Stony Brook University

X-ray Diffraction Microscopy: Computational Methods and Scanning-type Experiments

A Dissertation Presented

by

Jan Felix Steinbrener

to

The Graduate School

in Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

in

Physics

Stony Brook University

August 2010

Stony Brook University

The Graduate School

Jan Felix Steinbrener

We, the dissertation committee for the above candidate for the Doctor of Philosophy degree, hereby recommend acceptance of this dissertation.

Chris J. Jacobsen – Dissertation Advisor
Professor, Department of Physics and Astronomy

Alexandre G. Abanov – Chairperson of Defense
Professor, Department of Physics and Astronomy

Anand Sivaramakrishnan
Adjunct Professor, Department of Physics and Astronomy

Hanfei Yan
Associate Physicist, Brookhaven National Laboratory

This dissertation is accepted by the Graduate School.

Lawrence Martin
Dean of the Graduate School

Abstract of the Dissertation

X-ray Diffraction Microscopy: Computational Methods and Scanning-type Experiments

by

Jan Felix Steinbrener

Doctor of Philosophy

in

Physics

Stony Brook University

2010

X-ray Diffraction Microscopy (XDM) has been gaining in popularity for nanoscale imaging of biological and material science samples. Its high penetration depth (compared to electron microscopy) and its good dose efficiency (compared to its lens-based X-ray alternative) make it uniquely suited for imaging whole biological specimens, where radiation damage is a concern. Despite these advantages, XDM is still far from being a routine imaging tool. This is due to the computational challenge of reconstructing an image from recorded diffraction intensities as well as difficult-to-satisfy experimental requirements.

I address these challenges by improving on the computational methods and by implementing a more reliable experimental geometry for our existing diffraction microscope at the Advanced Light Source, Lawrence Berkeley Lab. First, a software library has been developed that streamlines the post-experiment processing of data and that improves on an important aspect of data analysis. Results

will be shown that illustrate the collective improvement to the reconstruction process. A modified version of a tool commonly used to assess the consistency of reconstructions is proposed and criteria of its validity are derived. Results show that it has improved utility for judging reconstruction quality. Second, a scanning-type experimental setup has been implemented for our existing diffraction microscope. Several possible geometries are discussed and preliminary results from recent experimental data are shown.

To my family.

Contents

List of Figures

List of Tables

Acknowledgements

After several years of graduate studies at Stony Brook, there are a number of people that I would like to thank for their contributions to my work and their support during stressful times.

I would like to start with my advisor, Chris Jacobsen. Thank you for giving me the opportunity to work in this exciting and diverse area of physics. Your motivation and your ability to put things in perspective have been a constant source of inspiration. I would like to thank the other members of the diffraction project: Johanna Nelson, Xiaojing Huang, and Joshua Turner. Thank you for all the fruitful discussions, and for making long hours at the beamline seem tolerable. Special thanks to Johanna for helping me debug my code; it's an ongoing process! I would like to thank Christian Holzner from our group for giving me the opportunity to join him for a great experiment at the APS. Thank you also for helping out during our beamtime and for many helpful conversations. I would also like to thank our new student, Rachel Mak, for coming out to California to help out during our beamtime. A big thank you goes to Sue Wirick; thank you for being a tremendous help, especially in the beginning of my graduate student career.

There are many former members of our group who I had the pleasure to work with. Andrew Stewart, Enju Lima and Huijie Miao helped me find my way into the diffraction project. Bjorg Larson, Holger Fleckenstein, Benjamin Hornberger and Mirna Lerotic provided many answers to my physics and programming questions.

I would like to thank the staff at the Advanced Light Source, Stefano Marchesini and David Shapiro, for their help as beamline scientists. Very special thanks to Janos Kirz who was involved in all phases of the experiments and provided countless useful hints. I would like to thank Rich Celestre and Tony Warwick for many helpful discussions and hands-on help with the experimental setup. Thank you to Ed Domning and Brian Smith for their help with the experimental control software.

I would also like to thank Pete Davis of Stony Brook for coating the beam attenuator windows and the machine shop at Stony Brook for many expedited

services.

On a personal level, I would like to thank my parents and my sister for their love and support during my many years abroad. Thank you for making the effort to visit me many times and for making me feel welcome at home. I am also very grateful to my wife, Heather. Thank you for all the happy times we have been having together and thanks for your support during stressful times.

Finally, I also wish to thank the Division of Materials Sciences and Engineering, Office of Basic Energy Sciences, at the Department of Energy for support of x-ray diffraction microscopy methods and instrumentation development under contract DE-FG02-07ER46128 and the National Institute for General Medical Services at the National Institutes for Health for support of the application of this method to biological imaging under contract 5R21EB6134.

Chapter 1

Introduction

1.1 Why X-rays?

Eukaryotic cells are complex biological systems. The dynamic processes involved in their life cycle pose many questions and are subject to intense research. High-resolution imaging of hydrated cells has been a powerful tool to study the interaction mechanisms between small functional substructures (proteins) that drive the evolution of the entire cell. Today, advanced fluorescence based visible light microscopy can deliver resolutions of up to 50 nm (see for example Hell [1]) while electron microscopy routinely achieves resolutions below 10 nm on thin sections or peripheral regions of whole hydrated cells. X-ray microscopy, using photons with wavelengths well below 10 nm, has the potential for higher spatial resolution than visible light microscopy. Furthermore, both x-ray and electron microscopy offer capabilities for elemental and chemical state mapping; however, this comes at the cost of damage to the specimen caused by the ionizing nature of the radiation. To minimize the effects of radiation damage, cryofixation of hydrated structures is used in electron and x-ray microscopy. It has been found [2] that the required radiation dose to image a small structure in ice depends strongly on the ice thickness, with x-rays offering lower dose than electrons if the thickness of the ice exceeds 500 nm. All together, this suggests that x-rays are uniquely suited for high resolution imaging of whole frozen-hydrated eukaryotic cells which typically are several micrometers in size.

1.2 X-ray Interactions with Matter

As primary interaction mechanisms of x-rays with matter I want to consider absorption, elastic scattering, and inelastic scattering; the latter is also referred

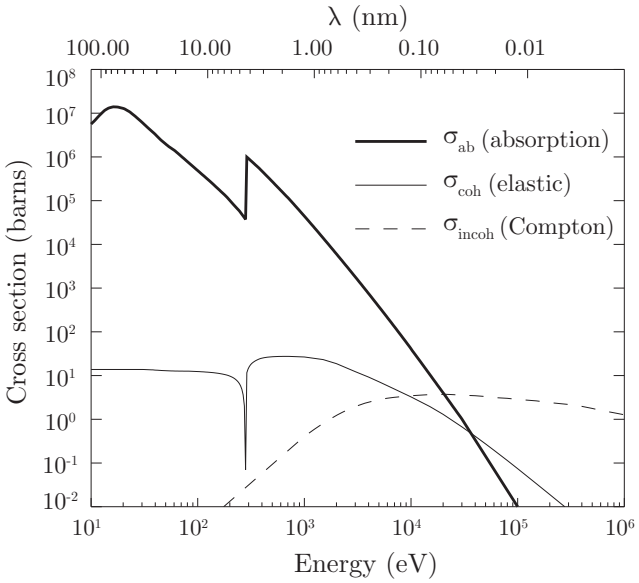

Figure 1.1: Cross sections for x-ray interactions with carbon atoms. The probability of Compton scattering below 1 keV is negligible compared to elastic scattering and absorption. Figure reprinted from Kirz et al. [3]

to as Compton scattering. Absorption occurs when the energy of an x-ray is fully transferred to an electron of an atom in the sample. The electron is then ejected, leaving an ionized atom behind. In elastic scattering, the wave field of the photon causes the bound electrons in the sample to oscillate. The oscillating electrons then reradiate the incoming photons in a new direction; the photon's energy is preserved in the process. Inelastic scattering transfers some momentum and energy of the photon to the electron. The photon scatters of in a different direction and with reduced energy. The total energy and momentum of photon, electron, and atom is conserved.

The probability for either event to occur is given by its cross sections σ. Figure 1.1 shows a plot of the cross sections of absorption σ_{ab}, elastic scattering σ_{coh}, and Compton scattering σ_{incoh} for a carbon atom. It can be seen that for the energies of interest (below 1 keV) inelastic scattering processes can be neglected and that absorption processes dominate. In what follows, I will therefore ignore all effects related to Compton scattering.

In the semi-classical picture, we can think of an atom as an N-times positively charged nucleus surrounded by a cloud of N negatively charged electrons. An incident electromagnetic wave will force the electrons to oscillate in the restoring potential of the nucleus. The equation of motion of the electrons

depends on the frequency of the incident field ω and their resonant frequency ω_r. Including dissipative forces (characterized by the damping ratio γ_r), we can write it as

$$m_e \frac{\mathrm{d}^2 \boldsymbol{x_r}}{\mathrm{d}t^2} + m_e \gamma_r \frac{\mathrm{d}\boldsymbol{x_r}}{\mathrm{d}t} + m_e \omega_r^2 \boldsymbol{x}_r = -e(\boldsymbol{E} + \boldsymbol{v}_r \times \boldsymbol{B}) \tag{1.1}$$

with solution

$$\boldsymbol{x}_r = \frac{e}{m_e} \frac{1}{\omega^2 - \omega_r^2 + i\gamma_r \omega} \boldsymbol{E} \tag{1.2}$$

$$\Rightarrow \boldsymbol{v}_r = \frac{e}{m_e} \frac{1}{\omega^2 - \omega_r^2 + i\gamma_r \omega} \frac{\mathrm{d}\boldsymbol{E}}{\mathrm{d}t}, \tag{1.3}$$

where I have dropped the term $\boldsymbol{v}_r \times \boldsymbol{B}$ on the right hand side of Eq. 1.1 and assumed that the system will oscillate with the frequency of the driving force ω. How do these oscillations affect the wavefield traveling through a material made up of many atoms? From Maxwell equations we can derive the wave equation assuming no net charge density ρ as

$$\left(\frac{\partial^2}{\partial t^2} - c^2 \nabla^2 \right) \boldsymbol{E}(\boldsymbol{r}, \boldsymbol{t}) = -\frac{1}{\epsilon_0} \frac{\partial \boldsymbol{J}_T(\boldsymbol{r}, t)}{\partial t}, \tag{1.4}$$

where $\boldsymbol{J}_T$ is the transverse current density; since the electric field vector $\boldsymbol{E}$ is perpendicular to the direction of propagation of the wave, only the transverse component of the current density matters. It can be calculated from the sum over all oscillating electrons as

$$\boldsymbol{J}_T(\boldsymbol{r}, t) = -e n_a \sum_r h_r \boldsymbol{v}_r(\boldsymbol{r}, t), \tag{1.5}$$

where n_a is the density of atoms in the material and h_r the number of electrons of an atom that have the resonance frequency ω_r; the factor h_r is referred to as oscillator strength. Plugging Eqs. 1.3 and 1.5 into Eq. 1.4 and rearranging, we get

$$\left(\frac{\partial^2}{\partial t^2} - \frac{c^2}{n^2(\omega)} \nabla^2 \right) \boldsymbol{E}(\boldsymbol{r}, \boldsymbol{t}) = 0 \tag{1.6}$$

which has the form of a wave equation again. The factor

$$\begin{aligned} n(\omega) &= \left[1 - \frac{e^2 n_a}{\epsilon_0 m_e} \sum_r \frac{h_r}{\omega^2 - \omega_r^2 + i\gamma_r \omega}\right]^{1/2} \\ &\simeq 1 - \frac{e^2 n_a}{2\epsilon_0 m_e} \sum_r \frac{h_r}{\omega^2 - \omega_r^2 + i\gamma_r \omega} \end{aligned} \tag{1.7}$$

is the complex index refraction of the material. The approximation in the last step is valid when the second term under the square root is much smaller than one. This is the case for x-rays where ω^2 is large compared to ${}^{e^2 n_a}/_{\epsilon_0 m_e}$. It is customary [3] to define a complex oscillator density f as

$$f(\omega) = f_1(\omega) + i f_2(\omega)$$

such that

$$n(\omega) = 1 - \frac{e^2 n_a}{2\epsilon_0 m_e} \left[f_1(\omega) + i f_2(\omega)\right]. \tag{1.8}$$

The quantity $f(\omega)$ is also referred to as complex atomic scattering factor. Its real and imaginary parts are related to the cross sections for absorption and elastic scattering by

$$\sigma_{\text{ab}} = 2 r_e \lambda f_2, \text{ and} \tag{1.9}$$

$$\sigma_{\text{coh}} = \frac{8}{3} \pi r_e^2 |f_1 + i f_2|^2, \tag{1.10}$$

where r_e is the classical radius of the electron ($r_e = 2.82$ fm) and λ the wavelength [3]. A description more common to x-ray physics [4] can be obtained when we express the complex index of refraction as

$$n = 1 - \delta - i\beta, \tag{1.11}$$

where for x-rays both δ and β are small. The relationship between δ and β and the complex electron density f is then given by

$$\delta = \frac{e^2 n_a}{2\epsilon_0 m_e} f_1(\omega), \text{ and} \tag{1.12}$$

$$\beta = \frac{e^2 n_a}{2\epsilon_0 m_e} f_2(\omega). \tag{1.13}$$

Using Eq. 1.11, the solution to the wave equation in a homogeneous material

with index of refraction n (Eq. 1.6) is given by

$$\psi(z) = \psi_0 e^{-inkz} = \psi_0 e^{-ikz} e^{-\beta kz} e^{i\delta kz}, \tag{1.14}$$

where I have assumed a wave traveling in z-direction and dropped the harmonic term. We now see that relative to propagation in vacuum (first factor), the wave gets attenuated (second factor) and phase shifted (third factor). The intensity of a wave traveling through such a medium can then be calculated as

$$I(z) = \psi(z) \cdot \psi^*(z) = \psi_0^2 e^{-2\beta kz}. \tag{1.15}$$

1.3 Synchrotron-based X-ray Microscopy

Methods used for generating x-ray beams have evolved considerably over the recent years. An early source that is still popular today are X-ray tubes. They produce x-rays by bombarding a target anode with energetic electrons from a source cathode. As the electrons are decelerated in the anode they produce *Bremsstrahlung* and a fluorescence peak, characteristic of the material of the anode. Nowadays, x-ray tubes find widespread use for low-resolution imaging applications in the field of medicine and material science; however, they are not suited for high resolution imaging due to their low brightness.

A more powerful x-ray source is a synchrotron in which electrons are accelerated to highly relativistic speeds and forced on a circular trajectory with the help of bending magnets. As the electrons are traveling through a bending magnet, they exhibit a constant transverse acceleration which gives rise to synchrotron radiation. The characteristics of bending magnet radiation are a long continuous spectral range of the emitted radiation and the high flux output measured in photons per second, per bandwidth, and per solid angle. Alternative sources to bending magnet radiation at a synchrotron are insertion devices: undulators and wigglers. Both consist of a number of pairs of magnets arranged in an alternating fashion along the trajectory of the electrons such that the electrons oscillate transversely to their direction of travel. Due to the highly relativistic motion of the electrons, the emitted radiation is strongly collimated in the forward direction. A characteristic of undulator radiation is that its spectrum has discrete peaks; that is, a fundamental energy and higher harmonics. It also is of higher brightness (flux per source area) than bending magnet radiation and thus provides more coherent light. The spectrum can be shifted to higher or lower energies by increasing or decreasing the gap between the two opposing magnetic structures. A detailed mathematical treatment of synchrotron radiation and radiation from insertion devices can be found for

example in Wiedemann [5].

1.3.1 The X-ray Diffraction Microscope at Beamline 9.0.1 of the Advanced Light Source

Our existing x-ray diffraction microscope is installed at beamline 9.0.1 of the Advanced Light Source at Lawrence Berkeley National Lab[1]. Since high resolution x-ray diffraction microscopy requires lots of coherent flux, our microscope is fed by the undulator U100[2]. Based on its design parameters, we adjust the gap to maximize the flux of the third harmonic at our target energy (around 520 eV or 750 eV). To further increase the temporal coherence of the beam, an off-axis zone plate monochromator has been included in the beamline. Its initial design was conceived by Howells et al. [6]. A schematic of the setup is shown in Fig. 1.2. An off-axis segment of a full zone plate is illuminated by the

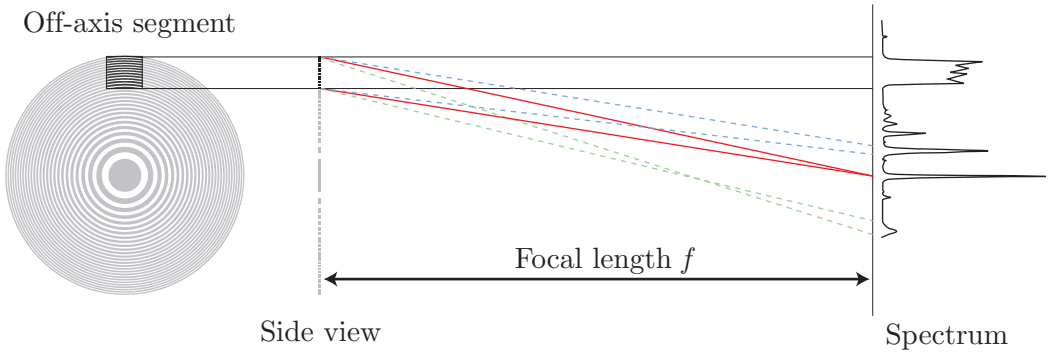

Figure 1.2: Schematic of the zone plate monochromator. An off-axis segment of a full zone plate focuses the incoming undulator radiation based on its energy. The peaks in the output spectrum can be identified by the undulator harmonic and zone plate focal order that produced them. Figure adapted from Lima [7].

radiation coming from the undulator. It focuses the different harmonics of the undulator radiation based on their energy. The output spectrum is comprised of different undulator harmonics and zone plate focal orders. The first order focal length of a zone plate can be calculated as

$$f(\lambda) = \frac{D_{\text{zp}} \cdot d_{r_N}}{\lambda}, \tag{1.16}$$

[1] http://www.als.lbl.gov/
[2] http://www.als.lbl.gov/als/curves/U10.html

where D_{ZP} is the diameter of the full zone plate, and d_{r_N} is the width of the outermost zone. We select a peak in the spectrum by placing a pinhole in the focal plane of the desired energy. The resolving power of the zone plate monochromator $\lambda/\Delta\lambda$ has been estimated to be about 500 [6].

The spatial coherence of the radiation that is transmitted through the pinhole can be calculated by the van Cittert-Zernike theorem (VCZ). It says that the complex degree of coherence between two points in a plane downstream of an incoherent source is equal to the normalized Fourier transform of the intensity function of the source if the distance between the source and the two points is much larger than the linear dimension of the source and the separation between the two points. The Fourier transform of a pinhole is an Airy pattern. Note that VCZ assumes an incoherent source which is not true for undulator radiation that can be considered partially coherent. A more exact analysis can be done by using for example the Gauss-Schell model. The coherence width can also be determined experimentally by observing the interference pattern of several pairs of pinholes with different separation distances that are illuminated with the beam from the pinhole. A detailed description of both theoretical and experimental analyses can be found in Miao [8]. The experimental results suggest that the coherence width of our illumination in the sample plane is around 6 μm in both the horizontal and vertical direction when using a 5 μm monochromator pinhole.

The x-rays will then enter the chamber that holds the sample and the detector. The original chamber was designed by Beetz et al. [9]. A main feature was the incorporation of a cryo-tilt stage that is used in electron microscopy as sample mount. The chamber received a major upgrade in the fall of 2008 with the help of the Experimental Systems Group at the Advanced Light Source. The upgrade retained the original sample mounting scheme but replaced the motor stages and extended the vacuum enclosure to accommodate a system of rails that hold all the optics involved in the experiment. The advantage of this is that the optics can be aligned to the optical path independently of the sample. An overview of the internal hardware is shown in Fig. 1.3 Two additional features of the upgraded chamber that are not shown in the drawing are an in-vacuum visible light microscope (VLM) and a cold shield for the sample called anti-contamination device (ACD). The VLM consists of an in-vacuum objective and light source and an out of vacuum CCD camera. It can be used to identify regions of interest on a sample that is already mounted in the chamber and has been described in detail in Nelson [10]. The ACD consists of a cold shield that surrounds the sample and that is cooled by an out of vacuum liquid nitrogen reservoir that is connected to the shield by a cold path. The ACD prevents that residual moisture in the vacuum condenses

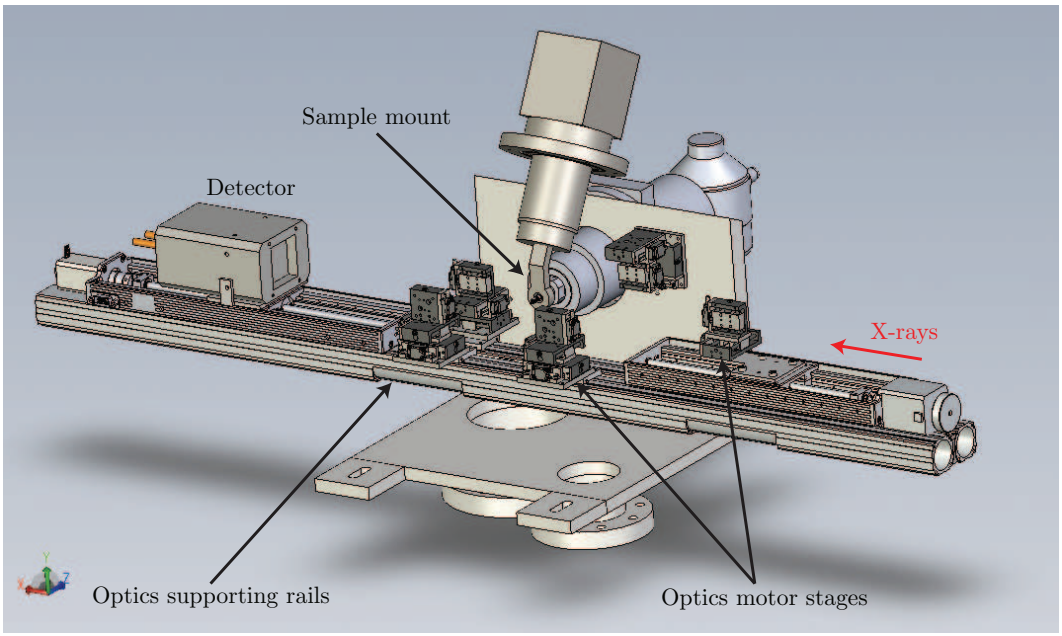

Figure 1.3: Overview over internal features of the diffraction chamber after the upgrade in fall 2008. The original sample mounting scheme was retained. All the optics are now mounted on a pair of rails which can be aligned to the x-rays independently of the sample. Solidworks drawing courtesy of Rich Celestre.

on the cold sample. It has been described in detail by Huang [11].

1.4 Contributions of This Dissertation

The work of this dissertation was carried out in the setting of a research group. The particular contributions I have made to the group's activities include the following:

- Leading the group's experiments in scanning x-ray diffraction microscopy.
- Modifying the apparatus of ALS beamline 9.0.1 as described in Sec. 4.1
- Writing an automated merging program (`AMP`) for diffraction data as described in Sec. 3.3
- Developing a common file format for diffraction imaging as explained in Sec. 3.1
- Developing a platform independent subroutine library with parallel processing capabilities for diffraction data analysis tools (Sec. 3.2)
- Writing a platform independent reconstruction code based on the subroutine library
- Writing a 3D data assembly program based on the subroutine library
- Contributing to a shared code archive of common post-reconstruction data analysis routines
- Developing improved methods to evaluate reconstructions, such as the wPRTF (Sec. 3.4)
- Programming a reconstruction code for scanning x-ray diffraction microscopy
- Analysis of SXDM data taken during an experiment at the APS (Sec. 4.2)
- Adapting the existing client-side experimental control software to work with a new server-side control software as part of the fall 2008 upgrade to our diffraction chamber

List of Publications

So far, my work as a PhD student has resulted in one lead-author publication that is currently under review. I have also coauthored several other publications from our group:

- **Jan Steinbrener**, Johanna Nelson, Xiaojing Huang, Stefano Marchesini, David Shapiro, Joshua Turner, and Chris Jacobsen, "High Fidelity Reconstructions for X-ray Diffraction Microscopy", *under review.*

- Johanna Nelson, Xiaojing Huang, **Jan Steinbrener**, David Shapiro, Janos Kirz, Aaron Neiman, Josh Turner and Chris Jacobsen, "High Resolution x-ray diffraction microscopy of specifically labelled yeast cells", *Proceedings of the National Academy of Science*, **107**, 7235, (2010).

- Xiaojing Huang, Johanna Nelson, Janos Kirz, Enju Lima, Stefano Marchesini, Huijie Miao, Aaron Neiman, David Shapiro, **Jan Steinbrener**, Andrew Stewart, Joshua Turner and Chris Jacobsen, "Soft X-ray diffraction microscopy of a frozen hydrated yeast cell", *Physical Review Letters*, **103**, 198101, (2009).

- Xiaojing Huang, Huijie Miao, **Jan Steinbrener**, Johanna Nelson, David Shapiro, Andrew Stewart, Joshua Turner and Chris Jacobsen, "Signal-to-noise and radiation exposure considerations in conventional and diffraction x-ray microscopy", *Optics Express*, **17**, 13541, (2009).

- Huijie Miao, Kenneth Downing, Xiaojing Huang, Chris Jacobsen, Janos Kirz, Stefano Marchesini, Johanna Nelson, David Shapiro, **Jan Steinbrener** and Andrew Stewart, "Cryo diffraction microscopy: ice condition and finite supports". X-ray Microscopy: Proceedings of the 9th International Conference, Zurich, Switzerland. *Journal of Physics: Conference Series (IOP)*, **186**, 012055, (2009).

Chapter 2

Basic Concepts in X-ray Diffraction Microscopy

In this chapter I will derive some basic results of diffraction theory (Sec. 2.1) and describe the experimental requirements and data analysis techniques of both diffraction microscopy modalities that are the subject of this dissertation, namely full-field X-ray Diffraction Microscopy (XDM, Sec. 2.2) and Scanning X-ray Diffraction Microscopy (SXDM or Ptychography, Sec. 2.3). I will talk about common tools to assess quality and resolution of resulting images (Sec. 2.4) and compare the two diffraction-based techniques with their lens-based alternatives (Sec. 2.5). Finally, in Section 2.6, I will motivate my work (described in the following chapters) based on the information presented here.

2.1 Scalar theory of diffraction

In Sec. 1.2 we described the basic interactions of x rays with matter in terms of the complex index of refraction. Now we will put it in the context of diffraction by deriving a relationship between an object [represented by its complex index of refraction $n(\boldsymbol{r},\omega)$] and the scattered complex wavefield $U^{\text{sc}}(\boldsymbol{r},\omega)$ that occurs at some plane downstream of the object when it is illuminated with light of a certain wavelength $U^{\text{inc}}(\boldsymbol{r},\omega)$. The following derivation is based on Ch. 13 of Born and Wolf [12].

The basic assumptions are that we have a monochromatic field $\propto e^{-i\omega t}$, incident on a linear, isotropic, nonmagnetic medium that occupies a finite domain volume V which contains no sources $(\rho,\boldsymbol{j})$, and finally, that the dielectric constant of the material $\epsilon(\boldsymbol{r})$ varies on length scales $\gg\lambda$.

Based on these assumptions we can derive a wave equation from the Maxwell equations that in the scalar approximation (*i. e.* no mixing in the components)

reduces to

$$\nabla^2 U(\boldsymbol{r},\omega) + k^2 U(\boldsymbol{r},\omega) = -4\pi F(\boldsymbol{r},\omega) U(\boldsymbol{r},\omega), \tag{2.1}$$

where $U(\boldsymbol{r})$ denotes the total wavefield at point $\boldsymbol{r}$, and the scattering potential (or "object function") $F(\boldsymbol{r},\omega)$ is given by

$$F(\boldsymbol{r},\omega) = \frac{1}{4\pi} k^2 [n^2(\boldsymbol{r},\omega) - 1]. \tag{2.2}$$

The total wavefield can be written as a superposition of incident and scattered wavefield in the form

$$U(\boldsymbol{r},\omega) = U^{\mathrm{inc}}(\boldsymbol{r},\omega) + U^{\mathrm{sc}}(\boldsymbol{r},\omega). \tag{2.3}$$

Since the incident wavefield satisfies the Helmholtz equation, Eq. 2.1 reduces to

$$(\nabla^2 + k^2) U^{\mathrm{sc}}(\boldsymbol{r},\omega) = -4\pi F(\boldsymbol{r},\omega) U(\boldsymbol{r},\omega). \tag{2.4}$$

With the Green's function of the Helmholtz operator

$$\left(\nabla^2 + k^2\right) G(\boldsymbol{r} - \boldsymbol{r}', \omega) = -4\pi \delta^{(3)}(\boldsymbol{r} - \boldsymbol{r}')$$

and with Green's theorem ignoring contributions of the resulting surface integral, we may write from Eq. 2.4

$$U^{sc}(\boldsymbol{r},\omega) = \int_V F(\boldsymbol{r}',\omega) U(\boldsymbol{r}',\omega) G(\boldsymbol{r} - \boldsymbol{r}', \omega) \mathrm{d}^3 r'. \tag{2.5}$$

Ignoring contributions from the surface integral is possible if we assume that the scattered field behaves like a spherical wave for $R \to \infty$. We therefore choose the outgoing free-space Green's function

$$G(\boldsymbol{r} - \mathbf{r}', \omega) = \frac{e^{ik|\boldsymbol{r} - \mathbf{r}'|}}{|\boldsymbol{r} - \mathbf{r}'|} \tag{2.6}$$

as solution to the Helmholtz equation which when plugged into Eq. 2.5 leads to

$$U^{sc}(\boldsymbol{r},\omega) = \int_V F(\boldsymbol{r}',\omega) U(\boldsymbol{r}',\omega) \frac{e^{ik|\boldsymbol{r} - \mathbf{r}'|}}{|\boldsymbol{r} - \mathbf{r}'|} \mathrm{d}^3 r'. \tag{2.7}$$

This represents a solution to the scattered wavefield we are interested in; however, it is not immediately solvable since one can notice that the scattered wavefield also appears in the integrand of the right hand side (see Eq. 2.3). Approximations are therefore necessary. The two most common ones are the Born and the Rytov approximations. Since the diffraction imaging techniques

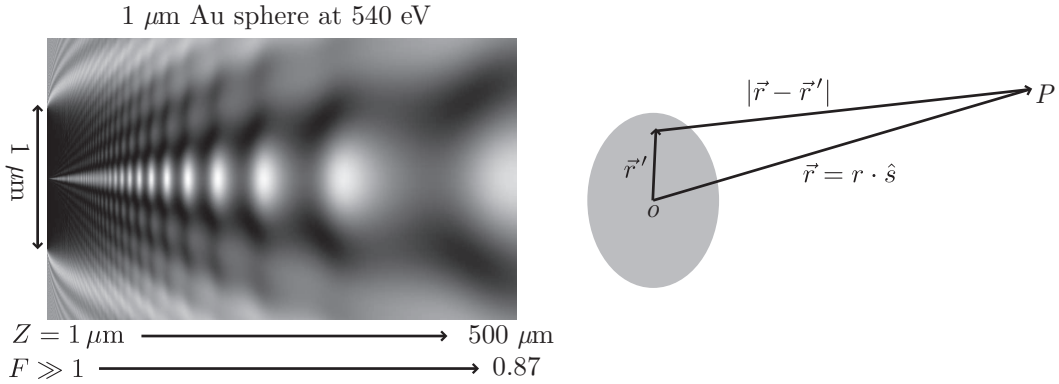

Figure 2.1: This figure illustrates the validity and the geometry of the far-field approximation. **Left:** Radial diffraction pattern of a 1 μm gold sphere at 540 eV as a function of distance and corresponding Fresnel number. Far-field diffraction occurs for $F < 1$. **Right:** Geometry of the far-field approximation.

subject to this dissertation only rely on the Born approximation, we will explain it in more detail below and mention the Rytov approximation only for completeness.

2.1.1 The 1st-Order Born Approximation in the Far-Field

If we assume a weak scatterer with $|n| \approx 1$ we can write $U(\boldsymbol{r}, \omega) \approx U^{\text{inc}}(\boldsymbol{r}, \omega)$ and Eq. 2.7 becomes

$$U^{\text{sc}}(\boldsymbol{r}, \omega) = \int_V F(\boldsymbol{r}', \omega) U^{\text{inc}}(\boldsymbol{r}', \omega) \frac{e^{ik|\boldsymbol{r}-\mathbf{r}'|}}{|\boldsymbol{r}-\mathbf{r}'|} \mathrm{d}\boldsymbol{r}'. \tag{2.8}$$

This approximate solution is referred to as the first-order Born approximation (corresponding to the first term of the perturbation expansion derived in Born [13]) Notice that this now represents a solvable solution to the scattered wavefield since the integrand now only contains the incident wavefield as a factor.

A further simplification can be achieved by looking at the scattered signal far away from the scatterer. This is called far-field or Fraunhofer diffraction. All data collected for the purpose of this dissertation was recorded in this

geometry. The approximation is valid if the so-called Fresnel number

$$F = \frac{a^2}{Z\lambda} \tag{2.9}$$

is less than 1 for our experimental geometry, where Z is the distance behind a diffracting object of diameter a. The relationship of scattered wavefield and Fresnel number is illustrated on the left of Fig. 2.1 which shows a radial cut through the scattered intensities of a 1 μm gold sphere at 540 eV as a function of distance. We are in the Fraunhofer regime when $F < 1$. At this point the angularly resolved scattered intensities will not change with the distance Z anymore; instead they only scale radially.

Mathematically the fact that we are in the Fraunhofer regime translates into a first order expansion of the Green function, see Eq. 2.6. The geometry is illustrated on the right side in Fig 2.1. The expansion assumes that $r' \ll r$, thus

$$\frac{e^{ik|\boldsymbol{r}-\boldsymbol{r}'|}}{|\boldsymbol{r}-\boldsymbol{r}'|} \approx \frac{e^{ikr}}{r} e^{-ik\hat{s}\cdot\boldsymbol{r}'},$$

where $\hat{s}$ is the direction of the scattered wave. Plugging this into Eq. 2.8 and leaving out frequency dependent terms, we obtain

$$U^{\mathrm{sc}} = f(\hat{s})\frac{e^{ikr}}{r},$$

where f is the so-called scattering amplitude of

$$f(\hat{s}) = \int F(\boldsymbol{r}')U^{\mathrm{inc}}(\boldsymbol{r}')e^{-ik\hat{s}\cdot\boldsymbol{r}'}\,\mathrm{d}^3r' = \mathcal{F}\left\{F(\boldsymbol{r}')U^{\mathrm{inc}}(\boldsymbol{r}')\right\}. \tag{2.10}$$

From this we see that there is a Fourier transform relationship between the product of the object with the incident wavefield, and the complex scattering amplitude in the far field (assuming the Born approximation). This is the basic idea of diffraction imaging: by measuring the scattering amplitudes we can obtain information about the object by a simple inverse Fourier transform!

Of course it is not that straightforward in real life. The first point is that the Fourier transform relationship of Eq. 2.10 still involves the product of object times incident wavefield. To solve for the object alone, one either has to make further assumptions regarding the incident wavefield (as is done for full-field XDM; see Sec. 2.2) or find a way of separating out the contributions of object and incident wavefield to the scattering amplitude (as is done for SXDM as described in see Sec. 2.3, and other diffraction imaging techniques mentioned in Sec. 2.3.5). The other major roadblock to successful Fourier

inversion is so well-known and fundamental that it deserves its own section.

2.1.2 The Crystallographic Phase Problem

When we are measuring the scattering amplitudes with a detector, we are measuring intensities and thus loose all information about the phases of the scattered wavefield. As it turns out, the phases are crucial to a successful reconstruction of the object function from a measurement of its diffraction pattern. This is commonly referred to as the "crystallographic phase problem" as it was first described in the x-ray crystallography community, but it also applies to our case of non-crystallographic objects as well as other imaging and signal processing techniques.

Figure 2.2 illustrates the effects of lost phase information with an example object. The object is shown on top, and its Fourier magnitudes and Fourier phases are shown in the middle on the right and left, respectively. Discarding the phases of the scattered wavefield and inverting the Fourier magnitudes alone results in an image that bears no resemblance to the original object, as can be seen on the bottom left of the figure. However, discarding the Fourier magnitudes and inverting the Fourier phases alone leads to a recognizable representation of the original object. In other words the lost phase information needs to be recovered before a successful inversion can be performed.

The task of reconstructing the original object becomes the task of phase retrieval. Various strategies have been developed for different techniques. In the field of x-ray diffraction imaging, phase retrieval is performed by a whole class of iterative algorithms. The first successful algorithm to be able to reconstruct an object function from intensity measurements in the Fourier plane alone was demonstrated by Fienup in 1978 [14]. It was based on an earlier algorithm invented to solve the inverse problem from intensity measurements in the Fourier and the real space domain, the "Gerchberg-Saxton" algorithm [15]. Several generalizations have since been developed [16, 17], all of which are based on the same principle of iteratively enforcing constraints in real and Fourier space. The type of constraint that is being used depends on the object that is being imaged and the experimental conditions. Suppose our measured diffraction pattern is an $N \times N$ array. This represents a set of N^2 equations we can use to solve for our object in real space. However, our object is complex and thus consists of $2 \cdot N^2$ unknowns. The basic idea behind additional constraints is to reduce the number of unknowns in our data space to have enough equations to solve for the missing phases.

The constraints that we used for our data analyses are the Fourier modulus constraint in inverse space, and the support constraint (for XDM data; see Sec. 2.2) or the overlap constraint (for SXDM data; see Sec. 2.3) in real space.

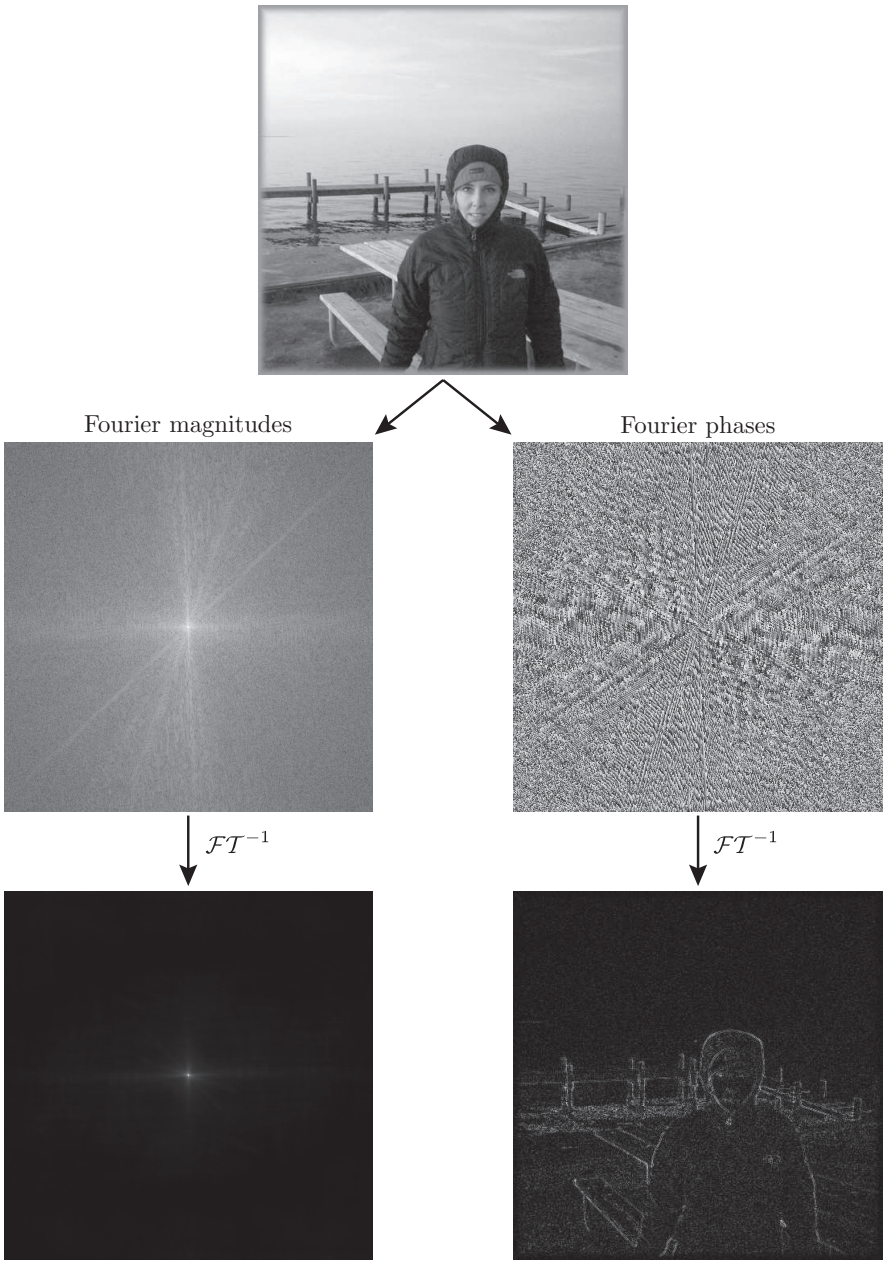

Figure 2.2: Illustration of the (non-)crystallographic phase problem: No recognizable image of the original object is obtained when discarding phase information and inverting the Fourier magnitudes alone (on left). However, discarding the Fourier magnitudes and inverting the Fourier phases alone leads to a recognizable representation of the real object (on right).

Other types of constraint (mainly used in x-ray crystallography) include the atomicity constraint and the object histogram constraint.

2.1.3 The Object Function $F(\boldsymbol{r})$

Before moving on to explaining the particulars of each experimental technique (XDM or SXDM), I would like to take a closer look at the "object function" (introduced in Eq. 2.2) that we are striving to recover from our diffraction measurements. Remembering Eq. 1.8 we see that

$$k^2\left(1-n^2\right) = 4\pi r_{\mathrm{e}}\rho(\mathbf{r}'),$$

where $\rho(\boldsymbol{r}') = n_{\mathrm{a}} f(\boldsymbol{r}')$ is the effective complex electron density; that is, what we are recovering from measuring the scattered wavefield

$$F(\mathbf{r}') = -r_{\mathrm{e}}\rho(\mathbf{r}') \tag{2.11}$$

is like an effective complex electron density. With Eq. 1.11, we can rewrite Eq. 2.2 as

$$F(\mathbf{r}') \approx \frac{2\pi}{\lambda^2}\left[\delta(\mathbf{r}') + i\beta(\mathbf{r}')\right]; \tag{2.12}$$

that is, the object function can be related to the absorption and phase shift coefficients of the elements it is made of.

2.2 Full-field X-ray Diffraction Microscopy

The idea of phase retrieval from recorded diffraction intensities alone was first conceived by Sayre in 1952 [18]. The first experimental demonstration of XDM was achieved by Miao et al. in 1999 on a fabricated test pattern [19]. Since then the technique has been successfully applied in 2D to biological [20–22] and material science samples [23], and in 3D to test structures [24], material science [25, 26] and biological [27] samples. Many details of this section are based on a very nice paper by Chapman et al. [24].

The full-field experimental geometry of x-ray diffraction microscopy operates on two underlying assumptions:

1. the incident wavefield is a plane wave and
2. the object to be imaged is surrounded by empty space.

I have already hinted at the first assumption after the derivation of Eq. 2.10. Let's see how this relationship is further simplified if the incident wavefield is

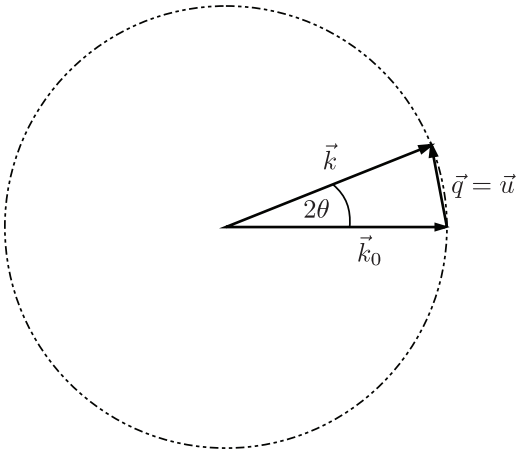

Figure 2.3: The wave vector transfer $\boldsymbol{q}$ in elastic scattering processes is equal to the volume grating spatial frequency $\boldsymbol{u}$. Recording the scattered wavefield at various positions $\hat{s}$ fills in the q-values on a sphere with radius k, the so-called Ewald sphere of diffraction.

a plane wave traveling in direction $\hat{s}_0$, $U^{\text{inc}} \propto e^{ik\cdot\hat{s}_0}$. Plugging this into Eq. 2.10 we obtain

$$f(\hat{s}_0, \hat{s}) = \int F(\mathbf{r}')e^{-ik(\hat{s}-\hat{s}_0)\cdot\mathbf{r}'}\mathrm{d}^3r' = \tilde{F}\left[k(\hat{s}-\hat{s}_0)\right]. \tag{2.13}$$

From this we see that in the far-field the scattering amplitude of the wavefield in direction $\hat{s}$ and for an incident wave with direction $\hat{s}_0$ is given by one and only one Fourier component of the object function $\tilde{F}(\mathbf{q})$, namely by the component $\mathbf{q} = k(\hat{s} - \hat{s}_0)$. We can think of the scattering object as being comprised of a number of volume gratings $\boldsymbol{u}$ each of which has a defined orientation, and magnitude. The magnitude is a measure of the periodicity of the volume grating; it is referred to as spatial frequency. The exact composition depends of course on the object and its internal features. For volume rather than planar grating scattering, the Bragg condition must be satisfied so that the wave vector transfer $\boldsymbol{q} = \boldsymbol{k} - \boldsymbol{k}_0$ must be equal to the grating spatial frequency $\boldsymbol{u}$ as can be seen in Fig. 2.3. In other words, by measuring the scattered wave at various different locations $\hat{s}$ we can fill in the q-space of the object function, such that for a given incident wave direction $\hat{s}_0$ we can measure Fourier components of the object function that lie on a sphere (the so-called Ewald sphere) with radius k which results in q-values ranging from 0 to $2k$, ideally. In reality we can only record a small range of q-values due to the

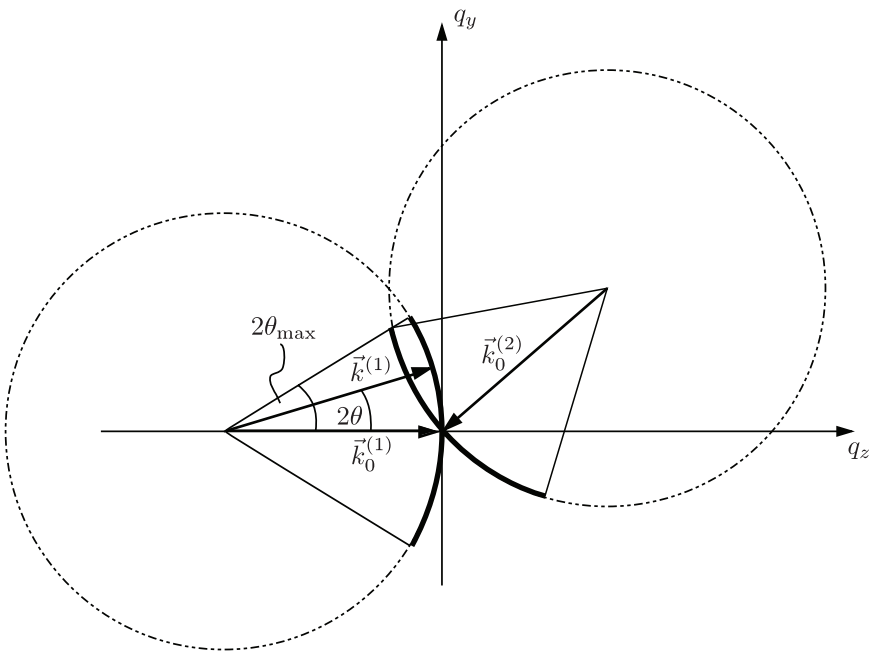

Figure 2.4: Rotating the object is equivalent with having the incident wave come from a different direction $\boldsymbol{k}_0^{(2)}$. Doing so provides access to more q-space values centered around $q = 0$. Due to limited solid angle acceptance of our experiment $2\theta_{\max}$ we are limited to small q-values. This is indicated by the solid black arcs.

limited solid angle acceptance of our detector $2\theta_{\max}$. This limits our resolution with which we can image details of the object as is explained in more detail in Sec. 2.2.1.

So far we have only recorded a diffraction pattern for one direction $\boldsymbol{k}_0$ of the incident wavefield, measuring mainly contributions from volume gratings that are perpendicular to the incident wavefield, $\boldsymbol{u} \perp \boldsymbol{k}_0$. It is possible to invert such a measurement and reconstruct something referred to as 2D exit wave leaving the object. The implications of reconstructions from such limited measurements are discussed in Sec. 2.2.3. For now I will focus on obtaining sufficient data to describe the object for all directions of its volume grating vectors $\boldsymbol{u}$. A practical way of collecting a different set of q-values is given by rotating the object in real space which is equivalent to the incident wavefield coming from a different direction $\boldsymbol{k}_0^{(2)}$. This situation is illustrated in Fig. 2.4. By rotating the object we rotate the Ewald sphere and thus can fill in our q-space.

2.2.1 Geometry of the experiment

We now want to quantitatively understand how we are mapping out the q-space when recording the diffraction pattern for different object orientations with a flat detector; for example a CCD. The left part of Fig. 2.5 shows the basic geometry of the scattering process. Again we call the direction of the incident wavefield $\boldsymbol{k}_0$ and the direction of the scattered wavefield $\boldsymbol{k}$. Let's assume the incident wave travels in the z-direction. The scattering angle 2θ describes the angle between $\boldsymbol{k}_0$ and $\boldsymbol{k}$. The components of the scattered wave vector can be written in terms of the angles 2θ and β as

$$\begin{aligned} k_{\mathrm{x}} &= \frac{2\pi}{\lambda} \sin 2\theta \cos\beta \\ k_{\mathrm{y}} &= \frac{2\pi}{\lambda} \sin 2\theta \sin\beta \\ k_{\mathrm{z}} &= \frac{2\pi}{\lambda} \cos 2\theta. \end{aligned}$$

From that it follows that the wave vector transfer which is equal to the spatial frequency of the scattering volume grating $\boldsymbol{u}$ is

$$\boldsymbol{q} = \boldsymbol{u} = \frac{2\pi}{\lambda} \begin{pmatrix} \sin 2\theta \cdot \cos\beta \\ \sin 2\theta \cdot \sin\beta \\ \cos 2\theta - 1 \end{pmatrix}.$$

Since our CCD is rather flat and not curved, the geometry of the actual experiment is slightly different and illustrated in the right part of Fig. 2.5. If we express the angles 2θ and β in terms of the measured quantities x, y and Z_0, where x and y are absolute distances in the detection plane and Z_0 is the distance between the object and the CCD, we find

$$\begin{aligned} \sin 2\theta &= \frac{r}{Z_0 + \Delta P} \\ \cos 2\theta &= \frac{Z_0}{Z_0 + \Delta P} \\ Z_0 + \Delta P &= \sqrt{Z_0^2 + r^2}, \end{aligned}$$

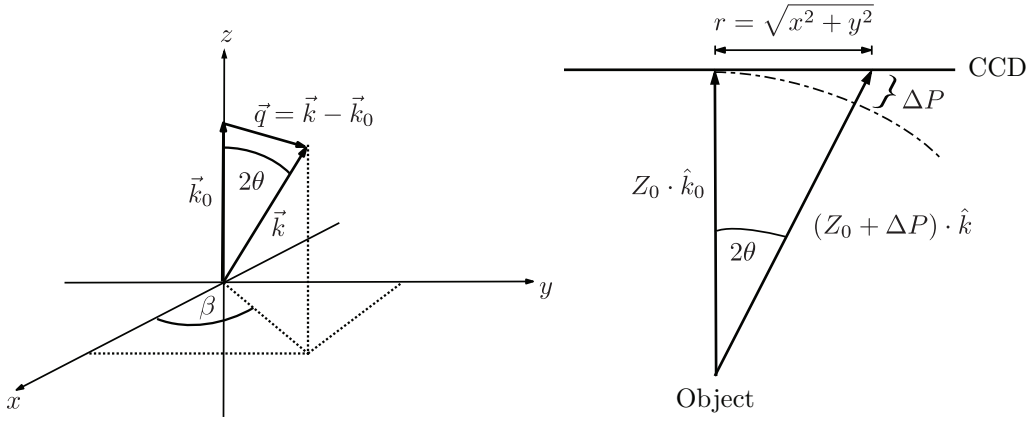

Figure 2.5: This figure illustrates the geometry of the scattering process in both spherical coordinates (left) and coordinates of the experiment (right). **Left:** The scattered wave vector $\boldsymbol{k}$ subtends an angle 2θ with the incoming wave vector $\boldsymbol{k}_0 \parallel \hat{z}$ and the wave vector transfer $\boldsymbol{q} = \boldsymbol{k} - \boldsymbol{k}_0$ is equal to the volume grating frequency. **Right:** We can express the scattering angle 2θ in terms of the geometry of the experiment.

so that we can express the wave vector transfer as

$$\boldsymbol{q} = \frac{2\pi}{\lambda} \begin{pmatrix} \frac{x}{\sqrt{Z_0^2+x^2+y^2}} \\ \frac{y}{\sqrt{Z_0^2+x^2+y^2}} \\ \frac{Z_0}{\sqrt{Z_0^2+x^2+y^2}} - 1 \end{pmatrix}. \tag{2.14}$$

Of course we are dealing with discrete arrays, where

$$\begin{aligned} x &= i \cdot p \\ y &= j \cdot p \end{aligned}$$

for a camera with pixelsize p. Equation 2.14 then becomes

$$\boldsymbol{q} = \frac{2\pi}{\lambda} \begin{pmatrix} \frac{i \cdot p}{\sqrt{Z_0^2+(i^2+j^2)p^2}} \\ \frac{j \cdot p}{\sqrt{Z_0^2+(i^2+j^2)p^2}} \\ \frac{Z_0}{\sqrt{Z_0^2+(i^2+j^2)p^2}} - 1 \end{pmatrix}. \tag{2.15}$$

Similarly, we want to express $\boldsymbol{q}$ in terms of its Miller indices in reciprocal space l, m, n as a function of camera indices i, j

$$\begin{aligned} q_x &= l \cdot \Delta q \\ q_y &= m \cdot \Delta q \\ q_z &= n \cdot \Delta q. \end{aligned}$$

Let's first determine the frequency sampling Δq: the highest spatial frequency $q_{\max}$ that we can record is determined by the geometry of the experiment. If we assume small angle approximation then we can write

$$q_{\max} = \frac{1}{d_{\min}} = \frac{p \cdot \frac{N}{2}}{Z_0 \cdot \lambda}, \tag{2.16}$$

where $d_{\min}$ is the size of the smallest volume grating we can record, p is the pixelsize of the camera and N is the number of pixels on the CCD. Note that $d_{\min}$ is the theoretical resolution limit of our reconstruction and is directly determined by the maximum spatial frequency that we can record. The actual resolution of a reconstruction d_{res} depends on the quality of the recorded data and real space constraints. It is

$$d_{\text{res}} \geq d_{\min}, \tag{2.17}$$

and we will look into ways of assessing the actual resolution of a reconstruction in Sec. 2.4.2.

Finally, the sampling rate in reciprocal space is given by

$$\Delta q = \frac{p}{Z_0 \lambda}. \tag{2.18}$$

If we divide Eq. 2.14 by Eq. 2.18 then we will obtain real dimensionless numbers which can be round to the nearest integer to give us the indices k, l, m of the q-space array as a function of the indices i, j of the measured array. In other words, Eq. 2.14 tells us how to map the measured array back onto the Ewald sphere on a per pixel basis.

There are two things left to worry about. Once we found the appropriate Miller indices l, m, n for a voxel in reciprocal space, its magnitude is found from the average of all combinations of detector pixels and orientations that contribute to this voxel. The other one is how to incorporate different orientations of the object with respect to the incident wavefield. The easiest way is to use Eq. 2.14 in conjunction with a 3D rotation matrix R that corresponds

to the rotation of the object (usually only around one axis), such that

$$\boldsymbol{q}(\phi) = R_\phi \boldsymbol{q} \tag{2.19}$$

2.2.2 Sampling Requirements

The array resulting from Eq. 2.19 represents a discretely sampled version of the continuous diffraction pattern described by Eq. 2.13. Whenever we are sampling a continuous function we have to be careful to avoid aliasing effects. The proper sampling of continuous signals has been discussed in many great publications; see for example Shannon [28] in general and Sayre [18], Bates [29] and Crowther et al. [30] specific to Fourier phase problems. In our case, the object function is recovered by inverse Fourier transform of Eq. 2.13, such that

$$F(\boldsymbol{r}) = \Delta\mathbf{q} \sum_j F(\mathbf{q_j}) e^{-2\pi i \boldsymbol{r}\cdot\mathbf{q_j}}.$$

In other words, the recovered object is a convolution of itself with the inverse transform of the sampling function. The result of this will only be close to the real object if the inverse transform of the sampling function has a sharp spike at the origin and is approximately 0 for all other values at least out to the radius of the object. This can only be accomplished with regular sampling functions. An important consequence of a regular sampling function is that its inverse transform is periodic with a period inversely proportional to the sample spacing. Due to the convolution effect, the recovered object will also be repeated with this period. It is desirable to have the period be larger than the width of the object which is equivalent of a sample spacing in inverse space of greater than $1/D$ for an object of width D. This is what happens in Shannon sampling [18]. If the phases were known, then Shannon sampling the Fourier-plane signal of the object would suffice to record all information within the object. Since we do not record the phases but merely intensities, the information we record is band-limited not by the extent of the actual object but by its autocorrelation

$$\mathcal{FT}^{-1}\{I(\mathbf{q})\} \propto F(\boldsymbol{r}) \circ F^*(-\boldsymbol{r}). \tag{2.20}$$

In general the autocorrelation will be twice as large as the object itself therefore leading us to conclude that we have to sample at $1/2D$ rather than $1/D$ to record all information present. This is sometimes referred to as "oversampling" but it really is just Shannon sampling for the autocorrelation of the object. In the lingo of oversampling, Shannon sampling corresponds to an oversampling factor of 2. From simulations it turns out that Shannon sampling a non-ideal

dataset (such as a dataset with noise) is not sufficient to reconstruct. Rather, an oversampling factor of around 2.6 or more is needed to recover the missing phases [31].

In the far-field we can derive a relationship between the pixel size of the detector and the size of the sample as a function of oversampling s. The frequency sampling in the actual experiment is determined by the detector position and pixel size according to Eq. 2.18. Thus, the oversampling in reciprocal space can be written as

$$s = \frac{1}{\Delta q \cdot D} = \frac{\lambda Z_0}{pD}. \tag{2.21}$$

We are in the far-field limit if the Fresnel number of Eq. 2.9 is less than one and thus

$$Z_0 > \frac{D^2}{\lambda}.$$

Plugging this into Eq. 2.21 leads to the expression

$$D < sp, \tag{2.22}$$

which says that the sample must be on the order of the CCD pixelsize to fulfill minimum oversampling requirements with $s = 2$.

Angular Sampling Requirements

Similar to the transverse coordinates described above, the width of a speckle due to an object of size D in the longitudinal coordinate is given by $1/D$. To record 3D data, the sample is rotated about one axis. At the highest spatial frequency $q_{\max}$ this leads to a separation of

$$\Delta q_{\max} = q_{\max} \cdot \Delta\phi$$

between two Ewald spheres. Using the same oversampling criterion as for the transverse coordinates we can achieve critical sampling ($s = 2$) when

$$\Delta\phi = \frac{\Delta q_{\max}}{q_{\max}} = \frac{\Delta x}{D},$$

where Δx is the real-space sampling interval given by $\Delta x = 1/2q_{\max}$. Again, this corresponds to Shannon sampling of a band-limited sample of width D as one can quickly verify by plugging in the relationship obtained above for $q_{\max}$. In other words, as long as we have 2 samples per longitudinal width of speckle ($1/D$) we fullfill the Shannon criterion. To be able to do so over the

full angular range we need

$$m = \frac{2\pi}{\Delta\phi} = \frac{2\pi \cdot D}{\Delta x}$$

different views of the specimen. A more thorough derivation of this theorem can be found in Crowther et al. [30]. In order to collect enough data to reconstruct an object of width $D = 5\,\mu$m down to a resolution of $2\Delta x = 10$ nm we need an angular spacing of approximately $\Delta\phi = 1$ mrad or $0.05°$. In reality such a small angular spacing is hardly achievable. Experimentally feasible are steps of $0.5°$ which will lead to a decrease in resolution for the 3D reconstruction.

Coherence

The required spatial coherence of the illumination is given by the size of the object D and the minimum oversampling factor ($s = 2$). It can be shown [32] that in order to avoid overlap from mutually incoherent speckle patterns due to different coherent patches in the illumination the spatial coherence width w_c should be

$$w_c > 2D. \tag{2.23}$$

To get an estimate on the required temporal coherence required, let's assume that we want the diffracted radiation from one volume grating element due to two slightly different wavelengths λ and $\lambda + \Delta\lambda$ to not be separated by more than a pixel on the detector (which corresponds to the half width of a speckle for critical sampling with $s = 2$), *i. e.*

$$\begin{aligned} \frac{y_1}{qZ_0} &= \lambda \\ \frac{y_2}{qZ_0} &= \lambda + \Delta\lambda \\ y_1 - y_2 &< p. \end{aligned}$$

Together with Eq. 2.18 this leads to the requirement that

$$\frac{\lambda}{\Delta\lambda} = \frac{q}{\Delta q} = \frac{N}{2},$$

where N is the number of pixels in our detector. The coherence properties of our beamline 9.0.1 at the Advanced Light Source have already been described in Sec. 1.3.1. We note that for samples of diameter $D = 3\,\mu$m and a CCD with $N = 1024$ pixels we satisfy both spatial and temporal coherence requirements.

2.2.3 2D imaging

When inverting the diffraction pattern obtained from a single projection of the specimen one obtains the complex wavefield leaving the object This is called an exit wave. Compared to 3D reconstructions there are some important limitations due to the nature of the Ewald sphere; most notably there will be defocus artifacts. A 2D diffraction pattern delivers information on spatial frequencies that lie on a sphere with

$$q_z = \frac{1}{\lambda} - \sqrt{\frac{1}{\lambda^2} - q_x^2 - q_y^2} \approx -\frac{\lambda}{2}\left(q_x^2 + q_y^2\right),$$

where $1/\lambda$ is the radius of the sphere and the origin $\boldsymbol{q} = (0, 0, 0)$ is on the sphere centered in the forward direction. The approximation is valid for small scattering angles. The numerical aperture of the diffraction pattern can be defined as $\mathrm{NA} = q_{x,\max}\lambda$ similar to imaging with a lens. With that the maximal longitudinal coordinate $q_{z,\max}$ is given as

$$q_{z,\max} = -\frac{\mathrm{NA}^2}{2\lambda}.$$

An object of width D will have speckles of width $1/D$ in longitudinal direction. Defocus artifacts occur if, for a given $q_{x,\max}$, the Ewald sphere does not cut through a speckle that is centered at $\boldsymbol{q} = (q_{x,\max}, q_{y,\max}, 0)$. This can be understood by looking at the extreme case of an object that cannot experience defocus artifacts, an object of width $D \to 0$. Such an object will scatter no information in the longitudinal direction and thus measuring its 2D diffraction pattern on a sphere is equivalent to measuring the 2D diffraction intensities on the plane $\boldsymbol{q} = (q_x, q_y, 0)$. In other words we can consider our sample to be free of defocus artifacts or truly 2D if its Ewald sphere departure for a given NA of the system is no more than $1/(4D)$ in the longitudinal coordinate. This leads to a maximum thickness of the object of

$$D < \frac{\lambda}{2\,\mathrm{NA}^2}, \tag{2.24}$$

so that the object thickness should be less than the depth of focus of a lens of the same numerical aperture to be considered a 2D object. If we want to measure out to very high spatial frequencies, the numerical aperture of the system will be high and the depth of focus of the imaging system will be small.

A 2D object that does not fulfill the requirement from Eq. 2.24 will experience defocus artifacts, where out of focus planes will acquire quadratic phase factors according to the Fresnel propagator [33]. This out-of-focus information

is especially distracting in a coherent imaging system. One can change the focal plane during the reconstruction process by adapting the support constraint to reflect the shape of the object in the desired focal plane [34] or after reconstruction by Fresnel-propagating the complex exit wave [35].

2.2.4 Experimental Setup at Beamline 9.0.1

Now that we have developed a firm understanding on how the object can be related to its scattered amplitudes and what the geometry of the scattering process is, let's switch gears and look at the experimental implementation of a diffraction setup at beamline 9.0.1 at the Advanced Light Source, Lawrence Berkeley Lab. The beamline itself has already been described in detail in Sec. 1.3.1 and we have also introduced the endstation in general terms in the same section. What we are concerned with here, are the parts of the microscope that are directly related to performing the diffraction experiment: the optics and the detector.

The Detector

As detector we use a charge-coupled device (CCD) from Princeton Instruments[1]. We have two different versions of the chip differing slightly in terms of pixel number and size, and well depth. The relevant parameters of each chip are summarized in Tab. 2.1. In this discussion we are mostly interested in how

	Number of pixels N	Pixel size p (μm)	Well depth ($10^3 e^-$)
1.3k chip	1300 × 1340	20.0	200
2k chip	2048 × 2048	13.5	100

Table 2.1: Parameters of our two CCD chips used to record the diffraction intensities.

many photons per pixel we can record before the CCD saturates. A simplified description of how a CCD works will help in understanding the two ways saturation can occur. Photons incident on a pixel on the CCD will produce a charge proportional to the total energy deposited in the pixel. Once the image has been recorded, the CCD is read-out by column-wise shifting the charges of all pixels by one row. The charges of pixels on the edge of the chip will end up in what's called the horizontal register. In the next step, the charges in the horizontal register are sent pixel-by-pixel to the ADC that converts their

[1] http://www.princetoninstruments.com/products/xraycam/pimte/

charges into counts using a certain conversion factor defined as

$$\text{gain} = \frac{\text{number of } e^-}{\text{counts}} \tag{2.25}$$

This process is then repeated until all rows have been read-out.

Saturation can occur due to the limited amount of charge that each pixel can store before being read-out (this is called the *well depth*) or due to the limited dynamic range of the ADC (our CCDs have a 16-bit ADC, corresponding to a maximum count of 2^{16}). For an unbinned CCD it makes sense to adjust the gain factor such that the two saturation levels are the same. For example, a gain factor of around 3 would saturate the ADC of the 1.3k chip when a pixel is charged to its well depth of 200 e^-.

Now let's look at how many of our photons we can record before the CCD is saturating. It takes an energy of 3.65 eV to create an electron-hole pair in the doped Silicon layer, so that the maximum number of photons that can be counted is

$$n_{\text{Photons}} = \frac{\text{well depth} \cdot 3.65}{E_{\text{Photon}}(eV)}, \tag{2.26}$$

where the well depth depends on the chip used (see Tab. 2.1). For the 2k-chip with 520 eV photons, saturation occurs at $n_{\text{Photons}} \simeq 700$ photons and for the 1.3k-chip at $n_{\text{Photons}} \simeq 1400$ photons. In other words, the detector has a dynamic range of 3 orders of magnitude.

Dynamic Range of the Scattered Intensities

How does this compare to the dynamic range of the scattered intensities that we want to record with our detector? Small angle x-ray scattering data suggests that the diffracted signal follows an inverse power law relationship with spatial frequency

$$I(q) \propto q^{-m} \tag{2.27}$$

($m = 4$ according to Porod's law). In our experiment we find the recorded power spectral densities to decline with exponents ranging from $m = 3$ to $m = 4$. This is also confirmed by other theoretical estimates in the literature, such as by Howells et al. [36] and Shen et al. [37]. The total dynamic range (DR) of the diffraction pattern can then be calculated as

$$\text{DR} = \left(\frac{q_{\max}}{q_{\min}}\right)^m = \left(\frac{d_{\max}}{d_{\min}}\right)^m, \tag{2.28}$$

where d_{min} is the desired resolution corresponding to the smallest feature to be imaged and d_{max} is the largest feature to be imaged (the size of the object). Assuming a 3 μm cell with desired resolution of 10 nm, we calculate a dynamic range of the diffracted signal to be DR $\approx 10^{10}$. This is far beyond the dynamic range of our CCD as described above. It is therefore necessary to record low and high spatial frequency regimes of the scattered signal separately to extend the dynamic range of our CCD.

Optical Components inside the Chamber

We use a movable piece of opaque silicon as a beamstop; this is mounted right in front of the CCD to block different parts of the intense low spatial frequency scatter. The entire optical setup is shown schematically in Fig. 2.6. The pinhole on the left serves as both beam defining aperture as well as en-

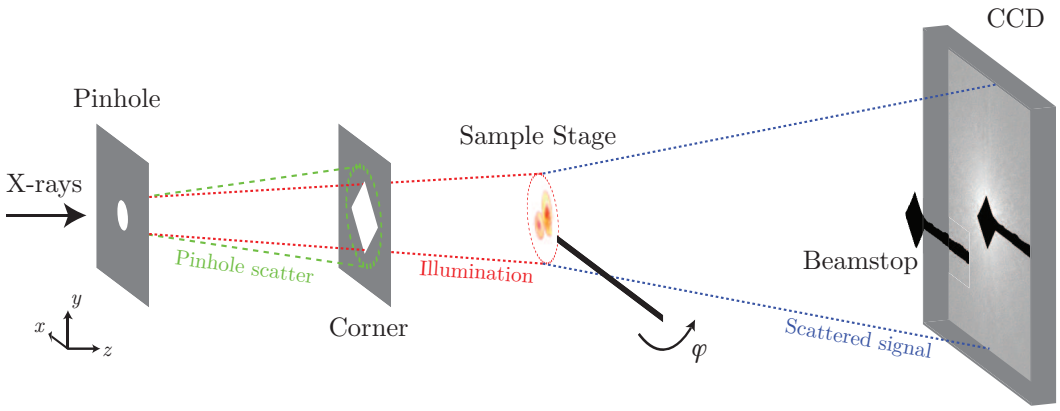

Figure 2.6: Schematic of the experimental setup for full-field XDM. The X rays first encounter a pinhole which acts as beam defining aperture as well as coherence and energy selecting aperture. Then they pass through the corner, a windowless Silicon Nitride frame that is meant to prevent scattered light from the pinhole to reach the detector. Then they hit the sample which is mounted on a rotation stage. Finally, the diffracted signal is recorded with a CCD, where a movable beamstop is used to block the intense 0-th order beam (not to scale).

ergy selecting and coherence defining aperture. We have already explained the relevant parameters in Sec. 1.3.1 but we recapitulate here: typically the experimental geometry is such that we have an Airy pattern of 30 μm diameter footprint in the plane of the sample and a coherence width of around 12 μm.

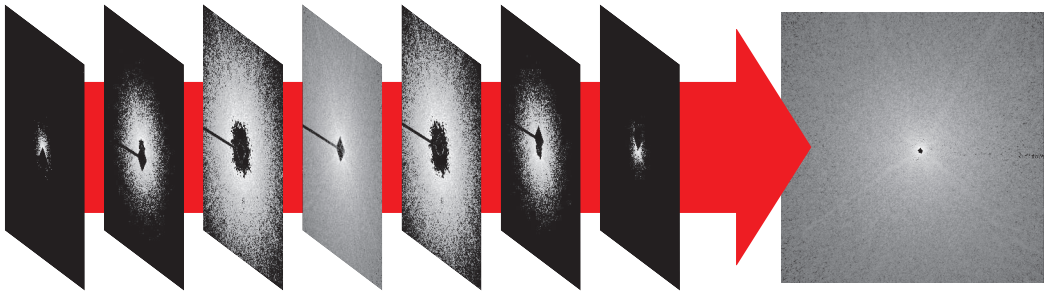

Figure 2.7: Due to dynamic range limitations in the detector, the diffracted signal has to be recorded in stages by blocking part of the intense low spatial frequency signal with a movable beamstop. The different parts then have to be merged afterwards to recreate the full 2D diffraction pattern (shown on right). To prevent damage to the CCD, a small area around $q = 0\,\mu\text{m}^{-1}$ cannot be recorded.

Just in front of the sample and downstream of the pinhole is mounted an "empty" Silicon Nitride window (in fact, only its frame is used) that removes higher order Airy rings and other pinhole scatter. Its size is determined by the condition that it have the same footprint on the CCD as the beamstop. Without a corner, we run the risk that the weak high spatial frequency signal from the sample is overpowered by pinhole scatter.

Since the pinhole-sample distance is fixed by the focal length of the monochromator zone plate (as described in Sec. 1.3.1), it is by varying the distance from sample to CCD, Z_0, that we can dial the experimentally important factors of oversampling s and maximum momentum transfer q_{max} that is recorded by the detector. Since $q_{\text{max}} \propto {}^1\!/\!_{Z_0}$ and $s \propto Z_0$, we have to find a trade-off position where we satisfy minimum oversampling criteria yet still record a high enough spatial frequency. For 520 eV X rays ($\lambda = 2.38$ nm) we typically use $Z_0 = 10$ cm. Plugging this into Eqs. 2.16 and 2.21 we see that we can record out to a maximum spatial frequency of $q_{\text{max}} = 55\,\mu\text{m}^{-1}$ at an oversampling factor of $s = 4$ (where we have assumed the 1.3k chip, see Tab. 2.1, and a sample of $3\,\mu$m diameter).

A Typical Data Collection Script

We use the in-vacuum VLM described in Sec. 1.3.1 to find a region of interest on the sample and then do a small 2D scan with the sample while monitoring the diffracted signal for a nice speckle pattern. Once a nicely scattering object has been found, we can start recording data on it. Due to the aforementioned

limited dynamic range of our CCD, we have to record each 2D diffraction pattern in several steps using the movable beamtsop arm to block different parts of the intense low spatial frequency signal while recording diffraction patterns at various different exposure times. Typically we record for at least 5 different beamstop positions and two to three different exposure times per position depending on the scattering of the object. The basic idea is as follows:

- We start with one corner of the beamstop (say the top corner) barely blocking the 0-th order and using shortest exposure times to measure the very low spatial frequency data
- We move the beamstop up a bit to block more of the intense low spatial frequency scatter and we record at slightly longer exposure times to fill in the intermediate spatial frequency range
- We center the beamstop on the beam to be able to record long exposure times to fill in the very high spatial frequency data
- The whole procedure has to be repeated for the bottom corner of the beamstop

Typical exposure times range from 1 msec for the low spatial frequencies to 60 sec or more for the high spatial frequencies. Additionally, we also record several frames for each position and exposure time to increase the signal-to-noise ratio for the scattered intensities. All these different exposures will have to be merged afterwards to recreate the full 2D diffraction pattern that we would have obtained in one shot if our detector had a high enough dynamic range. We have recently learned that this merging process is crucial to obtaining reconstructions of very high quality. This will be explained in more detail in Sec. 3.3.

Even with the help of a beamstop, there is usually a frequency range (typically spatial frequencies below $q = 1\,\mu\mathrm{m}^{-1}$) that is not accessible for CCDs because even shortest exposure times (on the order of 1 msec) would lead to damage to the chip. This is why one is left with an "empty" region in the center of the assembled diffraction pattern. The process is schematically illustrated in Fig. 2.7. On the left are shown representative diffraction patterns for all seven unique beamstop positions used for this particular data set. On the right is shown the full 2D diffraction pattern obtained by merging these seven positions. Note that the area around $q = 0\,\mu\mathrm{m}^{-1}$ could not be recorded to prevent damaging the CCD and is therefore black.

Recently, there has been an advance in Pixel Array Detectors (PAD) [38, 39] that have a dynamic range much higher than conventional CCDs, but high

pixel number detectors with good sensitivity in the soft x-ray energy regime are not yet widely available. Alternatively, one can try to attenuate the incident beam experimentally to be able to record the low spatial frequency data. We have recently implemented such a setup and show some preliminary results in Sec. 4.1.3.

2.2.5 The Reconstruction Process

Once the data have been recorded and merged into full 2D diffraction patterns, they can be assembled into a 3D diffraction pattern according to Eq. 2.15, if applicable. We then start the phase retrieval process by iteratively enforcing constraints in real and Fourier space as described below. Note that this procedure is also valid for single 2D projections whose limitations have been described in Sec. 2.2.3. In what follows, $I(\boldsymbol{q})$ refers to the set of measured diffraction intensities (mapped onto the Ewald-sphere for 3D), $\psi(\boldsymbol{x})$ to the current real space iterate, $\Psi(\boldsymbol{q})$ to the current iterate in Fourier space, and $\sigma(\boldsymbol{q})$ to the experimental error associated with the measured diffraction magnitudes ($\sqrt{I(\boldsymbol{q})}$).

Available Constraints

In order for any reconstruction to be successful we have to provide it with information that is known to be true, or at least assumed to be true for now subject to later improvement. These so-called constraints are enforced in both real and inverse space in an alternating fashion. Different reconstruction algorithms differ on how they update the current iterate based on their use of the constraints. The two most common constraints in non-crystallographic diffraction microscopy are modulus and support constraints.

A modulus constraint operator P_{M} is applied in inverse space and will project the modulus of the current guess of the Fourier amplitude $\Psi(\boldsymbol{u})$ onto the nearest measured diffraction amplitude while keeping the reconstructed phase,

$$P_{\mathrm{M}}\psi = \mathcal{F}\mathcal{T}^{-1}\begin{cases} \frac{\Psi(\boldsymbol{q})}{|\Psi(\boldsymbol{q})|}\left[\sqrt{I(\boldsymbol{q})}+\sigma(\boldsymbol{q})\right], & \text{if } |\Psi(\boldsymbol{q})| > \sqrt{I(\boldsymbol{q})}+\sigma(\boldsymbol{q}) \\ \frac{\Psi(\boldsymbol{q})}{|\Psi(\boldsymbol{q})|}\left[\sqrt{I(\boldsymbol{q})}-\sigma(\boldsymbol{q})\right], & \text{if } |\Psi(\boldsymbol{q})| < \sqrt{I(\boldsymbol{q})}-\sigma(\boldsymbol{q}) \\ \Psi(\boldsymbol{q}), & \text{otherwise, or if } \sqrt{I(\boldsymbol{q})}=0 \end{cases} \tag{2.29}$$

The first factor $\frac{\Psi(\boldsymbol{q})}{|\Psi(\boldsymbol{q})|}$ is responsible for keeping the reconstructed phase, while the second factor projects the magnitudes onto the closest point in the interval

Figure 2.8: From autocorrelation of the object shown on left to first support guess (middle). Usually the support is updated multiple times during the reconstruction process. A good support guess (shown on right for this data set) is crucial for successful reconstruction.

$\left[\sqrt{I(\boldsymbol{q})} - \sigma(\boldsymbol{q}), \sqrt{I(\boldsymbol{q})} + \sigma(\boldsymbol{q})\right]$. The modulus constraint operator will only project any $\Psi(\boldsymbol{q})$ that differ by more than one standard deviation from the measured data. It does nothing if it is already within that interval or if the intensities have not been measured for a particular $\boldsymbol{q}$, such as due to a limited number of angles recorded or missing central area from the beamstop.

The modulus constraint is usually supplemented by a *support constraint* in real space. The support constraint is the current guess of the outline of the object in real space and will set all magnitudes and phases of the current object guess outside that outline to zero while leaving everything on the inside of the outline unchanged.

$$P_{\mathrm{S}}\psi = \begin{cases} \psi(\boldsymbol{x}), & \text{if } \boldsymbol{x} \in S \\ 0, & \text{otherwise.} \end{cases}$$

If warranted by the object one can additionally impose positivity and reality constraints; that is, setting negative imaginary and/or real values in the current iterate in real space to 0 or setting its imaginary values to 0, respectively. Looking back at Eq. 2.12 we see that imposing a reality constraint is equivalent to neglecting absorption in our object. We assume that our specimen only introduces phase advances. Imposing a positivity constraint on the imaginary part only, or on both imaginary and real part of the object, limits the maximum phase shift introduced by each voxel in 3D, or between any two pixels in the exit wave in 2D, to be less than π or $\pi/2$, respectively.

Updating the Support Constraint

While the modulus constraint is given to us by the experimental data, the support constraint is not known *a priori* and usually has to be determined empirically. Being able to find a correct representation of the actual outline of the object is crucial to successful reconstruction. Typically, the support guess is updated several times during the reconstruction process. An example is illustrated in Fig. 2.8: The first guess is usually obtained by looking at the autocorrelation of the object given by Eq. 2.20 and shown on the left of Fig. 2.8. One then runs several iterations with the initial guess (shown in the middle) and then updates the support either by hand or in a semi-automated fashion using the `Shrinkwrap` algorithm [40]. One then repeats the procedure with the new support guess until one finds a support that minimizes the error metric discussed below.

Error Metric

During the phase retrieval process the quality of the current support guess is monitored using the error metric

$$E_{\mathrm{S}}^2 = \frac{\sum |\psi_n - P_{\mathrm{S}}\psi_n|^2}{\sum |P_{\mathrm{S}}\psi_n|^2} = \frac{\sum_{\mathbf{x}\notin S} |\psi_n(\boldsymbol{x})|^2}{\sum_{\boldsymbol{x}\in S} |\psi_n(\boldsymbol{x})|^2}. \tag{2.30}$$

It is a measure of how much power lies outside the support estimate S for a given reconstruction. High values indicate that the current support guess is inconsistent with the measured diffraction pattern. An example plot of the error metric for the reconstruction of an object with very well defined support is shown in Fig. 2.9. The error drops quickly after only a few iterations and then stays at a constant low level. Alternatively we can also define an error metric for the inverse space which measures the normalized distance between the current iterate and the modulus constraint set

$$E_{\mathrm{M}}^2 = \frac{\sum |\psi_n - P_{\mathrm{M}}\psi_n|^2}{\sum |P_{\mathrm{M}}\psi_n|^2}.$$

Phase Retrieval Algorithms

A variety of phase retrieval algorithms exist that all iteratively enforce the modulus and support constraints discussed above; they differ in how the next iterate is computed from the previous one. The ones we used for our reconstructions and that I will explain in more detail here are Fienup's Hybrid

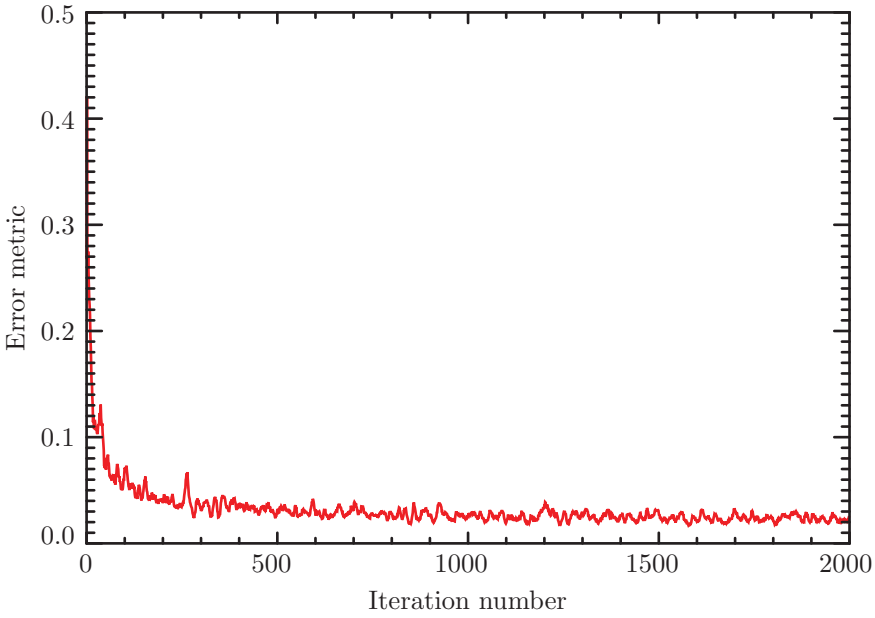

Figure 2.9: An example plot of the error metric for an object with well defined support. The error drops quickly and then stays at a low level.

Input-Output (HIO) algorithm and Elser's Difference Map algorithm (DM). A good comparison of most available algorithms in terms of convergence and likelihood of being stuck in local minima can be found in Marchesini [41].

Hybrid Input-Output (HIO) The hybrid input-ouput algorithm has been developed by Fienup [16]. In its general form it can be described as

$$\psi_{n+1} = \begin{cases} P_{\mathrm{M}}\psi_n, & \text{if } \boldsymbol{x} \in S' \\ (I - \beta P_{\mathrm{M}})\psi_n, & \text{otherwise,} \end{cases} \tag{2.31}$$

where β is the feedback or damping parameter and S' is the set of elements where $P_{\mathrm{M}}\psi_n$ satisfies the support and reality and positivity constraints if desired. A typical value for the feedback parameter that is known to lead to good convergence is $\beta = 0.9$. Note that the damping is applied in real space and only for elements outside the support S'. A schematic of how the next iterate is calculated in the case of intersecting one-dimensional constraint sets (S, M) is shown on the left of Fig. 2.10. After the desired number of iterations N, the final result is obtained after one more modulus constraint, *i. e.*

$$\psi = P_{\mathrm{M}}\psi_N. \tag{2.32}$$

The Difference Map (DM) This algorithm was developed by Elser [17]. It is defined by the iteration relation

$$\psi_{n+1} = \psi_n + \beta\left(\psi_{\mathrm{M}_n} - \psi_{\mathrm{S}_n}\right), \tag{2.33}$$

where

$$\begin{aligned} \psi_{\mathrm{M}_n} &= P_{\mathrm{M}}\left(\left(\gamma_{\mathrm{S}} + 1\right)P_{\mathrm{S}}\left(\psi_n\right) - \gamma_{\mathrm{S}}\psi_n\right) \\ \psi_{\mathrm{S}_n} &= P_{\mathrm{S}}\left(\left(\gamma_{\mathrm{M}} + 1\right)P_{\mathrm{M}}\left(\psi_n\right) - \gamma_{\mathrm{M}}\psi_n\right). \end{aligned}$$

An optimal value for the γ parameters is $\gamma_{\mathrm{S}} = \beta^{-1}$ and $\gamma_{\mathrm{M}} = -\beta^{-1}$. Note that with this choice and with $\beta = 1$, the difference map algorithm becomes the hybrid input-output algorithm described above. In our experiment we used a value of $\beta = -1.15$. It is also noteworthy that in its general form the difference map algorithm is computationally more demanding than HIO since it involves two time consuming modulus constraint projections per iteration. A schematic of how the next iterate is calculated in the case of intersecting one-dimensional constraint sets (S, M) is shown on the right of Fig. 2.10. After the desired number of iterations N, the final result is given by either the support

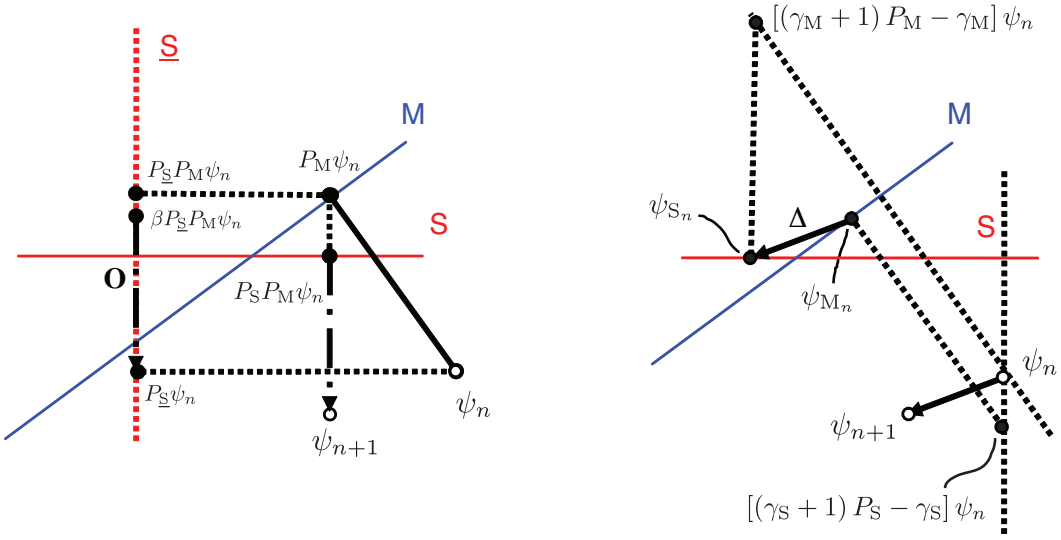

Figure 2.10: Calculation of the next iterate for the hybrid input–output (on left) and the difference map algorithm (on right) in the case of one-dimensional modulus and support constraint sets (S, M). The HIO algorithm applies the damping β in real space and for elements that are outside the support. This is represented by the perpendicular dotted line $\underline{S}$. The DM algorithm damps the application of both constraint operators $P_{M,S}$. The new iterate is calculated by evaluating the difference map $\Delta = (\psi_{M_n} - \psi_{S_n})$. Figure adapted from Marchesini [41].

estimate or the modulus estimate, *i. e.*

$$\psi = \psi_{S_n} \text{ or } \psi_{M_n}. \tag{2.34}$$

Note that in 3D, the final result ψ corresponds to the actual object function $F(\boldsymbol{r})$ as explained in Sec. 2.1.3. In 2D, it corresponds to the exit wave leaving the object as described in Sec. 2.2.3.

2.2.6 Post Reconstruction Data Analysis

As explained above, iterative phase retrieval in the far-field geometry works by finding a complex wavefield which satisfies real-space constraints such as the imposition of a finite support (and possibly others such as a limit on maximum phase variation), and the Fourier-space constraint of adjustment towards the measured diffraction magnitudes. Because one pixel in Fourier space affects

all pixels in real space and vice versa, one can start the algorithm with random phases and converge to a solution. Since the real-space constraints are not known perfectly, and since random and systematic errors are possible in the measurement of the Fourier plane magnitudes, one cannot find a single, numerically unique solution to the complex wavefield (though in "good" reconstructions the variations between different possible solutions are small). As a result, various iterate averaging procedures have been adopted [21, 24, 35], based on the idea that consistent phases add coherently, while inconsistent phases add incoherently. In order to make the result more statistically significant, one usually averages several independently obtained reconstruction results (thus eliminating any high frequency noise components) and then applies a high-pass filter to suppress the unconstrained low spatial frequency modes. We will take a detailed look into what constitutes sufficient averaging in Sec. 3.4.4; here we will look at how to average properly: Before averaging we have to remove any linear phase ramps and we have to set the global phase to be the same for each reconstruction.

Removing Linear Phase Ramps

The Fourier shift theorem states that translation of the object in real space corresponds to a phase ramp in Fourier space and vice versa. Thus if a diffraction pattern is not perfectly centered (in the sense that the pixel corresponding to $q = 0$ is at index $(0, 0, 0)$ in the array) it will lead to a phase ramp in the reconstruction. To remove the phase ramp, we go through the reconstructed phases line-by-line, calculate the average slope and subtract it from the entire array. We then repeat this procedure for the other dimensions. Note that this procedure will not work if we have a phase ramp that extends over more than 2π in one or more dimensions. In this case one first has to remove the phase wrap by properly centering the diffraction pattern, as has been described for example in Chapter 4 of Miao [8].

Adjusting the Global Phase

We want to set the global phase of all images so that they add up mostly coherently instead of canceling themselves. The choice of global phase for the first image then determines the global phase for subsequent images. There are different ways of doing this; we follow the approach of Chapman et al. [24]. The idea is to find the global phase that maximizes the real part of the first

image γ_0. We do this by maximizing the quantity

$$\begin{aligned}
\alpha &= \textstyle\sum_k \left\{\mathrm{Re}\left[\psi_0(k)e^{i\phi_0}\right]\right\}^2 \\
&= \textstyle\sum_k \left(\mathrm{Re}\left\{\mathrm{Re}[\psi_0(k)]\cos\phi_0 - \mathrm{Im}[\psi_0(k)]\sin\phi_0\right.\right. \\
&\qquad \left.\left. + i\left\{\mathrm{Re}[\psi_0(k)]\sin\phi_0 + \mathrm{Im}[\psi_0(k)]\cos\phi_0\right\}\right\}\right)^2 \\
&= \textstyle\sum_k \left\{\mathrm{Re}[\psi_0(k)]^2\cos{\phi_0}^2 - 2\mathrm{Re}[\psi_0(k)]\mathrm{Im}[\psi_0(k)]\cos\phi_0\sin\phi_0\right. \\
&\qquad \left. + \mathrm{Im}[\psi_0(k)]^2\sin{\phi_0}^2\right\} \\
&= \textstyle\sum_k \left\{2\left|\psi_0(k)\right|^2 + \left[\psi_0(k)\right]^2 e^{2i\phi_0} + \left[\psi_0(k)\right]^{*2} e^{-2i\phi_0}\right\}/4.
\end{aligned} \tag{2.35}$$

This can be achieved by either maximizing the second or the third term of Eq. 2.35. We are going for the second term and maximize it by setting $\phi_0 = -\phi/2$, where

$$\phi = \arctan\left[\frac{\mathrm{Im}\left\{\sum_k \left[\psi_0(k)\right]^2\right\}}{\mathrm{Re}\left\{\sum_k \left[\psi_0(k)\right]^2\right\}}\right].$$

The global phase ϕ_1 for subsequent images ψ is then determined by minimizing $\sum_k \left|\psi_0(k) - \psi(k)e^{i\phi_1}\right|^2$ which is equivalent to maximizing the real part of the mixed term $\sum_k \mathrm{Re}\left[\psi_0(k)\psi(k)e^{i\phi_1}\right]$, *i. e.* having them add up coherently. This is equivalent to the negative of the phase of $\sum_k \left[\psi_0(k)\right]^* \psi(k)$. Thus

$$\phi_1 = -\arctan\left[\frac{\mathrm{Im}\left\{\sum_k \left[\psi_0(k)\right]^* \psi(k)\right\}}{\mathrm{Re}\left\{\sum_k \left[\psi_0(k)\right]^* \psi(k)\right\}}\right].$$

Failure to adjust the global phase in the aforementioned manner will lead to less consistent phase retrieval.

Highpass Filtering

As described in Sec. 2.2.4, we have missing data in the low spatial frequency regime around $q = 0$. During the reconstruction process the affected pixels are left untouched; that is, we let the algorithm fill in the missing values without imposing any constraints on them. Independent reconstructions of the same data set will have slightly different values in these pixels. When looking at reconstruction results, this leads to variations in larger-scale features since diffraction data from spatial frequencies less than about $1\,\mu\mathrm{m}^{-1}$ were not recorded. These differences can be very confusing when comparing or averaging several reconstructions of the same data set. Since they can be considered artifacts inconsistent with the measured data it is permissible to remove the contributions of this low-q data by either calculating and constraining the data

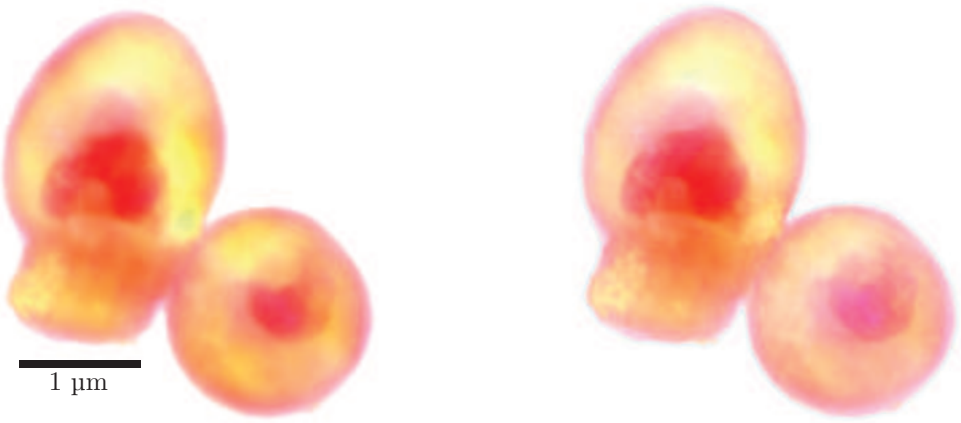

Figure 2.11: A qualitative illustration of the effects of applying a high-pass filter to a reconstruction of a budding yeast cell. The original result (on left) looks "fuzzier" due to unconstrained low spatial frequency features. When applying the high-pass filter, the contrast due to these unmeasured spatial frequencies is suppressed, resulting in improved image quality.

to the most likely modes that can exist in the missing region or by simply filtering out their contribution to the image with a high-pass filter. The former has been described in Thibault et al. [35].

Here we use the latter approach with a filtering function $f(q)$ similar to what has been described in Chapman et al. [24]. It is

$$f(q) = \begin{cases} w + (1-w)\left(\frac{q}{\sigma}\right)^4 \exp\left(2 - \frac{2q^2}{\sigma^2}\right), & \text{if } q < \sigma \\ 1, & \text{otherwise,} \end{cases} \tag{2.36}$$

where w is the depth of the filter and σ is the width of the filter. A typical value for the depth is $w = 0.5$ and for the width 10 % of the size of the array.

An example of the effects of the high-pass filter is shown in Fig. 2.11. Both panels show reconstruction results of diffraction data taken on a budding yeast cell. The image on the left is the original (unfiltered) reconstruction. Compared to the filtered image on the right it shows less contrast for internal features and has an overall "fuzzier" appearance.

2.3 Scanning X-ray Diffraction Microscopy

The full-field geometry discussed in the previous section requires the object to be isolated, *i. e.* surrounded by non-scattering empty space. This reduces the number of unknowns to less than the number of available equations. Alternatively, one can record a certain amount of redundant data and use this to reduce the number of unknowns. This is the case in Scanning X-ray Diffraction Microscopy (SXDM), also called "Ptychography".

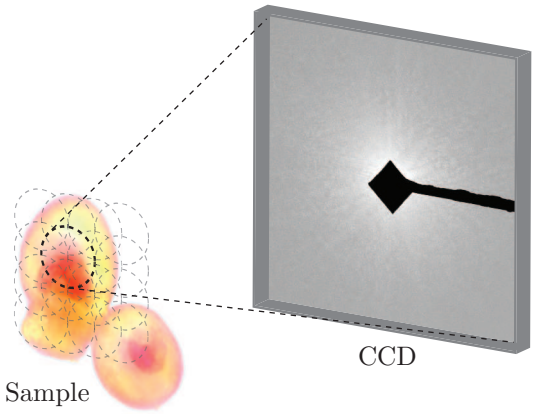

Figure 2.12: Illustration of scanning the probe across the sample in SXDM. For each position of the probe (shown as dashed circle) on the sample, a full 2D diffraction pattern is recorded. Redundancy in the data obtained through overlap in neighboring probe positions enables the reconstruction of extended objects.

The idea to solve the crystallographic phase problem in this way was originally developed in the electron microscopy community by Hegerl and Hoppe [42] in 1970. The idea was later picked up and proposed for super-resolution imaging by Bates and Rodenburg [43] in 1989 and it has been demonstrated soon after for visible light microscopy [44], electron microscopy [45], and x-ray microscopy [46]. The early methods relied on a non-iterative scheme using Wigner-deconvolution to solve for the missing phases. It was not until a few years later that iterative schemes were proposed [47]. Since then, a variety of phase retrieval algorithms have been developed; I will introduce the ones relevant to this work in Sec. 2.3.1.

The basic idea in SXDM is to scan a probe of finite extent across an object and to record the diffraction pattern at each scan position. We create redundancy in the data by stepping by less than the size of the probe so that

the information content between neighboring scan positions overlaps. The situation is illustrated in Fig. 2.12. During the reconstruction process, the redundant information is used as a constraint to reduce the number of unknowns. An advantage of this is that one no longer has to satisfy an isolated object requirement; in principle the object can be of unlimited size. However, remembering Eq. 2.10, we can now no longer assume a plane wave for the illumination and thus we will have to find a way to separate the contributions of the object $F(\boldsymbol{r})$ and the incident wavefield $U^{\text{inc}}(\boldsymbol{r})$ in order to be able to solve for the object function. One way is to know the incident wavefield from a different measurement (see *e. g.* Quiney et al. [48]) and to use this information to separate object and probe. Even if such a measurement can be done for the particular probe that one has to work with, it comes with the obvious disadvantage of having to perform a second experiment. Moreover, the amount of detail that one can achieve for the object function is limited by the amount of detail within the probe function which can make the probe-characterization experiment very challenging. In Sec. 2.3.1, I will describe a class of algorithms that are capable of reconstructing both the object and the probe function simultaneously [49, 50], requiring only a rough estimate of the probe function as a starting guess.

2.3.1 The Reconstruction Process

Let's assume we have a probe function $U^{\text{inc}}(\boldsymbol{r}-\boldsymbol{r}_j)$ incident on an object $F(\boldsymbol{r})$ and that we record the diffraction pattern for a number of different positions $\boldsymbol{r}_j$ by scanning the probe across the object. At each scan position we will have an exit wave leaving the object that is defined as

$$\psi_j(\boldsymbol{r}) = U^{\text{inc}}(\boldsymbol{r}-\boldsymbol{r}_j)F(\boldsymbol{r}). \tag{2.37}$$

What we measure with our detector in the far field is then the square of the absolute value of the Fourier transform of this quantity $I(\boldsymbol{q})$.

Available Constraints

Again, we can rely on a set of two constraints for the reconstruction. The first one, based on the measured diffraction intensities, is the same modulus constraint that is used for full-field XDM and its formalism is described in Eq. 2.29. Instead of the support constraint, the second constraint is the overlap constraint, making use of the redundant information from partially overlapping

regions on the object. It can be written as

$$P_{\mathrm{O}}\psi = \hat{U}^{\mathrm{inc}}(\boldsymbol{r} - \boldsymbol{r}_j)\hat{F}(\boldsymbol{r}),$$

where $\hat{U}$ and $\hat{F}$ are the current guesses of the probe and the object function, respectively.

Reconstruction Algorithms

There are several different ways of updating the object and probe guesses and of calculating the next iterate. I will restrict myself to the algorithm that was used to obtain the results discussed in Ch. 4.

Difference Map for SXDM The extension of the Difference Map algorithm to SXDM was conceived of by Thibault et al. [49]. The update rules for the object and the probe function, respectively, are

$$\begin{aligned}
\hat{F}(\boldsymbol{r}) &= \frac{\sum_j \hat{U}^*(\boldsymbol{r} - \boldsymbol{r}_j)\psi_j(\boldsymbol{r})}{\sum_j \left|\hat{U}(\boldsymbol{r} - \boldsymbol{r}_j)\right|^2} \\
\hat{U}(\boldsymbol{r}) &= \frac{\sum_j \hat{F}^*(\boldsymbol{r} + \boldsymbol{r}_j)\psi_j(\boldsymbol{r} + \boldsymbol{r}_j)}{\sum_j \left|\hat{F}(\boldsymbol{r} + \boldsymbol{r}_j)\right|^2},
\end{aligned}$$

where the sum is over all positions of the probe. For better legibility I have dropped the superscript of the probe function. The next iterate $\psi_j^{(n+1)}$ is calculated using the difference map as described in Eq. 2.33 but using the overlap constraint instead of the support constraint. With $\gamma_{\mathrm{S}} = \beta^{-1}$, $\gamma_{\mathrm{M}} = -\beta^{-1}$ and $\beta = 1$ it becomes

$$\psi_j^{(n+1)} = \psi_j^{(n)} + P_{\mathrm{M}}\left[2P_{\mathrm{O}}\left(\psi_j^{(n)}\right) - \psi_j^{(n)}\right] - P_{\mathrm{O}}\left(\psi_j^{(n)}\right), \tag{2.38}$$

where each exit wave for each probe position is updated. The entire reconstruction process is as follows:

1. Start with an initial guess of the probe function U as described above.
2. Initialize all exit waves ψ_j by setting them equal to the probe function.
3. Compute the object guess $\hat{F}$ according to the update rule above.
4. After some iterations, also update the probe function $\hat{U}$ by alternating between both update rules.

5. Calculate the next iterate of the exit waves $\psi_j^{(n+1)}$ according to Eq. 2.38.

6. Start over with step 3

Convergence is monitored with the error metric described above. After convergence has been observed, the solution is given by the current object and probe guesses, $\hat{F}$ and $\hat{U}$, respectively.

Initial Guess of the Probe Function

The initial guess of the probe function is a crucial step. A bad initial probe guess can lead to prolonged convergence times of the algorithms. In a discussion with Pierre Thibault, I found out that one can get reasonably good starting guesses by summing up all recorded diffraction patterns and taking an inverse Fourier transform, *i. e.*

$$U(\boldsymbol{r}) = \mathcal{F}\mathcal{T}^{-1}\left[\sum_j \sqrt{I_j(\boldsymbol{q})}\right]. \tag{2.39}$$

The idea is that by summing over all diffraction patterns, contributions of the object are averaged out and only contributions of the probe remain. Another option is to use knowledge of the probe defining aperture to come up with a guess of the probe function in the object plane, *e. g.* if the beam defining aperture is a pinhole, then one can get a good approximation of the probe by propagating the pinhole to the object plane.

Error Metric

In SXDM it is convenient to monitor the changes between two different iterates. Thus, the error metric is defined as

$$E_{n+1}^2 = \sum \left|\psi^{(n+1)} - \psi^{(n)}\right|^2. \tag{2.40}$$

In this case, ψ is the 3D vector of all 2D exit waves ψ_j.

2.3.2 Experimental Requirements

The experimental requirements are very similar to what has already been described for full-field XDM. We have the same oversampling requirements (see Sec. 2.2.2) but in the case of SXDM it is the autocorrelation of the probe function U^{inc} that needs proper sampling, rather than the autocorrelation of the object function F.

Since we are using the same CCD-type detector for our SXDM experiments, we also have the same dynamic range issues that have been described in Sec. 2.2.4. Again, we will resort to using a beamstop to block parts of the intense low spatial frequency signal, which also means that we have to assemble the full 2D diffraction pattern from several beamstop positions. I will talk about the optical setup in more detail in Ch. 4.

Overlap parameter

Experimentally important is the overlap parameter, *i. e.* the relative degree of overlap between two different probe positions. Detailed simulations and experimental data have shown [51] that a relative overlap parameter of 60 % is sufficient for high quality, highest resolution reconstructions. One can get away with an overlap parameter as little as 30 % if the main concern are low-dose imaging and fast scanning. To be able to dial the correct step size for the probe scans, one has to know the size of the footprint of the probe on the sample.

2.3.3 Stability of the Experimental Setup

A major difference between full-field XDM and SXDM is that in SXDM we are sensitive to small transverse shifts of the object with respect to the illumination function:

In the full-field case we measure the square of the Fourier transform of the object $F(\boldsymbol{r})$ alone. A transverse shift in object position that is small compared to the coherence width of the illumination will only produce a linear phase ramp in the complex scattering amplitudes in the far-field that is lost in the detection process. The same is true for small transverse vibrations of the object.

In SXDM, on the other hand, we are measuring the square of the Fourier transform of the product of object times probe $F(\boldsymbol{r})U^{\mathrm{inc}}(\boldsymbol{r}-\boldsymbol{r_j})$. A change in relative position Δx will change the diffraction pattern for length scales of up to Δx. In other words, we need to know the relative positioning of object and probe to at least the same resolution that we want to achieve in the reconstruction. It is possible to correct for some aberrations in the positions by treating them as additional variables to be recovered during the reconstruction process [52], however this only works for small deviations from the target positions. Again, the same is true for vibrations of the object with respect to the illumination. If these vibrations occur on time scales shorter than the longest exposure time, they cannot be corrected for. Obviously, if a resolution of 10 nm is to be achieved, one needs a very stable experimental

setup. I will talk about my attempts to improve the stability of our setup at beamline 9.0.1 at the Advanced Light Source in Sec. 4.1.4.

2.3.4 SXDM in 3D

What I have described so far will lead to a 2D reconstruction of an extended object. It is possible to record a series of 2D SXDM data sets while rotating the object and then perform a tomographic reconstruction on the set of 2D reconstructions obtained for each angle. It should also be possible to come up with a scheme to assemble the recorded 2D data sets into a 3D data cube directly in Fourier space (similar to Eq. 2.15) and to reconstruct a 3D representation of the object. However, it has not been shown yet.

2.3.5 Other Experimental Geometries

There are other experimental geometries besides full-field XDM and SXDM. I will only mention two of them here, and only for completeness since I have not used any of them for the results shown in later chapters. In Fresnel Coherent Diffractive Imaging [53, 54] one uses as probe a wavefield with nonzero phase curvature. As in full-field XDM one needs to have an isolated object. Due to the phase curvature of the illumination, however, rapid convergence to a unique solution is achieved.

In Keyhole Coherent Diffractive Imaging [55] the concept of Fresnel Coherent Diffractive Imaging is extended to objects of unlimited size by imposing an additional constraint that the illumination be of finite extent (such as a diverging focused beam). For both geometries the quality of the reconstruction depends critically on the knowledge of the illuminating beam which must be known to a resolution equal or better than the desired resolution of the reconstruction. The latter geometry has the advantage that when scanning an extended sample there needs to be no overlap in illumination between neighboring scan positions. This decreases the dose administered to the sample.

2.4 Resolution and Quality of a Reconstruction

2.4.1 Radiation Damage

Before I explain ways to estimate the resolution of a reconstruction, I want to talk about the effects of radiation damage first. It is important to know

about the effects of radiation damage and at what radiation levels they occur. This is true especially for biological specimen, where radiation induced structural changes happen at low doses and can make the interpretation of a reconstructed image difficult.

Dose Calculations

First, let's have a look at how to estimate the amount of radiation that an object received. This is commonly measured in terms of the dose, *i. e.* the amount of energy absorbed per unit mass. Remembering Eq. 1.15 we can calculate the decrease in intensity of the incident radiation per thickness as

$$-\frac{\mathrm{d}I}{\mathrm{d}z} = \mu I_0 e^{-\mu z}, \tag{2.41}$$

where z is the penetration depth into the object. We can calculate the energy absorbed at $z = 0$ (the skin dose) as

$$D = \frac{\mu\, n E_{\mathrm{photon}}}{(\Delta x)^2 \rho}, \tag{2.42}$$

where Δx is the pixel size of the reconstruction, n is the number of photons per pixel and ρ is the density of the object. This skin dose will be a measure of radiation that has been absorbed by the object. Unless otherwise noted, it will be calculated for protein with a stoichiometric composition of $H_{48.6}C_{32.9}N_{8.9}O_{8.9}S_{0.3}$ and density of $\rho = 1.35$ g/cm^3 [56]. When recording data for one 2D projection of the sample we typically administer a skin dose of $1 \cdot 10^9$ Gy at our microscope.

Effects of Radiation Damage

A nice overview of primary and secondary effects of radiation damage in organic specimens can be found in Talmon [57]. Here I will restrict myself to the coupled effects of C=O double-bond breaking and mass loss. The latter results from the former because the scission creates smaller molecules that can leave the object by sublimation. Mass loss is observed as structural changes in the object. Obviously this is what we want to avoid for imaging.

At what dose levels do these effects occur in representative biological specimens at the target resolution? There have been a number of experiments for both electron microscopy and x-ray microscopy; a good summary is given in Howells et al. [36]. Specifically for x-ray diffraction microscopy there has been work done by Shapiro [58] that looked into fractional changes in feature

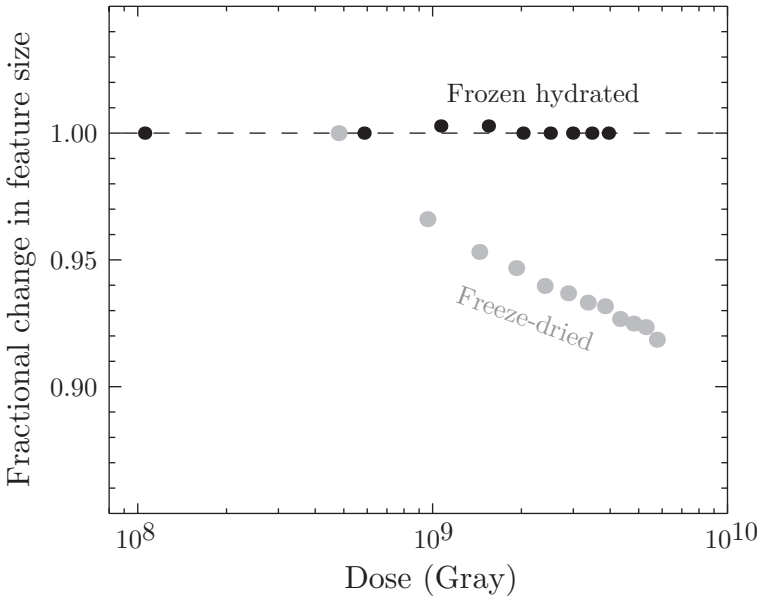

Figure 2.13: Changes in speckle position corresponding structural changes of features 25 nm in size are observed as a function of cumulative dose for room-temperature dried biological specimens. Frozen-hydrated samples do not show any change over the same range of doses. Reprinted from [58] and [22].

size of a specimen as a function of cumulative dose. The results of this work are shown in Fig. 2.13 (reprinted from [58] and [22]). Briefly summarized, the results indicate that freeze-dried samples are stable up to $\approx 5 \cdot 10^8$ Gy while frozen-hydrated samples are stable up to $3 \cdot 10^9$ Gy and beyond. This illustrates that we want to work with frozen samples in order to minimize the visible effects of radiation damage during prolonged 3D data collection, where the cumulative dose exceeds 10^9 Gy.

Minimizing Dose to the Object

When trying to minimize radiation exposure to biological specimen it is important to choose the energy of the x rays to maximize absorption (or phase) contrast of protein in ice. Figure 2.14 shows a plot of contour lines of equal dose required to image a 10 nm protein structure embedded in ice as a function of ice thickness and energy of the incident x rays. "Imaging" in this case means having a signal-to-noise ratio of greater than 5 according to the Rose criterion [59] and the plot shows the dose of either absorption contrast or phase contrast, whichever is lower. From the figure one can see that there are two interesting energy ranges. For samples of a few micron thickness (which are the ones we are interested in), energies in the "water window" [60], between the carbon and oxygen K-edges at 284.2 eV and 543.1 eV, respectively, provide very good contrast for protein in ice. Thus, in our experiments with frozen samples, we use x rays with an energy of 520 eV. The figure also shows that for thicker samples of several tens of microns it is better to work at higher energies.

It is important to note that the maximum allowable dose before structural changes occur is lower for smaller length scales than for larger ones. At the same time, the dose required to image small structures is larger than for imaging large structures. This means that radiation damage sets an ultimate limit to the resolution that can be achieved. Based on experimental data and theoretical calculations, this limit has been estimated to be around 10 nm for protein structures [36].

2.4.2 Two Common Measures of Reconstruction Quality; PRTF and PSD

In the following discussion I will assume a careful experimenter that kept dose levels at the sample below threshold levels for structural changes as discussed above. Let's say a reconstruction has been obtained and we want to know how good the resolution is. Intuitively, one could look at the assembled diffraction intensities and see what spatial frequency the scattered signal extends out

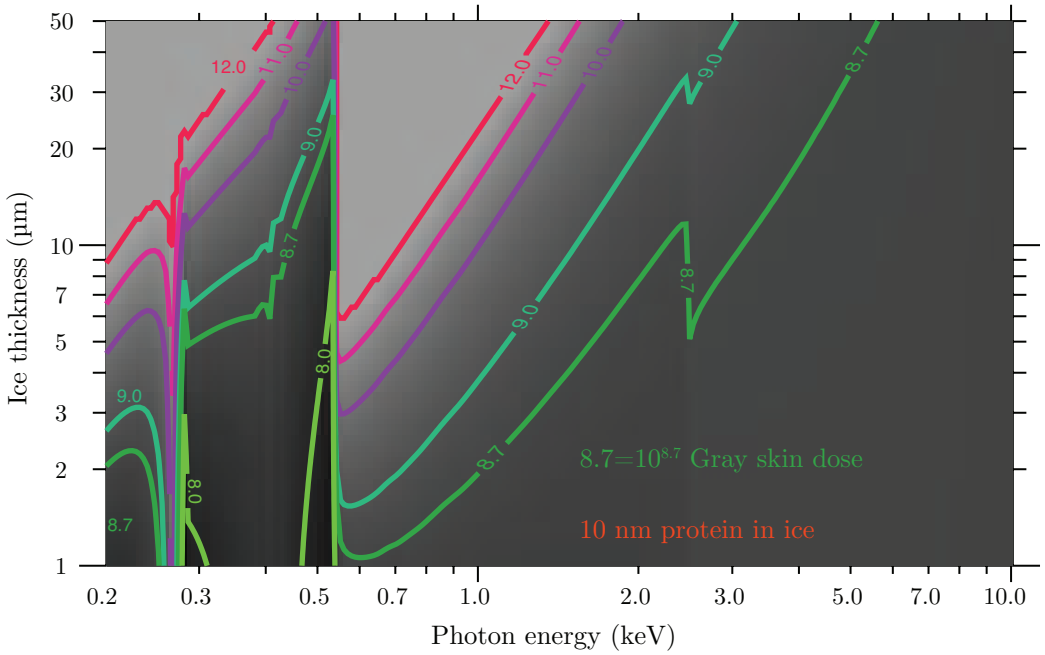

Figure 2.14: Contours of equal dose required to image a 10 nm protein structure in ice as a function of ice thickness and photon energy. Figure courtesy of Chris Jacobsen.

to. Assuming that this is the amount of detail that has been captured in the experiment, the resolution can then be calculated by Eq. 2.16. In fact, the first papers [19, 20] in x-ray diffraction microscopy did just that and used the presence of measured diffraction signal as a function of spatial frequency to estimate the resolution achieved in the reconstruction.

However, the simple presence of signal is only part of the story: one must consider the presence of noise, partial coherence in the beam, and the possible presence of small scatterers outside of the assumed support constraint. Taken together, these effects can lead to a decrease in the consistency of the estimated phases, and Fourier plane pixels which cannot be reliably retrieved will not contribute useful and reproducible information to the reconstructed image. To estimate the real resolution of a reconstruction, one typically does one of the following:

- Determine resolution from a line-out across edge of a sharp feature within object [22].
- Calculate point spread function (PSF) of the imaging system through

correlative microscopy [61].

- Look at the maximum information transfer in the Power Spectral Densities (PSD) of the reconstructed object [62].
- Determine the spatial frequency at which the Phase Retrieval Transfer Function (PRTF) has suffered a significant decline [21, 24, 35].

I want to dwell on the last two points since we will need it later on. The PSD is simply the radially averaged sum over the reconstructed Fourier intensities

$$\mathrm{PSD}(q) = \frac{\sum_{q=\text{const.}} |\Psi(q)|^2}{N}, \tag{2.43}$$

where N is the number of array elements for q = constant. An example of a PSD curve is shown on the right of Fig. 2.15. As I have already explained

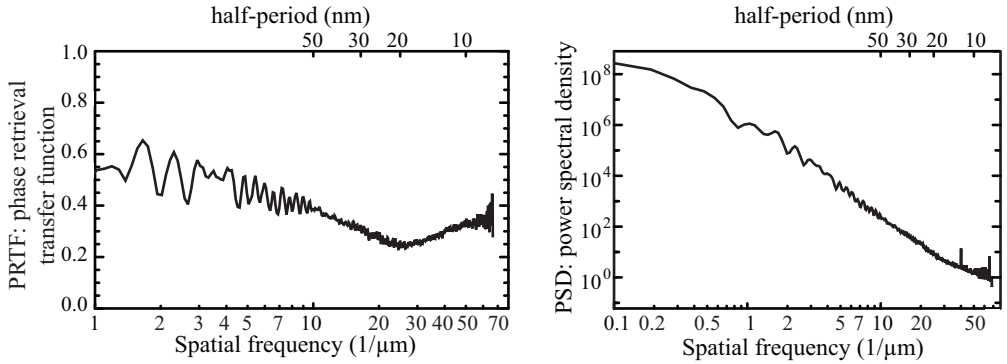

Figure 2.15: Examples of Phase Retrieval Transfer Function (PRTF) and Power Spectral Density (PSD) curves for a reconstructed object. The PSD gives an indication for the maximum information transfer, the PRTF measures consistency in the retrieved phases.

in Sec. 2.2.4, we expect the scattered power to decline as the inverse m-th ($m = 3 - 4$) power with spatial frequency. Plotted on a log-log scale this will result in a declining straight line as can be seen in Fig. 2.15. The interesting point is where the PSD deviates from the straight line and rolls off into a horizontal line. This suggests that noise is the predominant signal at this spatial frequency (we expect the PSD of shot noise to follow the Fourier transform of a delta function, *i. e.* a flat power spectrum) and gives an estimate as to the maximum information transfer within the reconstruction.

Estimating the resolution by itself is not sufficient however. One also needs to look into the "quality" of the phase retrieval process, *i. e.* how consistent is the reconstructed object with the measured data. A tool commonly used to measure the quality of a reconstruction is the Phase Retrieval Transfer Function (PRTF) defined as

$$\mathrm{PRTF}(q) = \frac{\sum_{q=\text{const.}} |\Psi(q)|}{\sum_{q=\text{const.}} \sqrt{I(q)}}. \tag{2.44}$$

It is the square root of the ratio of power spectral densities of reconstructed Fourier magnitudes over measured Fourier magnitudes (the intensity ratio [21, 35]). An example of a PRTF is plotted on the left of Fig. 2.15. The PRTF is a measure of consistency in phase retrieval. When averaging many reconstruction results as described in Sec. 2.2.6 the resulting magnitudes will be equal or smaller than the measured ones due to variations in retrieved phases. In the extreme cases of perfect phase retrieval or completely random phase retrieval, the PRTF will have a value of 1 or 0, respectively. The PRTF is an important tool for x-ray diffraction microscopy and also subject of my research. In Sec. 3.4 I will derive criteria of validity for the PRTF and look into ways of improving its utility for judging reconstruction quality.

2.5 Comparison of XDM to Lens-based Alternatives

Now that I have explained the experimental requirements, reconstruction procedures and post-reconstruction data analyses of x-ray diffraction microscopy, I want to compare it to its lens-based analogue, namely Transmission X-ray Microscopy (TXM). A typical TXM microscope setup is shown in Fig. 2.16. A condenser lens illuminates the sample and an objective lens placed further downstream of the sample focuses a magnified image of the sample onto a detector, for example a CCD. Because the index of refraction of most elements is slightly less than unity in the x ray regime, one relies on diffractive lenses for focusing rather than refractive ones. These Fresnel zone plates are comprised of alternating opaque and transparent rings that are of equal area and whose radii satisfy the condition that all transmitted light constructively interferes in the focal spot. For incoherent illumination of the object, the width of the outermost zone determines the resolution of the resulting image (see for example Michette [63]). In the soft x-ray regime, resolutions down to 12 nm [64] have been demonstrated for specific cases. However, at these wavelengths

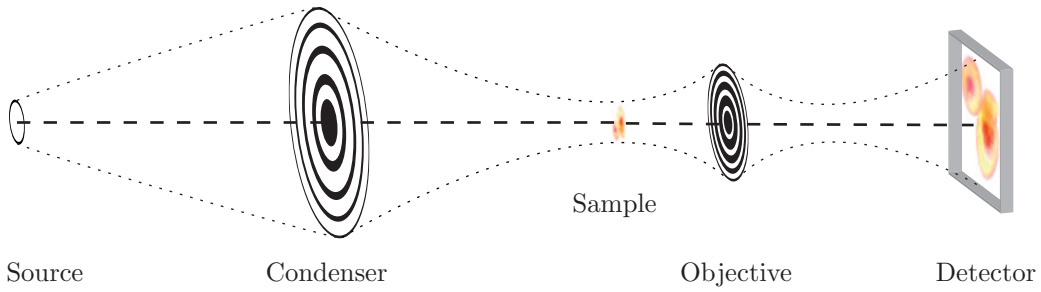

Figure 2.16: Schematic of a typical TXM system. The sample is illuminated by a condenser zone plate and the objective zone plate phases the light coming from the sample to form a real image on the detector.

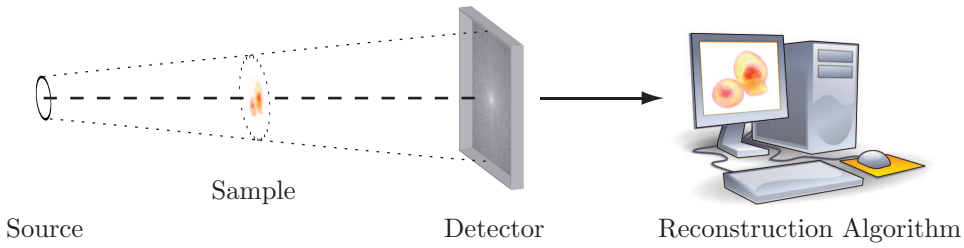

Figure 2.17: Typical setup of a XDM system. The sample is illuminated with coherent light and the diffracted signal is recorded with a detector. Without an objective lens image has to be reconstructed computationally.

Fresnel zone plates with outermost zone widths of less than 20 nm suffer from sub-millimeter focal lengths and diffraction efficiencies of less than 10%. The former poses a practical challenge of how to implement a 3D tomographic setup with the optic that close to the sample, while the latter results in an increase in dose to the specimen. In fact, recent simulations [65] have shown that to image an object with the same resolution, XDM is up to 50-times more dose efficient than a TXM system that uses zone plates. In the spirit of the discussion of Sec. 2.4.1, XDM is advantageous when working with radiation sensitive biological materials.

On the other hand, TXM is a direct imaging method. The objective zone plate phases the light coming from the sample and forms a real image on the detector. Besides alignment of the individual 2D projections and routine tomographic reconstruction, there is no computationally intensive phase retrieval process. When compared with a schematic of a typical diffraction setup as

is shown in Fig. 2.17, it becomes obvious that we have traded the inefficient objective lens with intensive post-experiment data processing requirements.

2.6 Motivation for Thesis Work

Based on the discussion above, XDM is a very attractive technique for high resolution imaging of biological specimens because a) unlike electron microscopy it can image thick unsliced biological specimen in their natural hydrated state and b) it is more dose efficient than its x-ray competitor TXM. At this point I would like to talk about some challenges in XDM and what I have done to address them.

Computational Methods The post-experiment processing of data in x-ray diffraction microscopy is often time-consuming and difficult; it is computationally demanding but results are also difficult to interpret. In Ch. 3, I show how to improve, automate and speed up the data analysis process and how to better judge the quality of a reconstruction.

Scanning X-ray Diffraction Microscopy Full-field XDM requires the object to be isolated. This is difficult to achieve for frozen-hydrated specimens. Scanning x-ray diffraction microscopy does not require an isolated object and is therefore an attractive alternative. In Ch. 4, I will present several implementations of a SXDM setup for our existing microscope and show preliminary results from recent experimental data.

Chapter 3

Computational Methods in XDM

The post-experiment processing of data in an XDM experiment poses some computational challenges. Apart from the actual reconstruction process one has to deal with the assembly of 2-dimensional diffraction patterns from a large set of raw diffraction data (this is done in order to increase the dynamic range of the imaging system, improve on signal-to-noise statistics, and to minimize the missing area in the center of the diffraction pattern as described in Sec. 2.2.4). Given the amount of data processing necessary, it is desirable to automate the assembly of raw data as much as possible, without introducing artifacts due to generalizations, and to integrate it into the general reconstruction process. This is especially important towards 3D imaging, where a number of 2D projections must be assembled before being merged into a 3D cube. Additionally, even if a preliminary result has been reconstructed, there is no definitive answer as to whether or not a better result with more consistently retrieved phases can still be obtained. While the PRTF (introduced in Sec. 2.4.2) is a common measure of the quality of a reconstruction, no consensus on its use has emerged in the community and its utility is limited by artifacts caused from incorrect support constraints, from errors in measurement of the Fourier plane intensity or from noise which produces consistent yet erroneous structure.

I address these challenges in several ways. In Sec. 3.1, I will introduce a file format for diffraction imaging that carries along all important information about a data set and thus integrates the various steps of data analysis. To further facilitate data analysis, I have developed a subroutine library with support for parallel processing that allows for platform independent programming of data analysis tools, as is described in Sec. 3.2. Together, the file format and the library provide a standardized way of sharing reconstruction results with

collaborators. I will further introduce a program that automates and improves the assembly process of 2D diffraction patterns; this is described in Sec. 3.3. Finally, in Sec. 3.4, I will present some criteria of validity for the PRTF and suggest a modified version that improves its utility for judging reconstruction quality. The collective improvement to the reconstruction process is illustrated with results from recent experimental data.

3.1 A File Format for Diffraction Experiments

The analysis process of diffraction data involves quite a number of different parameters that one needs to keep track of, from filenames of raw data to experimental parameters, to details of the reconstruction process. A typical 3D data set from our microscope will involve some 100 2D diffraction patterns and several thousand raw CCD frames. The amount of data produced at 4^{th} generation light sources (x-ray FELs) is far greater even, so the challenge to manage this amount of data will only get greater. In order to facilitate the data analysis process and to prevent confusion between similar data files, it is desirable to store all essential information together with the data itself into one binary file. The resulting file should be in a format that is flexible, scalable and supported by a wide variety of platforms and programming languages.

I propose here a file format for diffraction experiments based on the widely available, platform-independent HDF5 hierarchical data file format library[1]. The main structural elements of the HDF5 file structure are `group`, `datatype`, `dataspace`, and `dataset`. HDF5 `group`s are structures containing `dataset`s or more `group`s (similar to folders in the file structure of modern operating systems). `Dataset`s hold the data and supporting information. The header of a `dataset` contains the parameters `datatype` and `dataspace`. The `datatype` defines the type of the data (int, char, double, *etc*), the `dataspace` defines the dimensionality of the `dataset`. The advantages of HDF5 are that it is very flexible, very well supported and available for a wide variety of computing platforms and languages. Particularly interesting to us is that `dataset`s can be initialized with an unlimited `dataspace` and that they can be read into memory in small chunks of data. The former means that we can easily extend the size of a `dataset` after it has been created. This even works if the file has already been saved. The latter enables us to overcome memory shortages when reading large 3D arrays that can be as big as 16 GB.

[1] http://www.hdfgroup.org/HDF5/

3.1.1 General Structure of the File Format

This section is concerned with the overall structure of our file format; the library routines that perform the actual read and write processes are described in Sec. 3.2. The general idea for the file format is to identify and store the relevant data sets and additional important information that is needed during the reconstruction process and for bookkeeping. We typically start out with raw data which is currently recorded in a netCDF[2] format. We then assemble the raw data from different beamstop positions into one full 2D diffraction pattern which will be saved along with supporting information in the file format described here. If we have recorded a 3D data set, we take several of the assembled 2D diffraction patterns and merge them into a 3D data cube which will also be saved using the same file format. Having the same file format with (almost) the same information for 2D and 3D data sets makes sense since the basic data analysis techniques are the same for 2D and 3D. Using the structural elements provided by the HDF5 library, we define `groups` for each relevant data set that hold the actual data as well as supporting information. An overview over the general structure of the file format is shown in Fig. 3.1. The file format currently contains five groups each with a different purpose, as is explained below. Note that except for the `/comments` group, all groups have a version attribute (shown in red). This is important for backwards compatibility should the file format definition be updated in the future.

3.1.2 The `"/comments"` Group

This group will contain strings identifying the specimen that data was recorded on, the date and time of the assembly, and an array of comments meant to keep track of what type of analyses were performed on the particular data set. There is no explicit format requirement for the strings in the array of comments; however, for uniformity I suggest to always use the following string template:

```
'NAME_OF_PROGRAM, time of analysis, relevant parameters'
```

It is up to the programmer of a specific data analysis tool to make sure that the software adds an entry to the `/comments` group. A detailed definition of the `/comments` group is given in Appendix A.1.

[2] http://www.unidata.ucar.edu/software/netcdf/

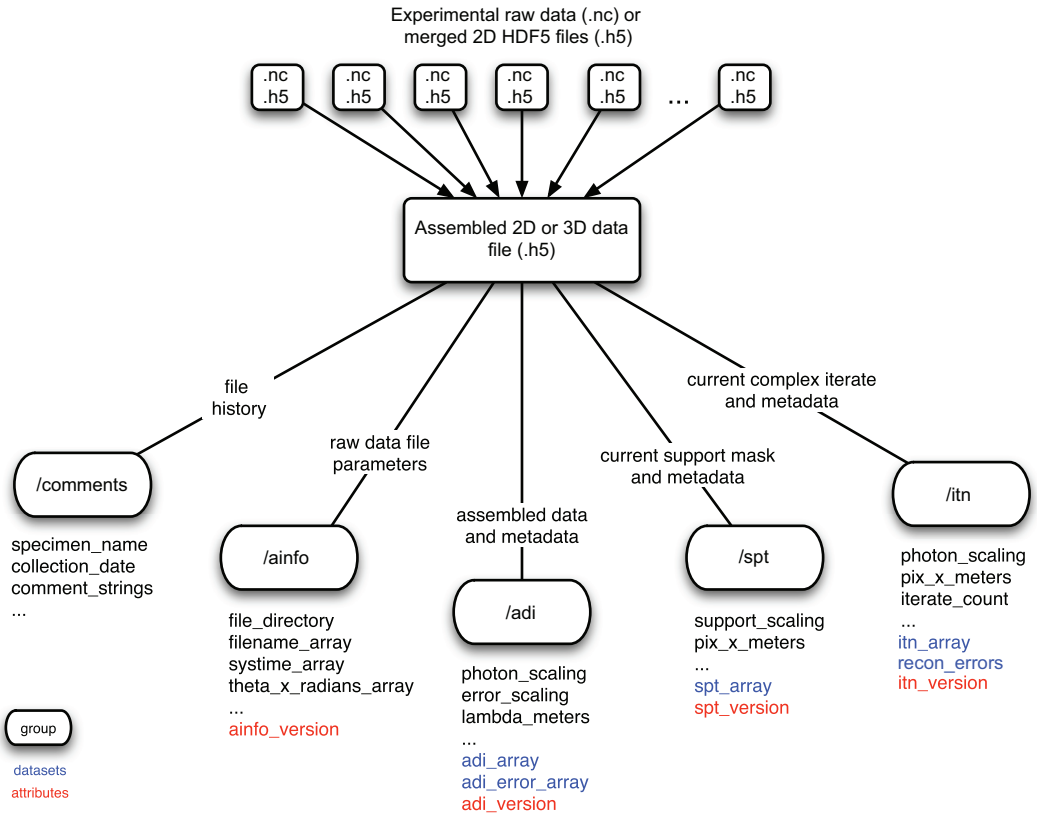

Figure 3.1: Schematic of the internal structure of the file format. The raw data will be assembled in an HDF5 file with predefined groups and data sets that are meaningful in the context of diffraction microscopy.

3.1.3 The Assembly Info Group - `"/ainfo"`

The `/ainfo` group keeps track of all the information that is not the same for all data files that were merged. For example, it will store the names of all files that were merged as well as their date of creation. Since the file format is supposed to be flexible and accommodate 2D as well as 3D data sets, not all of the structure tags of the `/ainfo` group will be used for each data file. For example, for 3D data, the `/ainfo` group will save an array of rotation angles (corresponding to diffraction patterns taken at different rotation angles of the specimen with respect to the x-ray beam) of the 2D data files that were merged into the 3D data cube. For a 2D data set we don't need to save this information because the rotation angle does not change. The file input-output

libraries (see Sec. 3.2) are written so that they handle missing structure tags gracefully. A detailed definition of the `/ainfo` group is given in Appendix A.2.

3.1.4 The Assembled Diffraction Intensity Group - "/adi"

This group holds the actual 2D or 3D data array that results from the assembly (called `adi_array`). It also stores experimental information that is the same for all data files that were merged, such as the wavelength that the experiment was carried out with or the sample-detector distance.

The optional array `adi_error_array` ($e_{i,j,k}$) can be used to store information on the error of each voxel in an assembled data file. It is meant to store an array of relative errors that when multiplied with the pixel values of the associated `adi_array` ($a_{i,j,k}$) results in an absolute error of each pixel of the form $\sigma_{i,j,k} = e_{i,j,k} \times a_{i,j,k}$; for example, the automated assembly program that I will introduce in Sec. 3.3 determines the error from CCD read-out and thermal noise plus photon statistics. This information will be used during reconstruction when applying the Fourier-modulus constraint according to Eq. 2.29. A detailed definition of the `/adi` group is given in Appendix A.3.

3.1.5 The Support Mask Group - "/spt"

The support mask (a real-space array of 1 byte per pixel) defines a support constraint in 2D or 3D. It is usually not present in the beginning but will be created once the user defined a support mask that he wants to use for the reconstruction. As I have explained in Sec. 2.2.5 the support guess is updated several times during the reconstruction process until a satisfactory result is obtained. To prevent confusion and for convenience, it is important to store the current support guess along with the current iterate amplitudes (see below) and the assembled data in one file. A detailed definition of the `/spt` group is given in Appendix A.4.

3.1.6 The Iterate Amplitude Group - "/itn"

This group holds the result of a reconstruction calculation (a complex amplitude in real space) in `itn_array`. It also stores an array of reconstruction errors (`recon_errors`) calculated for each iteration according to Eq. 2.30. Again, it is convenient to store the reconstructed result together with the assembled data. For example, to calculate the PRTF of a reconstruction according to Eq. 2.44 one needs both the assembled data and the reconstructed result. Keeping these

different arrays in one file minimizes potential confusion and error. Since the `itn_array` is a complex array, it requires a lot of memory, especially for 3D data sets. We therefore make use of the property of the HDF5 library that allows us to only read little chunks of data into memory. This is described in more detail in Sec. 3.2. A detailed definition of the `/itn` group is given in Appendix A.5.

3.2 A Platform Independent Subroutine Library with Parallel Processing Capabilities

Our subroutine library has two main parts. First, it consists of an array of functions that perform the read and write operations that are required to create and update files in a manner consistent with the definition of the file format (described in the previous section). Its name is the `dm_fileio` library and it is described in more detail in Sec. 3.2.2. Second, it defines a set of basic array manipulation routines that are common to reconstruction algorithms of diffraction data. This subset of functions is called the `dm_array` library, explained in more detail in Sec. 3.2.3. The entire library is available through Concurrent Versioning System (CVS) from our server[3] and a comprehensive user guide[4] is available online as well. Finally, I will give an overview of data analysis tools that make use of the library in Sec. 3.2.4.

The basic idea of the subroutine library is to provide high-level functions for common operations (such as reading and writing a file) that worry about lower-level details such as platform specific implementation or parallel processing. In other words, we want to present the user with a set of routines to the tune of `write_file(...)`, `get_phases(...)`, and so on whose outcome is obvious but which hide all the nitty-gritty detail from the user. That way, the user can write data analysis tools which will work automatically on all platforms that are supported by the subroutine library.

Internally, this is achieved by a) a `Makefile` that defines which libraries to use, where they are located for the platform the library is compiled on, and that sets appropriate `define`-commands and b) `#ifdef`-statements in the library that invoke the correct code depending on what `define`-commands were set in the `Makefile`. Currently, the library supports single processor machines using `fftw3`[5] for Fourier transforms, and multiple processor machines using

[3] http://xray1.physics.sunysb.edu/data/software.php
[4] http://xray1.physics.sunysb.edu/~micros/diffmic_recon/
[5] http://www.fftw.org/

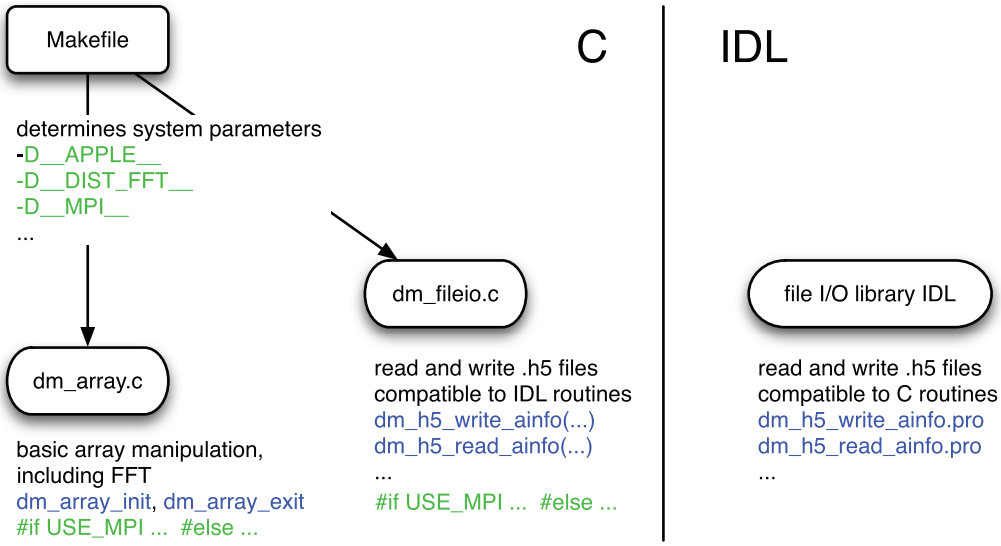

Figure 3.2: Schematic of the basic structure of the subroutine library. In the `C` implementation, a `Makefile` defines the system relevant parameters which (with the help of `#ifdef` statements) invoke different snippets of code in the library when compiled. The file input/output routines are also implemented in IDL which is used for visualization and post-reconstruction data analysis.

Message Passing Interface (MPI) [66, 67] and Apple's `dist_fft` library [68] for distributed Fourier transforms. Up to now, it is still the responsibility of the user to modify the `Makefile` according to the specifics of the platform being used. A schematic overview of the library is given in Fig. 3.2. It is advantageous to use fast `C` code for memory-intensive number crunching (reconstructions which involve many Fourier transforms on large arrays), and a higher-level programming language for convenient visualization and post-reconstruction data analysis (we use the Interactive Data Language (IDL) from ITT[6] which is very similar to Matlab, another popular choice). This is why the `dm_array` library only exists in `C` but the file input/output routines are written in both languages. The following sections and code examples are assumed to refer to the `C` implementation of the library unless otherwise noted.

[6] `http://www.ittvis.com/IDL`

3.2.1 Basic Conventions: the File `dm.h`

There are some simple conventions that have to be followed when using the library. Most of these conventions are defined in the header file `dm.h`; it has to be included in any program making use of the library. I will list here the most important coding conventions.

- Use of data types defined in `dm.h`. This ensures that we can work with double and single precision arrays and takes care of custom types defined by certain libraries, such as the complex-valued array types in `fftw` and `dist_fft`.

- Use of array structures defined in `dm.h`. The array structures will automatically use the correct data types and also hold additional information such as the number of array elements per process (needed for the `dm_array` library).

- Use of memory allocation and array-element selection macros defined in `dm.h`. This will ensure that the array structures are correctly initialized and that the memory is allocated according to the requirements of the Fast Fourier Transform (FFT) library that is being used. It also allows us to work with either

  – split complex arrays
  $(\mathrm{arr} = \{[\mathrm{Re}_0, \mathrm{Re}_1, \ldots, \mathrm{Re}_n], [\mathrm{Im}_0, \mathrm{Im}_1, \ldots, \mathrm{Im}_n]]\})$ or

  – interleaved complex arrays
  $(\mathrm{arr} = \{[\mathrm{Re}_0, \mathrm{Im}_0], [\mathrm{Re}_1, \mathrm{Im}_1], \ldots, [\mathrm{Re}_n, \mathrm{Im}_n]\})$

  in memory.

- Initialize and finalize the subroutine library by calling the functions `dm_init` and `dm_exit` before a routine from the library is called and after the last routine from the library has been called, respectively. These routines will gather important runtime dependent parameters, such as the number of processes and the current rank of each process when using MPI. Thus the basic structure of a user written program is as follows:

```
int my_rank,p;

/* Obtain info on number of processes and rank of current
 * process. Initialize MPI session if applicable.
 */
dm_init(&p,&my_rank);
```

```
/* call to library functions */
...

/* Finalize MPI session if applicable */
dm_exit();
```

3.2.2 File-Input/Output Routines: the `dm_fileio` Library

The `dm_fileio` library is the actual software implementation of our file format that I described in Sec. 3.1. It contains all functions that are needed to create, read and write the various parts of the file format. Its routines are defined in the header file `dm_fileio.h` which needs to be included in any program that wants to make use of the library. The names of the functions are as descriptive as possible to make clear what their main purpose is. There are two prefixes `"dm_"` and `"dm_h5_"` to the function names. The first one is used for routines that will not directly access a file on disk (read or write) while the second one indicates that a routine will do just that. The library functions are written to be very flexible and thus accommodating to the user. Here are the most important features:

- Floating point data arrays will be written according to the value of `typedef dm_array_real`: that is, as double-precision if `DM_ARRAY_DOUBLE` is specified at compile time (`C`) or as a keyword (IDL), or otherwise as a single-precision float. When reading the file, the file's native data type will be converted into float or double based on the type of `dm_array_real`.

- The `dm_fileio` routines automatically handle any byte-swapping that might be required. This enables one to read and modify files that were created on a big-endian system on a little-endian system and vice-versa. This is important since many data analyses involve the use of several computing systems, such as reconstruction on a remote Apple G5 cluster and visualization on a local Linux PC).

- Not all files will have all groups. The `dm_fileio` library handles read or write requests of missing groups gracefully. The same is true for missing values in one of the arrays of supporting information (such as the array of rotation angles in the `/ainfo` group).

- The functions are written to deal with large arrays by reading them in small slices rather than the whole array at once; this is a feature of the underlying HDF5 library. It is also important to prevent memory allocation issues and it also facilitates parallelization, as described below.
- The functions are capable of parallel processing using MPI. If applicable, both read and write operations can be split up among the processes according to the parameters determined by the `dm_init(...)` routine. In both cases, the process with rank 0 will be the only process that does the actual reading and writing but it will do so in slices that it sends/receives to/from other processes including itself.
- Provision for future additions is built-in through version attributes that are saved with each group (except the `/comments` group).
- The library exists in `C` and IDL. Both implementations are fully compatible.

3.2.3 Basic Array Manipulation Routines: the `dm_array` Library

The second part of the subroutine library is the `dm_array` library. It defines a set of basic array manipulation routines that are commonly used in reconstruction processes of diffraction data. Its functions are defined in the header file `dm_array.h` that needs to be included in every program making use of its routines. The function names are very descriptive and indicate the main purpose of each routine. All names have the prefix `"dm_array_"` to identify the library they are part of. The main features of this library are

- The library is fully integrated with the `dm_fileio` library described above by relying on the same data types and macros defined in `dm.h` as described in Sec. 3.2.1.
- All routines are capable of parallel processing using MPI.
- Since the `dm_array` library uses the types and macros of `dm.h` it automatically works with single- or double-precision, split or interleaved complex arrays.
- The library includes an FFT wrapper routine that invokes either the `fftw3` library for single processor machines or the `dist_fft` library for multi-node Apple machines. It provides the user with a simple interface while taking care of the idiosyncrasies of each FFT implementation.

3.2.4 The Code Archive

Based on the subroutine library, our group has written an extensive code archive of data assembly, reconstruction and post-reconstruction analysis tools. All of these tools are shared through Concurrent Versioning System (CVS) and all of them make use of the IDL or `C` implementations of the subroutine library. Because of this, they are all compatible, in that in principle any routine can be used on any file by anyone in the group without complications. I want to introduce some of them and close with a real life example.

The IDL archive `"dm_recon"` includes a number of routines for different stages of the data analysis process. There are routines that prepare data sets for reconstruction (centering of diffraction patterns, defining initial support guesses, *etc*), routines that help during reconstruction (redefining the support guess, removing phase ramps, *etc.*) as well as routines for post-reconstruction data analyses (averaging, adjusting global phase, high-pass filtering, calculation of phase retrieval transfer functions, and so on). Most of these routines have been written either by Johanna Nelson or me.

The 2D assembly program `AMP` was written to automate and improve the assembly of 2D diffraction patterns. I will explain it in more detail in Sec. 3.3. It saves its output into our custom file format defined above using the IDL implementation of our subroutine library. `AMP` was written by me and debugged mainly with the help of Johanna Nelson; it's an ongoing process!

Finally, I want to introduce two more programs that I wrote: the reconstruction program `"dm_recon.c"` and the 3D assembly program `"cewald.c"`. The program `dm_recon.c` was written in C making use of the C implementation of the subroutine library. Because of this, it is capable of reconstructing large 3D data sets using parallel processing on our 16-node Apple G5 cluster but it can also be used on single-processor machines. Its output is saved in the respective groups of our custom file format. The program `"cewald.c"` also makes use of the C implementation of our library. Its main purpose is to take a many 2D diffraction patterns produced by `AMP` and to assemble them into a 3D cube according to Eq. 2.15. In the process it will use a lot of experimental parameters such as wavelength, pixel size of the CCD, and angle of rotation that thanks to our custom file format have been stored with the 2D files created by `AMP`.

A Case Study: From 2D to 3D

To illustrate the convenience and advantages of the library and the code archive I want to describe an example data analysis process of actual experimental data that involved several scientists and several programs.

In August 2008 the experimental team was taking data on nano-porous glass particles. After they had found a promising scatterer they recorded a 2D diffraction pattern on it and immediately started to analyze it to see if it was worth spending several hours recording 3D data. Three quite similar reconstructions were obtained within 24 hours so it was decided to record 3D data on it (the details of the rapid 2D reconstructions are discussed elsewhere [69]). One of the three reconstructions was obtained by Johanna Nelson. She assembled the raw data using the automated merging program (`AMP`) introduced in Sec. 3.3. Because her own reconstruction code makes use of the IDL implementation of the subroutine library, it was straightforward for her to access this file and reconstruct it.

Based on the preliminary information obtained from the 2D assembly and reconstruction, Joshua Turner assembled all remaining 2D data sets of the 3D tilt series using `AMP`. He then centered and cropped each file using routines from the IDL `dm_recon` archive. After this was done, I assembled his 2D files into a 3D cube using `cewald.c` and started reconstructing the 3D data set using `dm_recon.c`. The whole process was straightforward thanks to the subroutine library and the file format.

3.3 `AMP` - An Automated Assembly Program

As I have mentioned in Sec. 2.2.4, the dynamic range of the scattered intensities can present challenges for many pixelated x-ray detectors. To recapitulate, in using direct detection on CCDs one generates several hundred electron-hole pairs per soft x-ray photon absorbed, which when coupled with a full-well charge capacity of 10^5–10^6 electrons means that a dynamic range of only something over three orders of magnitude can be achieved in a single recording. (Pixel array detectors are beginning to overcome these limitations, but high pixel number detectors with good sensitivity for soft x rays are not yet widely available). As a result, a common experimental strategy that I have already described in Sec. 2.2.4 is to to use a movable beamstop to block various parts of the strong, low spatial frequency signal while adjusting the exposure time to collect the weaker, high spatial frequency signals. These various intensity recordings must then be combined into a full 2D diffraction pattern. Typically, one merges multiple Fourier plane intensity recordings by using a per-dataset procedure based on manual adjustments of noise thresholds and requested exposure times. This is not only tedious but will also result in slight user-dependent variations in the assembled diffraction pattern. In this section I want to introduce a program that automates this merging process and introduces improvements in two key areas, resulting in improved reconstruction

quality [70].

3.3.1 Manual Data Assembly

First, lets look at a typical per-dataset assembly, where the following procedure is performed for each beamstop position. Saturated pixels (where the full-well capacity of the CCD had been reached) are first removed. Next, pixels with anomalously high values due to large charge deposition by cosmic ray events are found and removed, as are pixels with anomalous values due to either manufacturing flaws or radiation-induced damage. Individual recordings are then normalized to the synchrotron beam current, after which images with the same exposure time are averaged and a noise threshold floor is applied. The area behind the beamstop is then masked. Beam-normalized averages from the different exposure times are then scaled and averaged, taking care to include only those pixels with non-zero signal. In what follows I will refer to results obtained with this protocol as hand-assembled data sets.

3.3.2 Improvements Provided by `AMP`

The assembly performed by `AMP` improves upon this basic assembly protocol in several key areas. The first difference is a quantitative analysis of the CCD chip. Given a series of dark current images at different exposure times, `AMP` will calculate an average dark current and the variation in dark current either as an average for all pixels in the chip, or (if enough redundant dark current files are present) on a per-pixel basis. The variance in each pixel corresponds to the total CCD noise comprised of thermal noise and readout noise. From these data the scaling of average dark current and CCD noise with exposure time is determined from a linear fit. This dark current information is used twice: first to subtract an average dark current signal from each recorded image, and second to calculate an error value for each pixel. The latter is determined by the square root sum of CCD noise and noise due to initial photon statistics; this error array is kept throughout the entire assembly process and updated according to the rules of error propagation. It is a crucial ingredient to two other improvements that `AMP` introduces: weighted averages and weighted normalizations.

During the assembly process, arrays are frequently normalized with respect to some constant (such as exposure time or ring current) and subsequently averaged such that in the end there is only one data set containing all the information from all initially recorded images. Even though the arrays are normalized, problems may arise from insufficient knowledge of the normalization constants. We have found for instance that our shutter timing (which

ultimately determines the exposure time) is not very accurate at short exposure times. This will lead to scaling errors between different regions of the final assembled array. To overcome this problem, `AMP` calculates a normalization correction based on pixels that are common to the two arrays about to be averaged. This correction is applied just after the "regular" normalization (using beam current or exposure time), before the arrays are averaged. For both the calculation of the normalization correction and the averaging of two arrays, `AMP` makes use of their error arrays by weighting each pixel's influence on the result with its respective error. This makes good sense as we want pixels that we are more confident in, (with smaller error) to contribute more to the final result than pixels with higher uncertainties. Given two previously normalized arrays 1 and 2 with intensity values at the k-th pixel of $I_{k,1}$ and $I_{k,2}$ respectively, the normalization correction c is calculated from the minimum of the goodness-of-fit parameter

$$\chi^2 = \sum_k \frac{(I_{k,1} - c\, I_{k,2})^2}{\sigma_k^2}, \tag{3.1}$$

where σ_k is the effective total error for the k-th pixel. To calculate σ_k we express the intensity at the k-th pixel of the i-th array $I_{k,i}$ as the sum of true signal $I_{k,i}^{(\mathrm{true})}$ and error $\sigma_{k,i}$. With this, Eq. 3.1 becomes

$$\sum_k \left[I_{k,1}^{(\mathrm{true})} \pm \sigma_{k,1} - c \left(I_{k,2}^{(\mathrm{true})} \pm \sigma_{k,2} \right) \right]^2 = 0. \tag{3.2}$$

Rearranging the left hand side to

$$\sum_k \left[\left(I_{k,1}^{(\mathrm{true})} - c\, I_{k,2}^{(\mathrm{true}))} \right) \pm \sigma_{k,1} \pm c\, \sigma_{k,2} \right]^2 \tag{3.3}$$

and assuming $c \approx 1$, we arrive at

$$\sum_k (\sigma_{k,1} + \sigma_{k,2})^2 \approx \sum_k \left(\sigma_{k,1}^2 + \sigma_{k,2}^2 \right) \tag{3.4}$$

$$\Rightarrow \sigma_k = \sqrt{\left(\sigma_{k,1}^2 + \sigma_{k,2}^2 \right)}, \tag{3.5}$$

where we have assumed that the errors are uncorrelated, or $\sum \sigma_{k,1}\, \sigma_{k,2} = 0$. Now we can go back to the original idea and calculate the normalization

constant by taking the derivative of Eq. 3.1 with respect to c:

$$\frac{\mathrm{d}}{\mathrm{d}c}\sum_k \frac{|I_{k,1} - c\, I_{k,2}|^2}{\sigma_k^2} = 0. \tag{3.6}$$

Performing the derivation and solving for c, we end up with

$$c = \frac{\sum_k \frac{I_{k,1} I_{k,2}}{\sigma_k^2}}{\sum_k \frac{I_{k,2}^2}{\sigma_k^2}}. \tag{3.7}$$

Note that the sum above is performed over pixels that are defined; that is, greater than some threshold or zero and not saturated) in both arrays. After normalizing the arrays in the pre-described manner, we can average them. As with the normalization, we have to make sure that we give more weight to pixels with little uncertainty than to pixels with high uncertainty. Therefore `AMP` calculates for each pixel k the weighted average I_k^{avg} over all arrays $i = 0, ..., N$ as

$$I_k^{\mathrm{avg}} = \frac{\sum_i \frac{1}{\bar{\sigma}_{i,k}^2} I_{i,k}}{\sum_i \frac{1}{\bar{\sigma}_{i,k}^2}}, \tag{3.8}$$

where $\sigma_{i,k}$ is the previously normalized error of the k-th pixel in the i-th image. The new error σ_k for each pixel can then be calculated as the square root sum of all errors of pixels that were averaged, or

$$\sigma_k^2 = \frac{1}{\sum_i 1/\sigma_{i,k}^2} \tag{3.9}$$

which can be used in subsequent analysis.

Apart from providing a more rigorous defined and consistent assembly of the data, `AMP` was also written to facilitate and speed up the process of assembling a 2D diffraction pattern. A simple script file indicating the names of the raw data files to be assembled is sufficient to start `AMP`. Given such a basic script file, `AMP` will attempt to infer all information it needs directly from the data; for all else, it prompts the user for input. As the assembly progresses, `AMP` will write important data-set-specific parameters it determined to the script file for future reference. It also automatically saves information that can be re-used for a subsequent assembly of the same data and even for other data sets if applicable, such as if the same dark current parameters can be used for data sets recorded with the same CCD, or the same beamstop mask pattern for data sets that were recorded using the same beamstop. Finally, `AMP` will

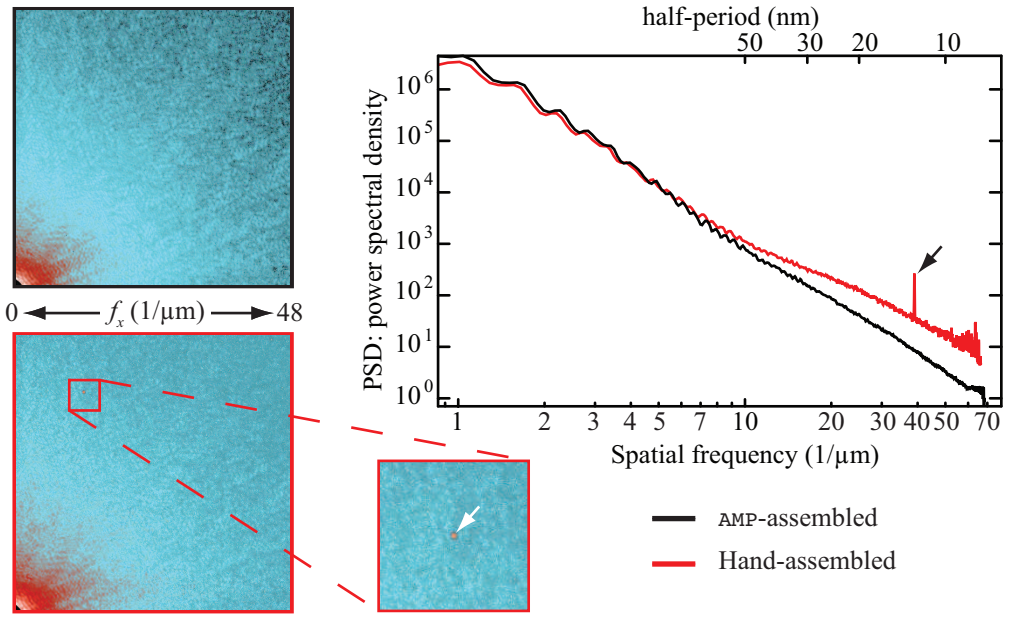

Figure 3.3: Assembled diffraction intensities of data collected on a freeze-dried labeled yeast cell. On the left, the same subsection of the entire array is shown on a false color scale for both AMP-assembled data (in black) and data assembled by hand (in red). On the right, the power spectral densities for both arrays are plotted on a log-log scale. The hand-assembled data follows a different power law and shows a peak around 40 μm^{-1} due to a cosmic ray recorded on the CCD that has not been removed during the assembly. The location of the cosmic ray is indicated by a white arrow in the magnified subset of the hand-assembled array.

save the final assembled diffraction intensities along with meta data important for reconstruction into our own custom file format that I have explained in Sec. 3.1. Automating these steps is especially important for data intensive three-dimensional x-ray diffraction microscopy [24, 71], where 2D diffraction patterns are recorded over a wide angular range with small angular steps prior to mapping the resulting Ewald spheres into a 3D data cube. I will refer to the 2D diffraction data assembled by the above automated procedure as "AMP-assembled" data in what follows.

Figure 3.3 illustrates some advantages AMP-assembled data has over hand-assembled data. Subsections of the assembled diffraction intensities for both AMP-assembled data (black) and hand-assembled data (red) on a logarithmic intensity scale are shown on the left. The x-axis spatial frequency range in each case is from 0 to $\approx 48\,\mu\mathrm{m}^{-1}$. Scaling issues are present in the hand-assembled array but not in the AMP-assembled array. This is illustrated by the plot of the power spectral densities (PSD) for each array, shown on the right on a log-log scale. While the PSD of AMP-assembled data (in black) follows a straight line as would be expected for most objects, the PSD of hand-assembled data (red) changes its slope at a spatial frequency of around $10\,\mu\mathrm{m}^{-1}$ suggesting that low and high spatial frequency data have not been properly scaled. Another prominent difference is the occurrence of a sharp peak at $\approx 40\,\mu\mathrm{m}^{-1}$ marked by a black arrow in the PSD of the hand-assembled data. This peak, presumably due to a cosmic ray incident on our CCD at the time of data collection, is found in one single exposure of the recorded raw data. Due to the large standard deviation of the affected pixels it is filtered out by weighted averaging early on in the AMP assembly process. This is not true for the hand-assembled data where the peak ends up in the final assembled array, as is indicated by the white arrow in the magnified inset of the hand-assembled data. I would like to point out that while a more careful assembly by hand is possible, it would be considerably more time consuming and its steps would have to be readjusted for each new data set.

3.3.3 Comparison of Reconstructed Images

The ultimate judgement of the quality of data assembly comes from seeing the quality of the reconstructed image. In this section I will compare images reconstructed from AMP-assembled versus hand-assembled diffraction data.

While Section 3.4.4 below involves a comparison between different iterate averaging procedures, for this result I used a variation of an already-demonstrated averaging procedure [35]. Reconstructions to determine the object's support were carried out first, starting with an initial support guess calculated from the autocorrelation of the diffraction pattern, then by application

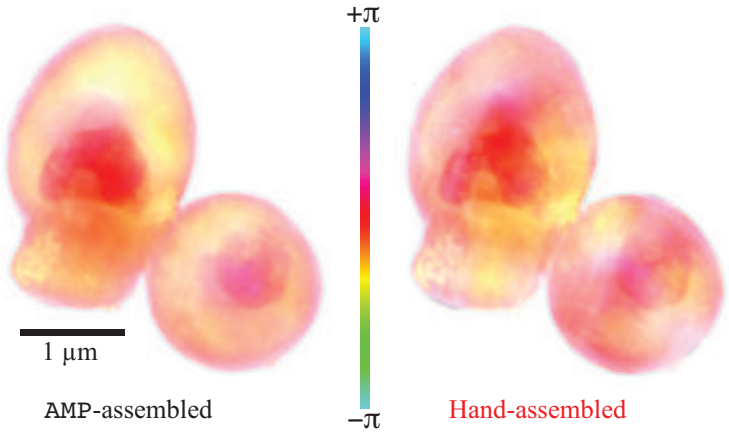

Figure 3.4: Final averaged reconstructions of both AMP-assembled diffraction intensities and hand-assembled diffraction intensities obtained in experiments on gold-labeled freeze-dried yeast [61]. Magnitude is represented as brightness and phase as hue according to the inset color bar. The hand-assembled reconstruction shows streaks and other variations in intensity that would not be expected in the cell; the AMP-assembled reconstruction provides an improved visual appearance.

of the `shrinkwrap` algorithm [40] with occasional by-hand adjustment as is described in Sec. 2.2.5. The final support guess was then used in 10 separate reconstructions with different random starting phases [72]. In each reconstruction, the difference map algorithm [17] was used with a positivity constraint on the imaginary part of the object's exit wave (this corresponds to a maximum phase shift of π or 1.5 μm of solid dry protein for X-rays of wavelength 1.65 nm), and a linear phase ramp was continuously removed (thus constraining the object to be centered in the real space array). In each of these separate reconstructions, every 50^{th} iterate from iterations 5,000 to 10,000 was set aside; the global or zero-spatial-frequency phase of each real-space iterate was adjusted to a common value as described in Sec. 2.2.6, and the complex iterates were then averaged together. Finally, the 10 separate reconstructions were averaged together, again with the global phase adjusted to a common value (the global phase has no effect on diffraction intensities and thus is unconstrained by measured data). This procedure yields a reconstruction with minimal sensitivity to those phases that are poorly constrained by the data.

The images reconstructed using the above procedure on both hand-assembled and AMP-assembled data are shown in Fig. 3.4. Magnitude is displayed as brightness, and phase as hue with a color bar illustrating the phase–hue

relationship. (Note that because the global or zero-spatial-frequency phase is unknown, the color bar serves only as an indicator for relative phase differences). Both reconstructions agree in key features; however, the AMP-assembled reconstruction shows less phase variation at low spatial frequencies. This is in better agreement with what would be expected from less-dense areas of the yeast cell which should have greater uniformity of projected thickness. It is also similar to the goal of maximum-entropy methods of image reconstruction, which seek to find the image with the least variation yet which is still consistent with the measured constraints. If we assume that the hand-assembled Fourier magnitudes have variations associated with erroneous assembly rather than with scattering properties of the specimen, then we would expect the reconstruction from hand-assembled magnitudes to give rise to more, but erroneous, contrast in the reconstructed image. This is consistent with the visual impression of Fig. 3.4.

3.3.4 Comparison of Reconstruction Consistency Using PRTF and PSD

Aside from the visual inspection discussed in the previous section, there are more quantitative means to assessing and comparing the "quality" of reconstructions. I have already introduced two common measures in Sec. 2.4.2, the Phase Retrieval Transfer Function (PRTF) and the Power Spectral Density (PSD). As discussed in this section, the PRTF is a measure of phase retrieval consistency while the PSD gives and indication of the maximum information transfer from measured data to reconstructed data. Figure 3.5 shows a plot of both the PRTF (on left) and the PSD (on right) for both reconstructed data sets, the AMP-assembly in black and the hand-assembly in red. The plot of PRTF curves on the left suggests that the AMP-assembled data set leads to reconstructions with more consistently retrieved phases than the hand-assembled data. Its PRTF curve stays above 0.5 up to a spatial frequency of slightly more than 40 μm^{-1} corresponding to a real space half-period of 13 nm; this is consistent with the resolution estimated from examining real-space features [61]. The hand-assembled data, however, has lower PRTF values for all spatial frequencies except for the very high spatial frequencies. It also behaves differently than the PRTF of the AMP-assembled data in that its values start to increase again after a spatial frequency of around 25 μm^{-1}. It seems counter-intuitive to have a better quality reconstruction at high spatial frequencies because one would assume that phase retrieval would get progressively harder for high spatial frequencies where the signal-to-noise ratio of the recorded data is lower. Instead it suggests that there are either systematic

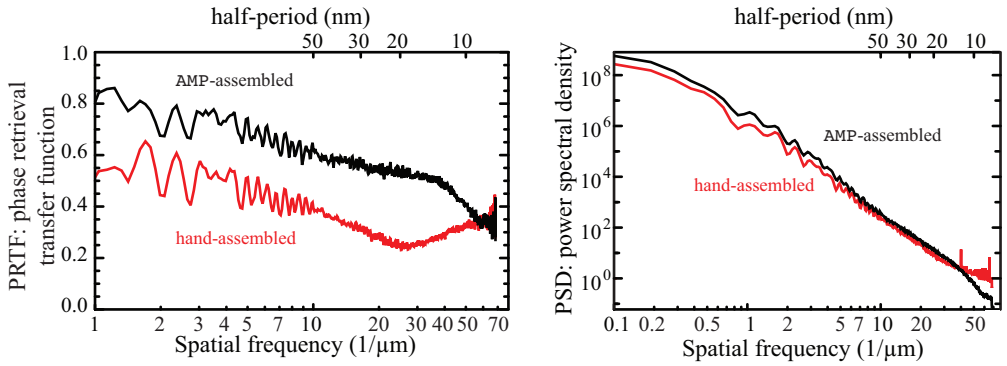

Figure 3.5: PRFT (left) and PSD (right) curves of both AMP-assembled (black) and hand-assembled reconstructions (red). Both plots suggest more consistent phase retrieval and higher maximum information transfer for the AMP-assembled reconstruction. The PRTF of the hand-assembled data displays an artifact that makes its interpretation ambiguous.

errors in diffraction intensity measurement or noise that lead to erroneous but consistent structure. This is the same conclusion that I already drew in Sec. 3.3.3.

The same view is also supported by looking at the plot of the PSD curves shown on right of Fig. 3.5. The reconstruction of the AMP-assembled data follows an inverse power law (as one would expect from Porod's law, see Sec. 2.2.4) out to about 50 μm^{-1} and then rolls off into a flat noise-like power spectrum. The reconstruction of the hand-assembled data, on the other hand, already rolls off at 25 μm^{-1} suggesting that its maximum information transfer is only out to this spatial frequency and thus less than a factor of two of what the AMP-assembled reconstruction achieves.

3.4 Improving the Utility of PRTF Curves for Judging Reconstruction Quality

As I have already explained in Sec. 2.4.2, measurement of the spatial frequency at which the PRTF has suffered a significant decline can be used to provide an estimate of the spatial resolution of the reconstructed image [21, 24, 35], since of course a PRTF value of 1 indicates perfectly reproducible phases while a value of 0 indicates completely random phases. Unfortunately no consensus has emerged on what PRTF value should be used to judge reconstructed

image resolution, with various authors using values of about 0.4 [21], 0.1 [73], "close to zero" [62], or unspecified values [27] as their criteria. In addition to that, the PRTF can exhibit strange behavior that make it hard to interpret. For example, Fig. 3.5 of the last section shows that the PRTF of the hand-assembled data increases again after a certain spatial resolution making the choice of a cutoff-value and its interpretation ambiguous. Finally, the PRTF measures phase retrieval consistency by averaging many independently obtained reconstructions, yet so far there has been no attempt to quantify how different averaging protocols affect the validity of the PRTF.

In this section, I will illustrate how the PRTF works with a simple example, propose a modified version of the PRTF that improves its utility for judging reconstruction quality, and experimentally derive some criteria of what constitutes sufficient averaging for the PRTF [70].

3.4.1 How the PRTF Works: A Simple Example

I mentioned above that there is no consensus in the community as to what value of the PRTF should be used as a cutoff to define the highest spatial frequency with (still) trustworthy phase retrieval. It is therefore instructive to relate the range of PRTF values to a measure of phase variation in the averaged pixels.

Let's assume that for a specific pixel, the ensemble of phases obtained in many reconstructions follows a Gaussian distribution about a fixed valus with a standard deviation of σ_θ. Figure 3.6 shows the relationship between the standard deviation of such an ensemble and the resulting magnitude obtained by averaging all pixels of the ensemble (the pixels were assumed to be of unit-magnitude without loss of generality). This plot can be used to compare a possible cutoff-criteria of the PRTF with the corresponding range of phase variations between the averaged iterates. For example, a value of the PRTF of 0.1 translates to a phase variation with a standard deviation of 123°. It seems unreasonable to use such a low value as cutoff for consistent phase retrieval. In fact, a classic quantity used as threshold for aberration free imaging in optics fabrication, the Rayleigh quarter-wave criterion, says that an imaging system can be considered free of aberrations if the maximum path length difference does not exceed $\lambda/4$ [74]. If we compare the results of Fig. 3.6 to this criterion then we will conclude that a value of 0.7 is a more reasonable cutoff value for the PRTF of Gaussian-distributed phase ensembles.

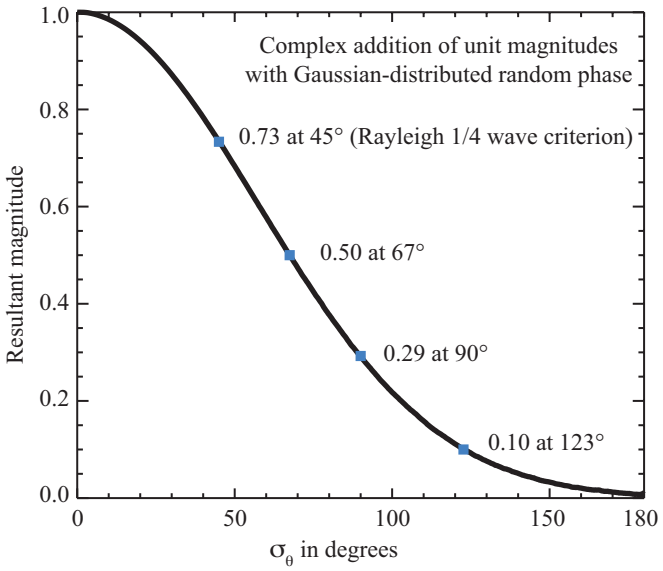

Figure 3.6: The magnitude resulting from a sum of random phases with Gaussian phase distribution characterized by σ_θ, calculated for a range of values of σ_θ. Since the phase retrieval transfer function or PRTF measures the magnitude (at a particular spatial frequency) of the average of many iterates, this figure provides insight into the range of phase variations between the iterates.

3.4.2 Wiener-filtered PRTFs

Even with a better sense of how to interpret different values of the PRTF, we still have to make sure to not get fooled by pathological behavior which can result from a too-small support constraint, or from errors in measurement of the diffraction intensities, or from noise which might place consistent but erroneous structure within the support constraint.

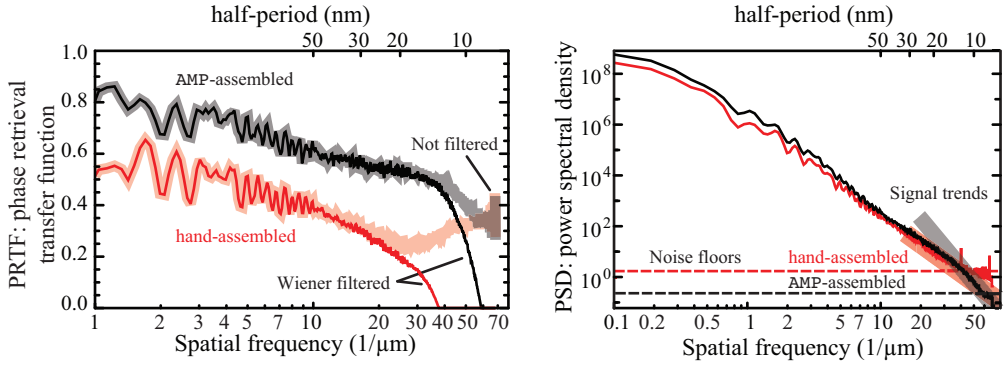

Figure 3.7: The use of a Wiener filter provides an improved measure of reconstructed image quality. This figure shows the phase retrieval transfer function (PRTF; shown in thick shaded lines at left) and power spectral density (PSD; right) curves corresponding to the reconstructions of Fig. 3.4. In the PSD curves, the square of the spatial-frequency-independent noise floor value $|N|^2$ is shown by a dashed line, and a linear fit to the square of the high spatial frequency signal trend $|S(q)|^2$ is shown with a thick shaded line. A Wiener filter function was then calculated according to Eq. 3.10 and applied to the PRTF curve, leading to the Wiener-filtered PRTF curve or wPRTF curve which is shown in thin, non-shaded lines at left. Applying the Wiener filter suppresses the artificially high PRTF-values of the hand-assembled reconstructed data (red) above a spatial frequency of about 25 μm^{-1} and underlines the higher quality of the reconstruction of the AMP-assembled data.

As mentioned in the introduction of this section, this can also make interpreting the PRTF difficult. Let's take another look at the data from the AMP-assembled vs. hand-assembled comparison of Sec. 3.3.4. Figure 3.7 shows again a plot of phase retrieval transfer functions (on left) and power spectral densities (on right) of both data sets, similar to Fig. 3.5. On the left, the original PRTFs of both the AMP-assembled (in black) and the hand-assembled results (in red) are plotted as shaded curves and labeled "Not filtered". As

discussed in Sec. 3.3.4, the shape of the hand-assembled PRTF curve suggests pathological behavior.

To make this a better tool for assessing reconstruction quality overall, we need a way to combine the PRTF's measure of reconstruction consistency with a measure of maximum information transfer. I propose that a Wiener filter [75] provides such a measure. If one can estimate the spatial frequency dependent trend $S(q)$ of the true signal, and the trend $N(q)$ in noise, the Wiener filter $W(q)$ is formed from

$$W(q) = \frac{|S(q)|^2}{|S(q)|^2 + |N(q)|^2} \tag{3.10}$$

so that it varies smoothly between 1 for signal dominated and 0 for noise dominated spatial frequencies. Note that what I refer to here as true signal $S(q)$ is actually a convolution of the true signal with the point-spread-function (PSF) of our imaging system, but since we do not apply the filter to the actual image, it does not make a difference. Since many noise sources (such as photon statistical noise, and thermal charge fluctuations in CCD detectors) are uncorrelated pixel-to-pixel, the power spectral density PSD of noise usually follows the form of the Fourier transform of a delta function: namely, a "flat" power spectrum consisting of a constant value at all spatial frequencies. The diffraction signal is much different; as was noted in Sec. 2.2.4, it tends to decline as $I \propto q^{-m}$ with $m \simeq 3$–4. We can therefore follow a straightforward procedure suggested by Press et al. [76] to generate a Wiener filter from the power spectral density of a measured diffraction pattern: If we assume the measured signal $C(q)$ to consist of true signal $S(q)$ plus frequency independent noise N, such that its power spectral density is given by

$$\mathrm{PSD}(q) = |C(q)|^2 = |S(q)|^2 + |N|^2 \tag{3.11}$$

then the trend of $|C(q)|^2$ can be found from a straight line fit in a log-log plot, while the square of the spatial frequency independent noise floor $|N|$ can be found from where the power spectral density rolls off to a constant at high spatial frequencies. From these two quantities, we can extrapolate the square of the true signal as $|S(q)|^2 = |C(q)|^2 - |N|^2$ and use this to determine the Wiener filter according to Eq. 3.10. This procedure is illustrated at right in Fig. 3.7; it has been used with success for image deconvolution [77] and phase contrast Fourier filtering [78] in lens-based x-ray microscopy.

Application of the Wiener filter to the phase retrieval transfer function (abbreviated here as wPRTF) provides an improved measure of the reconstructed image. In Fig. 3.7, the wPRTF for the `AMP`-assembled and hand-assembled reconstructed images of Fig. 3.4 is shown. While the non-filtered PRTF of the

hand-assembled data has a pathological *increase* at higher spatial frequencies, the wPRTF shows a sharp decrease. In addition, the wPRTF of the hand-assembled data is now below that of the AMP-assembled data for all spatial frequencies, which is consistent with the improved visual impression of the reconstructed images in Fig. 3.4.

3.4.3 wPRTF and Varied Specimen Exposures

To make sure that Wiener-filtered PRTFs are a reliable tool to assess the quality of reconstructions of a wide variety of diffraction data, I show here reconstructed diffraction data from a simulated object at various different incident photons per pixel values. The simulated object was designed to approximate a pair of frozen-hydrated biological cells in a 512^3 array with 15 nm pixel size, similar to simulated cells we have used in other computational studies [65]. The larger cell has an outer diameter of 2.1 μm while the smaller cell has an outer diameter of 1.27 μm. Together they are embedded in a 30 nm thin layer of ice. Both cells have a 45 nm thick double-layer cell membrane made from 50/50 protein and lipid, and are filled with a 10:1 ice and protein mixture. Several lipid balls of 60 nm diameter are distributed throughout the volume of both cells. Each cell also has a cell nucleus (assumed to be filled with chromatin) with a 15 nm thin single layer cell membrane made from the same composition as the outer cell membrane. Finally, each cell has a vacuole that is filled with ice and has a 15 nm thin lipid membrane. The values of the refractive index are calculated according to tabulated data of Henke *et al.* [79] assuming a stoichiometric composition of $H_{48.6}C_{32.9}N_{8.9}O_{8.9}S_{0.3}$ and density of $\rho = 1.35$ g/cm^3 for protein, $H_{62.5}C_{31.5}O_{6.3}$ with $\rho = 1.0$ g/cm^3 for lipid [56], and $H_{49.95}C_{24.64}N_{8.66}O_{15.57}P_{1.07}S_{0.03}$ and $\rho = 1.527$ g/cm^3 for chromatin [80]. Assuming an x-ray energy within the "water window" [60, 81] of 520 eV, an exit wave leaving the object was calculated using a multislice propagation process [35, 82] and then propagated to the far-field. Diffraction patterns were simulated for 11 different exposures with photons per pixel values ranging from 10^1 to 10^6, with simulated Poisson noise included [65]. The power spectral density of the highest exposure diffraction pattern shown in the grey curve at left in Fig. 3.8 indicates that the simulated cell showed strong scattering out to a spatial frequency of about 13 μm^{-1}. Each data set was reconstructed similar to what has been described in Sec. 3.3.3 for the experimental data, except that averaging was applied to every 10th iterate starting at 2,000 iterations up until a total of 10,000 iterations had been run.

Results from these simulations are shown at right in Fig. 3.8. As expected [65], lower exposures lead to poorer signal-to-noise values in the final reconstructions and thus poorer resolution. As a measure of the effective resolution

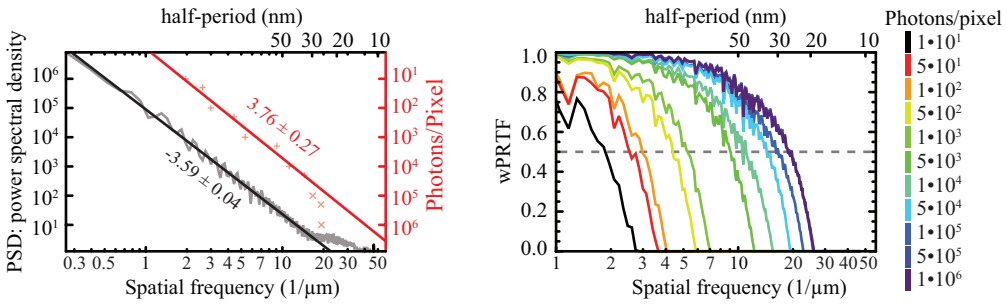

Figure 3.8: The Wiener-filtered phase retrieval transfer function (wPRTF) provides a good measure of reconstructed image quality over a wide range of photon exposures. Shown at right are a series of wPRTF curves for reconstructions of simulated data with several different photons per pixel values and simulated Poisson noise. The spatial frequency at which the wPRTF crosses the dashed 0.5 line is taken as effective resolution for each data set. On left, a power law fit to the power spectral density (PSD) of the data set with the highest photons per pixel value (black) is compared to a power law fit to the dose–resolution data (red) derived from the figure on the right. The magnitudes of both slopes agree within their error, indicating that the degree to which scattering decreases with spatial frequency in an object is equal to the degree at which reconstructed image resolution falls off with decreasing exposure [65].

of the reconstruction at each photon exposure value, I took the spatial frequency at which the filtered PRTF curve falls below 0.5. These values are plotted as red + marks at left in Fig. 3.8; plotting the values against the simulated photons per pixel values results in a power-law fit with an exponent of 3.76 ± 0.27 (after excluding resolution measures above the 13 μm^{-1} spatial frequency at which there was little signal present in the simulated object; these are shown as red $\times$ marks). Also shown in Fig. 3.8 is a fit to the power spectral density of the highest exposure diffraction pattern; this gave a slope of -3.59 ± 0.04. The magnitude of both exponents agree within error; this is as expected, since one needs signal at a spatial frequency to see structure over the corresponding length scale, so that achievable resolution should follow the same spatial frequency trend with exposure as the spatial frequency content of the object does [65]. The fact that the Wiener-filtered PRTF provides such a straightforward illustration of this result of this confirms the utility of the wPRTF measure.

3.4.4 wPRTF and Iterate Averaging

Iterate averaging provides a way to improve image reproducibility, and to measure the resolution via the Wiener-filtered phase retrieval transfer function (wPRTF). In this section I will consider how many iterates should be averaged, and at what frequency the iterates should be sampled. Sufficient averaging supresses the contribution of poorly-phased Fourier scattering to the real-space image, since phases which are not reliably retrieved will add up incoherently. A lack of sufficient averaging will lead to artificially high PRTF values, since not enough potentially inconsistent phases will have been sampled. In fact, in the extreme case of no averaging the PRTF will be unity for all spatial frequencies as follows from Eqs. 2.44 and 2.29.

How frequently should one sample particular iterates for averaging? To address this question, both the `AMP`-assembled experimental data, and the simulated data set described above, were reconstructed for 5,000 iterations of the difference map algorithm. Then the algorithm was run further, and a total of 100 iterates taken every i^{th} iteration ($i \in \{1, 5, 10, 20, 30, 40, 60, 70, 80, 90, 100, 200\}$) was averaged to obtain the final result. The resulting wPRTF curves shown in Fig. 3.9 are nearly identical for all different averaging intervals. This suggests that the frequency at which iterates are sampled is unimportant.

How many iterates should be averaged? One would expect that the result would depend on the quality of the data, since data with systematic errors should show more fluctuations in the reconstructed phase. Rather than plot a series of individual wPRTF curves as in Fig. 3.9, in this case the RMS residual

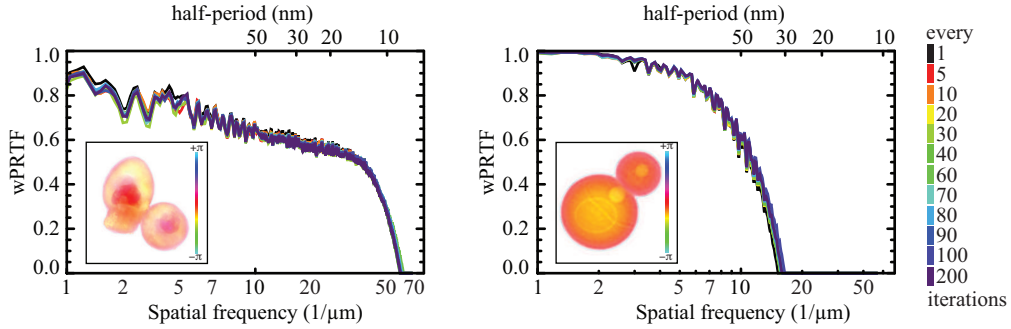

Figure 3.9: Illustration of the effect of choosing different iterate averaging frequencies. For both the AMP-assembled experimental data at left, and the simulated data set at right, the difference map algorithm was first run for 5,000 iterations. Next, 100 iterates were taken every i^{th} iteration ($i \in \{1, 5, 10, 20, 30, 40, 60, 70, 80, 90, 100, 200\}$) to obtain a final result. Wiener-filtered phase retrieval transfer function (wPRTF) curves are plotted for each of the iterate averaging frequencies. An example reconstructed image is shown as an inset. As judged by the wPRTF, all iterate sampling frequencies give essentially the same result.

change in the wPRTF as one went from i to $i+1$ averages was measured as

$$\text{RMS}_{i+1} = \sqrt{\frac{\sum \left(\text{wPRTF}_i - \text{wPRTF}_{i+1}\right)^2}{N}}, \tag{3.12}$$

where the sum extends over all N spatial frequencies up to the spatial frequency where the PSD rolls off to a steady noise floor for a given reconstruction. I calculated the RMS residual according to the above equation for the set of reconstructions with 12 different averaging frequencies that were already used for the analysis leading to Fig. 3.9. Since this analysis showed that the consistency in phase retrieval as measured by the wPRTF does not depend on the averaging frequency, I assume only statistical differences between these reconstructions and calculate the mean of all RMS residuals and their standard deviation as a function of number of iterates averaged. Based on examination of the resulting average RMS residual on a linear-log plot, the average RMS residual is fitted to a function of the form $y(x) = a\,x^b + c$ in order to characterize the residual trend.

Figure 3.10 shows graphs of this analysis for reconstructions of A) the AMP-assembled data set, B) the hand-assembled data set, and C) the simulated

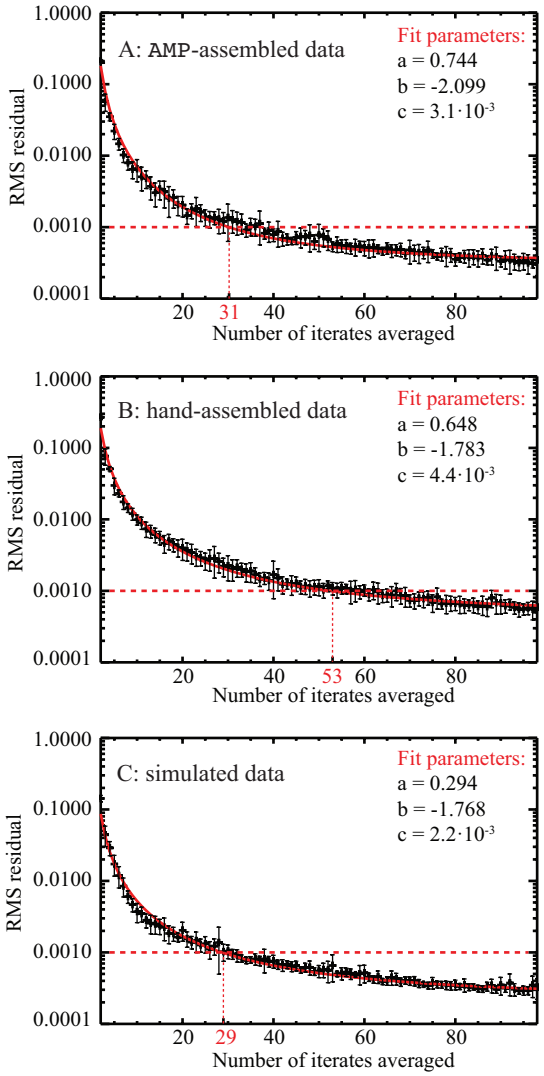

Figure 3.10: Changes in the wPRTF as a function of number of iterates averaged for A) AMP-assembled experimental data, B) hand-assembled experimental data, and C) simulated data. The RMS residual changes in the wPRTF (calculated using Eq. 3.12) as one goes from i to $i+1$ iterates averaged were then fitted to a function of the form $y(x) = a\,x^b + c$, plotted in red. The error bars indicate the standard deviation between 12 different averaging frequencies. A horizontal dashed line marks a value of RMS residual of 0.001 selected to compare the results for the three different data sets.

data set. The calculated means of all 12 reconstructions with different iterate averaging frequencies are plotted as crosses with error bars indicating their standard deviation. The fitted function is plotted in red and its fit parameters are indicated for each respective graph. An arbitrary threshold of 0.001 RMS residual was chosen to define convergence of the wPRTF; it is marked in the graph by a horizontal dashed line. The number of iterates at which the fitted function falls below the threshold (*i.e.*, the number of averages at which I declare the wPRTF to have converged) is indicated for each data set by a vertical dotted line. The reconstructions of both the AMP-assembled and the simulated data set converge after ≈ 30 averaged iterates, while the reconstruction of the hand-assembled data set converges only after about 50 averaged iterates. This result gives an estimate as to how many iterates need to be averaged for the PRTF to be a valid representation of the consistency in phase retrieval of a reconstruction. It also shows that AMP-assembly leads to data sets that have fewer systematic errors in the Fourier plane intensities.

Chapter 4

Scanning X-ray Diffraction Experiments

Full-field x-ray diffraction microscopy is limited by the requirement that the object be isolated within the coherent footprint of the illuminating beam. This puts constraints on what kind of object can be used in the experiment and how it has to be prepared. I commented in Sec. 2.4.1 that for biological specimen it is best to work with frozen-hydrated samples in order to minimize radiation damage. This poses an additional challenge since we have to make sure that scattering due to the always present ice layer outside the cell is negligible. Scattering from the ice background surrounding a specimen will violate the isolated object constraint and it will also amplify the effects of poor coherence in our illumination; if the coherence width is less than the illumination width then the recorded diffraction pattern will be a superposition of several mutually incoherent diffraction patterns due to different coherent patches in the illumination. It has been shown that partially coherent data does not properly reconstruct [83].

While others in our group have focused on improving specimen preparation conditions to minimize the effects of the ice layer, here I want to talk about an alternative experimental geometry (Scanning X-ray Diffraction Microscopy) that can overcome the above limitations. I have introduced the concept of SXDM along with reconstruction schemes and sampling requirements in Sec. 2.3. I also commented on the necessity of a stable experimental setup. In this Chapter I discuss my efforts to integrate a SXDM setup into our existing microscope at beamline 9.0.1 of the Advanced Light Source, Lawrence Berkeley Lab (Sec. 4.1). I will talk about the main design goals, possible optical geometries, improvements to the apparatus and show some preliminary results of recent experimental data. I also want to discuss the results of a SXDM experiment using a Pixel-Array-Detector (PAD) at beamline 2-ID-B of

the Advanced Photon Source, Argonne National Lab in Sec. 4.2.

4.1 Implementing SXDM at Beamline 9.0.1, ALS

4.1.1 Main Design Considerations

The main goal for the SXDM setup was to be able to record a properly sampled diffraction pattern of a frozen eukaryotic cell (using X rays with energies in the water window) that has no missing low spatial frequency data and extends out to at least 50 μm^{-1} (corresponding to a half-period resolution of 10 nm), and all this in a reasonable amount of time and with a minimum amount of dose to the sample. In this section I discuss how these main goals are related to each other. Some goals are complementary, some are competing. I will describe optimal choices to satisfy as many goals as possible. Note that I will ignore for now how limitations of the experimental hardware affect these goals; it will come up later again.

Beam Footprint on Sample

This is one of the most important parameters, and it is linked with all the other goals. By beam footprint I mean both the size and the shape of the beam. The shape of the beam footprint should be well defined. Ideally, there is a central bright area that is surrounded by lower intensity areas. However, this requirement is not as stringent as in the case of Keyhole Coherent Diffractive Imaging (Sec. 2.3.5) where the probe needs a well defined edge in the sense that the slope in intensity change going across the edge is very steep. Here, we can tolerate some "fuzziness" of the probe as long as our oversampling requirements are met (see below).

The size of the beam footprint primarily depends on the size of the sample. Our model biological specimen is a yeast cell of type *Saccharomyces cerevisiae* with *whi*-5 mutation. These types of cells are typically around 3 μm in diameter. To avoid overdosing, the size of the beam footprint should be significantly smaller than that, ideally less than 1 μm.

Oversampling and Resolution

The oversampling requirements in SXDM are the same as the ones I discussed for full-field XDM (Sec. 2.2.2) except that here the size of the autocorrelation of the beam footprint on the object (rather than of the object itself) is the

critical parameter. Let's assume that our beam footprint has diameter D; the oversampling can then be calculated according to Eq. 2.21 (repeated here for convenience) as

$$s = \frac{1}{\Delta q \cdot D} = \frac{\lambda Z_0}{pD},$$

where Z_0 is the distance between the object and the detector and p is the pixel size of the detector. We can see that to increase oversampling for a given detector and beam footprint, the detector has to be moved further back.

Increasing Z_0 will decrease the solid angle subtended by the detector, which decreases the theoretical maximum momentum transfer $q_{\max}$ that can be measured. The latter is derived in Eq. 2.16 (repeated here for convenience) as

$$q_{\max} = \frac{1}{d_{\min}} = \frac{p \cdot \frac{N}{2}}{Z_0 \cdot \lambda},$$

where N is the number of pixels of the detector. Clearly we need to find a good trade-off position between the two quantities.

Flux Considerations

Two things are of interest to us in this section. First, we want to calculate how long we need to expose the specimen to record scattered signal out to $q_{\max} = 50\,\mu\text{m}^{-1}$ with a signal-to-noise ratio of at least five [59]. Second, we want to know if we can record down to lowest spatial frequencies without saturating our detector. For both we need to look into how the total incident flux is distributed across the detector.

To answer the first question, we also need to estimate how scattered photons relate to signal-to-noise ratio. Let's first try to quantify the flux incident on the detector. At our beamline, we measure the photon flux n_{Ph} after the monochromator. We then have to take into account the efficiency of the beam-defining aperture η and the transmittance of the sample T_{sample}. With this we can calculate the rate of photons transmitted through the sample (*i. e.* the scattered signal for all spatial frequencies) $n_{\text{Ph}}^{(\text{trans})}$ as

$$n_{\text{Ph}}^{(\text{trans})} = n_{\text{Ph}} \cdot \eta \cdot T_{\text{sample}}. \tag{4.1}$$

Since we are actually measuring the photon flux with a photodiode that produces a current on the order of nano-Amperes, it makes more sense to rewrite

$$n_{\text{Ph}} = 2.28 \cdot 10^{10}\, \frac{I_{\text{PD}}(\text{nA})}{E_{\text{Ph}}(\text{eV})}, \tag{4.2}$$

where $2.28 \cdot 10^{10}$ is a conversion factor relating the photons incident on the photodiode with its output current in convenient units. With this, the flux behind the sample becomes

$$n_{\mathrm{Ph}}^{(\mathrm{trans})} = 2.28 \cdot 10^{10} \, \frac{I_{\mathrm{PD}}(\mathrm{nA}) \cdot \eta \cdot T_{\mathrm{sample}}}{E_{\mathrm{Ph}}(\mathrm{eV})}. \tag{4.3}$$

Now onto the question of the relationship between signal-to-noise and number of scattered photons. The noise in a pixel i on the CCD is given by the readout noise and the noise due to photon statistics, or

$$\mathrm{Noise}_i = \sqrt{\sigma_{\mathrm{CCD},i}^2 + N_{\mathrm{CCD},i}}, \tag{4.4}$$

where $N_{\mathrm{CCD},i}$ is the number of CCD counts produced by photons incident on this pixel and $\sigma_{\mathrm{CCD},i}^2$ is the read-out noise in pixel i of the detector which can be determined from the variance of a series of dark current images (as described in Sec. 3.3.2, this value is calculated by `AMP`). The signal-to-noise ratio is then defined as

$$\mathrm{SNR}_i = \frac{N_{\mathrm{CCD},i}}{\sqrt{\sigma_{\mathrm{CCD},i}^2 + N_{\mathrm{CCD},i}}}. \tag{4.5}$$

Solving for $N_{\mathrm{CCD},i}$ for $\mathrm{SNR}_i > 5$ yields

$$N_{\mathrm{CCD},i} > \frac{1}{2}\left(25 + \sqrt{625 + 100\,\sigma_{\mathrm{CCD},i}^2}\right), \tag{4.6}$$

where I dropped the negative solution. A typical value for the read-out noise of the 1.3k chip is

$$\sigma_{\mathrm{CCD},i} = 1.5 + 0.06 \cdot t_{\mathrm{exp}}. \tag{4.7}$$

With this I calculate that we need photons equivalent to at least thirty CCD counts ($N_{\mathrm{CCD},i} > 30$) to obtain a signal-to-noise ratio that satisfies the Rose-criterion. Remembering Eq. 2.26 along with the parameters of our two available detectors summarized in Tab. 2.1, we see that one 520 eV photon will produce around 46 (93) counts for the 1.3k (2k) chip, *i. e.* a single scattered photon should suffice for both CCDs. Note that this assumes that the photon noise linearly translates into "count-noise" on the CCD, which is only approximately true. It also does not take into account that the interference pattern of one photon at 10 nm half-period might be distributed over several pixels in which case we should calculate the SNR as an ensemble average of all pixels that contribute. Nonetheless we can use this simple argument to get some rough idea of what exposure time is required.

To get a sense of how the total transmitted intensity calculated in Eq. 4.3 is distributed across the detector, we can use Porod's law as described in Sec. 2.2.4. Using the fact that $I(q) \propto n_{\mathrm{Ph}}(q)$ and assuming $m = -4$, we can write for the number of photons scattered in a small frequency range $[q, q+\mathrm{d}q]$

$$n_{\mathrm{Ph}}(q)\,\mathrm{d}q = n_0 \cdot q^{-4}\mathrm{d}q, \tag{4.8}$$

We can calculate the constant n_0 from the requirement that the sum of scattered photon rates over all length scales add up to the total rate of photons transmitted through the specimen

$$n_{\mathrm{Ph}}^{(\mathrm{trans})} = \int_{q_{\mathrm{min}}}^{\infty} n_0\, q^{-4}\mathrm{d}q \tag{4.9}$$

$$\Rightarrow n_0 = 3\, n_{\mathrm{Ph}}^{(\mathrm{trans})} q_{\mathrm{min}}^3, \tag{4.10}$$

where q_{min} is the smallest spatial frequency in inverse space which corresponds to the largest length scale in the real space plane, *i. e.* the size of the beam footprint D. In particular, we note that for the lowest spatial frequency q_{min}

$$n_{\mathrm{Ph}}(q_{\mathrm{min}})\,\mathrm{d}q = \frac{3 \cdot n_{\mathrm{Ph}}^{(\mathrm{trans})}}{q_{\mathrm{min}}}\,\mathrm{d}q. \tag{4.11}$$

Remembering the dynamic range of scattered intensities according to Eq. 2.28, repeated here for convenience

$$\mathrm{DR} = \left(\frac{q_{\mathrm{max}}}{q_{\mathrm{min}}}\right)^4 = \left(\frac{d_{\mathrm{max}}}{d_{\mathrm{min}}}\right)^4,$$

we are now in a position to derive an equation that links the desired resolution d_{min} to the required exposure time t_{exp} by arguing that we need at least DR photons at spatial frequency q_{min} in order to get one scattered photon at our resolution length scale d_{min} ($\hat{=} q_{\mathrm{max}}$), *i.e.*

$$t_{\mathrm{exp}} \cdot n_{\mathrm{Ph}}(q_{\mathrm{min}})\,\mathrm{d}q = \mathrm{DR} = \left(\frac{q_{\mathrm{max}}}{q_{\mathrm{min}}}\right)^4. \tag{4.12}$$

The frequency interval $\mathrm{d}q$ in our case is given by the frequency sampling in the detector plane $\Delta q = {}^{2q_{\mathrm{max}}}/_{N}$. Using Eq. 4.11, we can rewrite in terms of

the total rate of photons transmitted through the specimen $n_{\mathrm{Ph}}^{(\mathrm{trans})}$ as

$$t_{\mathrm{exp}} = \frac{N \cdot q_{\mathrm{max}}^3}{6 \cdot n_{\mathrm{Ph}}^{(\mathrm{trans})} q_{\mathrm{min}}^3} = \frac{N \cdot d_{\mathrm{max}}^3}{6 \cdot n_{\mathrm{Ph}}^{(\mathrm{trans})} d_{\mathrm{min}}^3} \tag{4.13}$$

$$\Rightarrow d_{\mathrm{min}} = \frac{\sqrt[3]{N} d_{\mathrm{max}}}{\sqrt[3]{6 \cdot n_{\mathrm{Ph}}^{(\mathrm{trans})} \cdot t_{\mathrm{exp}}}}, \tag{4.14}$$

where I switched from frequencies $(q_{\mathrm{min}}, q_{\mathrm{max}})$ to their real space equivalents $(d_{\mathrm{max}}, d_{\mathrm{min}})$ in the last equation. Note that $n_{\mathrm{Ph}}^{(\mathrm{trans})}$ is given by Eq. 4.3. The last result is particularly useful to estimate what resolution can be obtained in a reasonable amount of time, given a certain photon flux and size of the beam footprint on the sample. I will show concrete examples when discussing the different optical geometries in Sec. 4.1.2. Note that due to the inverse third root relationship, one needs to increase the exposure time by a factor of 10^3 to gain a factor of 10 in resolution.

Finally, I wanted to calculate the longest exposure time that we can record before the detector saturates. The brightest pixels will be the ones measuring scattered photons of spatial frequency q_{min}. The number of pixels that sample this spatial frequency in one dimension is given by the oversampling factor s from Eq. 2.21. If we are assuming square speckles, then the total number of pixels sampling this spatial frequency will be s^2. Each of these s^2 pixels will again sample a frequency interval Δq which can be expressed in terms of the oversampling factor using the same equation; it is $\Delta q = {}^1/_{s \cdot d_{\mathrm{max}}}$. With that and Eq. 4.11, the number of photons per pixel N_{pixel} is given by

$$N_{\mathrm{pixel}} = \frac{3 \cdot n_{\mathrm{Ph}}^{(\mathrm{trans})}}{s^3} \cdot t_{\mathrm{exp}}. \tag{4.15}$$

As discussed in Sec. 2.2.4, both our CCDs have a 16-bit ADC allowing for a maximum count rate of $2^{16} - 1 = 65535$ and I have calculated before that each 520 eV photon produces 46 (93) counts for the 1.3k (2k) chip and each 750 eV photon produces 67 (135) counts. That means that the chip saturates at 1425 (705) incident photons of 520 eV and at 973 (487) incident photons of 750 eV. Calling the maximum number of photons for either case N_{max}, the maximum exposure time before saturation is then given as

$$t_{\mathrm{exp,max}} = \frac{N_{\mathrm{max}} \cdot s^3}{3 \cdot n_{\mathrm{Ph}}^{(\mathrm{trans})}} = \frac{N_{\mathrm{max}} \cdot \lambda^3 Z_0^3}{3 \cdot n_{\mathrm{Ph}}^{(\mathrm{trans})} p^3 d_{\mathrm{max}}^3}, \tag{4.16}$$

where I have used Eq. 2.21 in the last step. In words, the maximum exposure

time before saturation occurs goes as the inverse third power of the size of the beam footprint on the sample. Experimentally, the shortest exposure time that we can record is given by the speed with which we can open and close the shutter that blocks the x-rays from continuously exposing the sample. For our experiment this time is 1 msec. If the parameters of the experiment are such that $t_{\mathrm{exp,max}} < 1$ msec then we will not be able to record the low spatial frequencies without saturation and have to find other experimental solutions, such as the use of an absorbing beam attenuator, see Sec. 4.1.3.

Now that I have derived all important quantities, let's look into how they are related to each other in terms of the parameters to choose, d_{max} and Z_0:

- Oversampling s and resolution d_{min} are weakly linked competing goals. Both vary linearly with the sample detector distance Z_0; thus we cannot achieve larger oversampling and smaller resolution at the same time by adjusting the detector sample distance, see Eqs. 2.21 and 2.16. However, we can increase the oversampling by choosing a smaller footprint of the beam on the sample. This will not affect the resolution.

- Oversampling s and maximum exposure time before saturation $t_{\mathrm{exp,max}}$ are strongly linked complementary goals. Looking at Eqs. 2.21 and 4.16, we see that $t_{\mathrm{exp,max}} \propto s^3$. This is good, because we want large oversampling and long maximum exposure times.

- Oversampling s and required exposure time for a certain resolution t_{exp} are weakly linked complementary goals. From Eqs. 2.21 and 4.13 we can see that a smaller footprint of the beam increases the oversampling and decreases the exposure time for a certain resolution by decreasing the dynamic range of the signal. However, the exposure time is also dependent on the desired resolution which is a competing goal to both oversampling and exposure time for a certain resolution.

- Resolution d_{min} and maximum exposure time before saturation $t_{\mathrm{exp,max}}$ are indirectly linked competing goals. We can see from Eq. 4.16 that $t_{\mathrm{exp,max}} \propto s^3$ and I have commented above that resolution and oversampling are weakly competing goals.

- Resolution d_{min} and required exposure time for a certain resolution t_{exp} are strongly linked competing goals. As can be seen from Eq. 4.14, we have to increase the exposure time by a factor of 10^3 to increase the resolution by a factor of 10.

Since each goal also depends on a variety of other parameters, most notably the incident flux which can vary greatly with different optical geometries, I will

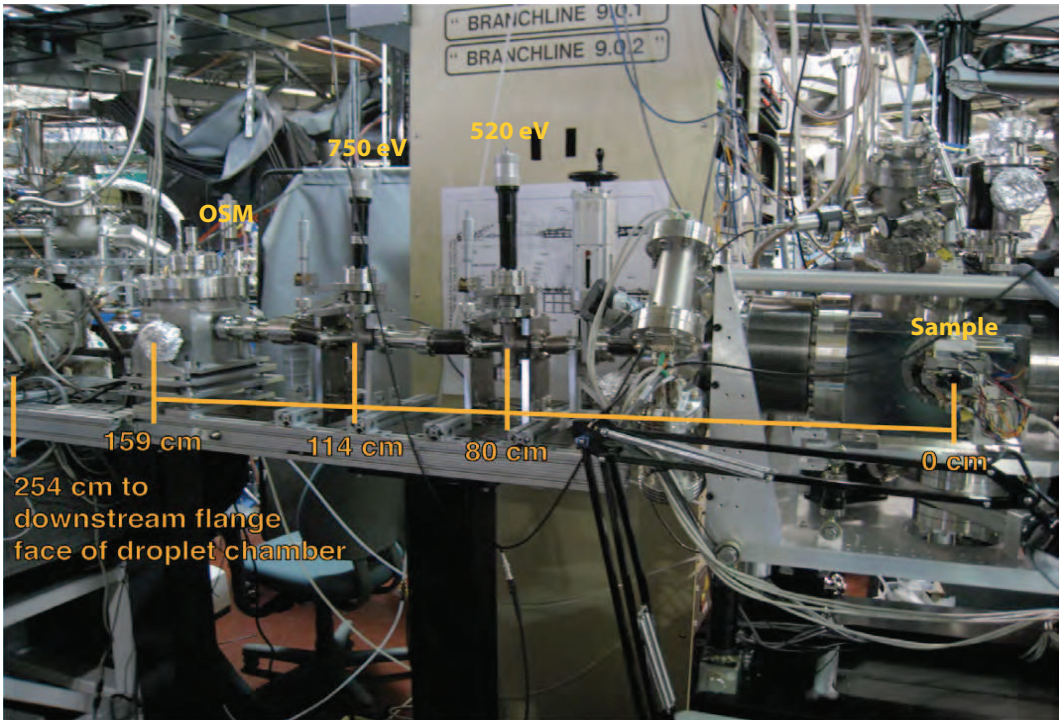

Figure 4.1: Distances between monochromator zone plates and the sample at our existing diffraction microscope at beamline 9.0.1 of the Advanced Light Source, Lawrence Berkeley Lab. Photo courtesy of Chris Jacobsen.

give specific examples of how to choose the best combination of parameters in the next section.

4.1.2 Different Optical Geometries

Now that we know how different experimental goals are related to each other I want to talk about the actual implementation of an SXDM setup for our existing microscope. Besides the goals above, we have additional constraints given to us by the existing apparatus. Most notably, we need to use the current monochromator design scheme and we want to be able to use the Anti-Contamination Device for cryogenic samples. Both features are described in more detail in Sec. 1.3.1.

An overview of the beamline distances is given in Fig. 4.1. The focal length for our zone plate monochromator is given by Eq. 1.16 and I have summarized the focal lengths for the two energies that we would like to be able to use at

Energy (eV)	Focal length (mm)
520	77.5
750	111.8

Table 4.1: Focal lengths of the monochromator zone plate for the two energies commonly used at our microscope.

this beamline in Tab. 4.1. If we compare these values for the focal lengths with the distances of the monochromator zone plates from the sample shown in Fig. 4.1 then we see that in either case the energy selecting pinhole will be about an inch from the sample. In order to get a sufficiently small beam footprint on the sample, we had to change this up a little bit. I will describe the three setup options that were used in experiments

Zone Plate as Beam-defining Aperture

The basic geometry is shown in Fig. 4.2, where the x rays are assumed to travel from left to right. A distance Z_{ZP} downstream of the monochromator pinhole of diameter D_{P} is a zone plate of outer diameter D_{ZP}. The sample sits in the focal plane of the zone plate with focal length Z_{f}. It is illuminated with a demagnified image of the monochromator pinhole which is of size D. The diffracted signal is then recorded onto a CCD detector which is a distance Z_0 further downstream of the sample. Apart from achieving our experimental goals outlined in the previous section, there are additional constraints that we have to fulfill: the zone plate should have a sufficient number of zones ($N > 50$) to avoid phase ringing effects but less zones than the temporal coherence of the beamline ($N < \lambda/\Delta\lambda \approx 500$) to avoid chromatic aberrations. The focal length of the zone plate Z_{f} should be long enough to accommodate an Order Sorting Aperture (OSA) between zone plate and cryo shield of the sample but not too long to exceed the extent of the vacuum enclosure. As discussed above, the size of the focal spot should be on the order of 1 μm. Finally, the distance between the zone plate and the monochromator pinhole Z_{ZP} should be long enough so that the zone plate is fully illuminated.

The main equations describing the relationship between the relevant parameters of zone plate diameter D_{ZP}, outermost zone width δ_{r_N}, wavelength λ, number of zones N, size of the focal spot D and focal length f are given by Eq. 1.16 and

$$\delta_{r_N} = \frac{D_{\mathrm{ZP}}}{4\,N} \tag{4.17}$$

$$D = 1.22 \cdot \delta_{r_N} \tag{4.18}$$

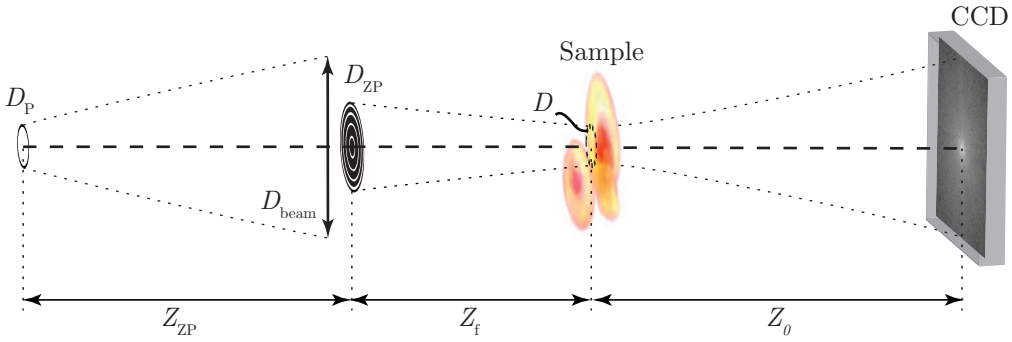

Figure 4.2: Schematic of illuminating the sample with a zone plate of diameter D_{ZP}

		$h\nu = 520$ eV	$h\nu = 750$ eV
$Z_{\mathrm{f}} = 1$ cm	$D = 500$ nm	$N = 36$	$N = 25$
		$D_{\mathrm{ZP}} = 59.0\,\mu$m	$D_{\mathrm{ZP}} = 41.0\,\mu$m
	$D = 1\,\mu$m	$N = 9$	$N = 6$
		$D_{\mathrm{ZP}} = 29.5\,\mu$m	$D_{\mathrm{ZP}} = 19.7\,\mu$m
$Z_{\mathrm{f}} = 5$ cm	$D = 500$ nm	$N = 180$	$N = 125$
		$D_{\mathrm{ZP}} = 295.0\,\mu$m	$D_{\mathrm{ZP}} = 205.0\,\mu$m
	$D = 1\,\mu$m	$N = 44$	$N = 35$
		$D_{\mathrm{ZP}} = 144.3\,\mu$m	$D_{\mathrm{ZP}} = 98.5\,\mu$m

Table 4.2: Example zone plate parameters. Only a zone plate with focal length of $Z_{\mathrm{f}} = 5$ cm and size of the focal spot of $D = 500$ nm satisfies both the minimum and maximum number of zones constraint.

From this I can calculate some possible combinations of zone plate parameters for both energies, which are shown in Tab. 4.2. Looking at these numbers we see that only a zone plate with focal length of $Z_{\mathrm{f}} = 5$ cm and focal spot size of $D = 500$ nm satisfies both number of zone requirements for both wavelengths. We can now use Eqs. 2.21 and 2.16 to estimate the oversampling and maximum measured spatial frequency for such a zone plate. I will do this for two representative detector sample distances Z_0 and for both CCDs whose parameters I have summarized in Tab. 2.1. The results are shown in Tab. 4.3 with the values for the 2k-chip in parentheses. All possible combinations satisfy the experimental goals of sufficient oversampling ($s > 3$) and high resolution ($\Delta x < 10$ nm). However, to be able to illuminate an entire zone plate of several hundred microns diameter with a monochromator pinhole of a few micron diameter (typically 5 μm) we need to move the entire monochromator

	$h\nu = 520$ eV	$h\nu = 750$ eV
$Z_0 = 10$ cm	$s = 24\,(35)$ $\Delta x = 9.2\,(8.65)$ nm	$s = 17\,(24)$ $\Delta x = 6.3\,(6.0)$ nm
$Z_0 = 5$ cm	$s = 12\,(18)$ $\Delta x = 4.6\,(4.3)$ nm	$s = 9\,(12)$ $\Delta x = 3.2\,(3.0)$ nm

Table 4.3: Oversampling and resolution as a function of sample detector distance and energy. Values in parentheses are for the 2k-chip.

assembly (pinhole and monochromator zone plate) further upstream. Looking at Fig. 4.1, one possibility is to move everything upstream by approximately 80 cm, that is, move the 750 eV zone plate to where the unused Order Sorting Mirrors (OSM) are, the 520 eV zone plate to where the 750 eV zone plate is and the monochromator pinhole to where the 520 eV zone plate is. This was done in preparation of the first SXDM experiment. I also used the occasion to improve on the stability and stiffness of the optical setup; these efforts are described in more detail in Sec. 4.1.4. With this change to the beamline, the distance between the beam-defining zone plate and the monochromator pinhole Z_{ZP} becomes roughly 77 cm. Assuming a monochromator pinhole with $D_P = 5\,\mu$m diameter we have a footprint of the beam on the zone plate of $D_{\mathrm{beam}} = 900\,\mu$m for 520 eV and $D_{\mathrm{beam}} = 620\,\mu$m for 750 eV. This corresponds to a linear overfill factor of 3 in each case.

We can now calculate the rate of photons in the focal spot according to Eq. 4.3. Assuming a typical zone plate diffraction efficiency of 10%, the total efficiency η (taking the overfill factor into account) is around 1%. A frozen-hydrated eukaryotic cell of 5 μm diameter transmits around 22% of the incident radiation of 520 eV. A typical reading on the photodiode for the photon flux after the monochromator pinhole is 75 nA (although it can vary in either direction by a factor of 2 or more). With this, I calculate the rate of photons in the focal spot to

$$n_{\mathrm{Ph}}^{(\mathrm{trans})}(520\,\mathrm{eV}) = 7.2 \cdot 10^6 \,\frac{\mathrm{photons}}{\mathrm{sec}} \tag{4.19}$$

$$n_{\mathrm{Ph}}^{(\mathrm{trans})}(750\,\mathrm{eV}) = 5.0 \cdot 10^6 \,\frac{\mathrm{photons}}{\mathrm{sec}} \tag{4.20}$$

We can now estimate the longest exposure time before saturation of our detector occurs using Eq. 4.16. The results for both energies and two sample detector distances Z_0 are summarized in Tab. 4.4, where values in parentheses are for the 2k-chip. All these times are substantially longer than the shortest exposure time that we can record with our experimental setup (1 msec), so I

	$h\nu = 520$ eV	$h\nu = 750$ eV
$Z_0 = 10$ cm	$t_{\mathrm{exp,max}} = 870\,(1420)$ msec	$t_{\mathrm{exp,max}} = 291\,(474)$ msec
$Z_0 = 5$ cm	$t_{\mathrm{exp,max}} = 111\,(178)$ msec	$t_{\mathrm{exp,max}} = 36\,(59)$ msec

Table 4.4: Longest exposure time before the detector saturates at both photon energies and 2 representative sample detector distances. Values for the 2k-chip are in parentheses.

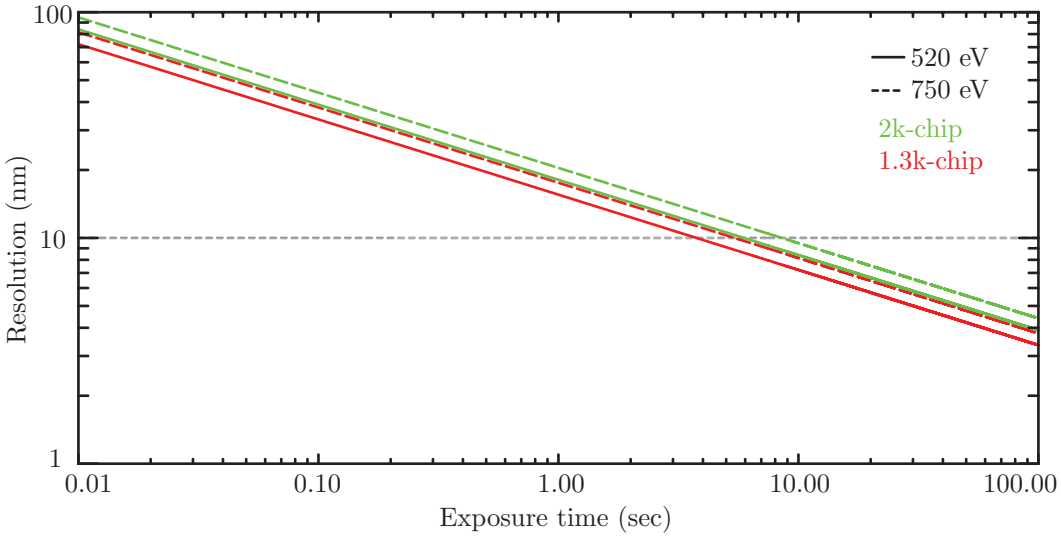

Figure 4.3: A log-log plot of Eq. 4.14 for the zone plate setup and both detectors. It illustrates that we can reach our experimental goal of recording statistically significant data out to 10 nm with an exposure time of $\lesssim 10$ sec for all combinations.

conclude that saturation effects are not a concern when using a zone plate as described above.

Finally, we want to estimate the exposure time required to record statistically significant data out to high spatial frequencies (corresponding to $d_{\mathrm{min}} < 10$ nm). Figure 4.3 shows a log-log plot of Eq. 4.14 for both detectors. We can see that an exposure time of about four seconds for the 1.3k chip and six seconds for the 2k chip at 520 eV is sufficient. The exposure times for 750 eV are slightly longer but of the same order of magnitude. In summary, the zone plate setup is able to achieve all of the experimental goals that I mentioned in the beginning of this section: it provides for sufficient oversampling, the spot size on the sample is of the desired size, we can record data of smallest spatial frequency q_{min} without having to worry about saturation, and we can

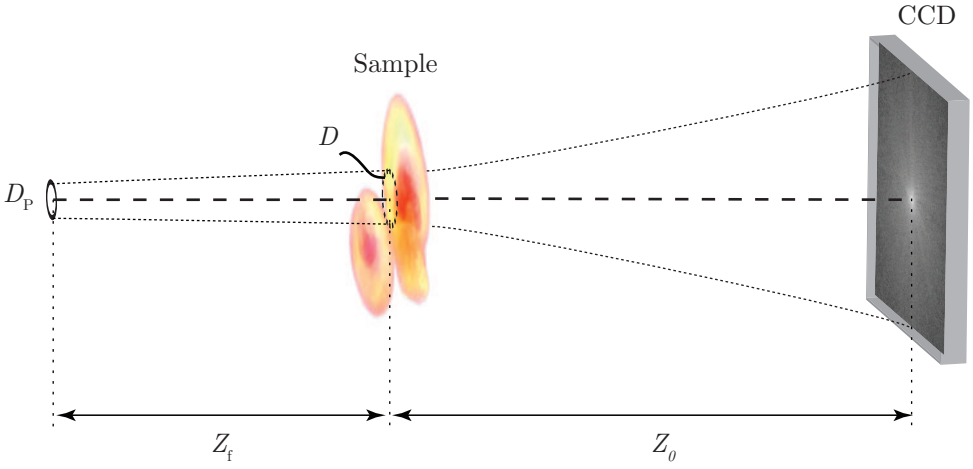

Figure 4.4: Schematic setup of SXDM with a pinhole as illumination source. The size of the beam footprint on the sample D due to a pinhole of size D_P can be determined numerically (similar to how I calculated the diffracted wavefield behind an Au sphere in Fig. 2.1).

record statistically significant data at our resolution target spatial frequency (corresponding to $d_{min} = 10$ nm) in only a few seconds of cumulative exposure time. Despite the requirement of having to rearrange the beamline to create the space that is needed for the additional zone plate, we decided to use this setup for our first SXDM experiment with our microscope. I will provide more detail on the experiment in Sec. 4.1.5.

Pinhole as Beam-defining Aperture

An alternative to using a zone plate as beam-defining aperture is to use a pinhole. A schematic of the setup is shown in Fig. 4.4. A pinhole of diameter D_P is a distance Z_f from the sample and illuminates it with a beam of size D. The detector then records the diffraction pattern a distance Z_0 behind the sample. To create a small beam footprint D on the sample with a pinhole it has to be very close to the sample, as I will explain in more detail below. The monochromator pinhole is located about an inch upstream of the sample; this is the location for 520 eV or 750 eV. We now have two options: a) introduce an additional pinhole after the monochromator pinhole that can be moved arbitrarily close to the sample or b) use the monochromator pinhole as beam-defining aperture. The latter involves a slight change in energy as we move the monochromator pinhole closer to the sample. This change is less than 3%

and thus tolerable. With that, the first option does not provide any additional advantages over the second one. Instead, introducing a second pinhole will severely limit the available flux. This is why I decided early on to go with option b) and I will also restrict my explanations here to this type of setup. The new energies considered in the following derivation are thus 535 eV and 770 eV.

First, I want to determine the best combination of pinhole size D_P and pinhole-sample separation Z_f in terms of size of the beam footprint on the sample D. Typical pinhole sizes are on the order of a few micrometers; smaller pinholes are difficult to produce and tend to close up fast due to carbon buildup from imperfect vacuum. In the far-field the shape of the beam that is originating from the pinhole will be an Airy-pattern with footprint of the central maximum of

$$D = \frac{2.44\,\lambda\,Z_\mathrm{f}}{D_\mathrm{P}}, \tag{4.21}$$

where D is always larger than D_P. Thus it seems unlikely that we will be able to achieve a spot on the sample of size $1\,\mu$m or less with pinholes that are $1\,\mu$m or more in diameter. If we bring the pinhole close to the sample such that we have Fresnel diffraction rather than far-field diffraction, the situation is slightly different. The transition from near to far-field is given by the dimensionless Fresnel number introduced in Eq. 2.9. We can see that for a pinhole of $3\,\mu$m diameter, near-field diffraction effects dominate up to a distance Z_f of 3.90 mm for 535 eV light. The behavior of the diffracted beam is now no longer described by Eq. 4.21. To estimate the footprint of a pinhole in the near-field, we can numerically propagate the wavefield immediately behind the pinhole to the plane of interest; this is similar to how I calculated the diffracted wavefield behind a gold sphere in Fig. 2.1. I have done such analysis for different size pinholes using our IDL routine `propagate.pro` which evaluates the propagator function derived in Collier et al. [84]. The results are shown in Fig. 4.5. On top, cross sections through the intensity profile of a wavefield emanating from a pinhole for different propagation distances are plotted as surface plot for a $2\,\mu$m diameter pinhole on left and a $5\,\mu$m diameter pinhole on right. The illuminating photons have an energy of 535 eV. We can see the transition from almost geometric projection for short propagation distances to near-field diffraction in both cases. For the $2\,\mu$m pinhole we can also see the far-field diffraction in form of a diverging Airy pattern for propagation distances of more than 1.7 mm which corresponds to a Fresnel number of less than one. On bottom, the ratio of intensity within a $1\,\mu$m spot over total intensity is plotted as a function of propagation distance for a $2\,\mu$m pinhole on left and a $5\,\mu$m pinhole on right. The simulations show that a $2\,\mu$m pinhole at

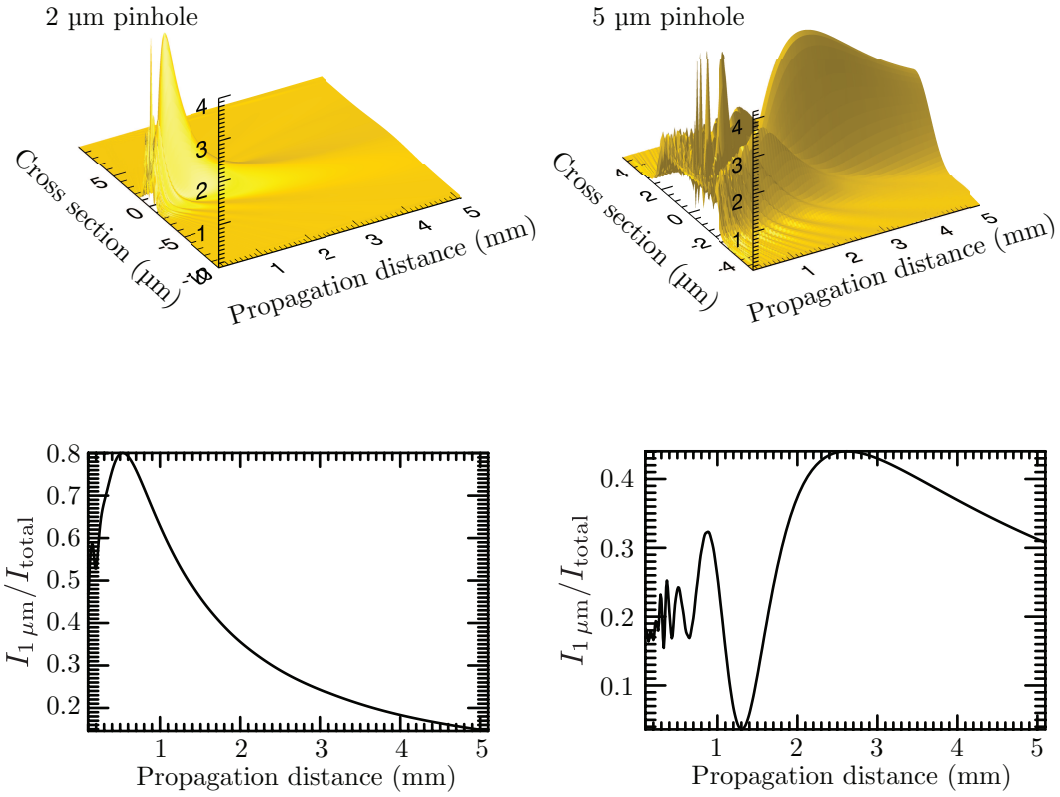

Figure 4.5: On Top: cross sections through the intensity profiles of the beam footprint produced by a pinhole as a function of propagation distance for two different size pinholes; 2 μm on left and 5 μm on right. The transition from geometric to near field diffraction can be seen. The photon energy is 535 eV. On bottom: ratio of intensity within a 1 μm spot over total intensity as a function of propagation distance for a 2 μm pinhole on left and a 5 μm pinhole on right.

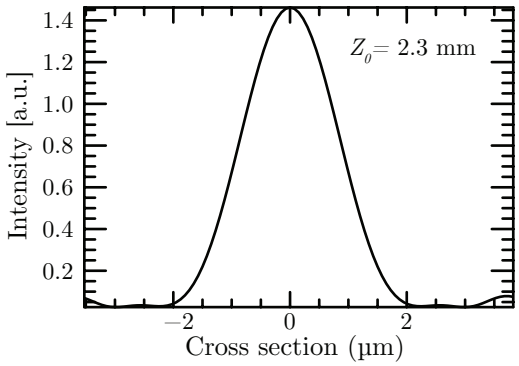

Figure 4.6: Cross section through the intensity profile of the beam footprint produced by a 3 μm pinhole at a distance of 2.3 mm. Most of the intensity is contained within a 4 μm spot.

a propagation distance of less than 1 mm produces a beam that has most of its intensity within a 1 μm footprint; however, a distance between pinhole and sample of less than 1 mm is not experimentally feasible. The wavefield behind a 5 μm pinhole (shown on right) has at most 40% of its intensity concentrated in a 1 μm spot. These results as well as repeating the same analysis for more pinhole sizes show that we won't be able to find a combination of pinhole size and pinhole-sample distance that results in a beam on the sample of size 1 μm or less. The same is true for photons of 770 eV (not shown).

We can work with a larger size beam on the sample if we can tolerate more dose to our specimen. Giving up on this goal, we can for example work with a 3 μm diameter pinhole that will produce a beam of approximately 4 μm in size at a distance of $Z_0 = 2.3$ mm for 535 eV light. A cross section through the intensity profile of the wavefield produced by a 3 μm pinhole at $Z_0 = 2.3$ mm is shown in Fig. 4.6. Most of the intensity is contained within a spot of 4 μm diameter. For 770 eV light a similar spot profile can be obtained at $Z_0 = 3.4$ mm. A 3 μm diameter pinhole will transmit less than the 5 μ pinhole that we typically use as monochromator pinhole but I will show below that it is still tolerable. A pinhole-sample distance of 2.3 mm is also experimentally convenient as it allows us to introduce a corner between pinhole and sample.

Similar to the zone plate setup, I now want to characterize the pinhole setup in terms of oversampling, resolution and flux for both energies and two representative sample-detector distances. The values for oversampling and resolution are again calculated using Eqs. 2.21 and 2.16 and are summarized in Tab. 4.5. Note that here I have chosen $Z_0 = 15$ cm as the second option

	$h\nu = 535$ eV	$h\nu = 770$ eV
$Z_0 = 10$ cm	$s = 3\,(4)$ $\Delta x = 8.9\,(5.8)$ nm	$s = 3\,(3)$ $\Delta x = 6.2\,(5.8)$ nm
$Z_0 = 15$ cm	$s = 5\,(6)$ $\Delta x = 13.3\,(8.7)$ nm	$s = 5\,(5)$ $\Delta x = 9.3\,(8.7)$ nm

Table 4.5: Oversampling and resolution as a function of sample detector distance and energy. Values in parentheses are for the 2k-chip.

	$h\nu = 535$ eV	$h\nu = 770$ eV
$Z_0 = 10$ cm	$t_{\mathrm{exp,max}} = 0.04\,(0.06)$ msec	$t_{\mathrm{exp,max}} = 0.01\,(0.02)$ msec
$Z_0 = 15$ cm	$t_{\mathrm{exp,max}} = 0.14\,(0.20)$ msec	$t_{\mathrm{exp,max}} = 0.03\,(0.07)$ msec

Table 4.6: Longest exposure time before the detector saturates at both photon energies and two representative sample detector distances. Values for the 2k-chip are in parentheses.

for the sample-detector distance. We can see from the results that we barely satisfy the oversampling requirements for the shorter distance. Working at the longer distance and with the 2k-chip we can still satisfy both the oversampling requirements and our resolution target of 10 nm.

We can estimate the flux transmitted through the pinhole using Eq. 4.3. Since we are dealing with a $3\,\mu$m pinhole now, we would expect a smaller reading on the photodiode, say 30 nA. At the same time, we have no efficiency losses as with the zone plate setup, thus $\eta = 1$. With this, the incident flux becomes

$$n_{\mathrm{Ph}}^{(\mathrm{trans})}(535\,eV) = 2.8 \cdot 10^8 \,\frac{\mathrm{photons}}{\mathrm{sec}} \tag{4.22}$$

$$n_{\mathrm{Ph}}^{(\mathrm{trans})}(770\,eV) = 2.0 \cdot 10^8 \,\frac{\mathrm{photons}}{\mathrm{sec}} \tag{4.23}$$

Using these numbers and Eq. 4.16, we can calculate the longest exposure time before saturation occurs for both energies and sample-detector distances. Due to the slight change in energies, the maximum number of photons before saturation occurs is now 1364 (682) for 535 eV and 948 (474) for 770 eV. The results are summarized in Tab. 4.6. All of these values are below the shortest exposure time that we can record with our experimental setup. This setup therefore requires the use of a beamstop as described in Sec. 2.2.4 or some means to attenuate the incident beam by at least a factor of 10 as described in Sec. 4.1.3.

Finally, I want to calculate the exposure time that is required to record

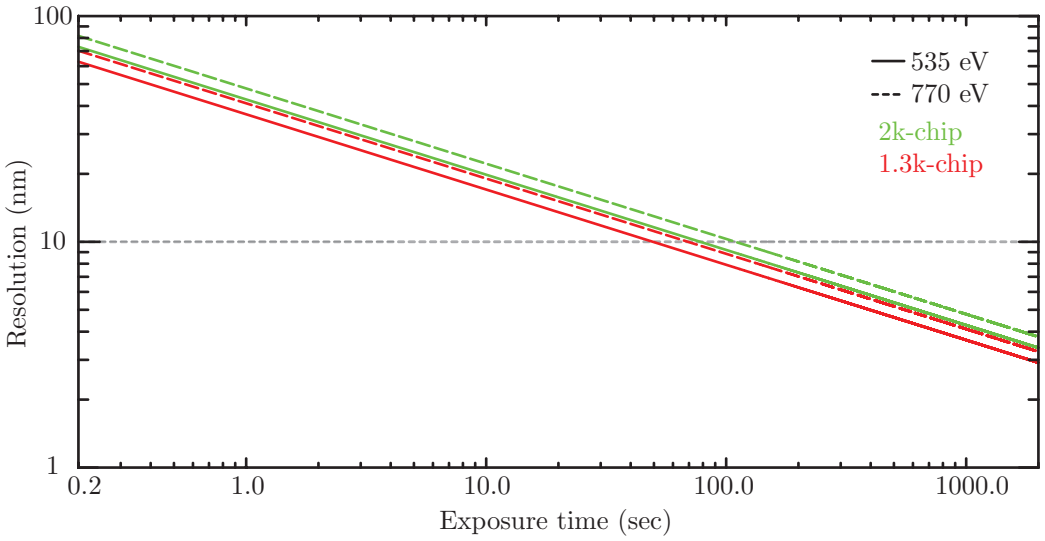

Figure 4.7: A log-log plot of Eq. 4.14 for the pinhole setup and both detectors. It illustrates that we can reach our experimental goal of recording statistically significant data out to 10 nm with an exposure time of $\lesssim$ 100 sec for all combinations.

statistically significant data at our resolution target length scale of 10 nm. Figure. 4.7 shows a log-log plot of Eq. 4.14 for the pinhole setup at both energies and for both detectors. To fulfill the goal of 10 nm resolution, we have to expose the sample from 50 seconds to 110 seconds depending on detector and energy. In summary, the pinhole setup does not fulfill all of our goals. We have to compromise on the spot size and we barely satisfy the oversampling requirements. Due to the increased flux and the small oversampling we cannot record the lowest spatial frequencies without saturating the detector. We are able to record data at our target resolution length scale but the required exposure times are an order of magnitude longer than with the zone plate setup. I therefore conclude that the zone plate setup is the better option. However, as I will describe in more detail in Sec. 4.1.5, the pinhole setup is experimentally more feasible.

4.1.3 Recording Low-spatial Frequency Data

We would like to be able to record down to lowest spatial frequencies. Simulations have shown [11] that missing low spatial frequency data results in

poor contrast in the reconstruction; even when a perfect support is assumed. The calculations presented above show that we can record the center speckles with the zone plate setup. To achieve the same for the pinhole setup I will introduce here an experimental solution.

Based on the maximum exposure times that are presented in Tab. 4.6 we can see that, in order to be able to record 1 msec exposures without saturation, we need to attenuate the incident beam by a factor of 10 to 100. The exact amount depends on the actual geometry of the setup and the flux that is available during the experiment. The amount of data that is typically lost when using a beamstop amounts to an area of 20×20 pixels on the 1.3k detector; that is, a square of $400 \times 400\,\mu$m. To attenuate the beam over such an area, I chose to coat a silicon nitride window with gold and mount it directly in front of the CCD. Standard windows are available in sizes of up to 5×5 mm; however, the window membrane is only about 150 nm thin so that smaller windows are more stable. To leave room for overlapping data between positions with and without the beam attenutator, I decided on a 1×1 mm window in a 6×6 mm frame. The thickness of the gold layer required to attenuate the incident beam by a factor of 10 can be calculated from Eq. 1.15. Looking up the tabulated value of β [79] for both gold and silicon-nitride at 535 eV, we find $\beta_{\mathrm{SiN}} = 7.36 \cdot 10^{-4}$ and $\beta_{\mathrm{Au}} = 4.188 \cdot 10^{-3}$. Solving

$$0.1 = \exp\left[-2k\left(\beta_{\mathrm{SiN}} t_{\mathrm{SiN}} + \beta_{\mathrm{Au}} t_{\mathrm{Au}}\right)\right] \tag{4.24}$$

for the thickness of the gold layer we find that $t_{\mathrm{Au}} \approx 75$ nm; I have assumed that $t_{\mathrm{SiN}} = 150$ nm. For photons of 770 eV we would need a thicker layer of gold but we explicitly designed it for the lower energy. The silicon-nitride windows were coated by Pete Davis from the Physics Department at Stony Brook University using an evaporator. To be on the safe side we deposited close to 100 nm of gold.

To record the low spatial frequency data with this device, we would first run a regular data acquisition script with beamstop (as described in Sec. 2.2.4), then center the beam attenuator on the direct beam and record a few exposures with it. To be able to move the beam attenuator in the vacuum chamber, I redesigned the mount for the beamstop so that it also accommodates the beam attenuator. An image of the mount, with both beamstop and beam attenuator glued to it, is shown in Fig. 4.8. In the actual experiment we noticed that the CCD was still slightly saturating even with the beam attenuator. We therefore had to narrow the horizontal beam defining aperture behind the undulator to further decrease the flux. With this we were able to record the whole diffraction pattern down to lowest spatial frequencies. An example diffraction pattern that was obtained by merging positions with and without beam attenuator

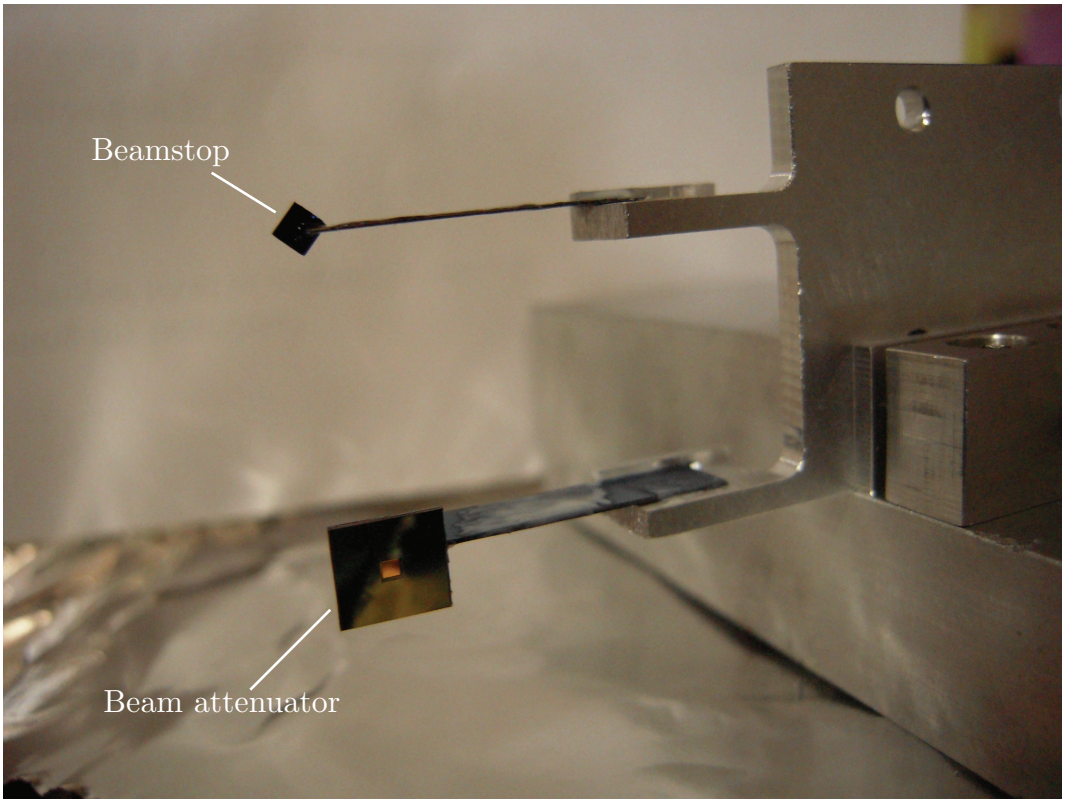

Figure 4.8: Combined mount to attach both beamstop and beam attenuator to a motor stage. Photo courtesy of Johanna Nelson.

is shown in Fig. 4.9. The newly developed merging program AMP (described

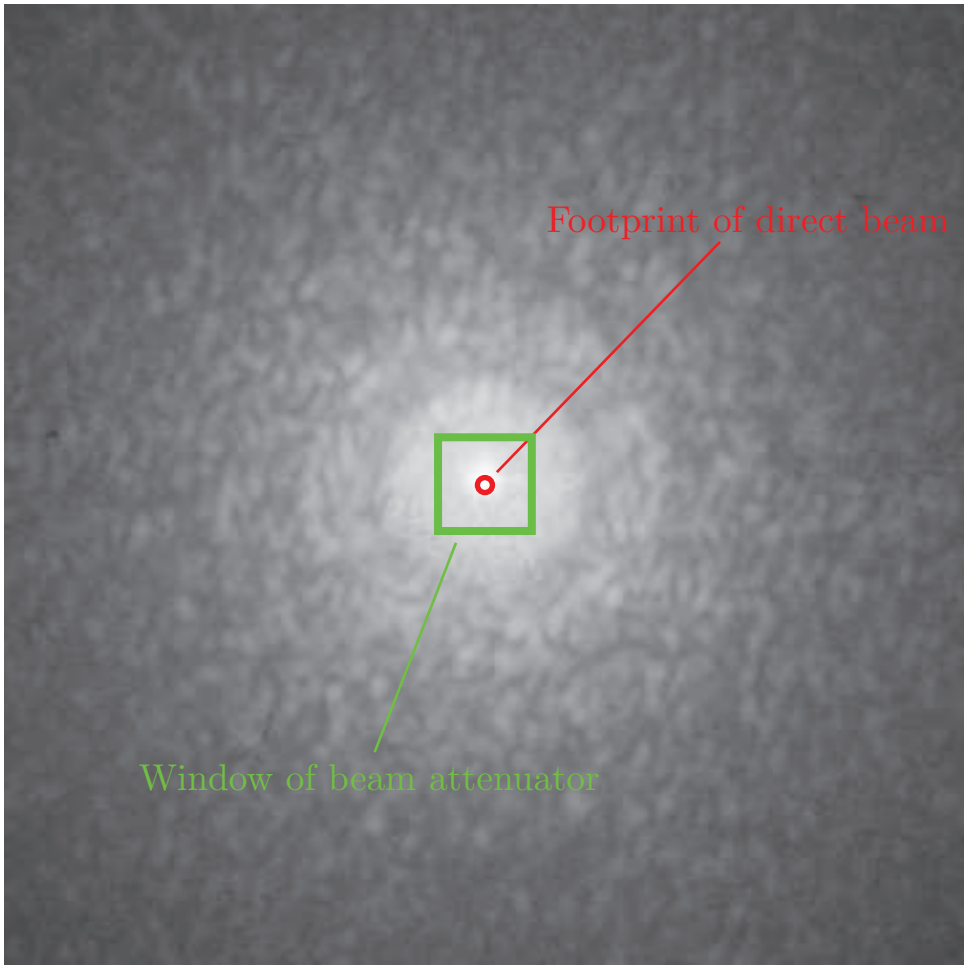

Figure 4.9: Diffraction pattern obtained by merging several positions with and without beam attenuator using AMP. The extent of the direct beam on the CCD is indicated as a red circle. The window of the beam attenuator is indicated as a green square.

in Sec. 3.3) was used to assemble the pattern from the various beamstop and beam attenuator positions. It automatically scales the different positions with respect to each other; this takes care of differences in flux incident on the detector for positions with and without beam attenuator. The footprint of the direct beam on the CCD is indicated by a red circle; the window of the beam attenuator is indicated by a green square. To convince ourselves that we can record the data and have AMP take care of the scaling differences, we can

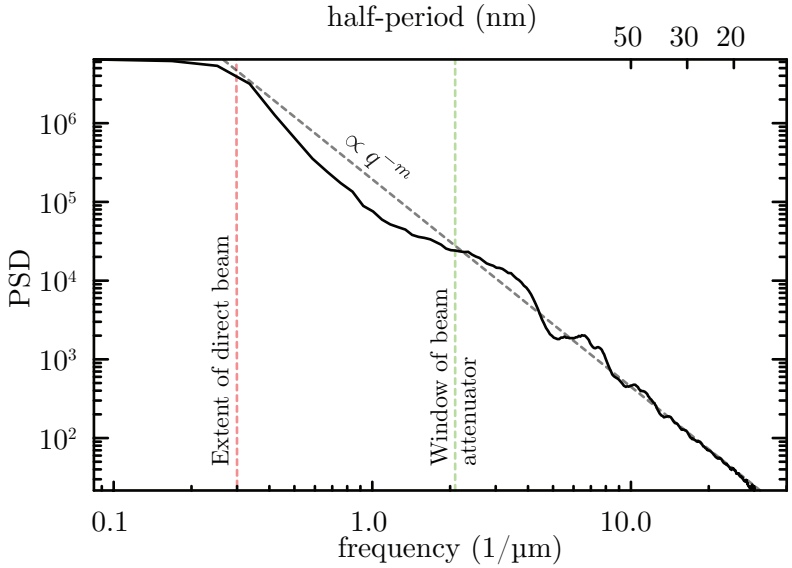

Figure 4.10: Power spectral densities (PSD) of the diffraction pattern shown in Fig. 4.9. The extent of the direct beam is indicated with a red dashed line; the extent of the beam attenuator with a green dashed line. As expected, the PSD follows an inverse power law approximated by the black dashed line in this log-log plot. The fact that the slope seems to be the same for all spatial frequencies confirms that `AMP` successfully deals with the scaling differences between data collected with beam attenuator and data collected without.

look at the power spectral densities (PSD) of the merged diffraction pattern. Figure 4.10 shows a plot of the PSD of the diffraction pattern shown in Fig. 4.9. For this particular data set, we used a pinhole of $5\,\mu$m diameter D_P which was mounted a distance $Z_\mathrm{f} + Z_0 = 12.5$ cm from the detector. Thus, the footprint of the direct beam on the CCD would be an Airy pattern of size

$$w = \frac{2.44\lambda \cdot (Z_\mathrm{f} + Z_0)}{D_\mathrm{P}} = 145\,\mu\mathrm{m} \tag{4.25}$$

which is equivalent to about 7 pixels on the 1.3k chip. Since our frequency sampling given by Eq. 2.18 is $\Delta q = 0.084\,\mu\mathrm{m}^{-1}$, we would expect to see effects of the direct beam to extend out to a spatial frequency of $3.5 \cdot 0.084\,\mu\mathrm{m}^{-1} \approx 0.3\,\mu\mathrm{m}^{-1}$; this frequency is indicated with a red dashed line in Fig. 4.10. As expected, there is a plateau in the PSD out to this frequency before it starts to decline. Similarly, the window of the beam attenuator should extend out to a spatial frequency of $25 \cdot 0.084\,\mu\mathrm{m}^{-1} = 2.1\,\mu\mathrm{m}^{-1}$; this is indicated with a

green dashed line. We can now confirm the scaling by looking at the trend of the PSD as a function of frequency. As discussed in Sec. 2.2.4, we would expect the PSD to decline as q^{-m} which in a log-log plot is represented as a straight line of slope $-m$. Approximating the average trend of the PSD with a straight line, we see that both ranges of spatial frequencies (recorded with and without beam attenuator) follow the same inverse power law. This confirms the validity of the approach.

4.1.4 Improving the Stability of the Existing Setup

One disadvantage of SXDM as compared to full-field XDM is the high stability requirements of the experimental setup. As discussed before, the uncertainty in relative position between sample and illumination limits the achievable resolution. We have to distinguish between absolute aberrations in positions (caused by limited resolution of motor encoders or slow thermal drifts of mechanical components during the course of a data acquisition script) and vibrations of optical components. The latter is more detrimental to the experiment. My main focus in improving the stability and stiffness of our optical setup was to a) decrease the length of the mechanical path that separates the sample and the beam-defining aperture, b) redesign the monochromator zone plate mounts to provide a more stable illumination of the monochromator pinhole.

Decreasing the mechanical path between the sample and the beam-defining aperture was achieved by moving the mounting point for the motor stage that controls the latter to the wall directly next to the sample interlock. Previously, the attachment point was on the double rails. This was convenient because both sample and illumination could be aligned independently; however, the mechanical path is orders of magnitude longer and involves in-vacuum as well as out of vacuum components. The location of the new and old mount point in relation to the sample and other components of the optical path are shown in Fig. 4.11. The actual bracket to hold the motor stage to the wall was designed by Rich Celestre of the Experimental Systems Group of the Advanced Light Source.

To provide a more stable illumination of the monochromator pinhole, and to be able to use a zone plate as beam defining aperture, I had to move and redesign the existing monochromator setup; as described in Sec. 1.3.1, the monochromator consists of a zone plate and a pinhole further downstream. The old monochromator zone plate mounts are shown in Fig. 4.1; they are labeled with 750 eV and 520 eV, respectively. The location of the monochromator pinhole is just in front of the sample inside the vacuum chamber. As I have already explained in Sec. 4.1.2, we have to shift everything upstream by about 80 cm to be able to use a zone plate. Thus, the new position of

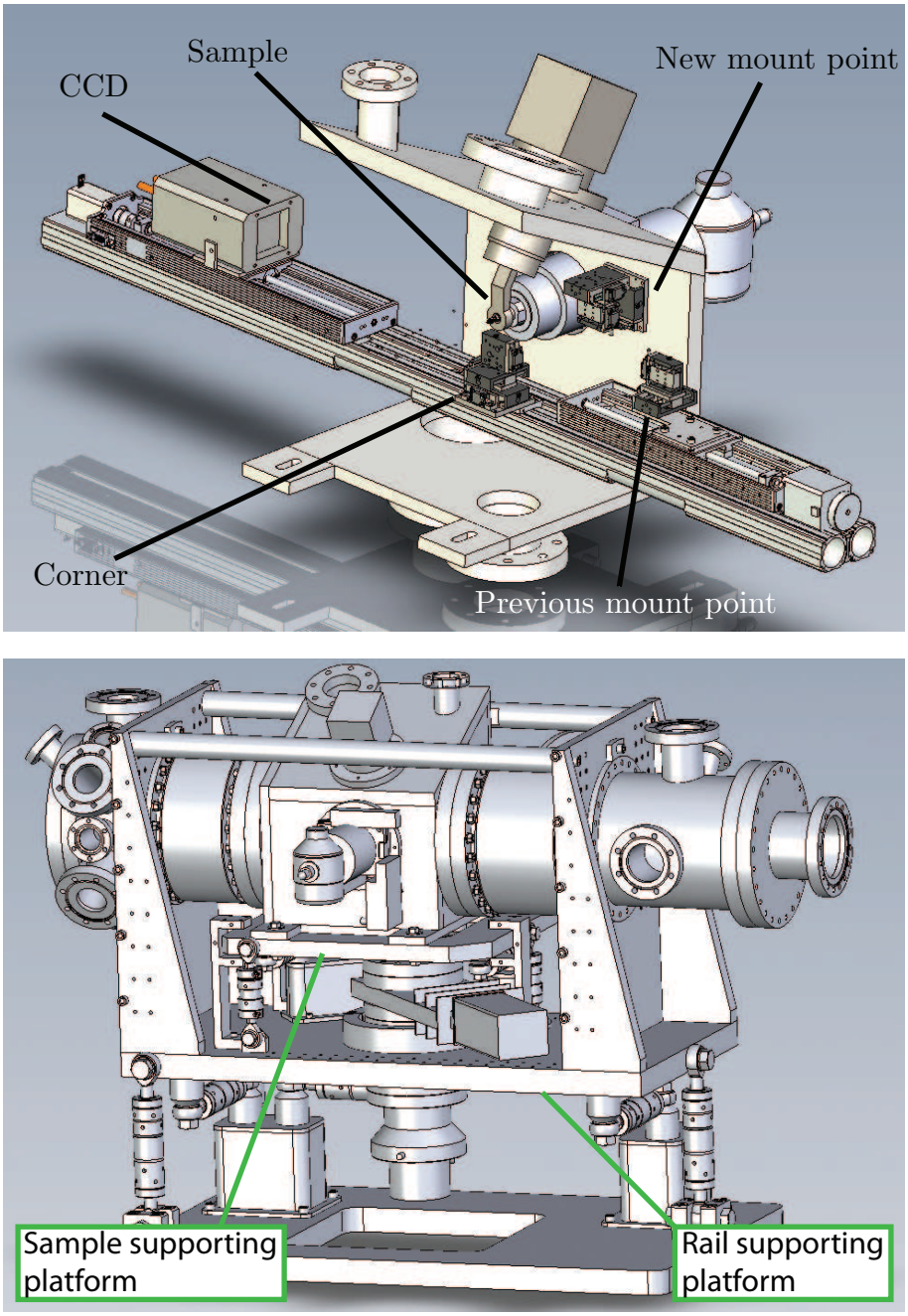

Figure 4.11: The mechanical path between the beam-defining aperture and the sample has been shortened by moving the mounting point of the latter next to the sample. The old mounting point on the rails was convenient but was separated by a long mechanical path from the sample. Solidworks drawing courtesy of Rich Celestre.

the monochromator pinhole will be where the current 520 eV zone plate is mounted, which in turn will be moved to the position of the 750 eV zone plate, which in turn will be moved to where the old order sorting mirrors (OSM) are. I took this opportunity to replace the old monochromator zone plate mounts which have been found in the past to be not very convenient to adjust and to drift slowly over time. They consist of an in-vacuum linear translation stage that is coupled with an out of vacuum tilt stage to provide x-y motion. The

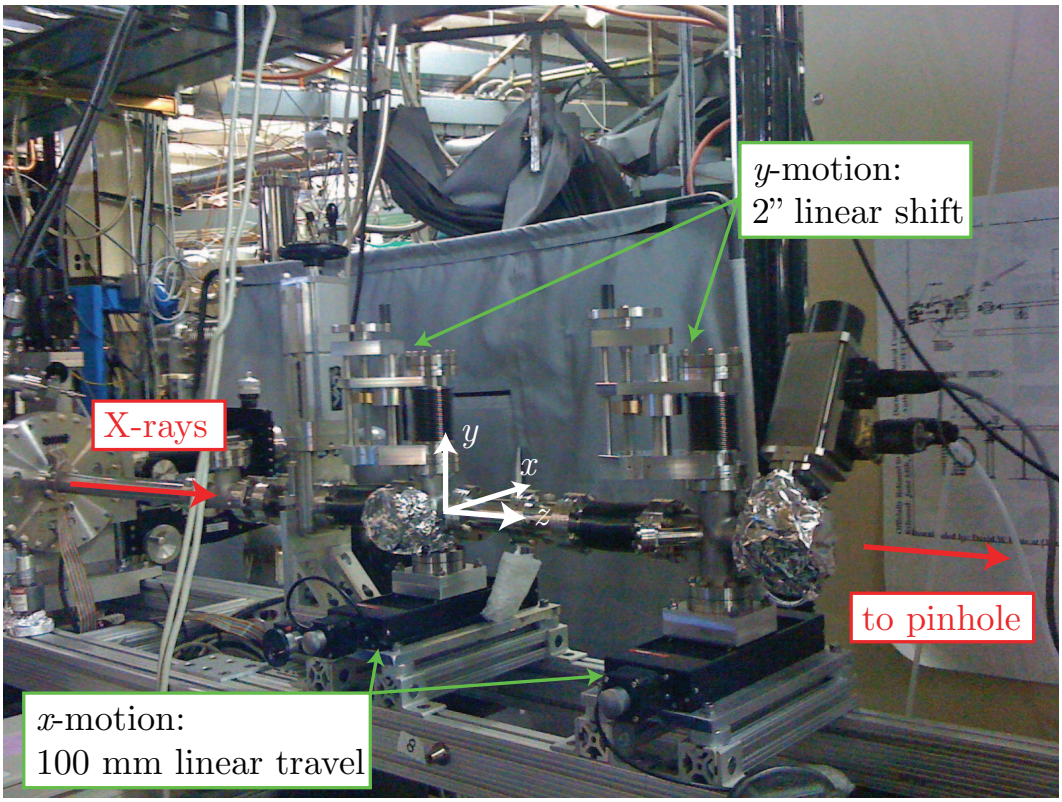

Figure 4.12: New monochromator zone plate mounts. They offer truly independent x, y-motion and are more stable than the previous ones. Note that during installation, the vacuum connection to the pinhole has been unmounted.

instability could be due to the design of the mounting brackets that clamp onto the flange of the 4-1/2" cross on the one end. On the other end they have a pair of vertical slits through which they are bolted onto the supporting rails. The clamping mechanism is bound to over-constrain the position of the zone plate mounts which will lead to competing equilibrium positions of the assembly. The vertical slits are convenient for adjusting the height of the zone plate

mounts; however it is likely that they will creep under the influence of gravity over time. To improve on this, I replaced both mounts with a new assembly that is shown in Fig. 4.12. The in-vacuum y-motion of the zone plates is performed by a 2" linear shift mechanism from Kurt J. Lesker company which is mounted on top of the 4-1/2" cross. The entire y-assembly is then bolted to a 100 mm linear travel stage from Thorlabs which is responsible for the motion in x. The Thorlabs stage is bolted top-down to the supporting rails. Aside from providing true independent x, y-motion, the new monochromator zone plate mounts are very stable; they can also be equipped with stepper motors for remote control. To mount the monochromator pinhole in its new position,

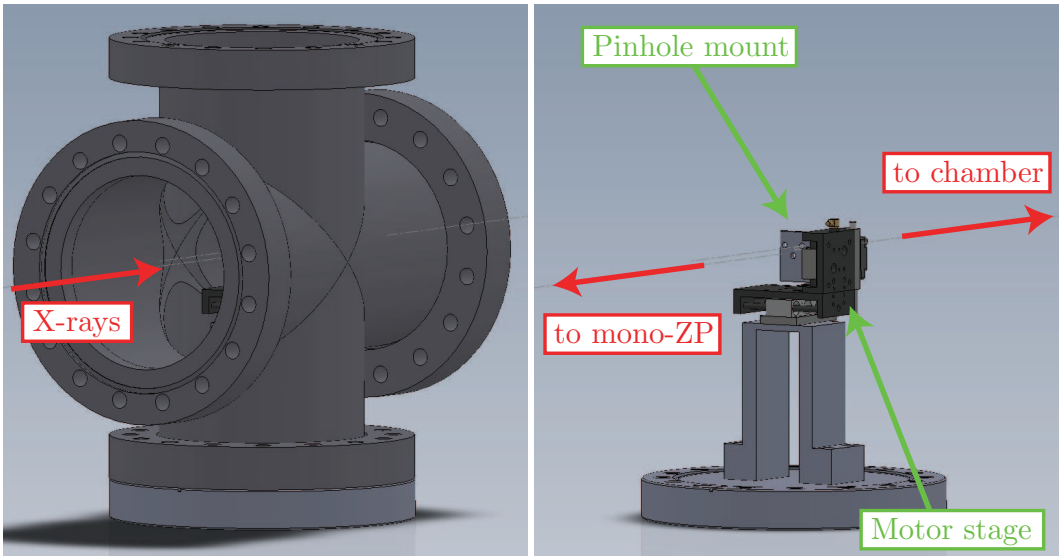

Figure 4.13: 3D rendering of the new monochromator pinhole mount. The motor stage that moves the mount sits in a 6" cross (shown on left, hidden on right), that is attached to the chamber by a 4" tube (not shown).

I decided to extend the vacuum chamber further upstream by mounting a 6" cross to a 4" tube that connects to the 10" upstream flange of the chamber; where the 6" cross houses the motorized x, y-stage for the monochromator pinhole. A 3D rendering of the pinhole stage inside the 6" cross is shown in Fig. 4.13; the cross is hidden on the right to show the pinhole mount inside it. Both the monochromator pinhole and the zone plate inside the chamber are now supported by the same platform which will minimize the relative drift between the two components. Finally, to still be able to use the old full-field setup, I designed the monochromator pinhole mount such that it can also hold

a monochromator zone plate; which is then at the correct distance for 520 eV with the monochromator pinhole an inch away from the sample.

4.1.5 Results of ALS Experiments

So far we have conducted three SXDM experiments using three different geometries; one used a zone plate as beam-defining aperture and two relied on a pinhole. Here, I want to summarize the main findings of all three runs.

Based on the results of Sec. 4.1.2 we implemented a zone plate setup first. Unfortunately we spent most of the time aligning the zone plate and order sorting aperture (OSA) with the x-rays. This was partly due to the fact that we had unforeseen motor problems but mostly because the chamber is not well equipped for precise alignment of optics. Even though we did get the zone plate and OSA aligned and focused on our sample it was not until the last day of beam time. We also noticed strange light leakage through the central stop of the zone plate presumably due to having higher order photon energies in our illumination; these are not filtered out with our current monochromator setup. In addition to that, the leaking light scattered off of the OSA and projected an image of our specimen onto the CCD. We misinterpreted the faint projection of the grid bars as speckles; thus the data that was recorded is of limited use. Note though, that we were able to use dwell times of up to 2 sec without saturation of the detector. Given the poor incident flux that we had on that day, these values are in agreement to the values estimated for zone plate illumination in Tab. 4.4.

For the next experiment we implemented the pinhole setup that was described in Sec. 4.1.2. I designed two mounts to hold the pinhole and a corner. Since both corner and pinhole had to be very close to the sample ($Z_\mathrm{f} = 2.3$ mm), we had to modify the ACD shield. A 3D rendering of the setup is shown in Appendix B.1. During the experiment we again had trouble with alignment of the optics with respect to the beam. Once we were aligned, we noticed the beamstop and the corner become misaligned over time due to drifting motor stages. Nonetheless we were able to record a ptychographic data set on a sample of gold balls. Unfortunately the exposure times were not properly matched so that there are missing data in the intermediate spatial frequency range. We also recorded some data on a frozen-hydrated sample but found that the scattered signal was coming from the plane of the corner instead of the sample; the closeness of the corner stage to the cold sample lead to a buildup of scatterers on the corner stage. Both data sets are thus of limited use. It also puts a question mark on whether this type of setup is useful for the imaging of frozen-hydrated specimen.

The most promising experiment so far was conducted in May of 2010. It

also involved a pinhole but this time the goal was to have a setup that can be used for both full-field XDM and SXDM. Working at 520 eV, we used a $5\,\mu$m diameter pinhole at a distance of $Z_{\mathrm{f}} = 2.5$ cm from the sample. In this geometry we would expect the beam footprint on the sample to have a diameter of about $30\,\mu$m. This should violate the oversampling requirements according to Eq. 2.21. However, since we were working with samples prepared for full-field, they were isolated so that the combined size of all scatterers within the illumination was small enough to satisfy oversampling requirements. We recorded data on a mixture of different size gold balls, as well as on freeze-dried and frozen-hydrated yeast cells. This was also the first SXDM experiment where we used the beam attenuator described in Sec. 4.1.3. I have already shown one example merged diffraction pattern of the gold ball data in Fig. 4.9. The entire data set consists of 4×4 positions where the step size between adjacent positions was $5\,\mu$m. The quality of the data is illustrated by the autocorrelation of the object at each scan position (the autocorrelation of the object is given by the inverse Fourier transform of the recorded diffraction pattern). Figure 4.14 shows the autocorrelation at each scan position. In each case the object is sufficiently oversampled and ghosts in the autocorrelation suggest the presence of smaller scatterers in the vicinity. The other recorded data sets are of similar quality; this is promising for good reconstruction results in the near future.

The experiments showed us that the zone plate setup is the best choice for SXDM on frozen-hydrated specimen. The flux estimates derived in Sec. 4.1.2 have been confirmed. Thus, out of the three experimental geometries that we tested, this setup seems the only one that is suited for dose-efficient and speedy data collection on frozen-hydrated samples. Unfortunately, the microscope chamber in its current form is ill equipped for zone plate imaging. The pinhole setup is not favorable in terms of spot size and dynamic range. A close pinhole interferes with the required shielding for frozen-hydrated imaging. A pinhole that is further away restricts ourselves to isolated samples to satisfy oversampling requirements and it is also not very dose efficient. However, we were able to record nice data with the help of the beam attenuator with this setup.

4.2 SXDM Experiment at the APS

To get some experience with SXDM experiments before implementing the setup described above, I also took part in a collaborative experiment where Christian Holzner was the lead researcher from our group. The experimental goal was to simultaneously record phase contrast and SXDM data using the

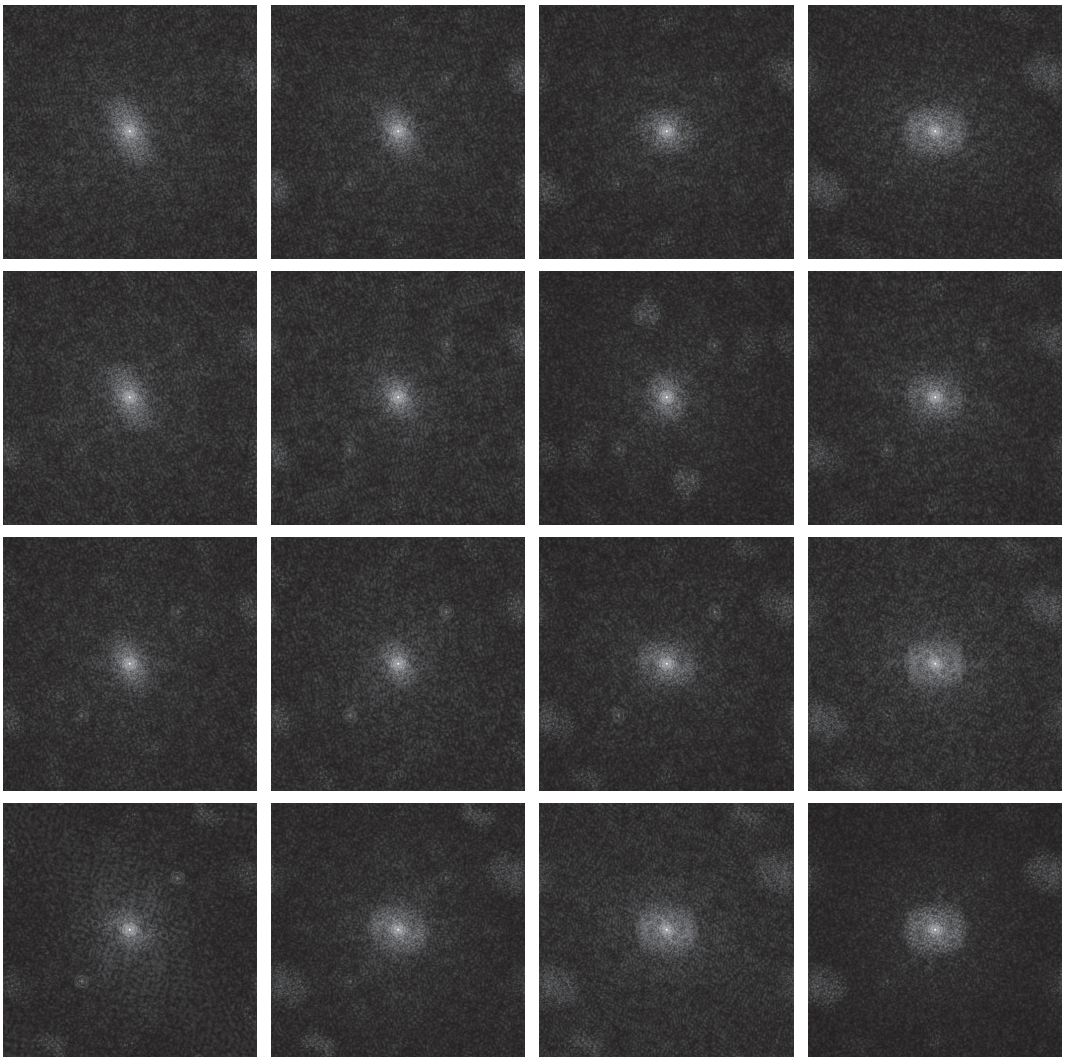

Figure 4.14: Autocorrelation of the object at each scan position of SXDM data taken on a mixture of different size gold balls. The data satisfy oversampling requirements. Ghosts in the autocorrelations suggest the presence of smaller scatterers in the vicinity.

pixel array detector (PAD) developed by Vernon et al. [38]. We used the existing Scanning Transmission X-ray Microscope (STXM) at beamline 2-ID-B of the Advanced Photon Source at Argonne National Lab. The experiment was a collaboration between our group, the Cornell detector group and the 2-ID-B staff. An overview of the setup is shown in Fig. 4.15. A pixel array

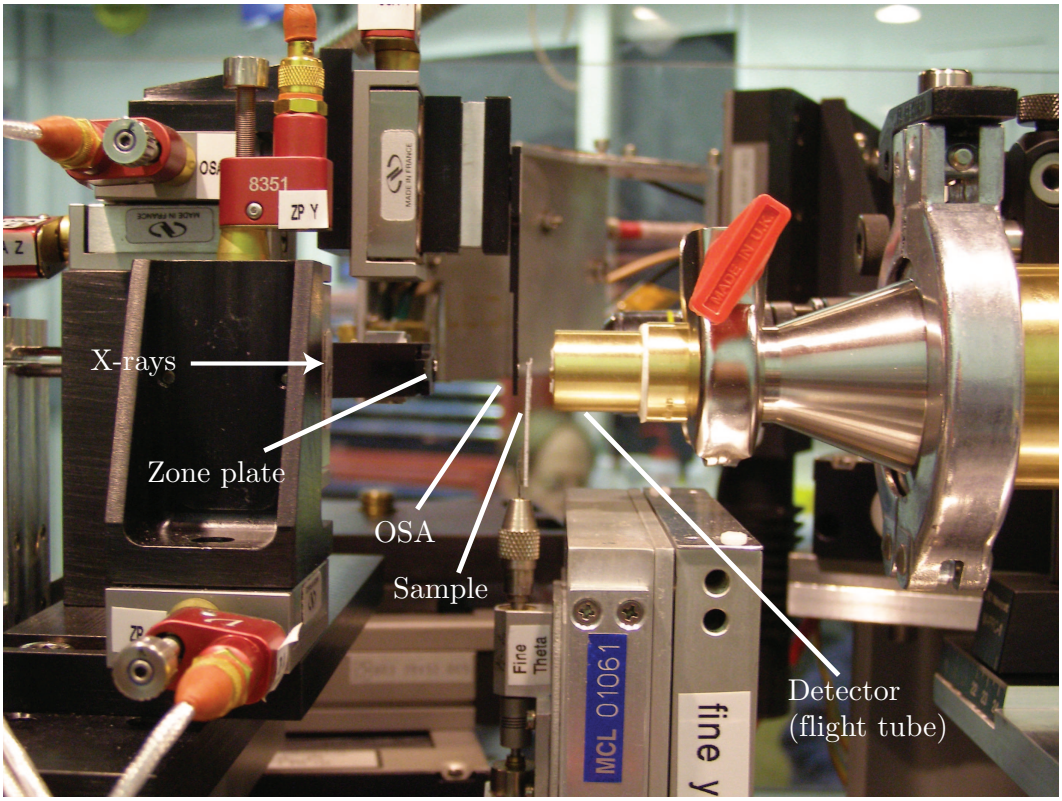

Figure 4.15: Side view of the SXDM setup at beamline 2-ID-B of the Advanced Photon Source. The x-rays were focused onto the sample using a zone plate and the scattered signal was recorded by a pixel array detector. Photo courtesy of Christian Holzner.

detector has the advantage of having a very large dynamic range, since each pixel has its own read-out electronics. The detector we used consisted of 128×128 pixels; each was $150 \times 150\,\mu$m in size. As zone plate we used a $160\,\mu$m diameter, $450\,\mu$m thick gold zone plate with an outermost zone width of 50 nm (thus a focal spot of 60 nm). The energy of the photons was 2850 eV and the sample-detector distance was about 22 cm. With these parameters, we can calculate the oversampling and the theoretical resolution using Eqs. 2.21

and 2.16 to $s = 10$ and $d_{\min} = 10$ nm, respectively.

We recorded SXDM data (and phase contrast data simultaneously) on several samples: two Siemens star test patterns and one material science sample. While the data analysis is still ongoing, I want to show a preliminary result here. Figure 4.16 shows the best reconstruction results of data taken on a 180 nm thin gold test pattern; only the phases are shown. The entire field of view of the scan is $2.5 \times 2.5\,\mu$m, the step size was 25 nm in each direction. At each position we exposed for 42 msec. The diffracted data was reconstructed with the algorithm described in Sec. 2.3.1; the code has been written in `IDL`. The data was reconstructed for only 20 iterations. This was possible because I previously reconstructed the probe from a small subset of the same scan and then used it for the reconstruction of the entire field of view; this saves computation time. Looking at Fig. 4.16 we can recognize the spoke pattern of the Siemens star. We also see that the small features in the top right quadrant are not clearly reproduced. We think that this is due to vibrations of the sample with respect to the beam. Further evidence for this is given in the absorption and phase contrast images of the same scan which also show the same "washed-out" structures in the top right quadrant. We plan on addressing the errors in the relative position of sample and probe by implementing a conjugate-gradient algorithm which has been shown to be able to right incorrect probe positions [52].

Figure 4.16: Phases of a gold test pattern reconstructed from SXDM data. The spokes are visible. Smaller features toward the top have a washed-out appearance. This indicates errors in the assumed relative positions of sample and probe (possibly due to vibrations of the sample during data collection). The entire field of view of the scan is $2.5 \times 2.5\,\mu$m.

Chapter 5

Conclusions and Outlook

The goal of 3D high resolution imaging of frozen-hydrated biological specimen with x-ray diffraction microscopy is challenging but promises new insights into the molecular processes of these cells. The contributions of this thesis address some of the remaining challenges to achieve that goal. Here, I want to summarize the main results and outline the next steps.

To facilitate data analysis, I have written an automated merging program (Sec. 3.3) that improves on an important step in the data analysis process and can improve reconstruction quality considerably. Its current implementation is in the proprietary `IDL` language and it is available through Concurrent Versioning System. In order to make it more accessible for other research groups I suggest to port it to an open source programming language such as Python[1] and to provide pre-compiled binaries for various platforms as downloads on our group's homepage. This also has the advantage that people can write their own extensions, such as support for different raw data file formats.

The merging program is integrated into the general reconstruction process through a common file format that I have described in Sec. 3.1. It is also meant to organize all the files and information that are important to a specific data set in one central place. Finally, it facilitates the exchange of data and reconstruction results between different scientists. With the community ever expanding and the advent of new sources that produce unprecedented amounts of data, the issue of sharing and organizing data cannot be overlooked. I suggest that we actively search the dialogue with members of the community and start a larger collaboration towards a standard file format for diffraction imaging; our working file format can be the starting point for the discussion.

The subroutine library discussed in Sec. 3.2 presents a platform independent implementation of the file format as well as a library of array manipulation

[1] `http://www.python.org/`

routines common to diffraction data analyses. It provides users with a simple interface to write data analysis tools that work on many different computing systems; including multiple node clusters using Message Passing Interface. To further facilitate its use, I suggest to provide a `configure` script that automates the installation process on different computing platforms. Recently, Graphics Processing Units (GPU) have become a very popular and powerful tool in scientific computing. With several hundred cores per graphics card, they provide a highly parallelized workstation with computation speeds far exceeding regular, multiple CPU clusters. A simple C-based library (CUDA[2]) exists to harness the power of GPUs for scientific computation and a successful implementation of CUDA code for diffraction data analysis can be found in Filipe Maia's `hawk`[3] library. I recommend that the `dm_array` library is extended to work on these types of systems by collaborating with Filipe and starting a common code archive.

An important question is how to assess the quality of a reconstruction in a reliable way. In Sec. 3.4, I introduced an improved version of a tool commonly used for this purpose. The Wiener-filtered phase retrieval transfer function removes artifacts that can make its interpretation ambiguous. I also derived some criteria that have to be satisfied for the PRTF to be valid. To make the PRTF even more meaningful I suggest to investigate the actual distribution of phase variations in a reconstruction; this can help to get a better sense of what different values of the PRTF mean. During my research I also found out that the convergence properties of the PRTF depend to some extent on the spatial frequency content of the object. This is something that should be investigated further.

Full-field XDM on frozen-hydrated specimens imposes some hard-to-satisfy experimental constraints, such as the requirement that the object be isolated in the coherent footprint of the beam. Scanning x-ray diffraction microscopy removes the constraint of having an isolated object. To investigate its suitability for imaging frozen-hydrated specimens, I implemented several experimental geometries for our existing diffraction microscope (as described in Sec. 4.1). Based on flux considerations, it was found that using a zone plate as beam defining aperture is the most promising option. However, the apparatus in its current form is ill-equipped for the precise alignment that is required for zone plate imaging. I suggest to look into simple upgrades that facilitate the alignment of optics and sample in the vacuum chamber; this should be followed by another experiment with a zone plate setup. We did collect some promising data sets with the second version of the pinhole setup and I am eager to work

[2] http://www.nvidia.com/object/cuda_home_new.html
[3] http://xray.bmc.uu.se/~filipe/

on reconstructing this data. Finally, from my experiences of the experiment at the APS (described in Sec. 4.2) I know of the advantages of a pixel array detector (PAD) over a regular CCD. Since PADs with good sensitivity in the soft x-ray regime are not yet widely available, I suggest starting a collaboration with pertinent research groups aiming at developing a PAD that meets our specific needs.

Bibliography

[1] S.W. Hell. Far-field optical nanoscopy. *Science*, 316:1153–1158, 2007.

[2] C. Jacobsen, R. Medenwaldt, and S. Williams. A perspective on biological x-ray and electron microscopy. In J. Thieme, G. Schmahl, E. Umbach, and D. Rudolph, editors, *X-ray Microscopy and Spectromicroscopy*, pages II–93–102, Berlin, 1998. Springer-Verlag.

[3] J. Kirz, C. Jacobsen, and M. Howells. Soft x-ray microscopes and their biological applications. *Quarterly Reviews of Biophysics*, 28(1):33–130, 1995. Also available as Lawrence Berkeley Laboratory report LBL-36371.

[4] D. T. Attwood and B. L. Henke, editors. *Low Energy X-ray Diagnostics*, volume 75, New York, 1981. American Institute of Physics. Monterey, 1981.

[5] H. Wiedemann. *Synchrotron Radiation*. Springer, 2002.

[6] M.R. Howells, P. Charalambous, H. He, S. Marchesini, and J.C.H. Spence. An off-axis zone-plate monochromator for high-power undulator radiation. In D.C. Mancini, editor, *Design and Microfabrication of Novel X-ray Optics*, volume 4783, pages 65–73. SPIE, 2002.

[7] E. Lima. *The Advancement of Biological Imaging Through X-ray Diffraction Microscopy*. PhD thesis, Department of Physics and Astronomy, Stony Brook University, 2006.

[8] H. Miao. *Studies on X-ray Diffraction Microscopy*. PhD thesis, Department of Physics and Astronomy, Stony Brook University, 2008.

[9] T. Beetz, M.R. Howells, C. Jacobsen, C.-C. Kao, J. Kirz, E. Lima, T.O. Mentes, H. Miao, C. Sanchez-Hanke, D. Sayre, and D. Shapiro. Apparatus for x-ray diffraction microscopy and tomography of cryo specimens. *Nuclear Instruments and Methods in Physics Research A*, 545(1-2):459–468, 2005.

[10] Johanna Nelson. *X-ray diffraction microscopy on frozen-hydrated specimens*. PhD thesis, Stony Brook University, 2010.

[11] X. Huang. *Cryo Soft X-ray Diffraction Microscopy with Biological Specimens*. PhD thesis, Department of Physics and Astronomy, Stony Brook University, Dec 2009.

[12] M. Born and E. Wolf. *Principles of Optics*. Cambridge University Press, Cambridge, seventh edition, 1999.

[13] M. Born. Quantenmechanik der Stoßvorgänge. *Zeitschrift für Physik*, 38: 803–827, 1926.

[14] J.R. Fienup. Reconstruction of an object from the modulus of its Fourier transform. *Optics Letters*, 3(1):27–29, 1978.

[15] R. W. Gerchberg and W. O. Saxton. A practical algorithm for the determination of phase from image and diffraction plane pictures. *Optik*, 35 (2):237–246, 1972.

[16] J.R. Fienup. Phase retrieval algorithms: a comparison. *Applied Optics*, 21(5):2758–2769, 1982.

[17] V. Elser. Phase retrieval by iterated projections. *Journal of the Optical Society of America A*, 20(1):40–55, 2003.

[18] D. Sayre. Some implications of a theorem due to Shannon. *Acta Crystallographica*, 5:843, 1952.

[19] J. Miao, P. Charalambous, J. Kirz, and D. Sayre. An extension of the methods of x-ray crystallography to allow imaging of micron-size noncrystalline specimens. *Nature*, 400:342–344, 1999.

[20] J. Miao, K.O. Hodgson, T. Ishikawa, C.A. Larabell, M.A. LeGros, and Y. Nishino. Imaging whole *escherichia coli* bacteria by using single-particle x-ray diffraction. *Proceedings of the National Academy of Sciences*, 100:110–112, 2003.

[21] D. Shapiro, P. Thibault, T. Beetz, V. Elser, M. Howells, C. Jacobsen, J. Kirz, E. Lima, H. Miao, A. M. Neiman, and D. Sayre. Biological imaging by soft x-ray diffraction microscopy. *Proceedings of the National Academy of Science*, 102(43):15343–15346, 2005.

[22] X. Huang, J. Nelson, J. Kirz, E. Lima, S. Marchesini, H. Miao, A.M. Neiman, D. Shapiro, J. Steinbrener, A. Stewart, J.J. Turner, and C. Jacobsen. Soft x-ray diffraction microscopy of a frozen hydrated yeast cell. *Physical Review Letters*, 103(19): 198101, Nov 2009. doi: 10.1103/PhysRevLett.103.198101. URL http://link.aps.org/doi/10.1103/PhysRevLett.103.198101.

[23] I.K. Robinson, I.A. Vartanyants, G.J. Williams, M.A. Pfeifer, and J.A. Pitney. Reconstruction of the shapes of gold nanocrystals using coherent x-ray diffraction. *Physical Review Letters*, 8719(19):art. no.–195505, 2001.

[24] H.N. Chapman, A. Barty, S. Marchesini, A. Noy, S. P. Hau-Riege, C. Cui, M.R. Howells, R. Rosen, H. He, J.C.H. Spence, U. Weierstall, T. Beetz, C. Jacobsen, and D. Shapiro. High resolution *ab initio* three-dimensional x-ray diffraction microscopy. *Journal of the Optical Society of America A*, 23:1179–1200, 2006.

[25] Jianwei Miao, Chien-Chun Chen, Changyong Song, Yoshinori Nishino, Yoshiki Kohmura, Tetsuya Ishikawa, Damien Ramunno-Johnson, Ting-Kuo Lee, and Subhash H. Risbud. Three-dimensional GaN-Ga_2O_3 core shell structure revealed by x-ray diffraction microscopy. *Physical Review Letters*, 97(21), 2006. ISSN 0031-9007. doi: {10.1103/PhysRevLett.97.215503}.

[26] A. Barty, S. Marchesini, H. N. Chapman, C. Cui, M. R. Howells, D. A. Shapiro, A. M. Minor, J. C. H. Spence, U. Weierstall, J. Ilavsky, A. Noy, S. P. Hau-Riege, A. B. Artyukhin, T. Baumann, T. Willey, J. Stolken, T. van Buuren, and J. H. Kinney. Three-dimensional coherent x-ray diffraction imaging of a ceramic nanofoam: Determination of structural deformation mechanisms. *Physical Review Letters*, 101(5):055501, 2008. ISSN 0031-9007. doi: {10.1103/PhysRevLett.101.055501}.

[27] Yoshinori Nishino, Yukio Takahashi, Naoko Imamoto, Tetsuya Ishikawa, and Kazuhiro Maeshima. Three-dimensional visualization of a human chromosome using coherent x-ray diffraction. *Physical Review Letters*, 102 (1):018101, 2009. ISSN 0031-9007. doi: 10.1103/PhysRevLett.102.018101.

[28] C.E. Shannon. Communication in the presence of noise. *Proceedings of the Institute of Radio Engineers*, 37(1):10–21, 1949.

[29] R. H. T. Bates. Fourier phase problems are uniquely solvable in more than one dimension. I. Underlying theory. *Optik*, 61:247–262, 1982.

[30] R.A. Crowther, D.J. DeRosier, and A. Klug. The reconstruction of a three-dimensional structure from projections and its application to electron microscopy. *Proceedings of the Royal Society of London A*, 317: 319–340, 1970.

[31] J. Miao, D. Sayre, and H. N. Chapman. Phase retrieval from the magnitude of the Fourier transforms of non-periodic objects. *Journal of the Optical Society of America A*, 15(6):1662–1669, 1998.

[32] J. C. H. Spence, U. Weierstall, and M. Howells. Coherence and sampling requirements for diffractive imaging. *Ultramicroscopy*, 101:149–152, 2004.

[33] J. W. Goodman. *Introduction to Fourier Optics*. McGraw-Hill, San Francisco, 1968.

[34] J.C.H. Spence, U. Weierstall, and M. Howells. Phase recovery and lensless imaging by iterative methods in optical, x-ray and electron diffraction. *Philosophical Transactions of the Royal Society of London A*, 360:875–895, 2002.

[35] P. Thibault, V. Elser, C. Jacobsen, D. Shapiro, and D. Sayre. Reconstruction of a yeast cell from x-ray diffraction data. *Acta Crystallographica A*, 62:248–261, 2006.

[36] M.R. Howells, T. Beetz, H.N. Chapman, C. Cui, J.M. Holton, C.J. Jacobsen, J. Kirz, E. Lima, S. Marchesini, H. Miao, D. Sayre, D.A. Shapiro, J.C.H. Spence, and D. Starodub. An assessment of the resolution limitation due to radiation-damage in x-ray diffraction microscopy. *Journal of Electron Spectroscopy and Related Phenomena*, 170:4–12, 2009.

[37] Q. Shen, I. Bazarov, and P. Thibault. Diffractive imaging of nonperiodic materials with future coherent x-ray sources. *Journal of Synchrotron Radiation*, 11:432–438, 2004.

[38] W. Vernon, M. Allin, R. Hamlin, T. Hontz, D. Nguyen, F. Augustine, S.M. Gruner, Ng.H. Xuong, D.R. Schuette, M.W. Tate, and L.J. Koerner. First results from the 128×128 pixel mixed-mode Si x-ray detector chip. In *SPIE proceedings 6703*, volume 6703. SPIE, 2007.

[39] Ch. Broennimann, E. F. Eikenberry, B. Heinrich, R. Horisberger, G. Huelsen, E. Pohl, B. Schmitt, C. Schulze-Briese, M. Suzuki, T. Tomizaki, H. Toyokawa, and A. Wagner. The pilatus 1m detector. *Journal of Synchrotron Radiation*, 13:120–130, 2006.

[40] S. Marchesini, H. He, H.N. Chapman, S.P. Hau-Riege, A. Noy, M.R. Howells, U. Weierstall, and J.C.H. Spence. X-ray image reconstruction from a diffraction pattern alone. *Physical Review B*, 68(14):140101, 2003.

[41] S. Marchesini. Invited article: A unied evaluation of iterative projection algorithms for phase retrieval. *Review of Scientific Instruments*, 78: 011301–1–10, 2007.

[42] R. Hegerl and W. Hoppe. Dynamische theorie der kristallstrukturanalyse durch elektronenbeugung im inhomogenen primärstrahlwellenfeld. *Berichte der Bunsengesellschaft*, 74(11):1148–1154, 1970.

[43] R. H. T. Bates and J. M. Rodenburg. Sub-angstrom transmission microscopy: A fourier-transform algorithm for microdiffraction plane intensity information. *Ultramicroscopy*, 31:303–308, 1989.

[44] B. C. McCallum and J. M. Rodenburg. Two-dimensional demonstration of Wigner phase-retrieval microscopy in the STEM configuration. *Ultramicroscopy*, 45:371–380, 1992.

[45] J.M. Rodenburg, B.C. McCallum, and P.D. Nellist. Experimental tests on double-resolution coherent imaging via STEM. *Ultramicroscopy*, 48: 304–314, 1993.

[46] H. N. Chapman. Phase-retrieval x-ray microscopy by Wigner-distribution deconvolution. *Ultramicroscopy*, 66:153–172, 1996.

[47] J. M. Rodenburg and H. M. L. Faulkner. A phase retrieval algorithm for shifting illumination. *Applied Physics Letters*, 85:4795–4797, 2004.

[48] H. M. Quiney, A. G. Peele, Z. Cai, D. Paterson, and K. A. Nugent. Diffractive imaging of highly focused x-ray elds. *Nature Physics*, 2:101–104, 2006.

[49] P. Thibault, M. Dierolf, A. Menzel, O. Bunk, C. David, and F. Pfeiffer. High-resolution scanning x-ray diffraction microscopy. *Science*, 321:379–382, 2008.

[50] A. M. Maiden and J. M. Rodenburg. An improved ptychographical phase retrieval algorithm for diffractive imaging. *Ultramicroscopy*, 109:1256–1262, 2009.

[51] O. Bunk, M. Dierolf, S. Kynde, I. Johnson, O. Marti, and F. Pfeiffer. Inuence of the overlap parameter on the convergence of the ptychographical iterative engine. *Ultramicroscopy*, 108:481–487, 2008.

[52] M. Guizar-Sicairos and J. R. Fienup. Phase retrieval with transverse translation diversity: a nonlinear optimization approach. *Optics Express*, 16:7264–7278, 2008.

[53] K. A. Nugent, A. G. Peele, H. M. Quiney, and H. N. Chapman. Diffraction with wavefront curvature: a path to unique phase recovery. *Acta Crystallographica A*, 61:373–381, 2005.

[54] G. J. Williams, H. M. Quiney, B. B. Dhal, C. Q. Tran, K. A. Nugent, A. G. Peele, D. Paterson, and M. D. de Jonge. Fresnel coherent diffractive imaging. *Physical Review Letters*, 97(2):025506, 2006. ISSN 0031-9007. doi: {10.1103/PhysRevLett.97.025506}.

[55] B. Abbey, K.A. Nugent, G.J. Williams, J.N. Clark, A.G. Peele, M.A. Pfeifer, M. de Jonge, and I. McNulty. Keyhole coherent diffractive imaging. *Nature Physics*, 4(5):394–398, 2008. ISSN 1745-2473. doi: {10.1038/nphys896}.

[56] R. A. London, M. D. Rosen, and J. E. Trebes. Wavelength choice for soft x-ray laser holography of biological samples. *Applied Optics*, 28:3397–3404, 1989.

[57] Y. Talmon. Electron beam radiation damage to organic and biological cryospecimens. In R. A. Steinbrecht and K. Zierold, editors, *Cryotechniques in Biological Electron Microscopy*, pages 64–84, Berlin, 1987. Springer-Verlag.

[58] D. Shapiro. *Biological imaging by soft x-ray diffraction microscopy*. PhD thesis, Department of Physics and Astronomy, Stony Brook University, 2004.

[59] A. Rose. Unified approach to performance of photographic film, television pickup tubes, and human eye. *Journal of the Society of Motion Picture Engineers*, 47:273–294, 1946.

[60] D. Sayre, J. Kirz, R. Feder, D. M. Kim, and E. Spiller. Transmission microscopy of unmodified biological materials: Comparative radiation dosages with electrons and ultrasoft x-ray photons. *Ultramicroscopy*, 2: 337–341, 1977.

[61] J. Nelson, X. Huang, J. Steinbrener, D. Shapiro, J. Kirz, S. Marchesini, A. M. Neiman, J. J. Turner, and C. Jacobsen. High-resolution x-ray diffraction microscopy of specifically labeled yeast cells. *Proceedings of*

the National Academy of Sciences, 107(16):7235–7239, 2010. doi: 10.1073/pnas.0910874107.

[62] K. Giewekemeyer, P. Thibault, S. Kalbfleisch, A. Beerlink, C. M. Kewish, M. Dierolf, F. Pfeiffer, and T. Salditt. Quantitative biological imaging by ptychographic x-ray diffraction microscopy. *Proceedings of the National Academy of Sciences*, 107(2):529–534, 2010.

[63] A. G. Michette. *Optical Systems for Soft X Rays*. Plenum, New York, 1986.

[64] W. Chao, J. Kim, S. Rekawa, P. Fischer, and E.H. Anderson. Demonstration of 12 nm resolution Fresnel zone plate lens based soft x-ray microscopy. *Optics Express*, 17(20):17669–17677, 2009.

[65] X. Huang, H. Miao, J. Steinbrener, J. Nelson, D. Shapiro, A. Stewart, J. Turner, and C. Jacobsen. Signal-to-noise and radiation exposure considerations in conventional and diffraction x-ray microscopy. *Optics Express*, 17(16):13541–13553, Jul 2009. doi: 10.1364/OE.17.013541.

[66] *MPI: A Message-Passing Interface Standard*, 1.1 edition, June 1995. URL http://www.mpi-forum.org/docs/mpi-11-html/mpi-report.html.

[67] *MPI-2: Extensions to the Message-Passing Interface*, 1997. URL http://www.mpi-forum.org/docs/mpi-20-html/mpi2-report.html.

[68] R. Crandall, E. Jones, J. Klivington, and D. Kramer. Gigaelement FFTs on Apple G5 clusters. Technical report, Advanced Computation Group, Apple Computer, 2004. URL http://images.apple.com/acg/pdf/20040827_GigaFFT.pdf.

[69] J. J. Turner, J. Nelson, X. Huang, J. Steinbrener, D. Shapiro, and C. Jacobsen. Rapid 'Lensless' Imaging of Nanoporous Glass with Soft X-rays. *Manuscript in preparation*.

[70] J. Steinbrener, J. Nelson, X. Huang, S. Marchesini, D. Shapiro, J. J. Turner, and C. Jacobsen. High Fidelity Reconstructions in X-ray Diffraction Microscopy. *Submitted*, 2010.

[71] J. Miao, T. Ishikawa, B. Johnson, E.H. Anderson, B. Lai, and K.O. Hodgson. High resolution 3D x-ray diffraction microscopy. *Physical Review Letters*, 89(8):088303, 2002.

[72] Chien-Chun Chen, Jianwei Miao, C. W. Wang, and T. K. Lee. Application of optimization technique to noncrystalline x-ray diffraction microscopy: Guided hybrid input-output method. *Physical Review B*, 76(6), 2007. ISSN 1098-0121. doi: {10.1103/PhysRevB.76.064113}.

[73] C. G. Schroer, R. Boye, J. M. Feldkamp, J. Patommel, A. Schropp, A. Schwab, S. Stephan, M. Burghammer, S. Schoeder, and C. Riekel. Coherent x-ray diffraction imaging with nanofocused illumination. *Physical Review Letters*, 101(9):090801, 2008. ISSN 0031-9007. doi: 10.1103/PhysRevLett.101.090801.

[74] A. E. Conrady. *Applied Optics and Optical Design*. Dover Publications, Inc., New York, 1957.

[75] N. Wiener. *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. The MIT Press, 1964.

[76] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C*. Cambridge University Press, 2nd edition, 1992.

[77] C. Jacobsen, S. Williams, E. Anderson, M. T. Browne, C. J. Buckley, D. Kern, J. Kirz, M. Rivers, and X. Zhang. Diffraction-limited imaging in a scanning transmission x-ray microscope. *Optics Communications*, 86: 351–364, 1991.

[78] B. Hornberger, M. Feser, and C. Jacobsen. Quantitative amplitude and phase contrast imaging in a scanning transmission X-ray microscope. *Ultramicroscopy*, 107:644–655, 2007. doi: 10.1016/j.ultramic.2006.12.006.

[79] B. L. Henke, E. M. Gullikson, and J. C. Davis. X-ray interactions: Photoabsorption, scattering, transmission, and reflection at E=50–30,000 eV, Z=1–92. *Atomic Data and Nuclear Data Tables*, 54:181–342, 1993.

[80] M. H. J. Koch, M. C. Vega, Z. Sayers, and A. M. Michon. The superstructure of chromatin and its condensation mechanism. *European Biophysics Journal*, 14:307–319, 1987.

[81] H. Wolter. Spiegelsysteme streifenden Einfalls als abbildende Optiken für Röntgenstrahlen. *Annalen der Physik*, 10:94–114, 286, 1952.

[82] J. M. Cowley and A. F. Moodie. Fourier images. I. The point source. *Proceedings of the Physical Society B*, 70(5):486–496, 1957.

[83] G.J. Williams, H.M. Quiney, A.G. Peele, and K.A. Nugent. Coherent diffractive imaging and partial coherence. *Physical Review B*, 75 (10):104102, MAR 2007. ISSN 1098-0121. doi: {10.1103/PhysRevB.75.104102}.

[84] R. J. Collier, C. B. Burckhardt, and L. H. Lin. *Optical Holography*. Academic Press, New York, 1971.

Appendix A

Detailed Definition of the File Format

A.1 The "/comments"-group

The /comments-group holds important information on the history of a particular data file. Within the /comments-group you will find:

> string_length tells the maximum length of each string. (We use fixed string lengths because HDF 5 variable length strings are not supported by IDL 6.2).
>
> n_strings tells how many strings are in the array.
>
> specimen_name is a string for a short description of the specimen.
>
> collection_date is a string that is meant to hold the C systime() string from when the original data was recorded.
>
> string_array is the array of strings, each string of which is no longer than string_length-1 characters in length.

A.2 The "/ainfo"-group

The /ainfo-group stores important information about the files that were used for the assembly of an adi-array. Within it you will find:

> The attribute ainfo_version is used to indicate to the dm_fileio routines any changes in the structure of information in the ainfo_group.

file_directory contains the path to the folder where all the files that were used for the assembly are stored.

filename_array contains an array of filenames of all the files used for the assembly.

systime_array stores the creation date and time for every file that was used for the assembly.

theta_x_radians_array: GMR position for each file .

xcenter_offset_pixels_array: number of pixels the center of the diffraction pattern is offset from the center of the array in x.

ycenter_offset_pixels_array: see above.

n_frames stores the number of files that were used for the assembly. This value is entered by the user and has to be checked carefully since it determines the size of the above arrays.

string_length: the minimum string length required to store the longest filename, systime, or file directory. Automatically determined in the IDL implementation. For C the user has to define a keyword.

no_background: either 1 or 0 depending on whether background has been subtracted during assembly or not

dk_by_pix: either 1 or 0 depending on whether dark current was subtracted on a per pixel basis or as average.

merge_first: either 1 or 0 depending on whether sample and background have been merged independently before subtraction or whether background was subtracted for each position separately.

A.3 The "/adi"-group

The /adi-group holds the array of assembled diffraction intensities and supporting information. Within the /adi-group, you will find:

The attribute adi_version is used to indicate to the dm_fileio routines any changes in the structure of information in the "/adi" group.

adi_array contains the actual data.

`adi_struct` contains a structure of parameters pertaining to the data.

`adi_error_array` is an optional array as described in Sec. 3.1.4.

The `"/adi"` group contains both the data, and a structure `adi_struct` of related parameters. The elements of this structure are as follows:

`photon_scaling`: multiply `adi_array` by this factor to convert it into photons per pixel. Set to zero if unknown.

`error_scaling`: optional scaling factor s_e to apply to `error_array`.

`lambda_meters`: x-ray wavelength in meters.

`camera_alpha/beta/gamma_radians`: optional rotation angles of the CCD camera in radians.

`median_filter_width`: width of median filter used to search for abnormally high pixel values in percent.

`median_filter_threshold`: threshold value in percent of the maximum of the median filtered array to determine outlier pixels.

`camera_z_meters`: CCD camera distance from sample.

`camera_x_pixelsize_meters`: size of camera pixels in the `nx` direction, in meters.

`camera_y_pixelsize_meters`: size of camera pixels in the `nx` direction, in meters.

`saturation_min`: minimum level of CCD counts below which we don't assume saturation effects to occur.

`saturation_max`: maximum level of CCD counts above which we don't assume saturation effects to occur.

`theta_x_radians`: GMR position (rotation about the horizontal transverse direction) to which the sample was set to be at. *Obsolete since already defined in* `"/ainfo"` *group!*

`theta_y_radians`: orientation of the specimen about the vertical transverse direction. Set to zero if unknown. *Should be moved to* `"/ainfo"` *group!*

`theta_z_radians`: orientation of the specimen about the beam axis direction. Set to zero if unknown. *Should be moved to* `"/ainfo"` *group!*

`xcenter_offset_pixels`: location of the true center of the array relative to pixel $n_x/2$. *Obsolete since already defined in* `"/ainfo"` *group!*

`ycenter_offset_pixels`: location of the true center of the array relative to pixel $n_y/2$. *Obsolete since already defined in* `"/ainfo"` *group!*

A.4 The "/spt"-group

The `/spt`-group holds the current support guess array as well as supporting information. Within the `/spt`-group you will find:

The attribute `spt_version` is used to indicate to the `dm_fileio` routines any changes in the structure of information in the `"/spt"` group.

`spt_array` contains the actual mask. This is usually a byte image.

`spt_struct` contains a structure of parameters pertaining to the mask.

The parameters contained in the associated `spt_struct` are:

`support_scaling` (s_s): multiply `spt_array` by this factor (or $s_s \cdot s_{i,j,k}$) to give a value which is 0 for fully outside the support, and 1 for fully inside the support. This allows one to use, for example a gray-scale byte support array by setting $s_s = 1/255$.

`pix_x_meters`: real space pixel size in the first dimension.

`pix_y_meters`: real space pixel size in the second dimension.

`pix_z_meters`: real space pixel size in the third dimension. The value will be ignored if the mask is 2D.

A.5 The "/itn"-group

The `/itn`-group holds the current complex iterate, an array of reconstruction errors, as well as supporting information. Within the `/itn`-group you will find:

The attribute `itn_version` is used to indicate to the `dm_fileio` routines any changes in the structure of information in the `"/itn"` group.

`itn_array` contains the actual complex iterate. This array will have C array ordering [`nz`,`ny`, `nx`,2] so that the index between interleaved real and imaginary parts will vary most rapidly.

`recon_errors`: an array that saves the current value of the error metric for each iteration.

`itn_struct` contains a structure of parameters pertaining to the mask.

The `itn_struct` information associated with these files is:

`pix_x_meters`: real space pixel size in the first dimension.

`pix_y_meters`: real space pixel size in the second dimension.

`pix_z_meters`: real space pixel size in the third dimension. This value must be present but is ignored for 2D arrays.

`photon_scaling` (s_p): multiply `itn_array` by this factor to convert it into photons per pixel. Set to zero if unknown.

`iterate_count` (i; uint32): counter of iteration number.

Appendix B

Details of the SXDM Setup

B.1 Design of the pinhole setup

This shows the design of the close scanning pinhole setup. Its design parameters have been described in Sec. 4.1.2.

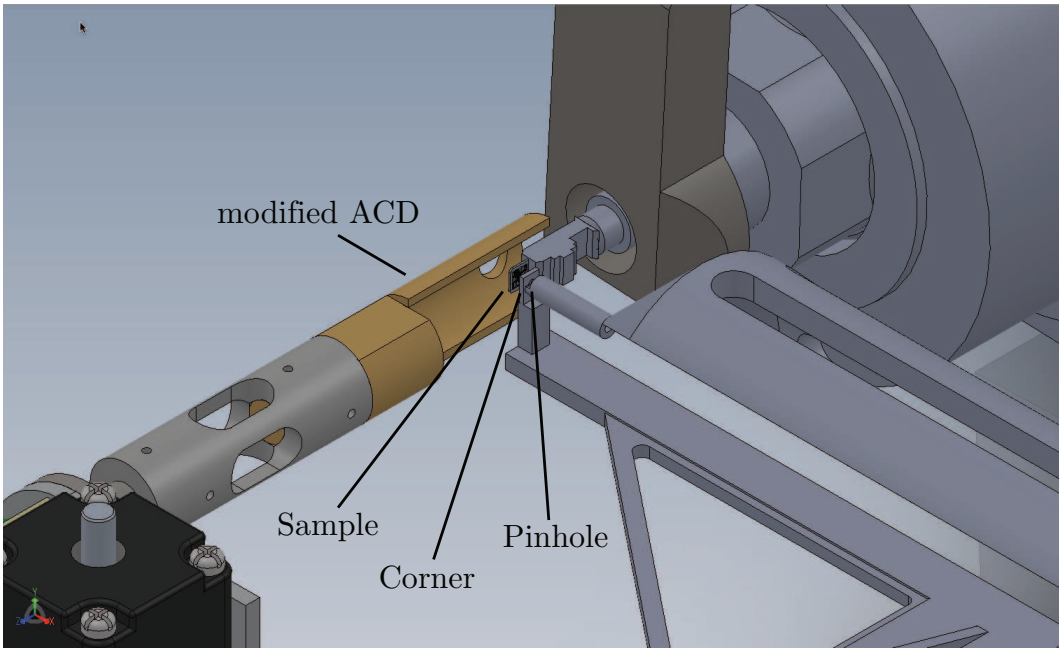

Figure B.1: 3D rendering of pinhole setup. To accommodate both pinhole and corner very close to the sample, the ACD shield had to be modified.