UNIVERSITY OF CALIFORNIA, SAN DIEGO

Dichroic Coherent Diffractive Imaging

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Physics

by

Ashish Tripathi

Committee in charge:

Oleg Shpyrko, Chair
Sunil Sinha, Co-Chair
Daniel Arovas
Eric Fullerton
Zhaowei Liu
Ian McNulty

2012

The dissertation of Ashish Tripathi is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____
Co-Chair

_____
Chair

University of California, San Diego

2012

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

Use of the Advanced Photon Source is supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract DE-AC02-06CH11357. Work at the University of California, San Diego, was supported by US Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract DE-SC0001805.

Part of the text of Chapter 3 is a preprint of material being prepared for submission for publication:
• A. Tripathi, I. McNulty, O. G. Shpyrko, "Ptychographic overlap constraint errors and diffraction limited imaging".

The text of Chapter 5, in full, is a reprint of the material as it appears in:
• A. Tripathi, J. Mohanty, S. Dietze, O. G. Shpyrko, E. Shipton, E. Fullerton, S.S. Kim and I. McNulty, "Dichroic Coherent Diffractive Imaging", *Proc. Natl. Acad. Sci. USA*, **108**, 13393-8 (2011).

VITA

2005	B. S. in Electrical Engineering, University of Illinois, Urbana
2008	M.S. in Physics University of California, San Diego
2012	Ph. D. in Physics, University of California, San Diego

PUBLICATIONS

A. Tripathi, J. Mohanty, S. Dietze, O. G. Shpyrko, E. Shipton, E. Fullerton, S.S. Kim and I. McNulty, "Dichroic Coherent Diffractive Imaging", Proc. Natl. Acad. Sci. USA, **108**, 13393-8 (2011)

A. Tripathi, O. G. Shpyrko, and I. McNulty, "Influence of Noise and Missing Data on Reconstruction Quality in Coherent X-ray Diffractive Imaging. AIP Conf. Proc. **1365**, 305 (2011).

S. O. Hruszkewycz, M. V. Holt, A. Tripathi, J. Maser, and P. H. Fuoss, "Framework for three- dimensional coherent diffraction imaging by focused beam x-ray Bragg ptychography. Opt. Lett. **36**, 2227 (2011).

M. M. Qazilbash, A. Tripathi, A. A. Schafgans, Bong-Jun Kim, Hyun-Tak Kim, Zhonghou Cai, M. V. Holt, J. M. Maser, F. Keilmann, O. G. Shpyrko, and D. N. Basov, "Nanoscale imaging of the electronic and structural transitions in vanadium dioxide" Phys. Rev. B. **83**, 165108 (2011).

ABSTRACT OF THE DISSERTATION

Dichroic Coherent Diffractive Imaging

by

Ashish Tripathi

Doctor of Philosophy in Physics

University of California, San Diego, 2012

Oleg Shpyrko, Chair
Sunil Sinha, Co-Chair

Understanding electronic structure at nanometer resolution is crucial to understanding physics such as phase separation and emergent behavior in correlated electron materials. Nondestructive probes which have the ability to see beyond surfaces on nanometer length and sub-picosecond time scales can greatly enhance our understanding of these systems and will impact development of future technologies, such as magnetic storage. Polarized x-rays are an appealing choice of probe due to their penetrating power, elemental and magnetic specificity, and high spatial resolution. The resolution of traditional x-ray microscopy is limited by the nanometer precision required to fabricate x-ray optics. In this thesis, a novel approach to lensless imaging of an extended magnetic nanostructure is presented. We

demonstrate this approach by imaging ferrimagnetic "maze" domains in a Gd/Fe multilayer with perpendicular anisotropy. A series of dichroic coherent diffraction patterns, ptychographically recorded, are numerically inverted using non-convex and non-linear optimization theory, and we follow the magnetic domain configuration evolution through part of its magnetization hysteresis loop by applying an external magnetic field. Unlike holographic methods, it does not require a reference wave or precision optics, and so is a far simpler experiment. In addition, it enables the imaging of samples with arbitrarily large spatial dimensions, at a spatial resolution limited solely by the coherent x-ray flux and wavelength. It can readily be extended to other non-magnetic systems that exhibit circular or linear dichroism. This approach is scalable to imaging with diffraction-limited resolution, a prospect rapidly becoming a reality in view of the new generation of phenomenally brilliant x-ray sources.

Chapter 1

Introduction

In this chapter we introduce concepts behind x-ray microscopy carried out at brilliant light sources known as synchrotrons. We begin with some traditional and typical types of x-ray microscopy, here the Transmission X-ray Microscope (TXM) and the Scanning Transmission X-ray Microscope (STXM). These microscopes offer nanoscale resolution which is primarily limited by the quality of x-ray optics used. We next go into a newer type of x-ray microscopy still under development known as the coherent x-ray diffraction microscope. This type of microscope replaces the x-ray optics used in TXM and STXM with "numerical optics", i.e. it is primarily a numerical procedure which takes measured diffraction intensity from an object and numerically inverts the diffraction to recover a high resolution image of the object. The spatial resolution is limited by factors which will be discussed. We next discuss analytical and numerical tools which allow us to interpret and reliably work with numerically the diffraction intensities measured with an area detector.

1.1 Synchrotron based x-ray microscopy

This section introduces and explores some examples of conventional x-ray microscopes which are widespread at synchrotron x-ray sources. Some very common examples are the Transmission X-ray Microscope (TXM) and the Scanning Transmission X-ray Microscope (STXM). These will be explored in section 1.1.1.

The primary experimental difficulties and limitations involved with these techniques will also be discussed. Section 1.1.2 will introduce an alternative to lens based microscopy known as Coherent X-ray Diffractive Imaging (CXDI), which is a lensless imaging technique, and is the primary topic of this thesis. A prominent feature of this technique is that the spatial resolution it is capable of is not determined by how well made (at nanometer scale precision) various optical elements are fabricated, as is the case in traditional microscopy, but rather by the quality of the diffraction measurement, e.g. the signal to noise of the diffraction. This in turn means that the spatial resolution is in practice determined by the quality of the x-ray area detector and in principle diffraction limited, i.e. close to the x-ray wavelength.

1.1.1 Traditional x-ray microscopy

Consider the two x-ray microscopes shown in Fig. 1.1. In TXM, x-rays first are monochromatized and focused down to a spot size on the sample of several microns by a condenser lens and then scatter from the sample. This scattered signal is then collected by an objective lens, which acts as an "inverse Fourier transform" when placed at an appropriate position downstream of the sample, and which projects and magnifies it onto a two dimensional pixel area detector. The TXM is essentially a direct analogue of visible light microscopes using condensing and objective opitcs, the main differences being the types of optical elements used. The Scanning Transmission X-ray Microscope (STXM) on the other hand uses a condenser optic to create a small spot size on the sample. The scattered signal from the small spot size is recorded using an integrating detector while the sample is raster scanned, and an image of the object is formed by raster scanning many locations on the sample. In both of these cases, the spatial resolution is determined by the quality of the focusing lenses; in TXM it is determined by how aberration free the objective lens is while in STXM it is determined by how small the focal spot size is.

Visible light microscopes use primarily refractive optics since visible light interacts strongly with matter. X-rays must generally use something else, e.g.

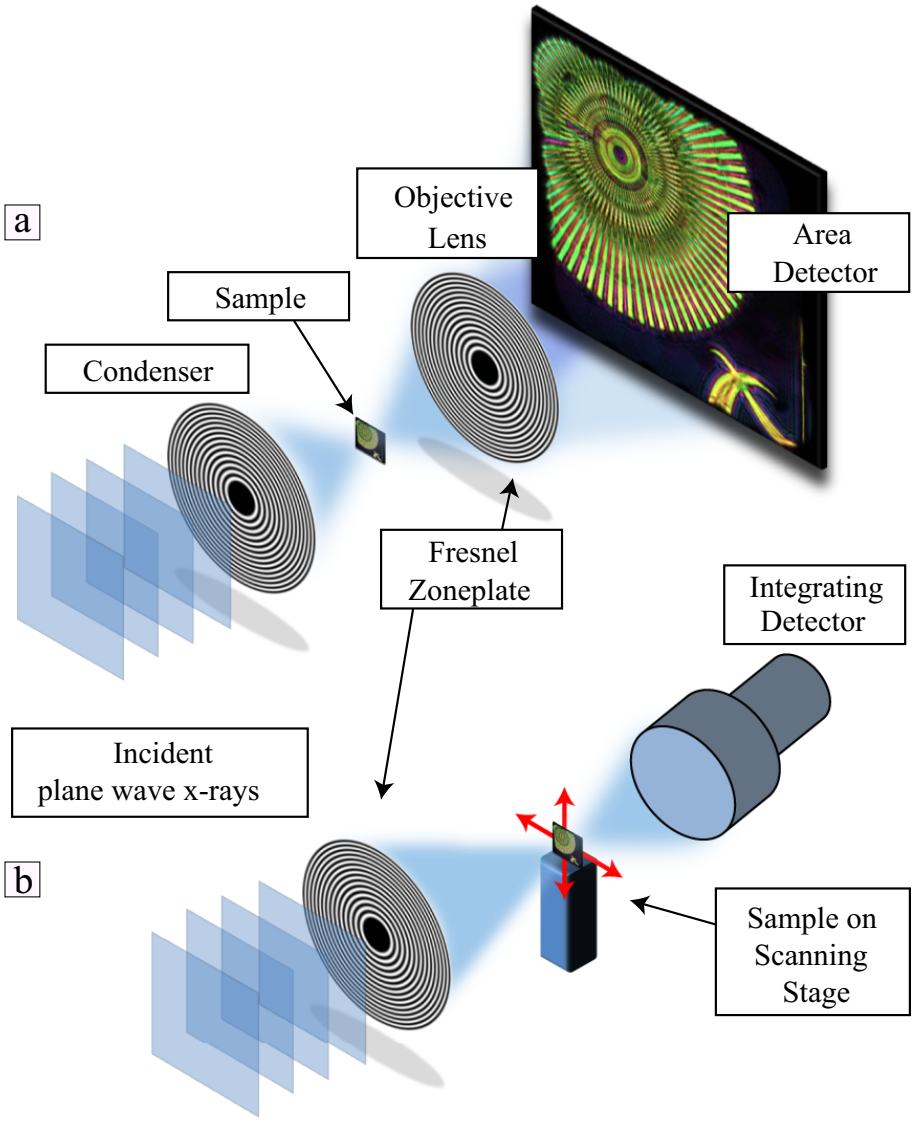

Figure 1.1: (a) The Transmission X-ray Microscope. X-rays are condensed onto the sample, scatter, and an objective lens is placed downsteam of the sample so that the "imaging condition" is satisfied and so the objective lens acts as an inverse Fourier transform of the scattered signal. An area detector is placed further downstream of the objective so that a magnified image of the sample is measured. (b) The Scanning Transmission X-ray Microscope. It uses a condenser optic to create a nanometer scale spot size on the sample, and the sample is raster scanned so that each scan location makes up a pixel of the image of the object after many locations on the sample are scanned.

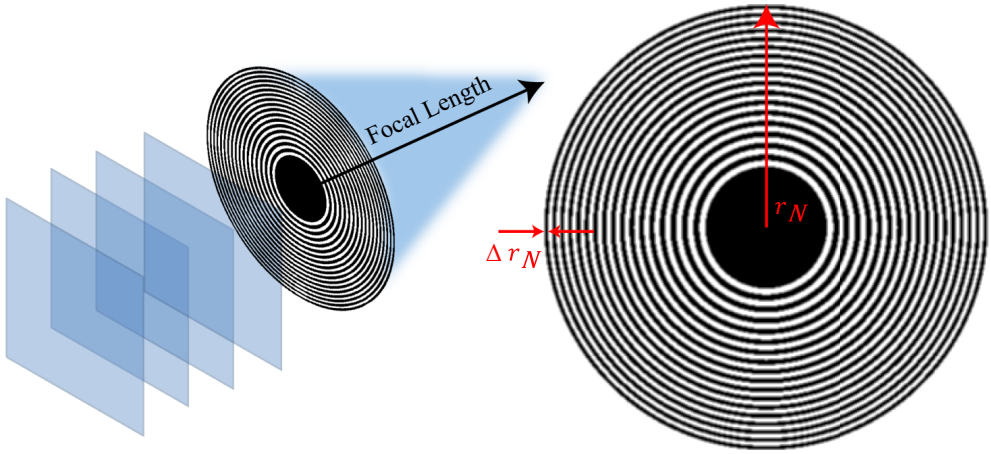

Figure 1.2: The Fresnel zoneplate, a popular diffractive optic used in x-ray microscopy. It is made by fabricating radially symmetric regions which alternate between opaque and transparent. X-rays incident on the zone plate will diffract around the opaque zones and the zones can be spaced so that the diffracted light constructively interferes at a desired focus.

diffractive optics, since x-ray refraction is generally very weak; recall the x-ray index of refraction is $n = 1 - \delta + i\beta = 1 - \Delta n$, where $\Delta n \sim (1 - i)10^{-5}$ or $(1 - i)10^{-6}$. A popular diffractive lens used in x-ray microscopy is the Fresnel zone plate, which is made by fabricating radially symmetric regions which alternate between opaque and transparent. X-rays incident on the zone plate will diffract around the opaque zones and the zones can be spaced so that the diffracted light constructively interferes at a desired focus; a cartoon of this is shown in Fig. 1.2. The focal length of a Fresnel zone plate is calculated using [1] the relation:

$$f = \frac{2r_N \Delta r_N}{\lambda}, \tag{1.1}$$

where Δr_N is the width of the outermost zone, r_N is the radius at which this outermost zone is located, and λ is the x-ray wavelength. The spot size at which the Fresnel zone plate can focus x-rays down to is given by:

$$\Delta s \approx 1.22\ \Delta r_N, \tag{1.2}$$

Fresnel zone plates are usually fabricated using a lithographic processes. The smallest size of features that can be fabricated is dependent on the sophistication level of the lithography technology, with the state of the art currently in the $\Delta r_N \sim$

10 nm range [2]. The prime difficulty the fabrication process encounters is that the zone plate needs to be sufficiently thick (in the direction the x-rays propagate) to provide acceptable diffraction efficiency (number of scattered photons which contribute to focusing); zone plate efficiencies are typically around 10% for the first diffraction order. This results in high aspect ratio nanostructures [3]; 3:1 or 4:1 are typical for the highest-resolution zone plates, but greatly limits the diffraction efficiency. The high aspect ratios also results in generally very fragile optics which can be easily damaged. It is possible for the focusing properties of Fresnel zone plates to degrade over time simply by being used; the x-rays themselves cause damage to the outermost zone and degrade the achievable spot size.

1.1.2 Lensless imaging

Rather than using a microscope which depends on time consuming and expensive x-ray optics fabrication, we can instead use a technique which does not use optics to image the sample but uses computational methods to synthesize real space information from diffraction patterns obtained by scattering a coherent x-ray beam from a sample. This is termed a "lensless" microscopy, or alternatively Coherent X-ray Diffractive Imaging (CXDI). Some typical experimental arrangements are shown in Fig. 1.3. In this way of doing things, the diffraction is usually considered "far-field" diffraction, or Fraunhofer diffraction, and so is related to the illuminated region on the sample by a simple Fourier transform. If we lived in Wonderland and had an area detector which in addition to measuring the complex valued wavefield modulus (the square root of the intensity) can also measure the *phase* of the wavefield, then we can get a high resolution image of the object by simply taking the inverse Fourier transform of the measured complex valued wavefield. However, since this detector is make believe and we can only measure the intensity, we must find another way to retrieve the wavefield phase information.

Fourier transform holography [4, 5] is one way of doing this phase retrieval: fabricate a sample and reference wave pinhole so that diffraction from the sample coherently interferes with the reference wave, and we will have phase information encoded in the measurement diffraction intensity. Indeed this method is a very

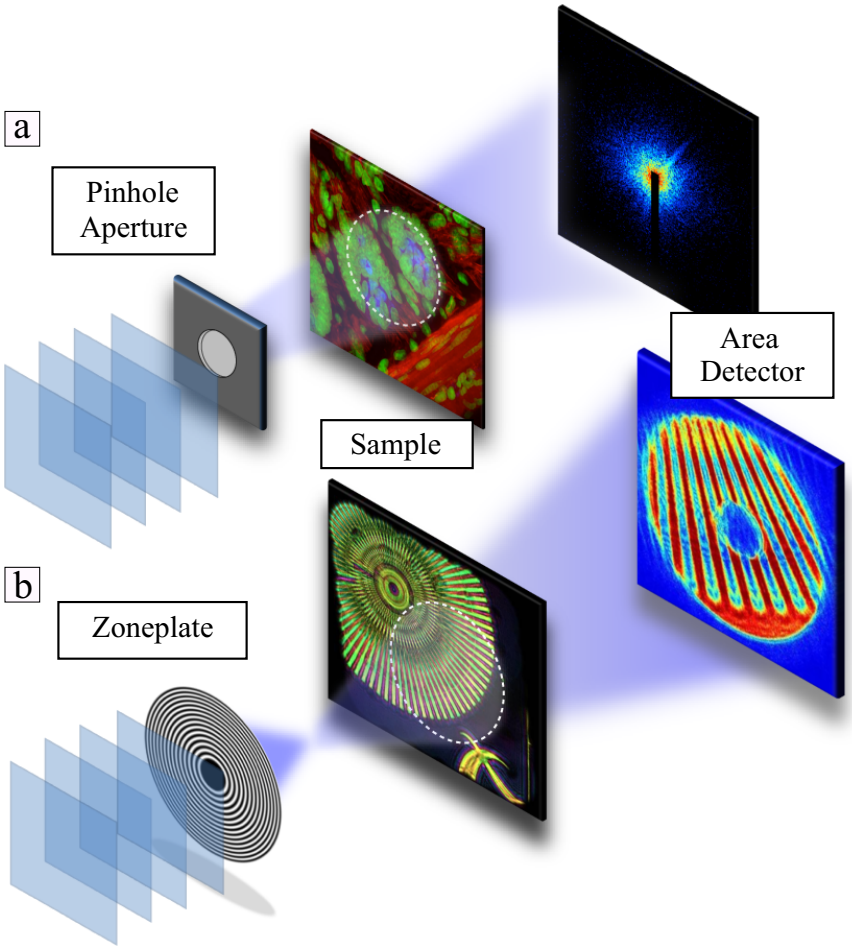

Figure 1.3: Some typical Coherent X-ray Diffractive Imaging (CXDI) experimental arrangements. (a-b) X-rays are incident on a pinhole aperture or a condenser optic, whose purpose is to select out a coherent volume of the incident partially coherent x-ray wavefield and to define the fully coherent x-ray illumination incident on the sample. An area detector is placed in the far field so that the Fraunhofer diffraction pattern is measured. Use of a condenser optic in (b) gives the x-ray wavefield known phase structure when the sample is placed upstream some distance of the focal plane, and will be explored later in future chapters.

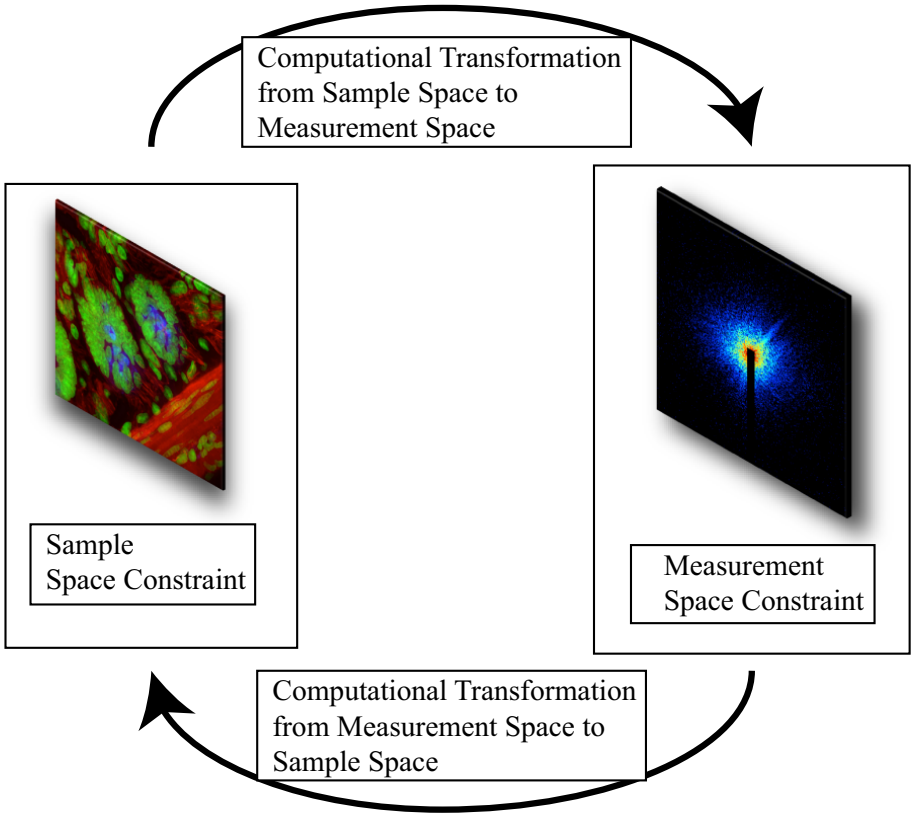

Figure 1.4: General "idea" behind iterative algorithms encountered in CXDI and this thesis. Starting from some initial guess for the sample, we iterate between sample space and measurement space computationally while enforcing *constraints* on the process in the hope of refining the initial guess so that it will eventually satisfy all constraints simultaneously, meaning we have likely recovered the phases of the wavefield lost upon measurement.

powerful lensless imaging technique. But issues come into play when attempting to push towards diffraction limited imaging: the reference wave pinhole must be fabricated, characterized, and known to nanoscale precision comparable to the desired spatial resolution. So we are back to the time consuming and experimentally difficult problem of fabricating a sample and reference pinhole composite object which must be made at nanoscale precision.

In CXDI, indirect phase retrieval is accomplished by the use of iterative algorithms derived from non-convex and non-linear optimization methods which can extract phase information from diffraction modulus information [6, 7, 8, 9, 10, 11, 12, 13, 14, 15]. In these algorithms, which will be covered in somewhat grue-

some detail in Chapter 2, we identify and computationally formulate *constraints* which are used to discover the phase of the measured wavefield. These constraints can be divided into two camps: sample space constraints and measurement space constraints. In Chapter 2, we will do some identification and formulation of sample space constraints; measurement space constraints are always the measured diffraction modulus. A typical flowchart of how CXDI iterative algorithms work is shown in Fig. 1.4. Here, starting from some initial guess ψ_0 for the sample we define a transformation which takes us from sample space to measurement space. Since the diffraction we measure is in the far-field, this transformation is the Fourier transform, or a close relative when using a condenser optics to define the x-ray illumination on the sample [16, 17]. So we take the Fourier transform of the guess sample $\Psi_0 = \mathcal{F}\{\psi_0\}$, and then enforce the measurement space constraint, the diffraction measurement. This is done by throwing away the modulus of Ψ_0 and replacing it by the diffraction measurement, but leaving the phase of Ψ_0 as it is. Once this is done we transform back to sample space, and apply any sample space constraints we have formulated. We then repeat this back and forth process of transforming to one space, apply a constraint, transform to another space, apply a constraint, etc., until a "solution" is found, i.e. we simultaneously satisfy all constraints. When the appropriately defined constraints are all simultaneously satisfied, it is exceedingly likely that we have successfully recovered the phase of the wavefield at the detector lost upon measurement [11].

1.2 Wavefield propagation

In this section, we mainly define analytical tools which will be used throughout this thesis. We start by defining the Fourier transform conventions used, for both continuous and discrete types. Also, as we will be dealing extensively with complex valued images when performing wavefield propagation, we present an effective and compact convention for rendering these complex valued images. Then, the mathematical formalism of x-ray diffraction is introduced by deriving the Helmholtz equation. Then various approximations are defined which make inter-

preting what exactly we are scattering x-rays from when performing experiments in the geometries shown in Fig. 1.3 computationally feasible. We also explore some analytical details and potential pitfalls of wavefield propagation, which are the transformations mentioned in the previous section required to do phase retrieval.

1.2.1 Fourier transform properties

We begin by defining the two dimensional spatial Fourier transform:

$$f(\mathbf{r}) = \mathcal{F}^{-1}[\tilde{f}(\mathbf{q})] = \frac{1}{2\pi}\int_{-\infty}^{\infty} \tilde{f}(\mathbf{q})\exp\left[i\mathbf{q}\cdot\mathbf{r}\right] d^2\mathbf{q} \tag{1.3}$$

$$\tilde{f}(\mathbf{q}) = \mathcal{F}[f(\mathbf{r})] = \frac{1}{2\pi}\int_{-\infty}^{\infty} f(\mathbf{r})\exp\left[-i\mathbf{q}\cdot\mathbf{r}\right] d^2\mathbf{r} \tag{1.4}$$

and also the discrete Fourier transforms:

$$\tilde{f}(q_x, q_y) = \frac{1}{\sqrt{MN}}\sum_{m,n} f(m,n)\exp\left[-2\pi i\left(\frac{mq_x}{M} + \frac{nq_y}{N}\right)\right] \tag{1.5}$$

$$f(m,n) = \frac{1}{\sqrt{MN}}\sum_{q_x,q_y} \tilde{f}(q_x, q_y)\exp\left[2\pi i\left(\frac{mq_x}{M} + \frac{nq_y}{N}\right)\right] \tag{1.6}$$

where (m,n) are pixel indices in the (x,y) directions, M is the array size in the y direction and N is the array size in the x direction. It is possible that sometimes we can have sample space ordering which causes diffraction to be extended out in reciprocal space more in one direction than the other but this is not considered in this thesis. In this thesis, we will always have that $N = M$ and the reason for this is mainly due to the fact that diffraction patterns for the samples we look at decay more or less in a radially symmetric way, necessitating that $N = M$. When using the discrete Fourier transforms computationally, we will need to relate pixel sizes between the measurement arrays and the sample arrays. This is done by the relation between pixel sizes in real space and Fourier space:

$$\Delta x \Delta q_x = 2\pi/N, \tag{1.7}$$

where N is the array size used in the discrete Fourier transforms. This relation tells us how sampling (pixel size) of the measurement (Fourier space) will affect

sampling of the object (real space) we scatter x-rays from. Later, we will see (in Fig. 1.8) that due to the experimental geometry used, we have that:

$$\frac{a}{z} \approx \frac{\Delta q_x}{k}, \tag{1.8}$$

where a is defined as the area detector pixel size (an experimentally fixed quantity), z is the detector to sample distance, and $k = 2\pi/\lambda$, with λ being the x-ray wavelength. This means that the sample space pixel size is

$$\Delta x = \frac{\lambda z}{Na}. \tag{1.9}$$

1.2.2 HSV color space

This thesis deals extensively with complex valued wavefields and so contains many complex-valued images. Showing the real and imaginary parts separately is not possible since most wavefields are only defined up to an overall constant phase factor and that depending on this phase factor, the real and imaginary parts are mixed up in some strange way. It is possible to display the wavefields by separately showing the modulus and phase, but generally this is a waste of space resulting in more "busy" figures for little gain in information. So, in this thesis complex valued wavefields are rendered as shown in Fig. 1.5. This scheme of encoding complex valued images is standard in the phase retrieval community and is known as HSV encoding, where HSV stands for Hue-Saturation-Value.

1.2.3 Helmholtz equation

Consider the forward scattering geometry shown in Fig. 1.6. When a material with index of refraction n is placed in the way of a source of x-ray plane waves of frequency ω propagating in the z direction, what does the x-ray wavefield look like after passing through the material? The answer to this question comes from Maxwell's equations:

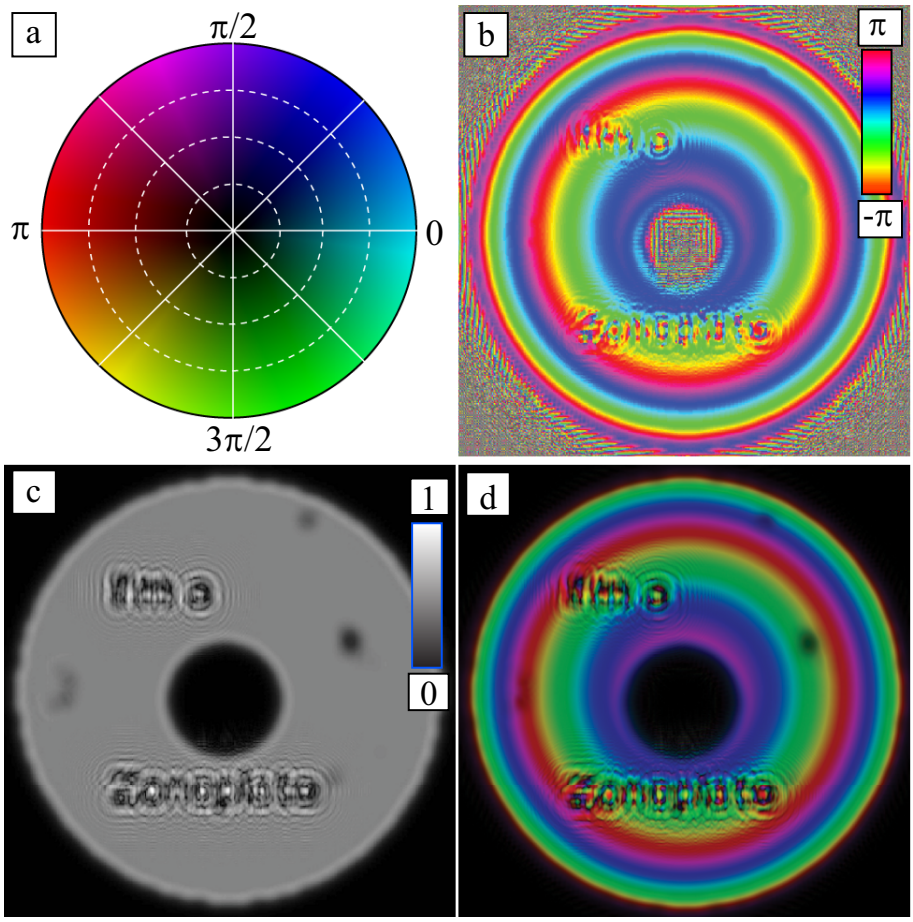

Figure 1.5: (a) The hue-saturation-value (HSV) colorwheel used to encode complex valued images. (b) The phase of an image, which can vary between $\pm\pi$, is encoded as a periodic color scheme as seen in (a). (c) These complex valued images have the modulus rescaled between 0 and 1 so that the brightest pixel has the maximum brightness at the edge of the wheel and the darkest pixel is at the center of the wheel. (d) The final HSV encoded complex valued image with the phase in (b) encoded as hue and modulus in (c) encoded as value (or brightness). The definition of zero phase is somewhat arbitrary as the absolute phases in this thesis are generally indeterminable; only relative phases are recoverable.

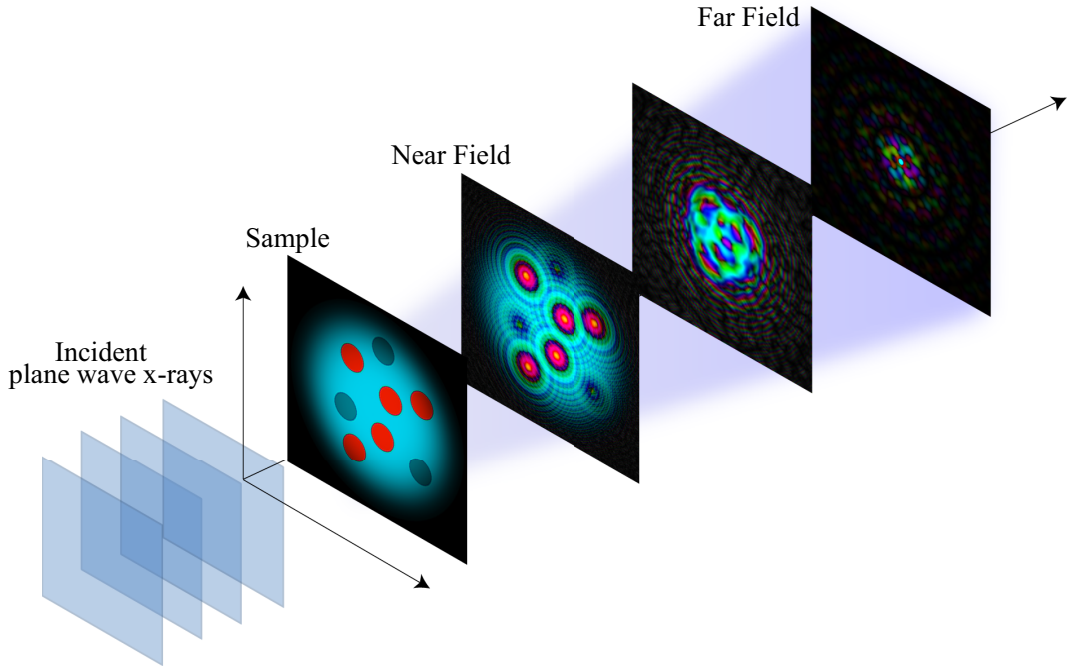

Figure 1.6: The forward scattering (or transmission) geometry. X-ray plane waves are incident on a thin sample (thin in the direction of propagation), and depending on the spatial configuration of the electron density will diffract. There are different "regimes" of diffraction encountered in this thesis: near field diffraction (also known as Fresnel diffraction) and far field diffraction (or Fraunhofer diffraction).

$$\nabla \times \mathbf{H} = -i\omega\epsilon(\mathbf{r},\omega)\mathbf{E} \tag{1.10a}$$

$$\nabla \cdot \mathbf{H} = 0 \tag{1.10b}$$

$$\nabla \times \mathbf{E} = i\omega\mu\mathbf{H} \tag{1.10c}$$

$$\nabla \cdot [\epsilon(\mathbf{r},\omega)\mathbf{E}] = 0, \tag{1.10d}$$

where ω is the frequency of the oscillating electromagnetic field, $\epsilon(\mathbf{r},\omega)$ is the complex electric permittivity tensor, and μ is the magnetic permeability and is assumed not a tensor but constant in space and frequency. One vital approximation [18] we use here is that the length scale of the spatial variations of $\epsilon(\mathbf{r},\omega)$ is very much longer than the x-ray wavelength λ. In other words, spatial derivatives of $\epsilon(\mathbf{r},\omega)$ are negligible when compared to spatial derivatives of $\mathbf{E}$ or $\mathbf{H}$. So, we get

that:

$$\begin{aligned}\nabla \times (\epsilon(\mathbf{r},\omega)\mathbf{E}) &= \nabla\epsilon(\mathbf{r},\omega) \times \mathbf{E} + \epsilon(\mathbf{r},\omega)\nabla \times \mathbf{E} \\ &\approx \epsilon(\mathbf{r},\omega)\nabla \times \mathbf{E} \\ &= i\omega\epsilon(\mathbf{r},\omega)\mu\mathbf{H}. \end{aligned} \tag{1.11}$$

Next, using the vector calculus identity:

$$\nabla \times \nabla \times \mathbf{H} = \nabla(\nabla \cdot \mathbf{H}) - \nabla^2\mathbf{H} = -\nabla^2\mathbf{H}, \tag{1.12}$$

and also that $\nabla \times \nabla \times \mathbf{H} = -i\omega\nabla \times \epsilon(\mathbf{r},\omega)\mathbf{E} = -i\omega\epsilon(\mathbf{r},\omega)\nabla \times \mathbf{E}$, we get:

$$\nabla^2\mathbf{H} + \omega^2\epsilon(\mathbf{r},\omega)\mu\mathbf{H} = 0. \tag{1.13}$$

Taking the curl of the above equation again, dropping all spatial derivatives of $\epsilon(\mathbf{r},\omega)$ and using Ampère's Law gives the same expression for the electric field:

$$\nabla^2\mathbf{E} + \omega^2\epsilon(\mathbf{r},\omega)\mu\mathbf{E} = 0. \tag{1.14}$$

As we do yet not care about order parameters within the sample which scatter x-rays differently depending on the direction of the incident electric field polarization (we will in a few chapters), we can ignore the electric field polarization direction and go from the vector partial differential equation above to a scalar partial differential equation known as the Helmholtz equation:

$$\nabla^2\Psi + k^2n^2\Psi = 0, \tag{1.15}$$

where $n^2 = c^2\epsilon(\mathbf{r},\omega)\mu$ is the index of refraction, c is the speed of light in a vacuum, and $k = \omega/c = 2\pi/\lambda$ is the vacuum wavenumber.

1.2.4 Projection approximation

What we measure downstream of the sample shown in Fig. 1.6 are solutions to the Helmholtz equation. Considering an incident plane wave on the sample as in Fig. 1.7 in the z direction and also assuming small angle scattering (scattered $\hat{\mathbf{k}}$ is $\simeq \hat{\mathbf{z}}$) we first factor out the highly oscillatory component of $\Psi(\mathbf{r}) = \psi(\mathbf{r})e^{-ikz}$,

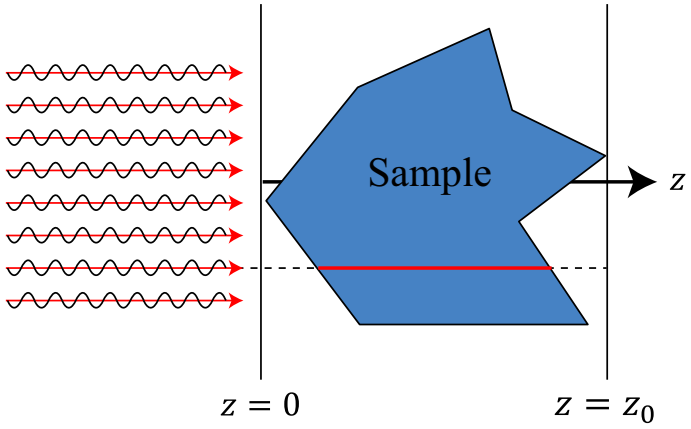

Figure 1.7: The projection approximation. In this approximation, it is assumed that the sample is thin enough and that x-rays interact weakly enough with electrons so that the path an x-ray travels through the sample is negligibly perturbed (in x and y) when compared to the path it would have taken had the sample been absent.

where $\psi(\mathbf{r})$ is slowly varying in z. Since $\psi(\mathbf{r})$ is slowly varying in z, derivatives of $\psi(\mathbf{r})$ in x and y dominate derivatives in z, so $\nabla^2 \approx \nabla_\perp^2 = \partial^2/\partial x^2 + \partial^2/\partial y^2$. In this case, Eq. 1.15 becomes:

$$\nabla_\perp^2 \psi + 2ik\frac{\partial \psi}{\partial z} + 2k^2 \Delta n \psi = 0, \tag{1.16}$$

where $n^2 = (1+\Delta n)^2 = (1-\delta+i\beta)^2 \approx 1 + 2\Delta n$ (Δn is assumed small). This is known as the paraxial inhomogeneous Helmholtz equation.

If we make a further assumption that x-ray electron interactions are weak enough that the path an x-ray travels through the sample is negligibly perturbed (in x and y) when compared to the path it would have taken had the sample been absent, then the x-ray wavefield at $z = z_0$ can be approximated as being solely determined by the phase and amplitude shifts it accumulates along this path connecting the entrance and exit surfaces. This is saying that if variations of the x-ray wavefront are negligible in x and y, then $\nabla_\perp^2 \psi$ can be dropped from Eq. 1.16. Thus, we have:

$$\frac{\partial \psi}{\partial z} - ik\Delta n \psi = 0, \tag{1.17}$$

and the solution to this is [19]:

$$\psi(\mathbf{r}, z_0) = \psi_0(\mathbf{r}) T(\mathbf{r}) = \psi_0(\mathbf{r}) \exp\left(ik \int_0^{z_0} (n(\mathbf{r}, z) - 1) dz\right), \tag{1.18}$$

where $\psi_0(\mathbf{r})$ is the incident wavefield at $z = 0$, and $T(\mathbf{r})$ is known as the sample transmission function.

1.2.5 Small angle approximation

In the small angle scattering geometry shown in Fig. 1.6, we measure the wavefield intensity somewhere downstream (i.e. for $z > z_0$) of the sample in a plane transverse to the propagation direction. To mathematically model this in a tractable way we separate the transverse (i.e. $\mathbf{r}_\perp = (x, y)$) and parallel z components of the downstream wavefield $\psi(\mathbf{r})$, so that $\psi(\mathbf{r}) = \psi(\mathbf{r}_\perp, z)$. The Fourier transform of Eq. 1.15 in this transverse plane $\mathbf{r}_\perp$ for some $z > z_0$ gives:

$$\left(-\mathbf{q}_\perp^2 + \frac{\partial^2}{\partial z^2} + k^2\right)\tilde{\psi}(\mathbf{q}_\perp, z) = 0. \tag{1.19}$$

where $\mathbf{q}_\perp = (q_x, q_y)$. The general solution to Eq. 1.19 is:

$$\tilde{\psi}(\mathbf{q}_\perp, z) = \tilde{\Phi}^-(\mathbf{q}_\perp)e^{-i\kappa z} + \tilde{\Phi}^+(\mathbf{q}_\perp)e^{+i\kappa z}, \tag{1.20}$$

where $\kappa = \sqrt{k^2 - \mathbf{q}_\perp^2}$ (see Fig. 1.8). In the small angle regime, we have $|\mathbf{q}_\perp| \ll k$ and so we make the approximation:

$$\sqrt{1 - \frac{\mathbf{q}_\perp^2}{k^2}} \approx 1 - \frac{\mathbf{q}_\perp^2}{2k^2}. \tag{1.21}$$

We assume that no part of the incident wavefield is backscattered, and so we set $\tilde{\Phi}^-(\mathbf{q}_\perp) \to 0$. The quantity $\tilde{\Phi}^+(\mathbf{q}_\perp)$ can thus be interpreted as the wavefield at some arbitrary z, and Eq. 1.20 gives how the wave evolves at any further z:

$$\psi(\mathbf{r}_\perp, z) = \mathcal{F}^{-1}\left\{\tilde{\Phi}^+(\mathbf{q}_\perp)\exp\left[ikz\left(1 - \frac{\mathbf{q}_\perp^2}{2k^2}\right)\right]\right\}. \tag{1.22}$$

To determine $\tilde{\Phi}^+(\mathbf{q}_\perp)$, we use the boundary condition at $z = z_0$ given by Eq. 1.18, which means that we can write that (see Section 1.4.2 in [19]):

$$\begin{aligned}\tilde{\Phi}^+(\mathbf{q}_\perp) &= \mathcal{F}\left\{\psi(\mathbf{r}_\perp, z_0)\right\} \\ &= \mathcal{F}\left\{\psi_0(\mathbf{r}_\perp)\,\exp\left(ik\int_0^{z_0}\Delta n(\mathbf{r}_\perp, z)dz\right)\right\}.\end{aligned} \tag{1.23}$$

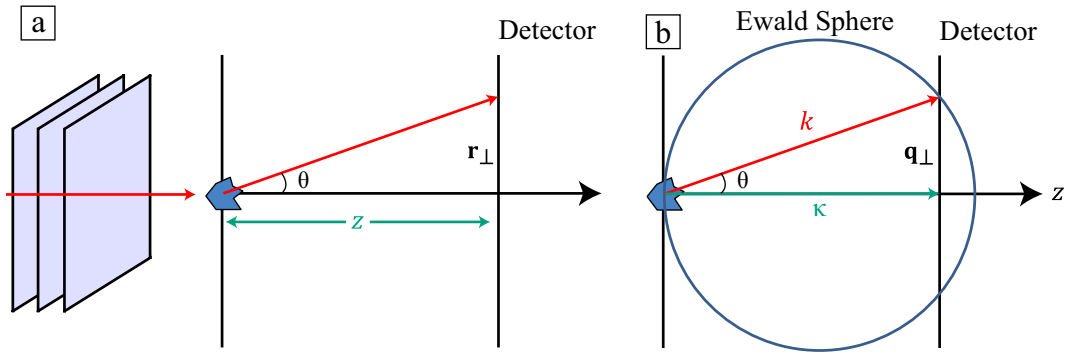

Figure 1.8: The small angle approximation. This approximation allows us to relate sampling pixel sizes at the measurement location and at the sample location in a simple way using the relation $\mathbf{r}_\perp/z \approx \mathbf{q}_\perp/\kappa \approx \mathbf{q}_\perp/k$.

Next, we write Eq. 1.22 as a convolution:

$$\psi(\mathbf{r}_\perp, z) = \psi(\mathbf{r}_\perp, z_0) * \Pi(\mathbf{r}_\perp, z), \tag{1.24}$$

where $\Pi(\mathbf{r}_\perp, z)$ is known as the Fresnel propagator:

$$\begin{aligned}\Pi(\mathbf{r}_\perp, z) &= \frac{1}{2\pi}\mathcal{F}^{-1}\left\{\exp\left[ikz\left(1 - \frac{\mathbf{q}_\perp^2}{2k^2}\right)\right]\right\} \\ &= \frac{-ik}{2\pi z}\exp\left[ikz\left(1 + \frac{\mathbf{r}_\perp^2}{2z^2}\right)\right].\end{aligned} \tag{1.25}$$

Using the two dimensional convolution theorem:

$$\mathcal{F}\{\psi(\mathbf{r}_\perp, z_0) * \Pi(\mathbf{r}_\perp, z)\} = 2\pi\mathcal{F}\{\psi(\mathbf{r}_\perp, z_0)\}\mathcal{F}\{\Pi(\mathbf{r}_\perp, z)\}, \tag{1.26}$$

we write Eq. 1.22 as:

$$\begin{aligned}\psi(\mathbf{r}_\perp, z) &= \frac{e^{ikz}}{i\lambda z}\int d^2\mathbf{r}'_\perp \psi(\mathbf{r}'_\perp, z_0)\exp\left[\frac{ik}{2z}(\mathbf{r}_\perp - \mathbf{r}'_\perp)^2\right] \\ &= \frac{e^{ikz}}{i\lambda z}\exp\left(\frac{ik\mathbf{r}_\perp^2}{2z}\right)\int d^2\mathbf{r}'_\perp \psi(\mathbf{r}'_\perp, z_0)\exp\left[\frac{ik\mathbf{r}_\perp'^2}{2z} - \frac{ik}{z}\mathbf{r}_\perp \cdot \mathbf{r}'_\perp\right].\end{aligned} \tag{1.27}$$

From the geometry of the scattering experiment shown in Fig. 1.8, we identify that $\mathbf{r}_\perp/z \approx \mathbf{q}_\perp/\kappa \approx \mathbf{q}_\perp/k$. With this, we finally get that:

$$\begin{aligned}\psi(\mathbf{r}_\perp, z) &= \frac{e^{ikz}}{i\lambda z}\exp\left(\frac{ik\mathbf{r}_\perp^2}{2z}\right)\int d^2\mathbf{r}'_\perp \psi(\mathbf{r}'_\perp, z_0)\exp\left[\frac{ik\mathbf{r}'^2_\perp}{2z} - i\mathbf{q}_\perp\cdot\mathbf{r}'_\perp\right] \\ &= \frac{e^{ikz}}{i\lambda z}\exp\left(\frac{ik\mathbf{r}_\perp^2}{2z}\right)\mathcal{F}\Big\{\psi(\mathbf{r}'_\perp, z_0)\exp\left[\frac{ik\mathbf{r}'^2_\perp}{2z}\right]\Big\}. \end{aligned} \tag{1.28}$$

The $\exp(-ik\mathbf{r}_\perp^2/2z)$ and $\exp(-ik\mathbf{r}'^2_\perp/2z)$ factors in Eq. 1.28 are known as a quadratic phase "curvature" terms; the $\exp(-ik\mathbf{r}_\perp^2/2z)$ term can be thought of as describing the phase curvature in the plane we just propagated to and the $\exp(-ik\mathbf{r}'^2_\perp/2z)$ term the phase curvature in the plane we originally started from. Equation 1.28 is known as the Fresnel diffraction integral, and is in a computationally useful form for calculating coherent diffraction patterns downstream of a sample since it is now formulated in terms of a single Fourier transformation, as opposed to Eq. 1.22, which requires a Fourier transformation of the exit wave right after the sample (i.e. $\psi(\mathbf{r}_\perp, z_0)$) and an inverse Fourier transformation.

1.3 Computational wavefield propagation

In this section, we explore how to computationally implement the results of the previous section. Computational wavefield propagation must be done carefully if numerical stability is to be retained. The most common issues related to computational wavefield propagation have to do with proper sampling of the quadratic phase curvature terms we encountered in the previous section. We go into some common problems which occur when sampling is done carelessly. We then present effective and stable computational tools for properly doing computation wavefield propagation.

1.3.1 Plane wave propagation

Now that we have an analytical expression in Eq. 1.28 for how the wavefield downstream of the sample in Fig. 1.6 diffracts, we now explore issues which arise when we must computationally work with it. We first assume that we have N

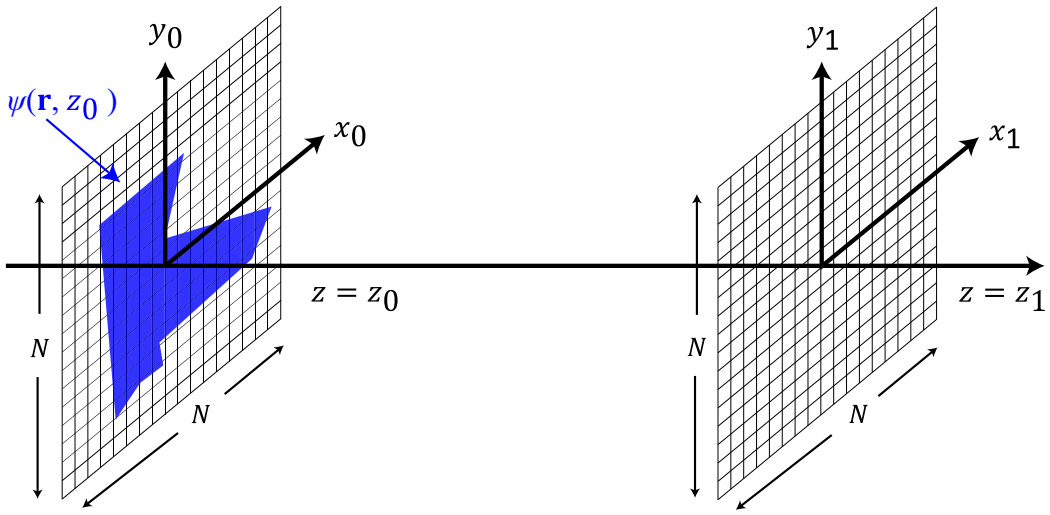

Figure 1.9: The way in which we view the forward scattering small angle experiments in computation. The side length for the initial plane on the left is L_0 and the side length for plane we will propagate to is L_1.

pixels (i.e. samples) in both the x and y directions, in both the input plane at z_0 and output plane at at z_1, as shown in Fig. 1.9. We define the pixel size at the output plane to be a; the reason for this definition is that the output plane is where we would put a pixel area detector in an experiment and the pixel size of the detector is known and constant. The field of view in the detector plane at z_1 is $L_1^2 = aN \times aN$. Thus $\mathbf{r}_\perp = (m\Delta x_1, n\Delta y_1) = (mL_1/N, nL_1/N)$, where L_1 is the side length of the array used in the output plane. Denote the distance between the input plane and output planes to be $z_{01} = z_1 - z_0$. Using the relation $\mathbf{r}_\perp/z \approx \mathbf{q}_\perp/k$ from Fig. 1.8, the Fourier space pixel size is found to be $\Delta q_x = \Delta q_y = ka/z_{01}$, and further using the relation between pixel sizes in real space and Fourier space given by $\Delta x \Delta q_x = 2\pi/N$, the real space pixel size at the input plane at z_0 is found to be $\Delta x_0 = \Delta y_0 = \lambda z_{01}/aN$. This also means that the side length of the array at the input plane is $L_0 = \lambda z_{01}/a = \lambda z_{01} N/L_1$.

A simulated numerical evaluation of Eq. 1.28 is shown in Fig. 1.10. One problem that we immediately encounter when z_{01} is small is aliasing and severe numerical artifacts at the edges of the array as seen in Fig. 1.10b. The cause of this (see reference [20] for more) can be seen by examining the expression $L_1 = \lambda z_{01} N/L_0$; if $z_{01} < L_0^2/\lambda N$ then $L_1 < L_0$. This means that it is possible that

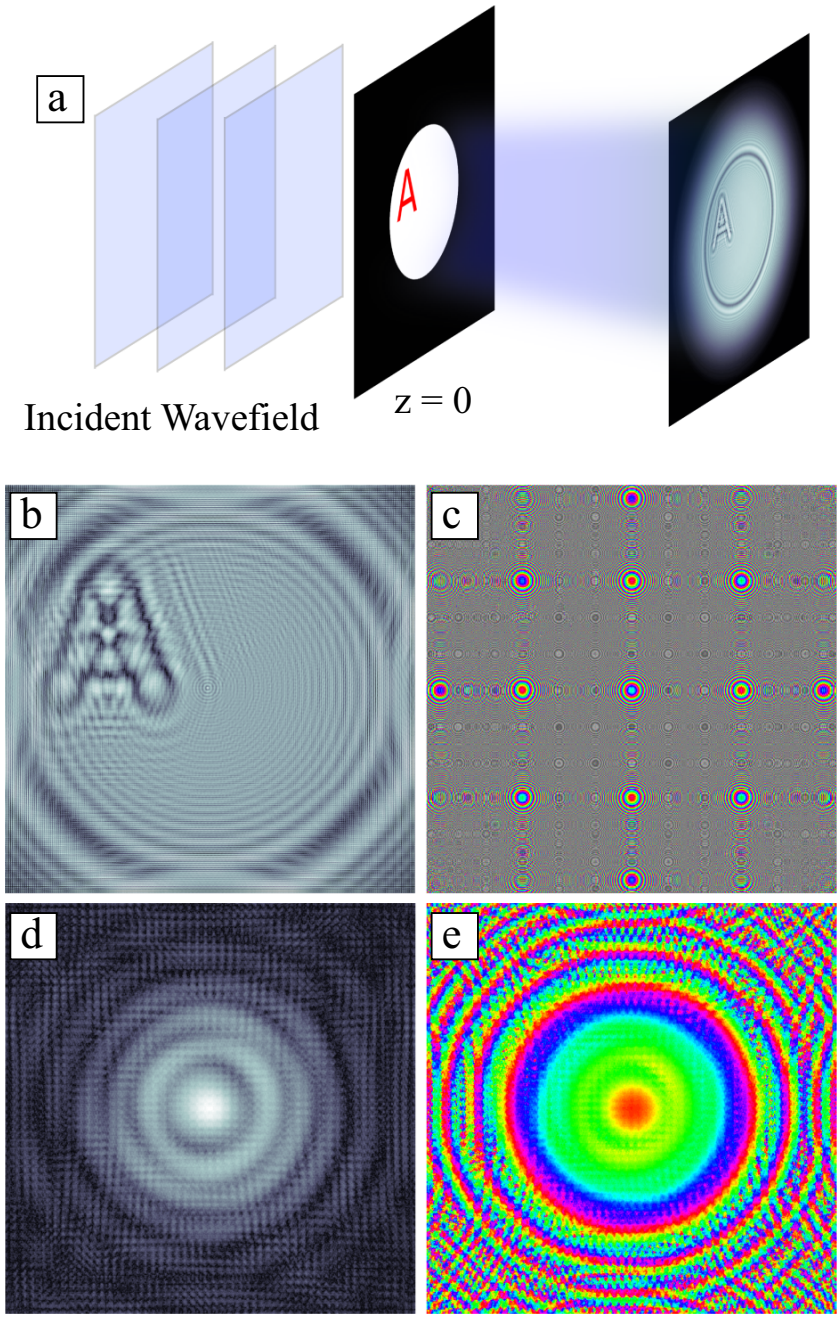

Figure 1.10: (a) Experimental geometry for plane waves incident on a sample. In (b-e), computational artifacts due to aliasing and insufficient sampling of the quadratic phase curvature terms in the Fresnel diffraction integral are shown. Defining $z_{crit} = L_0^2/\lambda N$, frame (b) shown numerical artifacts in the magnitude of the result using the "Direct" method for $z = 0.25\ z_{crit}$, while (c) is the phase. Frame (d) shown numerical artifacts in the magnitude of the result using the "Spectrum propagation" method for $z = 100\ z_{crit}$, while (f) is the phase.

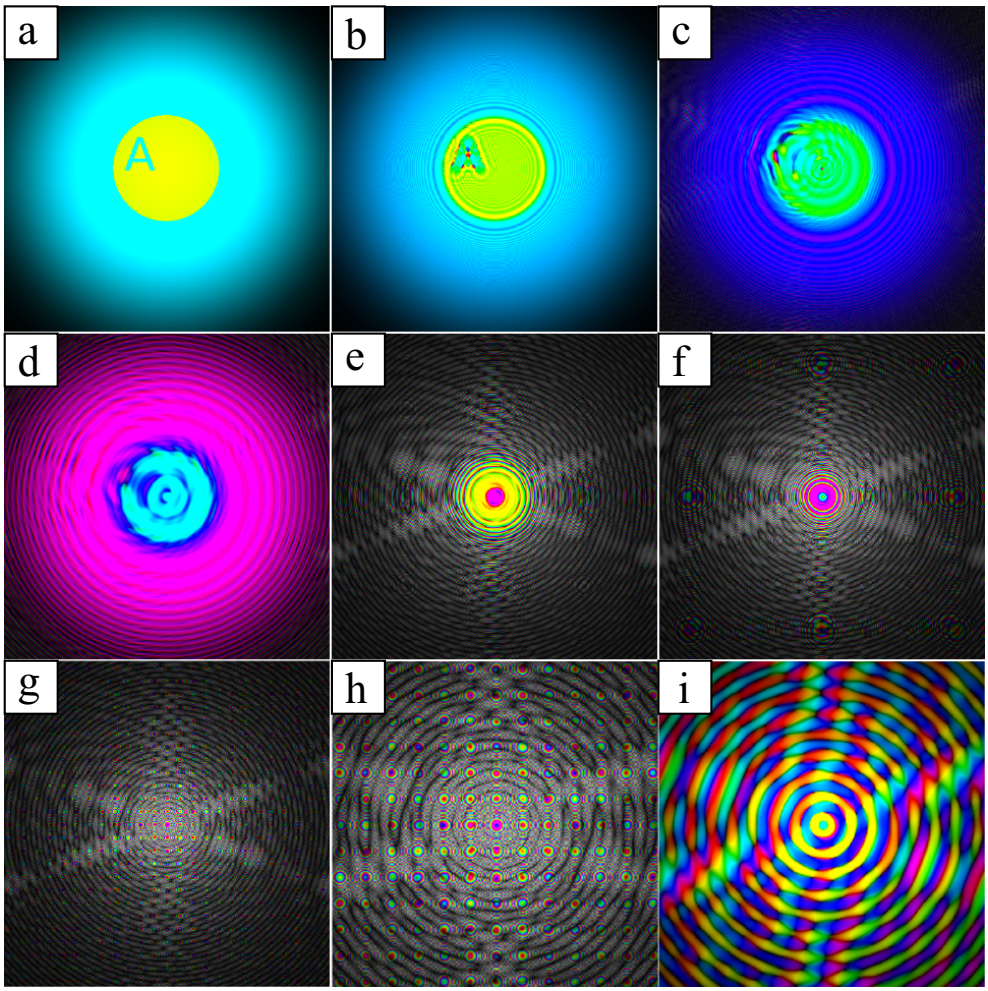

Figure 1.11: Defining $z_{crit} = L_0^2/\lambda N$, wavefield propagation is shown for (a) $z = 0$ (i.e. this is what is being propagated for $z > 0$), (b) $z = 0.1z_{crit}$, (c) $z = 0.5z_{crit}$, (d) $z = z_{crit}$, (e) $z = 5z_{crit}$, (f) $z = 10z_{crit}$, (g) $z = 100z_{crit}$. (h) is a close up of (g); when we begin to propagate more than $z \gtrsim 10z_{crit}$, Moire patterns due to phase aliasing in the quadratic phase curvature term outside of the Fourier transform in Eq. 1.29 begin to emerge. Frame (i) is also $z = 100z_{crit}$ only with the outside quadratic phase curvature term removed. Far from the aperture at $z = 0$, we would expect that quadratic phase curvature to be negligible (almost a flat or constant phase), so it should be removed when using the "Direct" method of propagation when $z \gg z_{crit}$.

the array size at the output plane can be too small to contain the all the spatial information present in the propagated wavefield, and aliasing of the FFT algorithm for computing the Fourier transform will occur. The aliasing effects present in the phase of the $\exp(ik\mathbf{r}'^2/2z)$ term inside the Fourier transform in Eq. 1.28 is especially devastating; as the phase of $\exp(ik\mathbf{r}'^2/2z)$ is pathological, the Fourier transform of the $\exp(ik\mathbf{r}'^2/2z)$ term multiplied by $\psi(\mathbf{r}_\perp, z_0)$ will also end up pathological. As a result, we must enforce that $z_{01} \geq L_0^2/\lambda N$ when using Eq. 1.28 [21].

In order to be able to propagate a wavefield when $z_{01} < L_0^2/\lambda N$, we should use Eq. 1.22 instead of Eq. 1.28. Since this expression uses a forward Fourier transform on $\psi(\mathbf{r}_\perp, z_0)$, multiplies with $\exp(-iz\mathbf{q}_\perp^2/2k)$, and inverse Fourier transforms the result, we have that $L_0 = L_1$ for all $z_{01} < L_0^2/\lambda N$ [20]. However, when using Eq. 1.22 for $z_{01} > L_0^2/\lambda N$ the diffracted wavefield will again have expanded so that it will have spatial features outside of the array at z_1, and we will again begin to see aliasing and severe numerical artifacts at the edges of the array; this can be seen in Fig. 1.10c-d where we have propagated a distance of $z_{01} = 100 \times L_0^2/\lambda N$. In summary we have that:

$$\psi(\mathbf{r}_\perp, z) = \begin{cases} \frac{e^{ikz}}{i\lambda z}\exp\left(\frac{ik\mathbf{r}_\perp^2}{2z}\right)\mathcal{F}\left\{\psi(\mathbf{r}'_\perp, z_0)\exp\left[\frac{ik\mathbf{r}'^2_\perp}{2z}\right]\right\} & \text{if } z_{01} \geq L_0^2/\lambda N, \\ e^{ikz}\mathcal{F}^{-1}\left\{\mathcal{F}\left\{\psi(\mathbf{r}'_\perp, z_0)\right\}\exp\left[\frac{-iz\mathbf{q}_\perp^2}{2k}\right]\right\} & \text{if } z_{01} < L_0^2/\lambda N. \end{cases} \tag{1.29}$$

and that $L_1 = L_0$ if $z_{01} < L_0^2/\lambda N$ but $L_1 = \lambda z_{01} N/L_0$ if $z_{01} >= L_0^2/\lambda N$; Fig. 1.11 shows the application of Eq. 1.29 for various z. For the remainder of this thesis, the method above using a single Fourier transform is termed the "Direct" method of propagating a wavefield while the method requiring two Fourier transforms is termed the "Spectrum propagation" method.

1.3.2 Focused beam propagation

Next we look at computational wavefield propagation in the case for which x-ray optics are used, the optics being the Fresnel zone plate introduced in section 1.1.1. Here we make the good approximation that these x-ray optics have similar properties as to those of a "thin lens", as seen in Fig. 1.12. The most important aspect of how a thin lens interacts with an incident plane wave is given

by [1]:

$$\psi(\mathbf{r}, z_0) = \psi(\mathbf{r}, 0)\Theta(\mathbf{r})\exp\Big[-\frac{ik}{2f}\mathbf{r}^2\Big] \tag{1.30}$$

where $\psi(\mathbf{r}, 0)$ is the wave incident on the lens, $\Theta(\mathbf{r})$ is known as the "pupil" function and describes the shape of the lens, $\psi(\mathbf{r}, z_0)$ is the exit wave after passing through the lens, and f is the *focal length* of the lens, i.e. the point at which the lens has concentrated photons into the smallest region of space. Here, it is assumed that $\psi(\mathbf{r}, 0)$ fully illuminates the pupil function, and so we have that $\psi(\mathbf{r}, 0)$ is constant across $\Theta(\mathbf{r})$. So, we can write that $\psi(\mathbf{r}, 0)\Theta(\mathbf{r}) = \psi_0\Theta(\mathbf{r})$, where ψ_0 is a constant proportional to the incident photon flux.

As a concrete example, first assume a pupil function as shown in Fig. 1.12, $N = 512$, $\lambda = 1.04$ nm, and some experimentally realistic parameters for a zone plate: a radius for the pupil function of $r_N = 0.08$ mm, $\Delta r_N = 30$ nm, $L_0 = 0.4$ mm, and so a focal length of $f = 4.6$ mm (see Eq. 1.1). With the parameters chosen here, we have that $f \ll L_0^2/(\lambda N)$. Thus at first glance, if we wish to propagate from the pupil function plane to the focal plane (or anywhere in between), it looks like we would need to use the "Spectrum propagation" method. The problem we would quickly run into is that the quadratic phase curvature term $\exp(-ik\mathbf{r}^2/2f)$ in Eq. 1.30 has a pathologically highly oscillatory phase using an array size of $N = 512$, meaning the Fourier transform of it multiplied by the pupil function would also be pathological. Increasing N to try to get out of this problem also is prohibitively expensive in terms of computer memory. We perhaps ironically are saved by using the "Direct" propagation method instead. Inserting Eq. 1.30 into Eq. 1.29, we find that:

$$\psi(\mathbf{r}, z) = \frac{e^{ikz}}{i\lambda z}\exp\left(\frac{ik\mathbf{r}^2}{2z}\right)\mathcal{F}\Big\{\psi_0\Theta(\mathbf{r}')\exp\Big[\frac{ik}{2}\Big(\frac{1}{z} - \frac{1}{f}\Big)\mathbf{r}'^2\Big]\Big\}. \tag{1.31}$$

From Eq. 1.31, we see that if $z = f$, i.e. we propagate the wavefield exiting the lens to the focal plane, the wavefield at the focus is

$$\psi(\mathbf{r}, f) = \psi_0\frac{e^{ikf}}{i\lambda f}\exp\left(\frac{ik\mathbf{r}^2}{2f}\right)\mathcal{F}\Big\{\Theta(\mathbf{r}')\Big\}. \tag{1.32}$$

Equation 1.32 says that the wavefield at the focus is, apart from the $\exp(-ik\mathbf{r}^2/2f)$ quadratic curvature term outside the Fourier transform and other constant phase

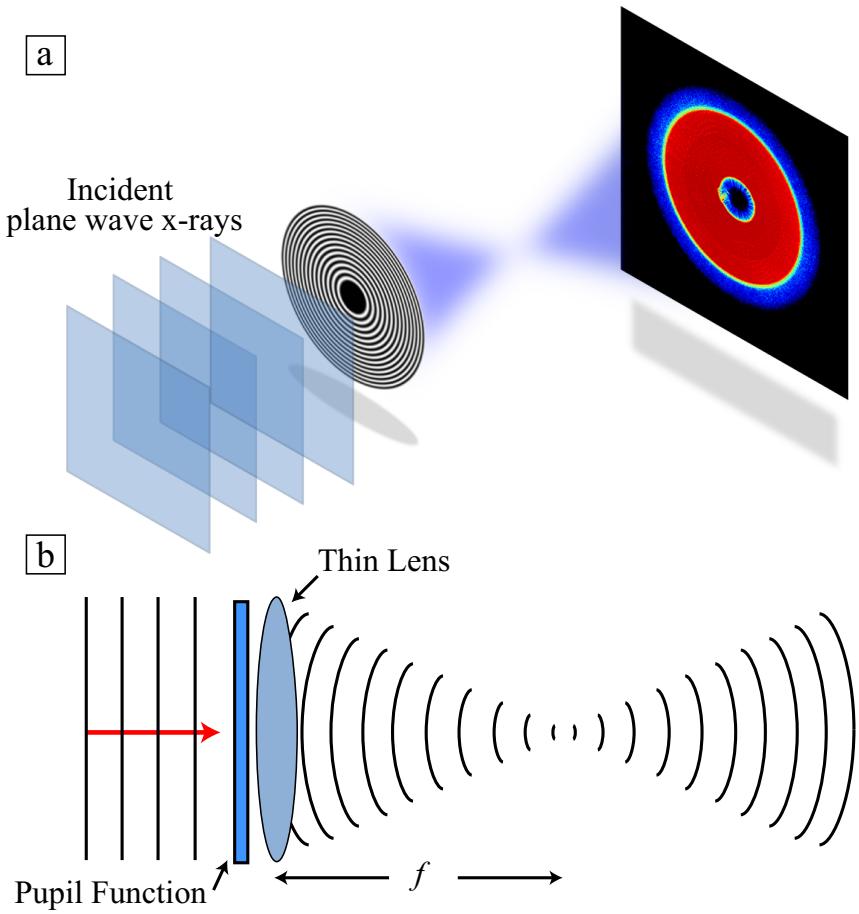

Figure 1.12: (a) Focused wavefield propogation. This kind of wavefield propagation is necessary when we use a condenser optic to focus the beam down so that we can place the sample past to focal spot so that the incident x-ray illumination will have phase structure, which is a useful sample space constraint. (b) The definition of the pupil function and how this and the quadratic phase curvature is used in the computational propagation method presented.

or scaling prefactors, just the Fourier transform of the lens pupil function. At the focal plane we have an array side length of $L_1 = \lambda f N/L_0 = 6$ μm, which means that the $\exp(-ik\mathbf{r}^2/2f)$ quadratic curvature term outside the Fourier transform in Eq. 1.32 is well behaved. If we wish to further propagate the wavefield at the focus to some other location, the criterion for whether we use the "Spectrum propagation" method is $z < L_1^2/\lambda N = 69$ μm, otherwise we use the "Direct" propagation method. Another computational tool for attacking this problem comes from the use of the "Fractional Fourier Transform". From references [22, 23], for an optical system like that in Fig. 1.12b, we can write the Fresnel diffraction integral Eq. 1.28 as:

$$\begin{aligned}\psi(\mathbf{r},z) &= \exp\Big[-i\frac{\pi\ \tan(\phi/2)}{N}\mathbf{r}^2 + i\phi\Big] \\ &\quad \times\mathcal{F}^{-1}\left\{\exp\Big[-i\frac{\pi\sin(\phi)}{N}\mathbf{q}^2\Big]\right. \\ &\quad \left.\times\mathcal{F}\left\{\Theta(\mathbf{r}')\exp\Big[-i\frac{\pi\tan(\phi/2)}{N}\mathbf{r}'^2\Big]\right\}\right\}, \end{aligned} \tag{1.33}$$

where $z = f\ \tan\phi$, and $\phi < \pi/4$. This allows us to propagate anywhere from the pupil function plane to the focal plane, i.e. for $0 \le z \le f$, with the array side length of $L_1 = L_0$ everywhere in this region. For locations $z > f$ past the focal plane, we start with the wavefield at $z = f$, i.e. simply replace $\Theta(\mathbf{r}')$ in Eq. 1.33 with $\psi(\mathbf{r}, z = f)$, and use Eq. 1.33 again, with $z = f + f\ \tan\phi$, $\phi < \pi/4$, and $f \le z \le 2f$. The array side length is also $L_1 = L_0$ for all $f \le z \le 2f$. Equation 1.33 cannot be used for $z > 2f$, and we must resort to the use of Eq. 1.29 in this regime; to do this we again start at the focal plane and use the "Direct" propagation method for all $z > 2f$.

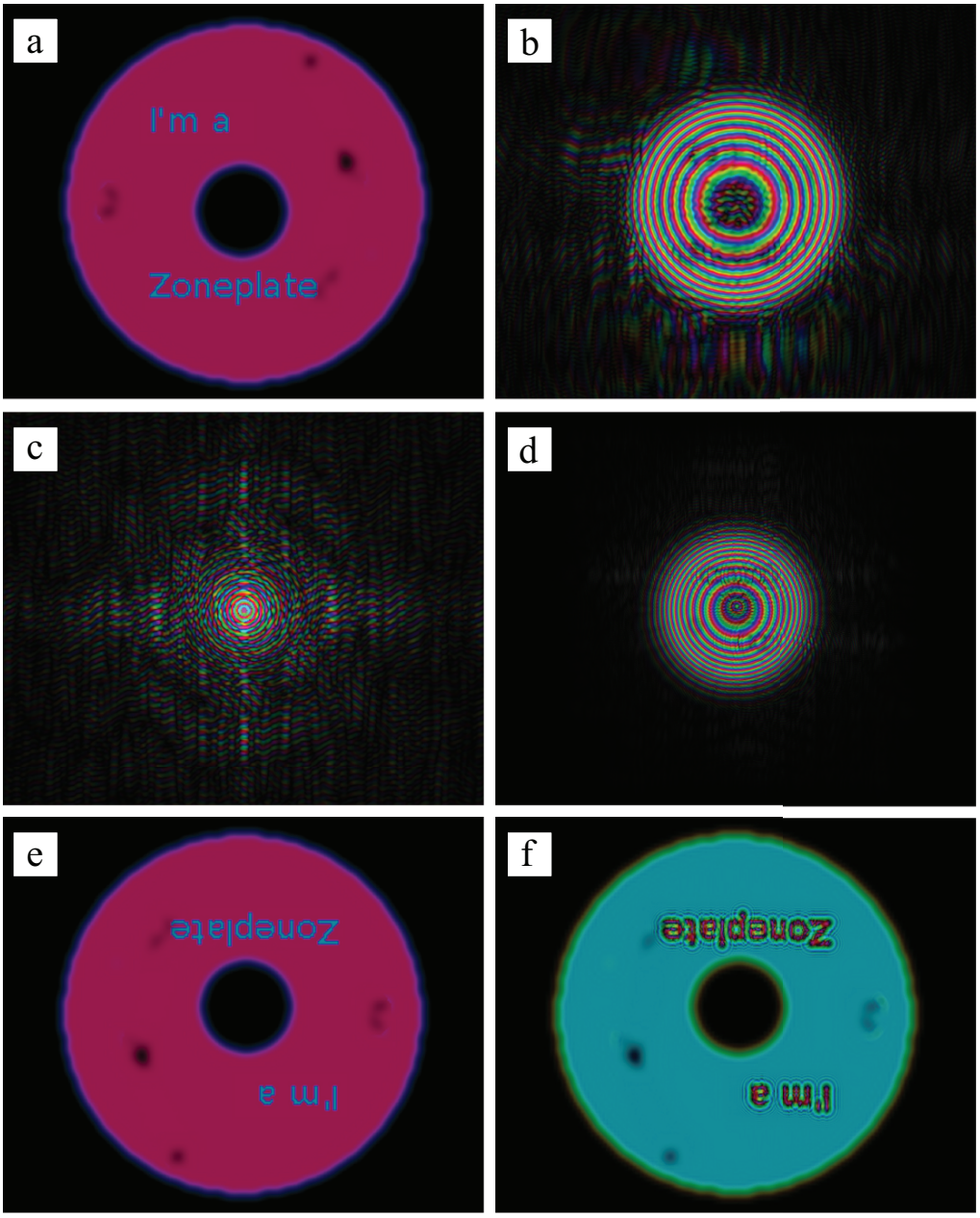

Figure 1.13: Application of the focused beam propagation method. (a) The pupil function. (b)The pupil function propagated downstream of the pupil function plane but before the focal plane. (c) The pupil function propagated to the focal plane. (d) The wavefield at the focal plane propagated downstream of the focal plane but before the plane located at $z = 2f$. (e) The wavefield at $z = 2f$. As expected, it is an inverted but unmagnified pupil function. (f) The wavefield for $z \gg 2f$. It is basically an inverted magnified pupil function, but some Fresnel fringing occurs.

Chapter 2

The phase problem

In the previous chapter, Coherent X-ray Diffractive Imaging (CXDI) was introduced as an imaging technique which numerically inverts a coherent diffraction pattern giving us a diffraction limited high resolution map of the sample under investigation. The numerical inversion process must solve the "phase problem" which arises due to an area detectors inability to measure a complex valued wavefield. Only the intensity, or magnitude squared, of the wavefield can be measured. This chapter delves into the mathematical tools necessary to solve the phase problem and do "phase retrieval". Phase retrieval is an extremely challenging optimization problem. It requires numerical solutions of nonlinear, non-convex, and very large scale problems ($\sim 10^6$ variables). This chapter goes into some algorithmic techniques which can effectively tackle this problem. The algorithms can be split into two families: *gradient descent algorithms*, which are covered in Section 2.1, and *projections onto constraint sets algorithms*, which are covered in Section 2.2. We explore these algorithms by applying them to some two dimensional optimization problems in order to gain some intuition on their characteristics and behavior. We then discuss how to apply the optimization algorithms to problems encountered in CXDI in Section 2.3.

2.1 Error metrics and gradient descent algorithms

Assuming $\psi(\mathbf{r})$ is the wavefield after passing through a sample as in the previous chapter, we begin by defining an error metric:

$$\epsilon_m = \|\pi_m \psi(\mathbf{r}) - \psi(\mathbf{r})\|^2, \tag{2.1}$$

where the π_m operator is known as the Fourier modulus operator and is defined as:

$$\begin{aligned} \pi_m \psi(\mathbf{r}) &= \mathcal{F}^{-1} \tilde{\pi}_m \mathcal{F}\{\psi(\mathbf{r})\} = \mathcal{F}^{-1} \tilde{\pi}_m \tilde{\psi}(\mathbf{q}) \\ &= \mathcal{F}^{-1} \tilde{\pi}_m |\tilde{\psi}(\mathbf{q})| e^{\phi(\mathbf{q})} = \mathcal{F}^{-1} \sqrt{I(\mathbf{q})} e^{\phi(\mathbf{q})}, \end{aligned} \tag{2.2}$$

where $I(\mathbf{q})$ is the appropriately oversampled [24] measured diffraction intensity, $\mathcal{F}$ denotes numerical propagation of $\psi(\mathbf{r})$ to the measurement location (and is typically just an Fourier transform for Fraunhofer diffraction), $\mathcal{F}^{-1}$ denotes numerical propagation from the measurement location back to the sample location (and is typically just an inverse Fourier transform for Fraunhofer diffraction), and the Euclidean length of a vector is defined as:

$$\|\rho\|^2 = \rho \cdot \rho^\dagger = \sum_{\mathbf{r}} |\rho(\mathbf{r})|^2 = \sum_{\mathbf{q}} |\tilde{\rho}(\mathbf{q})|^2, \tag{2.3}$$

where $\tilde{\rho}(\mathbf{q}) = \mathcal{F}\{\rho(\mathbf{r})\}$, and the last equality in the preceding equation is Parseval's theorem. With this, we can write the error metric in Eq. 2.1 as

$$\epsilon_m = \sum_{\mathbf{q}} \left\{ |\mathcal{F}\left[\psi(\mathbf{r}, z_0)\right]| - \sqrt{I(\mathbf{q})} \right\}^2. \tag{2.4}$$

The computational task we have here is to find a configuration for the elements of the $\psi(\mathbf{r})$ array which minimize the error of the metric given in Eq. 2.4. One possibility we can try to accomplish this starting from a guess for $\psi(\mathbf{r})$ is to use the steepest descent method:

$$\psi_{n+1}(\mathbf{r}) = \psi_n(\mathbf{r}) - \alpha \nabla_{\psi_n} \epsilon_m, \tag{2.5}$$

where α_n is a parameter which determines how far along the steepest descent direction (given by $-\nabla_{\psi_n} \epsilon_m$) we want to update ψ_{n+1} by. We can determine the

optimal value for α_n by doing an exact line search: this involves evaluating the error metric at a few trial values of α_n:

$$\epsilon_m(\alpha_n) = \sum_{\mathbf{q}} \left\{ |\mathcal{F}[\psi_n(\mathbf{r}) - \alpha_n \nabla_{\psi_n} \epsilon_m)]| - \sqrt{I_j(\mathbf{q})} \right\}^2, \tag{2.6}$$

and using the α_n which minimizes the error the most. Alternatively we can try an inexact line search using the Wolfe conditions [25, 26], or be lazy (and save on computational effort) and pick some value for α_n for all n. The laziness method many times works acceptably well, but will still require us to pick an α_n value that is appropriate to the scaling of $\psi_n(\mathbf{r})$ and $\nabla_{\psi_n}\epsilon_m$. This can be done by using the exact line search method once or twice every so often to see what kind of typical α_n values are returned, and use something on this order.

To get an intuitive idea of how the steepest descent method works in practice, we look at a reduced dimensionality problem to define the error metric. In the following algorithms, we look at the gradients of an error metric with respect to a *real* valued function ψ; if ψ is complex valued, what should be done is to compute the gradient of an error metric with respect to the real and inaginary parts of ψ, i.e. $\nabla_\psi \epsilon = (\nabla_{\psi_R} + i\nabla_{\psi_I})\epsilon$, where $\psi = \psi_R + i\psi_I$, $\mathrm{Re}\{\psi\} = \psi_R$ and $\mathrm{Im}\{\psi\} = \psi_I$. Here, we use as error metric the Rosenbrock "banana" function:

$$\epsilon(x, y) = (1 - x)^2 + 100(y - x^2)^2. \tag{2.7}$$

Some plots of this function are shown in Fig. 2.1. This function is widely used for exploring the performance of gradient descent optimization algorithms. It is highly non-convex (effectively meaning that there are many points which are approximately stationary, i.e. where the error metric gradient is close to zero) and is useful for performance testing gradient descent algorithms by looking at how well the algorithm under investigation can avoid being trapped in local minima in error metric space. The Rosenbrock "banana" function has a narrow parabolic valley, as seen in Fig. 2.1, with the global minimum being located at $(x, y) = (1, 1)$ where $\epsilon(x, y) = 0$.

A typical outcome of trying to use the steepest descent method is shown in Fig. 2.2; starting from outside the valley at where the cyan circle is located in

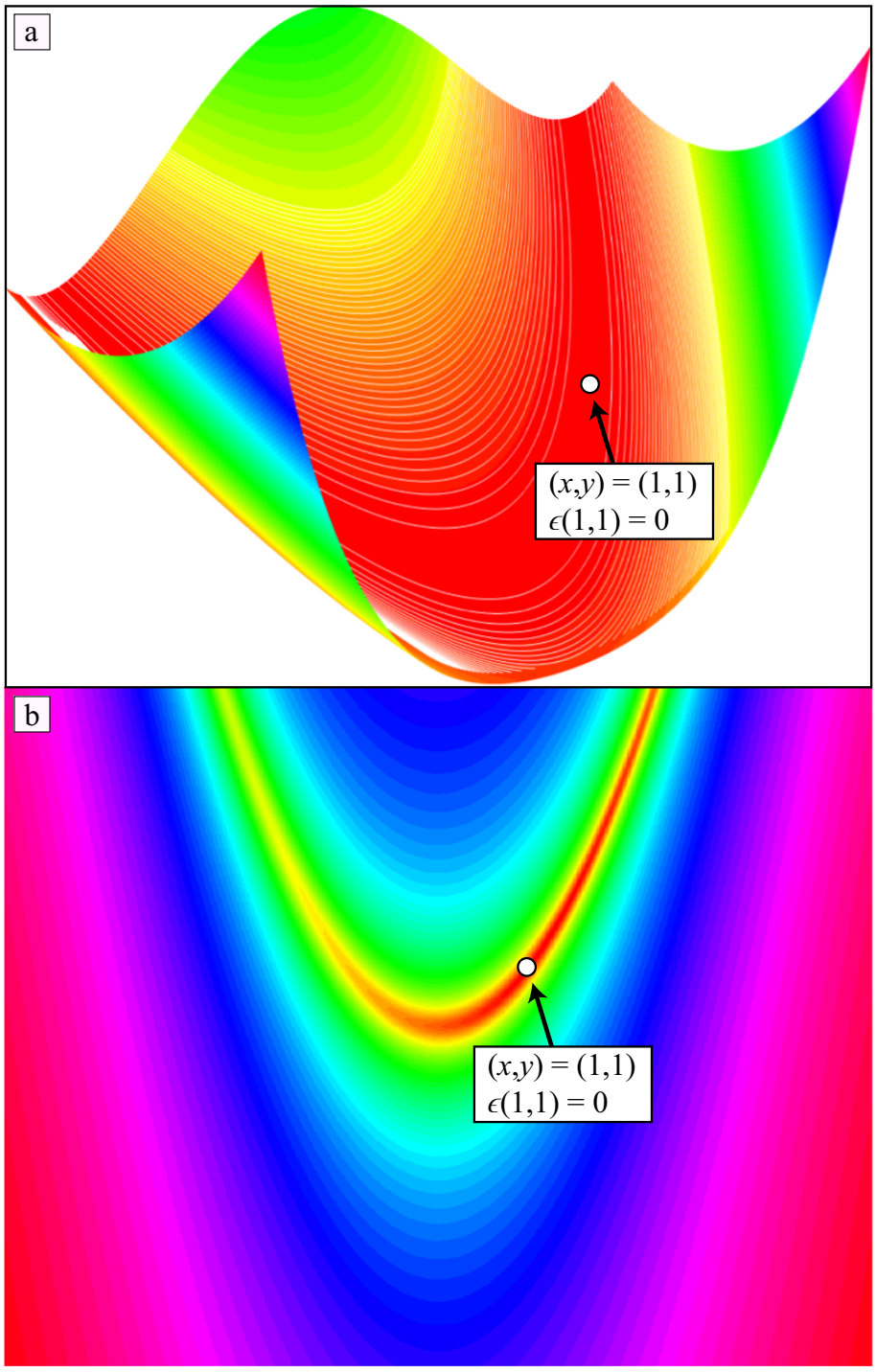

Figure 2.1: The Rosenbrock "banana" function. It is a standard test error metric for evaluating the performance of gradient descent algorithms. The top frame shows a three dimensional view of the function while the bottom frame shows a birds eye view. The global minima is located at $(x, y) = (1, 1)$ while all along the valley the gradient of the error metric is close to negligible.

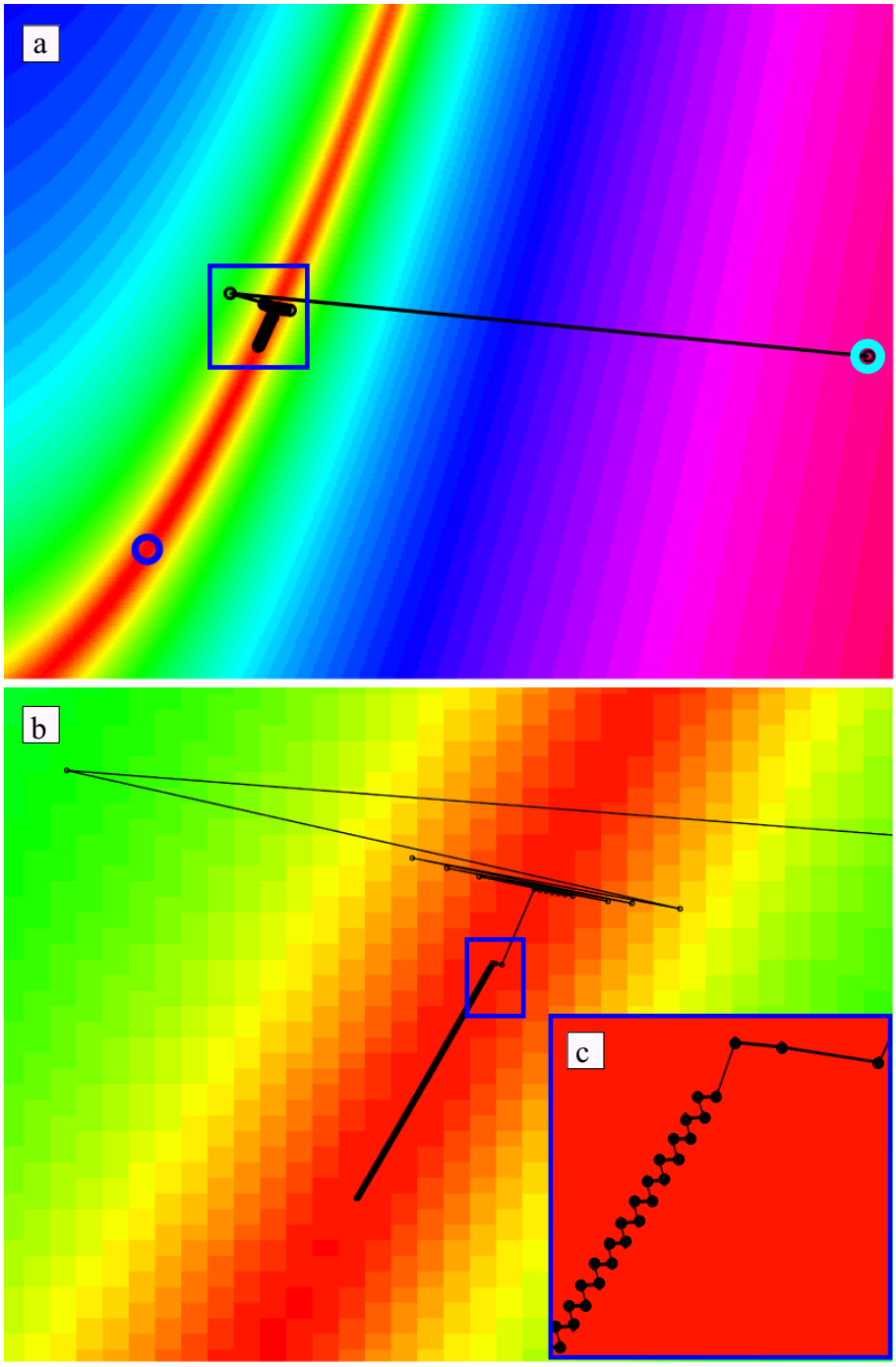

Figure 2.2: A typical trajectory of the steepest descent method of trying to get to the global minimum at $(x, y) = (1, 1)$ starting from a random location (here the cyan circle in (a)). We quickly travel into the valley, but get stalled there and barely make any progress due to the zigzagging trajectory the steepest descent methods takes us in.

Fig. 2.2a, we follow the steepest descent direction down into the valley with in a few iterations. What happens once we are in the valley is shown in Fig. 2.2b-c. Since inside the valley the error metric landscape is almost flat (i.e. is a continuous region of local minima), the trajectory of the steepest descent method becomes almost completely stalled. It proceeds in a zigzag fashion as seen in Fig. 2.2c, staying within the valley. This type of behavior of the steepest descent method is *very* typical when the error metric landscape is non-convex and has abundant local minima; for the type of optimization problems we encounter in CXDI the steepest descent method almost *always* becomes stagnated in this fashion making it next to useless.

In the steepest descent method, what often happens when stuck in local minima or valleys like in Fig. 2.2 is that we repeatedly move in the same direction over many different iterations. For example in Fig. 2.2c, the trajectory in the valley of the banana function is a zigzag shape and we end up going in almost the same direction every other iteration. What ends up happening there is that movement in one direction will partially or wholly undo the progress made in the movement of the previous iteration. A better way of traversing error metric space in situations like this is to pick a set of directions which minimizes this "destructive interference", i.e. find some "non-interfering" directions which will allow us to make better headway when stuck in error metric space features like the valley in Fig. 2.2. To do this, we introduce the *conjugate gradient* method [14, 15, 26, 27]. The conjugate gradient algorithm is given by:

$$\begin{aligned} \Lambda_0(\mathbf{r}) &= -\nabla_{\psi_0}\epsilon \\ \Lambda_n(\mathbf{r}) &= \beta_n\Lambda_{n-1}(\mathbf{r}) - \nabla_{\psi_n}\epsilon \\ \psi_{n+1}(\mathbf{r}) &= \psi_n(\mathbf{r}) + \alpha_n\Lambda_n(\mathbf{r}), \end{aligned} \tag{2.8}$$

where α_n is the step length taken in the conjugate direction Λ_n and this step length can be found in an identical manner as in the steepest descent method by using some type of line search routine. The conjugate directions Λ_n are determined by computing a β_n parameter given by the Fletcher-Reeves method:

$$\beta_n^{FR} = \frac{\Delta\psi_n^T\Delta\psi_n}{\Delta\psi_{n-1}^T\Delta\psi_{n-1}}, \tag{2.9}$$

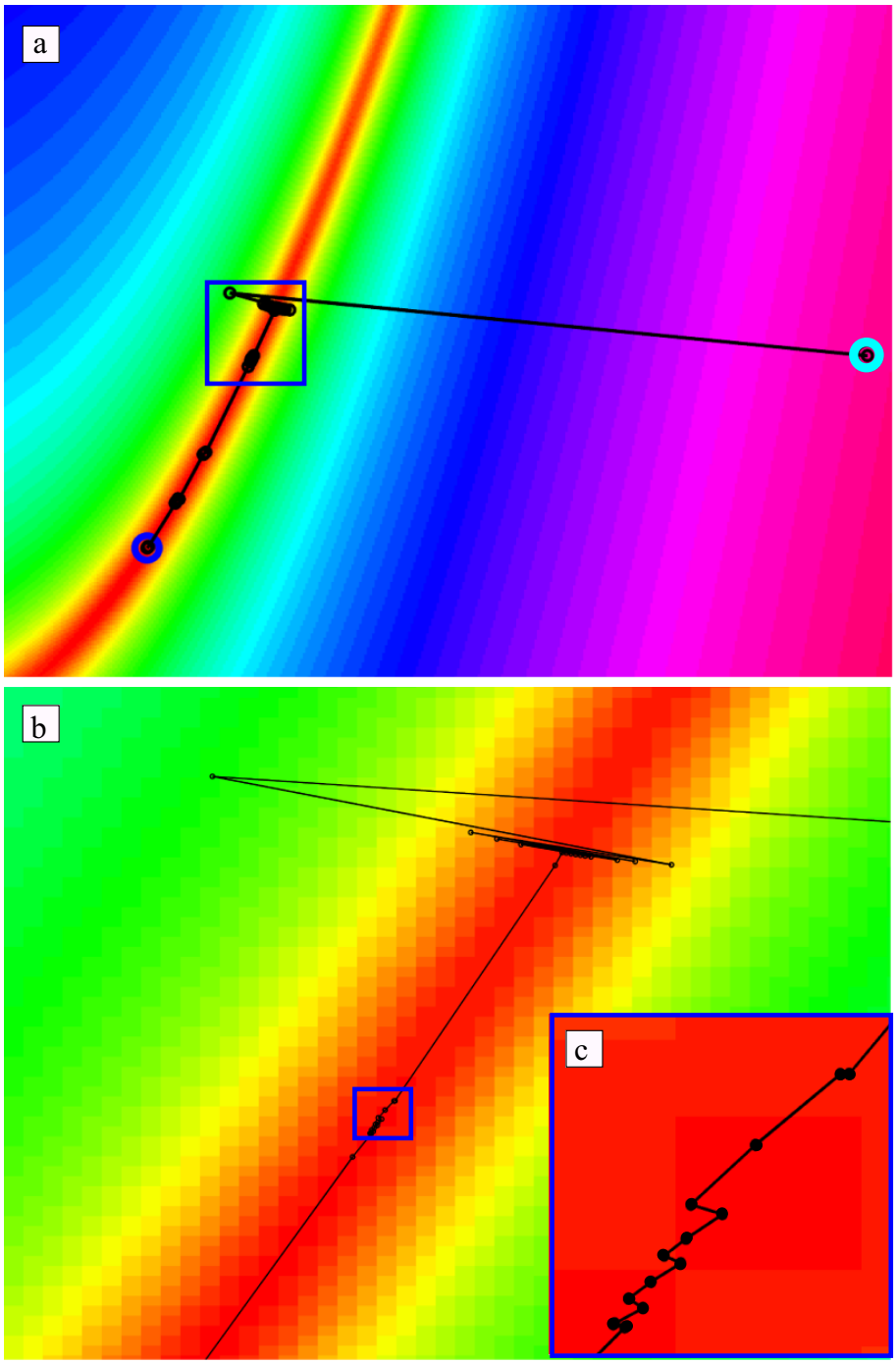

Figure 2.3: A typical trajectory of the conjugate gradient method when trying to get to the global minimum at $(x, y) = (1, 1)$ starting from a random location (here the cyan circle in (a)). We quickly travel into the valley, and due to only allowing steps in *conjugate* directions make much better progress than when using steepest descent directions.

or by the Polak-Ribière method:

$$\beta_n^{PR} = \frac{\Delta\psi_n^T(\Delta\psi_n - \Delta\psi_{n-1})}{\Delta\psi_{n-1}^T\Delta\psi_{n-1}} \tag{2.10}$$

where T denotes the transpose, $\Delta\psi_n = -\nabla_{\psi_n}\epsilon$, and for the Polak-Ribière method we check for and enforce that $\beta_n^{PR} = \max(0, \beta_n^{PR})$ each iteration, i.e. when β_n^{PR} becomes negative, we choose the descent direction to be the steepest descent for that iteration and begin calculating anew β_n^{PR} for subsequent iterations. Generally, the Polak-Ribière method performs better when dealing with non-quadratic features in error metric space, but in practice experimentation should be performed to see which methods works best.

Starting from the same location as in steepest descent case in Fig. 2.2 at the cyan circle, we again travel into the valley within a few iterations. What happens once we are in the valley is shown in Fig. 2.3b-c. Rather than following a zigzag trajectory in which we travel in nearly identical directions every other iteration, by only traveling in conjugate directions given by Λ_n instead we make much better headway and there are only a few locations within the valley where the trajectories are zigzagging and even then only for tens of iterations before finding a direction which gives very good progress.

The previous two gradient descent methods use only first derivative information to update a search direction. A popular class of algorithms which use both first and second derivative information are the "quasi-Newton" type methods, which are based on Newton's method for finding stationary points. The essence of Newton based optimization methods is to use local curvature of error metric space information to predict and travel a more direct route to stationary points. They are called quasi-Newton because second derivatives (the Hessian matrix) are not explicitly computed; instead the Hessian is iteratively computed and refined based on first derivative information obtained from gradient calculations. Probably the most popular quasi-Newton method is the Broyden-Fletcher-Goldfarb-Shanno (BFGS) method. The steps taken in the BFGS method are given in Eq. 2.11, where again α_n is a step length taken in the $\mathbf{p}_n$ direction and can be determined from some line search method. The inverse Hessian H_n^{-1} is determined in the last step in Eq. 2.11, with H_0^{-1} usually chosen to be the identity matrix. One issue

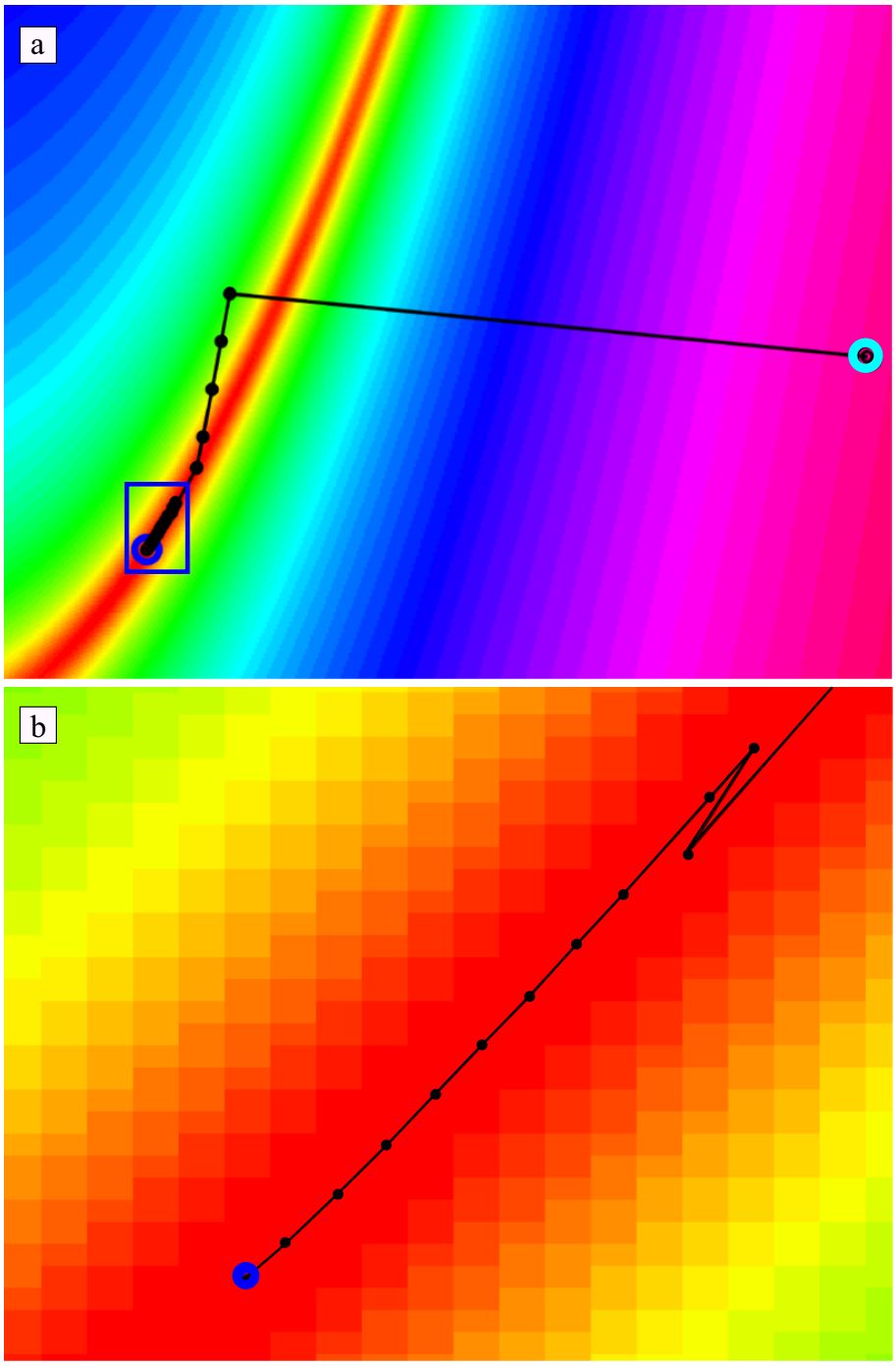

Figure 2.4: A typical trajectory of the BFGS method when trying to get to the global minimum at $(x, y) = (1, 1)$ starting from a random location (here the cyan circle in (a)).

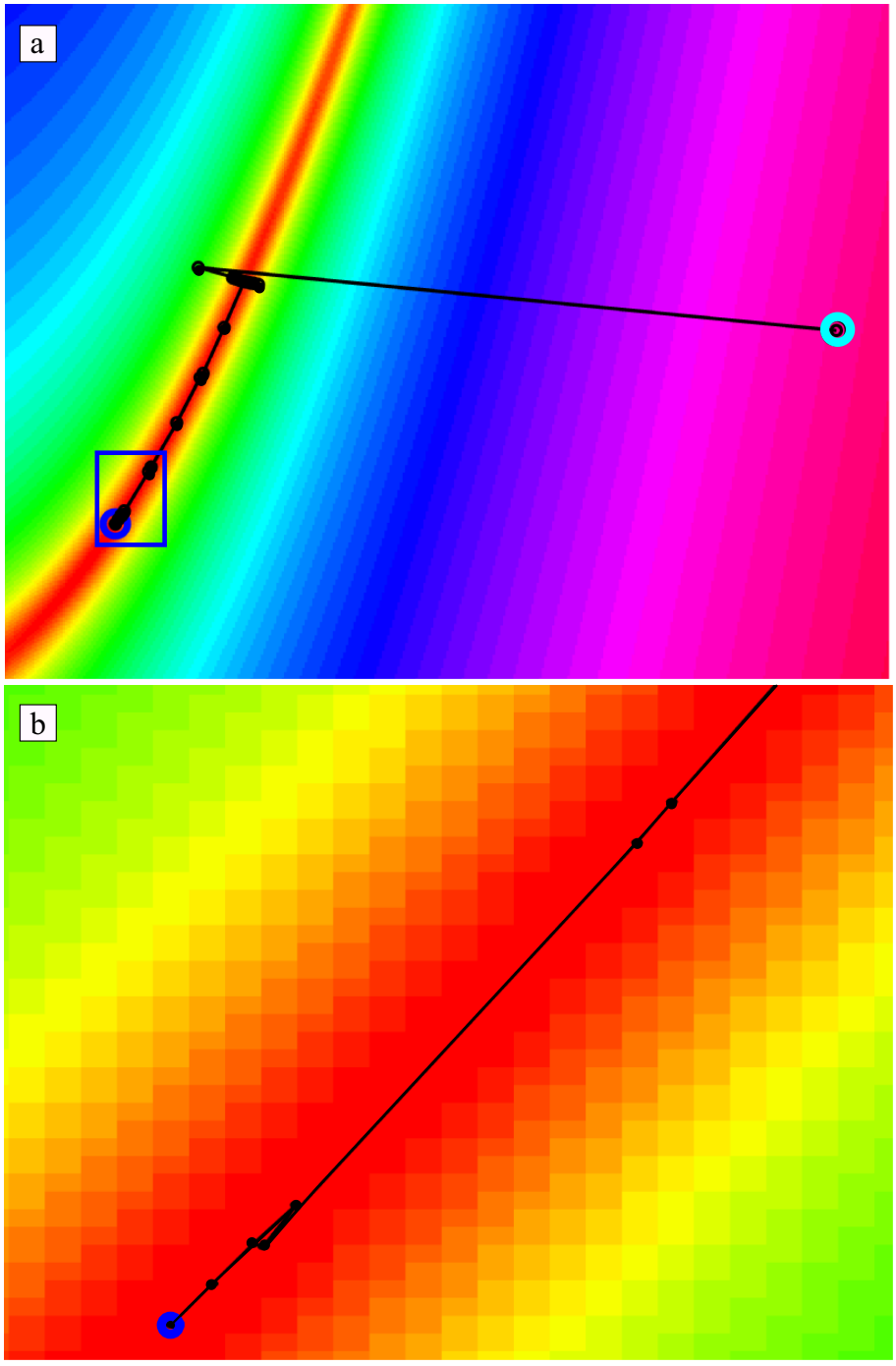

Figure 2.5: A typical trajectory of the L-BFGS method when trying to get to the global minimum at $(x, y) = (1, 1)$ starting from a random location (here the cyan circle in (a)).

that is immediately alarming is that for problems with many variables, like we have in CXDI, the memory requirements are approximately the square of that encountered using the previous two gradient descent methods, meaning that the BFGS method is possibly impractical for use in CXDI. However there is a limited memory version of the BFGS method called the L-BFGS method which does not attempt to compute an approximate Hessian every iteration. Instead the L-BFGS method stores in memory only a few vectors that represent an approximation for the Hessian implicitly and in a recursive fashion. It keeps track of a history of ψ_n and $\nabla_{\psi_n}\epsilon$ updates (usually 5 to 10 of them) for use in the recursive Hessian approximation; see reference [28] for details. The trajectory of the BFGS method when again starting from the cyan circle as in the steepest descent and conjugate gradient methods in shown in Fig. 2.4 and in Fig. 2.5 for the L-BFGS method.

$$
\begin{aligned}
&\mathbf{p}_n = -H_k^{-1}\nabla_{\psi_n}\epsilon \\
&\psi_{n+1}(\mathbf{r}) = \psi_n(\mathbf{r}) + \alpha_k \mathbf{p}_n \\
&\mathbf{s}_n = \alpha_k \mathbf{p}_n \\
&\mathbf{y}_n = \nabla_{\psi_{n+1}}\epsilon - \nabla_{\psi_n}\epsilon \\
&H_{n+1}^{-1} = H_n^{-1} + \left(1 + \frac{\mathbf{y}_n^T H_n^{-1}\mathbf{y}_n}{\mathbf{s}_n^T\mathbf{y}_n}\right)\frac{\mathbf{s}_n\mathbf{s}_n^T}{\mathbf{s}_n^T\mathbf{y}_n} - \frac{\mathbf{s}_n(H_n^{-1}\mathbf{y}_n)^T + H_n^{-1}\mathbf{y}_n\mathbf{s}_n^T}{\mathbf{s}_n^T\mathbf{y}_n}
\end{aligned}
\tag{2.11}
$$

2.2 Projections onto constraint sets

Another way of approaching the problem of finding the configuration for the elements of the $\psi(\mathbf{r})$ array which minimize the error metric given by Eq. 2.4 is to use a family of algorithms known as "projections onto constraint sets" (POCS) algorithms. In these, rather than computing gradients or approximate Hessian matrices, projections π_i are defined which enforce that some guess for $\psi(\mathbf{r})$ satisfy some constraint; for example by applying π_m, the Fourier modulus constraint, $\pi_m\psi(\mathbf{r})$ now belongs to a set of possibilities for $\psi(\mathbf{r})$ which when propagated to the measurement location have the measured Fourier modulus measurement. If we also have identified other constraints on the sample which we can enforce, then we can apply them using the "alternating projections" algorthim, also known as the

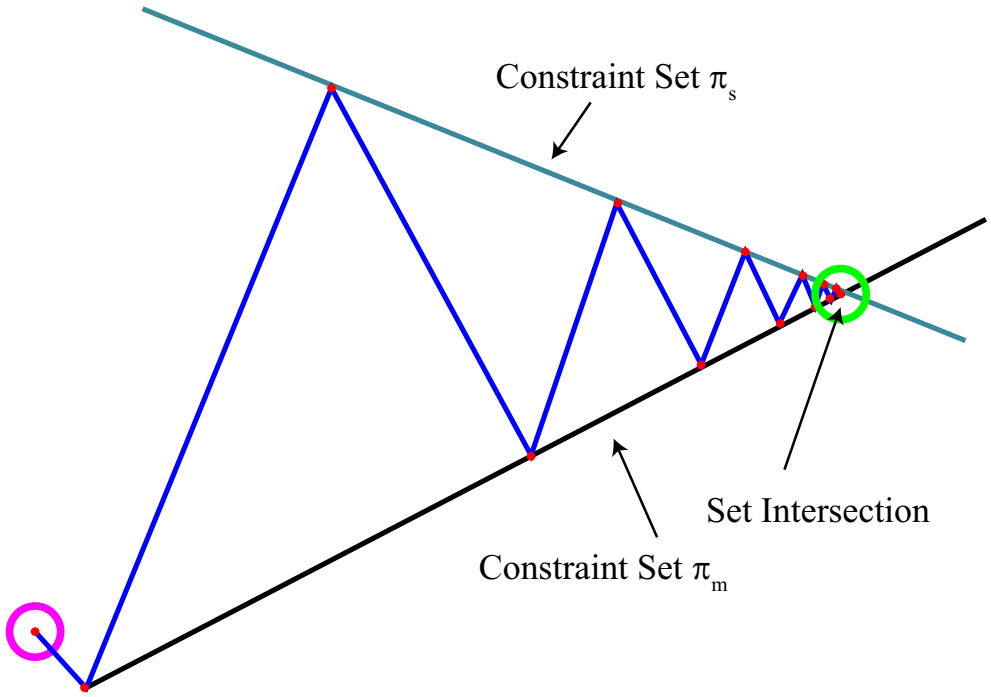

Figure 2.6: Example trajectory of the Error Reduction algorithm for simple convex constraint sets.

Error Reduction (ER) algorithm [7, 14, 29]:

$$\psi_{n+1}(\mathbf{r}) = \pi_s \pi_m \psi_n(\mathbf{r}), \tag{2.12}$$

where π_s is some constraint on the sample where the sample is located, as opposed to the Fourier modulus constraint which is a constraint on the sample exit wave-field when propagated to the measurement location; these "real space" constraints which π_s represents will be explored more in Section 2.3.

When the constraint sets are convex in Euclidian space, an example of how the ER method converges to the intersection of the sets is shown in Fig. 2.6. From the initial guess denoted by the magenta circle, it travels to the closest element of the π_s set and from there to the closest element of the π_m set, and repeats this until the intersection of the sets is reached. At this intersection both constraints are simultaneously satisfied, and for problems encountered in CXDI this generally means that a unique reconstruction of the sample has been achieved [9, 11]. The problem we run into when using the ER method in CXDI is that the Fourier modulus constraint set is *non-convex* [14], and it is also possible that we

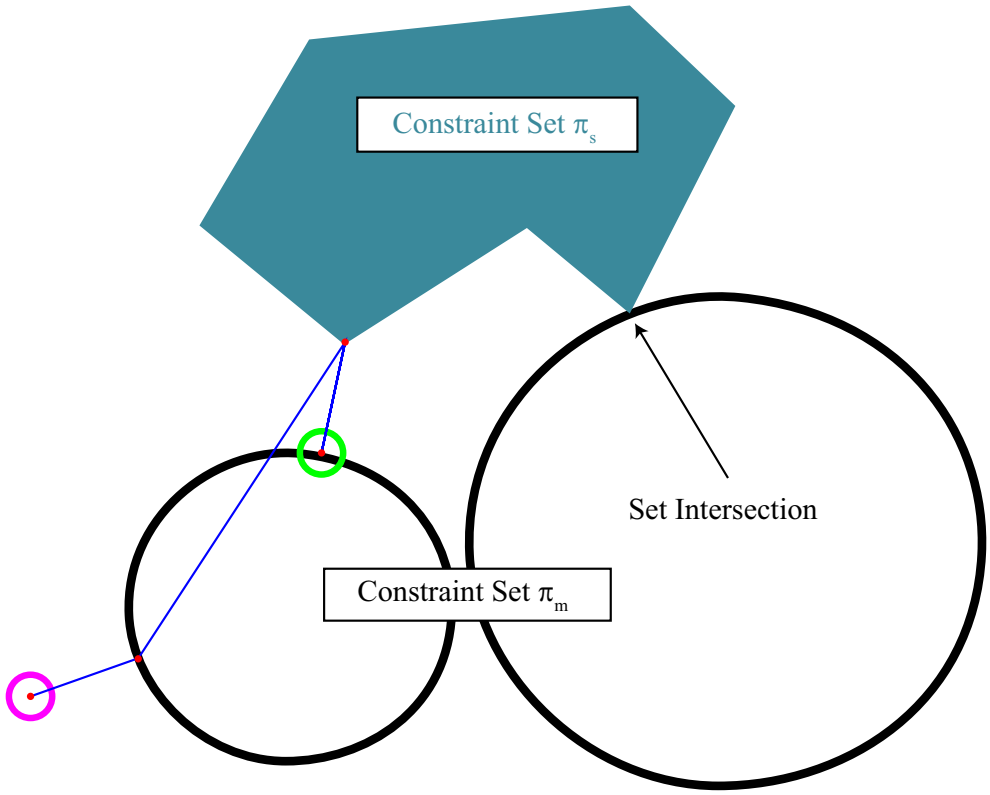

Figure 2.7: Example trajectory of the Error Reduction algorithm for non-convex constraint sets. Stagnation at some location where both constraint sets are close but not intersecting can occur.

have imperfect knowledge of real space constraints possibly causing the real space constraint set to also be non-convex. When this happens, ER usually gets trapped at a location where both constraint sets are close but not intersecting, as seen in Fig. 2.7. In context of the gradient of an error metric algorithms encountered in the previous section, non-convexity means that local minima exist and that like the steepest descent method, the ER algorithm gets easily stuck in these local minima and stagnates. The ER algorithm is actually identical to the steepest descent method when the optimal step length α_n is chosen which takes the current ψ_n iterate to the location of minimum error along the current descent direction [10, 14].

A POCS algorithm which copes much better with non-convexity is the Fienup Hybrid Input-Output (HIO) [8] algorithm. This algorithm originated from control-feedback theory and mixes the current iterate of $\pi_m\psi_n$ with the previous ψ_n

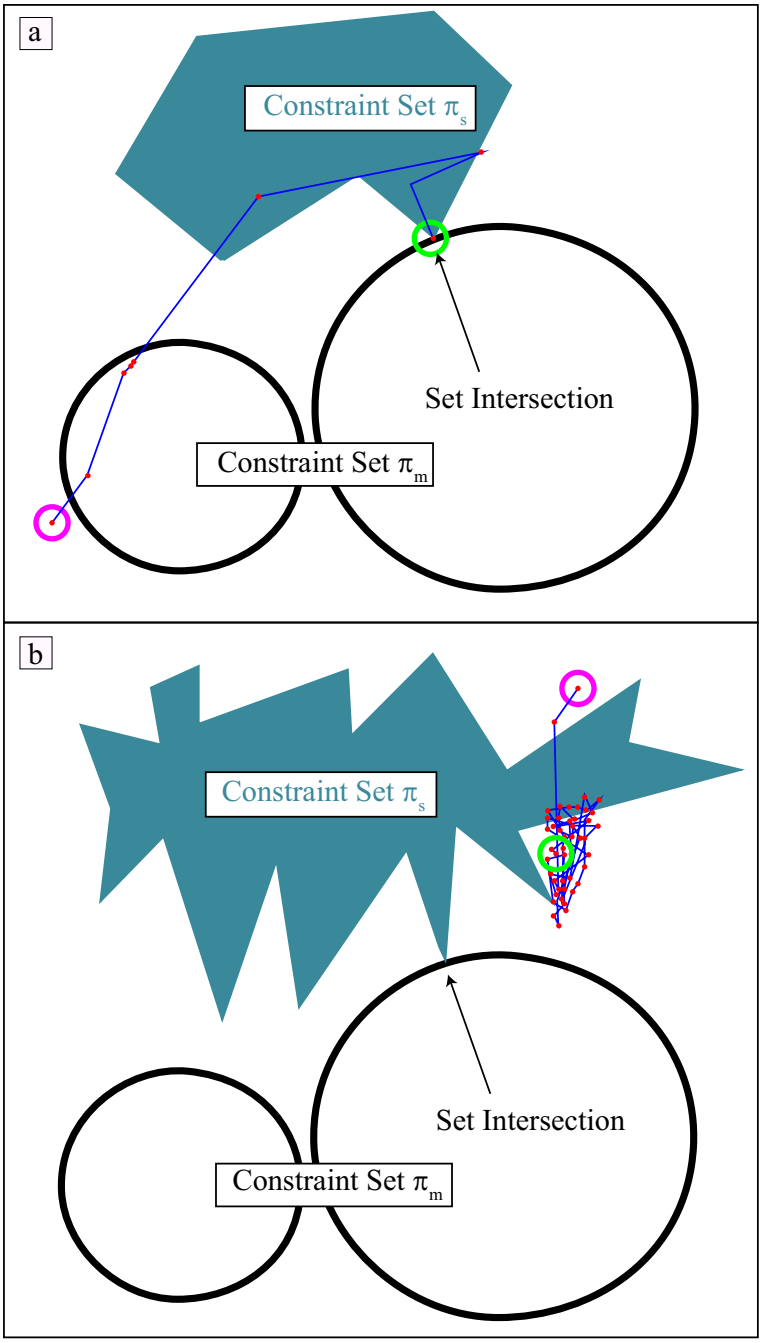

Figure 2.8: Example trajectories of the HIO algorithm for non-convex constraint sets. The top frame shows behavior when the non-convexity is not too severe, while the bottom frame shows behavior when the constraint sets are highly non-convex. In this case, even HIO can stagnate and not reach the intersection of the sets.

(via some β parameter) in such a way that positive feedback is built up when the current ψ_n is stuck in the current location for too long which in turn "encourages" the current ψ_n to move away from where it is currently stuck:

$$\psi_{n+1}(\mathbf{r}) = \begin{cases} \pi_m \psi_n(\mathbf{r}) & \mathbf{r} \in S, \\ (1 - \beta \pi_m \psi_n(\mathbf{r})) & \mathbf{r} \notin S. \end{cases} \tag{2.13}$$

Typically, β is chosen to be somewhere between 0.5 and 1.0, although it's possible to determine the optimal β every iteration by solving a sub-optimization problem [15]. This, combined with the fact that the trajectory the HIO algorithm causes ψ_n to take is not direct to either of the constraint sets but is a more "round about" trajectory means that HIO has the ability to avoid and get out of local minima in error metric space, as seen in Fig. 2.8. However it too can become stagnated when the degree of non-convexity is too great as seen in Fig. 2.8.

The current state of the art of POCS algorithms is the difference map (DM) [30, 31]:

$$\begin{aligned} \psi_{n+1}(\mathbf{r}) &= \psi_n(\mathbf{r}) + \beta \left[\pi_1 f_2 - \pi_2 f_1\right] \psi_n(\mathbf{r}) \\ f_1 &= (1 + \gamma_1) \pi_1 + \gamma_1 \\ f_2 &= (1 + \gamma_2) \pi_2 + \gamma_2 \end{aligned} \tag{2.14}$$

The DM algorithm can be viewed as a chaotic dynamical system with solutions defined as fixed points. It has a remarkable ability to escape local minima and by experimenting with the adjustable parameters β, γ_1, and γ_2, the DM algorithm is able to tackle a wide range of optimization problems, for example the graph coloring problem, Sudoku, spin glass ground states, and protein folding [31]. Additionally, it can handle any number of constraints, as opposed to only two as in the HIO method, by using a "divide and concur" strategy [32]. Some example trajectories of the difference map, using random numbers every iteration between 0 and -2 for γ_1, with $\gamma_1 = -\gamma_2$, and for β random numbers every iteration between 0 and 2, are shown in Fig. 2.9. The DM can be considered a meta algorithm, as many other specialized algorithms used in CXDI and non-convex optimization theory can be derived from it, based on the choices for β, γ_1, and γ_2 [14]. For example, with $\gamma_1 = -1$ and $\gamma_2 = \beta^{-1}$, $\pi_1 = \pi_s$ as the real space constraint and $\pi_2 = \pi_m$ as the Fourier modulus constraint, the DM algorithm reduces to the HIO algorithm.

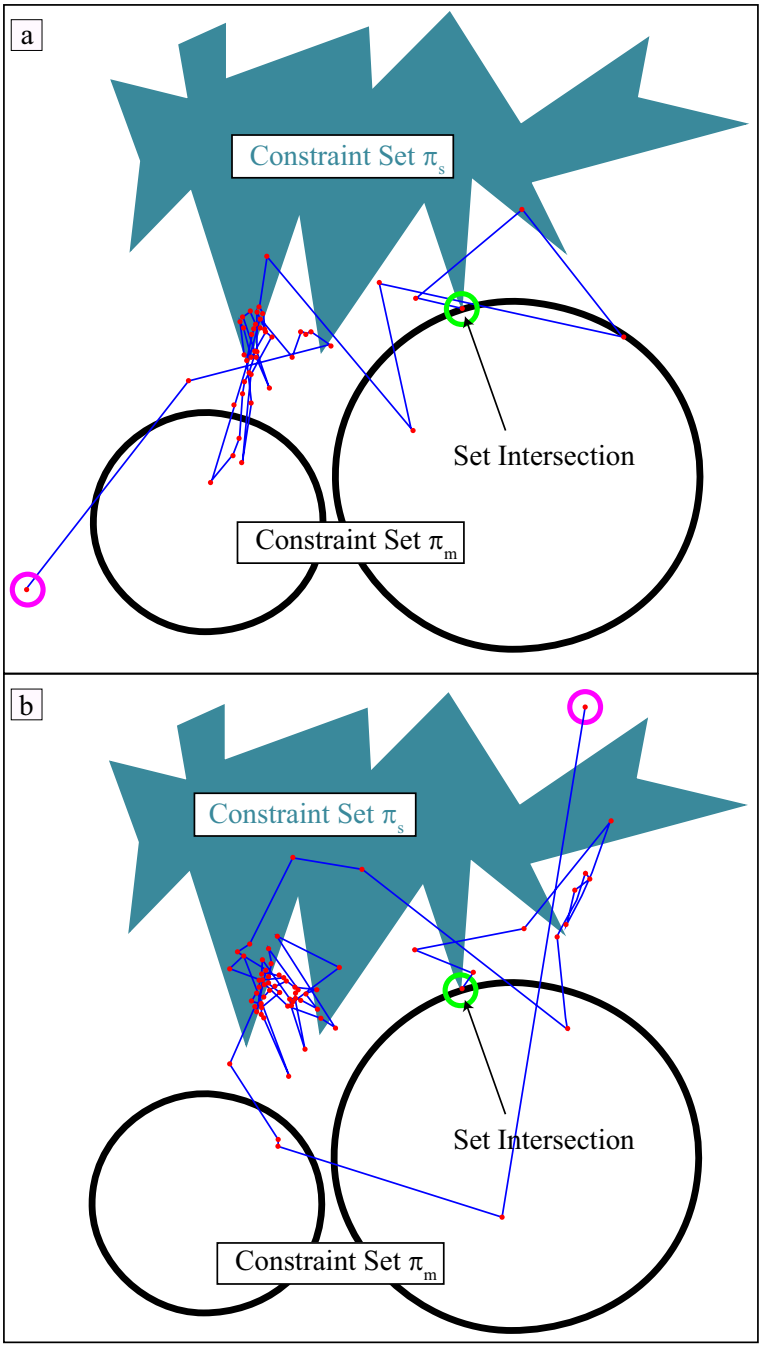

Figure 2.9: Example trajectories of the DM algorithm using random numbers every iteration between 0 and -2 for γ_1, with $\gamma_1 = -\gamma_2$, and for β random numbers every iteration between 0 and 2.

2.3 Real space constraints

Methods for iteratively solving the phase problem have been discussed in the previous two sections. Two strategies have been put forward: define an error metric and use gradient descent methods to locate the global minimum in error metric space, or to define constraint "projections" which when applied to a guess for the sample exit wave ψ_n modify the guess so that it satisfies this constraint (and is a member of the set of all possible ψ_n which satisfy the constraint), and iteratively find the intersection of the sets for all defined projections. The diffraction measurement is one constraint on the sample we will always be able to use. This section identifies and discusses other constraints on the sample, and how they are typically used and enforced in practice.

2.3.1 Support

When we are investigating a sample which is smaller than the illuminating x-ray beam, a constraint on the sample which can be used is the *support* constraint. The support constraint is simply knowledge about the maximum size of an object, for example the circle surrounding the sample in Fig. 2.10. Take for example the simulated sample in Fig. 2.10b. We can use the fact that we know the sample is smaller than a circle of some diameter as a support. Starting from a guess for the sample, as shown in Fig. 2.10c, and enforcing the Fourier modulus constraint π_m, we arrive at what is shown in Fig. 2.10d. What is done next is to enforce the support constraint by zeroing all pixels outside if the dashed blue circle from Fig. 2.10b, as seen in Fig. 2.10e. Generally the support constraint is only useful when it is "tight", i.e. it is close to the true sample shape; efficient and reliable methods exist for iteratively refining and determining the support [33].

2.3.2 Beam curvature

Another real space constraint available is to use some sort of structured beam to illuminate a sample and use this to modify the current iterate for the sample exit wave [16, 17]. This constraint can be used when the sample is both

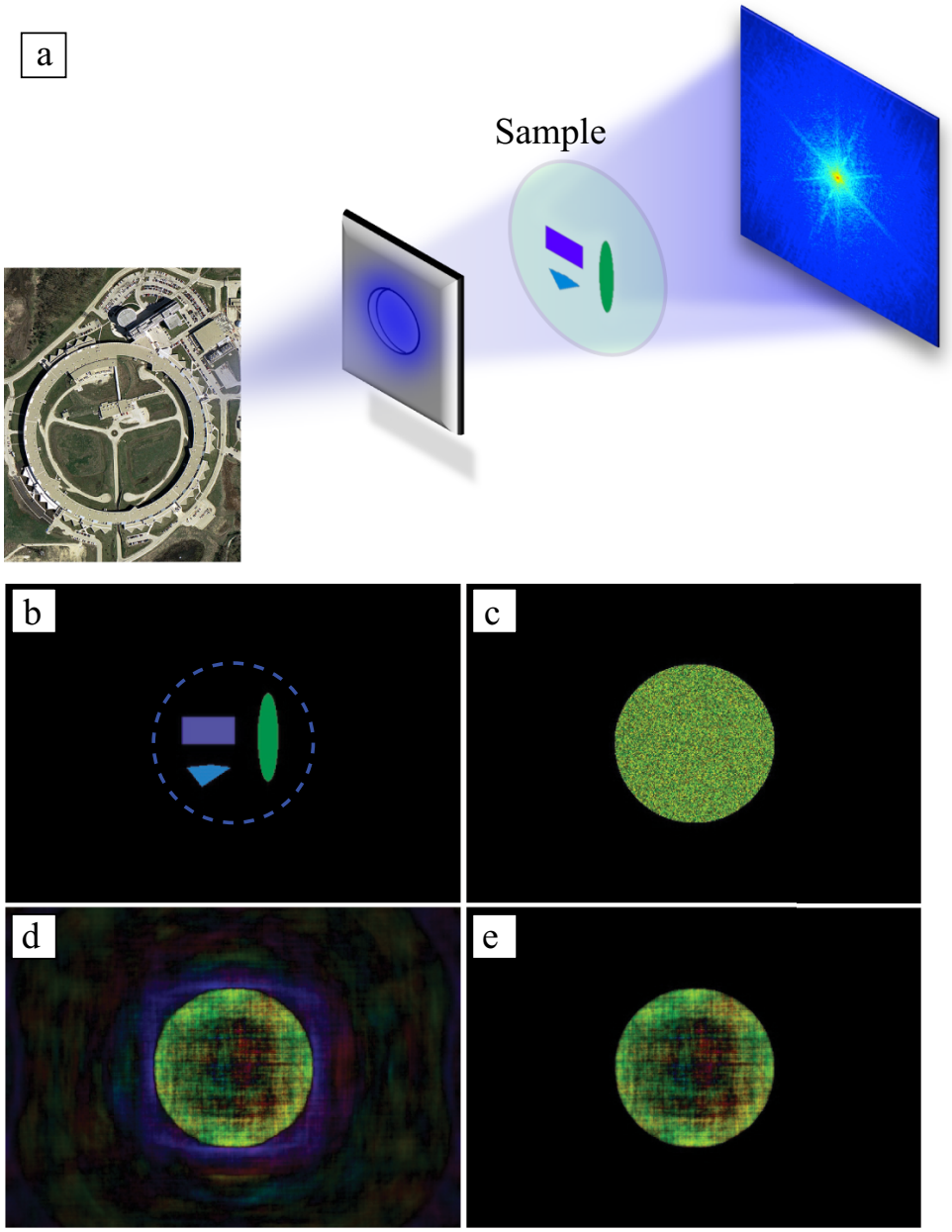

Figure 2.10: (a) Experimental geometry for when using the support constraint is possible: when the sample is smaller than the illuminating x-ray beam. (b) Usually we know that the sample cannot be larger than some maximum size (here the dashed blue circle), and so we can enforce that any iterates of the guessed exit wave ψ_n must satisfy this size limit. (c) A guess for the sample which satisfies the support constraint. (d) The result when the Fourier modulus constraint π_m is applied to the guessed sample in (c). (e) Enforcing the support constraint π_s on the result in (d).

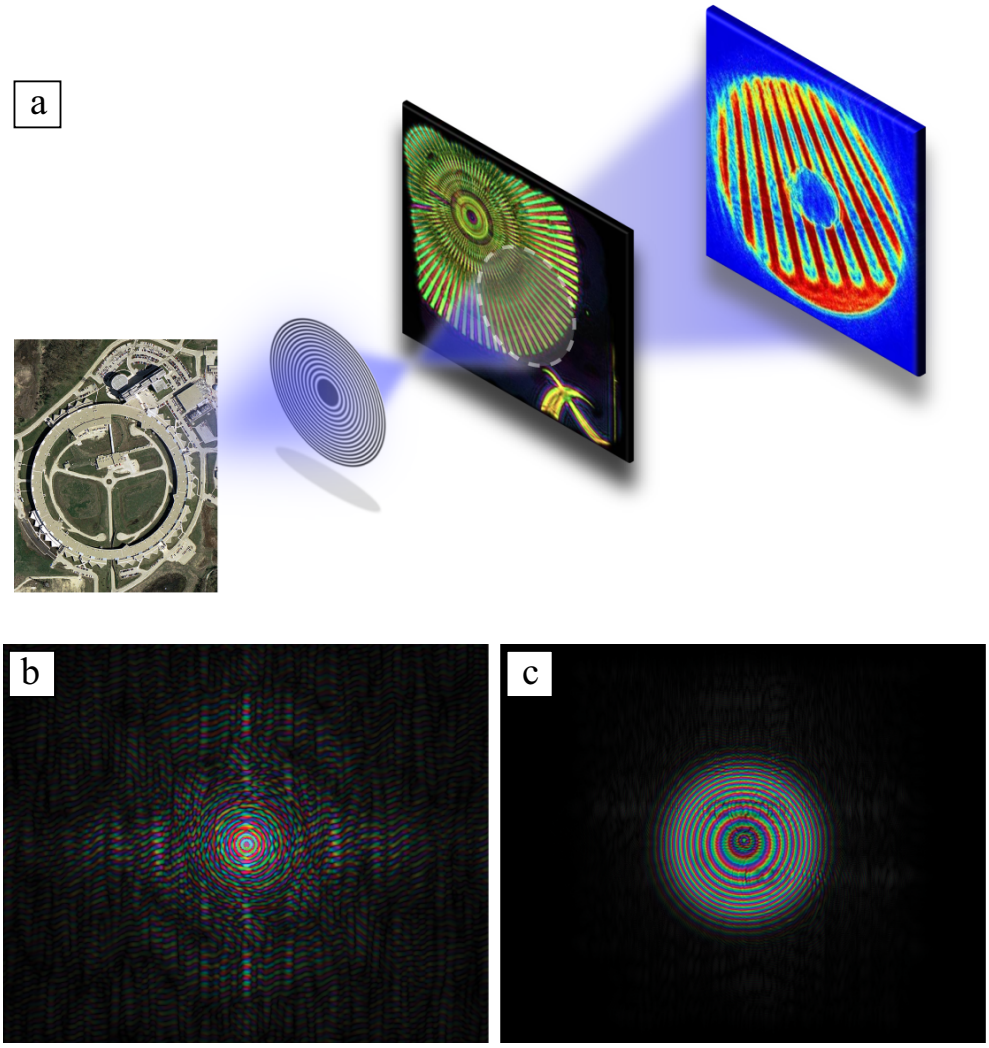

Figure 2.11: (a) Experimental geometry for when using the curved beam constraint is possible. In this case, the sample is larger than the x-ray illumination but it is also possible for the sample to be smaller than the x-ray beam. (b) The x-ray wavefield at the focus. At this focal plane, the wavefield is the Fourier transform of the wavefield just after the focusing optic, and so the quadratic phase curvature is gone. (c) However, by moving away from the focal plane downstream of the sample, the quadratic phase curvature becomes significant again. In (c) we see that the curvature is such that many 2π phase wraps occur. The effect of this is that many symmetries of the Fourier transform of the sample that exist when using plane wave illumination, such as translational and π rotation symmetries, are removed when using a beam with phase curvature. This in turns results in far greater convergence properties.

larger or smaller than the incident x-ray beam. What a structured beam means in this context mainly has to do with what its "phase" looks like. For example, a Fresnel zone plate introduces a quadratic phase curvature to the x-ray illumination function. In Fig. 2.11, a sample is placed a little ways downstream of the focal plane. This means that the quadratic phase curvature of the x-ray beam is significant; the effect of this is that many symmetries of the Fourier transform of the sample that exist when using plane wave illumination, such as translational and π rotation symmetries, are removed when using a beam with phase curvature. Due to the symmetries that are removed, numerical convergence rates of the iterative algorithms used are greatly enhanced; it appears that using curved beam illumination causes the constraint sets (or alternatively the error metric landscape) to be much less non-convex. An added effect of using an x-ray illumination which has phase curvature and is expanding is that a pseudo inline holographic region is created; this can be seen in the "donut" region in the detector plane in Fig. 2.11 where we see a magnified region of the illuminated region on the sample with additional Fresnel fringing effects. This is useful experimentally as we in effect have a low resolution image of the object and can quickly identify interesting features of the sample in almost real time. Once an interesting feature has been zeroed in on, we can then collect higher spatial frequency diffraction (located outside the "donut" region) which can be used to get at higher spatial resolution features on the sample by using the iterative algorithms in the previous section.

2.3.3 Ptychography

Another real space constraint that can be used is to scan the sample so that many adjacent and overlapping regions on the sample are illuminated and collect diffraction from each of these scan locations [13, 34, 35]; an example of this scanning scheme is shown in Fig. 2.12. This method of collecting data is known as "ptychography". The essence of this real space constraint is that there are regions on the sample from which we have collected many different and independent diffraction measurements. In other words, we can have multiple Fourier modulus constraints π_1, π_2, ..., π_N constraining regions on the sample which are covered

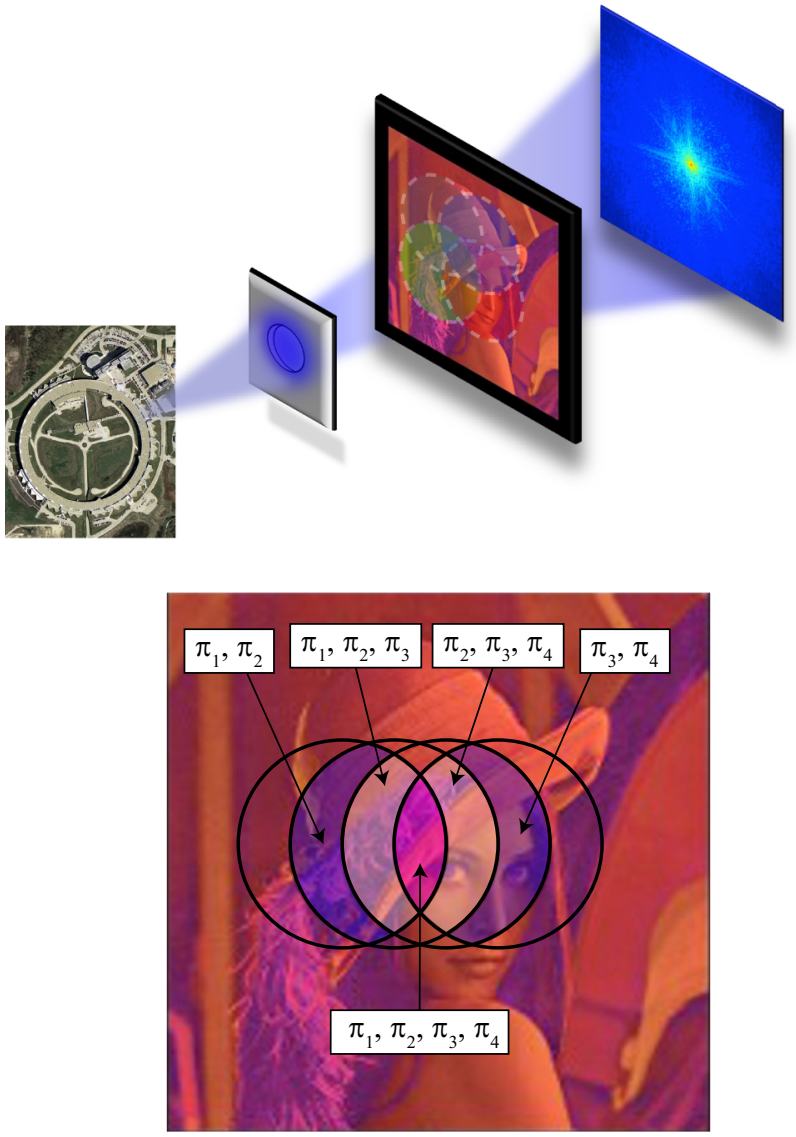

Figure 2.12: An experimental arrangement where multiple diffraction patterns are recorded by scanning at adjacent and overlapping regions on the sample. By recording diffraction in this way we have multiple Fourier modulus constraints π_1, π_2, ..., π_N which constrain a common region on the sample. For example, the purple common regions have two independent diffraction patterns constraining them, the white common regions have three, while the pink common region has four. Other than the true configuration of variables within these common regions, there are vanishingly small other possible configurations which simultaneously satisfy all the Fourier space constraints. It appears that ptychography causes the optimization problem to approach an almost convex type of problem.

by multiple scan locations. These common regions on the sample are highly constrained by the multiple diffraction patterns; there are very few potential solutions for these common regions that simultaneously satisfy all the corresponding Fourier modulus constraints. The next chapter will be completely devoted to further details of the ptychographic method of collecting diffraction, so nothing more will be said here.

Chapter 3

Ptychographic phase retrieval

This chapter goes into details on the ptychographic sample space (or alternatively termed "real space") constraint of overlapping scan locations on the sample. This causes a situation where we can have multiple diffraction patterns which constrain a particular region on the sample. We explore what happens to the performance of ptychography when the diffraction data set (the Fourier space constraint) is degraded in some realistic ways, for example when significant noise is present in the diffraction data, and when regions of the diffraction are missing due to having to use a beamstop (to prevent damage to the detector when using ultra-brilliant x-ray beams). We also explore the situation where the scan locations on the sample are not perfectly known. We may not have perfect knowledge of these scan location due to e.g. sample thermal drift and/or vibrations or if the stages we use to move the sample around so it can be ptychpgraphically scanned are installed in such a way as to not scan in the fashion we expect, and we explore the consequences of this and also explore an iterative method which allows us to discover the true scan positions.

3.1 Ptychographic phase retrieval

When the sample is larger than the incident x-ray beam, coherent diffraction can be collected ptychographically, and this type of experiment is termed ptychographic coherent diffractive imaging (pCXDI). In this scheme, the sample

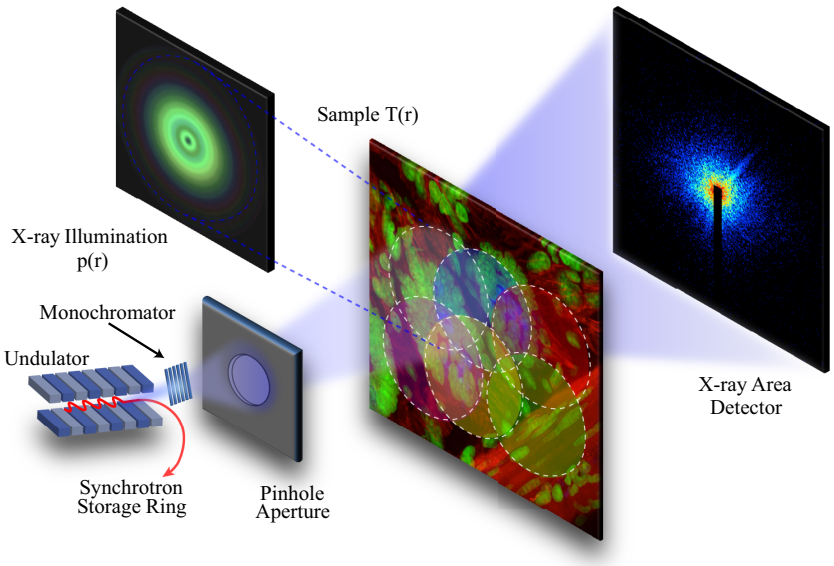

Figure 3.1: Schematic of an X-ray scanning CDI measurement in the forwards scattering geometry. Ptychographic diffraction patterns are recorded in the far-field by an X-ray area detector by scanning the sample, depicted as the overlapping circles.

is illuminated at multiple overlapping scan points, examples of which are shown in Fig. 3.1 as the overlapping circles. Here, a simulated x-ray wave field $p(\mathbf{r})$ is simulated by Fresnel propagating (using the methods of wavefield propagation from Chapter 1) a circular pinhole aperture, and it is incident on a sample with transmission function $T_0(\mathbf{r})$. The exit wave is defined as $\psi_0(\mathbf{r}) = p(\mathbf{r})T_0(\mathbf{r})$ (i.e. we assume the projection approximation), and is propagated to the detector by taking its Fourier transform. The CCD area detector then measures the modulus square of this; the phase is lost. The sample is then moved by some $\mathbf{r}_1$ so that a neighboring but overlapping region with exit wave $\psi_1(\mathbf{r}) = p(\mathbf{r})T_1(\mathbf{r})$ can be illuminated, where $T_1(\mathbf{r}) = T(\mathbf{r} - \mathbf{r}_1)$, and this can be repeated for further $\mathbf{r}_j$, $j \in \mathbb{Z}$, so that we have the desired total field of view on the sample.

The purpose of collecting diffraction from overlapping regions is that it results in a stringent constraint on the sample. We have multiple independent diffraction measurements constraining a particular common region on the sam-

ple. This drastically reduces spurious numerical artifacts produced by imperfect measurements of the diffraction, e.g. shot noise or missing data due to a beam stop. The enantiomorph problem associated with single view diffractive imaging is also removed as the known scheme of scanning the sample removes the Fourier transform symmetries from which the problem arises, allowing for vastly improved convergence time. Ptychographic diffraction collection allows for arbitrarily large field of view imaging of extended samples. It also allows for the simultaneous determination of the sample transmission function and the x-ray wave field illuminating the sample [34, 35]. Furthermore, when a known test sample is inserted into a beam from optical elements not well characterized, it allows for a powerful and robust method of determining the full complex valued wave field produced by any x-ray optics upstream of the sample [36].

The algorithms used in pCXDI to extract the sample exit wave from an initial guess using experimental constraints can be formulated in terms of some sort of gradient descent of an error metric [10, 14, 37] (and popularly known as the ePIE algorithm) or as a "projections onto constraint sets" algorithm [31] (here the DM algorithm is used; see the previous chapter). In gradient descent type algorithms, an error metric ϵ_j is defined:

$$\epsilon_j = \sum_{\mathbf{q}} \left\{ |\Psi_{j,n}(\mathbf{q})| - \sqrt{I_j(\mathbf{q})} \right\}^2, \tag{3.1}$$

where $\Psi_{j,n}(\mathbf{q}) = \mathcal{F}[\psi_{j,n}(\mathbf{r})] = \mathcal{F}[p_n(\mathbf{r})T_{j,n}(\mathbf{r})] = \mathcal{F}[p_n(\mathbf{r})T_n(\mathbf{r}-\mathbf{r}_j)]$ is the n^{th} iterate of the exit wave at position $\mathbf{r}_j$ propagated to the detector, and the sum over $\mathbf{q}$ only includes nonzero pixels for each diffraction intensity $I_j(\mathbf{q})$ measurement. The gradient of the error metric with respect to the sample transmission function or the x-ray illumination function is then performed analytically [37]. This will generate an "update function" [35] with which we modify $T_{j,n}(\mathbf{r})$ and $p_n(\mathbf{r})$ in such a way as to iteratively travel to a location in error metric space that has minimum error. The rules for updating $T_{j,n}(\mathbf{r})$ and $p_n(\mathbf{r})$ using the ePIE algorithm [35] in this way are:

$$p_{n+1}(\mathbf{r}) = p_n(\mathbf{r}) + \alpha_p \frac{T^*_{j,n}(\mathbf{r})}{|T_{j,n}(\mathbf{r})|^2_{max}} (\pi_m \psi_{j,n}(\mathbf{r}) - \psi_{j,n}(\mathbf{r})), \tag{3.2}$$

and

$$T_{j,n+1}(\mathbf{r}) = T_{j,n}(\mathbf{r}) + \alpha_T \frac{p_n^*(\mathbf{r})}{|p_n(\mathbf{r})|^2_{max}}(\pi_m \psi_{j,n}(\mathbf{r}) - \psi_{j,n}(\mathbf{r})), \tag{3.3}$$

where α_p and α_T are suitable step lengths along the steepest descent directions, $*$ is the complex conjugate, and π_m is defined by Eq. 2.2. To update $T_j(\mathbf{r})$ and $p(\mathbf{r})$ using the Difference Map (DM), we do:

$$\psi_{j,n+1}(\mathbf{r}) = \psi_{j,n}(\mathbf{r}) + \pi_m[2p_n(\mathbf{r})T_{j,n}(\mathbf{r}) - \psi_{j,n}(\mathbf{r})] - p_n(\mathbf{r})T_{j,n}(\mathbf{r}), \tag{3.4}$$

and from this $\psi_{j,n+1}(\mathbf{r})$ update:

$$T_n(\mathbf{r}) = \frac{\sum_j p_n^*(\mathbf{r}+\mathbf{r}_j)\psi_{j,n}(\mathbf{r}+\mathbf{r}_j)}{\sum_j |p_n(\mathbf{r}+\mathbf{r}_j)|^2}, \tag{3.5}$$

and

$$p_n(\mathbf{r}) = \frac{\sum_j T_n^*(\mathbf{r}-\mathbf{r}_j)\psi_{j,n}(\mathbf{r})}{\sum_j |T_n(\mathbf{r}-\mathbf{r}_j)|^2}. \tag{3.6}$$

The Fourier space constraints of diffraction from overlapping scan positions appear to be stringent enough to allow us to find the *global* minimum in error metric space. This is implied from the fact that, for diffraction not significantly degraded by Poisson shot noise or missing data regions, we always recover the same $T_j(\mathbf{r})$ and $p(\mathbf{r})$ even with very different initial guesses of these functions. That we are at a global minimum is further supported by the use of the DM algorithm to recover $T_j(\mathbf{r})$ and $p(\mathbf{r})$ since DM has well noted ability to escape local minima and find the global minimum [14, 31]. In practice, it is usually not a bad idea to use both the DM and ePIE. For example, we could run 100 iterations of ePIE initially and then follow with 100 iterations of DM, and the repeat this use of both methods one after the other until we have found a solution for $T_j(\mathbf{r})$ and $p(\mathbf{r})$, the criteria for which is if the error metric given by Eq. 3.1 has reached a suitably low value or has been at the same value for many iterations, indicating e.g. somehow stagnation in a local minimum or that we've truly found the global minimum.

3.2 Ptychography and noisy diffraction

One thing we always have to deal with regarding experimental data is noise. We explore here how diffraction that is degraded by noise, here Poisson "shot"

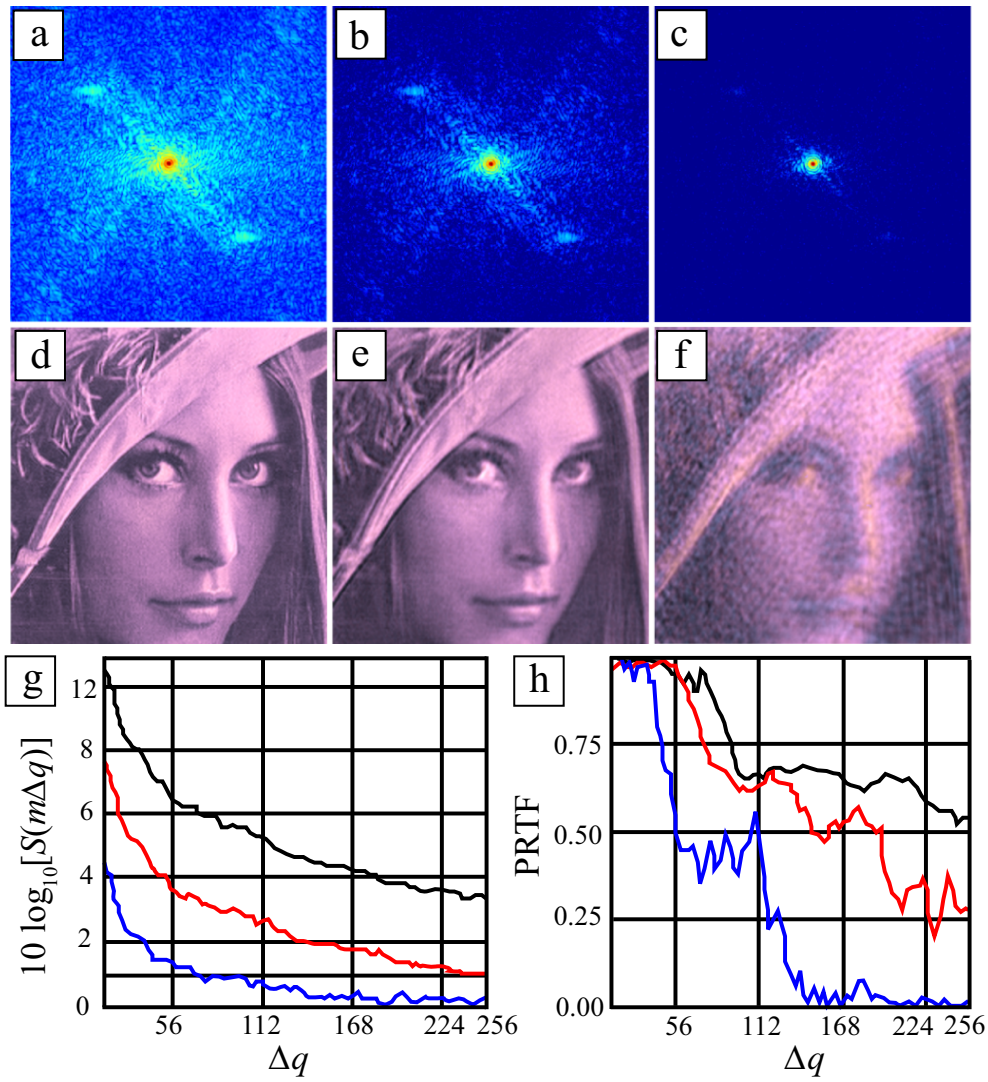

Figure 3.2: (a) An undegraded, noise-free diffraction pattern with high SNR which corresponds to the black curve in (g). (b) A degraded diffraction pattern with moderate SNR vs **q** (momentum transfer) which corresponds to the red curve in (g). (c) A heavily degraded diffraction pattern which corresponds to the blue curve in (g). (h) The PRTF of the reconstructions. Reconstructions (d, e, f) are from the high SNR diffraction (black curve in (g)), the moderate SNR diffraction (red curve in (g)), and the low SNR diffraction (blue curve in (g)), respectively.

noise, affects reconstruction quality. While there are many sources of noise which can affect the signal to noise vs. $\mathbf{q}$ (momentum transfer) of the measured diffraction, for example thermal noise of the electrons in the pixels of the CCD (also known as dark current), usually it is Poisson noise which is dominant. So, we define the signal to noise ratio (SNR) as

$$S(m\Delta q) = \frac{\sum_{m\phi} I_j(m\Delta q)}{\sum_{m\phi} \sqrt{I_j(m\Delta q)}}, \tag{3.7}$$

where $\Delta q = ak/z$, $k = 2\pi/\lambda$, λ is the x-ray wavelength, a is the detector pixel size, z is the distance from the sample to the detector, the sum over $m\phi$ means we sum over an annulus of 1 pixel width in $\mathbf{q}$ space, and m is a pixel index ranging from $m = 1$ (corresponding to the $\mathbf{q} = 0$ annulus) to $m = N/2$ (corresponding to the $\mathbf{q} = N\Delta q/2$ annulus). Using this definition, we look at three different SNR cases, as seen in Fig. 3.2, to generate ptychographic diffraction data sets, and reconstruct images for each of the SNR cases. Consider the three diffraction patterns in Fig. 3.2(a,b,c). The signal to noise vs. $\mathbf{q}$ is shown in Fig. 3.2g; the black curve corresponds the the diffraction in Fig. 3.2a, the red curve corresponds the the diffraction in Fig. 3.2b, and the blue curve corresponds the the diffraction in Fig. 3.2c. To quantify the quality of the reconstructions shown in Fig. 3.2(d,e,f) (corresponding the the diffraction patterns in Fig. 3.2(a,b,c) respectively), we use the concept of the Phase Retrieval Transfer Function (PRTF) [38], which is the ratio of the reconstructed diffraction magnitude (the absolute value of the Fourier transformation of the reconstructed specimen) to the measured diffraction magnitude as a function of $\mathbf{q}$:

$$\mathrm{PRTF}_j(m\Delta q) = \frac{\sum_{m\phi} |\mathcal{F}\{\langle T_j(\mathbf{r})p(\mathbf{r})\rangle\}|}{\sum_{m\phi} \sqrt{I_j(m\Delta q)}} \tag{3.8}$$

where $\langle\,\rangle$ denotes averaging over many independent reconstructions of $T_j(\mathbf{r})p(\mathbf{r})$. Typically a cutoff PRTF value is defined in the literature [39] as a value between 0.4 and 0.5, and here we take the value 0.5 as a cutoff. This means that any $\mathbf{q}$ values for which PRTF value fall below this cutoff we will not trust. In other words, this cutoff defines the spatial resolution of the reconstruction as $\pi/m_{max}\Delta q$, where $m_{max}\Delta q$ is the maximum value of $\mathbf{q}$ for which PRTF is above the cutoff value of

0.5. From Fig. 3.2h, the PRTF for the black curve is always greater than 0.5, which we would expect since this PRTF curve corresponds to the undegraded diffraction in Fig. 3.2a. This in turn means that the spatial resolution for this case is $\pi/(N\Delta q/2) = \Delta x$, where N is the array size used (for both rows and columns), and Δx is the pixel size in real space (see Eq. 1.2.1). The PRTF for the red curve (corresponding to the diffraction pattern shown in Fig. 3.2b) terminally falls below 0.5 at $m_{max}\Delta q \approx 195\Delta q$, meaning that we would trust spatial features up to $\pi/(195\Delta q) = 1.31\Delta x$. The SNR at this cutoff is ~ 1.5 dB. The PRTF for the blue curve (corresponding to the diffraction pattern shown in Fig. 3.2c) falls below 0.5 at about $56\Delta q$. Although there is a point where is rises back above 0.5 at about $112\Delta q$, this is likely due to the large variation of the PRTF between $56\Delta q$ and $112\Delta q$, so we ignore this point at $112\Delta q$. So for this case, we would trust spatial features up to $\pi/(56\Delta q) = 4.57\Delta x$. The SNR at this cutoff is also ~ 1.5 dB. This indicates that the DM is effective at reconstructing spatial features up to a SNR of between 1 to 2 dB.

3.3 Ptychography and missing data

In Fig. 3.3 we look at the effect of varying amounts of missing data. Missing data in diffraction measurements is common for intensely bright x-ray beams and the typical x-ray CCD detectors currently in use at synchrotrons. The type of (simulated) beamstop used in Fig. 3.3a is widespread using current direct-detection CCD detectors for x-ray coherent diffraction experiments; most experimenters using the scattering geometry shown in Fig. 3.1 would encounter something like this. The reconstruction corresponding to this diffraction pattern is shown in Fig. 3.3b. When the amount of missing data is not so severe, ptychography is usually able to "fill in" the information lost behind the beamstop. This is because typically the information content in ptychographic data sets is so rich, the information lost behind the beamstop is "in there somewhere" and the iterative algorithms are able to find it.

Next consider the beamstops in Fig. 3.3c and Fig. 3.3e. The amount of

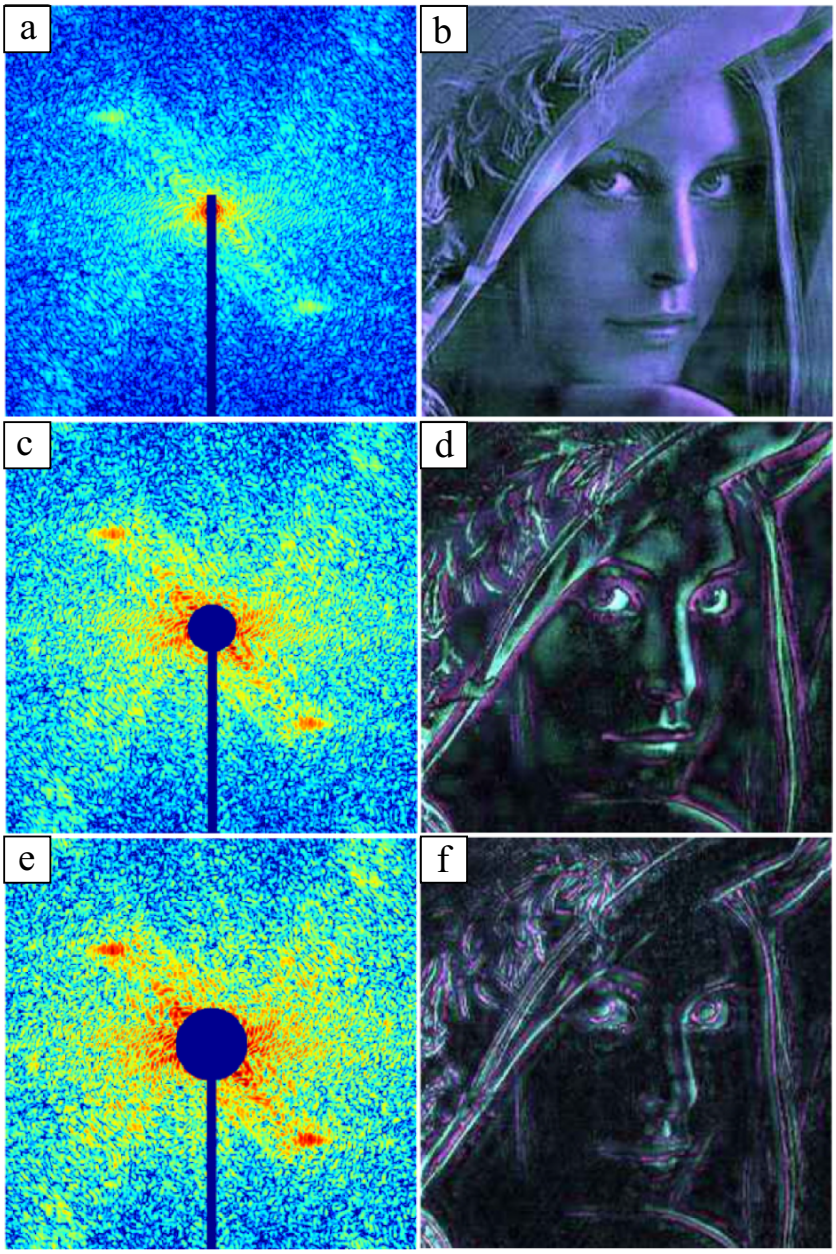

Figure 3.3: Performance of ptychography with varying amounts of missing data. The SNR for the diffraction patterns in frames (a, c, e) is the same as the black curve in Fig. 3.2g. Reconstructions in frames (b, d, f) correspond to representative diffraction patterns (a, c, e) respectively, and the diffraction patterns in frames (a, c, e) correspond to where 87%, 96% 97% of the integrated intensity has been lost, respectively. The performance of iterative algorithms to properly scale the missing data regions [40] for reconstructions in frames (d, f) is quite poor when the missing data regions get too large, and other methods should be explored.

missing data encountered in Fig. 3.3c and Fig. 3.3e might be encountered when we have a very weakly scattering sample and we must attenuate virtually the entire low spatial frequency signal to be able to take a long enough exposure to access the higher spatial frequency signal. For the reconstructions in Fig. 3.3d and Fig. 3.3f, the missing data regions due to the beamstop are treated as unconstrained, i.e. when applying the Fourier modulus constraint, we do not force these missing pixels to zero, but actually leave them alone, other than attempting to properly scale them using another iterative process which scales the missing data regions using scaling knowledge of the measured regions [40]. These reconstructions are clearly inadequate, and are in fact high pass filtered versions of the true object.

What is happening here is that in the missing data regions, algorithms we use to attempt to iteratively arrive at correct scaling [40] in the missing regions perform worse and worse as the missing data regions become larger. The method presented in reference [40] works very well if the missing data regions are relatively small, like in Fig. 3.3a, but begins to fail when we see relatively large missing data regions like in Fig. 3.3c and Fig. 3.3e. In practice, one could also overcome this iterative missing data scaling failure by using different beam stops, for example one large one like in Fig. 3.3e to collect high spatial frequencies with very long CCD exposures and another smaller beamstop with short exposures to collect the lower spatial frequency information. One can then use these two measurement regimes to initially update the object using the low frequency diffraction and after this use the high frequency diffraction to update the object, or alternatively, patch the two measurement regimes together into one diffraction pattern.

3.4 Scan positions errors and pCXDI

Errors of assumed scan positions $\mathbf{r}_j$ can be caused by unaccounted for thermal drift and expansion of the experimental equipment when undertaking experiments when temperature is varied significantly. If using a beam which is on the order of tens of nanometers, ambient vibrations can also compromise our knowledge of the scan positions significantly. The ability of pCXDI to quickly and robustly

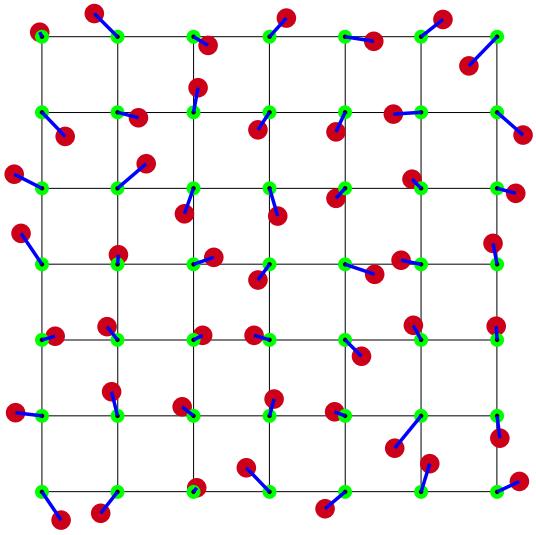

Figure 3.4: Schematic of how random errors are added to the assumed scan locations, denoted by the green circles. The red circles are the "true" scan locations. A uniform random number distribution is used to generate two random numbers within some bound, say ± 20 pixels, and each of these random numbers is added to the x and y component of the scan locations; this is done for all 49 scan locations.

converge to a solution for $T_j(\mathbf{r})$ and $p(\mathbf{r})$ is drastically degraded when errors in the assumed scan points come into play. This is explored in this section, as well as a scheme to iteratively solve for unknown scan position errors.

3.4.1 Effects of position errors on reconstructions

To illustrate this a simulated pCXDI experiment is performed. Diffraction is collected by scanning a simulated sample in an x-ray wavefield in the forward scattering geometry depicted in Fig. 3.1. A thin rectangular portion of each of these diffraction patterns is removed, to simulate the effect of a beam stop. Here, we donote a as the detector pixel size, λ as the x-ray wavelength, $k = 2\pi/\lambda$ the x-ray wavenumber, ℓ as the sample to detector distance, and the array size in the sample to detector Fourier transformation as $N^2 = 512 \times 512$. The field of view at the detector is $L_D = Na$, while the field of view at the sample plane is $L_S = \lambda\ell N/L_D$. Thus the real space pixel size at the sample plane is $\Delta x_S = L_S/N = \lambda\ell/Na$, and the Fourier space pixel size at the detector plane is found using the relation $\Delta q \Delta x_S = 2\pi/N$, resulting in $\Delta q = ka/\ell$. The x-ray illumination $p(\mathbf{r})$ has a

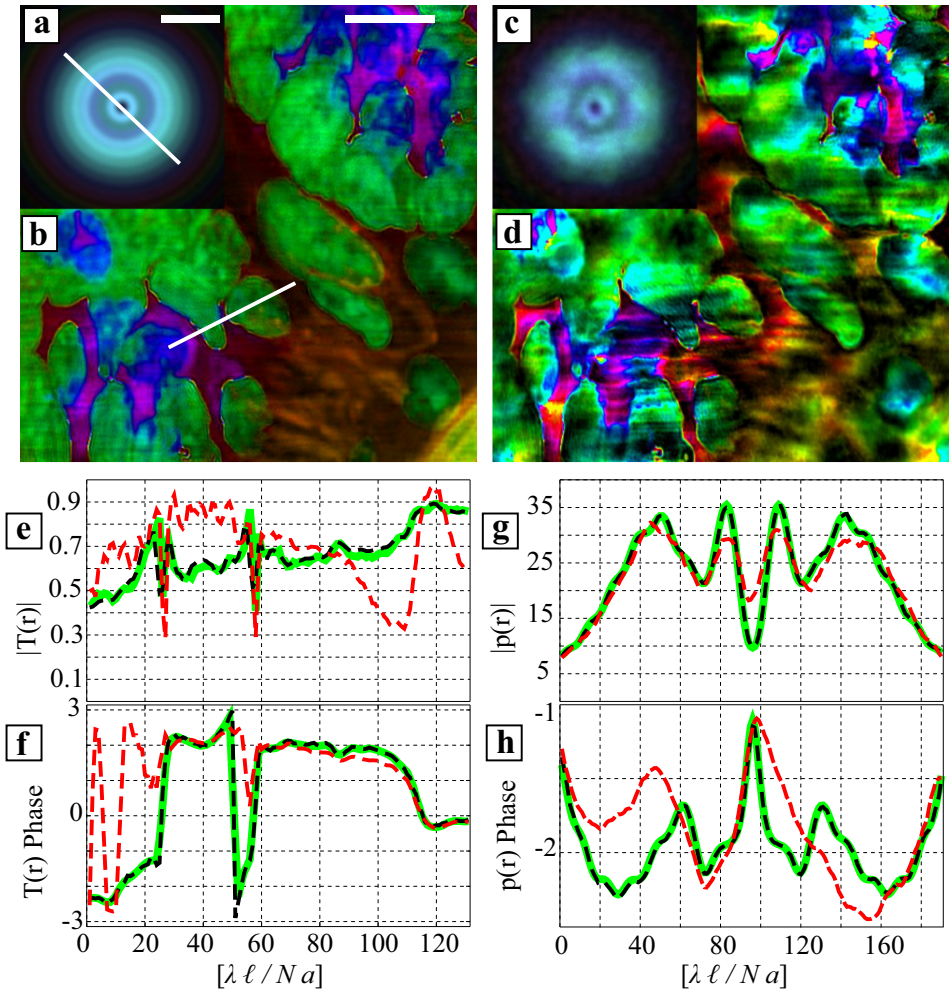

Figure 3.5: (a-b)Transmission function and probe reconstructions for when there is no error in scan position constraint. (c-d) Transmission function and probe reconstructions when random numbers over a range of $\pm$ 10 pixels are added to each scan position. (e) Lineout of $|T_j(\mathbf{r})|$. (f) Lineout of the phase of $T_j(\mathbf{r})$. (g) Lineout of $|p(\mathbf{r})|$. (h) Lineout of the phase of $p(\mathbf{r})$. In (e-f), the lineouts are along the white line indicated in (b), and in (g-h), the lineouts are along the white line indicated in (a). In (e-f), the red dashed line is the lineout for the reconstruction of the $T_j(\mathbf{r})$ with scan position errors, the black dashed line is the lineout for the reconstruction of the $T_j(\mathbf{r})$ without scan position errors, while the solid green line corresponds to the true $T_j(\mathbf{r})$. In (g-h), the red dashed line is the lineout for the reconstruction of the $p(\mathbf{r})$ with scan position errors, the black dashed line is the lineout for the reconstruction of the $p(\mathbf{r})$ without scan position errors, while the solid green line corresponds to the true $p(\mathbf{r})$. The scale bars in frames (a-b) are both $10^2 \lambda\ell/Na$.

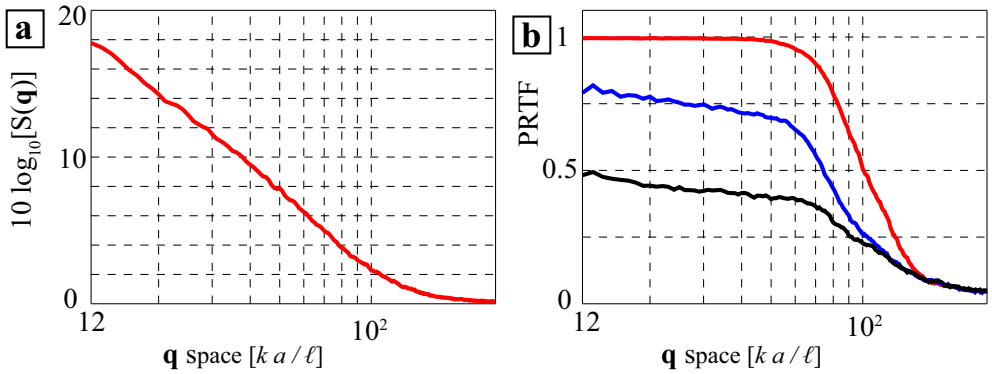

Figure 3.6: (a)The signal to noise of the ptychographic diffraction data set used in the reconstructions in Fig. 3.5. (b) The PRTF of the reconstructions. The red curve is for no errors, the blue curve is for a maximum ± 10 pixel random error added to the x and y components of each of the scan locations, and the black curve is for a maximum ± 20 pixel random error.

diameter of $\simeq$ 280 pixels, which in units of length is $D_{p(\mathbf{r})} \simeq 280\ L_S/N$. The overlap between adjacent scan positions is 75%, which corresponds to having the adjacent scan positions separated by a distance of $\simeq 70\ L_S/N$. In this way, 49 diffraction patterns are recorded scanning a 7×7 square grid.

When no errors in scan position knowledge are present, we ptychographically reconstruct $p(\mathbf{r})$ and $T(\mathbf{r})$ over all j as shown in Fig. 3.5a-b. The x-ray illumination function is shown in Fig. 3.5a, while the object transmission function is the image shown in Fig. 3.5b. In Fig. 3.5e-f, a lineout of $T(\mathbf{r})$ is taken (corresponding to the white line shown in Fig. 3.5b). The magnitude of $T(\mathbf{r})$ is shown in Fig. 3.5e while the phase of $T(\mathbf{r})$ is shown in Fig. 3.5f; both are shown as the dashed black line. The sold green curves in Fig. 3.5e-f are the true magnitude and phase of $T(\mathbf{r})$ respectively along the white line in Fig. 3.5b.

When scan position errors are present, we see significant degradation of the reconstructions of $T_j(\mathbf{r})$ and $p(\mathbf{r})$, as seen in Fig. 3.5c-d. What was done to simulate scan position errors is to add random numbers to the x and y coordinates for each of the 49 scan positions, and an example of what this might look like is shown in Fig. 3.4; in Fig. 3.5c-d, a uniform random number distribution over a range of ±10 pixels is used to add to the x and y coordinates, and so a maximum error of $10\sqrt{2}$ pixels is possible and the average position error is $\Delta r_{avg} \approx 8$ pixels.

Scan position errors have the effect of the degradation of the reconstruction over all spatial frequencies, as seen by computing the PRTF as seen in Fig. 3.6b. In this, the red curve is the PRTF of the reconstruction in Fig. 3.5a-b assuming no scan position errors, the blue curve is the PRTF of the reconstruction Fig. 3.5c-d assuming a maximum error of $10\sqrt{2}$ pixels, while the black curve is the PRTF of a reconstruction assuming a maximum error of $20\sqrt{2}$ pixels (reconstructions are not shown, they look too terrible). Referring back to Fig. 3.6b, we use a PRTF spatial frequency cutoff of ≈ 0.5, which for the red curve occurs at $\approx 100ka/\ell$. In Fig. 3.6a, the signal becomes lost in the noise back ground also at a momentum transfer value of $\approx 100ka/\ell$, telling us that the reconstruction spatial resolution is here signal to noise limited, as we would expect. However, with scan position errors, for the blue curve in Fig. 3.6b the 0.5 spatial frequency cutoff occurs at $75ka/\ell$ while the black curve never rises above the cutoff, illustrating how devastating even relatively small scan position errors can degrade reconstruction spatial resolution.

3.4.2 Iterative refinement of the overlap constraint

For the reconstructions shown in Fig. 3.5c-d, an average error in scan position of $\Delta r_{avg} \approx 8$ pixels is only $\sim 3\%$ of the 280 pixel probe diameter used, and is $\sim 11\%$ of the step size of 70 pixels used. For any experiment using a probe which is tens of nanometers in diameter, vibrations on the nanometer scale can be expected to seriously degrade the spatial resolution and quantitative index of refraction information of the reconstructions. However, ptychographic diffraction measurements are often information rich enough to allow us to take the gradient of Eq. 3.1 with respect to the assumed scan positions $\mathbf{r}_j$ and use an iterative process to correct for scan position errors [37]. We present here an effective method for incorporating iterative scan position correction into standard ptychographic reconstruction algorithms [34, 35]. The steepest descent method for correcting scan positions using an analytical calculation of the gradient of Eq. 3.1 with respect to the assumed scan positions $\mathbf{r}_j$ is given by:

$$\mathbf{r}_{j,n+1} = \mathbf{r}_{j,n} - \alpha_n \nabla_{\mathbf{r}_{j,n}} \epsilon_j, \tag{3.9}$$

For a little more computational effort but much better performance, we can also update the scan positions along the conjugate gradient descent directions:

$$\begin{aligned}\Lambda_{j,0} &= -\nabla_{\mathbf{r}_{j,0}}\epsilon_j \\ \Lambda_{j,n} &= \beta_n\Lambda_{j,n-1} - \nabla_{\mathbf{r}_{j,n}}\epsilon_j \\ \mathbf{r}_{j,n+1} &= \mathbf{r}_{j,n} + \alpha_n\Lambda_{j,n},\end{aligned} \tag{3.10}$$

where $\mathbf{r}_{j,0}$ are the initial assumed scan positions, α_n is a step length taken along the steepest descent or conjugate gradient directions, and β_n is calculated using either the Fletcher-Reeves or Polak-Ribière methods [14].

We integrate this conjugate gradient correction of the scan positions into the standard ptychographic reconstruction algorithms by using the $\mathbf{r}_{j,0}$ positions initially, and update $T_j(\mathbf{r})$ and $p(\mathbf{r})$ using either the DM or ePIE for some tens of iterations. Once this is done, we use these newly obtained $T_j(\mathbf{r})$ and $p(\mathbf{r})$ in the gradient calculation of Eq. 3.1, and update the $\mathbf{r}_j$ using Eq. 3.10. We choose the step length α_n by rescaling $\Lambda_{j,n}$ so that it is a unit vector, and either simply pick a value of α_n (say 2 pixels), or perform a line search along the $\Lambda_{j,n}$ direction by evaluating the error metric Eq. 3.1 at a few trial values of α_n:

$$\epsilon(\alpha_n) = \sum_{\mathbf{q}}\left\{|\mathcal{F}[p(\mathbf{r})T(\mathbf{r}-\mathbf{r}_{j,n}-\alpha_n\Lambda_{j,n})]| - \sqrt{I_j(\mathbf{q})}\right\}^2. \tag{3.11}$$

Since $\Lambda_{j,n}$ is a unit vector, some sensible trial values for α_n are say 1 pixel, 5 pixels, and 10 pixels along the conjugate gradient $\Lambda_{j,n}$ direction. Once we have found a value for α_n which gives us the smallest value of $\epsilon(\alpha_n)$, we use this α_n to update $\mathbf{r}_{j,n+1}$ as in Eq. 3.10. Next, we run the DM or ePIE to again update $T_j(\mathbf{r})$ and $p(\mathbf{r})$ for another ten iterations or so using the just updated $\mathbf{r}_{j,n+1}$, and repeat the above conjugate gradient scan position correction procedure again after this. That whole process is then repeated for however many iterations it takes to get the error metric to converge to zero.

It is inadvisable to update the $\mathbf{r}_{j,n}$ after every iteration of $T_j(\mathbf{r})$ and $p(\mathbf{r})$ updates; we have noticed that by doing this the $\mathbf{r}_{j,n}$ sometimes tend to travel back to previous values of $\mathbf{r}_{j,n}$. The topology of error metric space is crucially dependent on $T_j(\mathbf{r})$, $p(\mathbf{r})$, and $\mathbf{r}_{j,n}$, and by updating all three "back to back", we appear to

be more susceptible to stagnation in error metric space local minima. Also, the gradient of Eq. 3.10 with respect to the $\mathbf{r}_{j,n}$ requires four FFT evaluations (as opposed to the two FFT evaluations required by the DM and ePIE algorithms to update $T_j(\mathbf{r})$ and $p(\mathbf{r})$). So the time it takes to accomplish all the $\mathbf{r}_{j,n}$ updates is roughly twice as the $T_j(\mathbf{r})$ and $p(\mathbf{r})$ updates. By only updating the scan positions every ten or so iterations, the computational cost of incorporating $\mathbf{r}_{j,n}$ conjugate gradient updates is minimal, typically 10% to 20% longer than not incorporating the $\mathbf{r}_{j,n}$ updates.

3.4.3 Maximum recoverable scan position error

We next address the question of how well does the iterative scan position correction method in the previous section perform when the ptychographic diffraction data set is degraded in an experimentally realistic way. To explore this question, we explore in simulation two scenarios: one with a diffraction data set with good signal to noise out to almost the edge of the array, as seen in Fig. 3.7a, and another with diffraction which decays into the noise much quicker, as seen in Fig. 3.7b. The "good" signal to noise diffraction becomes lost in the Poisson distributed noise background at a momentum transfer value of $\approx 125ka/\ell$ where the signal to noise is between 1 to 2 dB, while the "poor" signal to noise diffraction becomes lost at $\approx 40ka/\ell$. For both diffraction data sets, a thin rectangular region extending up from the bottom is removed from the diffraction to simulate the effect of a beamstop; between %85 and %95 of the integrated diffraction intensity is lost behind the beamstop for both diffraction data sets.

Using the same simulation parameters as in Section 3.4, we start with using the "good" signal to noise diffraction and first use an assumed scan position step size of 70 pixels (which corresponds to 75% adjacent scan position overlap using a 280 pixel diameter x-ray illumination function). We use 1152 total iterations of the ePIE and DM algorithms solve for $T_j(\mathbf{r})$ and $p(\mathbf{r})$: (144 iterations of ePIE and 144 iterations of DM repeated 4 times each), and update probe positions every 12 iterations (total of 96 iterations). In this way, we discover that the maximum scan position error we can reliably solve for is when errors of $\approx 60\sqrt{2}$ pixels are added

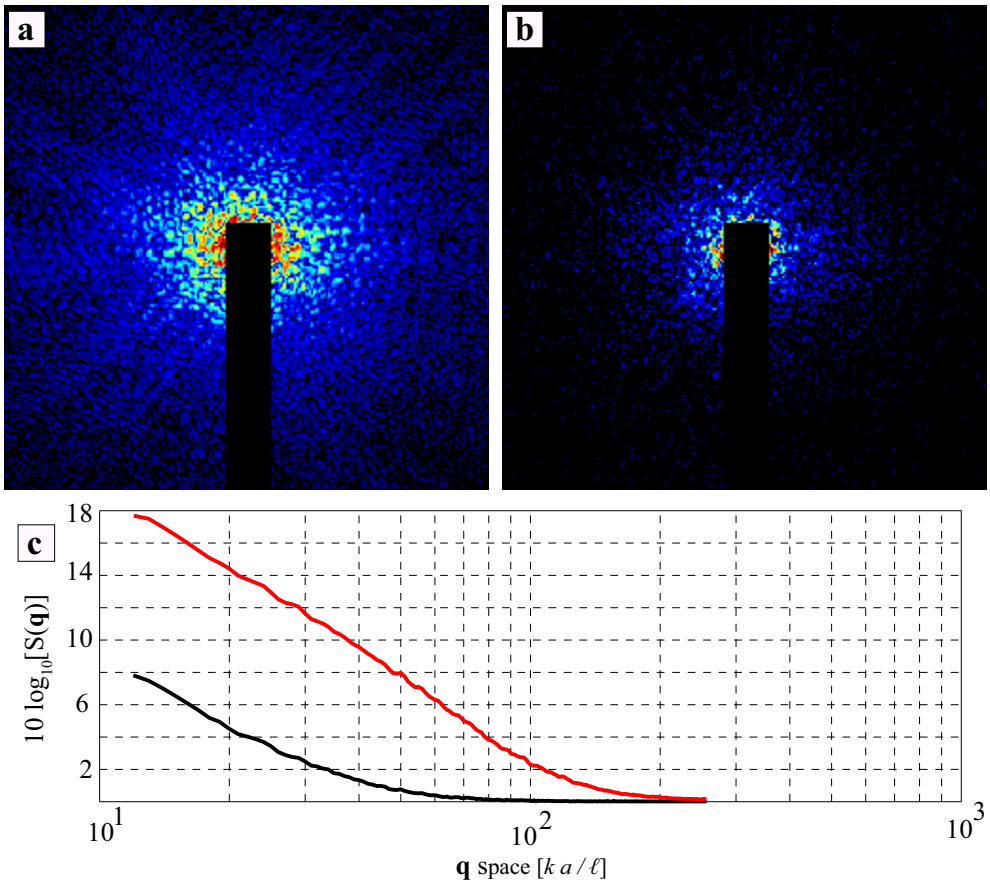

Figure 3.7: (a) A diffraction pattern with "good" signal to noise. The diffraction is dominated by Poisson "shot" noise, as is usual in x-ray synchrotron experiments, and the signal becomes lost in the noise back ground close to the edge of the measuring device. (b) A diffraction pattern with "poor" signal to noise. The signal becomes lost in the noise back ground relatively far away from the edge of the measuring device. In both diffraction patterns in (a-b), a rectangular region extending up from the bottom is removed from the diffraction to simulate the effect of a beamstop.

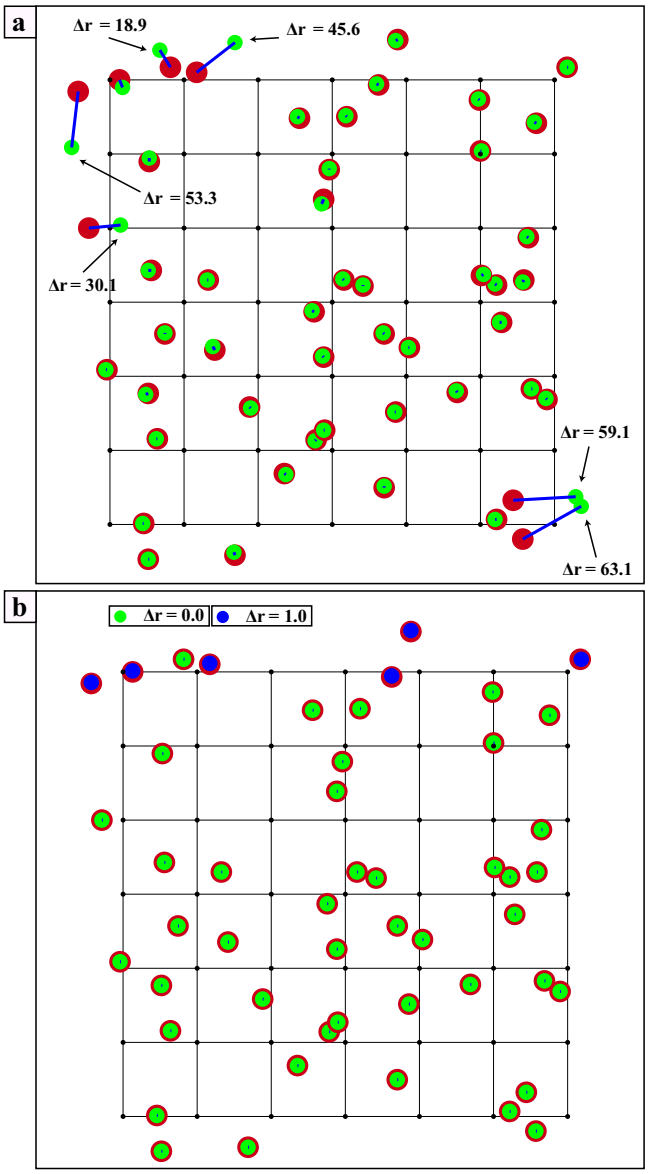

Figure 3.8: (a) The scan position configuration after application of the procedure given in Section 3.4.2 to a ptychographic diffraction data set with "good" signal to noise (as defined in Fig. 3.7), with 75% assumed adjacent scan position overlap, and random errors of up to $60\sqrt{2}$ pixels added to each scan location. The grid spacing between the small black dots is 70 pixels, the red dots are the true scan locations, the green dots are the solved for scan locations, and Δr refers the the distance between green and red dots. Most of the scan positions have been correctly solved for except a few at the periphery of the scan grid. (b) An average of 20 independent runs of the procedure given in Section 3.4.2. As a result of the averaging process, we have nearly perfectly recovered all the scan location positions.

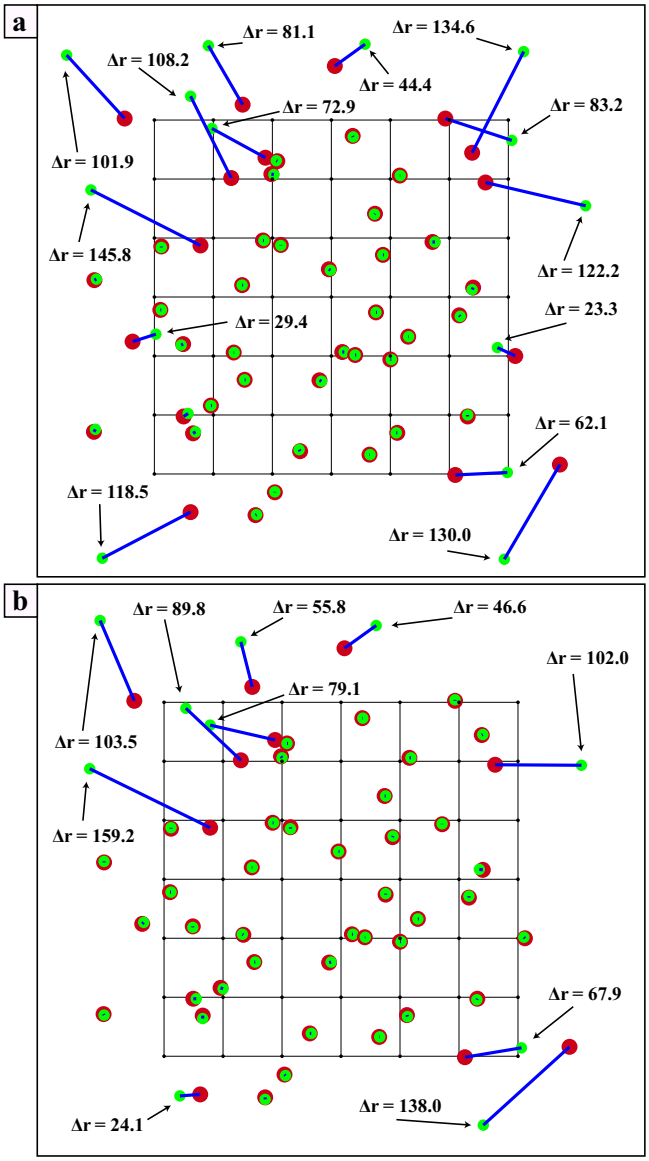

Figure 3.9: (a-b) The scan position configuration after application of the procedure given in Section 3.4.2 to a ptychographic diffraction data set with "good" signal to noise (as defined in Fig. 3.7), with 75% assumed adjacent scan position overlap, and random errors of up to $80\sqrt{2}$ pixels added to each scan location. The grid spacing between the small black dots is 70 pixels, the red dots are the true scan locations, the green dots are the solved for scan locations, and Δr refers the the distance between green and red dots. The errors in scan positions at the periphery of the scan grid in this case appear to be too severe for the conjugate gradient method to iteratively solve for. This is primarily a result of these locations having fewer diffraction patterns constraining them.

to each scan location (± 60 pixel error in x and the same for y), with an average error of $\Delta r_{avg} = 50.3$ pixels. This is 21% of the illumination function diameter and 143% of the scan position step size. The final scan position after running 1152 total iterations of the ePIE and DM, and 96 iterations of conjugate gradient scan position recovery, are shown in Fig. 3.8a. We see that all assumed scan positions (the green dots) have been corrected so that they are close to the true positions (the red dots), except on the top left and bottom right; this is characteristic of the method when the scan position errors are at the limit of the conjugate gradient algorithm performance. However, repeating the above process again will yield another configuration of scan positions, most of which have been correctly solved for and a few that have not been. If we repeat the above procedure say another 20 times, and average all of the obtained scan position configurations, we can finally reliably recover all scan positions as shown in Fig. 3.8b. Only a few scan positions at the top of Fig. 3.8b are slightly off from the true positions, by one pixel. In this example, it is *possible* to solve for errors larger than $\approx 60\sqrt{2}$ pixels, but usually cannot be done reliably; the scan positions at the periphery of the ptychographic data set are many times intractable for the conjugate gradient algorithm. This can be seen in Fig. 3.9a-b for maximum errors of $\approx 80\sqrt{2}$ pixels; here the centrally located scan positions have been solved for, but at the periphery the scan positions appear to never converge to the true locations, even after many averages as was done for the $\approx 60\sqrt{2}$ pixel case.

The maximum scan position error we are able to solve for is highly dependent of how tightly "clustered" the ptychographic scanning scheme is. The previous case discussed looked at an assumed scan position overlap of 75% and was able to recover errors of up to $\approx 60\sqrt{2}$ pixels reliably. If we instead look at an assumed scan position overlap of 85%, we find that we can reliably solve for up to $\approx 90\sqrt{2}$ pixels (which is 45% of the illumination function diameter and 303 % of the scan position step size of 42 pixels), as seen in Fig. 3.10a-b, while if the assumed scan position overlap is 65%, we can only reliably solve for up to $\approx 40\sqrt{2}$ pixels (which is 20% of the illumination function diameter and 58 % of the scan position step size of 98 pixels), as seen in Fig. 3.11a-b. In both cases, 20 indepen-

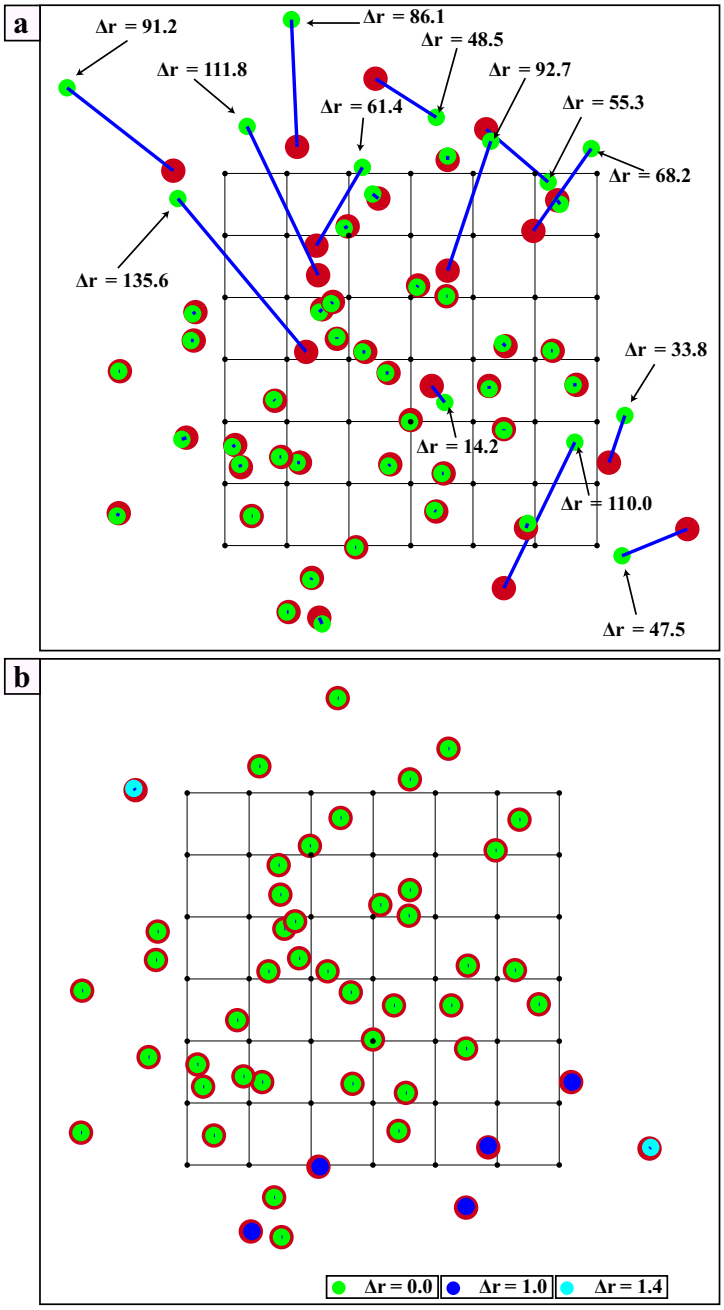

Figure 3.10: (a) The scan position configuration after application of the procedure given in Section 3.4.2 to a ptychographic diffraction data set with "good" signal to noise (as defined in Fig. 3.7), with 85% assumed adjacent scan position overlap, and random errors of up to $90\sqrt{2}$ pixels added to each scan location. The grid spacing between the small black dots is 42 pixels, the red dots are the true scan locations, the green dots are the solved for scan locations, and Δr refers the the distance between green and red dots. (b) An average of 20 independent runs of the scan position refinement procedure.

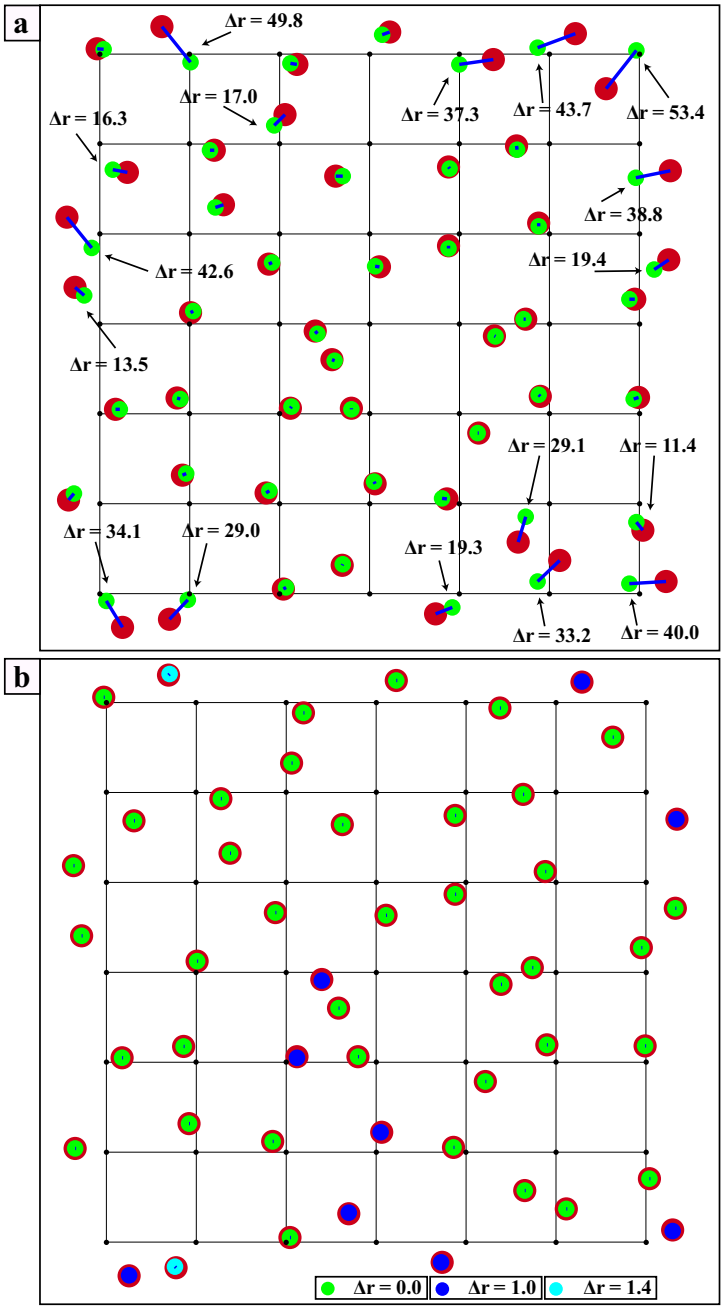

Figure 3.11: (a) The scan position configuration after application of the procedure given in Section 3.4.2 to a ptychographic diffraction data set with "good" signal to noise (as defined in Fig. 3.7), with 65% assumed adjacent scan position overlap, and random errors of up to $40\sqrt{2}$ pixels added to each scan location. The grid spacing between the small black dots is 98 pixels, the red dots are the true scan locations, the green dots are the solved for scan locations, and Δr refers the the distance between green and red dots. (b) An average of 20 independent runs of the scan position refinement procedure.

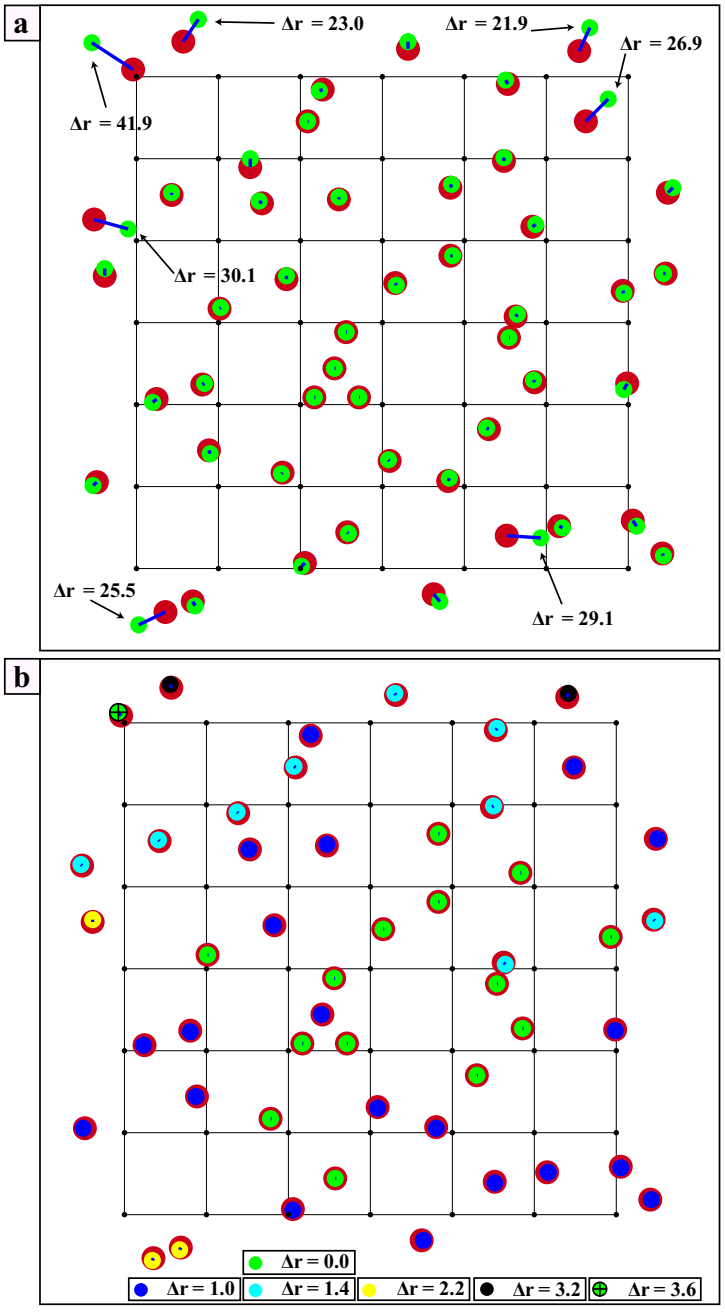

Figure 3.12: (a) The scan position configuration after application of the procedure given in Section 3.4.2 to a ptychographic diffraction data set with "poor" signal to noise (as defined in Fig. 3.7), with 75% assumed adjacent scan position overlap, and random errors of up to $40\sqrt{2}$ pixels added to each scan location. The grid spacing between the small black dots is 70 pixels, the red dots are the true scan locations, the green dots are the solved for scan locations, and Δr refers the the distance between green and red dots. (b) An average of 20 independent runs of the scan position refinement procedure.

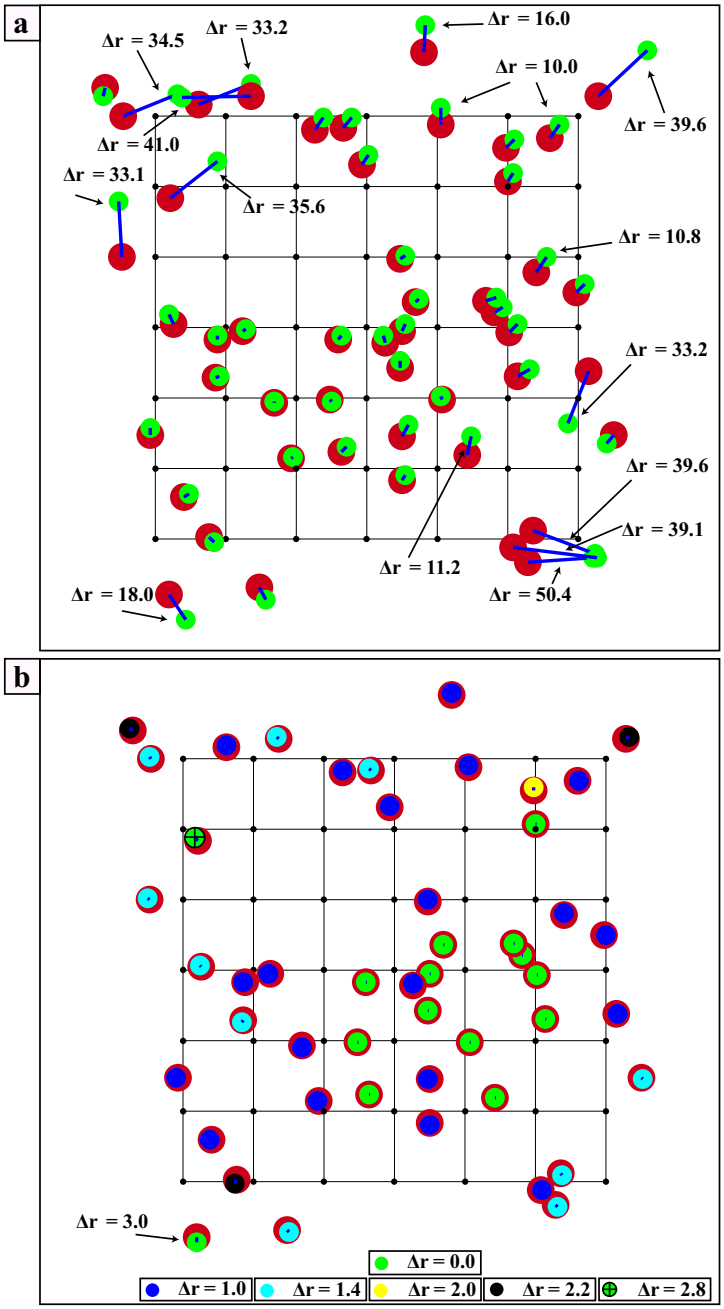

Figure 3.13: (a) The scan position configuration after application of the procedure given in Section 3.4.2 to a ptychographic diffraction data set with "poor" signal to noise (as defined in Fig. 3.7), with 85% assumed adjacent scan position overlap, and random errors of up to $60\sqrt{2}$ pixels added to each scan location. The grid spacing between the small black dots is 42 pixels, the red dots are the true scan locations, the green dots are the solved for scan locations, and Δr refers the the distance between green and red dots. (b) An average of 20 independent runs of the scan position refinement procedure.

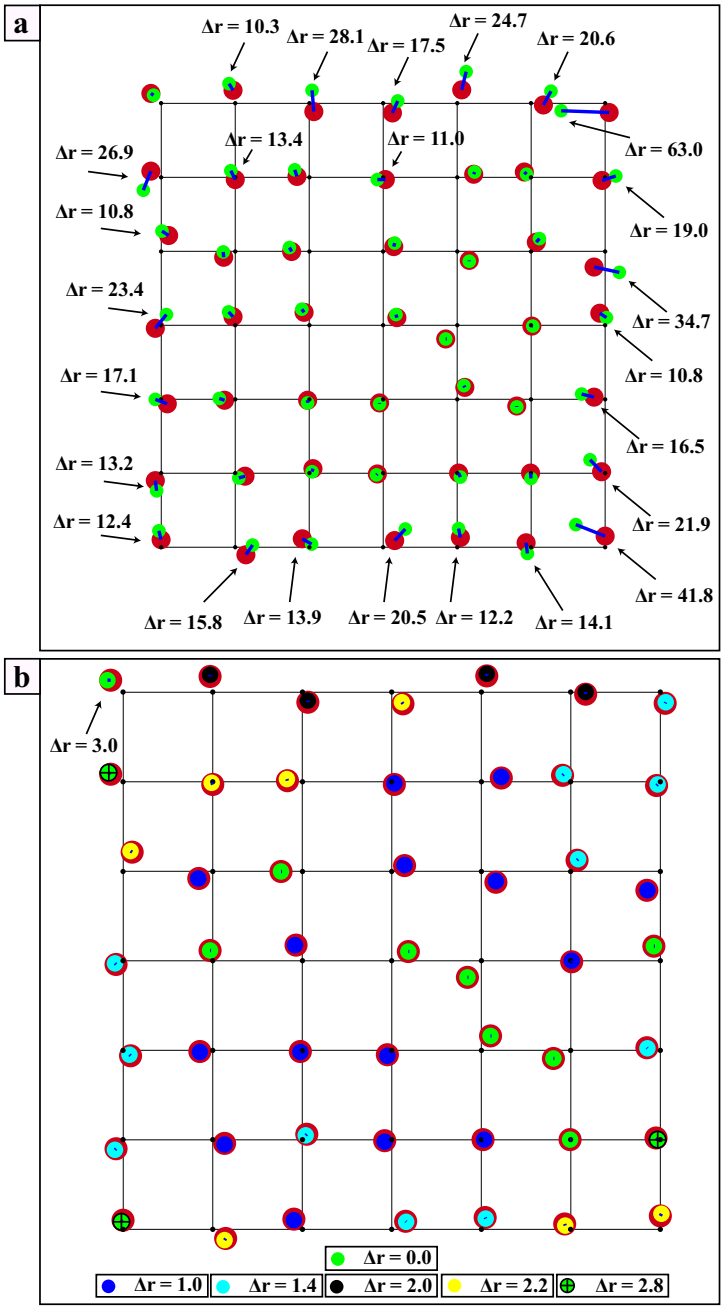

Figure 3.14: (a) The scan position configuration after application of the procedure given in Section 3.4.2 to a ptychographic diffraction data set with "poor" signal to noise (as defined in Fig. 3.7), with 65% assumed adjacent scan position overlap, and random errors of up to $20\sqrt{2}$ pixels added to each scan location. The grid spacing between the small black dots is 98 pixels, the red dots are the true scan locations, the green dots are the solved for scan locations, and Δr refers the the distance between green and red dots. (b) An average of 20 independent runs of the scan position refinement procedure.

dent runs using 144 iterations of ePIE and 144 iterations of DM repeated 4 times each, and scan positions updates every 12 iterations (for total of 96 iterations) are performed, and the scan positions are averaged at the end.

We next look at the "poor" signal to noise diffraction data as seen in Fig. 3.7b. As we would expect, since the diffraction is far less information rich when compared to the "good" signal to noise case, the maximum scan position errors we can solve for are significantly smaller. For 75% assumed overlap, the best we can do, using the same reconstruction recipe as for the "good" signal to noise diffraction above, is up to $\approx 40\sqrt{2}$ pixels (which is 20% of the illumination function diameter and 81% of the scan position step size of 70 pixels), as seen in Fig. 3.12a-b. For 85% assumed overlap, the best we can do is up to $\approx 60\sqrt{2}$ pixels (which is 30% of the illumination function diameter and 202% of the scan position step size of 42 pixels), as seen in Fig. 3.13a-b, and for 65% assumed overlap, the best we can do is up to $\approx 20\sqrt{2}$ pixels (which is 10% of the illumination function diameter and 29% of the scan position step size of 98 pixels), as seen in Fig. 3.14a-b. The number of scan positions for these cases which are not found perfectly is greatly increased for the "poor" signal to noise diffraction data. This however might be expected, since in the ptychographic diffraction data set we only have spatial frequency information up to about $q_{max} = 30\Delta q = 30ka/\ell$, which means that in real space we only have information up to about $\pi/q_{max} = N\Delta x_S/60 \approx 8.5$ pixels. For the 75%, 85%, and 65% assumed overlap cases presented in Fig. 3.12, Fig. 3.13, and Fig. 3.14, the typical error of a reconstructed scan position with respect to the true scan position ranges between $\Delta r_{typical} \approx 5$ to 15. That we are able to recover scan positions at a resolution *better* than 8.5 pixels is due to the averaging of many independent reconstructions, and is likely serendipitous.

3.4.4 Scan position errors and diffraction limited imaging

The next question we address is whether or not the ability to achieve diffraction limited spatial resolution and a quantitative index of refraction has been restored by iteratively solving for unknown scan positions as in the previous sections. Figure 3.15 shows reconstructions for 65% assumed overlap and using the "good"

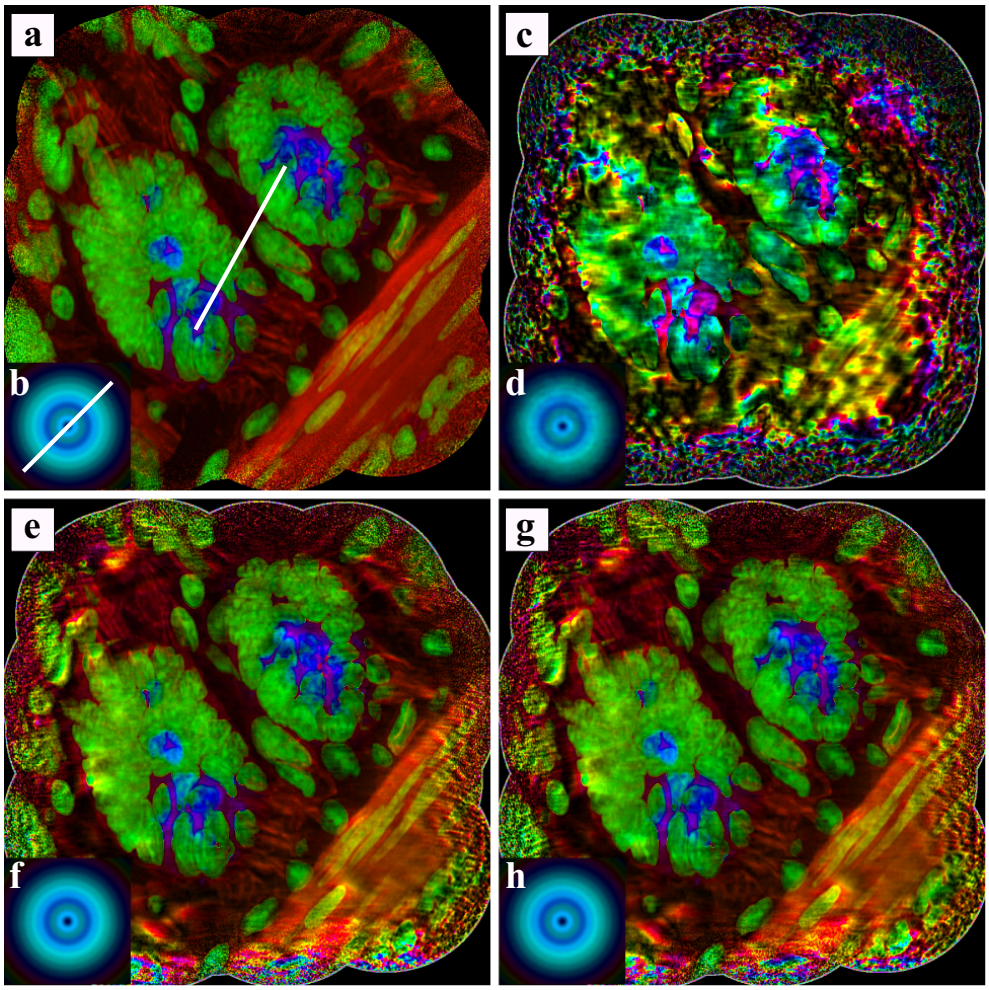

Figure 3.15: Reconstructions for $T_j(\mathbf{r})$ and $p(\mathbf{r})$, 65% assumed overlap, "good" signal to noise, $\pm 40\sqrt{2}$ pixel maximum scan position errors. The white line though the reconstructions refers to the lineouts shown in Fig. 3.16. (a-b) The true $T_j(\mathbf{r})$ and $p(\mathbf{r})$, assuming the phases of the noise degraded diffraction intensities are perfectly known. (c-d) Reconstructions for $T_j(\mathbf{r})$ and $p(\mathbf{r})$ using the scan positions from Fig. 3.11a. (e-f) Reconstructions for $T_j(\mathbf{r})$ and $p(\mathbf{r})$ using the *true* scan locations. (g-h) Reconstructions for $T_j(\mathbf{r})$ and $p(\mathbf{r})$ using the scan positions from Fig. 3.11b.

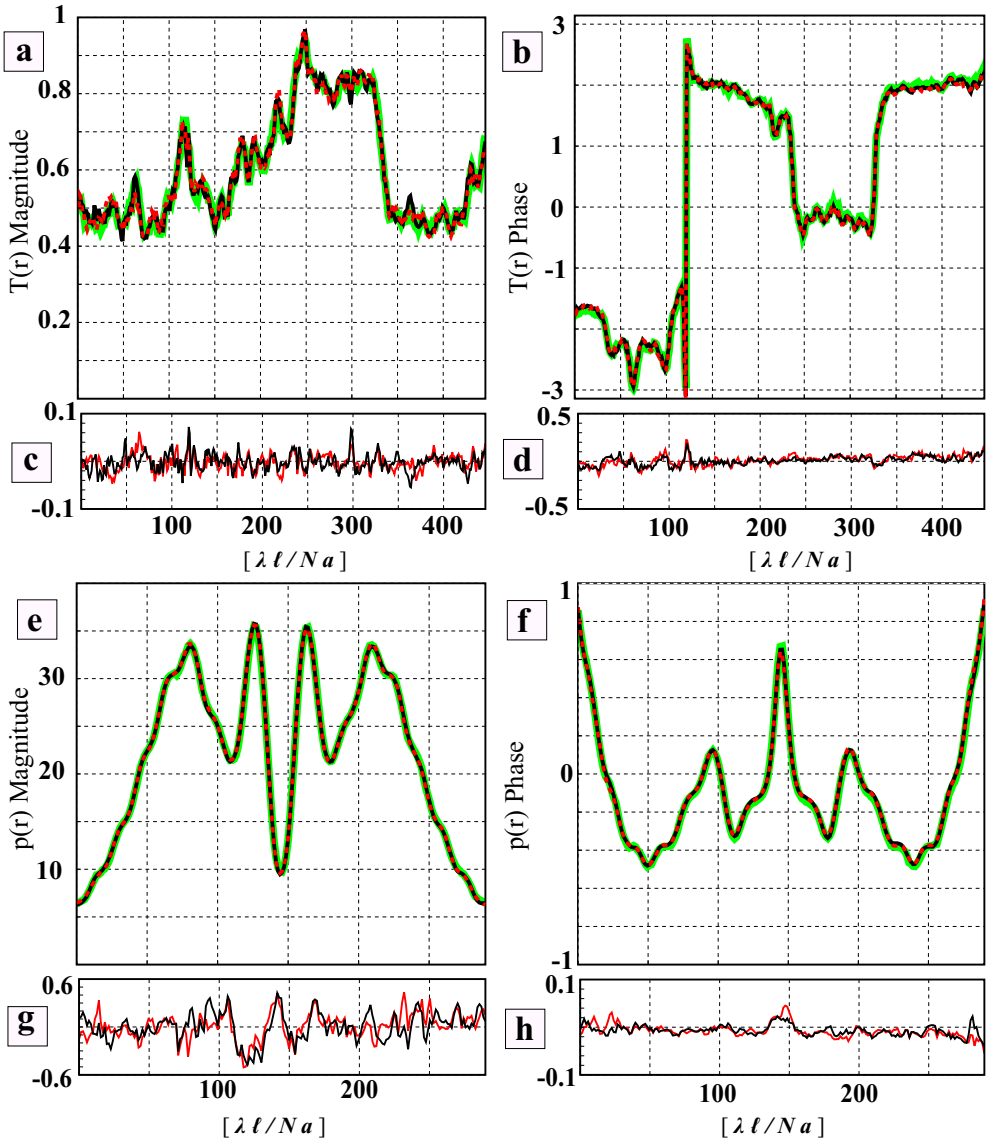

Figure 3.16: (a,b,e,f) Lineouts of the reconstructions along the white lines in Fig 3.17b. (a,b) are line outs for $T_j(\mathbf{r})$ while (e,f) are line outs for $p(\mathbf{r})$. In all frames, the red dashed curve is the lineout for the reconstruction using the recovered scan locations Fig. 3.8b, the black curve is for the reconstruction assuming the scan locations were perfectly known, and the green curve is for the true $T_j(\mathbf{r})$. (c,d,g,h) The red curve is the difference between the red dashed curve and the green curve in frames (a,b,e,f), while the black curve is the difference between the black curve and the green curve in frame (a,b,e,f). (a) Lineout of $|T_j(\mathbf{r})|$. (b) Lineout of the phase of $T_j(\mathbf{r})$. (c) Difference between line outs in frame (a). In both cases, the error between the reconstructions and the true $|T_j(\mathbf{r})|$ is less than ≈ 0.1. (d) Difference between line outs in frame (b). In both cases, the error between the reconstructions and the true phase of $T_j(\mathbf{r})$ is less than ≈ 0.5.

signal to noise diffraction. In Fig. 3.15a-b we have the the true $T_j(\mathbf{r})$ and $p(\mathbf{r})$, generated by using the true phases with the diffraction intensity and also without missing diffraction intensity due to the beam stop. In Fig. 3.15c-d a reconstruction is shown for the scan position configuration shown in Fig. 3.11a, in Fig. 3.15e-f a reconstruction is shown assuming that the scan positions are known without any error, and in Fig. 3.15g-h a reconstruction is shown using the iteratively solved for scan positions shown in Fig. 3.11b.

Since we know the true $T_j(\mathbf{r})$ and $p(\mathbf{r})$, we can compare the results shown in Fig. 3.15e-h to this, and this is done in Fig. 3.16. Here, a line out is taken along the white lines shown in Fig. 3.15a-b. In Fig. 3.16a and 3.16b, the magnitudes and phases of the reconstructed $T_j(\mathbf{r})$ are shown respectively; also in Fig. 3.16e and 3.16f, the magnitudes and phases of the reconstructed $p(\mathbf{r})$ are shown respectively. In all of these, reconstructions when the scan positions are assumed known are shown as the black curves, using the iteratively solved for scan positions are shown as the dashed red curves, and the true $T_j(\mathbf{r})$ magnitudes are shown as the green curves. In Fig. 3.16c, Fig. 3.16d, Fig. 3.16g, and Fig. 3.16h, the red curves are the difference between reconstructions using the iteratively solved for scan locations and the true $T_j(\mathbf{r})$ or $p(\mathbf{r})$ and the black curves are the difference between reconstructions when the scan locations are assumed known and the true $T_j(\mathbf{r})$ or $p(\mathbf{r})$.

All of this is also done for the 75% and 85% assumed overlaps in Fig. 3.17, Fig. 3.18, Fig. 3.19, and Fig. 3.20. We readily see that the reconstructions using the solved for scan locations and reconstructions using assumed known scan locations are essentially identical and that both have similar error bounds in all cases. The source of error here stems from the fact that the ptychographic diffraction data has Poisson noise associated with it, and that errors in Fourier space mean errors in real space. Also, we do not have complete diffraction measurements: the missing data can cause numerical artifacts due to unconstrained modes in the missing data region [41]. In order to get quantitative results for the reconstructions, we must effectively determine the proper scaling in this missing data region; this is done using a method given in [40]. The effects of Poisson noise and unconstrained

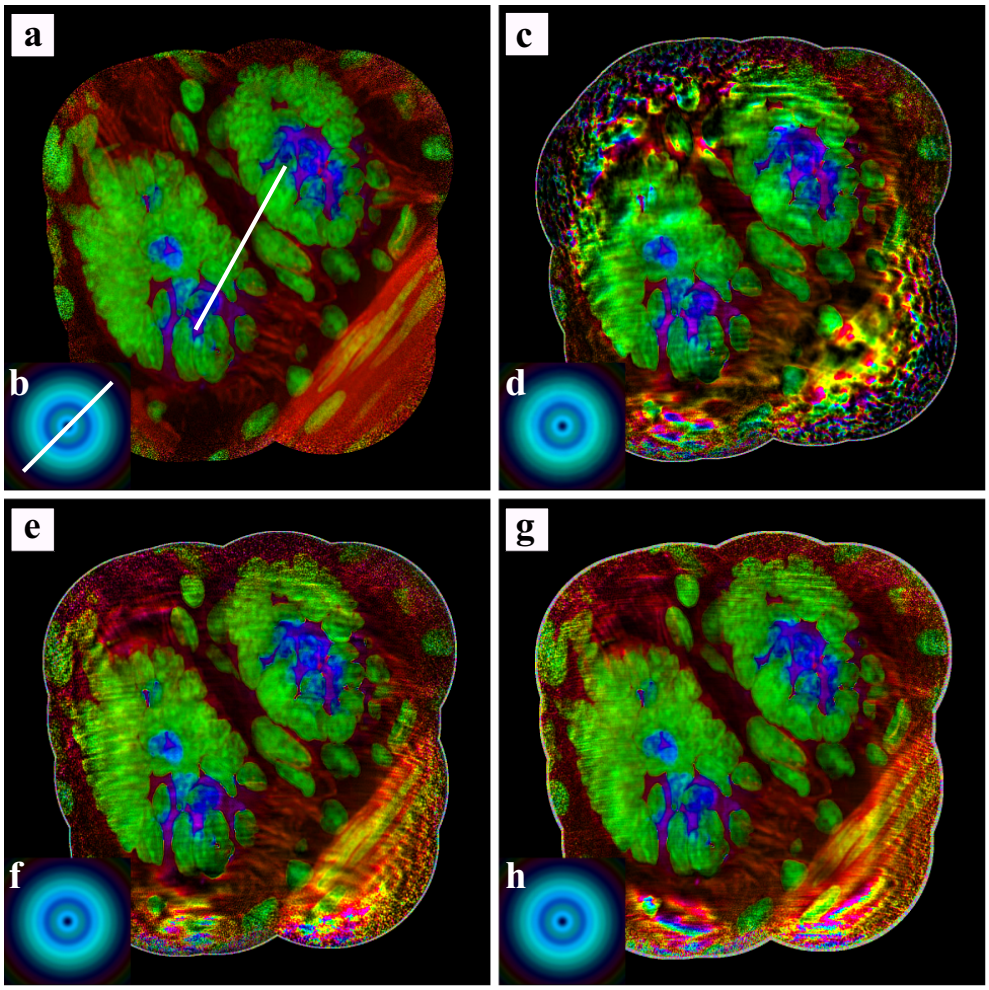

Figure 3.17: Reconstructions for $T_j(\mathbf{r})$ and $p(\mathbf{r})$, 75% assumed overlap, "good" signal to noise, $\pm 60\sqrt{2}$ pixel maximum scan position errors. The white line though the reconstructions refers to the lineouts shown in Fig. 3.18. (a-b) The true $T_j(\mathbf{r})$ and $p(\mathbf{r})$, assuming the phases of the noise degraded diffraction intensities are perfectly known. (c-d) Reconstructions for $T_j(\mathbf{r})$ and $p(\mathbf{r})$ using the scan positions from Fig. 3.8a. (e-f) Reconstructions for $T_j(\mathbf{r})$ and $p(\mathbf{r})$ using the *true* scan locations. (g-h) Reconstructions for $T_j(\mathbf{r})$ and $p(\mathbf{r})$ using the scan positions from Fig. 3.8b.

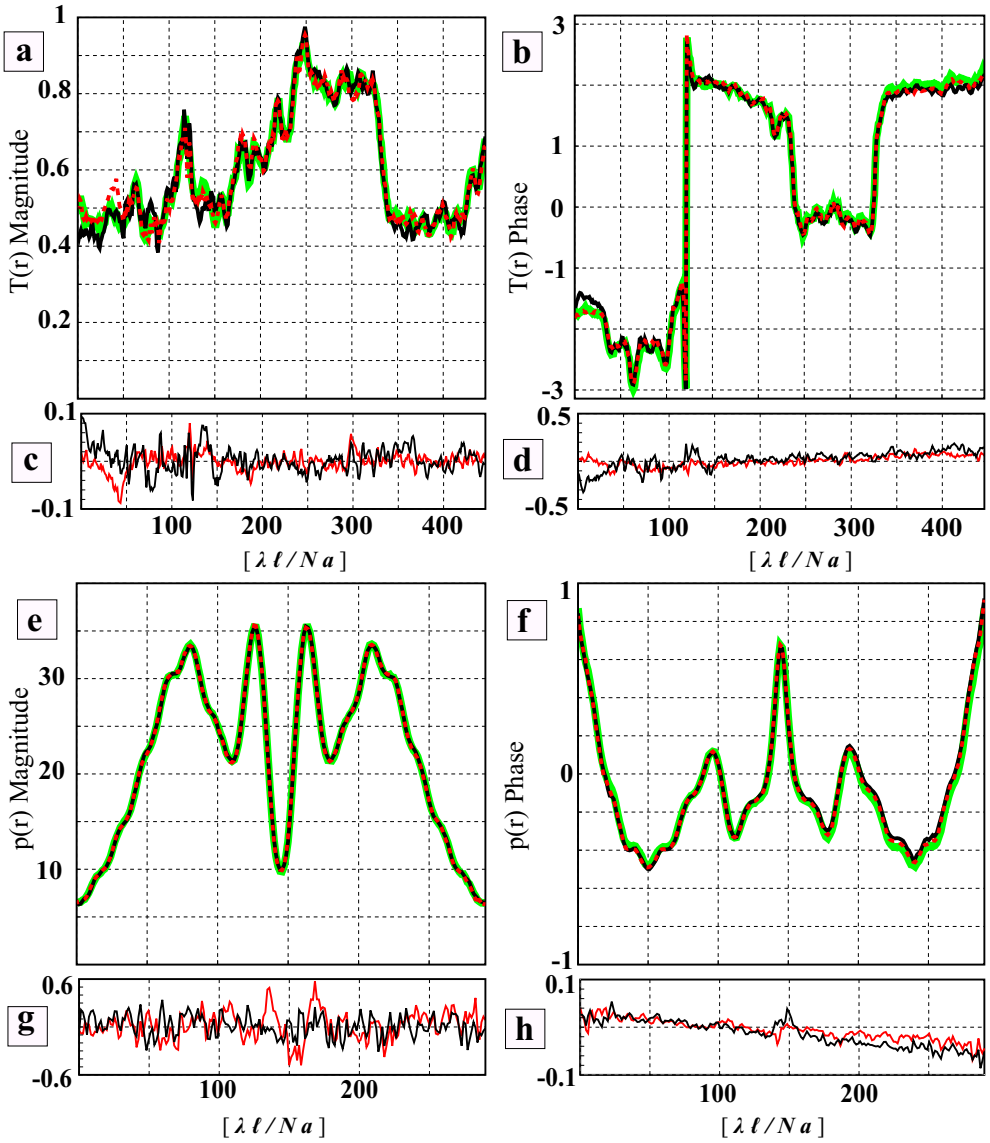

Figure 3.18: (a,b,e,f) Lineouts of the reconstructions along the white lines in Fig 3.17b. (a,b) are line outs for $T_j(\mathbf{r})$ while (e,f) are line outs for $p(\mathbf{r})$. In all frames, the red dashed curve is the lineout for the reconstruction using the recovered scan locations Fig. 3.8b, the black curve is for the reconstruction assuming the scan locations were perfectly known, and the green curve is for the true $T_j(\mathbf{r})$. (c,d,g,h) The red curve is the difference between the red dashed curve and the green curve in frames (a,b,e,f), while the black curve is the difference between the black curve and the green curve in frame (a,b,e,f). (a) Lineout of $|T_j(\mathbf{r})|$. (b) Lineout of the phase of $T_j(\mathbf{r})$. (c) Difference between line outs in frame (a). In both cases, the error between the reconstructions and the true $|T_j(\mathbf{r})|$ is less than ≈ 0.1. (d) Difference between line outs in frame (b). In both cases, the error between the reconstructions and the true phase of $T_j(\mathbf{r})$ is less than ≈ 0.5.

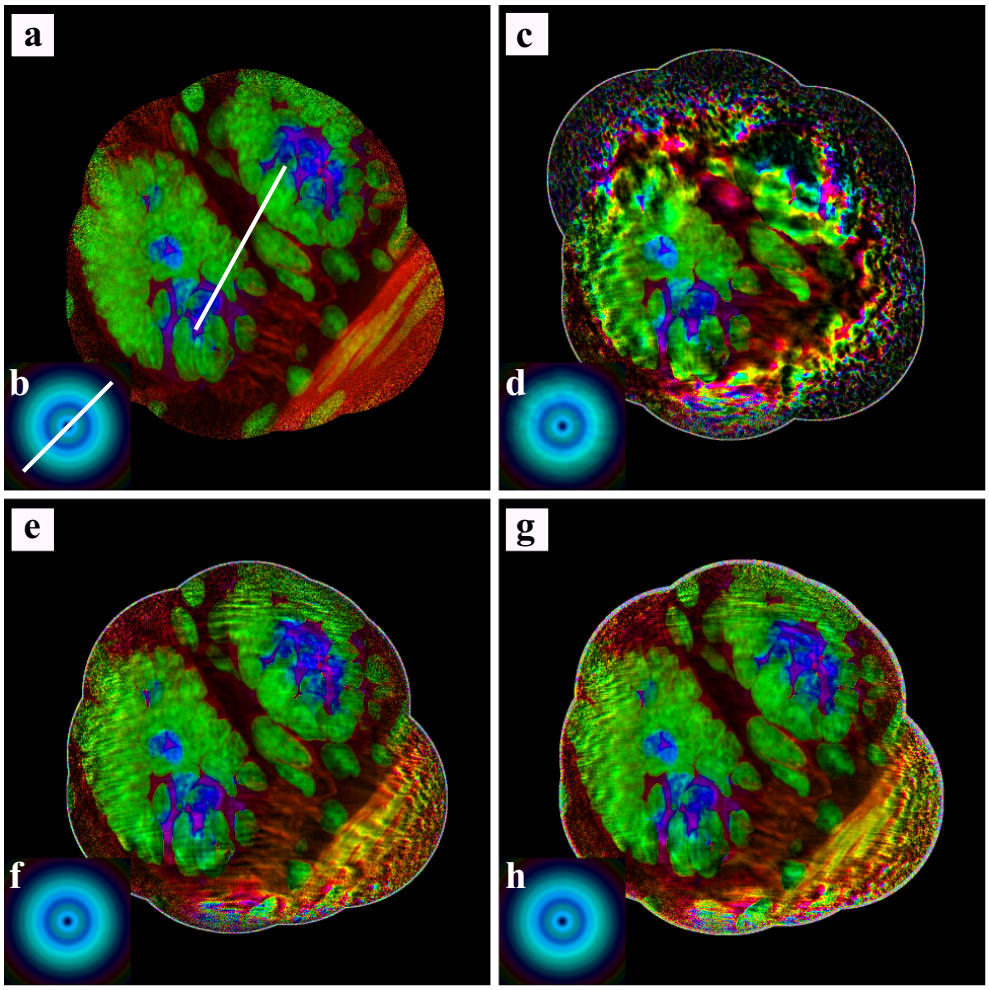

Figure 3.19: Reconstructions for $T_j(\mathbf{r})$ and $p(\mathbf{r})$, 85% assumed overlap, "good" signal to noise, $\pm 90\sqrt{2}$ pixel maximum scan position errors. The white line though the reconstructions refers to the lineouts shown in Fig. 3.20. (a-b) The true $T_j(\mathbf{r})$ and $p(\mathbf{r})$, assuming the phases of the noise degraded diffraction intensities are perfectly known. (c-d) Reconstructions for $T_j(\mathbf{r})$ and $p(\mathbf{r})$ using the scan positions from Fig. 3.10a. (e-f) Reconstructions for $T_j(\mathbf{r})$ and $p(\mathbf{r})$ using the *true* scan locations. (g-h) Reconstructions for $T_j(\mathbf{r})$ and $p(\mathbf{r})$ using the scan positions from Fig. 3.10b.

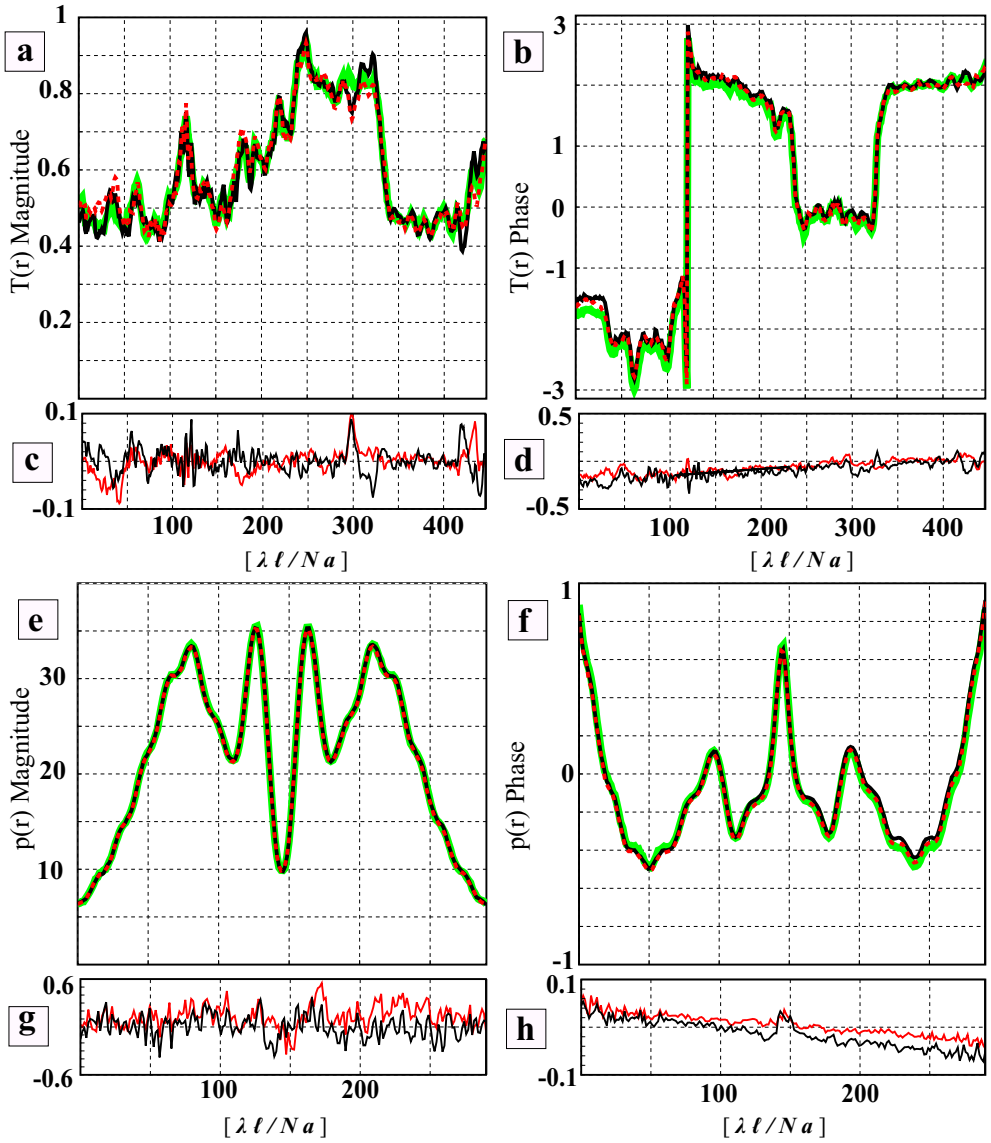

Figure 3.20: (a,b,e,f) Lineouts of the reconstructions along the white lines in Fig 3.17b. (a,b) are line outs for $T_j(\mathbf{r})$ while (e,f) are line outs for $p(\mathbf{r})$. In all frames, the red dashed curve is the lineout for the reconstruction using the recovered scan locations Fig. 3.8b, the black curve is for the reconstruction assuming the scan locations were perfectly known, and the green curve is for the true $T_j(\mathbf{r})$. (c,d,g,h) The red curve is the difference between the red dashed curve and the green curve in frames (a,b,e,f), while the black curve is the difference between the black curve and the green curve in frame (a,b,e,f). (a) Lineout of $|T_j(\mathbf{r})|$. (b) Lineout of the phase of $T_j(\mathbf{r})$. (c) Difference between line outs in frame (a). In both cases, the error between the reconstructions and the true $|T_j(\mathbf{r})|$ is less than ≈ 0.1. (d) Difference between line outs in frame (b). In both cases, the error between the reconstructions and the true phase of $T_j(\mathbf{r})$ is less than ≈ 0.5.

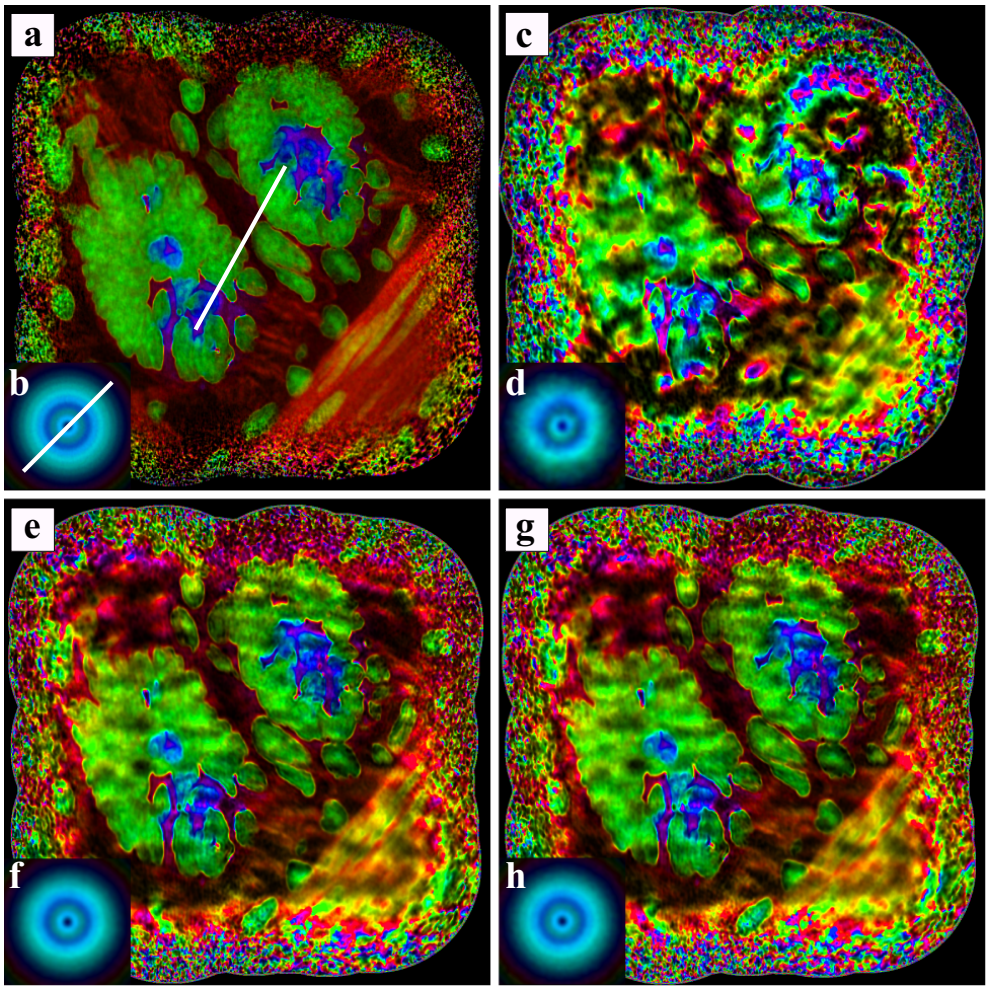

Figure 3.21: Reconstructions for $T_j(\mathbf{r})$ and $p(\mathbf{r})$, 65% assumed overlap, "poor" signal to noise, $\pm 20\sqrt{2}$ pixel maximum scan position errors. The white line though the reconstructions refers to the lineouts shown in Fig. 3.22. (a-b) The true $T_j(\mathbf{r})$ and $p(\mathbf{r})$, assuming the phases of the noise degraded diffraction intensities are perfectly known. (c-d) Reconstructions for $T_j(\mathbf{r})$ and $p(\mathbf{r})$ using the scan positions from Fig. 3.14a. (e-f) Reconstructions for $T_j(\mathbf{r})$ and $p(\mathbf{r})$ using the *true* scan locations. (g-h) Reconstructions for $T_j(\mathbf{r})$ and $p(\mathbf{r})$ using the scan positions from Fig. 3.14b.

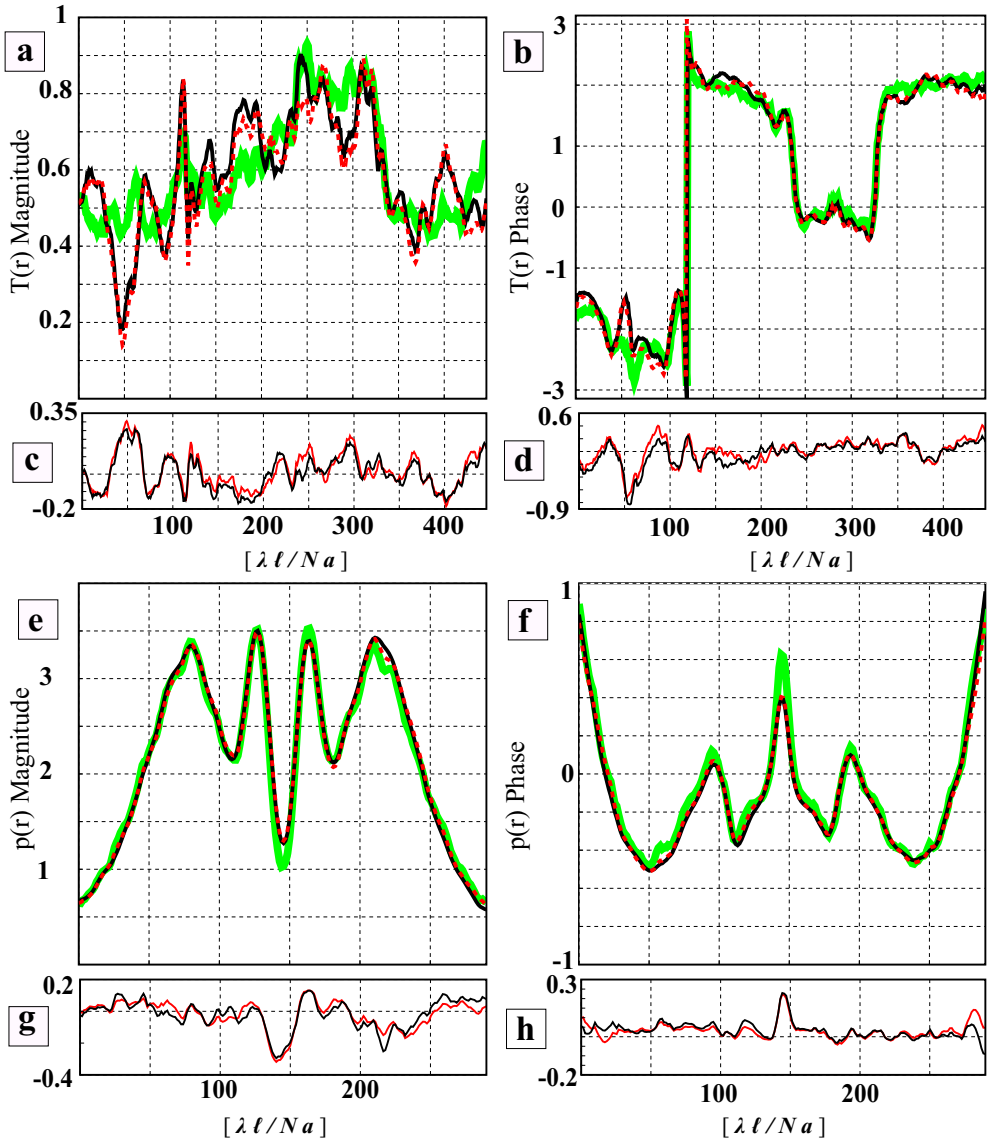

Figure 3.22: (a,b,e,f) Lineouts of the reconstructions along the white lines in Fig 3.17b. (a,b) are line outs for $T_j(\mathbf{r})$ while (e,f) are line outs for $p(\mathbf{r})$. In all frames, the red dashed curve is the lineout for the reconstruction using the recovered scan locations Fig. 3.8b, the black curve is for the reconstruction assuming the scan locations were perfectly known, and the green curve is for the true $T_j(\mathbf{r})$. (c,d,g,h) The red curve is the difference between the red dashed curve and the green curve in frames (a,b,e,f), while the black curve is the difference between the black curve and the green curve in frame (a,b,e,f). (a) Lineout of $|T_j(\mathbf{r})|$. (b) Lineout of the phase of $T_j(\mathbf{r})$. (c) Difference between line outs in frame (a). In both cases, the error between the reconstructions and the true $|T_j(\mathbf{r})|$ is less than ≈ 0.1. (d) Difference between line outs in frame (b). In both cases, the error between the reconstructions and the true phase of $T_j(\mathbf{r})$ is less than ≈ 0.5.

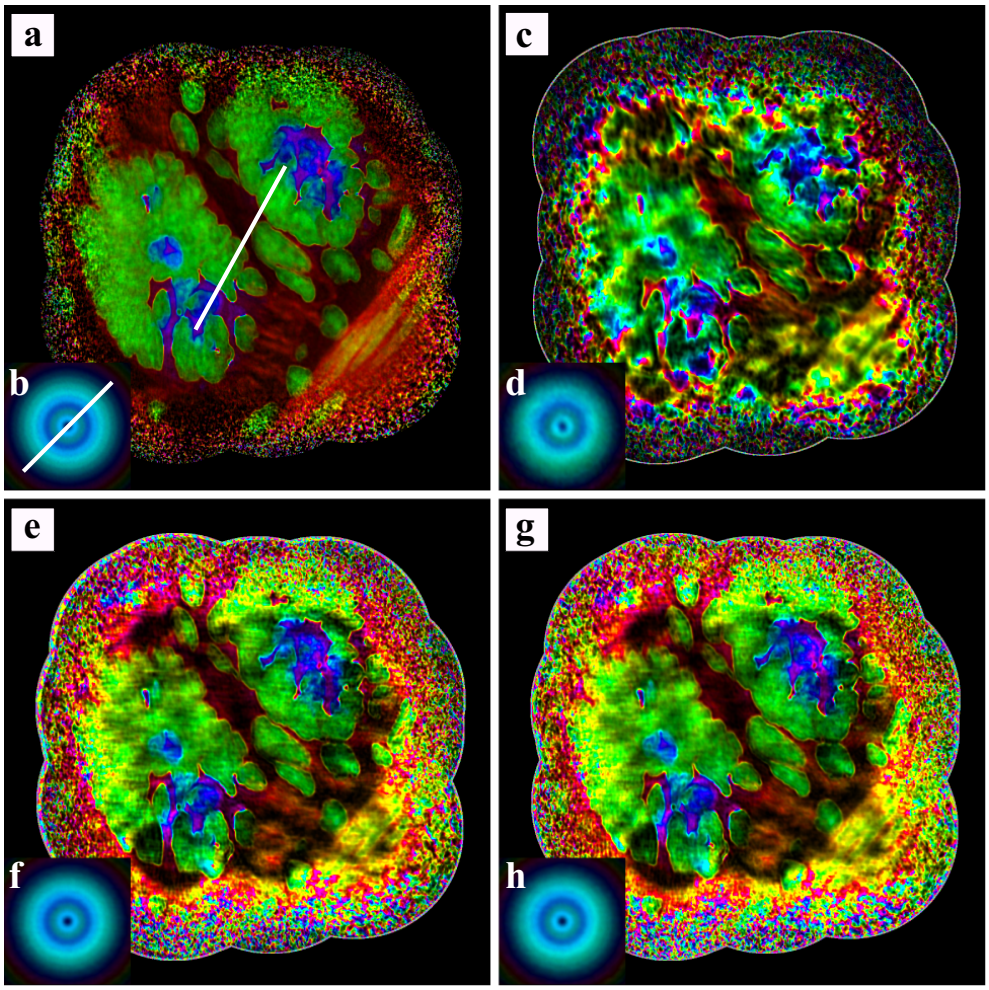

Figure 3.23: Reconstructions for $T_j(\mathbf{r})$ and $p(\mathbf{r})$, 75% assumed overlap, "poor" signal to noise, $\pm 40\sqrt{2}$ pixel maximum scan position errors. The white line though the reconstructions refers to the lineouts shown in Fig. 3.24. (a-b) The true $T_j(\mathbf{r})$ and $p(\mathbf{r})$, assuming the phases of the noise degraded diffraction intensities are perfectly known. (c-d) Reconstructions for $T_j(\mathbf{r})$ and $p(\mathbf{r})$ using the scan positions from Fig. 3.12a. (e-f) Reconstructions for $T_j(\mathbf{r})$ and $p(\mathbf{r})$ using the *true* scan locations. (g-h) Reconstructions for $T_j(\mathbf{r})$ and $p(\mathbf{r})$ using the scan positions from Fig. 3.12b.

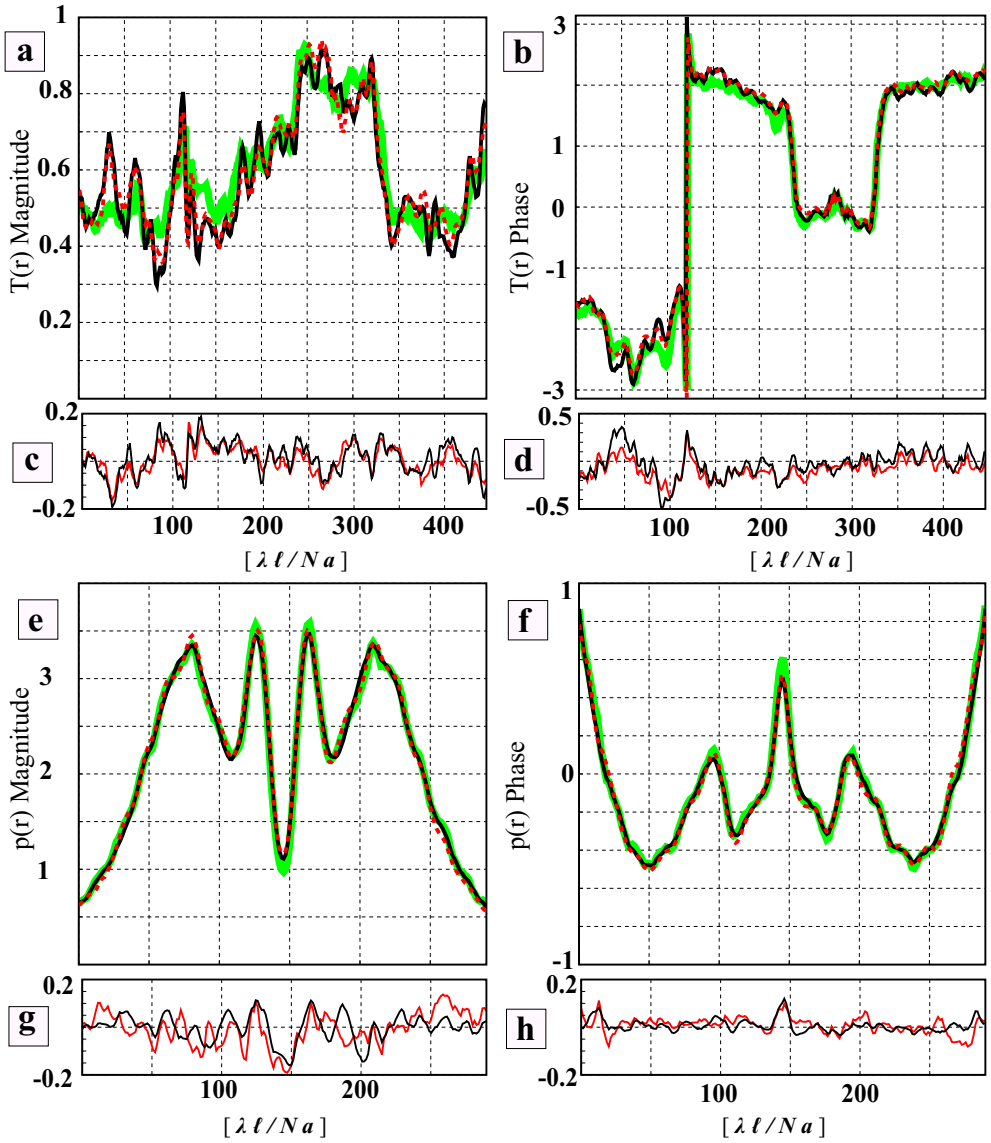

Figure 3.24: (a,b,e,f) Lineouts of the reconstructions along the white lines in Fig 3.17b. (a,b) are line outs for $T_j(\mathbf{r})$ while (e,f) are line outs for $p(\mathbf{r})$. In all frames, the red dashed curve is the lineout for the reconstruction using the recovered scan locations Fig. 3.8b, the black curve is for the reconstruction assuming the scan locations were perfectly known, and the green curve is for the true $T_j(\mathbf{r})$. (c,d,g,h) The red curve is the difference between the red dashed curve and the green curve in frames (a,b,e,f), while the black curve is the difference between the black curve and the green curve in frame (a,b,e,f). (a) Lineout of $|T_j(\mathbf{r})|$. (b) Lineout of the phase of $T_j(\mathbf{r})$. (c) Difference between line outs in frame (a). In both cases, the error between the reconstructions and the true $|T_j(\mathbf{r})|$ is less than ≈ 0.1. (d) Difference between line outs in frame (b). In both cases, the error between the reconstructions and the true phase of $T_j(\mathbf{r})$ is less than ≈ 0.5.

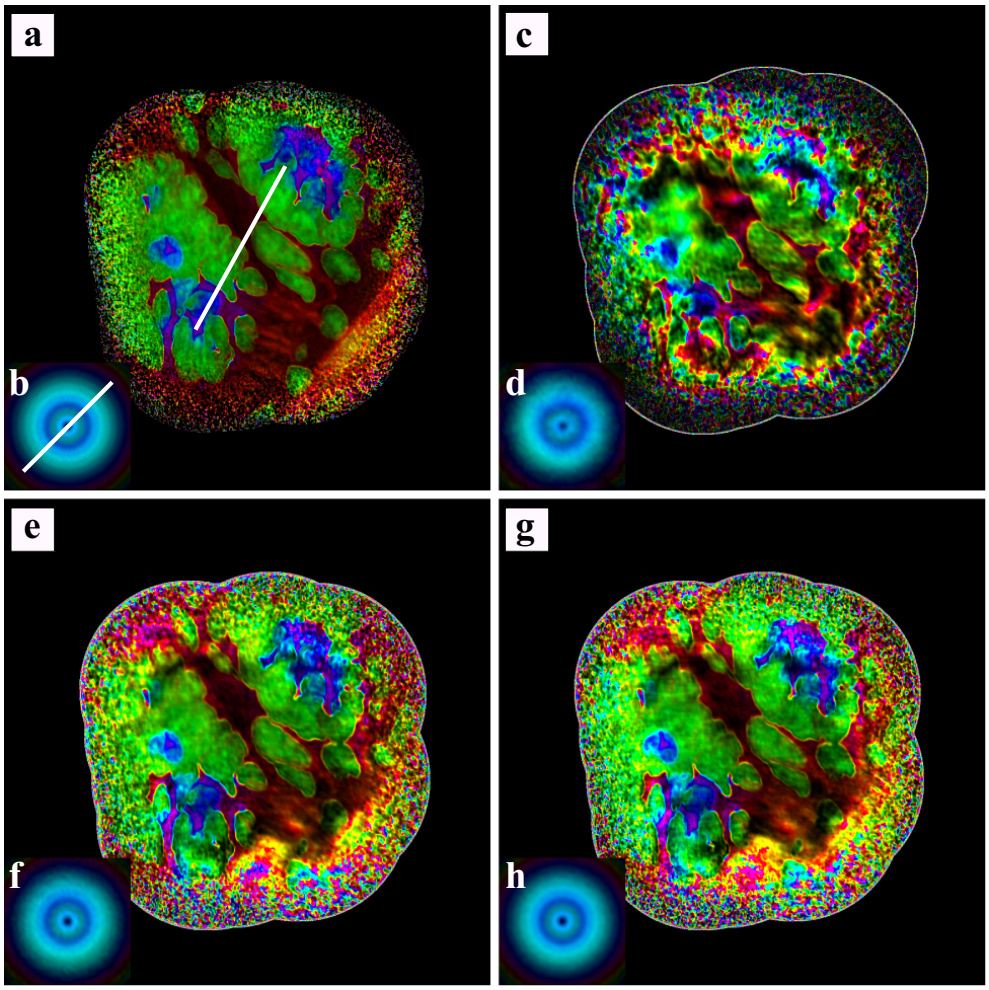

Figure 3.25: Reconstructions for $T_j(\mathbf{r})$ and $p(\mathbf{r})$, 85% assumed overlap, "poor" signal to noise, $\pm 60\sqrt{2}$ pixel maximum scan position errors. The white line though the reconstructions refers to the lineouts shown in Fig. 3.26. (a-b) The true $T_j(\mathbf{r})$ and $p(\mathbf{r})$, assuming the phases of the noise degraded diffraction intensities are perfectly known. (c-d) Reconstructions for $T_j(\mathbf{r})$ and $p(\mathbf{r})$ using the scan positions from Fig. 3.13a. (e-f) Reconstructions for $T_j(\mathbf{r})$ and $p(\mathbf{r})$ using the *true* scan locations. (g-h) Reconstructions for $T_j(\mathbf{r})$ and $p(\mathbf{r})$ using the scan positions from Fig. 3.13b.

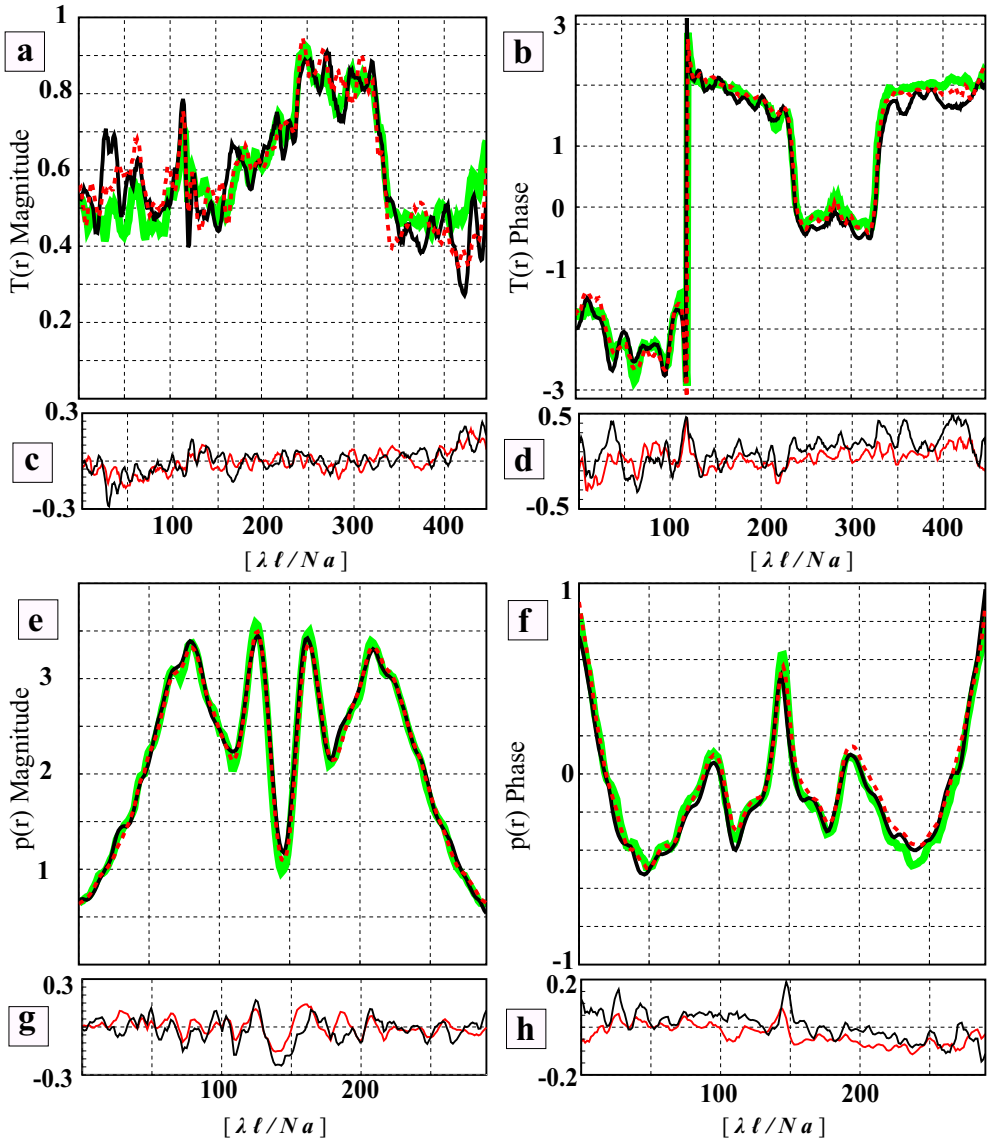

Figure 3.26: (a, b, e, f) Lineouts of the reconstructions along the white lines in Fig 3.17b. (a, b) are line outs for $T_j(\mathbf{r})$ while (e, f) are line outs for $p(\mathbf{r})$. In all frames, the red dashed curve is the lineout for the reconstruction using the recovered scan locations Fig. 3.8b, the black curve is for the reconstruction assuming the scan locations were perfectly known, and the green curve is for the true $T_j(\mathbf{r})$. (c, d, g, h) The red curve is the difference between the red dashed curve and the green curve in frames (a, b, e, f), while the black curve is the difference between the black curve and the green curve in frame (a, b, e, f). (a) Lineout of $|T_j(\mathbf{r})|$. (b) Lineout of the phase of $T_j(\mathbf{r})$. (c) Difference between line outs in frame (a). In both cases, the error between the reconstructions and the true $|T_j(\mathbf{r})|$ is less than ≈ 0.1. (d) Difference between line outs in frame (b). In both cases, the error between the reconstructions and the true phase of $T_j(\mathbf{r})$ is less than ≈ 0.5.

modes in the missing data region are even more pronounced when looking at the reconstructions using the "poor" signal to noise diffraction. Figure 3.21 shows reconstructions for 65% assumed overlap and using this "poor" signal to noise diffraction. Here, the reconstructions using the solved for and the assumed known scan locations are still very close to each other, but when compared to the true $T_j(\mathbf{r})$ or $p(\mathbf{r})$ as in Fig. 3.22, the relative error is greatly increased. The reason for this is clear: more information is lost behind the beam stop and the effect of Poisson noise is much greater, and as a result the algorithms used to numerically invert the diffraction are less effective; the Fourier space constraints contains much less information and the proper scaling of the missing data region is that much less effective. We see similar goings on with the 75% and 85% assumed overlap in Fig. 3.23, Fig. 3.24, Fig. 3.25 and Fig. 3.26.

We next examine the PRTF's of the reconstructions for both signal to noise cases; these are shown in Fig. 3.27-3.32. In these, the blue curves are the PRTF's when using the scan positions after only one application of the iterative scan position correction procedure in section 3.4.2 (corresponding to the top frames in Fig. 3.8 and Fig. 3.10-3.14), the dotted red curves are the PRTF's when using an average of 20 application of the iterative scan position correction procedure (corresponding to the bottom frames in Fig. 3.8 and Fig. 3.10-3.14), and the green curves are the PRTF's when assuming the scan positions are known. In these, we see that the dashed red PRTF curves are very close to the green PRTF curves, the difference between the two is shown in the bottom frame in Fig. 3.27-3.32. For the "good" signal to noise diffraction, the PRTF spatial frequency cutoff of 0.5 occurs at the same spatial frequency for all reconstructions, at about 125 ka/λ, and for the "poor" signal to noise diffraction, at about 40 ka/λ for all reconstructions. For all of these, the difference between the dashed red and the green PRTF curves is less than ~ 5 %, indicating that we have restored our ability to get spatial resolution which is diffraction limited.

In summary, we have explored how reconstruction quality degrades with uncertainties in the ptychographic constraint of known scan positions, positions of which cause the illumination function $p(\mathbf{r})$ to overlap with neighboring regions

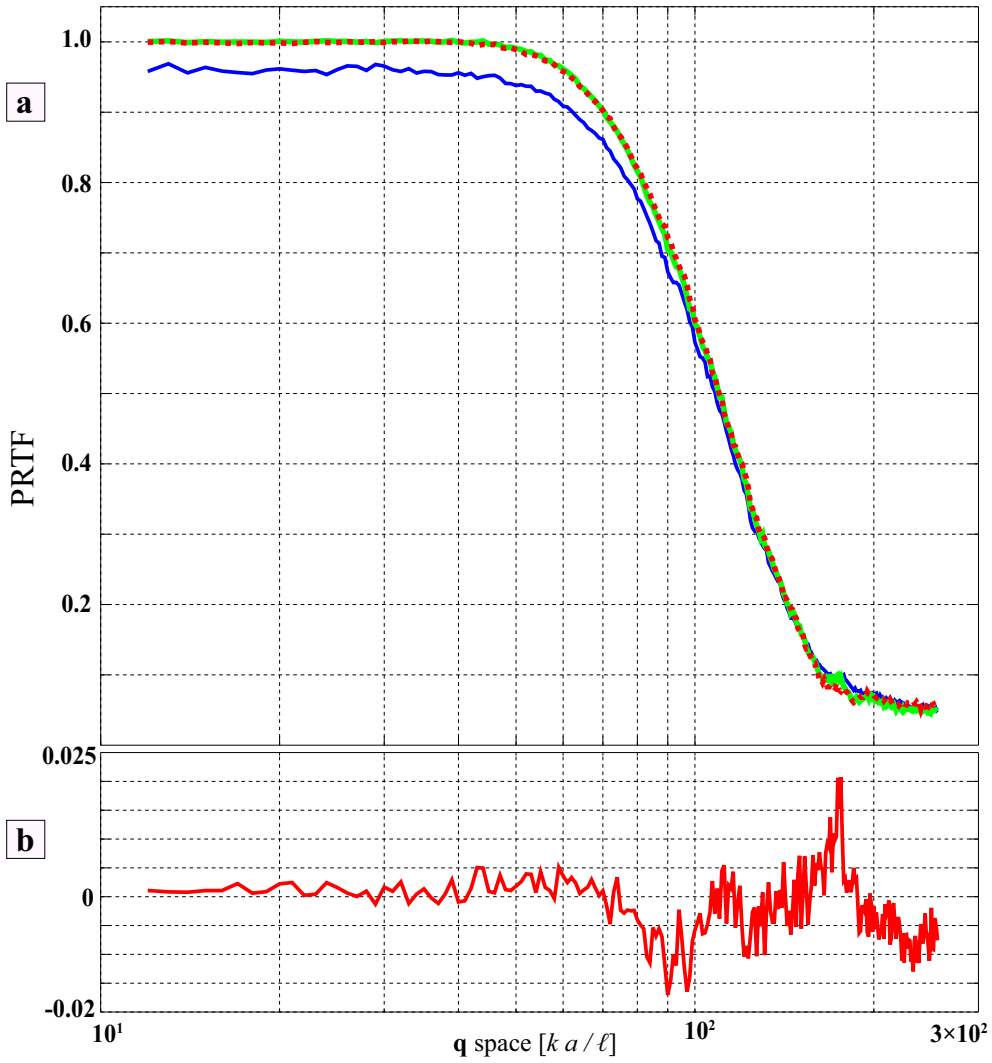

Figure 3.27: PRTFs of reconstructions when using 65% assumed overlap, "good" signal to noise, and with $\pm 40\sqrt{2}$ pixel maximum scan position errors. (a) The blue curve corresponds to the PRTF for the reconstruction in Fig. 3.15c-d, which in turn corresponds to using the recovered scan positions from Fig. 3.11a. The dashed red curve corresponds to the PRTF for the reconstruction in Fig. 3.15g-h, which in turn corresponds to using the recovered scan positions from Fig. 3.11b. The green curve corresponds to reconstructions for which the scan locations were assumed perfectly known. (b) The difference between the dashed red PRTF curve and the solid green PRTF curve; the percent error is less than 2.5 %.

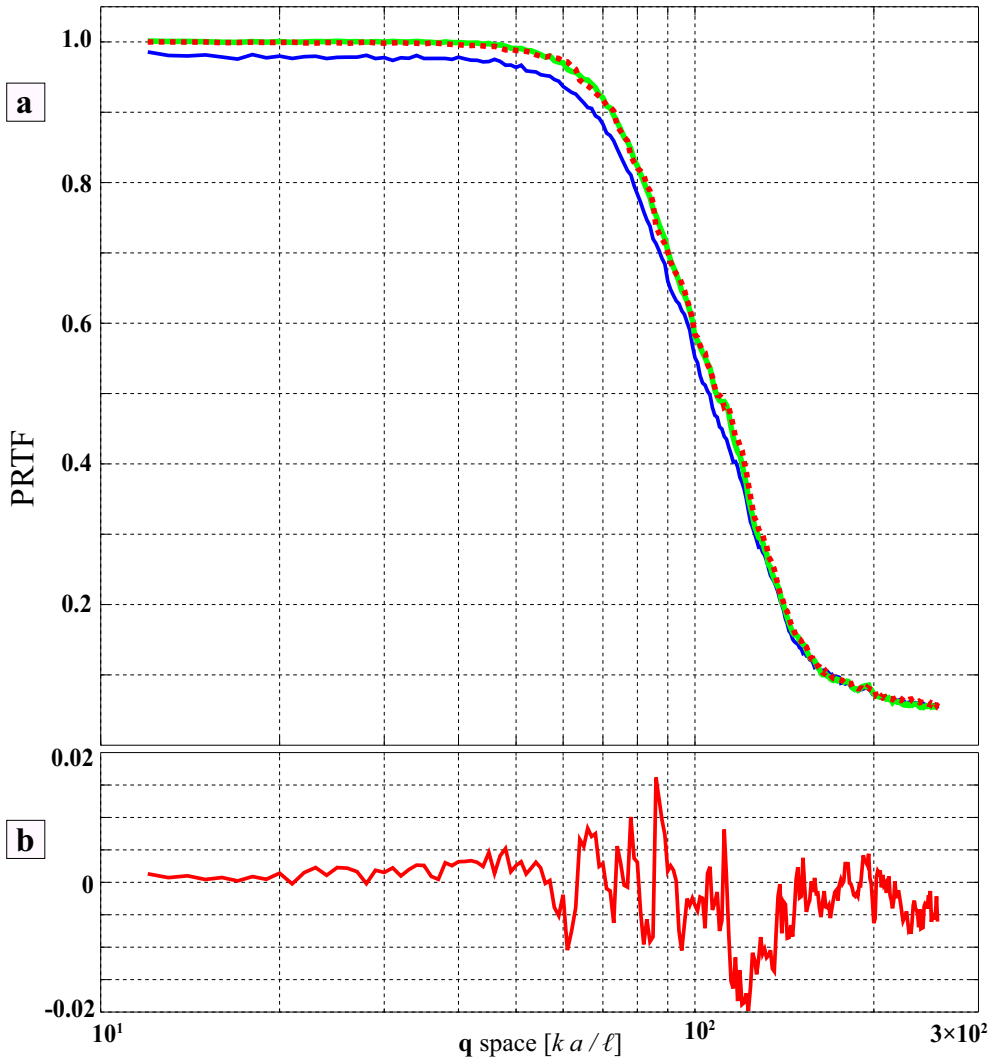

Figure 3.28: PRTFs of reconstructions when using 75% assumed overlap, "good" signal to noise, and with $\pm 60\sqrt{2}$ pixel maximum scan position errors. (a) The blue curve corresponds to the PRTF for the reconstruction in Fig. 3.17c-d, which in turn corresponds to using the recovered scan positions from Fig. 3.8a. The dashed red curve corresponds to the PRTF for the reconstruction in Fig. 3.17g-h, which in turn corresponds to using the recovered scan positions from Fig. 3.8b. The green curve corresponds to reconstructions for which the scan locations were assumed perfectly known. (b) The difference between the dashed red PRTF curve and the solid green PRTF curve; the percent error is less than 2.0 %.

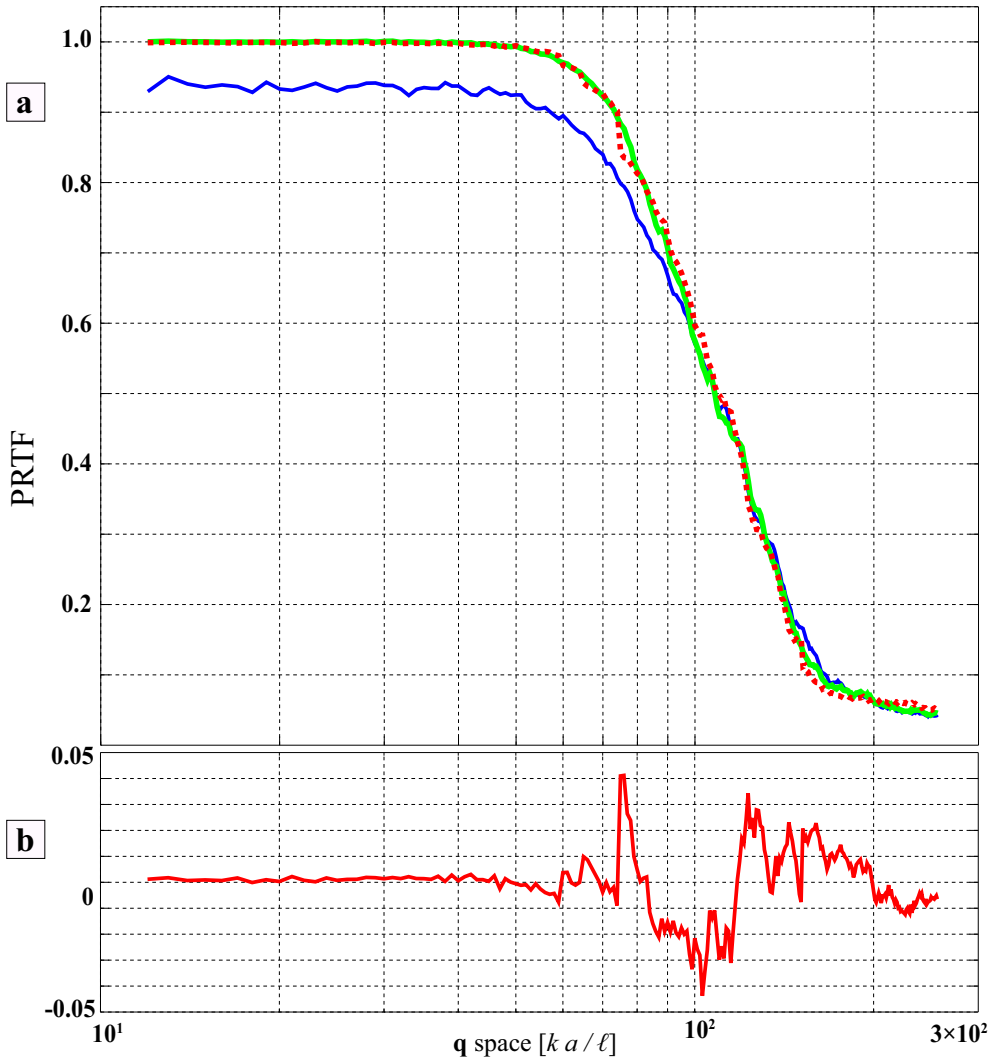

Figure 3.29: PRTFs of reconstructions when using 85% assumed overlap, "good" signal to noise, and with $\pm 90\sqrt{2}$ pixel maximum scan position errors. (a) The blue curve corresponds to the PRTF for the reconstruction in Fig. 3.19c-d, which in turn corresponds to using the recovered scan positions from Fig. 3.10a. The dashed red curve corresponds to the PRTF for the reconstruction in Fig. 3.19g-h, which in turn corresponds to using the recovered scan positions from Fig. 3.10b. The green curve corresponds to reconstructions for which the scan locations were assumed perfectly known. (b) The difference between the dashed red PRTF curve and the solid green PRTF curve; the percent error is less than 5.0 %.

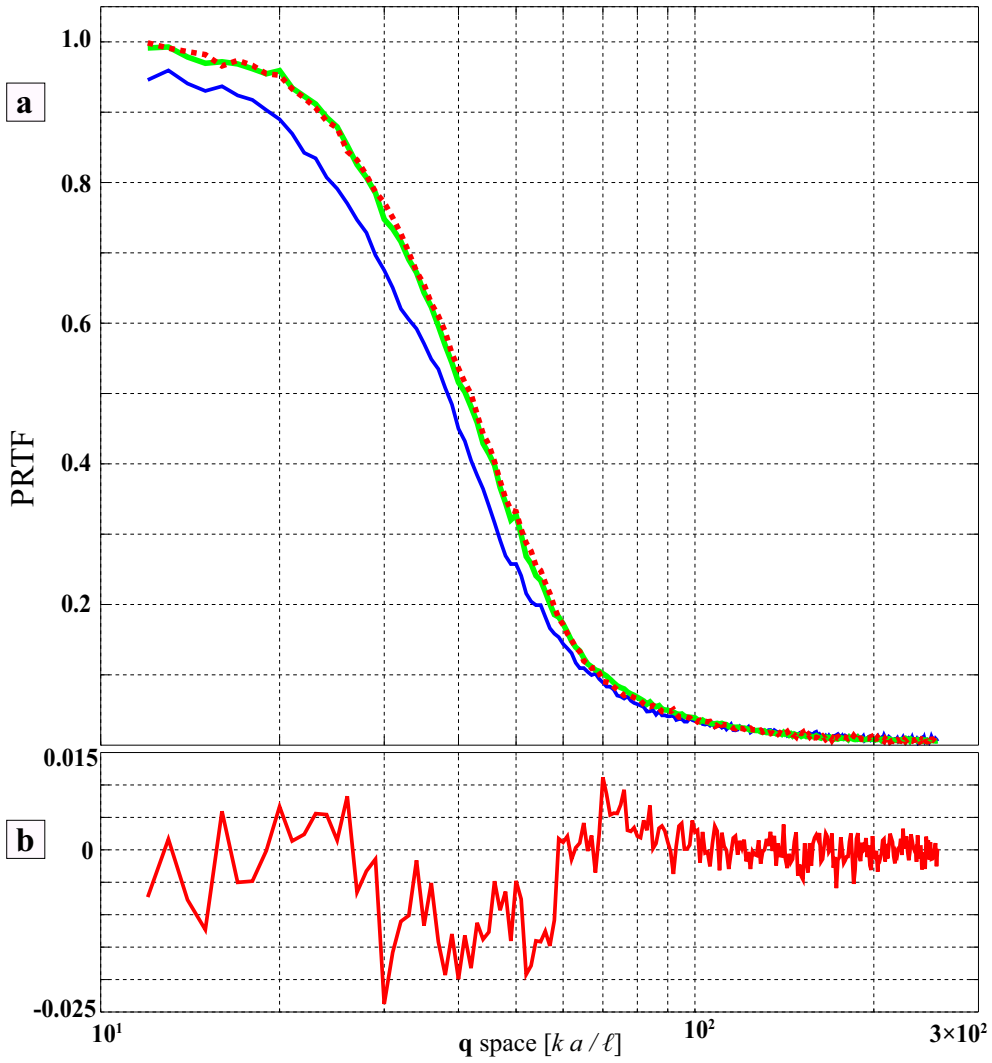

Figure 3.30: PRTFs of reconstructions when using 65% assumed overlap, "poor" signal to noise, and with $\pm 20\sqrt{2}$ pixel maximum scan position errors. (a) The blue curve corresponds to the PRTF for the reconstruction in Fig. 3.21c-d, which in turn corresponds to using the recovered scan positions from Fig. 3.14a. The dashed red curve corresponds to the PRTF for the reconstruction in Fig. 3.21g-h, which in turn corresponds to using the recovered scan positions from Fig. 3.14b. The green curve corresponds to reconstructions for which the scan locations were assumed perfectly known. (b) The difference between the dashed red PRTF curve and the solid green PRTF curve; the percent error is less than 1.5 %.

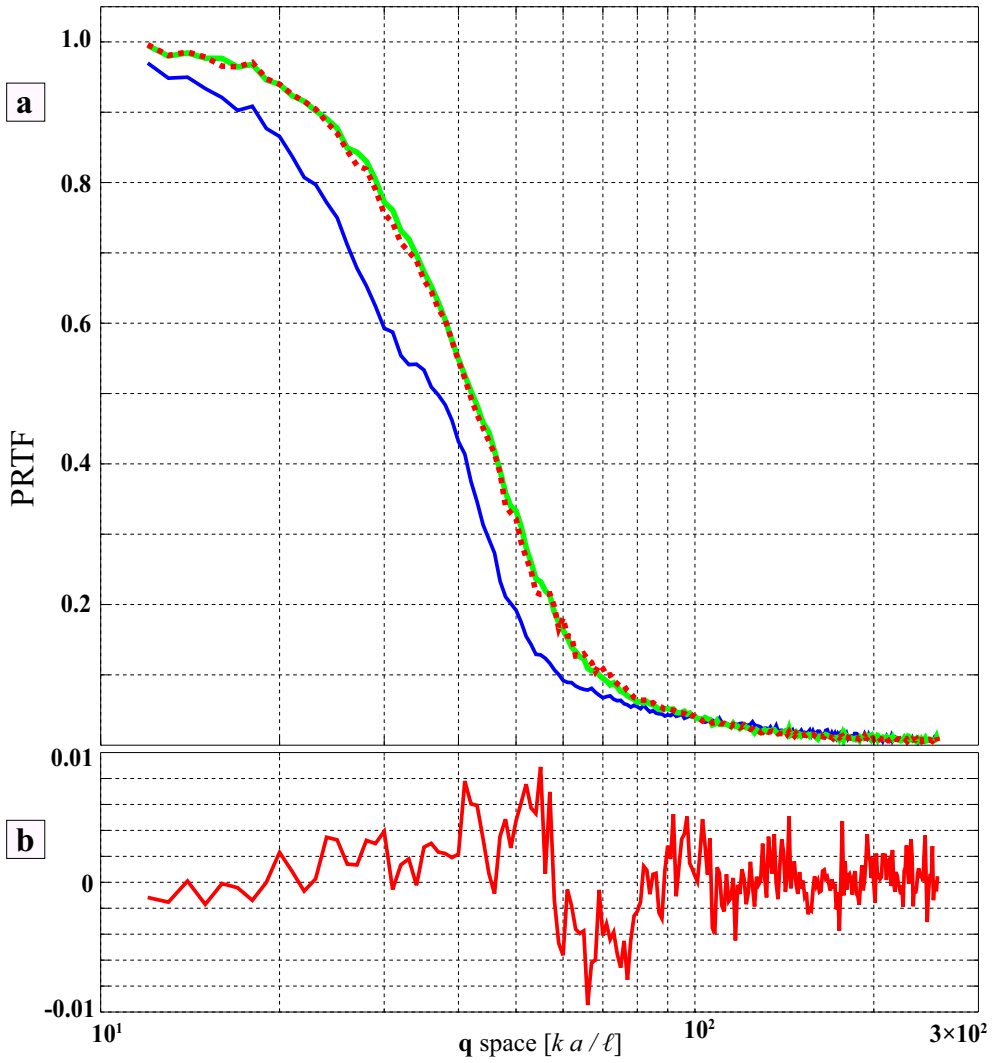

Figure 3.31: PRTFs of reconstructions when using 75% assumed overlap, "poor" signal to noise, and with $\pm 40\sqrt{2}$ pixel maximum scan position errors. (a) The blue curve corresponds to the PRTF for the reconstruction in Fig. 3.23c-d, which in turn corresponds to using the recovered scan positions from Fig. 3.12a. The dashed red curve corresponds to the PRTF for the reconstruction in Fig. 3.23g-h, which in turn corresponds to using the recovered scan positions from Fig. 3.12b. The green curve corresponds to reconstructions for which the scan locations were assumed perfectly known. (b) The difference between the dashed red PRTF curve and the solid green PRTF curve; the percent error is less than 1.0 %.

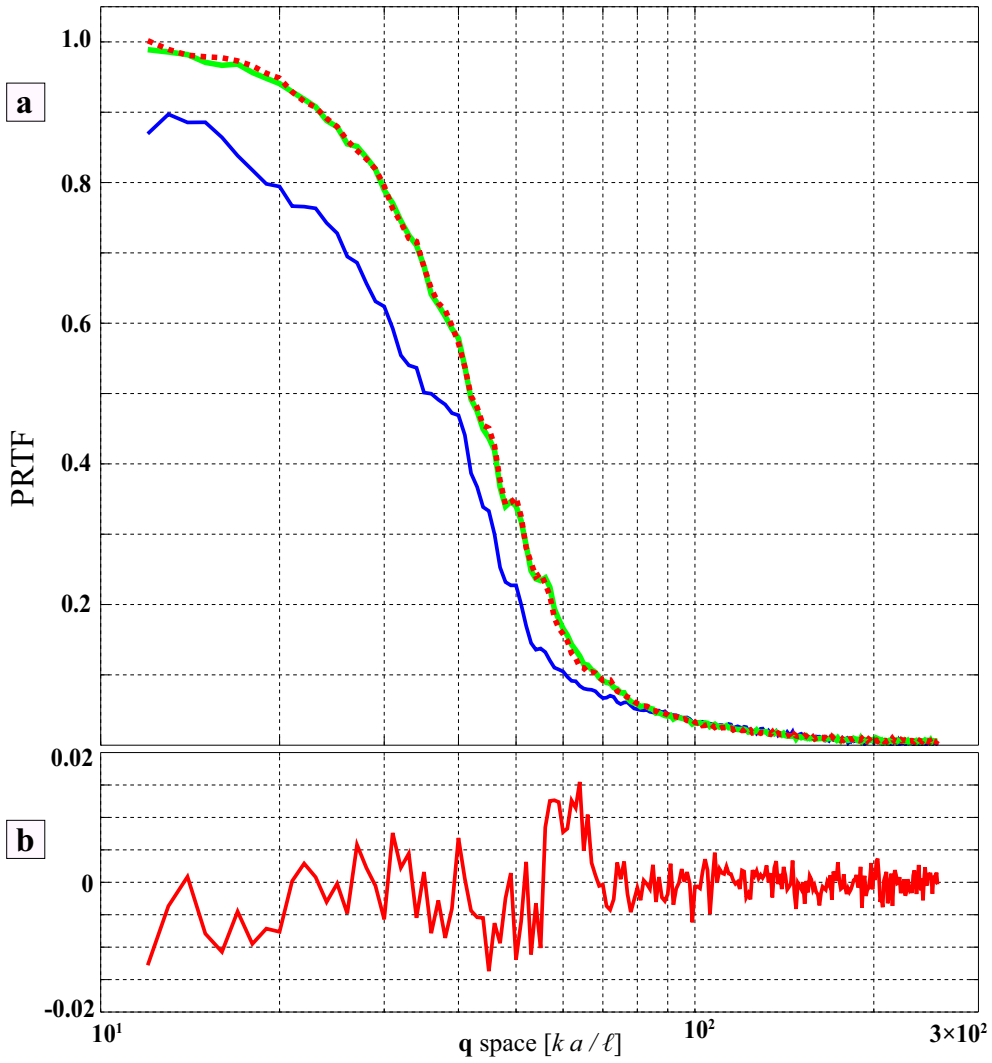

Figure 3.32: PRTFs of reconstructions when using 85% assumed overlap, "poor" signal to noise, and with $\pm 60\sqrt{2}$ pixel maximum scan position errors. (a) The blue curve corresponds to the PRTF for the reconstruction in Fig. 3.25c-d, which in turn corresponds to using the recovered scan positions from Fig. 3.13a. The dashed red curve corresponds to the PRTF for the reconstruction in Fig. 3.25g-h, which in turn corresponds to using the recovered scan positions from Fig. 3.13b. The green curve corresponds to reconstructions for which the scan locations were assumed perfectly known. (b) The difference between the dashed red PRTF curve and the solid green PRTF curve; the percent error is less than 2.0 %.

which are illuminated. A computationally simple method for correcting large errors in this scan position constraint is presented, which is easily integrated into existing algorithms used to numerically invert a ptychographic diffraction data set. Depending on how tightly "clustered" a scanned region on the sample is and the signal to noise of the diffraction, errors of up to ~ 45 % of the diameter of the illumination function $p(\mathbf{r})$ can be nearly perfectly recovered upon performing many independent runs of the iterative scan position refinement procedure and averaging the results. In this way, we have an effective method for achieve diffraction limited imaging even in experiments where vibration and sample drift can seriously degrade the spatial resolution and quantitative index of refraction of the reconstructions.

Chapter 3, in part, is a preprint of material being prepared for submission for publication: A. Tripathi, I. McNulty, O. G. Shpyrko, "Ptychographic overlap constraint errors and diffraction limited imaging". Ashish Tripathi is the primary investigator and author of this paper.

Chapter 4

X-ray interaction with magnetism

Experimentally, the type of samples this thesis will deal with are known as magnetic multilayers. What follows in the next few sections will be: magnetic multilayers will be introduced and physics of the multilayers will be explored. Next, the formalism for x-ray scattering from the magnetic moments comprising the multilayer will be laid out. In doing this, we will first cover classical x-ray electron charge scattering and derive an expression for the index of refraction we encountered in Eq 1.15. Next, the index of refraction for a quantum mechanical modeling of the interactions of x-rays with electron charge structure and magnetic moment structure will be introduced, and techniques for modeling x-ray diffraction from the multilayers will be explored.

4.1 Magnetic multilayers

Magnetic multilayers are engineered nanostructures which can display exotic magnetic domain configurations, phase separation, and reaction-diffusion behavior. They are made by depositing alternating layers of different types of magnetic elements, as seen in Fig. 4.1a. The magnetic domains in these samples arise from an ordering of atomic magnetic moments; the interactions between magnetic moments which cause this ordering are discussed here. The most important magnetic elements are the $3d$ and $4f$ elements. Some $4f$ elements are gadolinium (Gd), terbium (Tb) and holmium (Ho); these are known as rare earth (RE) elements.

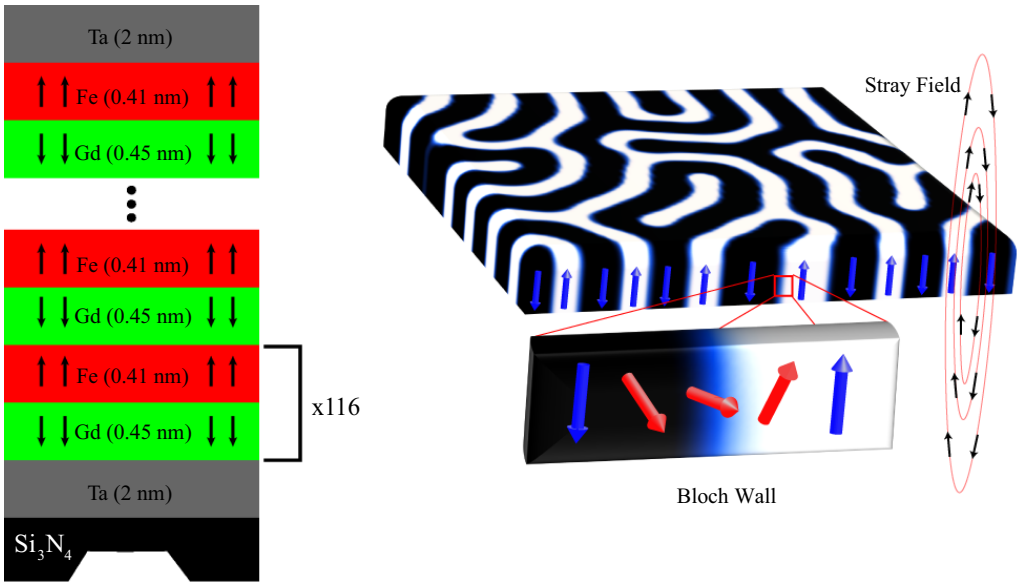

Figure 4.1: Magnetic multilayers. They are engineered nanostructures which can display exotic magnetic domain configurations, phase separation, and reaction-diffusion behavior

Some important $3d$ magnetic elements are iron (Fe), cobalt (Co), and nickel (Ni); these are known as the transition metals (TM). In RE-TM thin films (GdFe will be the type of film studied in this thesis) the magnetic domain structure splits into two ferromagnetic phases, one pointing up and the other pointing down as seen in Fig. 4.1. The interfaces between these oppositely aligned domains are known as Bloch domain walls; this is where the local magnetization direction changes from one alignment to the other. These films display what is known as perpendicular magnetic anisotropy (PMA) which means that the magnetization directions of the moments tend to align perpendicular to the surface of the film. The magnetic properties of these films can be changed by modifying their thickness, stoichiometry, and number of individual layers.

Domain formation in these types of films can be understood by the competition of various interactions in a Hamiltonian given by [42]:

$$\mathcal{H} = \int \left\{ A(\nabla \hat{\mathbf{m}})^2 + K_u(1 - m_z^2) - \mu_0 \mathbf{H}_d \cdot \mathbf{M} - \mu_0 \mathbf{H}_{ext} \cdot \mathbf{M} \right\} d^3\mathbf{r}, \tag{4.1}$$

where A is known as the exchange stiffness constant (a temperature dependent material parameter), $\hat{\mathbf{m}}$ is the magnetization unit vector that points in the direction of

the local magnetization $\mathbf{M}$, K_u is the uniaxial anisotropy constant which describes the strength of the perpendicular magnetic anisotropy and is dependent on sample growth parameters, m_z is the component of $\hat{\mathbf{m}}$ in the direction perpendicular to the film surface, μ_0 is the vacuum magnetic permeability, $\mathbf{H}_d$ is the stray magnetic field as a result of the Bloch walls, and $\mathbf{H}_{ext}$ is used to denote any external magnetic field present. Equation 4.1 models the energy of an infinitely extended defect-free film, and assumes that any magnetostriction or external strain is negligible. The $A(\nabla\hat{\mathbf{m}})^2$ term is known as the exchange interaction, and has the effect of aligning the magnetic moments in one direction (any direction). The exchange interaction term $A(\nabla\hat{\mathbf{m}})^2$ is inversely proportional to the nearest-neighbor distance of atoms. The exchange energy can be seen as an energy cost due to the deviation of the zero exchange energy state, i.e. a state of uniform magnetization (a single ferromagnetic domain). Typically the direction of the uniform magnetization doesn't affect this exchange energy term. The $K_u(1-m_z^2)$ term is known as the uniaxial anisotropy interaction and attempts to force the magnetic moments to point along the perpendicular direction. The $K_u(1-m_z^2)$ term reflects the fact that atoms in an anisotropic surrounding can result in a preferred magnetization direction and any moments not aligned in this direction can result in an anisotropy energy penalty. While the exchange interaction tries to align the magnetic moments in the material to any uniform direction, the anisotropy interaction tries to align them in a preferred magnetization direction known as the "easy" axis (or direction). The $\mu_0\mathbf{H}_{ext}\cdot\mathbf{M}$ term is known as the the stray field interaction, or alternatively the shape anisotropy interaction, and describes the energetic penalty of magnetic stray fields $\mathbf{H}_d$ outside a magnetized object. It favors in plane magnetization and has the effect of breaking up a uniform magnetization state into magnetic domains. The shape anisotropy interaction is dependent on the shape of the sample and is important in thin magnetic films due to a large area of free magnetic poles at surfaces. In the limit of an infinite thin film which is uniformly magnetized in the direction normal to the film surface, we have $\mathbf{H}_d = -\mathbf{M}$ and the shape anisotropy energy density has a maximum value of $K_d = \mu_0\mathbf{M}^2/2$, which is known as the stray field energy constant. To reduce this energetically unfavorable situation, the

magnetization direction breaks up into magnetic domains at the energetic cost of the formation of domain walls. This has the net effect of reducing $\mathbf{M}$ and thus the shape anisotropy energy. Lastly, $\mu_0 \mathbf{H}_{ext} \cdot \mathbf{M}$ terms in the Zeeman interaction which tries to align the magnetic moments along the direction of the external field.

Magnetic multilayers are interesting materials from a technological stand point; they are integral parts of next-generation magnetic data storage and spintronic technologies [43, 44]. They allow the realization of unique micromagnetic, magneto-optic, and magneto-electronic phenomena that cannot be achieved using conventional materials. In the realm of magnetic data storage, the orientation of the domain will act in the role of a magnetic bit (1 or 0). These magnetic bits can be coded as domains in a continuous storage layer, for example small regions where the magnetic moments are pointing in either one of two directions. In current hard drive technology, the domains are "written" by a device which applies a local magnetic field generated by an electrical current pulse through a coil. Reading of the domains is accomplished by measuring the electrical conductivity through a tiny sensing device consisting of a layered structure of various materials. Crucially, one of these layers in the reading device displays giant magnetoresistant properties. Giant magnetoresistance is the phenomenon in which conductivity is extremely dependent on the local magnetic field generated by the magnetic domain in which the reading device is currently looking at. In this scheme, changes in the magnetic domain direction of the storage layer causes a change in conductivity of the "read head" sensor which is interpreted as a 1, while if the domain is unchanging over a defined bit length size, it is interpreted as a 0. For this type of technology, magnetic multilayers offer a very tunable range of systems with which to explore giant magnetoresistance physics [45].

Magnetic multilayers also exhibit self organizing, reaction-diffusion, and pattern forming behavior not fully understood. Some partial differential equations with which these materials can be qualitatively modeled by are known as the Swift-Hohenberg equation [46]:

$$\frac{\partial u(x,y,t)}{\partial t} = ru(x,y,t) - (1+\nabla^2)^2 u(x,y,t) + N(u(x,y,t)) \tag{4.2}$$

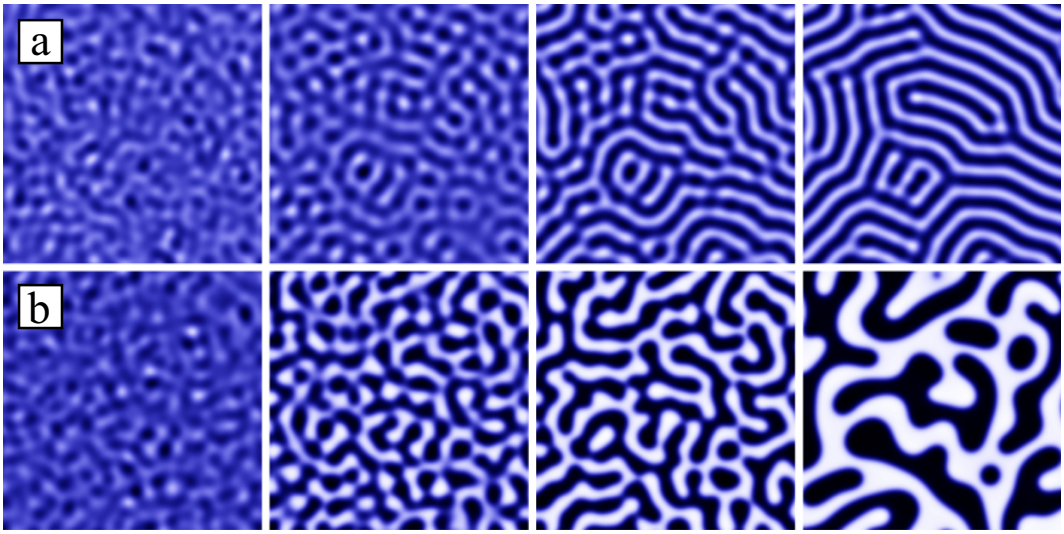

Figure 4.2: Reaction-diffusion and phase separation behavior of the Swift-Hohenberg (top) and Cahn-Hilliard equations (bottom).

and the Cahn-Hilliard equation [47]:

$$\frac{\partial u(x,y,t)}{\partial t} = D\nabla^2 \left[u(x,y,t)^3 - u(x,y,t) - \gamma\nabla^2 u(x,y,t)\right] \tag{4.3}$$

where $u = u(x,y,t)$ is a scalar function which describes an order parameter which undergoes phase separation, r is a real valued bifurcation parameter, $N(u)$ is some smooth nonlinearity function, D is a diffusion coefficient, and $\sqrt{\gamma}$ determines the width of the transition regions between oppositely aligned domains. The Swift-Hohenberg equation models more closely a relatively defect (or impurity) free magnetic multilayer while the Cahn-Hilliard more closely models a magnetic multilayer with impurities, e.g. a GdFe layer with Tb doping. A numerical solution of Eq. 4.2 for $u(x,y,t)$ as a function of t is shown in Fig. 4.2a while a numerical solution of Eq. 4.3 for $u(x,y,t)$ as a function of t is shown in Fig. 4.2b.

4.2 Resonant x-ray scattering

A good classical model for x-rays traveling through a material is the x-ray electric field polarization interacting with bound electrons [48]. An atom is modeled as a massive, positively charged (with charge $+Ne$) nucleus surrounded

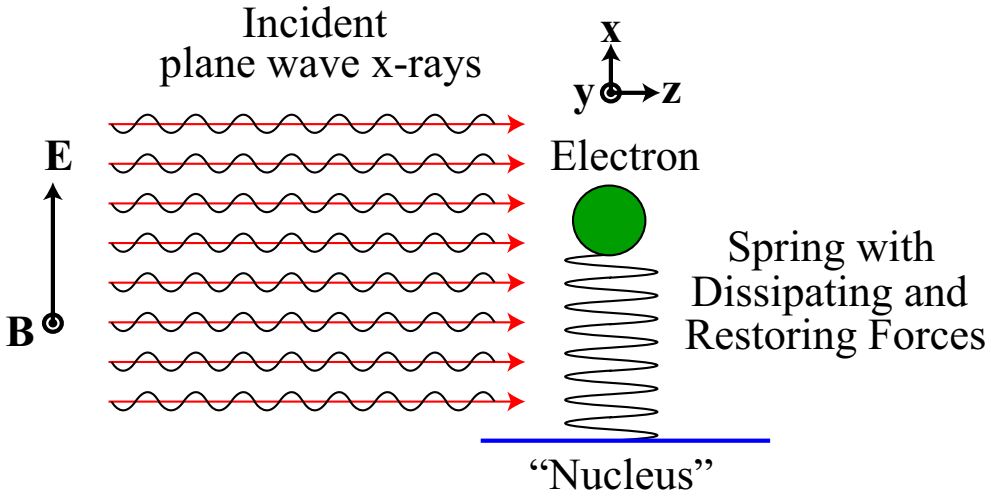

Figure 4.3: Schematic illustration of an electron "bound" to a nucleus via a spring (with restoring and dissipative forces), and being driven by an external electromagnetic field.

by N low mass, negatively charged electrons. When an external electromagnetic wave (i.e. the incident x-rays) $\mathbf{E}_i = \hat{\mathbf{x}} E_0 e^{-i\omega t}$ pass through this atom, as shown in Fig. 4.3, the bound electrons will begin to move under the influence of the x-ray polarization field, and begin oscillate, driven by the imposed electromagnetic field, but also damped by the nuclear restoring force. An electrons response to the x-ray polarization field depends on both the electrons resonant frequency ω_r and the frequency of the incident x-rays, denoted by ω. In this model, Newtons second law reads:

$$\frac{d^2\mathbf{x}}{dt^2} + \gamma \frac{d\mathbf{x}}{dt} + \omega_r^2 \mathbf{x} = -\frac{e}{m_e} \left(\mathbf{E}_i + \mathbf{v} \times \mathbf{B}_i \right), \tag{4.4}$$

where m_e is the mass of electron, e is the electron charge, $\mathbf{x}$ is the displacement, and γ is the damping ratio. The first term in Eq. 4.4 is the acceleration contribution, the second term is the dissipative force term which accounts for energy loss, and the third term is caused by the restoring force for an oscillator of resonant frequency ω_r. The term on the right side is the Lorentz force imposed by the external field. As the oscillation is driven by the incident electromagnetic field, the displacement $\mathbf{x}$, velocity $\mathbf{v} = d\mathbf{x}/dt$, and acceleration $d^2\mathbf{x}/dt^2$ all have the same time dependence $e^{-i\omega t}$ as the incident field, along with possible constant phase offsets. For non-relativistic oscillation velocities $\mathbf{v}$, the $\mathbf{v} \times \mathbf{B}_i$ term is negligible. In that case, Eq.

4.4 can be simplified to

$$\omega^2\mathbf{x} + i\omega\gamma\mathbf{x} - \omega_r^2\mathbf{x} = \frac{e}{m_e}\mathbf{E}_i, \tag{4.5}$$

and so the oscillation of the electron is

$$\mathbf{x} = \hat{\mathbf{x}}\frac{e}{m_e}\frac{1}{\omega^2 - \omega_r^2 + \omega\gamma}E_0 e^{-i\omega t}. \tag{4.6}$$

Going from a single electron to a bulk material with electron number density ρ (N electrons in a volume V), and comprised of molecules with f_i electrons with resonant frequency ω_j and damping γ_j, we next approximate the material polarization density as

$$\mathbf{P}(t) = \epsilon_0\chi\mathbf{E}(t) = (\epsilon - \epsilon_0)\mathbf{E}(t), \tag{4.7}$$

where $\chi = (\epsilon/\epsilon_0 - 1)$ is the complex electric susceptibility, and when the N electrons are displaced by $x(t)$, we have $P(t) = -x(t)Ne/V$, or

$$\frac{P(t)}{E(t)} = (\epsilon - \epsilon_0) = \frac{e^2\rho}{m_e}\sum_j \frac{f_j}{\omega^2 - \omega_{rj}^2 + \omega\gamma_j}. \tag{4.8}$$

An expression for the index of refraction $n = \epsilon/\epsilon_0$ can now be found, where the relative permeability μ/μ_0 is assumed unity:

$$n = \left[1 - \frac{e^2\rho}{\epsilon_0 m_e}\sum_j \frac{f_j}{\omega^2 - \omega_{rj}^2 + i\omega\gamma_j}\right]^{1/2}, \tag{4.9}$$

In the x-ray region, $\omega^2 \gg e^2\rho/\epsilon_0 m_e$, so we can expand Eq. 4.9 as

$$n = 1 - \frac{e^2\rho}{2\epsilon_0 m_e}\sum_j \frac{f_j}{\omega^2 - \omega_{rj}^2 + i\omega\gamma_j}. \tag{4.10}$$

This is generally written as

$$n = 1 - \delta + i\beta = 1 - \frac{e^2\rho}{2\epsilon_0 m_e}\left(f_1 - if_2\right), \tag{4.11}$$

with

$$f_1 = \sum_j \frac{f_j(\omega^2 - \omega_{rj}^2)}{(\omega^2 - \omega_{rj}^2)^2 + (\omega\gamma_j)^2}, \tag{4.12}$$

and

$$f_2 = \sum_j \frac{f_j\gamma_j\omega}{(\omega^2 - \omega_{rj}^2)^2 + (\omega\gamma_j)^2}. \tag{4.13}$$

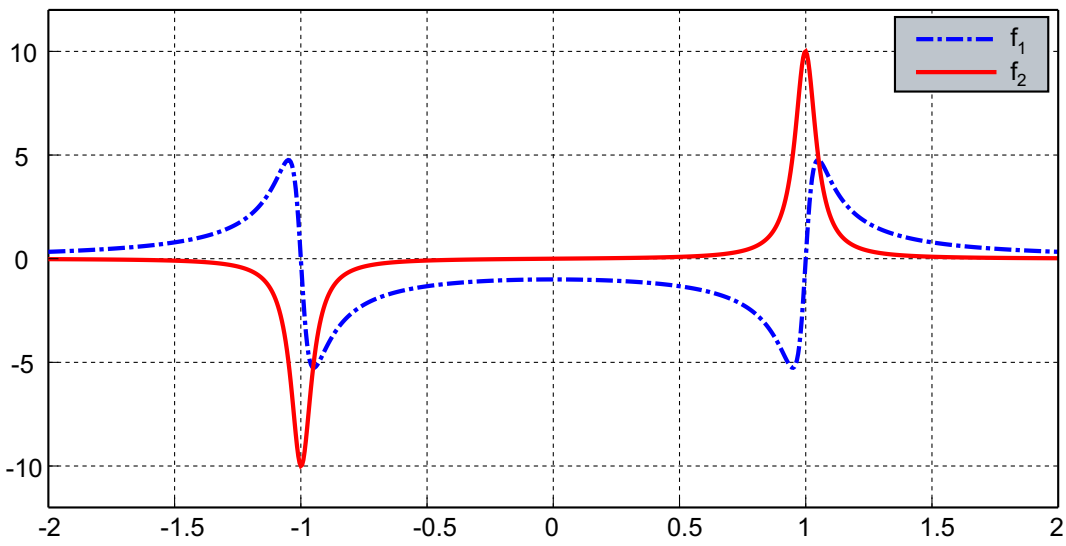

Figure 4.4: The real and imaginary parts of the dispersion correction to the index of refraction given by Eq. 4.10 as a function of ω/ω_r. Here, $\gamma = 0.1\ \omega_r$ and only one oscillator is assumed.

The $f_1 - if_2$ term above is known as the "dispersion" correction to the refractive index. It is known as this because away from resonances at ω_j, the index of refraction generally increases with increasing ω. In other words, interaction with electrons in a material become weaker as x-ray energy increases. This can be seen by looking at when $\omega \gg \omega_j \gg \gamma_i$

$$n = 1 - \frac{e^2\rho}{2\epsilon_0 m_e \omega^2}\sum_i f_i = 1 - \frac{e^2\rho}{2\epsilon_0 m_e \omega^2} Z = 1 - \frac{2\pi\rho r_e}{k^2} Z, \tag{4.14}$$

where $r_e = e^2/4\pi\epsilon_0 mc^2$ is the Thompson scattering length, $k = \omega/c$ is the vacuum wavenumber, and Z is the number of electrons in the molecules which comprise the material. Clearly, as $\omega \to \infty$, $n \to 1$, i.e. the x-ray electron interactions gets weaker. However, when in the vicinity of the resonances at ω_j, we begin to see features such as that shown in Fig. 4.4, which can increase the interactions of x-rays with bound electrons by factors with order of magnitude of 10 to 100, depending on the element (i.e. Z).

4.3 Resonant x-ray magnetic scattering

In this section the formalism for x-ray and magnetic moment interactions is explored. A quantum mechanical description of the interaction of light with matter is the non-relativistic approximation of the Dirac equation [49]. The Hamiltonian used for this description describes an atomic electron in a quantized electromagnetic field and is given by:

$$\begin{aligned}\mathcal{H} &= \overbrace{\frac{1}{2m}\left(\mathbf{p}-\frac{e}{c}\mathbf{A}\right)^2}^{\text{kinetic energy}} + \underbrace{V_0(\mathbf{r})}_{\text{potential energy}} - \overbrace{\frac{e\hbar}{2mc}\mathbf{s}\cdot\nabla\times\mathbf{A}}^{\text{Zeeman interaction}} \\ &= \underbrace{-\frac{e\hbar}{2(mc)^2}\mathbf{s}\cdot\mathbf{E}\times\left(\mathbf{p}-\frac{e}{c}\mathbf{A}\right)}_{\text{spin-orbit interaction}} + \underbrace{\sum_{\mathbf{k}\sigma}\hbar\omega_{\mathbf{k}}c^{\dagger}(\mathbf{k}\sigma)c(\mathbf{k}\sigma)}_{\text{free radiation field}},\end{aligned} \tag{4.15}$$

where $\mathbf{p}$ is the electron momentum operator, $\mathbf{A}$ is the vector potential, $\mathbf{s}$ is the electron spin angular momentum, $\mathbf{E}$ is the electric field of the atom, and $c^{\dagger}$ and c are photon annihilation and creation operators. From this Hamiltonian, a scattering amplitude can be derived [50, 51] which we will use in this chapter to find expressions for the differential scattering cross section. These in turn will be used to understand experimentally measured diffraction intensities encountered in Chapter 5.

4.3.1 Magnetic scattering amplitude

The previous section introduced the classical theory x-ray interaction with electrons "bound" to a nucleus by a very tiny "spring" to derive the index of refraction. This formalism is adequate for a wide range of problems, but when looking at physics which is quantum mechanical in nature (e.g. electron spin), quantum theory must be used to model the x-ray interaction with matter. The theory of x-ray polarization interactions with magnetically ordered material was first described in [52, 53, 54, 55, 56]. In these, relativistic quantum effects were included in the elastic scattering cross section which predicted that interaction between photon polarization and magnetic moment ordering in a material can have a strong resonant increase. When a polarized electromagnetic field propagates

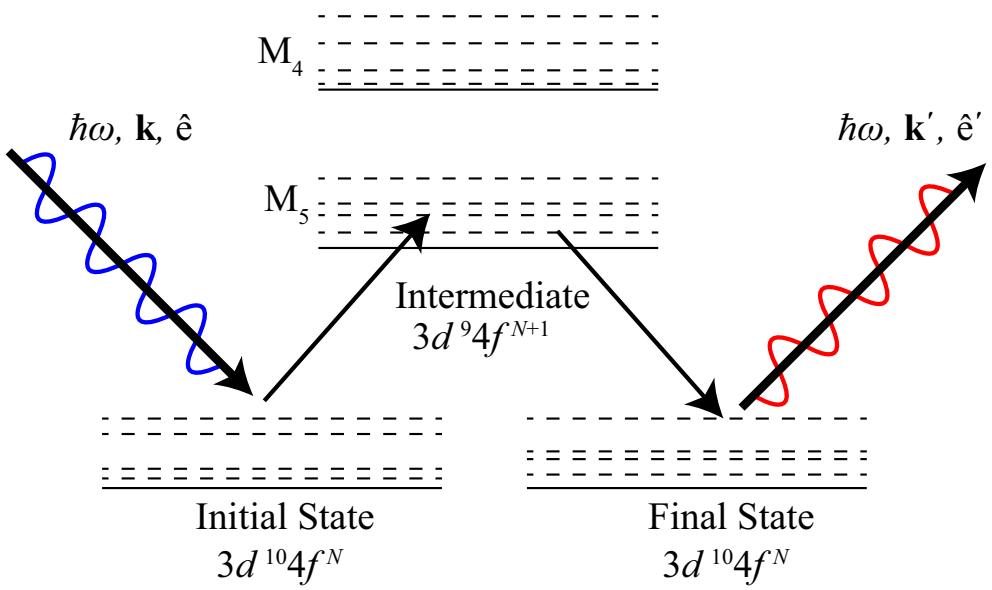

Figure 4.5: Cartoon of an electron in a $4f$ material (here Gd^{3+}) undergoing resonant elastic scattering at the M5 absorption edge. Incoming x-rays with polarization $\hat{\mathbf{e}}$ trigger virtual dipole transitions of a $3d$ electron to an empty $4f$ state above the Fermi level. The electron then instantly de-excites back to the initial state and an x-ray with the same energy but possibly different polarization $\hat{\mathbf{e}}'$ and propagation direction $\mathbf{k}'$.

through a material, the electrons bound to a nucleus begin to oscillate in an electric multipole transition. If the electrons in the atom have a magnetic moment and spin-orbit splitting, then angular momentum of the incident photons interacts with the spin and orbital angular momentum of the electrons. The spin-orbit effect is necessary because the photon polarization (i.e. angular momentum) state interacts directly with electron orbital angular momentum; it only indirectly interacts with electron spin via the orbital angular momentum. Conservation of energy and angular momentum must be enforced, and so only certain selection rules of the electric multipole transition are allowed. This leads to electronic ground to valence excitation probabilities which are dependent on photon angular momentum states and electron spin and orbital angular momentum states; a cartoon of these transitions is shown in Fig. 4.5.

When the incident photons have angular momentum (i.e. are right or left circularly polarized), and they interact with an atom with a magnetic moment

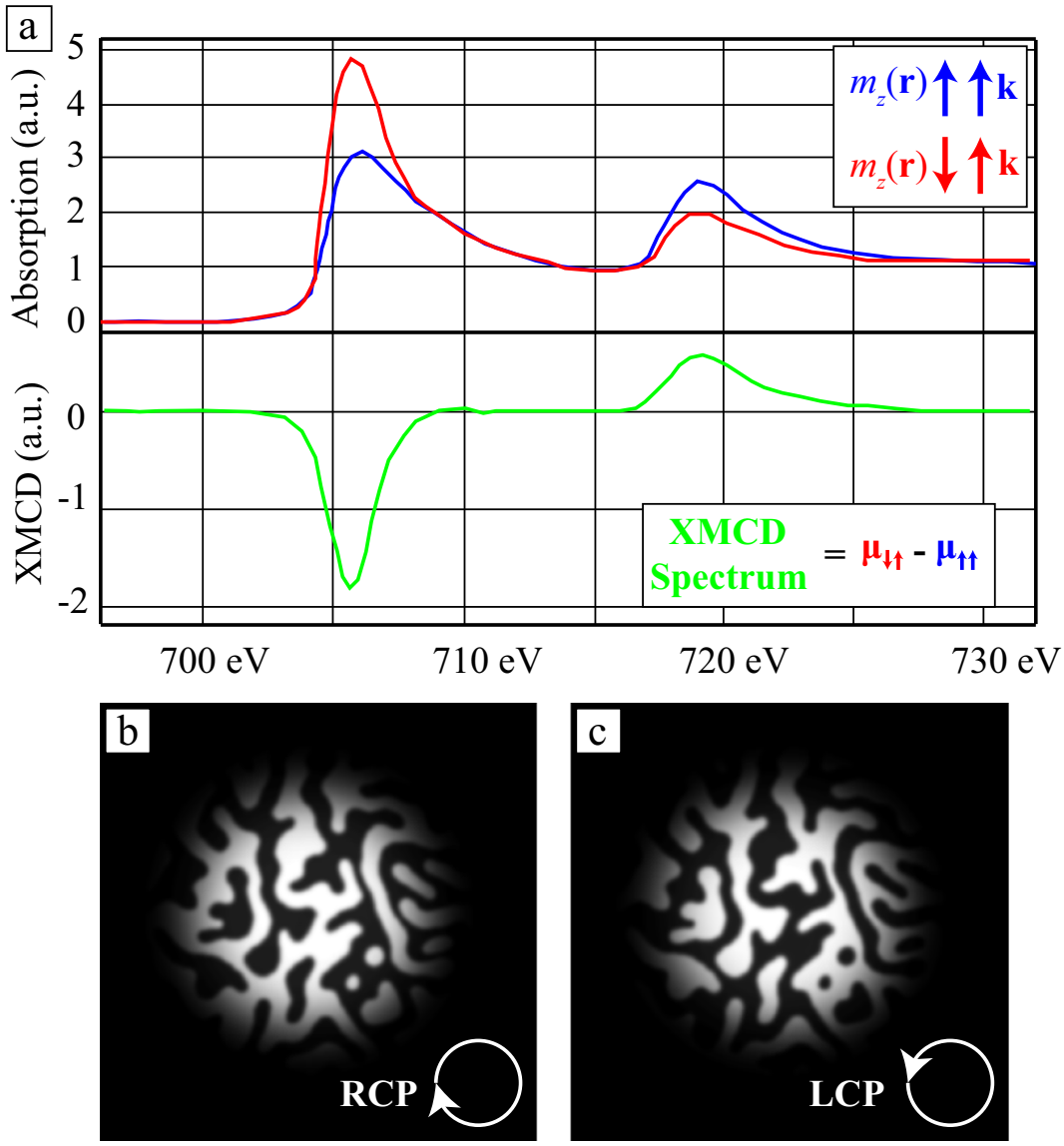

Figure 4.6: (a) Example of absorption differences when illuminating a magnetic sample which exhibits XMCD behavior. Depending on the orientation of $m_z(\mathbf{r})$ (for example $m_z(\mathbf{r}) = +1$ for the black domains in frame (b) while $m_z(\mathbf{r}) = -1$ for the white domains), we can have two different absorption values, due to different electron excitation selection rules will be dependent on whether the incident photon and valence state angular momenta are parallel or antiparallel. The XMCD spectrum is defined as the difference between the spectra for RCP and LCP light. Adapted from [57]. (b) A simulated sample with XMCD contrast for RCP incident light. (c) A simulated sample with XMCD contrast for LCP incident light. Notice that it is the inverse of that for RCP light.

(orbital and/or spin), photon angular momentum will be completely transferred to any excited electrons. Assuming that the x-ray energy is sufficient to excite this electron to a valence state, and if the valence state has a preferred spin/orbit angular momentum state (dependent on the shape of the Fermi surface of the material), photon-electron excitation selection rules will be dependent on whether the photon and valence state angular momenta are parallel or antiparallel. This is origin of the XMCD effect, and explains why there are different absorption spectra for right or left circularly polarized radiation. In Fig. 4.6a, a typical XMCD absorption spectra is shown. Also, in Fig. 4.6b, a simulated sample with XMCD contrast for RCP incident light is shown, while in Fig. 4.6c, a simulated sample with XMCD contrast for LCP incident light. Notice that it is the inverse of that for RCP light.

Another possibility when x-rays scatter from a magnetic sample is if the incident photons have zero angular momentum (i.e. are linearly polarized). The excited electrons must have the same total (spin + orbital) angular momentum state as the ground state, and the electric multiplet selection rules give a transition probability that is maximum when the electron total angular momentum is parallel or antiparallel to the photon angular momentum, and is minimum when perpendicular. This is the origin of the XMLD effect.

The analytical result [52, 53, 54, 55, 56] for the scattering amplitude for polarized photons resonantly interacting with a magnetically ordered material, which incorporates resonant charge scattering and both XMCD and XMLD scattering is given by

$$\begin{aligned} f^{res} = F^{(0)}(\omega)\hat{\mathbf{e}}' \cdot \hat{\mathbf{e}}\; c(\mathbf{r}) - iF^{(1)}(\omega)(\hat{\mathbf{e}}' \times \hat{\mathbf{e}}) \cdot \hat{\mathbf{m}}(\mathbf{r}) \\ + F^{(2)}(\omega)(\hat{\mathbf{e}}' \cdot \hat{\mathbf{m}}(\mathbf{r}))(\hat{\mathbf{e}} \cdot \hat{\mathbf{m}}(\mathbf{r})), \end{aligned} \tag{4.16}$$

where $\hat{\mathbf{e}}$ and $\hat{\mathbf{e}}'$ are polarization unit vectors of the incident and transmitted radiation respectively, $\hat{\mathbf{m}}(\mathbf{r})$ is the magnetization direction (and is a unit vector), as shown by the blue and red arrows in Fig. 4.1, and $c(\mathbf{r})$ is the resonant charge contrast function. The $F^{(0,1,2)}$ are complex (dispersive and absorbtive) resonant scattering strengths (in units of r_e, the Thompson scattering length), with $F^{(0)}(\omega)$ describing any resonant charge scattering, $F^{(1)}(\omega)$ describing the strength

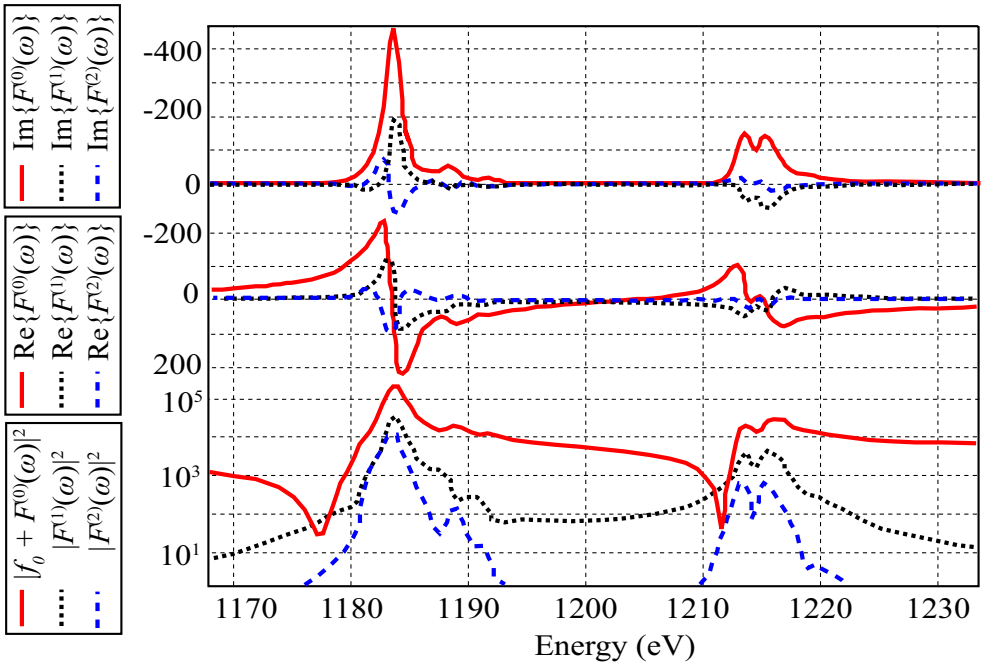

Figure 4.7: Calculated optical constants for Gd. Adapted from [58].

of XMCD and $F^{(2)}(\omega)$ describing the strength of XMLD. Calculated values of these for Gadolinium are shown in Fig. 4.7.

4.3.2 Stokes vectors and Poincaré representation

In order to analytically and numerically work with Eq. 4.16, some preliminary mathematical "machinery" must be introduced. We first start by defining the polarization ellipse, Stokes parameters and the Poincare representation for the polarization of light. Consider the experimental geometries in Fig 4.8; Fig 4.8a shows a diffraction experiment in the "transmission" geometry while Fig 4.8b shows a "reflection" geometry. Defining a plane for which both $\mathbf{k}$ and $\mathbf{k}'$ exist in (the light greenish planes in Fig 4.8a-b), we next express the incident electric field polarization as components parallel $\hat{\mathbf{e}}_\pi$ and perpendicular $\hat{\mathbf{e}}_\sigma$ to this plane. The outgoing electric field polarizations $\hat{\mathbf{e}}'_\pi$ and $\hat{\mathbf{e}}'_\sigma$ are defined similarly.

The polarization ellipse relates the components of the electric field polar-

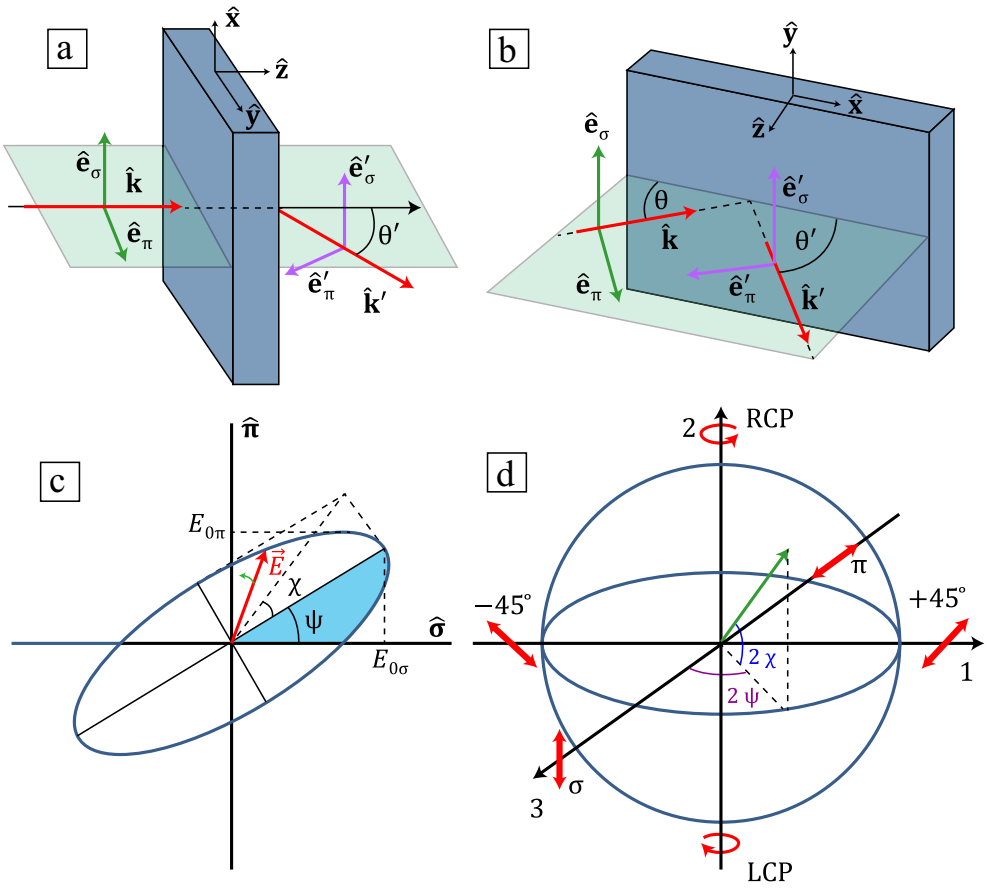

Figure 4.8: (a) A transmission geometry scattering experiment. An x-ray plane wave is incident perpendicular to a sample, interacts with the sample, and exits at some angle θ'. (b) A reflection geometry scattering experiment. An x-ray plane wave is incident at some angle θ to a sample, interacts with the sample, and reflects at some angle θ'. The light blue planes in both (a) and (b) are defined as the plane that is perpendicular to the cross product of the $\mathbf{k}$ and $\mathbf{k}'$ vectors. The $\hat{\mathbf{e}}_\sigma$ and $\hat{\mathbf{e}}'_\sigma$ vectors are perpenicular to this plane while the $\hat{\mathbf{e}}_\pi$ and $\hat{\mathbf{e}}'_\pi$ vectors are parallel. (c) The polarization ellipse. (d) The Poincaré sphere.

ization to each other:

$$\frac{E_\sigma(z,t)^2}{E_{0\sigma}^2}+\frac{E_\pi(z,t)^2}{E_{0\pi}^2}+\frac{E_\sigma(z,t)E_\pi(z,t)}{E_{0\sigma}E_{0\pi}}\cos\delta=\sin^2\delta \tag{4.17}$$

where $\delta=\delta_\pi-\delta_\sigma$ is the phase difference between the $E_\pi(z,t)$ and $E_\sigma(z,t)$ components. Furthermore we have:

$$\tan 2\psi=\frac{2E_{0\sigma}E_{0\pi}}{E_{0\sigma}^2-E_{0\pi}^2}\cos\delta, \tag{4.18}$$

where ψ is defined as the rotation angle with $0\leq\psi\leq\pi$ and

$$\sin 2\chi=\frac{2E_{0\sigma}E_{0\pi}}{E_{0\sigma}^2+E_{0\pi}^2}\sin\delta, \tag{4.19}$$

where χ is defined as the flatness angle with $-\pi/4\leq\psi\leq\pi/4$. This polarization ellipse is shown in Fig 4.8c. From these, we next introduce the Poincaré sphere in Fig 4.8d. The Poincaré sphere defines four parameters to describe all possible polarization states. These parameters can be determined by using the intensity difference between degenerate polarization states:

$$\begin{aligned}
P_0&\equiv I_0/I_0=1\\
P_1&\equiv(I_{+45LP}-I_{-45LP})/I_0\\
P_2&\equiv(I_{RCP}-I_{LCP})/I_0\\
P_3&\equiv(I_{LHP}-I_{LVP})/I_0,
\end{aligned} \tag{4.20}$$

where I_0 is the initial incident intensity, I_{+45LP} is the intensity which has $+45^\circ$ linear polarization, I_{-45LP} is the intensity which has -45° linear polarization, I_{RHC} is the intensity which has righthanded circular polarization, I_{LHC} is the intensity which has lefthanded circular polarization, I_{LHP} is the intensity which has linear horizontal polarization, and I_{LVP} is the intensity which has linear vertical polarization. From Fig 4.8d, we can write that:

$$\begin{aligned}
P_0&=1\\
P_1&=p\ \sin 2\psi\ \cos 2\chi\\
P_2&=p\ \sin 2\chi\\
P_3&=p\ \cos 2\psi\ \cos 2\chi,
\end{aligned} \tag{4.21}$$

where p is the degree of polarization ($0 \leq p \leq 1$). The parameters p, 2ψ and 2χ are the spherical coordinates of the Poincaré sphere. Now, we can define all polarization states:

$$\begin{aligned}
&\text{LHP (or } \sigma \text{ light)} = (P_1, P_2, P_3) = (0, 0, 1) \\
&\text{LVP (or } \pi \text{ light)} = (0, 0, -1) \\
&\text{RCP} = (0, 1, 0) \\
&\text{LCP} = (0, -1, 0) \\
&+45\text{LP} = (1, 0, 0) \\
&-45\text{LP} = (-1, 0, 0) \\
&\text{Unpolarized} = (0, 0, 0),
\end{aligned} \tag{4.22}$$

where P_1, P_2, and P_3 are known as the Stokes parameters.

4.3.3 Polarization density matrix

Next, the mathematical tools to work with Eq. 4.16 and the polarization states described by the Stokes parameters are developed. For a *pure state* (such as degenerate polarization states), there exists a state function $|\psi\rangle$ such that it completely determines the statistical behavior of the measurement. In other words, if $\hat{\mathbf{M}}$ is the associated observable operator, its expectation value is $\langle\psi|\hat{\mathbf{M}}|\psi\rangle$. For mixed states (like partially polarized states), there is no state function ξ to satisfy its expectation value $\langle\xi|\hat{\mathbf{M}}|\xi\rangle$. However there is a unique Hermitian operator (or density matrix) μ such that its expectation value $\langle\mu|\hat{\mathbf{M}}|\mu\rangle = \mathrm{tr}(\mu\hat{\mathbf{M}})$ and its general formulation is:

$$\mu = \sum_j p_j |\psi_j\rangle\langle\psi_j|, \tag{4.23}$$

where each of the pure states $|\psi_j\rangle$ occur with probability p_j. Next, consider the Pauli spin matrices:

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \qquad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{4.24}$$

As each Pauli matrix represents an observable operator for spin 1/2 particles in three spatial dimensions, we can describe statistically all polarization states in

terms of its measured quantity or its expectation values. Let the expectation values be P_1, P_2, and P_3. We can express a 2×2 polarization density matrix μ in terms of these parameters. Remember that μ is Hermitian and that $\mathrm{tr}(\mu\mathbf{I}) = 1$ by definition (where $\mathbf{I}$ is the identity matrix) since $\sum_j p_j = 1$. Thus μ should be of the form:

$$\mu = \begin{pmatrix} a & x+iy \\ x-iy & 1-a \end{pmatrix}, \tag{4.25}$$

where a, x, y are all real, and should satisfy $P_i = \mathrm{tr}(\mu\sigma_i)$. From this we can conclude that:

$$\mu = \frac{1}{2}\begin{pmatrix} 1+P_3 & P_1 - iP_2 \\ P_1 + iP_2 & 1-P_3 \end{pmatrix} = \frac{1}{2}\left(\mathbf{I} + P_1\sigma_1 + P_2\sigma_2 + P_3\sigma_3\right). \tag{4.26}$$

Next, define the state vectors of polarized a photon to be the Jones vectors. The state vectors for LHP (or σ light) and LVP (or π light) are:

$$|\psi_\sigma\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad\qquad |\psi_\pi\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \tag{4.27}$$

the state vectors for $+45$LP and -45LP are:

$$|\psi_{+45LP}\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix} \qquad\qquad |\psi_{-45LP}\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix}, \tag{4.28}$$

and the state vectors for RCP and LCP are:

$$|\psi_{RCP}\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ i \end{pmatrix} \qquad\qquad |\psi_{LCP}\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -i \end{pmatrix}. \tag{4.29}$$

As the probability that a photon has any given polarization state $|\psi\rangle$ is $\langle\psi|\mu|\psi\rangle$, we find that the probability to find LHP state is:

$$\langle\psi_\sigma|\mu|\psi_\sigma\rangle = \frac{1}{2}\begin{pmatrix} 1 & 0 \end{pmatrix}\begin{pmatrix} 1+P_3 & P_1 - iP_2 \\ P_1 + iP_2 & 1-P_3 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{1}{2}(1+P_3), \tag{4.30}$$

while for the LVP state:

$$\langle\psi_\pi|\mu|\psi_\pi\rangle = \frac{1}{2}\begin{pmatrix} 0 & 1 \end{pmatrix}\begin{pmatrix} 1+P_3 & P_1 - iP_2 \\ P_1 + iP_2 & 1-P_3 \end{pmatrix}\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \frac{1}{2}(1-P_3), \tag{4.31}$$

from which we can see that P_3 describes the linear polarization along $\hat{\mathbf{e}}_\sigma$ and $\hat{\mathbf{e}}_\pi$, i.e. $P_3 = 1$ means LHP state and $P_3 = -1$ means LVP state. Similarly, P_2 describes the mean value of the circular state ($P_2 = 1$ means RCP state and $P_2 = -1$ means LCP state), and P_1 describes the mean value of the $\pm$45LP state ($P_1 = 1$ means +45LP state and $P_2 = -1$ means $-$45LP state). So we identify the P_i as the Stokes parameters from the previous section.

4.3.4 Polarization density matrix properties

According to quantum mechanics, there is a scattering operator for an observable (scattering event). Let $\hat{\mathbf{M}}$ be this scattering matrix (or operator). Then the scattering cross section can be calculated by

$$\frac{d\sigma}{d\Omega} = \mathrm{tr}(\mu\hat{\mathbf{M}}\hat{\mathbf{M}}^\dagger) = \mathrm{tr}(\hat{\mathbf{M}}\mu\hat{\mathbf{M}}^\dagger), \tag{4.32}$$

where $\dagger$ is the Hermitian adjoint. In words: for a polarized beam with density matrix μ, the expectation value of the square of the scattering amplitude is given by Eq. 4.32. If we define a new density matrix $\mu' = \mathrm{tr}(\hat{\mathbf{M}}\mu\hat{\mathbf{M}}^\dagger)$ for the scattered beam, we can calculate the polarization of the scattered beam. However, μ' should be normalized first, or $\mathrm{tr}(\mu') = 1$. Therefore, the new density matrix for the scattered beam will be:

$$\mu' = \frac{\hat{\mathbf{M}}\mu\hat{\mathbf{M}}^\dagger}{d\sigma/d\Omega}, \tag{4.33}$$

and the polarization vector for the scattered beam will be

$$P_i' = \mathrm{tr}(\mu'\sigma_i), \tag{4.34}$$

where σ_i are the Pauli matrices.

4.3.5 X-ray magnetic scattering

In this thesis only the "transmission" geometry in Fig. 4.8a will be dealt with, and so we start by first assuming the polarization unit vectors $\hat{\mathbf{e}}$ and $\hat{\mathbf{e}}'$ in Eq. 4.16 are written in terms of the basis vectors $\hat{\mathbf{e}}_\sigma$, $\hat{\mathbf{e}}_\pi$, $\hat{\mathbf{e}}'_\sigma$, and $\hat{\mathbf{e}}'_\pi$ as seen in

Fig. 4.8a. Writing $\hat{\mathbf{e}}$ and $\hat{\mathbf{e}}'$ in this way means that there are only four possible scattering "channels" given by

$$\begin{pmatrix} \sigma \to \sigma' & \pi \to \sigma' \\ \sigma \to \pi' & \pi \to \pi' \end{pmatrix}, \tag{4.35}$$

and so we can write the $\hat{\mathbf{e}}' \cdot \hat{\mathbf{e}}$, $(\hat{\mathbf{e}}' \times \hat{\mathbf{e}}) \cdot \hat{\mathbf{m}}$ and $(\hat{\mathbf{e}'} \cdot \hat{\mathbf{m}})(\hat{\mathbf{e}} \cdot \hat{\mathbf{m}})$ in Eq. 4.16 as [55]:

$$\hat{\mathbf{e}}' \cdot \hat{\mathbf{e}} = \begin{pmatrix} \hat{\mathbf{e}}'_\sigma \cdot \hat{\mathbf{e}}_\sigma & \hat{\mathbf{e}}'_\sigma \cdot \hat{\mathbf{e}}_\pi \\ \hat{\mathbf{e}}'_\pi \cdot \hat{\mathbf{e}}_\sigma & \hat{\mathbf{e}}'_\pi \cdot \hat{\mathbf{e}}_\pi \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & \hat{\mathbf{k}}' \cdot \hat{\mathbf{k}} \end{pmatrix}, \tag{4.36}$$

$$(\hat{\mathbf{e}}' \times \hat{\mathbf{e}}) \cdot \hat{\mathbf{m}} = \begin{pmatrix} \hat{\mathbf{e}}'_\sigma \times \hat{\mathbf{e}}_\sigma & \hat{\mathbf{e}}'_\sigma \times \hat{\mathbf{e}}_\pi \\ \hat{\mathbf{e}}'_\pi \times \hat{\mathbf{e}}_\sigma & \hat{\mathbf{e}}'_\pi \times \hat{\mathbf{e}}_\pi \end{pmatrix} \cdot \hat{\mathbf{m}} = \begin{pmatrix} 0 & \hat{\mathbf{k}} \\ -\hat{\mathbf{k}}' & \hat{\mathbf{k}}' \times \hat{\mathbf{k}} \end{pmatrix} \cdot \hat{\mathbf{m}}, \tag{4.37}$$

and

$$(\hat{\mathbf{e}'} \cdot \hat{\mathbf{m}})(\hat{\mathbf{e}} \cdot \hat{\mathbf{m}}) = \begin{pmatrix} (\hat{\mathbf{e}}_\sigma \cdot \hat{\mathbf{m}})^2 & (\hat{\mathbf{e}}'_\sigma \cdot \hat{\mathbf{m}})(\hat{\mathbf{e}}_\pi \cdot \hat{\mathbf{m}}) \\ (\hat{\mathbf{e}}'_\pi \cdot \hat{\mathbf{m}})(\hat{\mathbf{e}}_\sigma \cdot \hat{\mathbf{m}}) & (\hat{\mathbf{e}}'_\pi \cdot \hat{\mathbf{m}})(\hat{\mathbf{e}}_\pi \cdot \hat{\mathbf{m}}) \end{pmatrix}. \tag{4.38}$$

From the geometry in Fig. 4.8a, assuming that $\hat{\mathbf{k}} = \hat{\mathbf{z}}$ and ϕ is the azimuthal angle in the x-y plane, we can identify that

$$\begin{aligned} \hat{\mathbf{e}}_\pi &= \cos\phi\ \hat{\mathbf{x}} + \sin\phi\ \hat{\mathbf{y}} \\ \hat{\mathbf{e}}_\sigma &= \sin\phi\ \hat{\mathbf{x}} - \cos\phi\ \hat{\mathbf{y}} \\ \hat{\mathbf{k}}' &= \sin\theta'\cos\phi\ \hat{\mathbf{x}} + \sin\theta'\sin\phi\ \hat{\mathbf{y}} + \cos\theta'\ \hat{\mathbf{z}} \\ \hat{\mathbf{e}}'_\pi &= \cos\theta'\cos\phi\ \hat{\mathbf{x}} + \cos\theta'\sin\phi\ \hat{\mathbf{y}} - \sin\theta'\ \hat{\mathbf{z}} \\ \hat{\mathbf{e}}'_\sigma &= \hat{\mathbf{e}}_\sigma. \end{aligned} \tag{4.39}$$

Using these, and making a further assumption that $\phi \approx 0$, we write Eq. 4.38, Eq. 4.37, and Eq. 4.36 as:

$$(\hat{\mathbf{e}'} \cdot \hat{\mathbf{m}})(\hat{\mathbf{e}} \cdot \hat{\mathbf{m}}) \approx \begin{pmatrix} m_y^2 & -m_x m_y \\ -m_x m_y\ \cos\theta' + m_z m_y \sin\theta' & m_x^2\cos\theta' - m_z m_x \sin\theta' \end{pmatrix}, \tag{4.40}$$

$$(\hat{\mathbf{e}}' \times \hat{\mathbf{e}}) \cdot \hat{\mathbf{m}} \approx \begin{pmatrix} 0 & m_z \\ -m_z\cos\theta' - m_x\sin\theta' & -m_y\sin\theta' \end{pmatrix}, \tag{4.41}$$

and

$$\hat{\mathbf{e}}' \cdot \hat{\mathbf{e}} \approx \begin{pmatrix} 1 & 0 \\ 0 & \cos\theta' \end{pmatrix}, \tag{4.42}$$

where $\hat{\mathbf{m}} = m_x\hat{\mathbf{x}} + m_y\hat{\mathbf{y}} + m_z\hat{\mathbf{z}}$. Lastly, the small angle scattering approximation as in Section 1.2.5 is assumed (i.e. $\theta' \approx 0$), so finally we can write:

$$(\hat{\mathbf{e}}' \cdot \hat{\mathbf{m}})(\hat{\mathbf{e}} \cdot \hat{\mathbf{m}}) \approx \begin{pmatrix} m_y^2 & -m_x m_y \\ -m_x m_y & m_x^2 \end{pmatrix}, \tag{4.43}$$

$$(\hat{\mathbf{e}}' \times \hat{\mathbf{e}}) \cdot \hat{\mathbf{m}} \approx \begin{pmatrix} 0 & m_z \\ -m_z & 0 \end{pmatrix}, \tag{4.44}$$

and

$$\hat{\mathbf{e}}' \cdot \hat{\mathbf{e}} \approx \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{4.45}$$

With these, we now write Eq. 4.16 as:

$$\begin{aligned} f^{res} = F^{(0)}(\omega)c(\mathbf{r}) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - iF^{(1)}(\omega)m_z(\mathbf{r}) \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \\ + F^{(2)}(\omega) \begin{pmatrix} m_y(\mathbf{r})^2 & -m_x(\mathbf{r})m_y(\mathbf{r}) \\ -m_x(\mathbf{r})m_y(\mathbf{r}) & m_x(\mathbf{r})^2 \end{pmatrix}. \end{aligned} \tag{4.46}$$

4.3.6 Differential scattering cross-section

We now have enough mathematical "machinery" to do some calculations with the scattering amplitude given by Eq. 4.46. We start by computing the differential scattering cross sections using Eq. 4.32. We make one further approximation in that the last term, the XMLD term, in Eq. 4.46 is assumed negligible compared to the first two terms. This is justified on the basis that the domain walls, which is "in plane" (i.e. has only x and y components) magnetization, and so undergoes XMLD scattering, are usually a small fraction of the size of the domains proper [59, 60, 61, 62]. Hence we would expect that XMCD and resonant charge scattering would overwhelm XMLD scattering. So, calculating the differential cross section

for linearly polarized light ($P_3 = \pm 1$) and for RCP ($P_2 = +1$) and LCP ($P_2 = -1$), and with the scattering matrix defined as $\hat{\mathbf{M}} = f^{res}$, for circularly polarized light, we have that:

$$\begin{aligned}\left(\frac{d\sigma}{d\Omega}\right)_{P_2=\pm 1} &= |F^{(0)}(\omega)\tilde{c}(\mathbf{q}) \pm F^{(1)}(\omega)\tilde{m}_z(\mathbf{q})|^2 \\ &= |F^{(0)}(\omega)\tilde{c}(\mathbf{q}))|^2 + |F^{(1)}(\omega)\tilde{m}_z(\mathbf{q}))|^2 \\ &\quad \pm 2\mathrm{Re}\left\{F^{(0)}(\omega)F^{(1)*}(\omega)\tilde{c}(\mathbf{q})\tilde{m}_z^*(\mathbf{q})\right\}, \end{aligned} \tag{4.47}$$

while for linearly polarized light, the differential scattering cross-section is given by:

$$\left(\frac{d\sigma}{d\Omega}\right)_{P_3=\pm 1} = |F^{(0)}(\omega)\tilde{c}(\mathbf{q})|^2 + |F^{(1)}(\omega)\tilde{m}_z(\mathbf{q})|^2, \tag{4.48}$$

where $\tilde{c}(\mathbf{q}) = \mathcal{F}[c(\mathbf{r})]$, $\tilde{m}_z(\mathbf{q}) = \mathcal{F}[m_z(\mathbf{r})]$, and $\mathcal{F}$ means propagation from the sample plane to the detector plane. In the case of linearly polarized light, by ignoring the XMLD term the dependence of the sign of P_3 is removed; i.e. only scattering from in plane magnetization $m_x(\mathbf{r})$ and $m_y(\mathbf{r})$ has any dependence on P_3. These expressions can be related to the diffraction intensity by using the relation:

$$\frac{I}{I_0} = |\mathcal{F}[T(\mathbf{r})]|^2 = \left|\mathcal{F}\left[\exp\left(ik\int_0^D (n-1)dz\right)\right]\right|^2 \propto \frac{d\sigma}{d\Omega}, \tag{4.49}$$

where I_0 is the incident intensity, $T(\mathbf{r})$ is the sample transmission function, n is the polarization dependent index of refraction, D is the sample thickness as seen along the beam propagation direction (i.e. $\hat{\mathbf{k}}$), and the projection approximation given by Eq. 1.18 is assumed. Examples of what diffraction from a sample with XMCD contrast would look like is shown in Fig. 4.9. If we wish to make an equality relation between the diffraction intensity and the differential cross section, we can use:

$$\frac{I}{I_0} = \varrho\Delta\Omega\frac{d\sigma}{d\Omega}, \tag{4.50}$$

where ϱ is the areal atomic number density as seen along the beam propagation direction, and $\Delta\Omega$ is some solid angle which the detector subtends. We should note also that when comparing Eq. 4.47 and Eq. 4.48 that they differ by the term given by:

$$\pm 2\mathrm{Re}\left\{F^{(0)}(\omega)F^{(1)*}(\omega)\tilde{c}(\mathbf{q})\tilde{m}_z^*(\mathbf{q})\right\}, \tag{4.51}$$

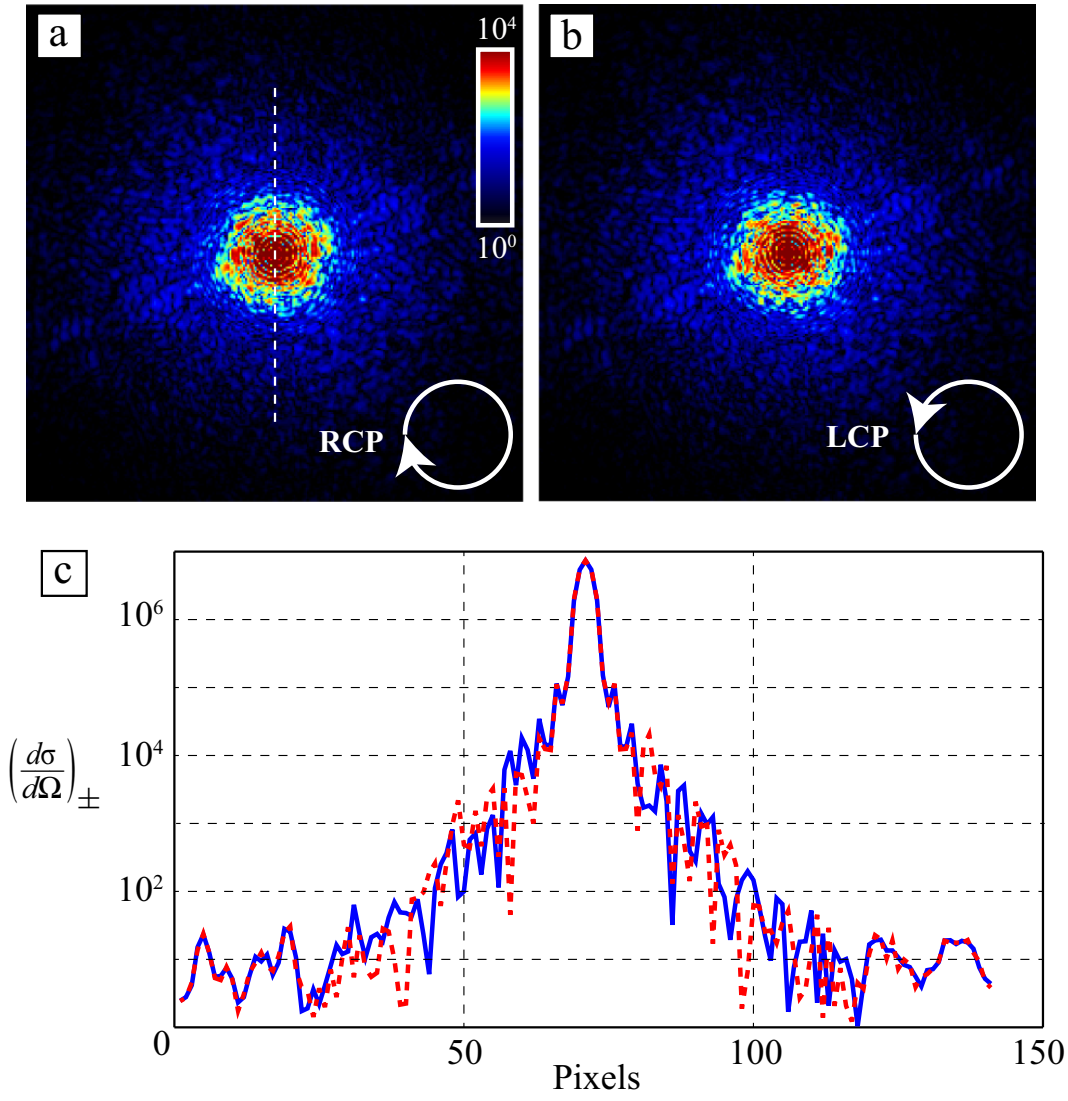

Figure 4.9: Diffraction intensities for RCP and LCP incident light. (a) Example diffraction intensity for RCP incident light using Eq. 4.47 with $P_2 = +1$. (b) Example diffraction intensity for LCP incident light using Eq. 4.47 with $P_2 = -1$. (c) A lineout of the diffraction patterns along the dotted white line in (a). The blue curve corresponds to RCP and the dashed red curve LCP. This illustrates that while at a distance they look very similar, the diffraction patterns actually are quite different from one another, and in fact are different by the interference term given in Eq. 4.51.

known as the "interference term" due to the product of $\tilde{c}(\mathbf{q})$ and $\tilde{m}_z^*(\mathbf{q})$ [63, 64], and interference here means that the diffracted wavefield due to charge scattering coherently interferes with the diffracted wavefield due to XMCD scattering. This interference between charge and magnetic scattering is only present when the incident radiation is circularly polarized; it does not occur when it is linearly polarized. The reason for this is that a linearly polarized electric field can be thought of as a sum of RCP and LCP electric fields (e.g. see the Jones vectors given in Eq. 4.27 and Eq. 4.29), and since the interference terms for RCP and LCP are equal but opposite in sign, they cancel leaving only the incoherent sum of charge and magnetic scattering given by Eq. 4.48. Indeed, we have that:

$$I_{lin} = \frac{1}{2}(I_{P_2=+1} + I_{P_2=-1}) \propto \left(\frac{d\sigma}{d\Omega}\right)_{P_2=+1} + \left(\frac{d\sigma}{d\Omega}\right)_{P_2=-1}. \tag{4.52}$$

4.3.7 Polarization dependent index of refraction

Having an expression for the polarization dependent index of refraction could also be useful. In the case when the incident radiation is circularly polarized ($P_2 = \pm 1$) and the experiment is in the forward scattering transmission geometry (i.e with $\theta' \approx 0$), it can be shown [63] that the scattered radiation from a sample with primarily out-of-plane magnetization (i.e. dropping the XMLD term in Eq. 4.16) will still be purely circularly polarized (i.e. $P_i' = \mathrm{tr}(\mu'\sigma_i)$ is still $(P_1', P_2', P_3') = (0, \pm 1, 0)$); this can be confirmed by evaluating Eq. 4.33 and Eq. 4.34. In this case, we can write the index of refraction as [65, 66]:

$$n_\pm = 1 - \frac{2\pi\rho r_e}{k^2} f_\pm = 1 - \frac{2\pi\rho r_e}{k^2}\left(f^0 + F^{(0)}(\omega)c(\mathbf{r}) \pm F^{(1)}(\omega)m_z(\mathbf{r})\right), \tag{4.53}$$

where n_+ refers to the refractive index for RCP incident x-rays, n_- refers to the refractive index for LCP incident x-rays, ρ is the atomic number density (N particles in a volume V), r_e is the Thompson scattering length, $k = 2\pi/\lambda$ is the wave number in vacuum, and f^0 is the nonresonant Thompson charge scattering amplitude, which is the electron density with relativistic corrections (i.e. $f^0(\mathbf{r}) = Z(\mathbf{r}) - (Z(\mathbf{r})/82.5)^{2.37}$), where $Z(\mathbf{r})$ is the number of electrons as a function of $\mathbf{r}$. With this, and assuming that the sample has approximately uniform

thickness (i.e. n is independent of z) we find that the sample exit wavefield (using the projection approximation again) is:

$$\begin{aligned}
\psi_{\pm}(\mathbf{r}, z = D) &\approx \psi_0 \exp\left(ik \int_0^D (n_{\pm} - 1) dz\right) \\
&\approx \psi_0 \; e^{ikD(n_{\pm}-1)} \\
&\approx \psi_0 \; (1 + ikD(n_{\pm} - 1)) \\
&= \psi_0 - \psi_0 \frac{2\pi i D \rho r_e}{k} \left(f^0 + F^{(0)}(\omega) c(\mathbf{r}) \pm F^{(1)}(\omega) m_z(\mathbf{r})\right) \\
&= \psi_{NR} - \psi_0 \frac{2\pi i D \rho r_e}{k} \left(F^{(0)}(\omega) c(\mathbf{r}) \pm F^{(1)}(\omega) m_z(\mathbf{r})\right). \qquad (4.54)
\end{aligned}$$

The first term $\psi_{NR} = \psi_0 \left(1 - 2\pi i D \rho r_e f^0/k\right)$ on the last line in Eq. 4.54 is the wavefield due to nonresonant charge scattering, and when on electronic resonances is overwhelmed by the resonant charge and magnetic scattering. Thus, we can drop this term and from here on assume we are on an electronic resonance. With this assumption, the wavefield in the far field is given by:

$$\tilde{\psi}_{\pm}(\mathbf{q}) = \mathcal{F}\left[\psi_0 \frac{2\pi D \rho r_e}{ik} \left(F^{(0)}(\omega) c(\mathbf{r}) \pm F^{(1)}(\omega) m_z(\mathbf{r})\right)\right], \qquad (4.55)$$

and the diffraction intensity would be:

$$\begin{aligned}
I_{\pm} &= \tilde{\psi}_{\pm}(\mathbf{q}) \tilde{\psi}_{\pm}(\mathbf{q})^* \\
&= I_0 \left(\frac{2\pi D \rho r_e}{k}\right)^2 \left(|F^{(0)}(\omega) \tilde{c}(\mathbf{q})|^2 + |F^{(1)}(\omega) \tilde{m}_z(\mathbf{q})|^2 \right. \\
&\quad \left. \pm 2\mathrm{Re}\left\{F^{(0)}(\omega) F^{(1)}(\omega)^* \tilde{c}(\mathbf{q}) \tilde{m}_z(\mathbf{q})^*\right\}\right), \qquad (4.56)
\end{aligned}$$

where we define $I_0 = \psi_0^2$ as the intensity of the incident wavefield propagated to the detector. Compare this with Eq. 4.47; we see that we again have the same thing other than the $I_0(2\pi D \rho r_e/k)^2$ prefactor.

Determining the index of refraction for linearly polarized light is more problematic; the problem we have here is that due to the incoherent sum of charge scattering diffraction intensity and magnetic scattering diffraction intensity given by Eq. 4.48, there is no phase relationship between charge and magnetic scattering. One type of scattering can be considered a source of "alien" signal for the other type of scattering because of this. For example, assuming the RCP case in

Eq. 4.55, and assuming we had a Wonderland detector which can also measure the phase of a wavefield, we would measure

$$\sqrt{I_+(\mathbf{q})}e^{\phi_+} = \mathcal{F}\left[\psi_0 e^{ikD(-1+n_+)}\right], \tag{4.57}$$

while for the linearly polarized case it would be something like

$$\begin{aligned}\sqrt{I_{lin}(\mathbf{q})}\exp[\phi_{lin}] &= \frac{\exp[\phi_{lin}]}{2}\sqrt{\left|\mathcal{F}\left[\psi_0 e^{ikD(-1+n_+)}\right]\right|^2 + \left|\mathcal{F}\left[\psi_0 e^{ikD(-1+n_-)}\right]\right|^2} \\ &\stackrel{?}{=} \mathcal{F}\left\{\psi_0 \exp\left[ikD(-1+n_{lin})\right]\right\}. \end{aligned} \tag{4.58}$$

We would be left with trying to interpret what this "quasi-sample" is exactly, which mixes the magnetic scattering signal and charge scattering signal is a very unclear fashion. One might naïvely try to simply sum the $\tilde{\psi}_+(\mathbf{q})$ and $\tilde{\psi}_-(\mathbf{q})$ wavefields at the detector, but notice that the XMCD term goes away by doing this, meaning that we would have no magnetic scattering, something which contradicts theory (i.e. Eq. 4.48) and experiments [64]. The problem we have here is that the polarization state of the sample exit wave is not the same as the incident wavefield polarization; the sample exit wave has gained an *elliptical* component as well as still having a linear component, and is known as Faraday rotation [67, 68, 69]. The helicity of this rotation is dependent on the sign of $m_z(\mathbf{r})$ (and so is dependent on $\mathbf{r}$ as well). This is why we are unable to simply sum the *scalar* wavefields for $\psi_+(\mathbf{r})$ and $\psi_-(\mathbf{r})$. By doing this we ignore exit wave electric field polarization state, which is dependent on $\mathbf{r}$; to be able to sum exit wavefields, we would need the $\psi_+(\mathbf{r})$ and $\psi_-(\mathbf{r})$ to include polarization information, i.e. make them vector wavefields (this information drops out when we compute intensities, and so can directly add the intensities for left and right circularly polarized light). Neverthless, if we know certain things about the sample, for example if the charge structure of the sample is independent of both $\mathbf{r}$ and z (as is the case usually for homogeneous thin films), then it is possible to subtract off the resonant charge scattering term in Eq. 4.48, leaving only the magnetic scattering part of the diffraction intensity:

$$I_{M,P_3=\pm 1} = I_0\left(\frac{2\pi D\rho r_e}{k}\right)^2\left(\cancelto{0}{|F^{(0)}(\omega)\tilde{c}(\mathbf{q})|^2} + |F^{(1)}(\omega)\tilde{m}_z(\mathbf{q})|^2\right). \tag{4.59}$$

Now we only have diffraction intensity from magnetic contrast, and so making sense of the wavefield at the detector is now easy. Using the Wonderland detector mentioned previously (it can measure the phase of a wavefield in addition to modulus), what we would "measure" in this case would be:

$$\tilde{\psi}_{M,P_3=\pm 1}(\mathbf{q}) = \psi_0 \left(\frac{2\pi D \rho r_e}{ik} \right) F^{(1)}(\omega) \tilde{m}_z(\mathbf{q}). \tag{4.60}$$

What a diffraction intensity pattern given by Eq. 4.59 might look like in practice is shown in Fig. 4.10; in Fig. 4.10a, a diffraction pattern I_{lin} is calculated assuming a linearly polarized incident wavefield using the diffraction patterns seen in Fig. 4.9b-c, and Eq. 4.52. In Fig. 4.10b, the diffraction pattern assuming a spatially homogeneous charge configuration shown. It is simply a scaled (attenuated) version of what we would see by taking the sample out of the beam, or alternatively introducing an external magnetic field and fully saturating the sample magnetically, so that $m_z(\mathbf{r})$ goes away. We can use this, plus knowledge of the sample thickness to determine the resonant attenuation scaling factor, to subtract off the resonant charge scattering diffraction to get what is shown in Fig. 4.10c. This diffraction pattern now contains only magnetic scattering intensity. Lastly, in Fig. 4.10d, we see what the interference term looks like; this expression is given in Eq. 4.51. It is this interference term which is subtracted out when we sum the diffraction intensities for RCP and LCP to obtain the diffraction pattern seen in Fig. 4.10a.

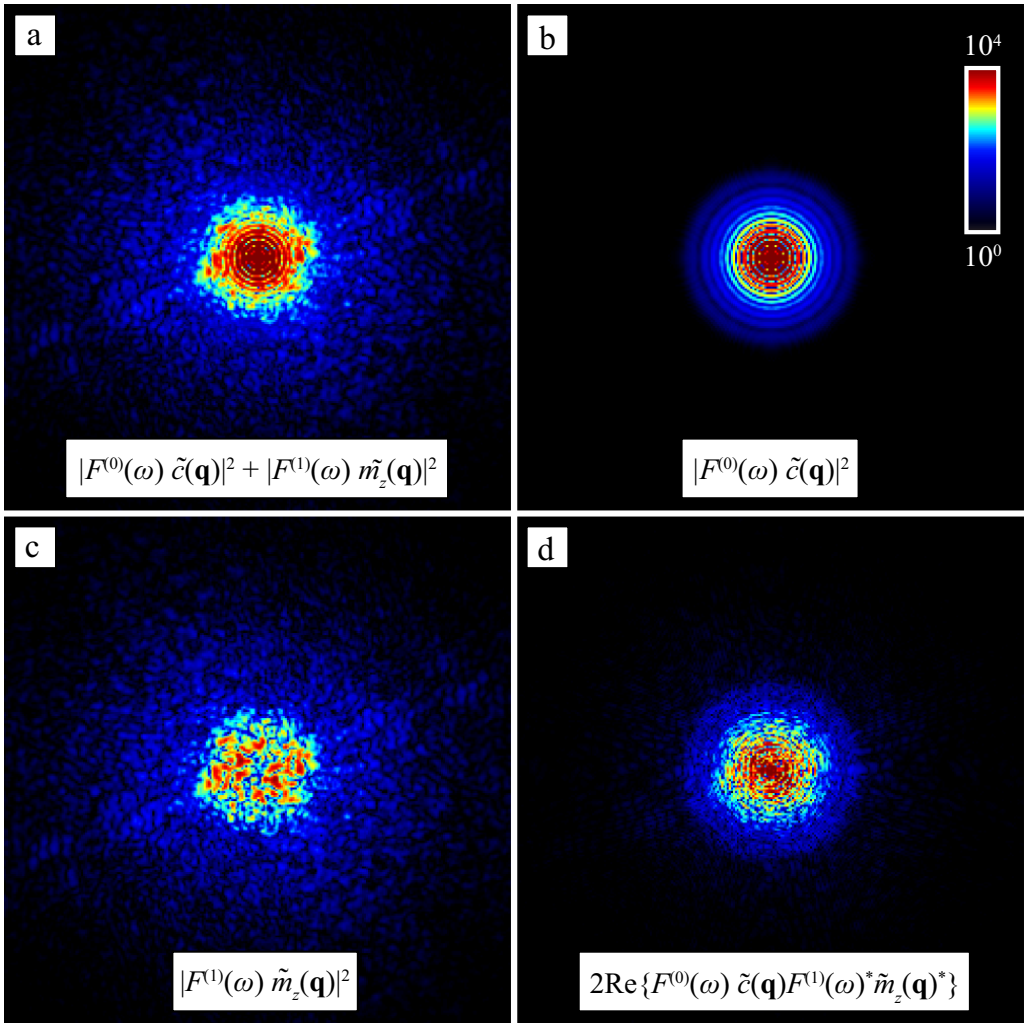

Figure 4.10: Subtraction of known charge diffraction intensity to isolate magnetic diffraction intensity. All frames have the same scale bar seen in frame (b). In (a), a simulated diffraction pattern assuming a linearly polarized incident wavefield; this pattern is proportional to the scattering cross-section given in Eq. 4.48. In (b), the diffraction pattern assuming a spatially homogeneous charge configuration (i.e. $c(\mathbf{r})$ is assumed to actually have no $\mathbf{r}$ dependence). It is simply a scaled version of what we would see by taking the sample out of the beam, or alternatively introducing an external magnetic field and fully saturating the sample magnetically (so that $m_z(\mathbf{r})$ goes away). In (c), we have the result of using the diffraction pattern in (b) to subtract from the diffraction pattern in (a), leaving only magnetic scattering, as in Eq. 4.59. In (d), we have the *difference* between the diffraction patterns for RCP and LCP incident wavefields; this is proportional to the interference between charge and magnetic scattering from Eq. 4.51, and it is this interference term which has been subtracted out when we sum the diffraction intensities for RCP and LCP to obtain the diffraction pattern seen in frame (a).

Chapter 5

Dichroic Coherent Diffrative Imaging

This chapter puts into practice everything we have gone over in previous four chapters. We now know theoretically how to numerically invert a diffraction pattern and how to interpret what exactly we are attempting to reconstruct when illuminating a magnetic multilayer with a linearly polarized x-ray beam. This chapter experimentally confirms this theory. Here we present a novel approach to lensless imaging of an extended magnetic nanostructure, in which a scanned series of dichroic coherent diffraction patterns is experimentally recorded and numerically inverted to map its magnetic domain configuration. Unlike holographic methods, it does not require a reference wave or precision optics. In addition, it enables the imaging of samples with arbitrarily large spatial dimensions, at a spatial resolution limited solely by the coherent x-ray flux, wavelength, and stability of the sample with respect to the beam. It can readily be extended to non-magnetic systems that exhibit circular or linear dichroism. We demonstrate this approach by imaging ferromagnetic "maze" in a GdFe multilayer with perpendicular anisotropy and follow the evolution of the domain structure through part of its magnetization hysteresis loop. This approach is scalable to imaging with diffraction-limited resolution, a prospect rapidly becoming a reality in view of the new generation of phenomenally brilliant x-ray sources.

5.1 Introduction

Materials such as magnetic multilayers and alloys, polymers, liquid crystals, biofibers and biominerals all exhibit self organizing, reaction-diffusion, and pattern forming behavior not fully understood. New schemes for directed domain formation in magnetic multilayers and alloys are integral parts of next-generation magnetic data storage and spintronic technologies [43, 44]. Controlled phase transitions and ordering dynamics in polymers and liquid crystals under applied electric fields, with consequent photonic bandgap shifts, play a major role in organic laser technology [70]. In the biological sciences, certain biofibers display tensile properties similar to that of steel yet are far more lightweight, properties thought to be the result of self-organized phase separation of molecular crystalline and amorphous regions within the fibers [71, 72]. Deeper understanding of biomineral growth and the interaction between inorganic material and organic macromolecule phases could enable use of similar techniques to fabricate novel synthetic materials.

Microscopy using dichroism as a contrast mechanism can reveal much about phase ordering, separation, and coexistence in these kinds of systems. All of these materials have an order parameter that scatters light differently depending on the direction or helicity of the photon polarization. Polarized x-rays are an ideal choice of probe for imaging buried magnetic structure due to strong resonant enhancement of the scattering at electronic transitions split by the spin-orbit effect. Since x-ray beams are unaffected by magnetic or electric fields, they are also well suited to studying phase transitions as a function of applied fields. Established techniques such as transmission x-ray microscopy and x-ray photoemission electron microscopy as well as novel holographic methods have dramatically improved our understanding of dichroic materials at the nanoscale [5, 73, 74, 75]. However, these approaches require focusing optics or apertures that must be made to a precision comparable to the desired spatial resolution. On the other hand, techniques such as magnetic force and electron microscopy offer nanoscale imaging, but only near surfaces and generally with weak external fields [76].

An alternative approach to nanoscale dichroic microscopy is resonant x-ray coherent diffractive imaging (CDI), whereby the diffraction pattern formed by

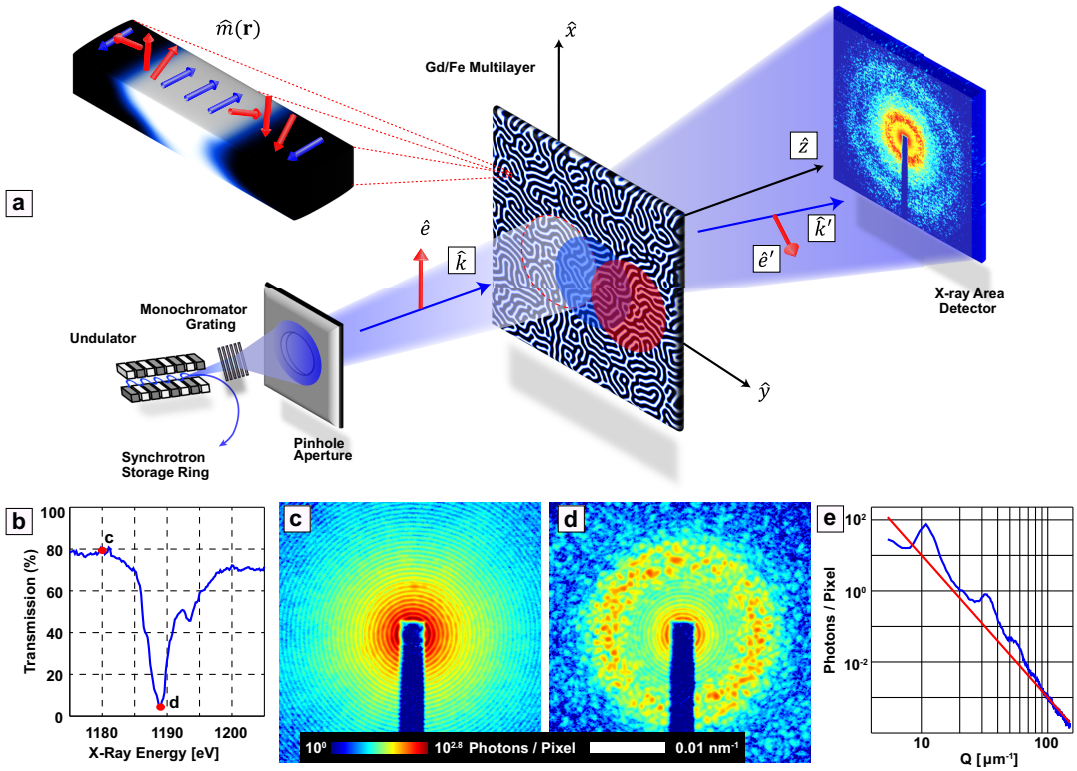

Figure 5.1: (a) Schematic of an x-ray scanning CDI measurement with a dichroic sample. For the sample under investigation, the magnetization direction $\hat{m}(\mathbf{r})$ is primarily parallel/antiparallel to the propagation vector $\hat{k}$. When linearly polarized synchrotron radiation with a polarization direction defined by $\hat{e}$ is used to illuminate the sample, contrast arises primarily from the x-ray magnetic circular dichroism (XMCD) effect. Diffraction patterns are recorded in the far-field by an x-ray area detector by scanning the sample, depicted as the overlapping white, blue and red circles. (b) X-ray energy transmission scan across the Gd M_5 edge, with red circles marking the x-ray energy values for which diffraction patterns (c) and (d) were collected. (c) Off-resonance diffraction pattern taken at an x-ray photon energy of 1180 eV has no discernible magnetic scattering and contains only Airy diffraction fringes from the circular pinhole. (d) On-resonance diffraction pattern taken at 1189 eV exhibits a magnetic scattering ring arising from magnetic stripe domains, in addition to the pinhole scattering. (e) Azimuthal average of diffraction intensity showing higher order magnetic diffraction rings. The red line illustrates a Q^{-4} power-law decay of the signal.

scattering a coherent x-ray beam from a sample is inverted numerically to form an image of the object [12]. In CDI, the spatial resolution is not limited by the quality of the optical elements, but by the highest spatial frequencies measured in the x-ray diffraction pattern. CDI can provide three-dimensional information by sufficiently mapping the Ewald sphere [38, 77, 78], and offers elemental selectivity near electronic resonances [79]. When combined with "ptychographic" methods, regions of arbitrary large lateral dimensions can be imaged quantitatively with a spatial resolution defined solely by the wavelength, coherent x-ray flux, and scan precision and stability [13, 30, 31, 34, 80]. In CDI the phases of the scattered amplitudes lost by recording the diffraction intensity are recovered computationally by an iterative algorithm. For the algorithm to converge to a unique solution, there must be sufficient constraints on the problem and the diffraction pattern must be adequately sampled with respect to the spatial Nyquist frequency [6]. This requirement, known as "oversampling" [24] is necessary to regain phase information in the complex-valued diffraction signal. The algorithm starts with a random guess for the phase of the measured diffraction, then iterates between the sample (real space) and the diffraction measurement (reciprocal space) using fast Fourier transforms, refining the initial guess by enforcing constraints in both real and reciprocal space at each step [7]. A solution is found when the initial guess is refined to a degree such that it satisfies all the constraints.

Here we demonstrate that lensless resonant scanning CDI can be used to image nanometer-scale ferrimagnetic domains in a multilayer GdFe film. In such multilayer systems, alternating layers of a transition metal and a rare-earth or noble metal are deposited to create an artificial ferrimagnet with perpendicular magnetic anisotropy. The balance between exchange, anisotropy, and dipolar energies results in stripe domain patterns as first described by Kittel [81] and further developed and studied experimentally by Kooy and Enz [82]. A schematic of this system is shown in Fig. 5.1a. Perpendicular anisotropy magnetic films are extensively studied to probe the physics of stripe-forming systems [83, 84], where local pinning sites, external magnetic fields, and temperature can strongly affect the domain structure. Interest in GdFe and related multilayer structures has grown

recently with demonstrations of laser-induced femtosecond switching, leading to the possibility of dramatically faster magnetic recording [85].

5.2 Dichroic diffraction pattern measurements

The present experiments take advantage of the brilliance, polarization, and tunability of x-ray undulator radiation. The scattering resulting from polarized photons interacting with the magnetic ordering is strongly enhanced at resonances [52, 53]; Fig. 5.1b shows an transmission spectrum for the sample in the vicinity of the Gd M_5 edge. This spectrum is highly sensitive to the photon polarization state and magnetic spin state of the Gd. For the GdFe multilayer studied here in the small angle scattering geometry, as shown in Fig. 5.1a, the scattered signal is maximal when the incident photon propagation $\hat{k}$ is perpendicular to the sample surface. Off resonance, at 1180 eV, the diffraction pattern is dominated by charge scattering as well as the unscattered direct beam from the pinhole (see Fig. 5.1c). The rectangular shadow in Figs. 5.1c and 5.1d is due to the beam stop used to protect the CCD from the bright, unscattered beam. Since this sample appears electronically homogeneous in projection, a diffraction pattern taken with the sample out of the x-ray beam, arising solely from the pinhole aperture, is identical to Fig. 5.1c up to a known absorption scale factor. The same diffraction pattern was obtained by saturating the magnetization by applying an out-of-plane magnetic field and coercing the sample into a single domain state.

A coherent diffraction pattern taken on resonance at 1189 eV is shown in Fig. 5.1d. The speckle ring at a wavevector of $\simeq 0.01$ nm^{-1} corresponds to spatial features of $\pi/0.01$ nm^{-1} $\simeq 300$ nm. There are several additional higher-order (odd) diffraction rings at wavevectors of $\simeq 0.03$ nm^{-1} and $\simeq 0.05$ nm^{-1}, which can be seen in the azimuthally averaged diffraction pattern shown in Fig. 5.1e. This suggests that the magnetic domains are monodisperse with a characteristic domain width of $\simeq 300$ nm. Beyond $\simeq 0.06$ nm^{-1}, the diffraction signal is difficult to distinguish from the noise background of the measurement. This maximum value of the momentum transfer over which the diffraction pattern can be sampled roughly

defines the spatial resolution in our experiments, and corresponds to spatial information down to $\pi/0.06$ nm$^{-1} \simeq 50$ nm. When using linearly polarized x-rays, the charge scattering does not interfere with the magnetic scattering [64]. As a result we can isolate the resonant magnetic scattering by straightforward subtraction of the magnetically saturated diffraction intensities from those of the non-saturated measurement, both taken on-resonance. This difference results in a diffraction pattern that arises solely from magnetic contrast. The resulting magnetic diffraction is used to reconstruct maps of the magnetic domain configuration; further details are provided in the *Materials and Methods* section.

Starting from a magnetically saturated state (i.e. the entire sample is in a single-domain ferrimagnetic state with the net magnetization parallel to the applied field), the external field was reduced until magnetic diffraction becomes visible; this occurs at about 44 mT as shown on the hysteresis curve in Fig. 5.2a. A reconstruction of the ptychographic diffraction data at this point is shown in Fig. 5.2c, while Fig. 5.2b shows the reconstructed x-ray wave incident on the sample. We then recorded a series of ptychographic measurements while reducing the field near zero, with representative reconstructions at selected points on the hysteresis curve (Fig. 5.2a) shown in Figs. 5.2d-g. Figs. 5.2h-l show the reconstructions of the Gd spin density at the same sample region as the field is increased towards saturation (corresponding to points h through l in Fig. 5.2a). As the magnetic field is decreased, the magnetic domain pattern undergoes classic evolution towards a labyrinthine stripe pattern. This type of pattern is commonly found in a wide class of uniaxially modulated condensed matter systems, including diblock copolymers, liquid crystals, Langmuir monolayers, adsorbates at metallic surfaces, incommensurate structures, membranes, vesicles, and ferrofluids, as well as dewetting, phase separation, dealloying and convection patterns [84, 87, 88]. The characteristic length scale corresponding to the width of the stripe is generally defined by the energetic balance between short- and long-range interactions. The ability to tune the relative strength of these interactions and to control the degree of orientational or translational order in many of these systems represents a pathway for driven or guided self-assembly at nanoscale, a concept especially appealing in areas of soft

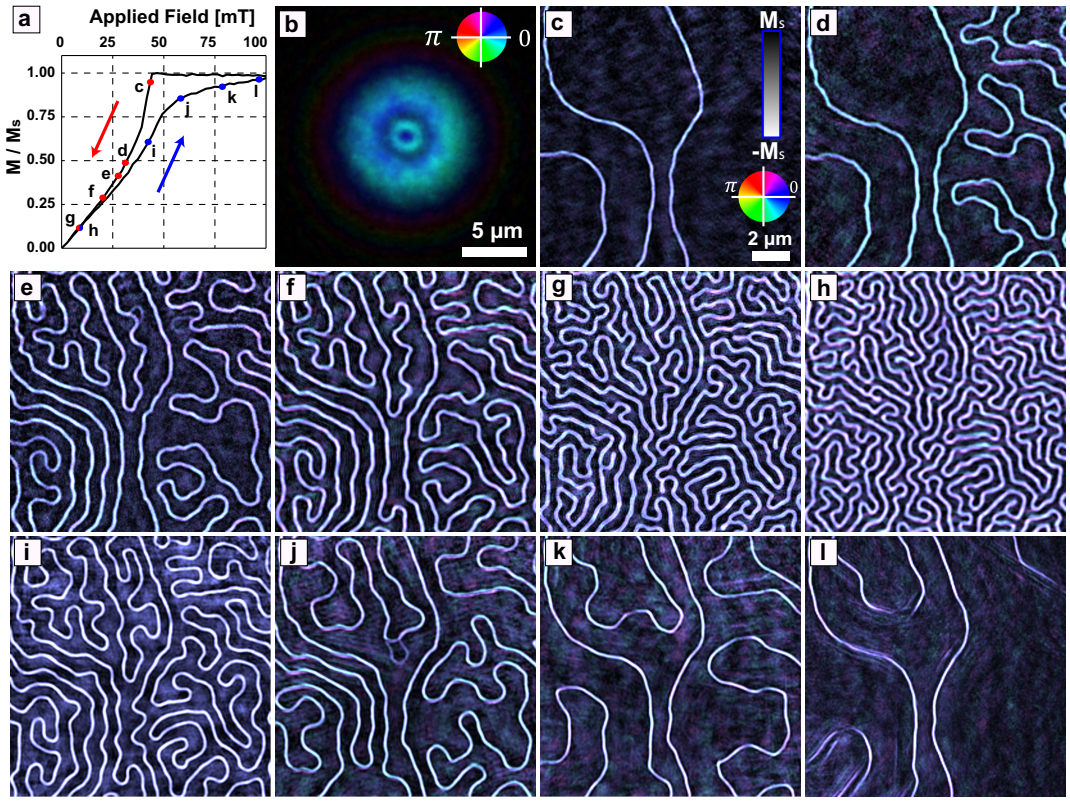

Figure 5.2: (a) Hysteresis loop of the sample magnetization as a function of magnetic field applied perpendicular to the film surface. (b) Reconstruction of the x-ray beam that illuminated the sample. The phase of the reconstruction is encoded as hue of the image, while the magnitude is encoded as brightness. The x-ray flux incident on the sample is estimated to be $2.5 \cdot 10^5$photons/s$\cdot \mu$m^2. (c-g) Gd spin density reconstructions at indicated points on the magnetic hysteresis curve shown in (a) while decreasing the external field. Bright domains are anti-parallel to the applied magnetic field. Phase is encoded as hue and magnitude is encoded as brightness, as for the probe reconstruction. The color bar shows the areal magnetization in terms of the saturation value, $M_s = 0.011$ emu/cm^2. Because only the Gd is resonantly scattering, the reconstructions only show the evolution of the Gd magnetic configuration. (h-l) Gd spin density reconstructions at indicated points on the magnetic hysteresis curve shown in (a) while increasing the external field. Note that near the top in (e) and towards the middle in (j) we observe that part of the bright domain appears to move, suggesting the domain configuration had not quite settled when the measurements began. Each reconstruction presented is the average of ten reconstructions using random initial phase guesses. The small variations visible in the images are artifacts caused by statistical noise due to the low photon count at large Q and by unconstrained modes in the reconstruction [86].

matter and magnetism.

5.3 Magnetization configuration behavior

The magnetization behavior with field of the GdFe multilayer studied here follows closely the classic theory of Kooy and Enz first developed in context of $BaFe_{12}O_9$ [82] and subsequent modeling of periodic stripe domains in an infinite plate. The sharp downturn in the magnetization with decreasing field (point c in Fig. 5.2a) occurs when isolated domains become unstable. These domains seed an avalanche of stripe domains that quickly fill the sample and a characteristic domain periodicity rapidly develops across the sample (as reflected by the ring in the magnetic scattering in Fig. 5.1d). The resulting morphology and ratio of "up" domains to "down" domains depends on the dipolar interactions that drive the domain formation, the field history, the density of nucleation sites, and the interaction of the domain walls with quenched disorder (e.g. pinning centers arising from local variations in the magnetic properties) in the sample. With decreasing field, the average domain period decreases as domains fill the space with the domain width remaining relatively constant. This indicates that domain evolution in GdFe films is qualitatively different from that in Co/Pd multilayer films where the domain period is relatively insensitive to field and only the relative ratio of up and down domain widths changes with field [89, 90]. The current results more closely follow the expectations of modeling [82] and suggest a relatively low defect density. However, the two early forming vertical domains seen in Fig. 5.2c appear more heavily pinned than other domains, retaining their position and shape down to the lowest magnetic field. These early forming domains define a boundary, dividing the imaged field of view into three distinct regions. Each region contains self-connected domains, which remain separated during their evolution [91]. Close comparison of Figs. 5.2e and 5.2f as well as Figs. 5.2g and 5.2h shows that the final and initial stages of labyrinthine growth are dominated by the unbinding of the disclination dipoles [83], resulting in formation of split-off Y-shaped branches and "cul-de-sac" domain lines that are not present in domain patterns shown in

Figs. 5.2c through 5.2d and Figs. 5.2i through 5.2k. The behavior for increasing and decreasing field in this region, where the divergence of the periodicity is the fundamental saturation mechanism, is nearly reversible.

5.4 Spatial resolution

To characterize the quality of the reconstructions in Fig. 5.2, we use the concept of the Phase Retrieval Transfer Function (PRTF) [38]. The PRTF is defined as the ratio of the reconstructed diffraction magnitude (the absolute value of the Fourier transform of the reconstruction) to the measured diffraction magnitude as a function of momentum transfer. The PRTF can be expressed as

$$\mathrm{PRTF}(Q) = \frac{|\mathcal{F}\{\langle \Phi(\mathbf{r}) p(\mathbf{r}-\mathbf{r}_j)\rangle\}|}{\sqrt{I_j(Q)}}, \tag{5.1}$$

where $I_j(Q)$ is the measured diffraction intensity at scan position $\mathbf{r}_j$, and $\langle\,\rangle$ denotes averaging over many independent reconstructions of $\Phi(\mathbf{r})$. For the PRTF shown in Fig. 5.3a, the Q axis pixel size is $\pi a/\lambda z \simeq 0.089\ \mu\mathrm{m}^{-1}$, and annuli 3 pixels wide in both the numerator and denominator of Eq. 5.1 were integrated from $Q = 0\ \mu\mathrm{m}^{-1}$ to $Q = 178\ \mu\mathrm{m}^{-1}$. The ratio of these two results is shown in Fig. 5.3a. The PRTF provides a quantitative estimate of the range of spatial frequencies over which the domain reconstructions can be trusted. Imperfections of the diffraction measurements due to photon shot noise and missing diffraction behind the beamstop, as well as imperfect real-space constraints, such as mechanical stage errors when scanning the sample, contribute to numerical artifacts and degradation of the reconstruction. A typical PRTF of the domain reconstructions is shown in Fig. 5.3a. The dips in the PRTF arise because of low signal to noise at spatial frequencies between the highly ordered magnetic scattering rings seen in Figs. 5.1d and 5.1e. Using a PRTF cutoff value of 0.5 as described previously [38] indicates that reconstruction is robust up to spatial frequencies of about $57\ \mu\mathrm{m}^{-1}$. This corresponds to a spatial resolution of $\pi/57\ \mu\mathrm{m}^{-1} \simeq 55$ nm, in good agreement with the estimate above based on the largest wavevector at which speckles are still visible among the noise. We can arrive at another estimate of the spatial

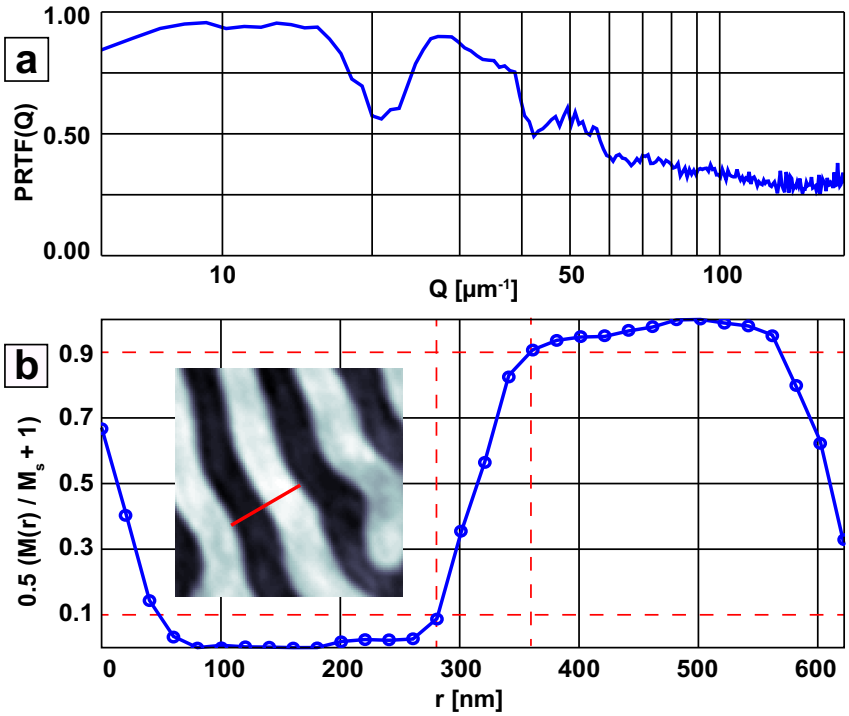

Figure 5.3: (a) The Phase Retrieval Transfer Function (PRTF) of a reconstruction. Defining a cutoff of 0.5 predicts a spatial resolution of $\pi/57\ \mu\text{m}^{-1} \simeq 55$ nm (b) Line cut (**r**) through two domains. The 10% to 90% edge response gives a resolution of $\simeq 80$ nm. The point spread function of the imaging process blurs the Bloch wall width of $\simeq 30$ nm, which accounts for the discrepancy.

resolution by taking a line scan across a reconstructed domain wall. As seen in Fig. 5.3b, the 10% to 90% edge response provides a width of 77 nm (corresponding to a standard deviation of 60 nm). The Bloch wall width for this system is $\simeq 30$ nm [59, 60, 61, 62]. Convolution of this domain wall width and a Gaussian point spread function with standard deviation $0.5 \cdot 55$ nm would result in a blurred Bloch wall with a width of 62 nm and 10% to 90% edge response of 80 nm, which accounts for the results obtained here. During data acquisition the positions of the magnetic speckles comprising the diffraction patterns did not move detectably even at the largest wavevector values, indicating that sample vibration within the illuminated region was negligible in our experiments. However, sample vibration and positioning errors in scanning could become practical limitations of the CDI approach at higher resolution.

5.5 Materials and Methods

This section contains mostly details of the experiment carried out and discussed above. It goes into sample growth parameters for the GdFe multilayer studied here, a discussion of how to properly interpret the measured diffraction intensity assuming a linearly polarized incident x-ray beam, details of data collection and experimental geometry used, details of the numerical inversion process, and how to extract quantitative information from the reconstructions.

5.5.1 Sample Preparation

The sample investigated here consisted of 116 layers of 0.41 nm of Fe and 0.45 nm of Gd, capped with 2 nm of Ta to prevent oxidation. The multilayer was deposited on a 100 nm thick Si_3N_4 membrane at room temperature at an Ar pressure of 3 mTorr with a deposition rate of 0.04 nm/s for Fe and 0.18 nm/s for Gd.

5.5.2 X-ray Diffraction Intensity

The resonant elastic scattering amplitude for polarized photons interacting with this dichroic sample is

$$\begin{aligned} f^{res} = F^{(0)}\hat{e}\cdot\hat{e}'c(\mathbf{r}) - iF^{(1)}(\hat{e}\times\hat{e}')\cdot\hat{m}(\mathbf{r}) \\ + F^{(2)}(\hat{e}'\cdot\hat{m}(\mathbf{r}))(\hat{e}\cdot\hat{m}(\mathbf{r})), \end{aligned} \tag{5.2}$$

where $\hat{e}$ and $\hat{e}'$ are Jones vectors of the incident and scattered radiation respectively, $\hat{m}(\mathbf{r})$ is the unit vector magnetization structure factor, as shown in Fig. 5.1a, and $c(\mathbf{r})$ is the resonant charge structure factor [52, 53]. The $F^{(0,1,2)}$ are complex (dispersive and absorptive) resonant scattering strengths, with $F^{(0)}$ describing any non-magnetic scattering, $F^{(1)}$ describing circular dichroic scattering and $F^{(2)}$ describing linear dichroic scattering. In our experiment (see Fig. 5.1a), the incident radiation was linearly polarized and the resonant Fraunhofer diffraction intensity

is given by

$$I \propto |F^{(0)}\mathcal{F}\{p(\mathbf{r})c(\mathbf{r})\}|^2 + |F^{(1)}\mathcal{F}[p(\mathbf{r})m_z(\mathbf{r})]|^2, \tag{5.3}$$

where $\mathcal{F}$ denotes the spatial Fourier transform, $p(\mathbf{r})$ is the x-ray probe (see Fig. 5.2b), and $m_z(\mathbf{r})$ describes magnetization direction parallel/antiparallel to the propagation vector $\hat{k}$. The scattering mechanism from the $m_x(\mathbf{r})$ and $m_y(\mathbf{r})$ components within Bloch walls is x-ray magnetic linear dichroism (XMLD), which is many times weaker than the XMCD signal [66]. Also, the Bloch wall width is a small fraction of the size of the ferrimagnetic domains, and so the scattering contribution from this in-plane magnetization is negligible.

5.5.3 Data Collection

The experiments were conducted at the 2-ID-B coherent imaging beamline at the Advanced Photon Source. The pinhole aperture used in Fig. 5.1a was 10 μm in diameter and positioned 6.5 mm away from the sample. We scanned a 4 by 4 grid with 3 μm step sizes to give a total illuminated field of view of 19 μm by 19 μm. A CCD camera with 2048 x 2048 pixels (pixel dimension $a^2 = 13.5$ μm by 13.5 μm) was placed $z = 455$ mm downstream from the sample. This gave a reconstructed square field of view of width $L = \lambda z/a = 35$ μm for the array sizes in the reconstructions and a pixel size of 35 μm/2048 = 17.1 nm, where the wavelength $\lambda = 1.04$ nm at 1189 eV. As the illuminated region on the sample has an area of $A = \pi(5\ \mu\text{m})^2$, the sampling ratio is about $L^2/A \simeq 15.6$. We collected 40 exposures at each of the 16 illuminated regions on the sample, each with an exposure time of 0.5 s, resulting in estimated flux per exposure of 2.5×10^5 photons / s $\cdot\mu$m^2. The 40 exposures were summed at each scan position to form a single diffraction pattern with improved counting statistics. Approximately 95 % of the radiation is absorbed on resonance (see Fig. 5.1b), resulting in an estimated radiation dose per exposure of $5.6 \cdot 10^4$ Gy.

5.5.4 Numerical Inversion

We used the difference map algorithm [30, 31, 34] combined with the ptychographic steepest descent method [13] to reconstruct the Gd spin density and x-ray probe. The square-root values of the integrated diffraction intensities at each scan position were used as constraints in reciprocal space. Unmeasured pixels behind the beam stop were left unconstrained. The known scan positions (with 70 % probe overlap between adjacent positions) were used as constraints in real space. In addition, because the dispersive part of the refractive index ideally vanishes on-resonance [66], the imaginary part of the reconstructed map (arising from the real component of the refractive index) can be constrained to be small. We therefore constrained the imaginary part of reconstructed density map between $\pm 0.05i$ for the first few hundred iterations and then gradually relaxed it to $\pm 0.5i$ for subsequent iterations. We solved the Gd spin density and probe simultaneously as in [34] using a random set of initial phases for the sample and a set of uniform amplitudes and phases over a circle of diameter 10 μm as the initial guess for the probe. The diffracted intensity was highest when the average sample magnetization was close to zero and the number of magnetic domains was greatest. At low applied fields the reconstruction of the Gd spin density and x-ray probe converged in about 500 iterations of the difference map followed by 100 iterations of the steepest descent method. When the sample was close to magnetic saturation, using a previously reconstructed probe, about 800 iterations of the difference map followed by 100 iterations of steepest descent were necessary to obtain a high quality reconstruction.

5.5.5 Quantitative Gd Spin Density Maps

Quantitative maps of the Gd magnetization $M(\mathbf{r})$ were obtained by scaling the reconstructions by the calculated XMCD contrast - the dependence of f^{res} on the relative direction of the Gd spin magnetization and photon helicity - based on the measured x-ray transmission. The x-ray intensity just exiting the sample can be written as

$$I_{\pm} = I_0 \exp\{-\mu_{\pm} D\} \tag{5.4}$$

where I_0 is the intensity incident on the sample, D is the net thickness of the Gd layers, $\pm$ represents the photon polarization helicity, and $\mu_{\pm}$ is the polarization dependent extinction coefficient for a given out-of-plane magnetization. The reconstructed amplitudes $\Phi(\mathbf{r})$ of the on-resonance diffraction patterns vary from $\Phi_C - \Phi_M$ to $\Phi_C + \Phi_M$, where Φ_C and Φ_M represent charge and magnetic scattering component amplitudes, respectively. The XMCD spectrum is defined as the difference between μ_+ and μ_-, the difference in extinction coefficients between the two out-of-plane magnetization directions for a given circular polarization:

$$\mu_+ - \mu_- = \frac{1}{D}\ln\frac{I_+}{I_-} = \frac{2}{D}\ln\frac{\Phi_C + \Phi_M}{\Phi_C - \Phi_M} \tag{5.5}$$

The absolute value of the magnetization is directly related to this XMCD contrast through the known magneto-optical constants of Gd [66]. From the reconstructions, $\Phi_M/\Phi_C = 0.49$, giving an XMCD atomic photoabsorption cross-section of 0.24 Å^2, in good agreement with XMCD spectra measured with a fully saturated Gd moment [66], corresponding to an areal magnetization of $M_s = 0.011$ emu/cm^2.

Note that due to the use of linearly polarized light, a negative of any reconstructed map of $M(\mathbf{r})$, i.e. $-M(\mathbf{r})$, would also satisfy the constraints. This two-fold degeneracy in the solution is lifted upon application of a magnetic field, since the Gd magnetization aligns in the direction of applied field. Measurements at the Gd M edges provide us with an elemental distribution of the Gd magnetization. The measurements would have to be repeated at the Fe M edges to obtain similar maps of the Fe magnetization. However, since Gd and Fe are known to be antiferromagnetically coupled, the Fe magnetization map is expected to be the negative of the Gd magnetization map, re-scaled by the ratio of the magnetic moments and total layer thicknesses for the two elements.

5.6 Conclusions

We have demonstrated x-ray microscopy of an extended disordered magnetic nanostructure using numerical inversion of a scanned series of resonant dichroic coherent diffraction patterns. Like x-ray holography and transmission x-ray microscopy, this approach can be used to image magnetic domains at nanometer resolution. In contrast to these techniques, the approach described here does not require circularly polarized light for magnetic contrast nor nano-fabricated optics or apertures to reach high spatial resolution. Indeed, it works equally well using circularly or linearly polarized light, allowing this approach to be expanded to image both magnetic and structural heterogeneity. As it replaces optical elements with a "numerical" lens, in addition to greatly simplifying the experiment, the spatial resolution is solely determined by the signal to noise at high spatial frequencies recorded in the coherent diffraction measurements and by the precision and stability with which the sample is scanned in the x-ray beam. It can be used to image an arbitrarily large area of interest, and the sample under investigation can be easily interchanged without the need for additional off-resonance calibration measurements. The technique can be adapted for imaging a wide range of dichroic materials. Study of order parameters on ultrafast time scales by this technique would require working below radiation damage thresholds due to the probe overlap requirement of the reconstruction algorithm used here. Use of additional real space constraints such as known material indices of refraction [92], compressed sensing [93], and probes with phase curvature [17] may enable extension of this approach to single-view dichroic imaging. While the measurements presented here were performed at a third-generation synchrotron, this experimental technique will benefit tremendously as the new generation of high brilliance x-ray sources that produce beams with nearly full transverse coherence, including ultra low-emittance synchrotrons [94], energy recovery linacs [95], and free electron lasers [96] become available.

Chapter 5, in full, is a reprint of the material as it appears in A. Tripathi, J. Mohanty, S. Dietze, O. G. Shpyrko, E. Shipton, E. Fullerton, S.S. Kim and I. McNulty, "Dichroic Coherent Diffractive Imaging", *Proc. Natl. Acad. Sci. USA*,

108, 13393-8 (2011). Ashish Tripathi was the primary investigator and author of this paper.

Bibliography

[1] J. W. Goodman, *Introduction to Fourier Optics, 3rd Edition*, Roberts & Company, Englewood, 2005

[2] 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*, Opt. Exp. **17**, 17669 (2009)

[3] A. Sakdinawat and D. Attwood, *Nanoscale X-ray imaging*, Nat. Photonics **4**, 840 (2010)

[4] I. McNulty , J. Kirz, C. Jacobsen, E. H. Anderson, M. R. Howells, and D. P. Kern, *High-Resolution Imaging by Fourier Transform X-ray Holography*, Science **256**, 1009 (1992)

[5] S. Eisebitt, J. Lüning, W. F. Schlotter, M. Lörgen, O. Hellwig, W. Eberhardt, and J. Stöhr, *Lensless imaging of magnetic nanostructures by X-ray spectro-holography*, Nature **432**, 885 (2004)

[6] D. Sayre, *Some implications of a theorem due to Shannon*, Acta Cryst. **5**, 843 (1952).

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

[8] J. Fienup, *Reconstruction of an object from the modulus of its Fourier transform*, Opt. Lett. **3**, 27 (1978)

[9] Y.M. Bruck and L.G. Sodin, *On the ambiguity of the image reconstruction problem*, Opt. Comm. **30**, 304 (1979)

[10] J. R. Fienup, *Phase retrieval algorithms: a comparison*, App. Opt. **21**, 2758 (1982)

[11] R. H. T. Bates, *Fourier phase problems are uniquely solvable in more than one dimension. I. Underlying theory*, Optik **61**, 247 (1982)

[12] J. Miao, P. Charalambous, J. Kirz, and D. Sayre, *Extending the methodology of X-ray crystallography to allow imaging of micrometre-sized non-crystalline specimens*, Nature **400**, 342 (1999)

[13] J.M. Rodenburg and H.M.L. Faulkner, *A phase retrieval algorithm for shifting illumination*, Appl. Phys. Lett. **85**, 4795 (2004)

[14] S. Marchesini, *A unified evaluation of iterative projection algorithms for phase retrieval*, Rev. Sci. Instrum. **78**, 011301 (2007)

[15] S. Marchesini, *Phase retrieval and saddle-point optimization*, J. Opt. Soc. Am. A **24**, 3289 (2007)

[16] G. J. Williams, H. M. Quiney, B. B. Dhal, C. Q. Tran, K. A. Nugent, A. G. Peele, D. Paterson, and M. D. de Jonge, *Fresnel Coherent Diffractive Imaging*, Phys. Rev. Lett. **97**, 025506 (2006)

[17] B. Abbey, K. A. Nugent, G. J. Williams, J. N. Clark, A. G. Peele, M. A. Pfeifer, M. D. de Jonge and I. McNulty, *Keyhole coherent diffractive imaging* Nat. Physics **4**, 394 (2008)

[18] T. J. Davis, *A unified treatment of small-angle X-ray scattering, X-ray refraction and absorption using the Rytov approximation*, Acta Cryst. A **50**, 686 (1994)

[19] D. M. Paganin, *Coherent X-Ray Optics*, Oxford University Press, New York, 2006

[20] D. Mas, J. Pérez, C. Hernández, C. Vázquez, J.J. Miret, and C. Illueca, *Fast numerical calculation of Fresnel patterns in convergent systems*, Opt. Comm. **227**, 245 (2003)

[21] M. Sypek, C. Prokopowicz, and M. Gorecki, *Image multiplying and high-frequency oscillations effects in the Fresnel region light propagation simulation*, Opt. Eng. **42**, 3158 (2003)

[22] D. Mas, J. Garcia, C. Ferreira, L. M. Bernardo, and F. Marinho, *Fast algorithms for free-space diffraction patterns calculation*, Opt. Comm. **164**, 233 (1999)

[23] J. García, D. Mas, and R. G. Dorsch, *Fractional-Fourier-transform calculation through the fast-Fourier-transform algorithm*, App. Opt. **35**, 7013 (1996)

[24] J. Miao, D. Sayre, and H.N. Chapman, *Phase retrieval from the magnitude of the Fourier transforms of nonperiodic objects*, J. Opt. Soc. Am. A **15**, 1662 (1998)

[25] P. Wolfe, *Convergence conditions for ascent methods*, SIAM Rev. **11**, 226 (1969)

[26] J. Nocedal and S. J. Wright, *Numerical optimization*, Springer Verlag, New York, NY, 1999

[27] J.R. Shewchuk, *An introduction to the conjugate gradient method without the agonizing pain*, (Tech. Rep. CMU-CS-94-125) Pittsburgh, PA: School of Computer Science, Carnegie Mellon University, (1994)

[28] J. Nocedal, *Updating Quasi-Newton Matrices with Limited Storage*, Math. of Comp. **35**, 773 (1980)

[29] A. Levi and H. Stark, *Image restoration by the method of generalized projections with application to restoration from magnitude*, J. Opt. Soc. Am. A **1**, 932 1984

[30] V. Elser, *Phase retrieval by iterated projections*, J. Opt. Soc. Am. A **20**, 40 (2003)

[31] V. Elser, I. Rankenburg, and P. Thibault, *Searching with iterated maps*, Proc. Natl. Acad. Sci. USA **104**, 418 (2007)

[32] S. Gravel and V. Elser *Divide and concur: A general approach to constraint satisfaction*, Phys. Rev. E **78**, 036706 (2008)

[33] 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*, Phys. Rev. B **68**, 140101 (2003)

[34] P. Thibault, M. Dierolf, A. Menzel, O. Bunk C. David, and F. Pfeiffer, *High-Resolution Scanning X-ray Diffraction Microscopy*, Science **321**, 379 (2008)

[35] A.M. Maiden and J.M. Rodenburg, *An improved ptychographical phase retrieval algorithm for diffractive imaging*, Ultramicroscopy **109**, 1256 (2009)

[36] C.M. Kewish, M. Guizar-Sicairos, C. Liu, J. Qian, B. Shi, C. Benson, A.M. Khounsary, J. Vila-Comamala, O. Bunk, J.R. Fienup, A.T. Macrander, and L. Assoufid, *Reconstruction of an astigmatic hard X-ray beam and alignment of K-B mirrors from ptychographic coherent diffraction data*, Opt. Exp. **18**, 23420 (2010)

[37] M. Guizar-Sicairos and J.R. Fienup, *Phase retrieval with transverse translation diversity: a nonlinear optimization approach*, Opt. Exp. **16**, 7264 (2008)

[38] 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*, J. Opt. Soc. Am. A **23**, 1179 (2006)

[39] J. Steinbrener, J. Nelson, X. Huang, S. Marchesini, D. Shapiro, J.J. Turner, and C. Jacobsen, *Data preparation and evaluation techniques for x-ray diffraction microscopy*, Opt. Exp. **18**, 18598 (2010)

[40] Y. Nishino, J. Miao, and T. Ishikawa, *Image reconstruction of nanostructured nonperiodic objects only from oversampled hard x-ray diffraction intensities*, Phys. Rev. B **68**, 22010 (2003)

[41] P. Thibault, V. Elser, C. Jacobsen, D. Shapiro and D. Sayre, *Reconstruction of a yeast cell from X-ray diffraction data*, Acta Cryst. A **62**, 248 (2006)

[42] A. Hubert and R. Schäfer, *Magnetic Domains: The Analysis of Magnetic Microstructures*, Springer, (2000)

[43] A. Moser, K. Takano, D. T. Margulies, M. Albrecht, Y. Sonobe, Y. Ikeda, S. Sun and E. E Fullerton, *Magnetic Recording: Advancing into the Future*, J. Phys. D: Appl. Phys. **35**, R157 (2002)

[44] C. Chappert, F. Fert, F.N. Van Dau, *The emergence of spin electronics in data storage*, Nat. Materials **6**, 813 (2007)

[45] M.A.M Gijs and M Okada *Magnetoresistance and exchange coupling of Fe/Cr magnetic multilayers: dependence on temperature and Cr thickness*, J. Magn. Magn. Mater. **113**, 105 (1992)

[46] J. Swift and P. C. Hohenberg, *Hydrodynamic fluctuations at the convective instability*, Phys. Rev. A **15**, 319 (1977)

[47] J. W. Cahn and J. E. Hilliard, *Free energy of a nonuniform system. I. Interfacial energy*, J. Chem. Phys. **28**, 258 (1958)

[48] D.J. Griffiths, *Introduction to Electrodynamics (3rd Edition)*, Benjamin Cummings, (1999)

[49] A. Messiah, *Quantum Mechanics, Volume II.* North-Holland Publishing Company, Amsterdam, 4th edition (1966).

[50] M. Blume, *Magnetic scattering of x rays*, J. Appl. Phys. **57**, 3615 (1985)

[51] S.W. Lovesey and S.P. Collins, *X-ray Scattering and Absorption by Magnetic Materials*, Oxford University Press, New York, 1996

[52] F. de Bergevin, M. Brunel, *Diffraction of X-rays by Magnetic Materials. I. General Formulae and Measurements on Ferro- and Ferrimagnetic Compounds*, Acta. Cryst. A **37**, 314 (1981)

[53] J.P Hannon, G.T. Tramell, M. Blume, D. Gibbs, *X-ray Resonance Exchange Scattering*, Phys. Rev. Lett. **61**, 1245 (1988)

[54] M. Blume and D. Gibbs, *Polarization dependence of magnetic x-ray scattering*, Phys. Rev. B **37**, 1780 (1988)

[55] J.P. Hill and D.F. McMorrow *X-ray Resonant Exchange Scattering: Polarization Dependence and Correlation Functions*, Acta Cryst. A **52**, 236 (1996)

[56] S.W. Lovesey, *Photon scattering by magnetic solids*, Rep. Prog. Phys., 257-326 (1993)

[57] S.J.P. Konings, *Pinning of magnetic domains studied with resonant x-rays*, Ph.D. thesis, Universiteit van Amsterdam (2007)

[58] J.M. Soriano, *Static and dynamic X-ray resonant magnetic scattering studies on magnetic domains*, Ph.D. thesis, Universiteit van Amsterdam (2005)

[59] Y. Mimura, N. Imamura, T. Kobayashi, A. Okada, and Y. Kushiro, *Magnetic properties of amorphous alloy films of Fe with Gd, Tb, Dy, Ho, or Er*, J. Appl. Phys. **49**, 1208 (1978)

[60] D. Raasch, J. Reck, C. Mathieu, and B. Hillebrands, *Exchange stiffness constant and wall energy density of amorphous GdTb-FeCo thin films* J. Appl. Phys. **76**, 1145 (1994)

[61] C. Mathieu, B. Hillebrands, and D. Raasch, *Exchange stiffness constant and effective gyromagnetic factor of Gd, Tb, and Nd containing, amorphous rare earth-transition metal film*, IEEE T. Magn. **30**, 4434 (1994)

[62] S. Mangin, L. Thomas, F. Montaigne, W. Lin, T. Hauet, and Y. Henry, *Angle dependence of the interface magnetic configuration in a model antiferromagnetically coupled ferrimagnetic/ferrimagnetic bilayer GdFe/TbFe*, Phys. Rev. B **80**, 224424 (2009)

[63] Marcus Lörgen, *Coherent Soft X-Ray Magnetic Scattering and Lensless Imaging*, Ph.D. thesis, Technischen Universität Berlin, (2004)

[64] S. Eisebitt, M. Lörgen,W. Eberhardt, J. Lüning, J. Stöhr, C. T. Rettner, O. Hellwig, E. E. Fullerton, and G. Denbeaux, *Polarization effects in coherent scattering from magnetic specimen: Implications for x-ray holography, lensless imaging, and correlation spectroscopy*, Phys. Rev. B **68**. 104419 (2003)

[65] J. F. Peters, *Resonant soft x-ray scattering studies of the magnetic nanostructure of stripe domains*, Ph.D. thesis, Universiteit van Amsterdam (2003)

[66] J. F. Peters, J. Miguel, M.A. de Vries, O.M. Toulemonde, J.B. Goedkoop, S.S. Dhesi, N.B. Brookes, *Soft x-ray resonant magneto-optical constants at the Gd $M_{4,5}$ and Fe $L_{2,3}$ edges*, Phys. Rev. B **70**, 224417 (2004)

[67] J. B. Kortright and S. K. Kim, *Resonant magneto-optical properties of Fe near its 2p levels: Measurement and applications*, Phys. Rev. B **62**, 12216 (2000)

[68] H. C. Mertins, F. Schafers, X. Le Cann, A. Gaupp, and W. Gudat, *Faraday rotation at the 2p edges of Fe, Co, and Ni*, Phys. Rev. B **61**, R874 (2000)

[69] H. C. Mertins, O. Zaharko, A. Gaupp, F. Schafers, D. Abramsohn, and H. Grimmer, *Soft x-ray magneto-optical constants at the Fe 2p edge determined by Bragg scattering and Faraday effect*, J. Magn. Magn. Mater. **240**, 451 (2002)

[70] H. Coles and S. Morris, *Liquid-crystal lasers*, Nat. Photonics **4**, 676 (2010)

[71] H. Ade and B. Hsiao, *X-ray Linear Dichroism Microscopy*, Science **262**, 1427 (1993)

[72] H. Ade and H. Stoll *Near-edge X-ray absorption fine-structure microscopy of organic and magnetic materials*, Nat. Materials **8**, 281 (2009)

[73] P. Fischer, D. H. Kim, W. Chao, J. A. Liddle, E. H. Anderson, and D. T. Attwood, *Soft X-ray microscopy of nanomagnetism*, Materials Today **9**, 26 (2006)

[74] R. A. Metzler, M. Abrecht, R. M. Olabisi, D. Ariosa, C. J. Johnson, B. H. Frazer, S. N. Coppersmith, and P. U. P. A. Gilbert, *Architecture of Columnar Nacre, and Implications for Its Formation Mechanism*, Phys. Rev. Lett. *98*, 268102 (2007)

[75] J. Stöhr, M. G. Samant, J. Lüning, A. C. Callegari, P. Chaudhari, J. P. Doyle, J. A. Lacey, S. A. Lien, S. Purushothaman, and J. L. Speidell, *Liquid Crystal Alignment on Carbonaceous Surfaces with Orientational Order*, Science **292**, 2299 (2001)

[76] M.R. Freeman and B.C. Choi, *Advances in Magnetic Microscopy*, Science **294**, 1484 (2001)

[77] J. Miao, T. Ishikawa, B. Johnson, E. H. Anderson, B. Lai, and K. O. Hodgson, *High Resolution 3D X-Ray Diffraction Microscopy*, Phys. Rev. Lett. **89**, 088303 (2002)

[78] M.A. Pfeifer, G. J. Williams, I. A. Vartanyants, R. Harder and I. K. Robinson *Three-dimensional mapping of a deformation field inside a nanocrystal.* Nature **442**, 63 (2006)

[79] C. Song, R. Bergstrom, D. Ramunno-Johnson, H. Jiang, D. Paterson, M. D. de Jonge, I. McNulty, J. Lee, K. L. Wang, and J. Miao, *Nanoscale Imaging of Buried Structures with Elemental Specificity Using Resonant X-Ray Diffraction Microscopy*, Phys. Rev. Lett. **100**, 025504 (2008)

[80] 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*, Proc. Natl. Acad. Sci. USA **107**, 529 (2010)

[81] C. Kittel, *Physical Theory of Ferromagnetic Domains*, Rev. Mod. Phys. **21**, 541 (1949)

[82] C. Kooy and U. Enz, *Experimental and theoretical study of the domain configuration in thin layers of* $BaFe_{12}O_9$, Philips Res Rep **15** 181 (1960)

[83] M. Seul and R. Wolfe, *Evolution of disorder in two-dimensional stripe patterns: "Smectic" instabilities and disclination unbinding*, Phys. Rev. Lett. **68**, 2460 (1992)

[84] M. Seul and D. Andelman, *Domain Shapes and Patterns: The Phenomenology of Modulated Phases*, Science **267**, 476 (1995)

[85] B. Koopmans, G. Malinowski, F. Dalla Longa, D. Steiauf, M. Fähnle, T. Roth, M. Cinchetti and M. Aeschlimann, *Explaining the paradoxical diversity of ultrafast laser-induced demagnetization*, Nat. Materials **9**, 259 (2010)

[86] 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*, Proc Natl Acad Sci USA **102**, 15343 (2005)

[87] G. Malescio and G. Pellicane, *Stripe phases from isotropic repulsive interactions*, Nat. Materials **2**, 97 (2003)

[88] R. Plass, J.A. Last, N.C. Bartelt, G.L. Kellogg, *Nanostructures: Self-assembled domain patterns*, Nature **412**, 875 (2001)

[89] O. Hellwig, G. Denbeaux, J.B. Kortright, and E.E. Fullerton, *X-ray studies of aligned magnetic stripe domains in perpendicular multilayers*, Physica B **336**, 136 (2003)

[90] J. E. Davies, O. Hellwig, E. E. Fullerton, G. Denbeaux, J. B. Kortright, and K. Liu, *Magnetization reversal of Co/Pt multilayers: Microscopic origin of high field magnetic irreversibility*, Phys. Rev. B **70**, 224434 (2004)

[91] J. E. Davies, O. Hellwig, E. E. Fullerton, M. Winklhofer, R. D. Shull, and K. Liu, *Frustration driven stripe domain formation in Co/Pt multilayer films*, Appl. Phys. Lett. **95**, 022505 (2009)

[92] J. N. Clark, C. T. Putkunz, M. A. Pfeifer, A. G. Peele, G. J. Williams, B. Chen, K. A. Nugent, C. Hall, W. Fullagar, S.S. Kim, and I. McNulty, *Use of a complex constraint in coherent diffractive imaging*, Opt. Exp. **18**, 1981 (2010)

[93] W. L. Chan, M. L. Moravec, R. G. Baraniuk, and D. M. Mittleman, *Terahertz imaging with compressed sensing and phase retrieval*, Opt. Lett. **33**, 974 (2008)

[94] D. H. Bilderback, P. Elleaume, and E. Weckert, *Review of third and next generation synchrotron light sources*, J Phys. B: At. Mol. Opt. Phys. **38**, S773 (2005)

[95] S. M. Gruner, D. Bilderback, I. Bazarov, K. Finkelstein, G. Krafft, L. Merminga, H. Padamsee, Q. Shen, C. Sinclair, and M. Tigner, *Energy recovery linacs as synchrotron radiation sources*, Rev. Sci. Instrum. **73**, 1402 (2002)

[96] K. J. Gaffney and H.N. Chapman, *Imaging atomic structure and dynamics with ultrafast X-ray scattering*, Science **316**, 1444 (2007)