifted by 10 pixels in both the x- and y-direction, respectively, hence they do not overlap.
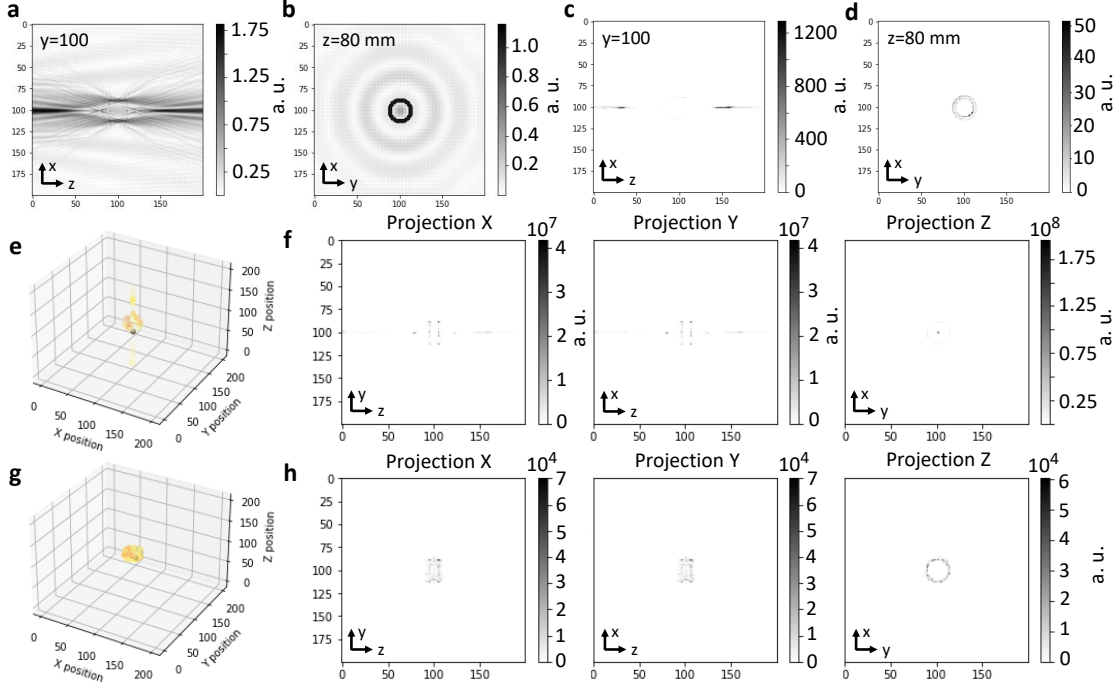

Figure 5.5: **Iterative deconvolution: extended objects II** Reconstruction by iterative deconvolution of the hologram of a ring of outer radius 15 pixels and inner radius 10 pixels located at $z = 80\,\text{mm}$ with (**g-h**) and without (**e-f**) the application of spatial filters. **a** $x - z$-cut through the reconstructed object volume before deconvolution at $y = 100$ pixels. **b** 2D reconstruction (U_o) of the object in the plane corresponding to $z = 80\,\text{mm}$. **c** $x - z$-cut through the reconstructed object volume after deconvolution (25 iterations) at $y = 100$ pixels. **d** Reconstruction of the object ($O^{(n)}$) after deconvolution (25 iterations) in the plane corresponding to $z = 80\,\text{mm}$. **e** 3D plot of the reconstructed object $O^{(n)}$ after deconvolution (25 iterations) without the application of spatial filters. **f** Projections along the x-, y- and z-axes of the 3D plot shown in **e**. **g** 3D plot of the reconstructed object $O^{(n)}$ after deconvolution (25 iterations) with the application of spatial filters. The artefacts along the optical axis are removed. **h** Projections along the x-, y- and z-axes of the 3D plot shown in **g**.

While the z-extent is significantly reduced, as in the previous examples (Fig. 5.6a, b), a contribution of the out-of-focus object remains after the deconvolution (Fig. 5.6d), although only by the object itself, other out-of-focus artefacts, such as fringes, are removed when compared to the reconstruction before the deconvolution (Fig. 5.6c). In the 3D plot and the corresponding projections (Fig. 5.6e, f), both objects appear as a double-signal like the single disk in Fig. 5.4. Increasing the number of iterations from 10 to 25 seems to increase the resolution in z, the two signals corresponding to each disks are narrower and better separated, as shown in Fig. 5.6g, h. However, the lateral dimensions also seem to decrease, as can be seen when comparing the z-projections in Fig. 5.6f and h. This apparent decrease could be due to the extreme increase in maximum amplitude

166

value which could induce a change in contrast rendering the smaller amplitude values indiscernible from the background. Thus, it appears that amplitude spikes can occur during deconvolution and can be aggravated by a higher number of iterations.
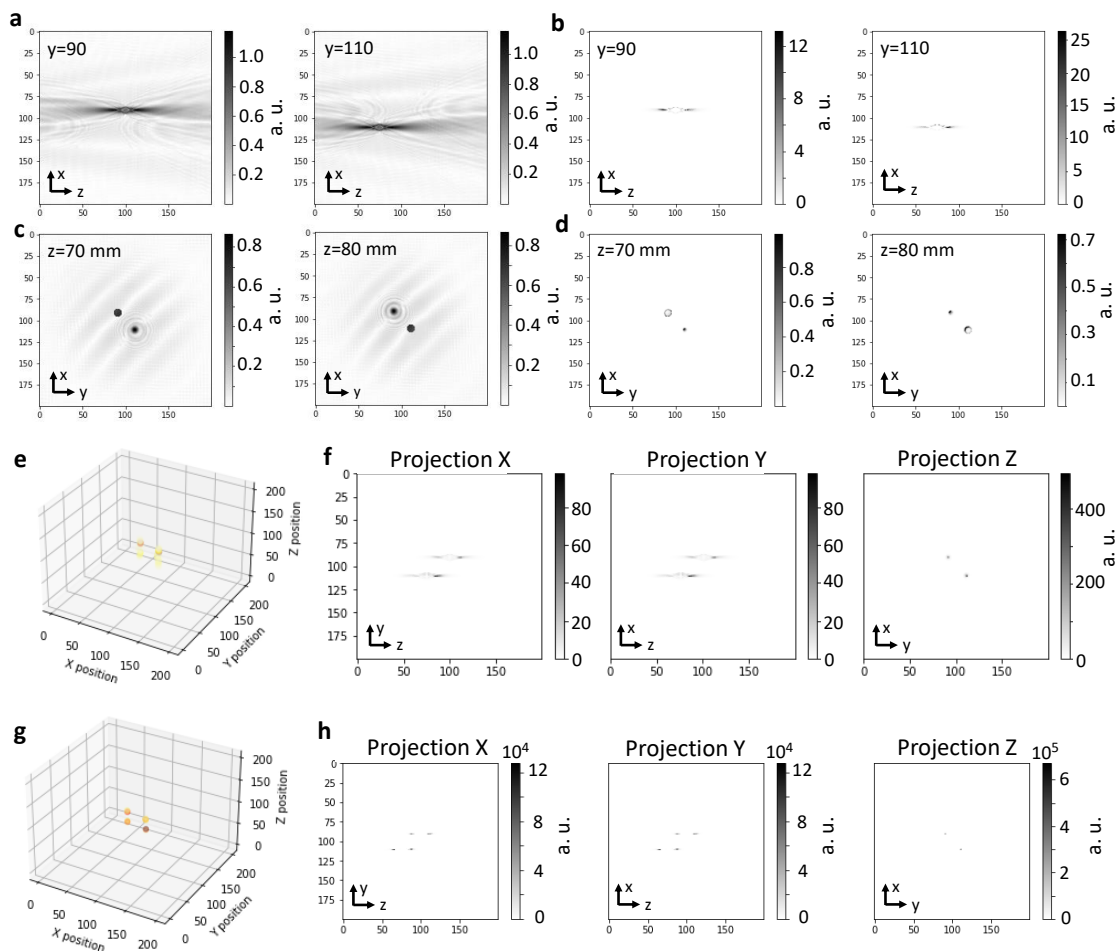

Figure 5.6: **Iterative deconvolution: extended objects III** Reconstruction by iterative deconvolution of the hologram of two disks of radius 5 pixels located at $z = 70\,\text{mm}$ and $z = 80\,\text{mm}$. **a** $x - z$-cuts through the reconstructed object volume before deconvolution at $y = 90$ and $y = 110$ pixels. **b** $x - z$-cuts through the reconstructed object volume after deconvolution (10 iterations) at $y = 90$ and $y = 110$ pixels. **c** 2D reconstructions (U_o) in the respective focal planes of the two disks. **d** Reconstructions ($O^{(n)}$) after iterative deconvolution (10 iterations) of the z-planes in which the objects are located. **e** 3D plot of the reconstructed object $O^{(n)}$ after deconvolution (10 iterations). **f** Projections along the x-, y- and z-axes of the 3D plot shown in **e**. **g** 3D plot of the reconstructed object $O^{(n)}$ after deconvolution (25 iterations). **h** Projections along the x-, y- and z-axes of the 3D plot shown in **g**.

A similar dependence on the number of iterations is also apparent in the case of two laterally overlapping disks as shown in Fig. 5.7. The disks are identical to the ones in Fig. 5.6, as are their locations in the axial direction, but they have been moved closer to the centre of the image to allow a degree of overlap. When performing the iterative deconvolution with 10 iterations, the z-planes of the two objects can be distinguished

167

(Fig. 5.7d), even though the overlap region provides the strongest signal. After 25 iterations, however, only the overlap region is visible (Fig. 5.7g).
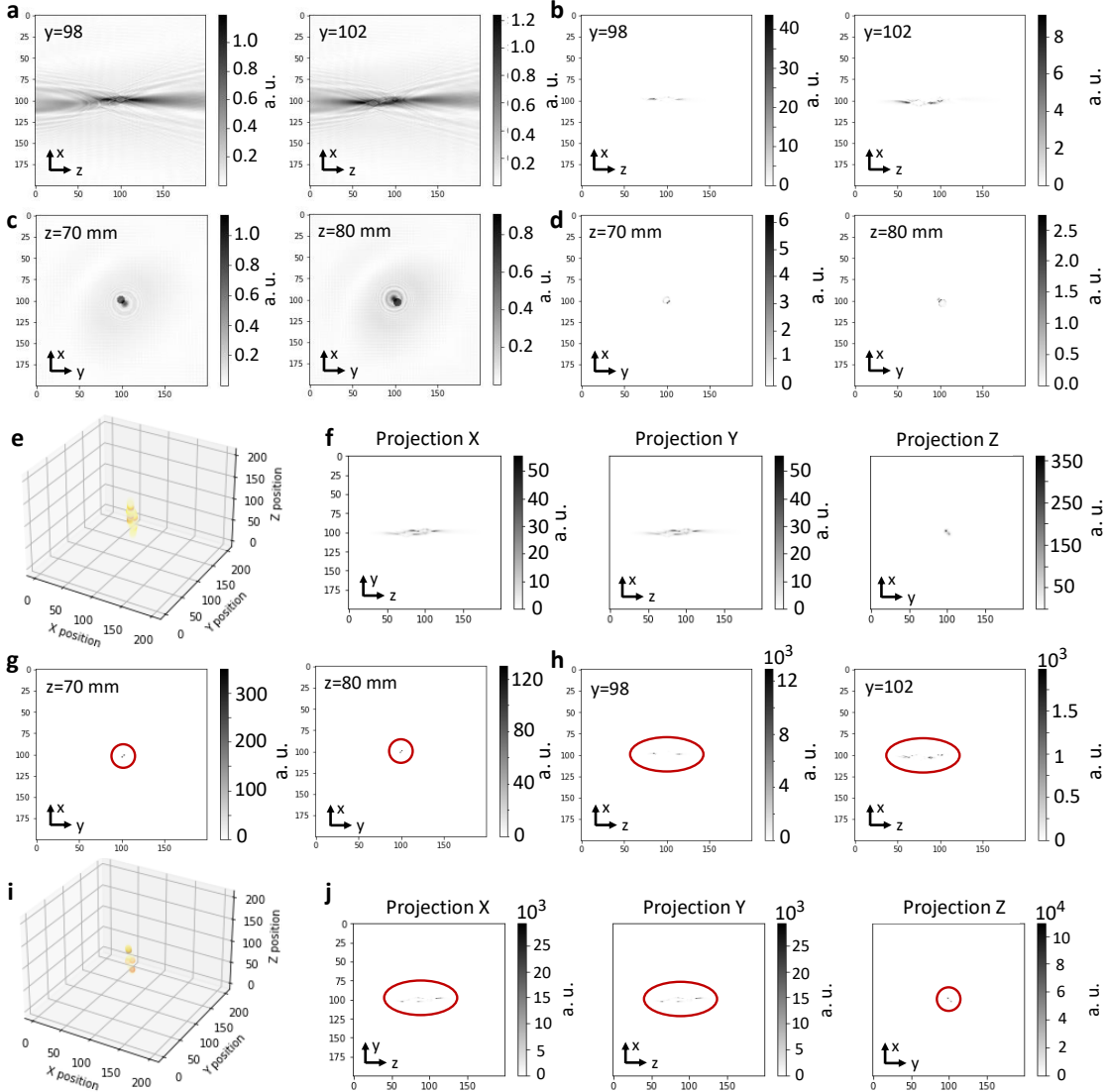

Figure 5.7: **Iterative deconvolution: overlapping objects** Reconstruction by iterative deconvolution of the hologram of two disks of radius 5 pixels located at $z = 70\,\text{mm}$ and $z = 80\,\text{mm}$ with slight lateral overlap. **a** $x - z$-cuts through the reconstructed object volume before deconvolution at $y = 98$ and $y = 102$ pixels. **b** $x - z$-cuts through the reconstructed object volume after deconvolution (10 iterations) at $y = 98$ and $y = 102$ pixels. **c** 2D reconstructions (U_o) in the respective focal planes of the two disks. **d** Reconstructions ($O^{(n)}$) after iterative deconvolution (10 iterations) of the z-planes in which the objects are located. **e** 3D plot of the reconstructed object $O^{(n)}$ after deconvolution (10 iterations). **f** Projections along the x-, y- and z-axes of the 3D plot shown in **e**. **g** Reconstructions ($O^{(n)}$) after iterative deconvolution (25 iterations) of the z-planes in which the objects are located. **h** $x - z$-cuts through the reconstructed object volume after deconvolution (25 iterations) at $y = 98$ and $y = 102$ pixels. **i** 3D plot of the reconstructed object $O^{(n)}$ after deconvolution (25 iterations). **j** Projections along the x-, y- and z-axes of the 3D plot shown in **i**.

Since large molecules, such as proteins, in general feature a high degree of lateral overlap, it is of interest to study two objects in different z-planes that fully overlap laterally. Fig. 5.8 shows the example of two disks of radius 5 pixels, located at $z = 70$ mm and $z = 80$ mm while fully overlapping laterally. While the $x - z$-profile after deconvolution shows two signals that can be attributed to the two disks, the signal is the strongest in between the two contributions from the disks (Fig. 5.8c). This is also reflected by the darker part in the centre of the 3D plot in Fig. 5.8e. Hence, although the two reconstructed z-planes show one disk each, it is hard to tell whether this is in fact only the contribution of the respective focal plane. The results from partially overlapping objects in Fig. 5.7 seem to suggest that a contribution from the out-of-focus plane remains.
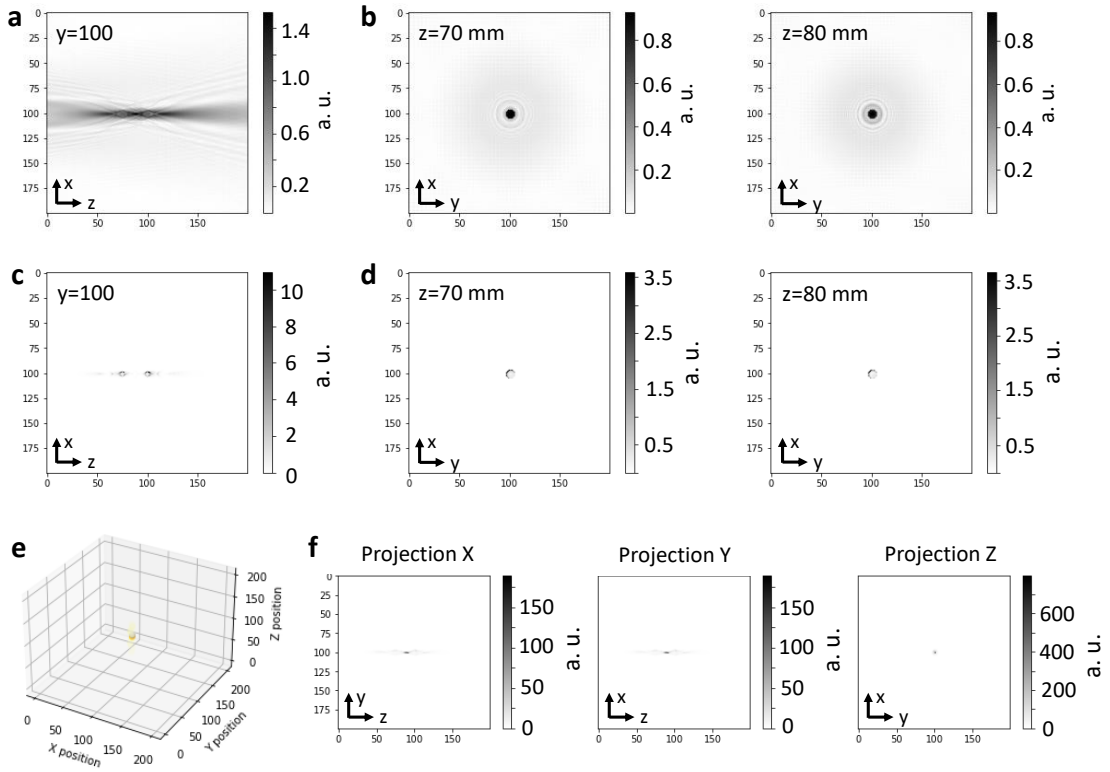

Figure 5.8: **Iterative deconvolution: overlapping objects II** Reconstruction by iterative deconvolution of the hologram of two disks of radius 5 pixels located at $z = 70$ mm and $z = 80$ mm with full lateral overlap. **a** $x - z$-cut through the reconstructed object volume before deconvolution at $y = 100$ pixels. **b** 2D reconstructions (U_o) in the respective focal planes of the two disks. **c** $x - z$-cut through the reconstructed object volume after deconvolution (10 iterations) at $y = 100$ pixels. **d** Reconstructions ($O^{(n)}$) after iterative deconvolution (10 iterations) of the z-planes in which the objects are located. **e** 3D plot of the reconstructed object $O^{(n)}$ after deconvolution (10 iterations). **f** Projections along the x-, y- and z-axes of the 3D plot shown in **e**.

The previous examples have all been simulated and reconstructed using a plane wave propagator for simplicity and to facilitate the comparison to the existing literature [211, 207]. The experimental set-up, however, operates with incident spherical waves. Hence, in order to assess whether the algorithm is applicable to experimentally acquired

holograms, the performance of the algorithm for spherical incident waves has to be evaluated.

Fig. 5.9 shows an example analogous to the one in Fig. 5.7, with two partially overlapping disks of radius 3 pixels. The parameters have been chosen such that they map a realistic imaging situation for our experimental setup: a source-to-sample distance of 290 nm and 300 nm for the respective object planes, a sample-to-detector distance of 11.5 cm, an object size of 70 nm, corresponding to the side length of the slice, and a wavelength of 1 Å.

The general performance of the algorithm is very similar when comparing between input generated from incident plane and spherical waves. The differences in performance can likely be attributed to the fact that the objects are closer together axially in the spherical waves example. In both cases, the extent of the axial contributions is reduced significantly by the deconvolution process (Fig. 5.9b). The z-planes in which the two objects are in focus are identifiable (Fig. 5.9d), however, the object's shape is not fully recovered. As in the plane wave examples, the object's edge appears to contribute disproportionately, although the double-signal structure is less pronounced in the spherical case.
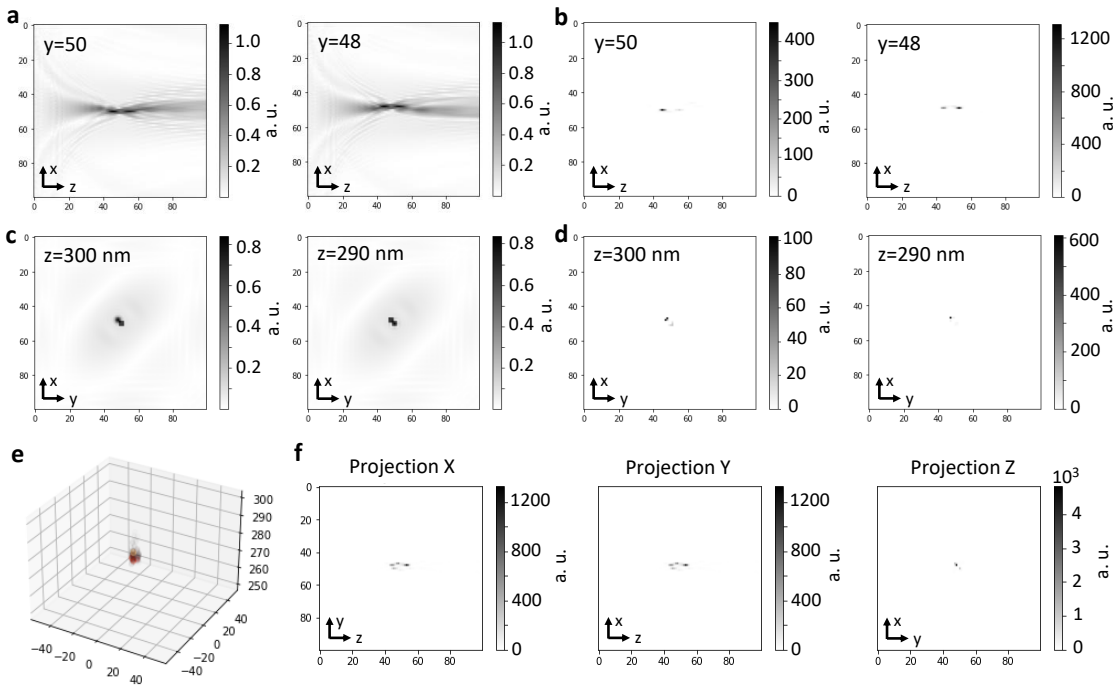

Figure 5.9: **Iterative deconvolution: spherical waves** Reconstruction by iterative deconvolution of the hologram of two disks of radius 3 pixels located at $z = 290$ nm and $z = 300$ nm with partial lateral overlap. The hologram has been simulated and reconstructed using a Fresnel-Kirchhoff propagator for spherical waves. **a** $x - z$-cuts through the reconstructed object volume before deconvolution at $y = 50$ and $y = 48$ pixels. **b** $x - z$-cuts through the reconstructed object volume after deconvolution (7 iterations) at $y = 50$ and $y = 48$ pixels. **c** 2D reconstructions (U_o) in the respective focal planes of the two disks. **d** Reconstructions ($O^{(n)}$) after iterative deconvolution (7 iterations) of the z-planes in which the objects are located. **e** 3D plot of the reconstructed object $O^{(n)}$ after deconvolution (7 iterations). **f** Projections along the x-, y- and z-axes of the 3D plot shown in **e**.

While this discussion of three-dimensional deconvolution as a method for retrieving 3D information from single holograms is by no means exhaustive, some conclusions can be drawn. The iterative three-dimensional deconvolution algorithm accurately retrieves the z-positions of isolated objects, such as point objects and small disks, contributions from other object planes are removed. The recovery of the correct lateral shapes is, however, less accurate, evidence from the simulations suggests that this could be due to strong edge artefacts that appear during the iterative process and can be exacerbated by larger numbers of iterations. It has been demonstrated that the addition of filters during the iteration can to some degree counteract the appearance of these artefacts. Comparison to literature [211, 212] suggests that optimizing filters and the number of iterations can further improve the reconstruction and reduce artefacts.

The z-position of laterally overlapping objects is more difficult to recover. Thus, it appears that three-dimensional deconvolution might not be the method of choice for the 3D-reconstruction of large, extended objects such as proteins. One reason for this could be that the condition of linearity that describes the independence of the scatterers and was one of the assumptions from which the deconvolution algorithm was derived, might not be valid in the case of proteins given that multiple scattering processes are likely to occur. Furthermore, the algorithm, as presented here, does not include phase information. To include this, the algorithm would have to be adapted, possibly by first reconstructing the complex wave field along the z-axis, as described in Chapter 3, and then using this as input for the deconvolution step.

Additionally, deconvolution is computationally costly, especially when using high-resolution data, i. e. a large number of pixels, which is necessary for experimental data.

In practice, three-dimensional deconvolution appears to be more suitable in experimental contexts involving small, isolated objects, such as in particle tracking [249], than in the 3D-reconstruction of large, extended molecules.

5.2   Outlook: towards LEEH tomography

The previous considerations suggest that the information stored in a single hologram is not sufficient for a full 3D reconstruction of an extended molecule such as a protein. To overcome this, the data from several holograms encoding different spatial information has to be combined.

A widely applied technique for retrieving three-dimensional information is tomography, which extracts the three-dimensional structure of an object from a series of projections obtained by imaging the object at different angles [250, 251]. In holography, this can be implemented by rotating the sample around an axis, producing a tilt series.

A projection can be defined as a set of integrated values of an object parameter, such as density or beam attenuation, along a straight line through the object [251]. Mathematically, this can be represented by the line integral

$$P_\theta = \int_{l(\theta,t)} o(x,y)\mathrm{d}s, \tag{5.17}$$

where $o(x,y)$ denotes the object distribution and the integral is evaluated along the line $l(\theta,t)$ defined by the line equation $x\cos\theta + y\sin\theta = t$. The integral given in eq. (5.17) is referred to as the Radon transform [252].

Most commonly, a projection is generated from a set of line integrals along parallel rays (parallel beam geometry). This case is also the most straightforward regarding the reconstruction of the object distributions from the projections since the Radon transform can be directly related to slices in two-dimensional Fourier space. This relationship is described by the Fourier Slice theorem [251], which states that the $n-1$-dimensional Fourier transform of a parallel projection $P_{\theta(t)}$ of an n-dimensional object corresponds to a $(n-1)$-dimensional slice through the n-dimensional Fourier transform of the object distribution defined by the angle θ with the Fourier coordinate axes. This relation allows the Fourier transform of the object to be generated slice by slice from the measured projections, which in turn allows the retrieval of the object distribution by inverse Fourier transform. This conceptually simple reconstruction scheme is, however, met with difficulties in the numerical implementation since the Fourier slices are defined in a polar coordinate system, while the Fast Fourier transform (FFT) used for numerical evaluation is defined in a rectangular coordinate system. To circumvent the problems caused by that, the Fourier space contributions can be weighted by the introduction of a frequency filter in the Fourier domain before the inversion step. This reconstruction routine is called *filtered backprojection* and is very commonly employed in tomographic reconstruction [251, 253, 254] and can also be employed in conjuction with holographic tomography [255].

In the case of low-energy electron holography with spherical waves, there are two reasons why the filtered backprojection cannot be directly applied: on the one hand, due to the incident spherical waves, the experimental set-up corresponds to a cone beam geometry rather than a parallel beam geometry. Solutions to this have been suggested in [256, 257]. On the other hand, the imaged objects are of the same lengths scale as the wavelength of the probing beam, hence diffraction effects cannot be neglected. Since diffraction dominates in this regime, the ray optical considerations employed in the derivation of projections via line integrals are not valid [251]. Thus, for a tomographic reconstruction, the equivalent to the projections in non-diffractive tomography has to be generated based on the diffracted wave field [251, 254]. With this, an analogue to the Fourier Slice theorem, the Fourier Diffraction theorem, can be employed in the tomographic reconstruction [251, 258]. Since the wave equation for propagation through a medium can in general not be solved exactly, approximations such as the Born or the Rytov approximations, which are valid under different conditions, have to be applied [251, 258, 254].

Several studies have shown that optical diffraction tomography can successfully be used to image biological matter, especially when using the Rytov approximation [259, 260, 261, 262], and can be combined with holographic imaging [260, 262]. Although further adjustments will have to be made to account for the use of an electron beam, this could be a promising approach to tomography with LEEH.

6 Conclusion and Outlook

The results presented in this thesis show that low-energy electron holography can be employed to image a wide range of proteins on the single-molecule level at sub-nanometer resolution, from small proteins (cytochrome C, 12 kDa) to large protein complexes (GroEL, ca. 800 kDa). Structural features, such as subunits, can be identified and different molecular orientations and conformations of the proteins can be distinguished in the reconstructed images retrieved from the low-energy electron holograms. Comparisons with molecular models from the Protein Data Bank show that molecular sizes and shapes determined by LEEH imaging are in accordance with those of the models, implying that the proteins in general remain intact during the sample preparation and imaging process of LEEH.

Single-molecule imaging by LEEH allows the mapping of the conformational variability of intrinsically flexible proteins as exemplarily shown in a study of the monoclonal antibody Herceptin. A detailed statistical analysis revealed that the molecules observed on the surface can be classified into two categories: extended conformations related to the solution structure and compact conformations related to the gas-phase structure. The recovery of the extended structures occurs during the landing process and strongly depends on the landing energy.

While one-step amplitude reconstructions with a Fresnel-Kirchhoff propagator yield insights into structural details as well as molecular shapes and sizes, proteins in general exhibit both absorptive and phase-shifting properties when interacting with electron waves. Hence, reconstructing the phase shift distribution in the object plane is of relevance for retrieving the full information about the imaged protein that is encoded in the hologram.

An iterative phase retrieval algorithm based on ref. [193] has been presented, characterized in detail and applied to a wide range of simulated and experimentally acquired holograms. Multislice and scattering simulations as well as extensive data analysis have been performed to propose possible interpretations of the phase results, which point towards a connection between the molecular density and the induced phase shift.

First steps towards a three-dimensional reconstruction have been explored, both in the form of a deconvolution-based reconstruction from a single hologram and a tomographic reconstruction from holograms obtained from a series of tilt angles.

Although low-energy electron holography has only recently emerged as a method for single-molecule protein imaging, the results presented in this thesis show that it can provide relevant structural information and be considered as a complementary approach to other biomolecular imaging techniques. Yet, the method is still being developed, which requires further investigation and improvement both on the experimental and on the theoretical side.

Outlook

Structural analysis of proteins

Having established that LEEH can image both rigid and flexible proteins and can map conformational changes on the larger structural level, the next step is to apply the method to relevant biological questions. Such questions could for example pertain to the binding behaviour and associated structural changes of certain classes of molecules.

Structural changes in the gas phase, such as a partial collapse of the three-dimensional protein structure, which has been observed in the case of antibodies, are also a topic for further study with proteins of different mechanical properties. LEEH imaging of samples prepared by ES-IBD allows the direct imaging of gas phase-related structures on the surface, which could yield insights into the capabilities of LEEH as a structural biology technique as well as complement the structural information obtained by gas phase analysis methods.

Dosing water onto a protein sample in situ would also be of interest since it would allow for the possibility of observing the interaction with a solvent. LEEH imaging could reveal whether this induces structural changes in the proteins, giving insights into the differences between the native protein structures and the structures observed in UHV. Furthermore, water dosing could be a tool for exploring changes in the protein surface, which could potentially be observable in the phase reconstruction.

Interpreting phase data

Due to the high level of complexity of proteins, the interpretation of the reconstructed phase data from protein holograms is not straightforward. Despite the strong indication of a relation between phase shift and molecular density, a degree of ambiguity remains since other factors such as charges, changes in electric potential and changes in scattering strength likely also contribute.

Thus, test systems to separately explore the different possible contributions to the phase need to be defined and investigated. Since small metal-bearing molecules like porphyrins proved to be difficult in terms of sample preparation, custom-made structures like DNA origami could be a promising next step. By engineering certain molecular properties such as molecular thickness or the incorporation of metal atoms, different phase contributions could become separable.

In simple molecular systems, where the effect of a single metal centre should be visible, one might have to take an additional phase contribution due to spin into account, especially, when imaging magnetic centres at very low emission voltages.

The interpretation of phase data could also be advanced by improving the phase retrieval algorithm. So far, phase retrieval has been done iteratively from a single hologram, as a comparison, other phase retrieval techniques such as near-field ptychography [263, 79] could be employed. Ptychography can also be combined with multislice approaches [264, 265, 266] and with tomography [267] to retrieve three-dimensional information about the imaged objects.

3D analysis

For large, extended objects, multi-image approaches such as tomography are necessary for retrieving three-dimensional structural information from the measured data. Our current LEEH microscope head features a piezo rotation stage that allows the rotation of the sample around one axis by $\pm 20°$.

In addition to the development of a tomographic reconstruction algorithm compatible with LEEH imaging, the corresponding tomographic measurement technique needs to be established. The latter involves devising a calibration for the piezo rotation stage, since the tilt angle of each image needs to be known for reconstruction, and developing a scheme for keeping the imaged molecule at the centre of the illuminated spot during the rotation in order to be able to align the holograms for reconstruction.

Improvements of the experimental set-up

An increased resolution of the reconstructed images will be central for tackling all the challenges mentioned above, thus, the experimental set-up needs to be continuously improved. To improve the resolution, the cooling of tip and sample seems to be the most promising step. The cooling of the tip increases the coherence of the emitted electrons, which is directly related to the achievable resolution, while the cooling of the sample reduces molecular motion, vibrations of the graphene substrate as well as radiation damage, which all have adverse effects of the resolution.

A cryostat has recently been added to the LEEH set-up, which allows the cooling of tip and sample to temperatures of approximately $50\,\text{K}$. In preliminary tests, the resolution of molecules imaged at room temperature and low temperature was comparable. To optimize the measuring at low temperatures, cooling shields need to be added to avoid contamination of the sample.

Another step in improving the experimental set-up to increase the quality of the measured holograms is to implement magnetic shielding since stray magnetic fields can create distortions. Such magnetic shielding could be achieved by placing a μ-metal foil around tip and sample, which would be able to mitigate the effects from both external fields, such as the earth's magnetic field, and from local, spatially non-uniform magnetic fields that can arise due to magnetized components of the chamber.

Bibliography

[1] Sandeep Kumar, Buyong Ma, Chung Jung Tsai, Haim Wolfson, and Ruth Nussinov. Folding funnels and conformational transitions via hinge-bending motions. *Cell Biochemistry and Biophysics*, 31(2):141–164, 1999. ISSN 10859195. doi: 10.1007/BF02738169.

[2] Ozlem Keskin. Binding induced conformational changes of proteins correlate with their intrinsic fluctuations: a case study of antibodies. *BMC Structural Biology*, 7 (31), 2007.

[3] Sara E. Dobbins, Victor I. Lesk, and Michael J.E. Sternberg. Insights into protein flexibility: The relationship between normal modes and conformational change upon protein-protein docking. *Proceedings of the National Academy of Sciences of the United States of America*, 105(30):10390–10395, 2008. ISSN 00278424. doi: 10.1073/pnas.0802496105.

[4] Robert Huber and William S. Bennett. Functional significance of flexibility in proteins. *Pure and Applied Chemistry*, 54(12):2489–2500, 2013. ISSN 13653075. doi: 10.1351/pac198254122489.

[5] S. W. Schneider, S. Nuschele, A. Wixforth, C. Gorzelanny, A. Alexander-Katz, R. R. Netz, and M. F. Schneider. Shear-induced unfolding triggers adhesion of von Willebrand factor fibers. *Proceedings of the National Academy of Sciences of the United States of America*, 104(19):7899–7903, 2007. ISSN 00278424. doi: 10.1073/pnas.0608422104.

[6] Ross T.A. MacGillivray, Stanley A. Moore, Jie Chen, Bryan F. Anderson, Heather Baker, Yaoguang Luo, Maria Bewley, Clyde A. Smith, Michael E.P. Murphy, Yili Wang, Anne B. Mason, Robert C. Woodworth, Gary D. Brayer, and Edward N. Baker. Two high-resolution crytal sructures of the recombinant N-lobe of human tranferrin reveal a structure change implicated in iron release. *Biochemistry*, 37 (22):7919–7928, 1998. ISSN 00062960. doi: 10.1021/bi980355j.

[7] Ralph Weissleder and Matthias Nahrendorf. Advancing biomedical imaging. *Proceedings of the National Academy of Sciences of the United States of America*, 112 (47):14424–14428, 2015. ISSN 10916490. doi: 10.1073/pnas.1508524112.

[8] Laurent Maveyraud and Lionel Mourey. Protein X-ray crystallography and drug discovery. *Molecules*, 25(5), 2020. ISSN 14203049. doi: 10.3390/molecules25051030.

Bibliography

[9] Tom Ceska, Chun-wa Chung, Rob Cooke, Chris Phillips, and Pamela A Williams. Cryo-EM in drug discovery. *Biochemical Society Transactions*, 0(August 2018): 281–293, 2019. doi: 10.1042/BST20180267.

[10] Jean Paul Renaud, Ashwin Chari, Claudio Ciferri, Wen Ti Liu, Hervé William Rémigy, Holger Stark, and Christian Wiesmann. Cryo-EM in drug discovery: Achievements, limitations and prospects. *Nature Reviews Drug Discovery*, 17(7): 471–492, 2018. ISSN 14741784. doi: 10.1038/nrd.2018.77. URL http://dx.doi.org/ 10.1038/nrd.2018.77.

[11] Takanori Nakane, Abhay Kotecha, Andrija Sente, Greg McMullan, Simonas Masiulis, Patricia M.G.E. Brown, Ioana T. Grigoras, Lina Malinauskaite, Tomas Malinauskas, Jonas Miehling, Tomasz Uchański, Lingbo Yu, Dimple Karia, Evgeniya V. Pechnikova, Erwin de Jong, Jeroen Keizer, Maarten Bischoff, Jamie McCormack, Peter Tiemeijer, Steven W. Hardwick, Dimitri Y. Chirgadze, Garib Murshudov, A. Radu Aricescu, and Sjors H.W. Scheres. Single-particle cryo-EM at atomic resolution. *Nature*, 587(7832):152–156, 2020. ISSN 14764687. doi: 10.1038/s41586-020-2829-0. URL http://dx.doi.org/10.1038/s41586-020-2829-0.

[12] Slavica Jonic and Catherine Vénien-Bryan. Protein structure determination by electron cryo-microscopy. *Current Opinion in Pharmacology*, 9(5):636–642, 2009. ISSN 14714892. doi: 10.1016/j.coph.2009.04.006.

[13] Joachim Frank. Single-particle reconstruction of biological macromolecules in electron microscopy – 30 years. *Quarterly Reviews of Biophysics*, 42(3):139–158, 2009. doi: 10.1017/S0033583509990059.

[14] Alberto Bartesaghi, Alan Merk, Soojay Banerjee, Doreen Matthies, Xiongwu Wu, Jacqueline L S Milne, and Sriram Subramaniam. 2.2 Å resolution cryo-EM structure of. *Science*, 348(6239):1147–1152, 2015.

[15] Matthew K. Higgins and Susan M. Lea. On the state of crystallography at the dawn of the electron microscopy revolution. *Current Opinion in Structural Biology*, 46:95–101, 2017. ISSN 1879033X. doi: 10.1016/j.sbi.2017.06.005. URL http://dx.doi.org/10.1016/j.sbi.2017.06.005.

[16] Diego F. Gauto, Leandro F. Estrozi, Charles D. Schwieters, Gregory Effantin, Pavel Macek, Remy Sounier, Astrid C. Sivertsen, Elena Schmidt, Rime Kerfah, Guillaume Mas, Jacques Philippe Colletier, Peter Güntert, Adrien Favier, Guy Schoehn, Paul Schanda, and Jerome Boisbouvier. Integrated NMR and cryo-EM atomic-resolution structure determination of a half-megadalton enzyme complex. *Nature Communications*, 10(1):1–12, 2019. ISSN 20411723. doi: 10.1038/s41467-019-10490-9. URL http://dx.doi.org/10.1038/s41467-019-10490-9.

[17] Andrea Cavalli, Xavier Salvatella, Christopher M. Dobson, and Michele Vendruscolo. Protein structure determination from NMR chemical shifts. *Proceedings of the National Academy of Sciences of the United States of America*, 104(23):9615–9620, 2007. ISSN 00278424. doi: 10.1073/pnas.0610313104.

[18] Anthony K Mittermaier and Lewis E Kay. Observing biological dynamics at atomic resolution using NMR. *Trends in Biochemical Sciences*, 34(12):601–611, 2009. doi: 10.1016/j.tibs.2009.07.004.

[19] Fa-an Chao and R Andrew Byrd. Protein dynamics revealed by NMR relaxation methods. *Emerging Topics in Life Science*, 2:93–105, 2018.

[20] Ka Man Yip, Niels Fischer, Elham Paknia, Ashwin Chari, and Holger Stark. Atomic-resolution protein structure determination by cryo-EM. *Nature*, 587(7832): 157–161, 2020. ISSN 14764687. doi: 10.1038/s41586-020-2833-4. URL http://dx.doi.org/10.1038/s41586-020-2833-4.

[21] Megan J. Dobro, Linda A. Melanson, Grant J. Jensen, and Alasdair W. McDowall. Plunge Freezing for Electron Cryomicroscopy. In Grant J. Jensen, editor, *Methods in Enzymology*. Academic Press, 2010.

[22] Michael A Cianfrocco and Elizabeth H Kellogg. What Could Go Wrong? A Practical Guide to Single-Particle Cryo-EM: From Biochemistry to Atomic Models. *Journal of Chemical Information and Modeling*, 60(5):2458–2469, may 2020. ISSN 1549-9596. doi: 10.1021/acs.jcim.9b01178. URL https://pubs.acs.org/doi/10.1021/acs.jcim.9b01178.

[23] Robert M Glaeser. How good can cryo-EM become? *Nature Methods*, 13(1):28–32, 2016.

[24] Jean-Nicolas Longchamp, Stephan Rauschenbach, Sabine Abb, Conrad Escher, Tatiana Latychevskaia, Klaus Kern, and Hans-Werner Fink. Imaging proteins at the single-molecule level. *Proceedings of the National Academy of Sciences*, 114(7):1474–1479, 2017. ISSN 0027-8424. doi: 10.1073/pnas.1614519114. URL https://www.pnas.org/content/114/7/1474.

[25] Matthias Germann, Tatiana Latychevskaia, Conrad Escher, and Hans Werner Fink. Nondestructive imaging of individual biomolecules. *Physical Review Letters*, 104 (9):6–9, 2010. ISSN 00319007. doi: 10.1103/PhysRevLett.104.095501.

[26] J. Spence, W. Qian, and X. Zhang. Contrast and radiation damage in point-projection electron imaging of purple membrane at 100 V. *Ultramicroscopy*, 55(1): 19–23, 1994. ISSN 03043991. doi: 10.1016/0304-3991(94)90076-0.

[27] Dennis Gabor. A New Microscopic Principle. *Nature*, 161:777–778, 1948.

[28] Werner Stocker, Hans Werner Fink, and Roger Morin. Low-energy electron and ion projection microscopy. *Ultramicroscopy*, 31(4):379–384, 1989. ISSN 03043991. doi: 10.1016/0304-3991(89)90336-7.

[29] Hans-Werner Fink. Coherent point source electron beams. *Journal of Vacuum Science and Technology B: Microelectronics and Nanometer Structures*, 8(6):1323, 1990. ISSN 0734211X. doi: 10.1116/1.584911.

Bibliography

[30] Hans-Werner Fink, Werner Stocker, and Heinz Schmid. Holography with low-energy electrons. *Phys. Rev. Lett.*, 65(10):1204–1206, sep 1990. doi: 10.1103/PhysRevLett. 65.1204. URL https://link.aps.org/doi/10.1103/PhysRevLett.65.1204.

[31] J. Y. Mutus, L. Livadaru, J. T. Robinson, R. Urban, M. H. Salomons, M. Cloutier, and R. A. Wolkow. Low-energy electron point projection microscopy of suspended graphene, the ultimate microscope slide. *New Journal of Physics*, 13, 2011. ISSN 13672630. doi: 10.1088/1367-2630/13/6/063011.

[32] Jean Nicolas Longchamp, Tatiana Latychevskaia, Conrad Escher, and Hans Werner Fink. Low-energy electron holographic imaging of individual tobacco mosaic virions. *Applied Physics Letters*, 107(13), 2015. ISSN 00036951. doi: 10.1063/1.4931607. URL http://dx.doi.org/10.1063/1.4931607.

[33] Stephan Rauschenbach, Markus Ternes, Ludger Harnau, and Klaus Kern. Mass Spectrometry as a Preparative Tool for the Surface Science of Large Molecules. *Annual Review of Analytical Chemistry*, 9:473–498, 2016. ISSN 19361335. doi: 10.1146/annurev-anchem-071015-041633.

[34] Dennis Gabor. Holography, 1947-71, Nobel lecture 1971, 1971.

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

[36] E Ruska. Das Elektronenmikroskop als Übermikroskop. *Forschungen und Fortschritte*, 10, 1934.

[37] O. Scherzer. Über einige Fehler von Elektronenlinsen. *Zeitschrift für Physik*, 101 (9-10):593–603, 1936. ISSN 1434-6001.

[38] Dennis Gabor. Microscopy by reconstructed wave-fronts. *Proceedings of the Royal Society of London Series A - Mathematical and Physical Sciences*, 197(1051), 1949.

[39] Max Haider, Harald Rose, Stephan Uhlemann, Eugen Schwan, Bernd Kabius, and Knut Urban. A spherical-aberration-corrected 200 kV transmission electron microscope. *Ultramicroscopy*, 75(1):53–60, 1998. ISSN 03043991. doi: 10.1016/ S0304-3991(98)00048-5.

[40] Max Born, Emil Wolf, A. B. Bhatia, P. C. Clemmow, D. Gabor, A. R. Stokes, A. M. Taylor, P. A. Wayman, and W. L. Wilcock. *Principles of Optics*. Cambridge University Press, 1999. ISBN 9780521642224. doi: 10.1017/CBO9781139644181.

[41] T. H. Maiman. Stimulated Optical Radiation in Ruby. *Nature*, 187:493–494, 1960.

[42] A. Javan, W. R. Bennett, and D. R. Herriott. Population inversion and continuous optical maser oscillation in a gas discharge containing a He-Ne mixture. *Physical Review Letters*, 6(3):106–110, 1961. ISSN 00319007. doi: 10.1103/PhysRevLett.6. 106.

[43] Pierre Marquet, Benjamin Rappaz, Pierre J. Magistretti, Etienne Cuche, Yves Emery, Tristan Colomb, and Christian Depeursinge. Digital holographic microscopy: a noninvasive contrast imaging technique allowing quantitative visualization of living cells with subwavelength axial accuracy. *Optics Letters*, 30(5):468, 2005. ISSN 0146-9592. doi: 10.1364/ol.30.000468.

[44] Wenbo Xu, M. H. Jericho, I. A. Meinertzhagen, and H. J. Kreuzer. Digital in-line holography for biological applications. *Proceedings of the National Academy of Sciences of the United States of America*, 98(20):11301–11305, 2001. ISSN 00278424. doi: 10.1073/pnas.191361398.

[45] S. Eisebitt, J. Lüning, W. F. Schlotter, M. Lörgen, O. Hellwig, W. Eberhardt, and J. Stöhr. Lensless imaging of magnetic nanostructures by X-ray spectro-holography. *Nature*, 432(7019):885–888, 2004. ISSN 00280836. doi: 10.1038/nature03139.

[46] Martin Krenkel, Mareike Toepperwien, Frauke Alves, and Tim Salditt. Three-dimensional single-cell imaging with X-ray waveguides in the holographic regime. *Acta Crystallographica Section A: Foundations and Advances*, 73(4):282–292, 2017. ISSN 20532733. doi: 10.1107/S2053273317007902.

[47] M. Bartels, M. Krenkel, J. Haber, R. N. Wilke, and T. Salditt. X-ray holographic imaging of hydrated biological cells in solution. *Physical Review Letters*, 114(4): 1–5, 2015. ISSN 10797114. doi: 10.1103/PhysRevLett.114.048103.

[48] Henry N. Chapman, Stefan P. Hau-Riege, Michael J. Bogan, Saša Bajt, Anton Barty, Sébastien Boutet, Stefano Marchesini, Matthias Frank, Bruce W. Woods, W. Henry Benner, Richard A. London, Urs Rohner, Abraham Szöke, Eberhard Spiller, Thomas Möller, Christoph Bostedt, David A. Shapiro, Marion Kuhlmann, Rolf Treusch, Elke Plönjes, Florian Burmeister, Magnus Bergh, Carl Caleman, Gösta Huldt, M. Marvin Seibert, and Janos Hajdu. Femtosecond time-delay X-ray holography. *Nature*, 448(7154):676–679, 2007. ISSN 14764687. doi: 10.1038/nature06049.

[49] Hannes Lichte and Michael Lehmann. Electron holography - Basics and applications. *Reports on Progress in Physics*, 71(1), 2008. ISSN 00344885. doi: 10.1088/0034-4885/71/1/016102.

[50] Fengshan Zheng, Jan Caron, Vadim Migunov, Marco Beleggia, Giulio Pozzi, and Rafal E. Dunin-Borkowski. Measurement of charge density in nanoscale materials using off-axis electron holography. *Journal of Electron Spectroscopy and Related Phenomena*, 241(June):146881, 2020. ISSN 03682048. doi: 10.1016/j.elspec.2019. 07.002. URL https://doi.org/10.1016/j.elspec.2019.07.002.

[51] Leonardo Pierobon, András Kovács, Robin E. Schäublin, Stephan S.A. Gerstl, Jan Caron, Urs V. Wyss, Rafal E. Dunin-Borkowski, Jörg F. Löffler, and Michalis Charilaou. Unconventional magnetization textures and domain-wall pinning in Sm–Co magnets. *Scientific Reports*, 10(1):1–11, 2020. ISSN 20452322. doi: 10.1038/s41598-020-78010-0. URL https://doi.org/10.1038/s41598-020-78010-0.

[52] Tarjei Bondevik, Heine H. Ness, Calliope Bazioti, Truls Norby, Ole Martin Løvvik, Christoph T. Koch, and Øystein Prytz. Investigation of the electrostatic potential

of a grain boundary in Y-substituted BaZrO3 using inline electron holography. *Physical Chemistry Chemical Physics*, 21(32):17662–17672, 2019. ISSN 14639076. doi: 10.1039/c9cp02676b.

[53] Hans Werner Fink, Werner Stocker, and Heinz Schmid. Holography with low-energy electrons. *Physical Review Letters*, 65(10):1204–1206, 1990. ISSN 00319007. doi: 10.1103/PhysRevLett.65.1204.

[54] Emmett N. Leith and Juris Upatnieks. Reconstructed Wavefronts and Communication Theory. *Journal of the Optical Society of America*, 52(10):1123, 1962. ISSN 0030-3941. doi: 10.1364/josa.52.001123.

[55] Michael Lehmann and Hannes Lichte. Tutorial on off-axis electron holography. *Microscopy and Microanalysis*, 8(6):447–466, 2002. ISSN 14319276. doi: 10.1017/S1431927602029938.

[56] D. K. Saldin and J. C.H. Spence. On the mean inner potential in high- and low-energy electron diffraction. *Ultramicroscopy*, 55(4):397–406, 1994. ISSN 03043991. doi: 10.1016/0304-3991(94)90175-9.

[57] Zhenyu Tan, Yueyuan Xia, Mingwen Zhao, and Liu Xiangdong. Electron stopping power and inelastic mean free path in amino acids and protein over the energy range of 20 – 20,000 eV. *Radiation and Environmental Biophysics*, 45:135–143, 2006. doi: 10.1007/s00411-006-0049-0.

[58] Pawel Mozejko and Léon Sanche. Cross section calculations for electron scattering from DNA and RNA bases. *Radiation and Environmental Biophysics*, 42(3): 201–211, 2003. ISSN 0301634X. doi: 10.1007/s00411-003-0206-7.

[59] C. J. Russo and R. F. Egerton. Damage in electron cryomicroscopy: Lessons from biology for materials science. *MRS Bulletin*, 44(12):935–941, 2019. ISSN 08837694. doi: 10.1557/mrs.2019.284.

[60] Lindsay A. Baker and Joel L. Rubinstein. Radiation damage in electron cryomicroscopy. *Methods in Enzymology*, 481:371–388, 2010.

[61] R. F. Egerton, P. Li, and M. Malac. Radiation damage in the TEM and SEM. *Micron*, 35(6):399–409, 2004. ISSN 09684328. doi: 10.1016/j.micron.2004.02.003.

[62] R. F. Egerton. Mechanisms of radiation damage in beam-sensitive specimens, for TEM accelerating voltages between 10 and 300 kV. *Microscopy Research and Technique*, 75(11):1550–1556, 2012. ISSN 1059910X. doi: 10.1002/jemt.22099.

[63] Y Talmon. Electron Beam Radiation Damage to Organic and Biological Cryospecimens. In *Cryotechniques in Biological Electron Microscopy*. 1987. ISBN 9783642728174. doi: 10.1007/978-3-642-72815-0.

[64] Manikandan Karuppasamy, Fatemeh Karimi Nejadasl, Milos Vulovic, Abraham J. Koster, and Raimond B.G. Ravelli. Radiation damage in single-particle cryo-electron microscopy: Effects of dose and dose rate. *Journal of Synchrotron Radiation*, 18 (3):398–412, 2011. ISSN 09090495. doi: 10.1107/S090904951100820X.

[65] J. C.H. Spence. XFELs for structure and dynamics in biology. *IUCrJ*, 4:322–339, 2017. ISSN 20522525. doi: 10.1107/S2052252517005760.

[66] Henry N. Chapman, Carl Caleman, and Nicusor Timneanu. Diffraction before destruction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1647), 2014. ISSN 14712970. doi: 10.1098/rstb.2013.0313.

[67] Richard Henderson. The Potential and Limitations of Neutrons, Electrons and X-Rays for Atomic Resolution Microscopy of Unstained Biological Molecules. *Quarterly Reviews of Biophysics*, 28(2):171–193, 1995. ISSN 14698994. doi: 10.1017/S003358350000305X.

[68] David C. Joy and Carolyn S. Joy. Low voltage scanning electron microscopy. *Micron*, 27(3-4):247–263, 1996. ISSN 09684328. doi: 10.1016/0968-4328(96)00023-6.

[69] M Isaacson. Inelastic scattering and beam damage in biological molecules. In *Physical Aspects of Electron Microscopy and Microbeam Analysis*. 1975.

[70] A. Howie, F. J. Rocca, and U. Valdrè. Electron beam ionization damage processes in p-terphenyl. *Philosophical Magazine B: Physics of Condensed Matter; Statistical Mechanics, Electronic, Optical and Magnetic Properties*, 52(3):751–757, 1985. ISSN 13642812. doi: 10.1080/13642818508240634.

[71] Michael R. Stevens, Qing Chen, Uwe Weierstall, and John C.H. Spence. Transmission Electron Diffraction at 200 eV and Damage Thresholds below the Carbon K Edge. *Microscopy and Microanalysis*, 6(4):368–379, 2000. ISSN 1431-9276. doi: 10.1007/s100050010041.

[72] Th Schmidt, H. Marchetto, P. L. Lévesque, U. Groh, F. Maier, D. Preikszas, P. Hartel, R. Spehr, G. Lilienkamp, W. Engel, R. Fink, E. Bauer, H. Rose, E. Umbach, and H. J. Freund. Double aberration correction in a low-energy electron microscope. *Ultramicroscopy*, 110(11):1358–1361, 2010. ISSN 03043991. doi: 10.1016/j.ultramic.2010.07.007.

[73] Tatiana Latychevskaia and Hans-Werner Fink. Practical algorithms for simulation and reconstruction of digital in-line holograms. *Applied Optics*, 54(9):2424, 2015. ISSN 1559-128X. doi: 10.1364/ao.54.002424.

[74] Dennis Gabor. Microscopy by reconstructed wave fronts: II. *Proceedings of the Physical Society. Section B*, 64(6):449–469, 1951. ISSN 03701301. doi: 10.1088/0370-1301/64/6/301.

[75] Francesco Maria Grimaldi. *Physico mathesis de lumine, coloribus, et iride, aliisque annexis libri duo*. 1665.

[76] Augustin Fresnel. Mémoire sur la diffraction de la lumière. *Oeuvres complètes d'Augustin Fresnel, vol. 1*, 1866.

[77] Christiaan Huygens. *Traité de la Lumière*. 1690.

[78] D. Paganin. *Coherent X-ray Optics*. Oxford University Press, New York, 2006.

[79] A. L. Robisch, K. Kröger, A. Rack, and T. Salditt. Near-field ptychography using lateral and longitudinal shifts. *New Journal of Physics*, 17(7), 2015. ISSN 13672630. doi: 10.1088/1367-2630/17/7/073033.

[80] A. Santos, C. Ortiz De Solórzano, J. J. Vaquero, J. M. Peña, N. Malpica, and F. Del Pozo. Evaluation of autofocus functions in molecular cytogenetic analysis. *Journal of Microscopy*, 188(3):264–272, 1997. ISSN 00222720. doi: 10.1046/j.1365-2818. 1997.2630819.x.

[81] Frank Dubois, Cédric Schockaert, Natacha Callens, and Catherine Yourassowsky. Focus plane detection criteria in digital holography microscopy by amplitude analysis. *Optics Express*, 14(13):5895, 2006. ISSN 1094-4087. doi: 10.1364/oe.14. 005895.

[82] Lisa J. Harris, Steven B. Larson, Karl W. Hasel, and Alexander McPherson. Refined structure of an intact IgG2a monoclonal antibody. *Biochemistry*, 36(7):1581–1597, 1997. ISSN 00062960. doi: 10.1021/bi962514+.

[83] J.W. Goodman. *Statistical Optics*. Wiley Classics Library, New York, 2000.

[84] Eugene Hecht. *Optics*. Addison Wesley, 2002.

[85] Friso Van Der Veen and Franz Pfeiffer. Coherent x-ray scattering. *Journal of Physics Condensed Matter*, 16(28):5003–5030, 2004. ISSN 09538984. doi: 10.1088/ 0953-8984/16/28/020.

[86] Tatiana Latychevskaia. Spatial coherence of electron beams from field emitters and its effect on the resolution of imaged objects. *Ultramicroscopy*, 175:121– 129, 2017. ISSN 18792723. doi: 10.1016/j.ultramic.2016.11.008. URL http: //dx.doi.org/10.1016/j.ultramic.2016.11.008.

[87] Dominik Ehberger, Jakob Hammer, Max Eisele, Michael Krüger, Jonathan Noe, Alexander Högele, and Peter Hommelhoff. Highly Coherent Electron Beam from a Laser-Triggered Tungsten Needle Tip. *Physical Review Letters*, 114(22):1–5, 2015. ISSN 10797114. doi: 10.1103/PhysRevLett.114.227601.

[88] J. C. H. Spence, W. Qian, and M. P. Silverman. Electron source brightness and degeneracy from Fresnel fringes in field emission point projection microscopy. *Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films*, 12(2): 542–547, 1994. ISSN 0734-2101. doi: 10.1116/1.579166.

[89] B. Cho, T. Ichimura, R. Shimizu, and C. Oshima. Quantitative evaluation of spatial coherence of the electron beam from low temperature field emitters. *Physical Review Letters*, 92(24):1–4, 2004. ISSN 00319007. doi: 10.1103/PhysRevLett.92.246103.

[90] Soichiro Matsunaga and Yasunari Sohda. Reducing virtual source size by using facetless electron source for high brightness. *AIP Advances*, 9(6), 2019. ISSN 21583226. doi: 10.1063/1.5098528. URL http://dx.doi.org/10.1063/1.5098528.

[91] Josh Y Mutus, Lucian Livadaru, Radovan Urban, Jason Pitters, A. Peter Legg, Mark H. Salomons, Martin Cloutier, and Robert A. Wolkow. Nanoscale structuring of tungsten tip yields most coherent electron point-source. *New Journal of Physics*, 15(073038), 2013. doi: 10.1088/1367-2630/15/7/073038.

[92] Tatiana Latychevskaia. Lateral and axial resolution criteria in incoherent and coherent optics and holography, near- and far-field regimes. *Applied Optics*, 58(13): 3597, 2019. ISSN 1559-128X. doi: 10.1364/ao.58.003597.

[93] Tatiana Latychevskaia and Hans Werner Fink. Resolution enhancement in in-line holography by numerical compensation of vibrations. *Optics Express*, 25(17), 2017. ISSN 23318422. doi: 10.1364/oe.25.020109.

[94] Ryuta Mizutani, Rino Saiga, Susumu Takekoshi, Chie Inomoto, Naoya Nakamura, Masanari Itokawa, Makoto Arai, Kenichi Oshima, Akihisa Takeuchi, Kentaro Uesugi, Yasuko Terada, and Yoshio Suzuki. A method for estimating spatial resolution of real image in the Fourier domain. *Journal of Microscopy*, 261(1): 57–66, 2016. ISSN 13652818. doi: 10.1111/jmi.12315.

[95] M. H. Jericho and H. J. Kreuzer. *Point Source Digital In-Line Holographic Microscopy*. Springer, Berlin, Heidelberg, 2011. ISBN 978-3-642-15812-4.

[96] Tatiana Latychevskaia and Hans-Werner Fink. Resolution enhancement in digital holography by self-extrapolation of holograms. *Optics Express*, 21(6):7726, 2013. ISSN 1094-4087. doi: 10.1364/oe.21.007726.

[97] Precision Laboratory. URL https://www.fkf.mpg.de/2489208/01_Precision_ Laboratory.

[98] Tatiana Latychevskaia. Phase retrieval methods applied to coherent imaging, 2014.

[99] Ralph Howard Fowler and Lothar Wolfgang Nordheim. Electron Emission in Intense Electric Fields. *Proceedings of the Royal Society of London Series A - Mathematical and Physical Sciences*, 119(781):173–181, 1928.

[100] Lothar Wolfgang Nordheim. The effect of the image force on the emission and reflexion of electrons by metals. 121(788):626–639, 1928.

[101] Richard G. Forbes and Jonathan H.B. Deane. Transmission coefficients for the exact triangular barrier: An exact general analytical theory that can replace Fowler & Nordheim's 1928 theory. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 467(2134):2927–2947, 2011. ISSN 14712946. doi: 10.1098/rspa.2011.0025.

[102] Claudia Lenk, Steve Lenk, Mathias Holz, Elshad Guliyev, Martin Hofmann, Tzvetan Ivanov, Ivo W. Rangelow, Mahmoud Behzadirad, Ashwin K. Rishinaramangalam, Daniel Feezell, and Tito Busani. Experimental study of field emission from ultra-sharp silicon, diamond, GaN, and tungsten tips in close proximity to the counter electrode. *Journal of Vacuum Science & Technology B*, 36(6):06JL03, 2018. ISSN 2166-2746. doi: 10.1116/1.5048518. URL http://dx.doi.org/10.1116/1.5048518.

[103] Lucian Livadaru, Josh Mutus, and Robert A. Wolkow. Limits of elemental contrast by low energy electron point source holography. *Journal of Applied Physics*, 110 (9), 2011. ISSN 00218979. doi: 10.1063/1.3658250.

[104] J. A. Panitz. Field-ion microscopy - A review of basic principles and selected applications. *Journal of Physics E: Scientific Instruments*, 15(12):1281–1294, 1982. ISSN 00223735. doi: 10.1088/0022-3735/15/12/004.

[105] Anne-Sophie Lucier. *Preparation and Characterization of Tungsten Tips Suitable for Molecular Electronics Studies*. PhD thesis, 2004.

[106] Allan J. Melmed. The art and science and other aspects of making sharp tips. *Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures*, 9(2):601, 1991. ISSN 0734211X. doi: 10.1116/1.585467.

[107] Bing Feng Ju, Yuan Liu Chen, and Yaozheng Ge. The art of electrochemical etching for preparing tungsten probes with controllable tip profile and characteristic parameters. *Review of Scientific Instruments*, 82(1), 2011. ISSN 00346748. doi: 10.1063/1.3529880.

[108] Erwin W. Müller. Das Feldionenmikroskop. *Zeitschrift für Physik*, 131:136–142, 1951.

[109] Erwin W. Müller and Kanwar Bahadur. Field Ionization of Gases at a Metal Surface and the Resolution of the Field Ion Microscope. *Physical Review*, 102(624), 1956.

[110] R. R. Nair, P. Blake, J. R. Blake, R. Zan, S. Anissimova, U. Bangert, A. P. Golovanov, S. V. Morozov, A. K. Geim, K. S. Novoselov, and T. Latychevskaia. Graphene as a transparent conductive support for studying biological molecules by transmission electron microscopy. *Applied Physics Letters*, 97:153102, 2010. doi: 10.1063/1.3492845.

[111] Jean-Nicolas Longchamp, Conrad Escher, and Hans-Werner Fink. Ultraclean freestanding graphene by platinum-metal catalysis. *Journal of Vacuum Science and Technology B*, 31, 2013.

[112] Helena Hernández and Carol V. Robinson. Determining the stoichiometry and interactions of macromolecular assemblies from mass spectrometry. *Nature Protocols*, 2(3):715–726, 2007. ISSN 17542189. doi: 10.1038/nprot.2007.73.

[113] Idlir Liko, Timothy M. Allison, Jonathan TS Hopper, and Carol V. Robinson. Mass spectrometry guided structural biology. *Current Opinion in Structural Biology*, 40:136–144, 2016. ISSN 1879033X. doi: 10.1016/j.sbi.2016.09.008. URL http://dx.doi.org/10.1016/j.sbi.2016.09.008.

[114] Howard G. Barth and Barry E. Boyes. Size Exclusion Chromatography. *Analytical Chemistry*, 62(12):268–303, 1990. ISSN 15206882. doi: 10.1021/ac00211a020.

[115] Richard R. Burgess. A brief practical review of size exclusion chromatography: Rules of thumb, limitations, and troubleshooting. *Protein Expression and Purification*, 150(May):81–85, 2018. ISSN 10465928. doi: 10.1016/j.pep.2018.05.007. URL https://doi.org/10.1016/j.pep.2018.05.007.

[116] John B Fenn, Matthias Mann, Chin Kai Meng, Shek Fu Wong, and Craig M Whitehouse. Electrospray Ionization for Mass Spectrometry of Large Biomolecules. *Science*, 246(4926):64–71, 1989.

[117] John B Fenn, Matthias Mann, Chin Kai Meng, Shek Fu Wong, and Craige M Whitehouse. Electrospray ionization -principles and practice. *Mass Spectrometry Reviews*, 9:37–70, 1990.

[118] Joost Snijder, Rebecca J. Rose, David Veesler, John E. Johnson, and Albert J.R. Heck. Studying 18 MDa virus assemblies with native mass spectrometry. *Angewandte Chemie - International Edition*, 52(14):4020–4023, 2013. ISSN 14337851. doi: 10.1002/anie.201210197.

[119] Leonhard H. Urner, Idlir Liko, Hsin Yung Yen, Kin Kuan Hoi, Jani Reddy Bolla, Joseph Gault, Fernando Gonçalves Almeida, Marc Philip Schweder, Denis Shutin, Svenja Ehrmann, Rainer Haag, Carol V. Robinson, and Kevin Pagel. Modular detergents tailor the purification and structural analysis of membrane proteins including G-protein coupled receptors. *Nature Communications*, 11(1), 2020. ISSN 20411723. doi: 10.1038/s41467-020-14424-8. URL http://dx.doi.org/10.1038/s41467-020-14424-8.

[120] Hsin Yung Yen, Jonathan T.S. Hopper, Idlir Liko, Timothy M. Allison, Ya Zhu, Dejian Wang, Monika Stegmann, Shabaz Mohammed, Beili Wu, and Carol V. Robinson. Ligand binding to a G protein–coupled receptor captured in a mass spectrometer. *Science Advances*, 3(6):1–7, 2017. ISSN 23752548. doi: 10.1126/sciadv.1701016.

[121] Helena Hernández and Carol V Robinson. Dynamic Protein Complexes: Insights from Mass Spectrometry. *The Journal of Biological Chemistry*, 276(50):46685–46688, 2001. doi: 10.1074/jbc.R100024200.

[122] Hortense Videler, Leopold L Ilag, Adam R C Mckay, Charlotte L Hanson, and Carol V Robinson. Mass spectrometry of intact ribosomes. *FEBS Letters*, 579: 943–947, 2005. doi: 10.1016/j.febslet.2004.12.003.

[123] Geoffrey Taylor. Disintegration of water drops in an electric field. *Proceedings of the Royal Society of London Series A - Mathematical and Physical Sciences*, 280 (1382), 1964.

[124] Denis Duft, Tobias Achtzehn, Rene Müller, Bernd A. Huber, and Thomas Leisner. Rayleigh jets from levitated microdroplets. *Nature*, 421:128, 2003.

[125] Paul Kebarle and Michael Peschke. On the mechanisms by which the charged droplets produced by electrospray lead to gas phase ions. *Analytica Chimica Acta*, 406(1):11–35, 2000. doi: 10.1016/S0003-2670(99)00598-X.

Bibliography

[126] Paul Kebarle. A brief overview of the present status of the mechanisms involved in electrospray mass spectrometry. *Journal of Mass Spectrometry*, 35:804–817, 2000. doi: 10.1007/s12517-014-1353-7.

[127] M Labowsky, J B Fenn, and J La Fernandez Mora. A continuum model for ion evaporation from a drop. *Analytica Chimica Acta*, 406:105–118, 2000.

[128] J. V. Iribarne and B. A. Thomson. On the evaporation of small ions from charged droplets. *The Journal of Chemical Physics*, 64:2287–2294, 1976. ISSN 00219606. doi: 10.1063/1.432536.

[129] B. A. Thomson and J. V. Iribarne. Field induced ion evaporation from liquid surfaces at atmospheric pressure. *The Journal of Chemical Physics*, 71(11):4451–4463, 1979. ISSN 00219606. doi: 10.1063/1.438198.

[130] Christopher J. Hogan, James A. Carroll, Henry W. Rohrs, Pratim Biswas, and Michael L. Gross. Combined charged residue-field emission model of macromolecular electrospray ionization. *Analytical Chemistry*, 81(1):369–377, 2009. ISSN 00032700. doi: 10.1021/ac8016532.

[131] Lars Konermann, Elias Ahadi, Antony D Rodriguez, and Siavash Vahidi. Unraveling the Mechanism of Electrospray Ionization. *Analytica Chimica Acta*, 85:2–9, 2013.

[132] J. Fernandez de la Mora. Electrospray ionization of large multiply charged species proceeds via Dole's charged residue mechanism. *Analytica Chimica Acta*, 406: 93–104, 2000.

[133] Carol V. Robinson and Sheena E. Radford. Weighing the evidence for structure: electrospray ionization mass spectrometry of proteins. *Structure*, 3(9):861–865, 1995. ISSN 09692126. doi: 10.1016/S0969-2126(01)00221-0.

[134] Aneika C. Leney and Albert J.R. Heck. Native Mass Spectrometry: What is in the Name? *Journal of the American Society for Mass Spectrometry*, 28:5–13, 2017. ISSN 18791123. doi: 10.1007/s13361-016-1545-3. URL http://dx.doi.org/10.1007/s13361-016-1545-3.

[135] Cherokee S. Hoaglund-Hyzer, Anne E. Counterman, and David E. Clemmer. Anhydrous Protein Ions. *Chemical Reviews*, 99(10):3037–3079, 1999. ISSN 00092665. doi: 10.1021/cr980139g.

[136] Thomas B Parsons, Weston B Struwe, Joseph Gault, Keisuke Yamamoto, Thomas A Taylor, Ritu Raj, Kim Wals, Shabaz Mohammed, Carol V Robinson, Justin L P Benesch, and Benjamin G Davis. Optimal Synthetic Glycosylation of a Therapeutic Antibody. *Angewandte Chemie - International Edition*, 55:2361–2367, 2016. doi: 10.1002/anie.201508723.

[137] Albert J.R. Heck. Native mass spectrometry: A bridge between interactomics and structural biology. *Nature Methods*, 5(11):927–933, 2008. ISSN 15487091. doi: 10.1038/nmeth.1265.

[138] Kelvin Anggara, Yuntao Zhu, Martina Delbianco, Stephan Rauschenbach, Sabine Abb, Peter H. Seeberger, and Klaus Kern. Exploring the Molecular Conformation Space by Soft Molecule-Surface Collision. *Journal of the American Chemical Society*, 142(51):21420–21427, 2020. ISSN 15205126. doi: 10.1021/jacs.0c09933.

[139] Kelvin Anggara, Yuntao Zhu, Giulio Fittolani, Yang Yu, Theodore Tyrikos-Ergas, Martina Delbianco, Stephan Rauschenbach, Sabine Abb, Peter H. Seeberger, and Klaus Kern. Identifying the origin of local flexibility in a carbohydrate polymer. *Proceedings of the National Academy of Sciences of the United States of America*, 118(23):1–6, 2021. ISSN 10916490. doi: 10.1073/pnas.2102168118.

[140] X. Wu, M. Delbianco, K. Anggara, T. Michnowicz, A. Pardo-Vargas, P. Bharate, S. Sen, M. Pristl, S. Rauschenbach, U. Schlickum, S. Abb, P. H. Seeberger, and K. Kern. Imaging single glycans. *Nature*, 582(7812):375–378, 2020. ISSN 14764687. doi: 10.1038/s41586-020-2362-1.

[141] Natalie J. Thompson, Sara Rosati, Rebecca J. Rose, and Albert J.R. Heck. The impact of mass spectrometry on the study of intact antibodies: From post-translational modifications to structural analysis. *Chemical Communications*, 49(6):538–548, 2013. ISSN 1364548X. doi: 10.1039/c2cc36755f.

[142] Stephan Rauschenbach, Frank L Stadler, Eugenio Lunedei, Nicola Malinowski, Sergej Koltsov, Giovanni Costantini, and Klaus Kern. Electrospray Ion Beam Deposition of Clusters and Biomolecules. *Small*, 2(4):540–547, 2006. doi: https://doi.org/10.1002/smll.200500479. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/smll.200500479.

[143] Zhitao Deng, Nicha Thontasen, Nikola Malinowski, Gordon Rinke, Ludger Harnau, Stephan Rauschenbach, and Klaus Kern. A close look at proteins: Submolecular resolution of two- and three-dimensionally folded cytochrome c at surfaces. *Nano Letters*, 12(5):2452–2458, 2012. ISSN 15306984. doi: 10.1021/nl3005385.

[144] Stephan Rauschenbach, Gordon Rinke, Rico Gutzler, Sabine Abb, Alyazan Albarghash, Duy Le, Talat S. Rahman, Michael Dürr, Ludger Harnau, and Klaus Kern. Two-Dimensional Folding of Polypeptides into Molecular Nanostructures at Surfaces. *ACS Nano*, 11(3):2420–2427, 2017. ISSN 1936086X. doi: 10.1021/acsnano.6b06145.

[145] Julia Laskin, Peng Wang, and Omar Hadjar. Soft-landing of peptide ions onto self-assembled monolayer surfaces: An overview. *Physical Chemistry Chemical Physics*, 10(8):1079–1090, 2008. ISSN 14639076. doi: 10.1039/b712710c.

[146] Grant E. Johnson, Don Gunaratne, and Julia Laskin. Soft- and reactive landing of ions onto surfaces: Concepts and applications. *Mass Spectrometry Reviews*, 35(3):439–479, 2016. ISSN 10982787. doi: 10.1002/mas.21451.

[147] Zheng Ouyang, Zoltán Takáts, Thomas A. Blake, Bogdan Gologan, Andy J. Guymon, Justin M. Wiseman, Justin C. Oliver, V. Jo Davisson, and R. Graham Cooks. Preparing protein microarrays by soft-landing of mass-selected ions. *Science*, 301(5638):1351–1354, 2003. ISSN 00368075. doi: 10.1126/science.1088776.

[148] Lukas Krumbein, Kelvin Anggara, Martina Stella, Tomasz Michnowicz, Hannah Ochner, Sabine Abb, Gordon Rinke, André Portz, Michael Dürr, Uta Schlickum, Andrew Baldwin, Andrea Floris, Klaus Kern, and Stephan Rauschenbach. Fast molecular compression by a hyperthermal collision gives bond-selective mechanochemistry. *Phys. Rev. Lett.*, 126(5):056001, 2021. ISSN 23318422. doi: 10.1103/physrevlett.126.056001.

[149] Gordon Rinke, Stephan Rauschenbach, Ludger Harnau, Alyazan Albarghash, Matthias Pauly, and Klaus Kern. Active conformation control of unfolded proteins by hyperthermal collision with a metal surface. *Nano Letters*, 14(10):5609–5615, 2014. ISSN 15306992. doi: 10.1021/nl502122j.

[150] Kathrin Breuker and Fred W Mclafferty. Stepwise evolution of protein native structure with electrospray into the gas phase, 10^{-12} to 10^2 s. *Proceedings of the National Academy of Sciences*, 105(47):18145–18152, 2008.

[151] Stephan Warnke, Gert Von Helden, and Kevin Pagel. Protein Structure in the Gas Phase: The Influence of Side-Chain Microsolvation. *Journal of the American Chemical Society*, 135:1177–1180, 2013.

[152] Kjetil Hansen, Andy M. Lau, Kevin Giles, James M McDonnell, Weston B Struwe, Brian J Sutton, and Argyris Politis. A Mass-Spectrometry-Based Modelling Workflow for Accurate Prediction of IgG Antibody Conformations in the Gas Phase. *Angewandte Chemie*, 130:17440–17445, 2018.

[153] Martina Palamini, Anselmo Canciani, and Federico Forneris. Identifying and Visualizing Macromolecular Flexibility in Structural Biology. *Frontiers in Molecular Biosciences*, 3, 2016. doi: 10.3389/fmolb.2016.00047.

[154] Yufeng Hu, Zhenhang Chen, Yanjun Fu, Qingzhong He, Lun Jiang, Jiangge Zheng, Yina Gao, Pinchao Mei, Zhongzhou Chen, and Xueqin Ren. The amino-terminal structure of human fragile X mental retardation protein obtained using precipitant-immobilized imprinted polymers. *Nature Communications*, 6(6634), 2015. doi: 10.1038/ncomms7634.

[155] Pau Bernardó, Efstratios Mylonas, Maxim V Petoukhov, Martin Blackledge, and Dmitri I. Svergun. Structural characterization of flexible proteins using Small-Angle X-ray Scattering. *Journal of the American Chemical Society*, 129:5656–5664, 2007.

[156] Joshua Holcomb, Nicholas Spellmon, Yingxue Zhang, Maysaa Doughan, Chunying Li, and Zhe Yang. Protein crystallization: Eluding the bottleneck of X-ray crystallography. *AIMS Biophysics*, 4(4):557–575, 2017. ISSN 23779098. doi: 10.3934/biophy.2017.4.557.

[157] Sara Sandin, Lars-Göran Öfverstedt, Ann-Charlotte Wikström, Örjan Wrange, and Ulf Skoglund. Structure and Flexibility of Individual Immunoglobulin G Molecules in Solution. *Structure*, 12(3):409–415, 2004. ISSN 0969-2126. doi: https://doi.org/10.1016/j.str.2004.02.011. URL http://www.sciencedirect.com/science/article/pii/S096921260400053X.

[158] Xing Zhang, Lei Zhang, Huimin Tong, Bo Peng, Matthew J. Rames, Shengli Zhang, and Gang Ren. 3D Structural Fluctuation of IgG1 Antibody Revealed by Individual Particle Electron Tomography. *Scientific Reports*, 5:1–14, 2015. ISSN 20452322. doi: 10.1038/srep09803.

[159] Hans-Werner Fink, Heinz Schmid, Eugen Ermantraut, and Torsten Schulz. Electron holography of individual DNA molecules. *Journal of the Optical Society of America A*, 14(9):2168, 1997. ISSN 1084-7529. doi: 10.1364/josaa.14.002168.

[160] Tatiana Latychevskaia, Conrad Escher, William Andregg, Michael Andregg, and Hans Werner Fink. Direct visualization of charge transport in suspended (or free-standing) DNA strands by low-energy electron microscopy. *Scientific Reports*, 9(1):1–8, 2019. ISSN 20452322. doi: 10.1038/s41598-019-45351-4. URL http://dx.doi.org/10.1038/s41598-019-45351-4.

[161] Gregory B. Stevens, Michael Krüger, Tatiana Latychevskaia, Peter Lindner, Andreas Plückthun, and Hans Werner Fink. Individual filamentous phage imaged by electron holography. *European Biophysics Journal*, 40(10):1197–1201, 2011. ISSN 01757571. doi: 10.1007/s00249-011-0743-y.

[162] C Janeway, P Travers, and M Walport. *The structure of a typical antibody molecule. In: Immunobiology: The Immune System in Health and Disease.*

[163] Patrick Chames, Marc Van Regenmortel, Etienne Weiss, and Daniel Baty. Therapeutic antibodies: Successes, limitations and hopes for the future. *British Journal of Pharmacology*, 157(2):220–233, 2009. ISSN 00071188. doi: 10.1111/j.1476-5381. 2009.00190.x.

[164] Ruei Min Lu, Yu Chyi Hwang, I. Ju Liu, Chi Chiu Lee, Han Zen Tsai, Hsin Jung Li, and Han Chung Wu. Development of therapeutic antibodies for the treatment of diseases. *Journal of Biomedical Science*, 27(1):1–30, 2020. ISSN 14230127. doi: 10.1186/s12929-019-0592-z.

[165] Bo Wang, Sachith Gallolu Kankanamalage, Jianbo Dong, and Yue Liu. Optimization of therapeutic antibodies. *Antibody Therapeutics*, 4(1):45–54, 2021. ISSN 2516-4236. doi: 10.1093/abt/tbab003.

[166] Nicola L. Lawson, Carly I. Dix, Paul W. Scorer, Christopher J. Stubbs, Edmond Wong, Liam Hutchinson, Eileen J. McCall, Marianne Schimpl, Emma DeVries, Jill Walker, Gareth H. Williams, James Hunt, and Craig Barker. Mapping the binding sites of antibodies utilized in programmed cell death ligand-1 predictive immunohistochemical assays for use with immuno-oncology therapies. *Modern Pathology*, 33(4):518–530, 2020. ISSN 15300285. doi: 10.1038/s41379-019-0372-z. URL http://dx.doi.org/10.1038/s41379-019-0372-z.

[167] Mohammed M. Al Qaraghuli, Karina Kubiak-Ossowska, Valerie A. Ferro, and Paul A. Mulheran. Antibody-protein binding and conformational changes: identifying allosteric signalling pathways to engineer a better effector response. *Scientific Reports*, 10(12696), 2020. ISSN 20452322. doi: 10.1038/s41598-020-70680-0. URL https://doi.org/10.1038/s41598-020-70680-0.

Bibliography

[168] Rosie Upton, David Firth, and Perdita E Barran. Hybrid mass spectrometry methods reveal lot-to-lot differences and delineate the effects of glycosylation on the tertiary structure of Herceptin. *Chemical Science*, 10:2811–2820, 2019. doi: 10.1039/c8sc05029e.

[169] Julien Marcoux, Thierry Champion, Olivier Colas, Elsa Wagner-Rousset, Nathalie Corvaïa, Alain Van Dorsselaer, Alain Beck, and Sarah Cianférani. Native mass spectrometry and ion mobility characterization of trastuzumab emtansine, a lysine-linked antibody drug conjugate. *Protein Science*, 24(8):1210–1223, 2015. ISSN 1469896X. doi: 10.1002/pro.2666.

[170] Hannah Ochner, Sven Szilagyi, Sabine Abb, Joseph Gault, Carol V Robinson, Luigi Malavolti, Stephan Rauschenbach, and Klaus Kern. Low-energy electron holography imaging of conformational variability of single-antibody molecules from electrospray ion beam deposition. *Proceedings of the National Academy of Sciences*, 118(51), 2021. doi: 10.1073/pnas.2112651118/-/DCSupplemental.Published.

[171] Lisa J. Harris, Eileen Skaletsky, and Alexander McPherson. Crystallographic structure of an intact IgG1 monoclonal antibody. *Journal of Molecular Biology*, 275(5):861–872, 1998. ISSN 00222836. doi: 10.1006/jmbi.1997.1508.

[172] J G Vilhena, A C Dumitru, Elena T Herruzo, Jesús I Mendieta-Moreno, Ricardo Garcia, P A Serena, and Rubén Pérez. Adsorption orientations and immunological recognition of antibodies on graphene. *Nanoscale*, 8(27):13463–13475, 2016. doi: 10.1039/C5NR07612A. URL http://dx.doi.org/10.1039/C5NR07612A.

[173] Paul W A Devine, Henry C Fisher, Antonio N Calabrese, Fiona Whelan, Daniel R Higazi, Jennifer R Potts, David C Lowe, Sheena E Radford, and Alison E Ashcroft. Investigating the Structural Compaction of Biomolecules Upon Transition to the Gas-Phase Using ESI-TWIMS-MS. *Journal of the American Society for Mass Spectrometry*, 28(9):1855–1862, 2017. doi: 10.1007/s13361-017-1689-9.

[174] Kamila J Pacholarz, Massimiliano Porrini, Rachel A Garlish, Rebecca J Burnley, Richard J Taylor, Alistair J Henry, and Perdita E Barran. Dynamics of Intact Immunoglobulin G Explored by Drift-Tube Ion- Mobility Mass Spectrometry and Molecular Modeling Angewandte. *Angewandte Chemie - International Edition*, 53: 7765–7769, 2014. doi: 10.1002/anie.201402863.

[175] Massimiliano Porrini, Frederic Rosu, Clemence Rabin, Leonardo Darre, Hansel Gomez, Modesto Orozco, and Valerie Gabelica. Compaction of Duplex Nucleic Acids upon Native Electrospray Mass Spectrometry. *ACS Central Science*, 3: 454–461, 2017. doi: 10.1021/acscentsci.7b00084.

[176] Amber D. Rolland and James S. Prell. Computational Insights into Compaction of Gas-Phase Protein and Protein Complex Ions in Native Ion Mobility-Mass Spectrometry. *Trends in Analytical Chemistry*, 119:282–291, 2019.

[177] Erik G. Marklund, Matteo T. Degiacomi, Carol V. Robinson, Andrew J. Baldwin, and Justin L.P. Benesch. Collision cross sections for structural proteomics. *Structure*,

23(4):791–799, 2015. ISSN 18784186. doi: 10.1016/j.str.2015.02.010. URL http://dx.doi.org/10.1016/j.str.2015.02.010.

[178] Sherman Picardo and Remo Panaccione. Anti-MADCAM therapy for ulcerative colitis. *Expert Opinion on Biological Therapy*, 20(4):437–442, 2020.

[179] Seiji Arihiro, Haruo Ohtani, Manabu Suzuki, Masahiro Murata, Chieko Ejima, Motoji Oki, Yoshitaka Kinouchi, Kouhei Fukushima, Iwao Sasaki, Shiro Nakamura, Takayuki Matsumoto, Akira Torii, Gotaro Toda, and Hiroshi Nagura. Differential expression of mucosal addressin cell adhesion molecule-1 (MAdCAM-1) in ulcerative colitis and Crohn's disease. *Pathology International*, 52(5-6):367–374, 2002. ISSN 13205463. doi: 10.1046/j.1440-1827.2002.01365.x.

[180] Kevin Cowtan. Phase Problem in X -ray Crystallography, and its Solution. *eLS*, pages 1–5, 2003. doi: 10.1038/npg.els.0002722.

[181] Robert W. Harrison. Phase problem in crystallography. *Journal of the Optical Society of America A*, 10(5):1046, 1993. ISSN 1084-7529. doi: 10.1364/josaa.10.001046.

[182] Garry Taylor. The phase problem. *Acta Crystallographica - Section D Biological Crystallography*, 59(11):1881–1890, 2003. ISSN 09074449. doi: 10.1107/S0907444903017815.

[183] R. W. Gerchberg and W. O. Saxton. Practical Algorithm for the Determination of Phase From Image and Diffraction Plane Pictures. *Optik*, 35(2):237–250, 1972. ISSN 0030-4026.

[184] Tatiana Latychevskaia, Jean Nicolas Longchamp, and Hans Werner Fink. When holography meets coherent diffraction imaging. *Digital Holography and Three-Dimensional Imaging, DH 2012*, 2012. doi: 10.1364/dh.2012.dw1c.3.

[185] Tatiana Latychevskaia and Jan Pieter Abrahams. Inelastic scattering and solvent scattering reduce dynamical diffraction in biological crystals. *Acta Crystallographica Section B: Structural Science, Crystal Engineering and Materials*, 75:523–531, 2019. ISSN 20525206. doi: 10.1107/S2052520619009661.

[186] Dennis Gabor. Theory of Electron Interference Experiments. *Reviews of Modern Physics*, 28(3):260–276, 1956.

[187] Yong Keun Park, Christian Depeursinge, and Gabriel Popescu. Quantitative phase imaging in biomedicine. *Nature Photonics*, 12(10):578–589, 2018. ISSN 17494893. doi: 10.1038/s41566-018-0253-x. URL http://dx.doi.org/10.1038/s41566-018-0253-x.

[188] Lukas Wesemann, Jon Rickett, Jingchao Song, Jieqiong Lou, Elizabeth Hinde, Timothy J. Davis, and Ann Roberts. Nanophotonics enhanced coverslip for phase imaging in biology. *Light: Science and Applications*, 10(1), 2021. ISSN 20477538. doi: 10.1038/s41377-021-00540-7. URL http://dx.doi.org/10.1038/s41377-021-00540-7.

[189] Brent Fultz and James M. Howe. *Transmission Electron Microscopy and Diffractometry of Materials.* 2008. ISBN 9783540738855. doi: 10.1007/978-3-662-04901-3_2.

[190] Akira Tonomura. Applications of electron holography. *Reviews of Modern Physics*, 59(3):639–669, 1987. ISSN 00346861. doi: 10.1103/RevModPhys.59.639.

[191] Rafal E. Dunin-Borkowski, Takeshi Kasama, Alexander Wei, Steven L. Tripp, Martin J. Hytch, Etienne Snoeck, Richard J. Harrison, and Andrew Putnis. Off-axis electron holography of magnetic nanowires and chains, rings, and planar arrays of magnetic nanoparticles. *Microscopy Research and Technique*, 64(5-6):390–402, 2004. ISSN 1059910X. doi: 10.1002/jemt.20098.

[192] Tatiana Latychevskaia. Iterative phase retrieval for digital holography. *Journal of the Optical Society of America*, 36(D31), 2019. ISSN 23318422.

[193] Tatiana Latychevskaia and Hans-Werner Fink. Solution to the Twin Image Problem in Holography. *Physical Review Letters*, 98(233901), 2007. doi: 10.1103/PhysRevLett.98.233901.

[194] Gabriel Koren, François Polack, and Denis Joyeux. Iterative algorithms for twin-image elimination in in-line holography using finite-support constraints. *Journal of the Optical Society of America A*, 10(3):423, 1993. ISSN 1084-7529. doi: 10.1364/josaa.10.000423.

[195] J. R. Fienup. Phase retrieval algorithms: a comparison. *Applied Optics*, 21(15): 2758, 1982. ISSN 0003-6935. doi: 10.1364/ao.21.002758.

[196] T. R. Crimmins and J. R. Fienup. Phase Retrieval for Functions With Sufficiently Disconnected Support. 4(1):124–134, 1983.

[197] Jiaqi Hu, Qi Li, and Yi Zhou. Support-domain constrained phase retrieval algorithms in terahertz in-line digital holography reconstruction of a nonisolated amplitude object. *Applied Optics*, 55(2):379, 2016. ISSN 0003-6935. doi: 10.1364/ao.55.000379.

[198] Tatiana Latychevskaia and Hans-Werner Fink. Reconstruction of purely absorbing, absorbing and phase-shifting, and strong phase-shifting objects from their single-shot in-line holograms. *Applied Optics*, 54(13):3925, 2015. ISSN 0003-6935. doi: 10.1364/ao.54.003925.

[199] John R. Taylor. *Scattering Theory The Quantum Theory of Nonrelativistic Collisions.* 1972.

[200] Anna-Lena Robisch. *Phase retrieval for object and probe in the optical near-field.* 2015. ISBN 9783863952525. URL https://ediss.uni-goettingen.de/handle/11858/00-1735-0000-0028-868D-F.

[201] Aimin Yan and Xizeng Wu. Robustness of phase retrieval methods in x-ray phase contrast imaging : A comparison. *Medical Physics*, 38(9):5073–5080, 2011. doi: 10.1118/1.3618731.

[202] J. M. Cowley and A. F. Moodie. The Scattering of Electrons by Atoms and Crystals. I. A New Theoretical Approach. *Acta Crystallographica*, 1957. doi: 10.1107/S0365110X57002194.

[203] P. Goodman and A. F. Moodie. Numerical evaluations of N-beam wave functions in electron scattering by the multi-slice method. *Acta Crystallographica Section A*, 30(2):280–290, 1974. ISSN 16005724. doi: 10.1107/S056773947400057X.

[204] Earl J. Kirkland. *Advanced Computing in Electron Microscopy*. Springer, New York, 2010. ISBN 9781441965325.

[205] Wouter Van Den Broek and Christoph T. Koch. Method for retrieval of the three-dimensional object potential by inversion of dynamical electron scattering. *Physical Review Letters*, 109(24):1–5, 2012. ISSN 00319007. doi: 10.1103/PhysRevLett.109. 245502.

[206] L. M. Peng, S.L. Dudarev, and M. J. Whelan. *High-Energy Electron Diffraction and Microscopy*. Oxford Science Publications, 2003.

[207] Tatiana Latychevskaia. Three-Dimensional Structure from Single Two-Dimensional Diffraction Intensity Measurement. *Physical Review Letters*, 127(6):63601, 2021. ISSN 10797114. doi: 10.1103/PhysRevLett.127.063601. URL https://doi.org/10. 1103/PhysRevLett.127.063601.

[208] J. A. Ratcliffe. Some aspects of diffraction theory and their application to the ionosphere. *Reports on Progress in Physics*, 19(1):188–267, 1956. ISSN 00344885. doi: 10.1088/0034-4885/19/1/306.

[209] J.W. Goodman. *Introduction to Fourier Optics*. 2004.

[210] Soung Hun Roh, Corey F. Hryc, Hyun Hwan Jeong, Xue Fei, Joanita Jakana, George H. Lorimer, and Wah Chiu. Subunit conformational variation within individual GroEL oligomers resolved by Cryo-EM. *Proceedings of the National Academy of Sciences of the United States of America*, 114(31):8259–8264, 2017. ISSN 10916490. doi: 10.1073/pnas.1704725114.

[211] Tatiana Latychevskaia, Fabian Gehri, and Hans-Werner Fink. Depth-resolved holographic reconstructions by three-dimensional deconvolution. *Optics Express*, 18(21):22527, 2010. ISSN 1094-4087. doi: 10.1364/oe.18.022527.

[212] Tatiana Latychevskaia. Three-dimensional volumetric deconvolution in coherent optics and holography. *Applied Optics*, 60(5):1304, 2021. ISSN 1559-128X. doi: 10.1364/ao.412736.

[213] Francesc Salvat, Aleksander Jablonski, and Cedric J Powell. ELSEPA- Dirac partial-wave calculation of elastic scattering of electrons and positrons by atoms, positive ions and molecules. *Computer Physics Communications*, 165:157–190, 2005. doi: 10.1016/j.cpc.2004.09.006.

[214] D W Walker. Relativistic effects in low energy electron scattering from atoms. *Advances in Physics*, 20(85):257–323, 1971. doi: 10.1080/00018737100101251.

Bibliography

[215] Claude Cohen-Tannoudji, Bernard Diu, and Franck Laloë. *Quantenmechanik Band 1 und 2*. Berlin, 4. auflage edition, 2010.

[216] David J. Griffiths. *Introduction to Quantum Mechanics*. 1995.

[217] Ramamurti Shankar. *Principles of Quantum Mechanics*. New York, second edi edition, 1994. ISBN 0306447908.

[218] Aleksander Jablonski, Francesc Salvat, and Cedric J Powell. *NIST Electron Elastic-Scattering Cross-Section Database - Version 3.2*. National Institute of Standards and Technology, Gaithersburg, MD, 2010.

[219] Tatiana Latychevskaia and Hans-Werner Fink. Simultaneous reconstruction of phase and amplitude contrast from a single holographic record. *Optics Express*, 17 (13):10697, 2009. ISSN 1094-4087. doi: 10.1364/oe.17.010697.

[220] M. R. McCartney and M. Gajdardziska-Josifovska. Absolute measurement of normalized thickness, t/λ_i, from off-axis electron holography. *Ultramicroscopy*, 53 (3):283–289, 1994. ISSN 03043991. doi: 10.1016/0304-3991(94)90040-X.

[221] Tatiana Latychevskaia, Flavio Wicki, Conrad Escher, and Hans Werner Fink. Imaging the potential distribution of individual charged impurities on graphene by low-energy electron holography. *Ultramicroscopy*, 182:276–282, 2017. ISSN 18792723. doi: 10.1016/j.ultramic.2017.07.019.

[222] David Nečas and Petr Klapetek. Gwyddion: an open-source software for SPM data analysis. *Cent. Eur. J. Phys.*, 10(1):181–188, 2012. URL http://gwyddion.net/.

[223] Martin K. Safo and Donald J. Abraham. The X-ray structure determination of bovine carbonmonoxy hemoglobin at 2.1 Å resoultion and its relationship to the quaternary structures of other hemoglobin crystal forms. *Protein Science*, 10(6): 1091–1099, 2001. ISSN 09618368. doi: 10.1110/ps.48301.

[224] Gordon W. Bushnell, Gordon V. Louie, and Gary D. Brayer. High-resolution three-dimensional structure of horse heart cytochrome c. *Journal of Molecular Biology*, 214(2):585–595, 1990. ISSN 00222836. doi: 10.1016/0022-2836(90)90200-6.

[225] Sai Rohit Guntupalli, Zhuang Li, Leifu Chang, Bryce V. Plapp, and Ramaswamy Subramanian. Cryo-Electron Microscopy Structures of Yeast Alcohol Dehydrogenase. *Biochemistry*, 60(9):663–677, 2021. ISSN 15204995. doi: 10.1021/acs.biochem.0c00921.

[226] Alberto Bartesaghi, Doreen Matthies, Soojay Banerjee, Alan Merk, and Sriram Subramaniam. Structure of β-galactosidase at 3.2-Å resolution obtained by cryo-electron microscopy. *Proceedings of the National Academy of Sciences*, 111(32): 11709–11714, 2014. doi: 10.1073/pnas.1402809111.

[227] Manajit Hayer-Hartl, Andreas Bracher, and F. Ulrich Hartl. The GroEL-GroES Chaperonin Machine: A Nano-Cage for Protein Folding. *Trends in Biochemical Sciences*, 41(1):62–76, 2016. ISSN 13624326. doi: 10.1016/j.tibs.2015.07.009. URL http://dx.doi.org/10.1016/j.tibs.2015.07.009.

[228] M. P. Seah and W. A. Denc. Quantitative Electron Spectroscopy of Surface: A Standard Data Base for Electron Inelastic Mean Free Paths in Solids. *Surface and Interface Analysis*, 1(1):2, 1979.

[229] Alberto Bartesaghi, Cecilia Aguerrebere, Veronica Falconieri, Soojay Banerjee, Lesley A. Earl, Xing Zhu, Nikolaus Grigorieff, Jacqueline L.S. Milne, Guillermo Sapiro, Xiongwu Wu, and Sriram Subramaniam. Atomic Resolution Cryo-EM Structure of β-Galactosidase. *Structure*, 26(6):848–856.e3, 2018. ISSN 18784186. doi: 10.1016/j.str.2018.04.004. URL https://doi.org/10.1016/j.str.2018.04.004.

[230] Minji Wang, Tsz Pui Lai, Li Wang, Hongmin Zhang, Nan Yang, Peter J. Sadler, and Hongzhe Sun. "Anion clamp" allows flexible protein to impose coordination geometry on metal ions. *Chemical Communications*, 51(37):7867–7870, 2015. ISSN 1364548X. doi: 10.1039/c4cc09642h.

[231] PyMOL. The PyMOL Molecular Graphics System, Version 2.5, Schrödinger, LLC.

[232] Tim K Esser, B Jan, Paul Fremdling, Mark T Agasid, Adam Costin, Kyle Fort, Albert Konijnenberg, Alan Bahm, Alexander Makarov, Carol V Robinson, Justin L P Benesch, Lindsay Baker, Tanmay A M Bharat, Joseph Gault, and Stephan Rauschenbach. Mass-selective and ice-free cryo-EM protein sample preparation via native electrospray ion-beam deposition. *bioRxiv 2021.10.18.464782*, 2021.

[233] Christopher J Hogan, Brandon T Ruotolo, Carol V Robinson, Juan Fernandez, and De Mora. Tandem Differential Mobility Analysis-Mass Spectrometry Reveals Partial Gas-Phase Collapse of the GroEL Complex. *Journal of Physical Chemistry B*, 115:3614–3621, 2011.

[234] D. R. Hall, J. M. Hadden, G. A. Leonard, S. Bailey, M. Neu, M. Winn, and P. F. Lindley. The crystal and molecular structures of diferric porcine and rabbit serum transferrins at resolutions of 2.15 and 2.60 Å, respectively. *Acta Crystallographica Section D: Biological Crystallography*, 58(1):70–80, 2002. ISSN 09074449. doi: 10.1107/S0907444901017309.

[235] Swarup Dey, Chunhai Fan, Kurt V. Gothelf, Jiang Li, Chenxiang Lin, Longfei Liu, Na Liu, Minke A. D. Nijenhuis, Barbara Saccà, Friedrich C. Simmel, Hao Yan, and Pengfei Zhan. DNA origami. *Nature Reviews Methods Primers*, 2021. doi: 10.1038/s43586-020-00009-8.

[236] D. S. Auld and T. Bergman. Medium- and short-chain dehydrogenase/reductase gene and protein families: The role of zinc for alcohol dehydrogenase structure and function. *Cellular and Molecular Life Sciences*, 65(24):3961–3970, 2008. ISSN 1420682X. doi: 10.1007/s00018-008-8593-1.

[237] J. Günter Grossmann, Margarete Neu, Robert W. Evans, Peter F. Lindley, Helmut Appel, and Samar S. Hasnain. Metal-induced Conformational Changes in Transferrins. *Journal of Molecular Biology*, 299(3):585–590, 1993.

[238] Masaaki Hirose. The structural mechanism for iron uptake and release by transferrins. *Bioscience, Biotechnology and Biochemistry*, 64(7):1328–1336, 2000. ISSN 13476947. doi: 10.1271/bbb.64.1328.

[239] Philip D. Jeffrey, Maria C. Bewley, Ross T.A. MacGillivray, Anne B. Mason, Robert C. Woodworth, and Edward N. Baker. Ligand-induced conformational change in transferrins: Crystal structure of the open form of the N-terminal half-molecule of human transferrin. *Biochemistry*, 37(40):13978–13986, 1998. ISSN 00062960. doi: 10.1021/bi9812064.

[240] Piyali Guha Thakurta, Debi Choudhury, Rakhi Dasgupta, and J. K. Dattagupta. Tertiary structural changes associated with iron binding and release in hen serum transferrin: A crystallographic and spectroscopic study. *Biochemical and Biophysical Research Communications*, 316(4):1124–1131, 2004. ISSN 0006291X. doi: 10.1016/j.bbrc.2004.02.165.

[241] Mariko Kitahara, Satoshi Fudo, Tomoki Yoneda, Michiyoshi Nukaga, and Tyuji Hoshino. Anisotropic Distribution of Ammonium Sulfate Ions in Protein Crystallization. *Crystal Growth and Design*, 19(11):6136–6140, 2019. ISSN 15287505. doi: 10.1021/acs.cgd.9b00256.

[242] Michael B. Cammarata and Jennifer S. Brodbelt. Structural characterization of holo- and apo-myoglobin in the gas phase by ultraviolet photodissociation mass spectrometry. *Chemical Science*, 6(2):1324–1333, 2015. ISSN 20416539. doi: 10.1039/c4sc03200d. URL http://dx.doi.org/10.1039/C4SC03200D.

[243] N. Vats, S. Rauschenbach, W. Sigle, S. Sen, S. Abb, A. Portz, M. Dürr, M. Burghard, P. A. Van Aken, and K. Kern. Electron microscopy of polyoxometalate ions on graphene by electrospray ion beam deposition. *Nanoscale*, 10(10):4952–4961, 2018. ISSN 20403372. doi: 10.1039/c8nr00402a.

[244] Tatiana Latychevskaia, Flavio Wicki, Jean Nicolas Longchamp, Conrad Escher, and Hans Werner Fink. Direct Observation of Individual Charges and Their Dynamics on Graphene by Low-Energy Electron Holography. *Nano Letters*, 16(9):5469–5474, 2016. ISSN 15306992. doi: 10.1021/acs.nanolett.6b01881.

[245] Ian T. Young. Image Fidelity: Characterizing the Imaging Transfer Function. *Methods in Cell Biology*, 30:1–45, 1989.

[246] W. Wallace, L. H. Schaefer, and J. R. Swedlow. A workingperson's guide to deconvolution in light microscopy. *BioTechniques*, 31(5):1076–1097, 2001. ISSN 07366205. doi: 10.2144/01315bi01.

[247] James G. McNally, Tatiana Karpova, John Cooper, and José Angel Conchello. Three-dimensional imaging by deconvolution microscopy. *Methods: A Companion to Methods in Enzymology*, 19(3):373–385, 1999. ISSN 10462023. doi: 10.1006/meth.1999.0873.

[248] Raymond Gold. An Iterative Unfolding Method for Response Matrices. Technical report, Argonne National Laboratory, 1964.

[249] Tatiana Latychevskaia and Hans-Werner Fink. Holographic time-resolved particle tracking by means of three-dimensional volumetric deconvolution. *Optics Express*, 22(17):20994, 2014. ISSN 1094-4087. doi: 10.1364/oe.22.020994.

[250] Autoridad Nacional del Servicio Civil. *Electron Tomography Methods for Three-Dimensional Visualization of Structures in the Cell.* Springer, New York, NY, 2006. ISBN 9780387312347.

[251] Avinash C. Kak and Malcolm Slaney. *Principles of Computerized Tomographic Imaging.* IEEE Press, 1988.

[252] Johann Radon. Über die Bestimmung von Funktionen durch ihre Integralwerte längs gewisser Mannigfaltigkeiten. *Berichte über die Verhandlungen der Königlich-Sächsischen Akademie der Wissenschaften zu Leipzig, Mathematisch-Physische Klasse*, 1917.

[253] Martin J. Willemink and Peter B. Noël. The evolution of image reconstruction for CT—from filtered back projection to artificial intelligence. *European Radiology*, 29 (5):2185–2195, 2019. ISSN 14321084. doi: 10.1007/s00330-018-5810-7.

[254] Jonathan M. Blackledge. *Digital Image Processing Mathematical and Computational Methods.* 2006.

[255] Aike Ruhlandt and Tim Salditt. Three-dimensional propagation in near-field tomographic X-ray phase retrieval. *Acta Crystallographica Section A: Foundations and Advances*, 72:215–221, 2016. ISSN 20532733. doi: 10.1107/S2053273315022469.

[256] L A Feldkamp. Practical cone-beam algorithm. *Journal of the Optical Society of America*, 1(6):612–619, 1984.

[257] Bruce D. Smith. Cone-beam tomography: recent advances and a tutorial review. *Optical Engineering*, 29(5):524, 1990. ISSN 00913286. doi: 10.1117/12.55621.

[258] Paul Müller, Mirjam Schürmann, and Jochen Guck. The Theory of Diffraction Tomography. *arXiv 1507.00466*, 2016.

[259] Paul Müller. *Optical Diffraction Tomography for single cells*. PhD thesis, Technische Universität Dresden, 2016.

[260] Vinoth Balasubramani, Arkadiusz Kuś, Han-Yen Tu, Chau-Jern Cheng, Maria Baczewska, Wojciech Krauze, and Małgorzata Kujawińska. Holographic tomography: Techniques and biomedical applications. *Applied Optics*, 60(10), 2021. ISSN 1559-128X. doi: 10.1364/ao.416902.

[261] Florian Charrière, Anca Marian, Frédéric Montfort, Jonas Kuehn, Tristan Colomb, Etienne Cuche, Pierre Marquet, and Christian Depeursinge. Cell refractive index tomography by digital holographic microscopy. *Optics Letters*, 31(2):178–180, 2006. ISSN 21622701. doi: 10.1364/dh.2008.dtuc3.

[262] I. Bergënd, C. Arfire, N. Pavillon, and C. Depeursinge. Diffraction tomography for biological cells imaging using digital holographic microscopy. *Laser Applications in Life Sciences*, 7376, 2010. ISSN 0277786X. doi: 10.1117/12.871061.

[263] Marco Stockmar, Peter Cloetens, Irene Zanette, Bjoern Enders, Martin Dierolf, Franz Pfeiffer, and Pierre Thibault. Near-field ptychography: Phase retrieval for inline holography using a structured illumination. *Scientific Reports*, 3:1–6, 2013. ISSN 20452322. doi: 10.1038/srep01927.

[264] A. M. Maiden, M. J. Humphry, and J. M. Rodenburg. Ptychographic transmission microscopy in three dimensions using a multi-slice approach. *Journal of the Optical Society of America A*, 29(8):1606, 2012. ISSN 1084-7529. doi: 10.1364/josaa.29. 001606.

[265] T. M. Godden, R. Suman, M. J. Humphry, J. M. Rodenburg, and A. M. Maiden. Ptychographic microscope for three-dimensional imaging. *Optics Express*, 22(10): 12513, 2014. ISSN 1094-4087. doi: 10.1364/oe.22.012513.

[266] Si Gao, Peng Wang, Fucai Zhang, Gerardo T. Martinez, Peter D. Nellist, Xiaoqing Pan, and Angus I. Kirkland. Electron ptychographic microscopy for three-dimensional imaging. *Nature Communications*, 8(1), 2017. ISSN 20411723. doi: 10.1038/s41467-017-00150-1. URL http://dx.doi.org/10.1038/s41467-017-00150-1.

[267] Peng Li and Andrew Maiden. Multi-slice ptychographic tomography. *Scientific Reports*, 8(1):1–10, 2018. ISSN 20452322. doi: 10.1038/s41598-018-20530-x. URL http://dx.doi.org/10.1038/s41598-018-20530-x.

Acknowledgements

Acknowledgements

Curriculum Vitae

Hannah Ochner
Katzenbachstraße 122
70563 Stuttgart
Hannah.Ochner@gmx.de

Date of birth: December 7, 1992   in Waiblingen, Germany

Education

since Sept 2017	**Doctoral Studies (PhD)** École Polytechnique Fédérale de Lausanne, Switzerland Doctoral School: EDPY Physics Research Assistant in the group of Prof. Klaus Kern, Max Planck Institute for Solid State Research Stuttgart
2015–2017	**Master Studies in Theoretical and Mathematical Physics** Ludwig-Maximilians-Universität München, Germany Final grade: 1,2 (very good, passed with high distinction)
2012–2015	**Bachelor Studies in Physics** Ludwig-Maximilians-Universität München, Germany Final grade: 1,13 (very good)
2012–2014	**Bachelor Studies in Philosophy** Ludwig-Maximilians-Universität München, Germany Final grade: 1,02 (very good)
Jun-Sept 2012	**Summer Internship** Max Planck Institute for Solid State Research Stuttgart Electrospray Ion Beam Deposition of Nanoparticles
2011–2012	**Bachelor Studies Physics and Philosophy** University of Nottingham, England
2003–2011	**Gymnasium (High School)** Friedrich-Schiller-Gymnasium Fellbach Final grade (Abitur) 1.0

Teaching Experience: 2014–2016: Tutor "Mathematics for Physicists I and III"

Scholarships: 2013-2017: Studienstiftung des deutschen Volkes

Publications

Hannah Ochner, Sven Szilagyi, Sabine Abb, Luigi Malavolti,
Stephan Rauschenbach, and Klaus Kern
Phase reconstruction of low-energy electron holograms of proteins
in preparation

Sven Szilagyi, **Hannah Ochner**, Moritz Edte, Sabine Abb, Jean-Nicolas Longchamp,
Luigi Malavolti, Stephan Rauschenbach, and Klaus Kern
*A novel low temperature low-energy electron holography microscope with in situ
electrospray ion beam deposition sample preparation*
in preparation

Hannah Ochner, Sven Szilagyi, Sabine Abb, Joseph Gault, Carol V. Robinson,
Luigi Malavolti, Stephan Rauschenbach, and Klaus Kern
*Low-energy electron holography imaging of conformational variability of single-antibody molecules
from electrospray ion beam deposition*
PNAS 118, e2112651118 (2021)

Nilesh Vats, Devendra S. Negi, Deobrat Singh, Wilfried Sigle, Sabine Abb, Suman Sen,
Sven Szilagyi, **Hannah Ochner**, Rajeev Ahuja, Klaus Kern,
Stephan Rauschenbach, and Peter A. van Aken
Catalyzing bond-dissociation in graphene via alkali-iodide molecules
Small, 2102037 (2021)

Lukas Krumbein, Kelvin Anggara, Martina Stella, Tomasz Michnowicz, **Hannah Ochner**,
Sabine Abb, Gordon Rinke, André Portz, Michael Dürr, Uta Schlickum, Andrew Baldwin,
Andrea Floris, Klaus Kern, and Stephan Rauschenbach
Fast Molecular Compression by a Hyperthermal Collision Gives Bond-Selective Mechanochemistry
Phys. Rev. Lett. 126, 056001 (2021)

Nilesh Vats, Yi Wang, Suman Sen, Sven Szilagyi, **Hannah Ochner**, Sabine Abb,
Marko Burghard, Wilfried Sigle, Klaus Kern, Peter A. van Aken, and Stephan Rauschenbach
Substrate-selective morphology of cesium-iodide clusters on graphene
ACS Nano 14, 4626 (2020)