Title

Inference from Incomplete Data in Coherent Diffraction Imaging

Author

Salha, Sara

Publication Date

2014

UNIVERSITY OF CALIFORNIA
Los Angeles

Inference from Incomplete Data in Coherent Diffraction Imaging

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

by

Sara Salha

2014

Abstract of the Dissertation

Inference from Incomplete Data in Coherent Diffraction Imaging

by

Sara Salha

Doctor of Philosophy in Physics
University of California, Los Angeles, 2014
Professor Jianwei (John) Miao, Chair

Progress in nanotechnology and biotechnology are propelled by our ability to manipulate and resolve the structure of matter on fine scales. As imaging at higher resolution is limited by the probing light source and the numerical aperture, lensless imaging offers an advantage over lensed microscopy. Dispensing with lenses allows one to overcome certain intrinsic aberrations and to bypass fabrication costs, in the optical and the X-ray regimes. The long penetration depth of X-rays renders coherent X-ray diffraction imaging (CXDI) the method of choice for high resolution structure determination with broad applications from materials science to biology; moreover, the same methodology is extensible to electrons, optical photons, or even gamma rays or neutrons. Since coherent diffraction imaging (CDI) bypasses the need for focusing optics, it relies upon computer algorithms to reconstruct the structure of the scattering object. Currently, one of the main obstacles to nanometer resolution of biological imaging is noisy, incomplete data due to radiation damage. With the rapid development of new light source facilities and the advancement in image reconstruction techniques, determining the structure of individual virons or cells at high resolution is becoming more feasible. In particular, the femtosecond pulse of a free electron laser (FEL) is shorter than the coulomb explosion of the specimen, and thus, it is possible to collect diffraction data prior to radiation damage. However, to fully exploit the computational aspect of lensless imaging, prior knowledge about the object should be incorporated into the image reconstruction process and yet so far such methods are generally lacking. In this thesis, we develop tools that incorporate prior knowledge and reduce the amount of necessary data to recover the structure. We begin by a brief overview of lensless imaging and its place in the natural sciences. we then review the process of image formation in coherent X-ray scattering, the corresponding phase problem and the current state of image recovery. The contributions to this field are two fold. We first demonstrate that three dimensional information can be extracted from a two dimensional diffraction pattern collected at a high numerical aperture. Second, we present a framework for image discovery through Bayesian inference, where we introduce four general constraints: symmetry, sparsity and bounded local and total variation. Using simulated noisy, incomplete data, we recover the solution in situations where traditional algorithms fail. We anticipate that these results will encourage the broader application of Bayesian learning into the phase retrieval problem from noisy, incomplete diffraction data and further enhance the possibility of single shot three dimensional structure determination.

The dissertation of Sara Salha is approved.

Pietro Musumeci

Pei-Yu (Eric) Chiou

Jianwei (John) Miao, Committee Chair

University of California, Los Angeles

2014

To my mother and my grandmother
who never tired of asking
“when are you going to finish”.
To my brother and father
who supported me along the way.
...
To the golden eagle
whose genuine empathy
taught how to fish.
...
To you
I am forever
grateful.

Table of Contents

List of Figures

LIST OF TABLES

ACKNOWLEDGMENTS

Thank you UCLA and Eugene V. Cota-Robles Fellowship for the opportunity to do research. Thank you Professor John Miao and group for the opportunity to learn about lensless imaging. Thank you Kevin S. Raines for your enlightened wisdom, objective support and creative approach.

Vita

2011	Lawrence Berkeley National Lab, summer internship.
2008-2011	Coherent X-ray Diffraction Imaging, UCLA.
2006-2011	Teacher assistant, UCLA.
2006-2007	M.S. Astronomy, UCLA.
2005	Physical Optics Coorporation, Torrance CA.
2002-2004	B.S Physics, UCLA. California Institute of Technology, summer internship
2000-2002	Loyola Marymount University, Summer internship. A.A. Mathematics, Fullerton Community College (FCC). High school Tutor, FCC.

Publications

1. S. Salha, K. S. Raines, "Adaptive Phase Retrieval for Noisy and Incomplete Data", 2014 (Manuscript in preparation)
2. C. Chen, H. Jiang, L. Rong, S. Salha, R. Xu, T. Mason, J. Miao"3D Imaging of a Phase Object from a Single Sample Orientation Using an Optical Laser". (In review, submitted to PRL: arXiv:1107.3276)

3. R. Xu, S. Salha, K. S. Raines, H. Jiang, C.-C. Chen, Y. Takahashi, Y. Kohmura, Y. Nishino, C. Song, T. Ishikawa, J. Miao, ”Coherent Diffraction Microscopy at SPring-8: Instrumentation, Data Acquisition and Data Analysis”, J. Synch. Rad. 18, 293-298 (2011).
4. K. S. Raines, S. Salha, R. L. Sandberg, H. Jiang, J. A. Rodriguez, B. P. Fahimian, H. C. Kapteyn, J. Du and J. Miao, ”Three-Dimensional Structure Determination From a Single View”, Nature 463, 214-217 (2010).

CHAPTER 1

Introduction

From the 11th century to the present time, our ability to resolve structures on fine scales has been broadening our understanding of nature. The magnification power of a lens was first investigated by Alhazen (1021 AD); six hundred years later, the systematic use of a microscope to study objects unresolved by the naked eye was first documented by Robert Hooke who also coined the word *cell* as a building block of life. Subsequently, the discovery of bacteria by Leeuwenhoek strengthened the evidence of the germ theory of diseases. Fast forwarding to recent decades, with the development of higher resolution imaging using light and electron microscopy, details at the mesoscopic/atomic world are becoming more accessible to technological developments. Take, for example, how the structure determination of an isolated single layered graphene had prompted new areas of research from high speed electronics to low cost solar cells [1].

In this thesis, we focus on a relatively new imaging modality, lensless imaging. This type of imaging has the potential of determining structures of cells at nanometer resolutions and beyond. The structure of this chapter is as follows: we begin with a brief history of lensless imaging, the related phase problem, and how the revolutionary work in information theory inspired a solution to this problem. With the emergence of high brilliance coherent X-ray sources, new research territories and challenges arise; in particular, the problem of structure recovery from insufficient data. To address this issue, we review current developments aimed at suppressing the main problem of radiation damage, overview our proposed algorithmic methods: first by using high numerical aperture data and second by the incorporation of prior knowledge through Bayesian inference.

Lensless imaging is a type of microscopy that dispenses with focusing optics. To form an image, partially coherent light is scattered off the object creating a diffraction pattern which is then synthesized using a computer algorithm or a reference beam. The principle of lensless imaging is founded on the observation of Ernest Abbe (1873) who explained the process of image formation in a compound microscope in two steps: analysis at the focal plane of the objective lens and a synthesis at the eyepiece [2]. As the partially coherent light enters the first lens, it forms a diffraction pattern at the back focal plane; the scattered rays then converge to form a magnified image as shown in figure 1.1. The pattern formed at the focal plane was overlooked by most microscopists; it took nearly a century later, before Gabor investigated the pattern closely and invented *holography* (1947), where he removed the objective lenses to eliminate spherical aberration in the transmission electron microscope (TEM). This new lensless microscope was first demonstrated with partially coherent optical light. Its success relied on the ingenious idea that the wavefront phases are preserved by interfering the primary incident wave with the secondary scattered wave. By re-creating the full wavefield, it is possible to obtain the three dimensional structure of the object [3, 4]. The source size, however, limits the resolution from reaching atomic resolution, unless the atoms themselves are used as the source of the reference wave [5, 6].

Another familiar form of lenless imaging is *crystallography*. With the discovery of X-ray by Roentgen (1895), and the demonstration of X-ray diffraction by Max Von Laue (1912), crystallography played a fundamental role in understanding atomic and molecular objects under crystallization. This form of imaging is limited to materials that can be crystallized, where the phase problem is solved through various methods, i.e molecular replacement or using the Patterson function, etc. For a general object, amorphous, un-crystalized object, however, the

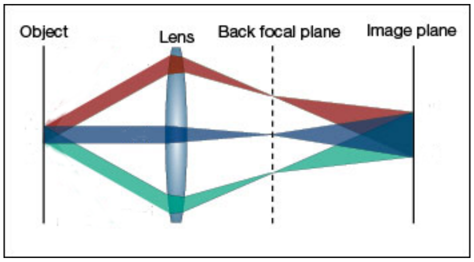

Figure 1.1: Image formation using a simple microscope. For a compound microscope, a second lens (eyepiece), placed upstream from the image plane, magnifies the image formed by the first lens.

solution to the phase problem requires a new approach.

With the development of crystallography during the middle of the last century, another field was born outside of physics that would prove to have a significant impact upon science in general and lensless imaging in particular, laying the foundation for the solution to the phase problem. With the foundational work of Claude Shannon, the field of information theory rapidly progressed, where among many other contributions, he proved the sampling theorem that bears his name, which gives the conditions under which the exact recovery of a bandlimited signal from the discrete measurements is possible [7, 8]. A few years later, David Sayre, while working on protein crystallography, was inspired by Shannon's work and realized that the main limitation to X-ray crystallography is the limited sampling measurements due to Bragg peaks which, according to Shannon's theorem, are insufficient to determine the electron density. Subsequently, he wrote an insightful note about the possibility of reconstructing a signal from the phase-less diffraction data, under the condition of extra sampling, namely *oversampling*, using a non-crystalline object, where the diffraction pattern is continuous, unlike the discrete pattern of crystals [9].

In addition to the oversampling idea, coherent X-ray diffraction imaging (CDI) from general objects became possible due to progress in two different fields: the light source technology and algorithmic methods. With the emergence of high brilliance coherent X-ray sources [10, 11, 12], it became possible to obtain excellent quality X-ray diffraction data from amorphous objects. Furthermore, the rapid development of computational power, facilitated the implementation of iterative projection algorithms. These algorithms recursively apply the constraints in multiple spaces, i.e measurement and image space, discussed in chapter 3. Initiated by the work of Saxton and Gerchberg for two intensity measurements [13] and later vastly accelerated by the work of James Fienup in his attempt to correct for lens aberration on Hubble telescope [14, 15], the error reduction and the hybrid input output algorithms were also implemented as phase retrieval algorithms on lensless imaging data. With the first experimental demonstration of CDI, where a fabricated object was reconstructed from the oversampled diffraction pattern [16], this lensless form of microscopy became an important tool for imaging biological objects, such as bacteria, whole cells, cellular organelles, viruses, and biomaterials where the highest resolution of < 15 nm is been achieved on biological samples [17, 18, 19] and a resolution of ~ 2 nm is obtained with inorganic materials [20, 21, 22, 23, 24, 25].

Although conventional phase retrieval algorithms perform well in the case of diffraction data with adequate signal to noise ratio, the case of noisy and incomplete data remains challenging. Such data is prevalent in high resolution diffraction patterns, where the extreme cases correspond to single molecule and protein structure determinations [26, 27]. At fixed beam energy, to increase the resolution, a larger scattering angle is required, which translates into either higher exposure time or higher radiation level, each of which causes radiation damage. As the main causes for limited data are due to finite flux and radiation sensitivity of the samples [28], various methods had been developed to overcome this effect. Some of the experimental developments are: freeze hydrating the sample to increase the radiation dose tolerance [29], the use of electrospray-generated particles to obtain multiple diffraction patterns of identical copies [30], building femto-second X-ray sources where the samples' electronic destruction time is longer than the pulse duration [31], namely *diffract and destroy* [32]. Simultaneously, algorithmic methods for the noisy incomplete data are also being developed. For example, the field of compressed sensing is recently being extended to phase retrieval where several algorithms are being proposed based upon specific sparsity constraints [33, 34].

The two main contributions in this thesis are ankylography (chapters 4 and 5) and the adaptive phase retrieval algorithm (APR) (chapter 8). Ankylography is as a high numerical aperture lensless imaging method which extrapolates three dimensional information from high numerical aperture data by accurately accounting for the geometrical distribution of the two dimensional measurements (chapter 4). Using ankylography, we successfully reconstructed a simulated glass particle at atomic resolution and a polio virus at < 2 nm resolution. Furthermore, using experimental soft X-ray and optical diffraction data of thin objects, we obtained three dimensional information by accurately recovering the corresponding tilt angles, in reference to the incident beam [35, 36]. Although this method is currently limited to small sized objects ($\sim 14^3$ voxels), with the implementation of robust prior constraints, we anticipate that practical sized entities can be reconstructed as well.

To complement on the previous work of ankylography, we present a method that improves the performance of projection based phase retrieval algorithms when the data are noisy and incomplete (chapter 8). We propose a framework for inferring accurate image estimates from the data through a recursive feedback loop. We extract information subject to certain known constraints from the data, we then apply a non-parametric model from which we extract new constraints. The numerical studies are performed on two dimensional ($2D$) objects, whereas the generalization to $3D$ will be the topic of subsequent research (chapter 9). The three dimensional ($3D$) phase retrieval problem differs from the $2D$ case as it is plagued by noisy, incomplete data to a much greater extent than in $2D$ [37, 38, 39, 40], primarily due to radiation sensitivity, finite flux, and practical experimental concerns; however, the extreme undersample $2D$ case forms a good testing ground.

The chapters in this thesis focus on the phase problem in CDI with an emphasis on noisy, incomplete data. Chapter 2 reviews the theory that lead to understanding the scattering phenomena, chapter 3 presents current phase retrieval algorithms, chapters 4 and 5 report on the experiments and lastly chapters $6-9$ focus on algorithmic methods for overcoming the inverse problem in the case of noisy, incomplete data.

CHAPTER 2

Elements of CDI

In this chapter we present the theoretical elements of a coherent diffraction imaging (CDI) microscope. We review the interaction between the beam and the scattering object, the beam quality and the effects of measurements on the wavefield which give rise to the phase problem whose solution is further discussed in the algorithmic component of CDI in chapter 3.

2.1 Interaction between beam and object

Here we focus on elastic scattering using monochromatic incident beam, where we follow the treatment of Born and Wolf [41]. As we are mainly interested in isotropic, non-magnetic materials, we can later justify the scalar approximation of the wavefield. To derive the scattering equation, we start by Maxwell's equations in matter, also known as the macroscopic equations, where $\vec{E}$ and $\vec{B}$ are the electric and magnetic induction fields, while $\vec{D}$ and $\vec{H}$ are the electric displacement and the magnetic fields. Given a scattering medium with volume V and surface S (Fig. 2.1), with free charge density ρ_f, and a magnetic permeability μ (~ 1 for non-magnetic materials), and a scalar dielectric constant ϵ independent of direction (isotropic material), we have:

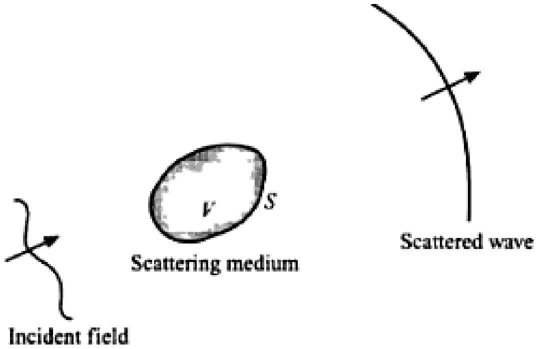

Figure 2.1: A simple sketch of a scattering event; figure obtained from [41].

$$\nabla \cdot \vec{D} = 4\pi\rho_f \quad \text{Gauss's law} \tag{2.1}$$

$$\vec{D} = \epsilon\vec{E} \tag{2.2}$$

$$\nabla \times \vec{E} = -\frac{1}{c}\frac{\partial \vec{B}}{\partial t} \quad \text{Faraday's law} \tag{2.3}$$

$$\vec{B} = \mu\vec{H} \tag{2.4}$$

$$\nabla \times \vec{H} = \frac{4\pi}{c}\vec{J}_f + \frac{1}{c}\frac{\partial \vec{D}}{\partial t} \quad \text{Amper's law} \tag{2.5}$$

$$\vec{J}_f \text{ is the current density}$$

After applying the curl and replacing $\vec{B}$ with $\vec{H}$, Faraday's law yields:

$$\nabla \times (\frac{1}{\mu}\nabla \times \vec{E}) + \frac{1}{c}\nabla \times \frac{\partial \vec{H}}{\partial t} = 0 \tag{2.6}$$

The time derivative of Ampere's law yields:

$$\nabla \times \frac{\partial \vec{H}}{\partial t} = \frac{\epsilon}{c}\ddot{\vec{E}} \tag{2.7}$$

Here we assumed that the material is not a source of charge and hence the continuity equation holds; thus, $\frac{\partial \vec{J}_f}{\partial t} = 0$.

Plugging Eq. 2.7 in 2.6, we find:

$$\begin{aligned}\nabla \times (\frac{1}{\mu}\nabla \times \vec{E}) + \frac{\epsilon}{c^2}\ddot{\vec{E}} = 0\\ \nabla \times \nabla \times \vec{E} + \mu\nabla\frac{1}{\mu} \times \vec{E} + \frac{\epsilon}{c^2}\ddot{\vec{E}} = 0\\ \nabla^2\vec{E} - \nabla(\nabla\cdot\vec{E}) + \nabla ln(\mu) \times \vec{E} - \frac{\epsilon}{c^2}\ddot{\vec{E}} = 0\end{aligned} \tag{2.8}$$

Using Gauss's law, we can find an expression for $\nabla(\nabla \cdot \vec{E})$:

$$\begin{aligned}\nabla \cdot \epsilon\vec{D} = \epsilon\nabla\cdot\vec{E} + \vec{E}\cdot\nabla\epsilon = 4\pi\rho_f\\ \text{Take the gradient}\\ \nabla(\epsilon\nabla\cdot\vec{E}) + \nabla(\vec{E}\cdot\nabla\epsilon) = \nabla(4\pi\rho_f) = 0\\ \text{then divide by } \epsilon\\ \nabla(\nabla\cdot\vec{E}) + \nabla(\vec{E}\cdot\nabla ln\epsilon) = 0\end{aligned} \tag{2.9}$$

Again by plugging Eq. 2.9 in 2.8, we have:

$$\nabla^2\vec{E} + \nabla(\vec{E}\cdot\nabla ln\epsilon) + \nabla ln(\mu) \times \vec{E} - \frac{\epsilon}{c^2}\ddot{\vec{E}} = 0 \tag{2.10}$$

As we mentioned at the beginning of this section, we are interested in non-magnetic isotropic materials, where $\mu \approx 1$ and ϵ varying slowly on the order of the oscillations $\omega = \frac{2\pi}{\lambda} = kc$ of the monochromatic wave $\vec{E}e^{-i\omega t}$. As a result, we can neglect variations in both ϵ, μ terms in Eq. 2.10, and hence the field equation simplifies to:

$$\nabla^2\vec{E} + \epsilon k^2\vec{E} = 0 \tag{2.11}$$

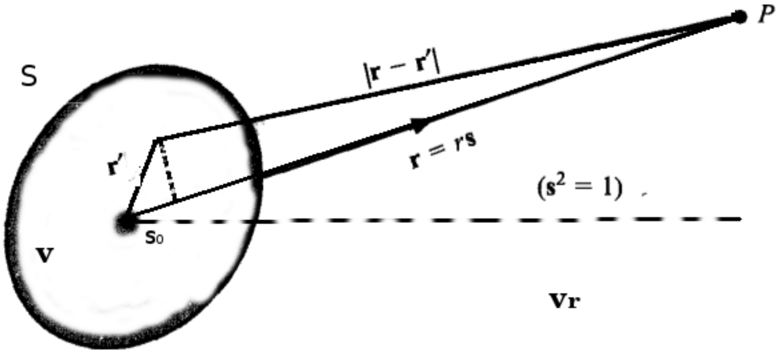

Figure 2.2: The elastic scattering approximation; sketch obtained from [41].

Unlike equation 2.10, equation 2.11 does not couple spatial directions, and hence it is possible to treat the vector equation, component wise. Thus; hereafter, we use the scalar field equation ψ, as the component representation of the original field $\vec{E}$:

$$\nabla^2\psi(\vec{r},\omega) + k^2n^2(\vec{r},\omega)\psi(\vec{r},\omega) = 0 \qquad \text{where } n^2 = \epsilon(\vec{r},\omega) \tag{2.12}$$

where r is the spatial coordinate (Fig. 2.2), and n is the index of refraction.

For convenience, we can express the above equation in a form that is mathematically equivalent to the time independent Schroedinger equation for non-relativistic particles:

$$\begin{aligned}\nabla^2\psi(\vec{r},\omega) + k^2(\vec{r},\omega)\psi(\vec{r},\omega) &= -4\pi F(\vec{r},\omega)\psi(\vec{r},\omega)\\ \text{where } F(\vec{r},\omega) &= \frac{k^2(n(\vec{r},\omega)^2-1)}{4\pi}\end{aligned} \tag{2.13}$$

is the scattering potential of the medium.

Physically, $\psi(\vec{r},\omega)$ is the field resulting from the interaction between the incident wave and the medium V, hence it is the superposition of the total field: incident $\psi^{(i)}$ and scattered $\psi^{(s)}$ fields.

$$\psi(\vec{r},\omega) = \psi^{(i)}(\vec{r},\omega) + \psi^{(s)}(\vec{r},\omega) \tag{2.14}$$

By decomposing ψ into these two components, we can further simplify the subsequent treatment, as the $\psi^{(i)}$ (a plane wave) obeys the helmholtz equation.

$$(\nabla^2 + k^2)\psi^i(\vec{r},\omega) = 0 \tag{2.15}$$

And hence, the scattering equation reduces to :

$$(\nabla^2 + k^2(\vec{r},\omega))\psi^s(\vec{r},\omega) = -4\pi F(\vec{r},\omega)\psi(\vec{r},\omega) \tag{2.16}$$

By introducing the symmetric Green's function $G(\vec{r}-\vec{r}',\omega)$ as the solution to the non-homogeneous helmholtz equation, we have:

$$(\nabla^2 + k^2(\vec{r},\omega))G(\vec{r}-\vec{r}',\omega) = -4\pi\delta^{(3)}(\vec{r}-\vec{r}') \tag{2.17}$$

Multiplying Eq 2.17 and 2.16 by $\psi^{(s)}$ and G respectively, and subtracting the corresponding equations from each other we find:

$$\begin{aligned} &\psi^s(\vec{r},\omega)\nabla^2 G(\vec{r}-\vec{r}',\omega) - G(\vec{r}-\vec{r}',\omega)\nabla^2\psi^s(\vec{r},\omega) \\ &= 4\pi F(\vec{r},\omega)\psi(\vec{r},\omega)G(\vec{r}-\vec{r}',\omega) - 4\pi\delta^{(3)}(\vec{r}-\vec{r}')\psi^s(\vec{r},\omega) \end{aligned} \tag{2.18}$$

To easily solve this equation, we will integrate over the whole space, using the volume Vr. While r' is the coordinate confined to the small volume V, r extends over Vr, as in Fig 2.2.

$$\begin{aligned}&\int_{V_R}[\psi^s(\vec{r},\omega)\nabla^2 G(\vec{r}-\vec{r}',\omega)-G(\vec{r}-\vec{r}',\omega)\nabla^2\psi^s(\vec{r},\omega)]dr^3\\ =&\int_{V_R}[4\pi F(\vec{r},\omega)\psi(\vec{r},\omega)G(\vec{r}-\vec{r}',\omega)-4\pi\delta^{(3)}(\vec{r}-\vec{r}')\psi^s(\vec{r},\omega)]dr^3\end{aligned} \tag{2.19}$$

By integrating over the $\delta^{(3)}(\vec{r}-\vec{r}')$ function and re-arranging the terms above, one can express ψ^s as:

$$\begin{aligned}\psi^s(\vec{r}',\omega)=&\int_{V_R}F(\vec{r},\omega)\psi(\vec{r},\omega)G(\vec{r}-\vec{r}',\omega)]dr^3\\ -\frac{1}{4\pi}&\int_{V_R}[\psi^s(\vec{r},\omega)\nabla^2 G(\vec{r}-\vec{r}',\omega)-G(\vec{r}-\vec{r}',\omega)\nabla^2\psi^s(\vec{r},\omega)]dr^3\end{aligned}$$

Subsequently, we convert the second volume integral to surface integral using Green's theorem, taking note that $F(\vec{r},\omega)$ is confined within V:

$$\begin{aligned}\psi^s(\vec{r}',\omega)=&\int_{V}F(\vec{r},\omega)\psi(\vec{r},\omega)G(\vec{r}-\vec{r}',\omega)]dr^{3\prime}\\ -\frac{1}{4\pi}&\int_{S_R}[\psi^s(\vec{r},\omega)\frac{\partial G(\vec{r}-\vec{r}',\omega)}{\partial n'}-G(\vec{r}-\vec{r}',\omega)\frac{\partial\psi^s(\vec{r},\omega)}{\partial n'}]dS_R\end{aligned}$$

Where S_R is the surface corresponding to V_R, and $\vec{n}'$ is the corresponding normal vector. The Green's function is the solution to the point source wave equation Eq. 2.17, which can be expressed as:

$$G(\vec{r}-\vec{r}',\omega)=\frac{e^{ik|\vec{r}-\vec{r}'|}}{|\vec{r}-\vec{r}'|} \tag{2.20}$$

The surface integral S_R is negligible for $R\rightarrow\infty$. Hence the scattering formula reduces to:

$$\psi^{(s)}(\vec{r},\omega) = \int_V F(\vec{r}',\omega)\psi(\vec{r}',\omega)\frac{e^{ik|\vec{r}-\vec{r}'|}}{|\vec{r}-\vec{r}'|}d^3r' \tag{2.21}$$

with $\psi^{(i)} = e^{ik\vec{s}_0\cdot\vec{r}}$ ($\vec{s}_0$ is the incident beam direction), we obtain the potential scattering equation:

$$\psi(\vec{r},\omega) = \psi^i(\vec{r},\omega) + \int_V F(\vec{r}')\psi(\vec{r}',\omega)\frac{e^{ik|\vec{r}-\vec{r}'|}}{|\vec{r}-\vec{r}'|}d^3r' \tag{2.22}$$

Once the total field $\psi(\vec{r},\omega)$ inside the scattering volume integral V is *determined*, then one can find the total field outside the volume. So this is really a tricky embedded solution. In order to solve it, we use the first order perturbation approximation of Born. This approximation is justified for the case of elastic scattering that we are mainly interested in, where a high incident energy interacts with a weak potential well.

2.1.1 Born approximation

For a medium that scatters weakly, we can use the method of perturbation to solve the scattering equation, and hence inside the integral, the total field can be simply approximated by the incident field. For large R, we have:

$$\begin{aligned} |r - r'| &= r - \vec{s}\cdot\vec{r}' \\ \vec{r} &= r\vec{s} \quad \vec{s}^2 = 1 \end{aligned} \tag{2.23}$$

For the denominator in Eq. 2.22, we can use $|r - r'| \approx r$, while the exponent term is more sensitive to such an approximation, and thus we keep both terms. As a result, we now have:

$$\psi(\vec{r},\omega) = \psi^i(\vec{r},\omega) + \frac{e^{ikr}}{r}\int_V F(\vec{r}',\omega)\psi^i(\vec{s}\cdot\vec{r}',\omega)e^{-ik\vec{s}\cdot r'}d^3r' \tag{2.24}$$

Therefore, for an incident beam that goes through a single scattering event, the output field $\psi(\vec{r}, \omega)$ is a combination of the original incident field and a spherical wave $\frac{e^{ikr}}{r}$ with a scattering amplitude f:

$$\begin{aligned} f(\vec{s}, \vec{s}_0, \omega) &= \int_V F(\vec{r'}, \omega)\psi^i(\vec{r}, \omega)e^{-ik\vec{s}\cdot r'} d^3r' \\ &= \int_V F(\vec{r'}, \omega)e^{-ikr'(\vec{s}-\vec{s}_0)} d^3r' \end{aligned} \tag{2.25}$$

Which can be recognized as the Fourier transform of the scattering potential. Recall Eq. 2.13 for the relation to the macroscopic behaviour of the index of refraction, which in turn can be linked to the electron density by:

$$n = 1 - \frac{N\lambda^2 r_e(f_1 + if_2)}{2\pi} \tag{2.26}$$

where r_e is the radius of the electron, N is the number of atoms in unit volume, and $f_1 + if_2$ is the atomic scattering factor of the material, related to the atomic number and the wavelength λ of the incident beam [42]. In the case where the incident radiation has frequency much smaller than the natural frequency of the material, example visible light scattering off a piece of glass, the scattering factor is negative, and the index of refraction is larger than unity. However, in the X-ray range, the frequency is usually larger than the natural frequency of the material, and hence the index of refraction is negative. The index of refraction is complex, when the two frequencies are within the same range, which is the case for anomalous scattering, or light absorption. In this thesis, we are mainly interested in elastic coherent scattering, with real index of refraction.

As part of the image recovery discussion, we seek the inversion of Eq. 2.25, namely relating $f(\vec{s}, \vec{s}_0, \omega)$ to the measurements.

2.2 The coherence of the incident beam

For the information encoded in the diffraction pattern to be exclusive of the scattering object, the intrinsic beam parameters should satisfy stringent coherence criteria; across the sample and along the propagation distance between the scatterer and the measurement plane. As the light source generator plays a role in shaping the beam's characteristics, we categorize the coherence study into two sections: visible lasers and undulators, applicable to optical and synchrotron X-ray radiations, respectively. Coherent sources are not restricted to X-ray and optical, but these were the two light sources I worked with during my graduate study [36, 43]. A more thorough study would include X-ray free-electron laser and table top soft X-ray sources. For more information about these systems and facilities please see [44, 45, 46].

2.2.1 Optical laser

Light generated by a laser cavity with planar mirrors produces plane wave. Unfortunately, such mirrors configuration is a highly unstable system. As a result, confocal cavities with spherical mirrors are generally used [47]. Here we will follow the treatment provided by Saleh and Teich [48], in describing the beam generated in such a system.

The solution to the wave equation satisfying the boundary conditions of confocal mirrors, and under the paraxial approximation of the Helmholtz wave equation, is a Gaussian field:

$$U(\vec{r}) = A_0 \frac{W_0}{W(z)} \exp^{-(\frac{r}{W(z)})^2} \exp^{-i(kz + \frac{r^2}{2R(z)} - \xi(z))} \tag{2.27}$$

where z_0 is the Rayleigh range along which the beam cross section doubles, as measured from the waist W_0 location ($z = 0$), with

$$W_0 = (\frac{\lambda z_0}{\pi})^{\frac{1}{2}} \qquad W(z) = W_0(1 + (\frac{z}{z_0})^2)^{\frac{1}{2}}$$
$$R(z) = z + \frac{z_0^2}{z} \qquad r^2 = x^2 + y^2 \qquad \xi = \tan^{-1}\frac{z}{z_0} \tag{2.28}$$

Hence the spatial profile of the Gaussian beam imparts not only three dimensional amplitude variations, but an additional phase shift in both the transverse and the longitudinal spatial beam profile.

For a continuous wave laser, the temporal coherence and monochromaticity are related by the Heisenberg uncertainty principle $\tau_c \Delta f \approx 1$. Temporal coherence is practically a measure of the longitudinal coherence, and can be designated by $l_c = c\tau_c$. For a propagation path length $l_{max} << l_c$, coherence is satisfied.

2.2.2 Undulators

Accelerated electrons produce radiation; when the acceleration is originated by a magnetic field, the resulting radiations are called *synchrotron radiation*. Using radio frequency (RF) type of accelerator, built from a sequence of bending magnets, undulators form an important tool of synchrotron radiation at the third generation light sources [49]. The bending magnets produce a sinusoidal magnetic field $\vec{B}$ which acts on the accelerating electrons, and thus confine them into a circular trajectory. The resulting electric field is not static, and hence it becomes possible to accelerate the electrons to a high energy (hard X-rays $\sim 0.5\mathring{A}$). As the electrons circulating the synchrotron are ultra relativistic, the emitted radiation, in the laboratory frame, has a small divergence angle (fig 2.3), with respect to the electron travelling at a velocity v, such that:

$$\tan\theta = \frac{1}{\beta\gamma} \tag{2.29}$$

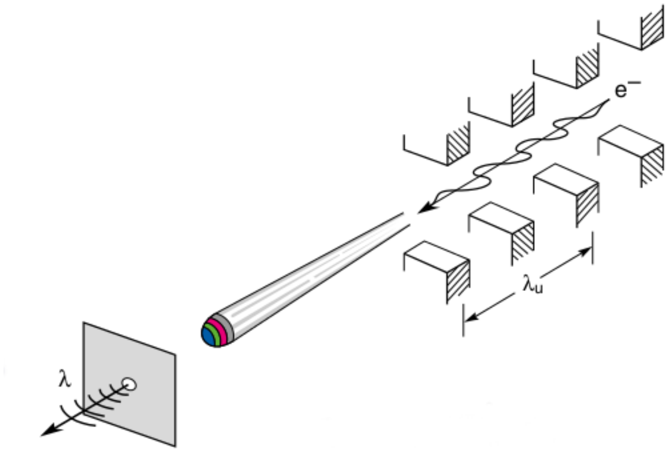

Figure 2.3: A sketch for the accelerated electrons and emitted X-rays in a synchrotron facility; figure obtained from [50].

where $\beta = v/c$ and $\gamma >> 1$. The undulator parameters, namely the magnetic field $\vec{B}$ and the oscillation period of the bending magnets λ_u, both determine the synchrotron radiation wavelength λ, expressed as:

$$\lambda = \frac{\lambda_u}{2\gamma^2}(1 + \frac{K^2}{2} + \gamma^2\theta^2) \tag{2.30}$$

Where $K = \frac{e\hat{B}\lambda_u}{2\pi mc}$, a dimension less quantity. For a more in depth review of third generation radiation facilities, please see [49].

To computed the spatial coherence, let us examine a point emitter located at the central optical axis, with size d and a half opening angle of $\Delta\theta$. Following Heisenberg's uncertainty principle, we have:

$$\Delta x \Delta p = d\Delta p < \hbar/2 = d\Delta k < 1/2$$

By forming a triangle with k and Δk as adjacent and opposite sides, we find

$\Delta k = k\Delta\theta$. Thus, the spatial frequency criterion is:

$$d\Delta\theta < \frac{\lambda}{2\pi} \tag{2.31}$$

For a higher spatial coherence, smaller source size d is desirable, which could be achieved with the use of pinholes.

The temporal coherence should satisfy the same criterion for laser light, namely $l_{max} << l_c$, where l_c can be found, following Attwoods' derivation [50]. Let l_c be the coherence length, where the wave with wavelength λ travels N cycles before it becomes π out of phase with the wave $\lambda + \Delta\lambda$, then we have:

$$\begin{aligned} l_c &= N\lambda \\ l_c &= (N - \frac{1}{2})(\lambda + \Delta\lambda) \end{aligned}$$

And hence $N = \lambda/(2\Delta\lambda)$, as $\lambda >> \Delta\lambda$. Thus the coherence length is:

$$l_c = \frac{\lambda^2}{2\Delta\lambda} \tag{2.32}$$

To achieve a suitable temporal coherence at the expenses of light brilliance, a series of monochromators can be added, with very narrow bandwidth.

2.3 The phase problem

The relationship between the sample and the diffracted wave is to a first approximation through the Fourier transform (Eq. 2.25). The act of measurement is only restricted to the intensity values $I(\vec{k}) = |f(\vec{s}, \vec{s}_0)|^2 = |F(\vec{k})|^2$. According to Emil Wolf, even if we were able to actually measure the phases, it is not possible to agree on a specific phase for a given intensity point, as no laboratory light source produces a strictly monochromatic wave [51]. Thus, the wavefield can only be measured statistically, where the phase information are washed out by the averaging process. To retrieve the sample's structure, we formulate the phase retrieval as an inverse optimization problem and discuss the algorithmic tools in the following chapters.

CHAPTER 3

Iterative Projection Algorithms

The role of a lens in imaging is to synthesize the analytical Fourier transform of the scattered waves and form an image. In CDI, computer algorithms take the lens' role by numerically reconstructing the structure from the scattered waves' amplitudes. In this chapter, we examine the phase problem in detail, in particular why it is considered a hard optimization problem §3.2. As a result, iterative projection algorithms, instead of direct methods, are used to compute a solution. The computational wavefront reconstruction using iterative methods was first introduced by Gerchberg and Saxton for the two intensity measurements [13], where the modulus of a complex valued image as well as the modulus of its Fourier transform are measured, as in electron microscopy and wavefront sensing; unlike the phase problem which is a one intensity measurement. Fienup showed that the Gerchberg and Saxton algorithm can be extended to the one intensity measurement [52]. In §3.2, 3.3, we review the conventional constraints and the current phase retrieval algorithms. To gain more insight into the most successful type of phase retrieval methods which employs a feedback mechanism, we study the performance of such method on a toy model §3.4.

3.1 Notation

Given a real density $\rho(\vec{x})$, where $\vec{x} \in \mathbf{R}^3$, we can represent ρ as an n-dimensional vector, i.e $\rho(\vec{x}) \in \mathbf{R}^n$. Where n is the number of pixels in a given structure.

Similarly, let $\sqrt{I(\vec{k})}$ be a measurement point, where $\vec{k}$ refers to the coordinates in reciprocal space; $I(\vec{k})$ spans an m-dimensional vector space, where m is the number of measured data points; $m = o_d \times n$ where o_d is the oversampling degree defined in the next chapter.

We also introduce the constraint set $\mathcal{C}_i$ and the corresponding projection operators $\mathbf{P}_i$. For example, if ρ is bounded, then $\mathcal{C}_i$ is the set of points that are within the given bounds. $\mathbf{P}_i$ is the operator that maps a general point $\rho^{(\nu)}$ to the constraint set by minimizing the Euclidean distance as:

$$\mathbf{P}_i \rho^{(\nu)} = \operatorname{argmin} |\rho^{(\nu)} - \mathcal{C}_i|. \tag{3.1}$$

ν is the iteration number.

Let ρ_e refer to the exact density and ρ refers to the solution obtained by phase retrieval.

3.2 Constraint sets, projections and error metrics

For a smooth, bounded and positive valued density $\rho(\vec{x})$, the forward model in CDI, is an elastic scattering process under the Born approximation, where the far field scattered amplitudes $\sqrt{I}$ or $|\hat{\rho}_e|$ are related through the Fourier transform $\mathcal{F}$ as [53] [41]:

$$\mathcal{F}\rho_e(\vec{x}) = \sqrt{I}(\vec{k}) \exp^{i\phi(\vec{k})} = \hat{\rho}_e(\vec{k}) \tag{3.2}$$

Since the sample size can be estimated *a priori*, the approximate support of ρ_e is usually assumed, where sampling above the Nyquist rate pads ρ_e with zeros. Solving for $\phi(\vec{k})$ is then formalized as a constrained optimization problem using the modulus and the support projection operators $\mathbf{P}_m, \mathbf{P}_{s+}$ [54]:

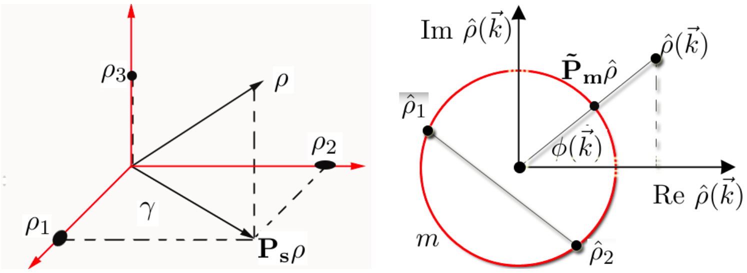

Figure 3.1: Low dimensional support set (left): $\rho(\vec{x}) \in \mathbf{R}^3$, where ρ is non zeros only in $\mathbf{R}^2$. Modulus Set for one complex point in reciprocal space (right)

The modulus projector: Each measurement is represented in the complex plane by a phase and a magnitude (Fig 3.1, right). The projection operator $\tilde{\mathbf{P}}_{\mathbf{m}}$ enforces the known magnitude, by minimizing the Euclidean distance between a given iterate and the modulus set:

$$\mathbf{P}_m[\rho] = \mathcal{F}^{-1}\tilde{\mathbf{P}}_m\mathcal{F}[\rho] = \mathcal{F}^{-1}\tilde{\mathbf{P}}_m\hat{\rho} \tag{3.3}$$

where $\tilde{\mathbf{P}}_m\hat{\rho} = \frac{\sqrt{I}}{|\hat{\rho}|}\hat{\rho}$. Assuming $|\hat{\rho}| \neq 0$, and hence avoiding the additional complication of a multivalued $\tilde{\mathbf{P}}_m$ [55].

The non-convexity of the modulus constraints creates an energy landscape with local minima. This is not obvious from the reciprocal space diagram (Fig. 3.1 right). Hence, to understand the problem more easily, we examine the effects of non-convexity in the object domain (Fig. 3.2). Each of the measurement points is a function of all the density points and forms an equation, namely the autocorrelation function:

$$I(\vec{k}) = \sum_{\vec{x},\vec{x}'} \rho(\vec{x})\rho(\vec{x}')\cos(\vec{k}\cdot(\vec{x}-\vec{x}')) \tag{3.4}$$

This equation can be visualized as an ellipse/hyperbola for the simplified three-pixel system, and as an ellipsoid for higher dimensions. Local minima are intersec-

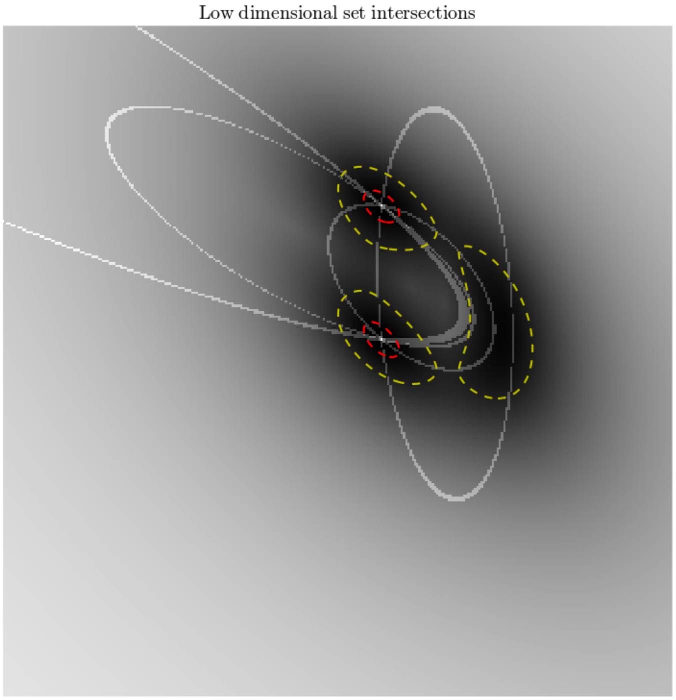

Figure 3.2: Set intersections. Each ellipse/parabola is a real space solution satisfying one equation 3.4. The global minimum and its twin correspond to the intersection among all sets (located in the red contour).

tions between sets; the global minimum is the intersection among all sets. Areas between intersections are high error regions. If a given density ρ belongs to these area, it does not satisfy any of the equations. Hence, the geometrical formation of local minima, in the energy landscape can be appreciated by studying the set intersections. From this simplified diagram (Fig.3.2), we can also deduce that the number of local minima increases with the size of the problem, and hence further enriching the complexity of the problem.

The support and positivity projector:

$$\mathbf{P}_{s+}\rho(x) = \begin{cases} \rho(x) & \text{for } x \in \gamma \\ 0 & \text{otherwise.} \end{cases} \tag{3.5}$$

Where γ denotes the set of points that satisfy the support and positivity constraint. Convergence is monitored through one of the scalar metrics: fidelity or compactness (positivity) terms:

$$\epsilon_m^2(\rho) = ||(\mathbf{I}-\mathbf{P}_m)\rho||^2 \qquad \epsilon_{s+}^2(\rho) = ||(\mathbf{I}-\mathbf{P}_{s+})\rho||^2 \tag{3.6}$$

$\mathbf{I}$ is the identity operator. Hereafter, to clear our notation, we will denote $\rho(x)$ by ρ_i.

The goal is to minimize the distance norm between a potential image ρ and the corresponding measurements $\mathbf{P}_m\rho$, i.e $\epsilon_m^2(\rho)$. The presence of oversampled data facilitates this step by providing a complementary real space constraint. The earliest phase retrieval algorithm, error reduction, follows the gradient:

$$-\nabla\epsilon_m^2(\rho) = -2(\mathbf{I}-\mathbf{P}_m)\rho \qquad -\nabla\epsilon_{s+}^2(\rho) = -2(\mathbf{I}-\mathbf{P}_{s+})\rho$$

However, due to the presence of the local minima, strictly minimizing the gradient is not ideal as the iterate would stagnate at the local minima. In the next section, we briefly discuss the error reduction method, then we review the feedback mechanism which provides a solution to the stagnation problem.

3.3 Feedback mechanisms

Phase retrieval algorithms had their initial success with the error reduction algorithm, which is a generalization of the Gerchberg Saxton algorithm. The error reduction algorithm proceeds in two steps, starting from a random density distribution $\rho^{(0)}$:

1. Apply the modulus constraint:

$$\hat{\rho}'^{(\nu)} = \mathbf{P}_m[\rho^{(\nu)}] \tag{3.7}$$

2. Apply the object domain constraint i.e a support projection P_{s+}:

$$\rho^{(\nu)} = \mathbf{P}_{s+}\rho'^{(\nu)} \tag{3.8}$$

This process is repeated till the error metric ϵ_m^2 stagnates. The error reduction algorithm is a steepest descent method; these genre of gradient algorithms have slow convergence and stagnate in local minima. There are two reasons for the stagnation: 1) the constant step size which is not optimized at each iteration, 2) the reduced partial gradient ∇_ρ along P_{s^+} [54]. The feedback mechanism alleviates the stagnation and convergence problem, as shown in the following sections.

3.3.1 Hybrid Input Output (HIO)

Before we explain the HIO algorithm, let us detail the relationship of the error metric gradient $\nabla\epsilon_m^2$ and the projection operator $\mathbf{P}_m$ following Fienup's [52]. Assume a complete measurement in reciprocal space, the scalar error metric ϵ_m can be computed as:

$$\epsilon_m^2 = N^{-2}\sum_{k=0}^{N}|\hat{\rho}_k - \tilde{\mathbf{P}}_m\hat{\rho}_k|^2 = N^{-2}\sum_{k=0}^{N}[|\hat{\rho}_k| - \sqrt{I_k}]^2 \tag{3.9}$$

Where N is the size of the full dimensional Fourier transform space, or the oversampled region. Note that $\hat{\rho}_k$ and $\tilde{\mathbf{P}}_m\hat{\rho}_k$ have the same phases, hence we needed to compute only the difference in their moduli. The corresponding gradient term is:

$$\nabla \epsilon_m^2 = \frac{\partial \epsilon_m^2}{\partial \vec{\rho}} = 2N^{-2} \sum_{k=0}^{N} [|\hat{\rho}_k| - \sqrt{I_k}] \frac{\partial |\hat{\rho}_k|}{\partial \vec{\rho}} \tag{3.10}$$

From $\hat{\rho}_k = \sum_{j=0}^{N} \rho_j \exp^{-2\pi i \frac{kj}{N}}$, we have $\frac{\partial \hat{\rho}_k}{\partial \rho_j} = \exp^{-2\pi i \frac{kj}{N}}$. Thus now we can compute $\frac{\partial |\hat{\rho}_k|}{\partial \vec{\rho}}$:

$$\begin{aligned} \frac{\partial |\hat{\rho}_k|}{\partial \rho_j} &= \frac{\partial [|\hat{\rho}_k|^2]^{1/2}}{\partial \rho_j} = \frac{\partial [\hat{\rho}_k \hat{\rho}_k^\star]^{1/2}}{\partial \rho_j} = \frac{1}{2|\hat{\rho}_k|} [\hat{\rho}_k^\star \frac{\partial \hat{\rho}_k}{\partial \rho_j} + \hat{\rho}_k \frac{\partial \hat{\rho}_k^\star}{\partial \rho_j}] \\ &= \frac{1}{2|\hat{\rho}_k|} [\hat{\rho}_k^\star \exp^{-2\pi i \frac{kj}{N}} + \hat{\rho}_k \exp^{2\pi i \frac{kj}{N}}] \end{aligned}$$

Now, let us insert this information back into the gradient equation, by computing the element wise terms:

$$\begin{aligned} \frac{\partial \epsilon_m^2}{\partial \rho_j} &= 2N^{-2} \sum_{k=0}^{N} [|\hat{\rho}_k| - \sqrt{I_k}] \frac{\partial |\hat{\rho}_k|}{\partial \rho_j} \\ &= 2N^{-2} \sum_{k=0}^{N} [|\hat{\rho}_k| - \sqrt{I_k}] \frac{1}{2|\hat{\rho}_k|} [\hat{\rho}_k^\star \exp^{-2\pi i \frac{kj}{N}} + \hat{\rho}_k \exp^{2\pi i \frac{kj}{N}}] \\ &= N^{-2} \sum_{k=0}^{N} [\hat{\rho}_k^\star \exp^{-2\pi i \frac{kj}{N}} + \hat{\rho}_k \exp^{2\pi i \frac{kj}{N}} - \frac{\sqrt{I_k} \hat{\rho}_k^\star}{|\hat{\rho}_k|} \exp^{-2\pi i \frac{kj}{N}} - \frac{\sqrt{I_k} \hat{\rho}_k}{|\hat{\rho}_k|} \exp^{2\pi i \frac{kj}{N}}] \end{aligned}$$

Also note that we are dealing with real functions, hence:

$$\rho_j^\star = N^{-1} \sum_{k=0}^{N} \hat{\rho}_k^\star \exp^{-2\pi i \frac{kj}{N}} = \rho_j = N^{-1} \sum_{k=0}^{N} \hat{\rho}_k \exp^{2\pi i \frac{kj}{N}} \tag{3.11}$$

Which leads us to conclude:

$$\frac{\partial \epsilon_m^2}{\partial \rho_j} = 2[\rho_j - \rho'_j] \tag{3.12}$$

In the steepest descent method, a half step in the opposite gradient direction is implemented, hence:

$$\rho^{(\nu+1)} - \rho^{(\nu)} = -\frac{\nabla \epsilon_m^2}{2} = [\rho' - \rho^{(\nu)}] = [\mathbf{P}_m - \mathbf{I}]\rho^{(\nu)} \tag{3.13}$$

$$\rho^{(\nu+1)} = \rho' \tag{3.14}$$

To speed up the convergence, Fienup devised the Hybrid input output algorithm as a form of nonlinear feedback. As a result, great flexibility in the iterative system is introduced by relaxing the object domain constraint and forming the iterate $\rho^{(\nu)}$ unconfined to a subspace but as linear combination. To examine the details of this step, we follow [56]'s analysis:

A small change in the input ($\Delta\rho$) results in a change of the output ($\Delta\rho'$) in the same direction:

$$\Delta\rho = \beta\Delta\rho' \tag{3.15}$$

Hence for pixels $i \notin \gamma$, one needs to modify the input such that the next output $\rho'^{(\nu+1)}$ would be driven to zero:

$$\Delta\rho' = \cancelto{0}{\rho'^{(\nu+1)}} - \rho'^{(\nu)} = -\rho'^{(\nu)} \tag{3.16}$$

Thus,

$$\Delta\rho = \rho^{(\nu+1)} - \rho^{(\nu)} = \beta\Delta\rho' = -\beta\rho'^{(\nu)} \tag{3.17}$$

The flexibility in choosing an input that drives to a more desirable output, gave the algorithm its name, where both of $\rho(x)$ and $\rho'(x)$ are used at each iteration. Combining Eq. 3.13 and Eq. 3.17, we find:

$$\rho^{(\nu+1)}(x) = \begin{cases} \rho'^{(\nu)}(x) & x \in \gamma \\ \rho^{(\nu)}(x) - \beta\rho'^{(\nu)}(x) & x \notin \gamma \end{cases} \tag{3.18}$$

Where β is a relaxation parameter.

To understand HIO's empirical sucess, we examine the work of Marchesini where he expressed HIO as a saddle point operations §3.3.2 and the work of Elser where he showed that the difference maps are a generalization of HIO §3.3.3.

3.3.2 Saddle point optimization

Here we follow Marchesini's derivation expressing HIO as a function of the Lagrangian of the two error metrics [54]. Using Eq. 3.6, we define the Lagrangian as the difference between the two square metrics:

$$\mathcal{L} = \epsilon_m^2 - \epsilon_s^2 \tag{3.19}$$

The Gradient expression is:

$$\begin{aligned} \nabla\mathcal{L} &= \nabla\epsilon_m^2 - \nabla\epsilon_s^2 \\ &= 2[\mathbf{P}_s - \mathbf{P}_m]\rho \end{aligned} \tag{3.20}$$

Let us define the following projection:

$$\underline{\mathbf{P}_s} = \mathbf{I} - \mathbf{P}_s \tag{3.21}$$

Now we are ready to express HIO in terms of $\mathcal{L}$:

$$\begin{aligned}
\rho^{(\nu+1)} &= \underline{\mathbf{P}_s}\rho^{(\nu)} - \beta\underline{\mathbf{P}_s}\mathbf{P}_m\rho^{(\nu)} + \mathbf{P}_s\mathbf{P}_m\rho^{(\nu)} \\
&= [\mathbf{I} - \mathbf{P}_s]\rho^{(\nu)} - \beta\underline{\mathbf{P}_s}\mathbf{P}_m\rho^{(\nu)} + \mathbf{P}_s\mathbf{P}_m\rho^{(\nu)} \\
&= \rho^{(\nu)} + \mathbf{P}_s[\mathbf{P}_m - I]\rho^{(\nu)} - \beta\underline{\mathbf{P}_s}\mathbf{P}_m\rho^{(\nu)} \\
&= \rho^{(\nu)} + \mathbf{P}_s[\mathbf{P}_m - \mathbf{P}_s]\rho^{(\nu)} - \beta\underline{\mathbf{P}_s}[\mathbf{P}_m - \mathbf{P}_s]\rho^{(\nu)} \quad \text{as} \quad \mathbf{P}_s\mathbf{P}_s = \mathbf{P}_s \quad \text{and} \quad \underline{\mathbf{P}_s}\mathbf{P}_s = 0 \\
&= \rho^{(\nu)} + [\mathbf{P}_s - \beta\underline{\mathbf{P}_s}][\mathbf{P}_m - \mathbf{P}_s]\rho^{(\nu)} \\
&= \rho^{(\nu)} + \frac{1}{2}[\mathbf{P}_s - \beta\underline{\mathbf{P}_s}][\nabla\mathcal{L}]\rho^{(\nu)} \qquad (3.22)
\end{aligned}$$

Hence, by minimization the subspace of object constraint and maximization the complementary subspace, HIO is recast as a saddle point optimization. HIO's ability to escape local minima is conceptually understood by this minmax approach.

3.3.3 Difference map

Veit Elser showed that HIO belongs to a generalized map, called the difference map D, where:

$$D = 1 + \beta\Delta \qquad (3.23)$$

$$\Delta = \mathbf{P_i} \circ f_j - \mathbf{P_j} \circ f_i$$

$$f_i(\rho) = (1 + \gamma_i)\mathbf{P_i}(\rho) - \gamma_i\rho$$

Where γ, β are relaxation parameters. The form of f_i, which is a generalized line passing through the two points $(1 + \gamma_i)\mathbf{P_i}(\rho) - \gamma_i\rho$ and $\gamma_i\rho$ and parametrized by γ_i, enhances convergence to a fixed point $\rho^\star$ such that:

$$(\mathbf{P_1} \circ f_2)\rho^\star = (\mathbf{P_2} \circ f_1)\rho^\star \tag{3.24}$$

Note, however, a fixed point is not necessarily a global solution due to the non-convexity of the $\mathbf{P_m}$ constraint, and hence Eq. 3.3 does not necessarily lead to a solution:

$$\mathbf{P_1}\rho_{sol} \neq \mathbf{P_2}\rho_{sol} \tag{3.25}$$

On the other hand, a solution is a fixed point. But that is not an insightful statement, as once the solution is found, we anticipate it to be a unique fixed point.

For HIO, we have the following parameters:

$$\begin{aligned} \mathbf{P}_1 &= \mathbf{P}_s \qquad \mathbf{P}_2 = \mathbf{P}_m \\ \gamma_1 &= -1 \qquad \gamma_2 = \beta^{-1} \end{aligned} \tag{3.26}$$

Locally $\mathbf{P}_s, \mathbf{P}_m$ can be considered orthogonal, and the choice of γ is more optimized if [57]:

$$\gamma_1 = \beta^{-1} \qquad \gamma_2 = -\beta^{-1} \tag{3.27}$$

Here also, the main characteristic of the difference map, is its ability to escape local minima, as each iterate is not a member of a particular subspace defined by the constraint, but a combination of the respective constraints.

3.3.4 RAAR

Luke devised the Relaxed Averaged Alternating Reflections (RAAR) to eliminate one disadvantage of HIO [58]: as the iterate becomes closer to the solution, HIO map occasionally wanders away with continuing iterations while RAAR anchors the solution back to the nearest optimal point. The map $\mathcal{V}$ can be expressed as follows:

$$\begin{aligned}\mathcal{V}(\tau_\star, \beta) &= \beta\tau_\star + (1-\beta)\mathbf{P_m} \\ \tau_\star &= \frac{1}{2}(\mathbf{R}_{S_+}\mathbf{R}_M + \mathbf{I}) \\ \mathbf{R}_C &= 2\mathbf{P}_C - \mathbf{I} \end{aligned} \tag{3.28}$$

For $\beta = 1$, RAAR, HIO and D, all yield the same sequence. The optimal β used in RAAR, is a variable β_n that approaches unity, as the iterate becomes closer to the solution.

3.4 Insight

To further understand the behavior of difference map algorithms in general and HIO in particular, we examine the iterative process in the case of two simple constraint sets. A more accurate representation with constraints that correspond to the phase retrieval problem is presented in §6.2.1. Nonetheless, the simplicity of the current discussion provides a deep insight into the dynamics of these iterative maps, as explained by Rankenburg and Elser [59]. In analogy with the support and modulus constraints, we define:

1. $\mathcal{S}_g$ is a convex set with an associated geometric projection operator $\mathbf{P_g}$
2. $\mathcal{S}_e$ is a non-convex set constraint with an associated energy projection op-

erator $\mathbf{P_e}$

With $\beta = 1$, in Eq 3.23, the difference map can be expressed as follows:

$$\Delta = d_g - d_e$$
$$d_e = \mathbf{P_e}\rho \quad \text{and} \quad d_g = \mathbf{P_g}(2\mathbf{P_e} - \mathbf{I})\rho$$

d_g is the projection of the mirror reflection of ρ across the boundary $\mathcal{S}_e$ and onto $\mathcal{S}_g$.

3.4.1 Local minima, fixed points and the global minimum

The difference map iterate $\rho^{(\nu+1)}$ does not satisfy either constraints and is defined as follows:

$$\begin{aligned} \rho^{(\nu+1)} &= \rho^{(\nu)} + \Delta \\ &= \rho^{(\nu)} + d_g - d_e \\ &= (\mathbf{I} + \mathbf{P_g}(2\mathbf{P_e} - \mathbf{I}) - \mathbf{P_e})\rho^{(\nu)} \end{aligned}$$

At a local minima, the iterate moves steadily away from the two constraints with a constant step $\Delta = \rho^{(\nu+1)} - \rho^{(\nu)} = \min |\mathcal{S}_e - \mathcal{S}_g|$ as shown in figure 3.3. Note the alternation between d_e, d_g in table 3.1, where $d_g\rho^{(\nu)} = d_e\rho^{(\nu+1)}$ from 2 → 10, as the iterate hovers above a local minimum. Instead of stagnating at a local minimum, the difference map iterate is repelled monotonically, with a step Δ. A stagnation point occurs when $\mathbf{P_g}(2\mathbf{P_e} - \mathbf{I})\rho^{(\nu)} = \mathbf{P_e}\rho^{(\nu+1)}$, as depicted in figure 3.4, where the distance between the white squares corresponds to $\mathbf{P_g}(2\mathbf{P_e} - \mathbf{I})\rho^{(\nu)}$, and the distance between the yellow dots corresponds to $\mathbf{P_e}\rho^{(\nu+1)}$.

Figure 3.5 shows that for larger distances between the sets, the algorithm converges faster, as the step size Δ increases.

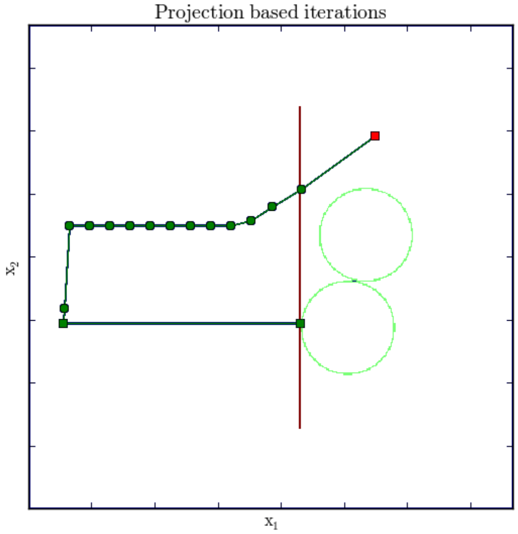

Figure 3.3: Difference Map with a successful path: $\mathcal{S}_e$ is the non-convex double circular surfaces. $\mathcal{S}_g$ is the one dimensional vertical line. Red dot is the starting point. Green dots traces the difference map iterates, which are repelled by the local minimum, and finally fall into a fixed point, which is a point satisfying the following equation:
$\mathbf{P_g}(2\mathbf{P_e} - \mathbf{I})\rho^\star = \mathbf{P_e}\rho^\star$.

n	d_e	d_g	$\lvert d_g - d_e \rvert$
1	30	45	15
2	42	57	15
3	57	73	16
4	73	89	16
5	89	105	16
6	105	121	16
7	121	137	16
8	137	153	16
9	153	169	16
10	169	185	16
11	184	177	7
12	174	175	1
13	175	176	1
Last	$\mathbf{P_e}\rho^\star$		$\lvert d_g - d_e \rvert$
14	176	176	0

Table 3.1: The difference map iterates. A local solution corresponds to $d_g\rho^{(\nu)} = d_e\rho^{(\nu+1)}$, as can be seen in iterations $2 \rightarrow 10$. In this case the step Δ between iterates is equal to the distance between the sets, and thus $|d_g - d_e| = const \neq 0$. A fixed point is reach when $|d_g - d_e| \rightarrow 0$. A fixed point is one step away from the solution, namely
$\rho^{(sol)} = \mathbf{P_g}\rho^\star = \mathbf{P_e}\rho^\star$

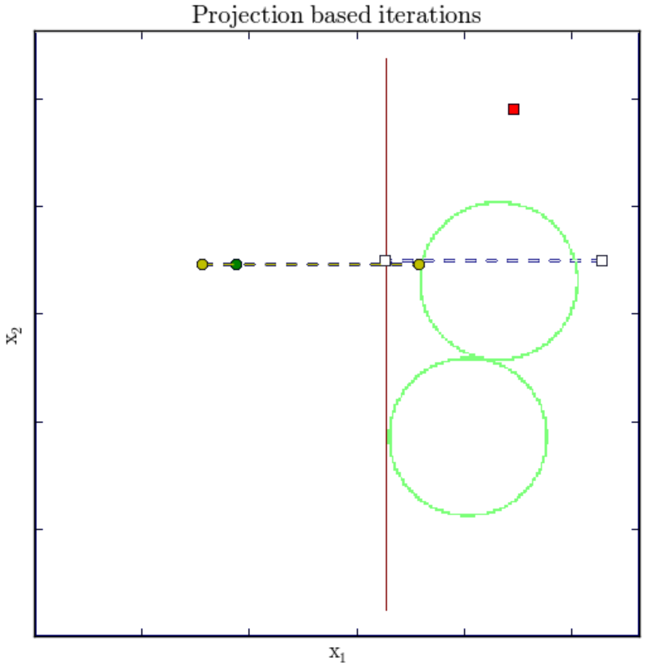

Figure 3.4: A stagnation point occurs when $\mathbf{P_g}(2\mathbf{P_e} - \mathbf{I})\rho^{(\nu)} = \mathbf{P_e}\rho^{(\nu+1)}$. For two consecutive iterates $\rho^{(\nu)}, \rho^{(\nu+1)}$, depicted by the neighboring green and yellow dots, the white square on the far right is $(2\mathbf{P_e} - \mathbf{I})\rho^{(\nu)}$. $\mathbf{P_g}(2\mathbf{P_e} - \mathbf{I})\rho^{(\nu)}$ is the distance between the two white squares. $\mathbf{P_e}\rho^{(\nu+1)}$ is the distance between the two yellow dots. Hence a stagnation occurs when the distance between the two white squares is equal to the distance between the two yellow dots; in other words, when the projection of the mirror image $(2\mathbf{P_e} - \mathbf{I})\rho^{(\nu)}$ on $\mathcal{S}_g$ is equal to the projection of $\rho^{(\nu+1)}$ on $\mathcal{S}_e$. For this case, the step between iterates Δ (yellow and green circles) is equal to the separation between the constraints: $\min|\mathcal{S}_e - \mathcal{S}_g|$.

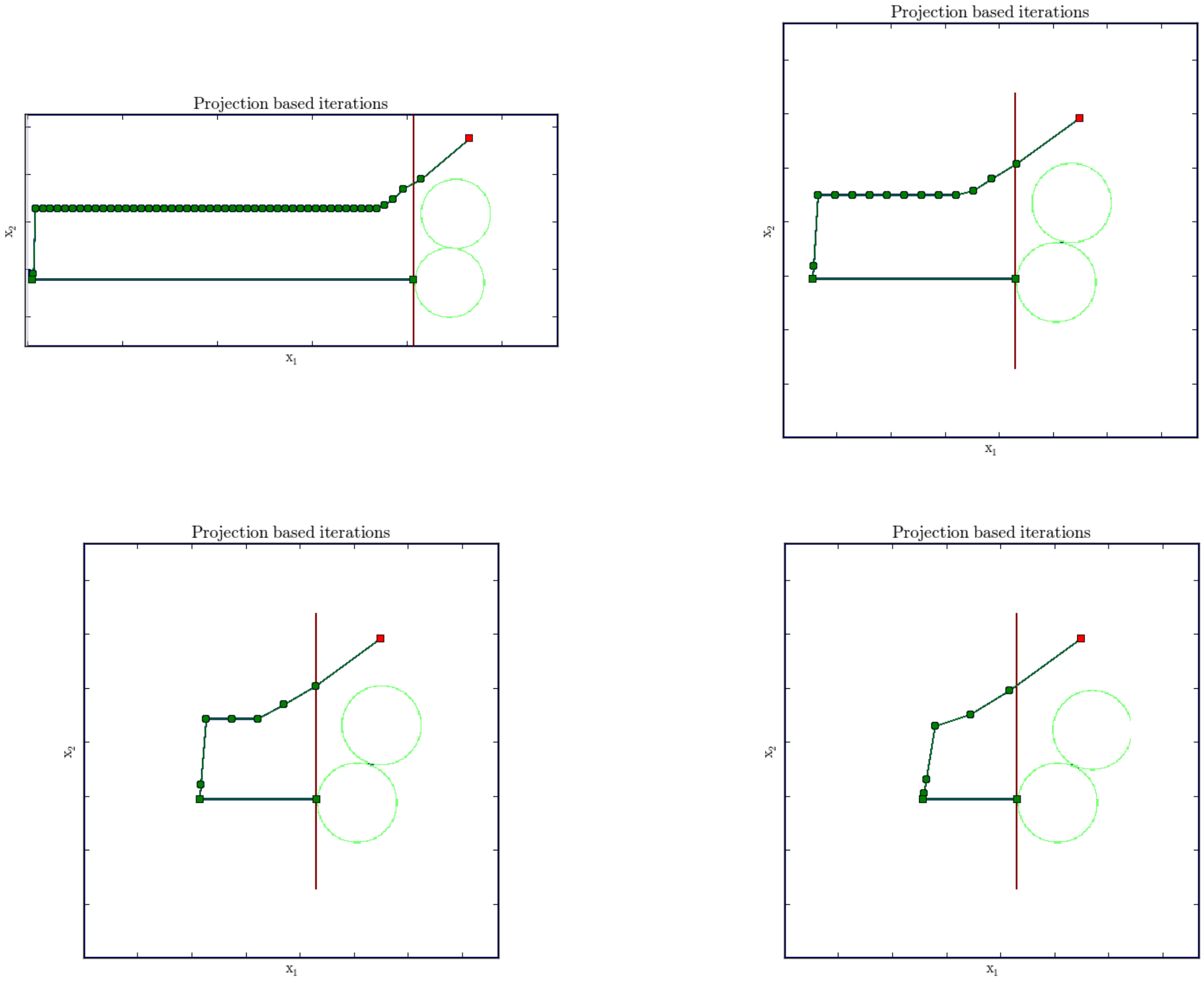

Figure 3.5: Clockwise, starting top left: as the local minimum depth increases, the number of iterations required for convergence decreases.

Fixed point and global minimum: The mechanism of escaping a local minimum, relies on the proximity of the non-convex surfaces. In figure 3.6, the iterate $\rho^{(\nu-1)}$ represented by the blue dot is closer to the upper circle, while $\rho^{(\nu)}$ is closer to the lower circle. The global solution is at the lower surface, and hence, the next step projects the iterate to a fixed point, where $\Delta \to 0$. With $\Delta = 0$, the iterative process stagnates. In this simple example, the global solution is one step away from the fixed point, and hence we have:

$$\rho^{(sol)} = \mathbf{P_g}\rho^{\star} = \mathbf{P_e}\rho^{\star} \tag{3.29}$$

3.4.2 When difference map fails

Fig 3.7 and table 3.2 show a case where the difference map's iterate fails to emerge out of stagnation and oscillates between three positions. This particular arrangement of the constraint surfaces is analogous to a standing wave. Figures 3.8, 3.9 and 3.10 show how the geometrical distribution of the constraint sets affects the ability of the iterate to escape stagnation points. The green line is the difference map trajectory, while the yellow line is the trajectory after the addition of a prior knowledge constraint. The prior constraint provides enough perturbations to alleviate stagnation. Figures 3.8 and 3.9 show more cases where the difference map algorithm fails to emerge out of stagnation which is due the constraint boundaries: at the verge of escaping a local minimum, the third surface constraint perpetually reflects the iterate to a previous stagnation point. The boundaries in figure 3.10 are less restrictive, and hence an escape out of the local minimum was possible. In conclusion, even though difference map algorithm performs remarkably well by escaping local minima, in certain conditions, the boundary constraints form resonance cavity as in fig 3.8, 3.9. It is for these cases that we introduce the adaptive phase retrieval (APR) algorithm in chapter 8 as an improved map that breaks

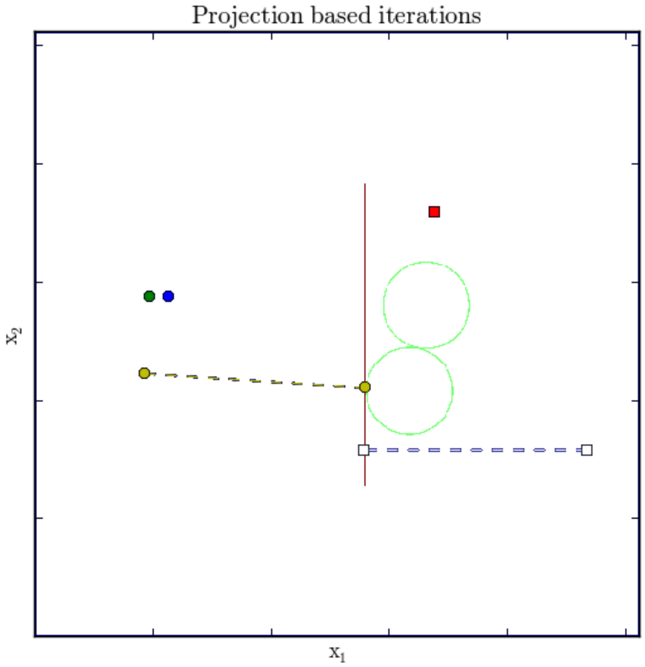

Figure 3.6: How does the difference map escape local minima? This figure encapsulates the answer. At a local minimum, the difference map algorithm is not passive; at each iteration, ρ inches away from the constraints with a step Δ. The iterates $\rho^{(\nu-1)}, \rho^{(\nu)}$ are represented by the blue/green dots. When the lower circle in $\mathcal{S}_e$ becomes the closer projection, $(2\mathbf{P_e} - \mathbf{I})\rho^{(\nu)}$ gets projected down (the far right white square, corresponds to the reflection of the green dot across the lower circle). At this iteration, the step Δ corresponds to $(\mathbf{P_g}(2\mathbf{P_e} - \mathbf{I}) - \mathbf{P_e})\rho^{(\nu)}$, which is the distance between the yellow dot and the white square, also equal to the distance between the two subsequent iterates (green and yellow circles). The yellow dot is a fixed point $\rho^{\star}$, as the distance between the sets at this location is 0. At this point, we simply project the iterate $\rho^{\star}$ to find the global solution ρ^{sol}.

such stagnation, by approaching an approximate solution.

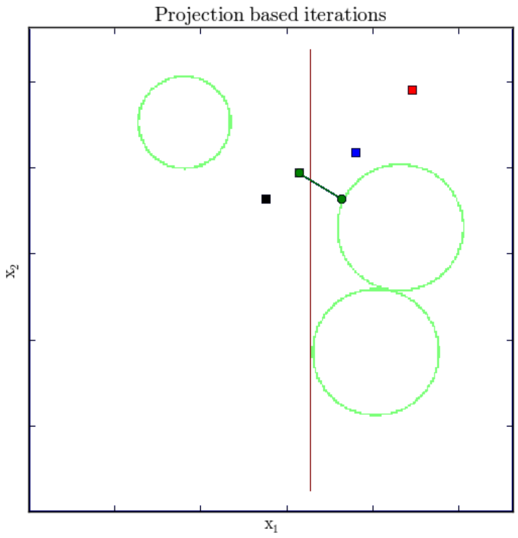

Figure 3.7: The difference map fails when it enters into endless oscillations (oscillations among blue, black squares and the green dot. Red square is the starting location.). Intuitively, as the iterate emerges from a given stagnation, it is reflected back into stagnation by another constraint surface.

n	d_e	d_g	$\lvert d_e - d_g\rvert$
0	43	56	13
1	25	43	17
2	38	54	16
3	34	75	41
4	15	39	24
5	31	48	17
6	35	79	44
7	13	41	28
8	31	46	15
9	37	79	42
10	13	40	27
11	30	45	15
12	37	80	43
13	13	39	26
14	29	44	15
15	38	79	41
16	13	39	26
$\vdots$	$\vdots$	$\vdots$	$\vdots$
296	29	44	15
297	38	79	41
298	13	39	26
Last	$\mathbf{P_e}\rho^\star$		$\lvert d_e - d_g\rvert$
299	29	44	15

Table 3.2: A three point oscillations that yields a deep stagnation. The pattern observed from $n = 14 \to 16$, repeats for endless iterations. The oscillations are between three distances $15, 41, 26$.

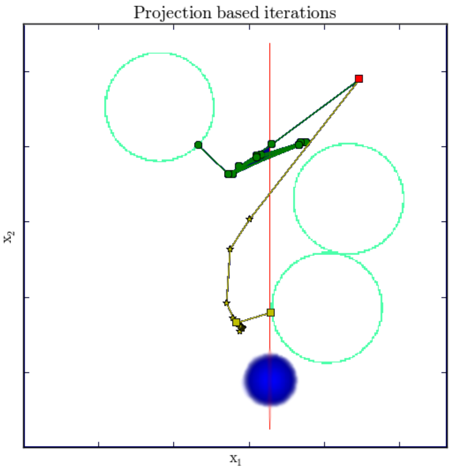

Figure 3.8: The geometric location of the third non convex surface affects the oscillations or the standing wave. A three point oscillations that yields a deep stagnation, where the difference map is unable to emerge successfully (green trace). By adding prior knowledge (blue Guassian), the map is perturbed enough to emerge out of stagnation (yellow trace).

n	d_e	d_g	$\|d_e - d_g\|$
0	43	56	13
1	25	42	17
2	30	70	40
3	9	35	26
4	27	43	16
5	30	66	36
6	12	37	25
7	30	46	16
8	28	67	39
9	11	36	25
10	29	46	17
11	28	67	39
12	11	36	25
13	29	46	17
Last	$\mathbf{P_e}\rho^\star$		$\|d_e - d_g\|$
14	28	46	18

Table 3.3: Here we have $39, 25, 17$ as the repeating pattern.

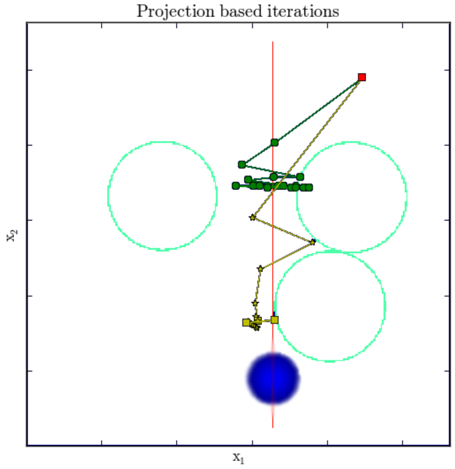

Figure 3.9: A two point oscillations due to the symmetry between the constraint sets. The difference map is unable to emerge successfully. With the addition of prior knowledge (blue Gaussian), the iterates (yellow line) emerge out of stagnation.

n	d_e	d_g	$\|d_e - d_g\|$
0	43	56	13
1	25	42	17
2	21	58	37
3	1	17	16
4	16	33	17
5	22	58	36
6	4	13	9
7	12	28	16
8	25	62	37
9	8	9	1
Last	$\mathbf{P_e}\rho^\star$		$\|d_e - d_g\|$
10	8	9	1

Table 3.4: As the constraint surfaces are re-aranged, we obtain a double standing wave at $16, 37$ as the repeating pattern.

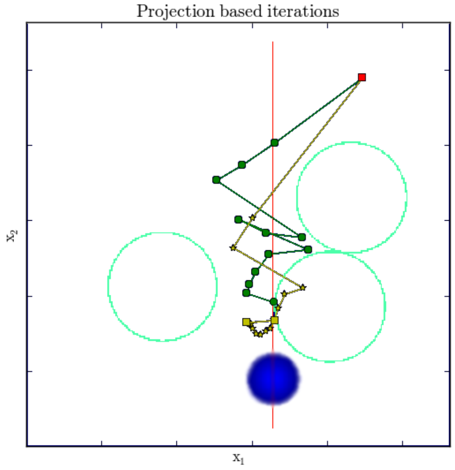

Figure 3.10: In this case, both the difference map and the map with prior knowledge are successful.

n	d_e	d_g	$\|d_g - d_e\|$
0	43	56	13
1	25	42	17
2	38	54	16
3	43	77	34
4	5	28	23
5	24	40	16
6	31	70	39
7	4	28	24
8	16	20	4
9	17	19	2
10	18	19	1
Last	$\mathbf{P_e}\rho^\star$		$\|d_g - d_e\|$
11	19	19	0

Table 3.5: A chaotic tumbling towards the global minimum.

CHAPTER 4

Ankylography

Ankylography is a coherent diffraction imaging modality whose inception was inspired by high numerical aperture data. At high numerical aperture, the measurements have a natural curvature which is a direct result of the elastic scattering phenomenon §4.1; in other words, the data is written (*graphy*) on a curved surface (*ankylos*). Ankylography requires fewer tilt angles, in comparison to conventional CDI data, as the informational content in the acquired measurements is higher, due to the non-linear distribution of information in Fourier space. The initial goal of Ankylography, was three dimensional structures determination *from a single diffraction pattern*, but was later extended to multiple patterns for cases of insufficient information [60]. Single shot imaging, however, is still an active goal which is possible using additional constraints, as we explore in chapter 8. Ankylography was first demonstrated using soft x-ray table top source [35] and optical laser [36]. In §4.2, we examine the resolution gained from such method; in §4.3, we detail the required data analysis steps and constraints. In §4.4, we demonstrate ankylography on simulated objects: a glass structure and a polio virus; in the next chapter we discuss the experimental results.

4.1 Elastic scattering

For a weakly scattering semi transparent material, and under coherent illumination with wavelength λ, the measured amplitudes $|F(\vec{k})|$ are related to the density of the material by the Fourier Transform.

$$|F(\vec{k})| = |\int \rho(\vec{r}) e^{-i\vec{k}\cdot\vec{r}} d^3\vec{r}| \tag{4.1}$$

$\vec{r}$ spans the density coordinates, also known as real space. In the measurement space (reciprocal space), $\vec{k}$ lies on a spherical surface, as for elastic scattering the magnitude $|\vec{k}|$ is preserved.

4.2 Resolution

Following the geometry in figure 4.1, we compute the transverse resolution as:

$$dx = \frac{\lambda}{\sin\theta_{max}} \tag{4.2}$$

The longitudinal resolution can be found as:

$$\sin\frac{\theta_{max}}{2} = \frac{\frac{l}{2}}{\frac{1}{\lambda}} \qquad \text{where } l^2 = \frac{1}{dz^2} + \frac{1}{dx^2}$$

$$dz = \frac{\lambda}{2\sin^2\frac{\theta}{2}}$$

At large scattering angles, where $\max(k_z) \approx \max(k_{xy})$, the geometrical distribution of the data along the Ewald sphere carries more information about the structure than the corresponding flat panel i.e where $k_z \approx 0$.

4.3 Data analysis and constraints

At the present time, all experimental coherent diffraction data are collected on a planar detector. To account for the scattering geometry, we perform two step analysis on the data: flux normalization and planar to spherical interpolation.

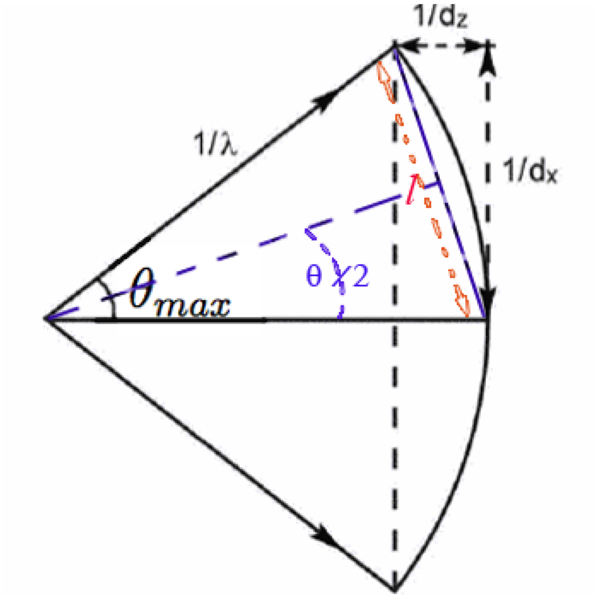

Figure 4.1: High numerical aperture diffraction geometry. (dx, dz) correspond to the transverse and longitudinal resolutions. θ_{max} is the half-angle maximum scattering. λ is the wavelength of the coherent beam. The curvature of the Ewald sphere is proportional to $\frac{1}{\lambda}$.

4.3.1 Flux normalization

For a detector with spherical geometry, the solid angles per measurement pixel are equivalent. However, for the case of planar detector, the solid angle per pixel decreases with higher scattering angle. To account for this discrepancy, the intensity measurements I_M are first normalized to I_N as follows:

$$I_N(k_x^d, k_y^d) = \frac{\Delta\Omega(0,0)}{\Delta\Omega(k_x^d, k_y^d)} I_M(k_x^d, k_y^d) \tag{4.3}$$

[61]

k_x^d, k_y^d are unitless integers representing the Cartesian coordinates on the detector. $\Delta\Omega(0,0), \Delta\Omega(k_x^d, k_y^d)$ are the solid angles formed at the central pixel and a general pixel k_x^d, k_y^d, respectively (Figure 4.2). $\Delta\Omega$ can be computed by integrating the corresponding surface area as:

$$\Delta\Omega(k_x^d, k_y^d) = R \int_{k_x^d-1/2}^{k_x^d+1/2} \int_{k_y^d-1/2}^{k_y^d+1/2} \frac{dk_x dk_y}{(k_x^2 + k_y^2 + R^2)^{3/2}} \tag{4.4}$$

R is a unitless measure of the distance between the scatterer and the detector, which can be computed as follows:

$$\sin\theta = \frac{n\Delta p}{L} = \frac{n}{R} \tag{4.5}$$

and Δp is the physical size of a CCD pixel, L is the distance from the detector and n is the number of pixels measured at θ angle.

4.3.2 Interpolation

After projecting the equiangular spherical coordinates k_x^c, k_y^c, k_z^c onto the planar detector, we obtain a new set of points $k_x^{d'}, k_y^{d'}$:

$$k_x^{d'} = \frac{Rk_x^c}{R - k_z^c} \qquad k_y^{d'} = \frac{Rk_y^c}{R - k_z^c} \tag{4.6}$$

To find the intensity at these locations, we conduct a spline interpolation from $I_N(k_x^d, k_y^d)$. With the normalized intensity, the measurement at $I_N(k_x^{d'}, k_y^{d'})$ corresponds to $I_N(k_x^c, k_y^c, k_z^c)$.

To prepare the data for three dimensional reconstruction, the spherical pattern is embedded into the corresponding $3D$ array. For a real electron density, where x-ray absorption is negligible, we can also include the centro-symmetric pattern. Through the phase retrieval process, only points on the spherical patterns are enforced, the rest of the points in reciprocal space are treated as unknown and updated at each iteration.

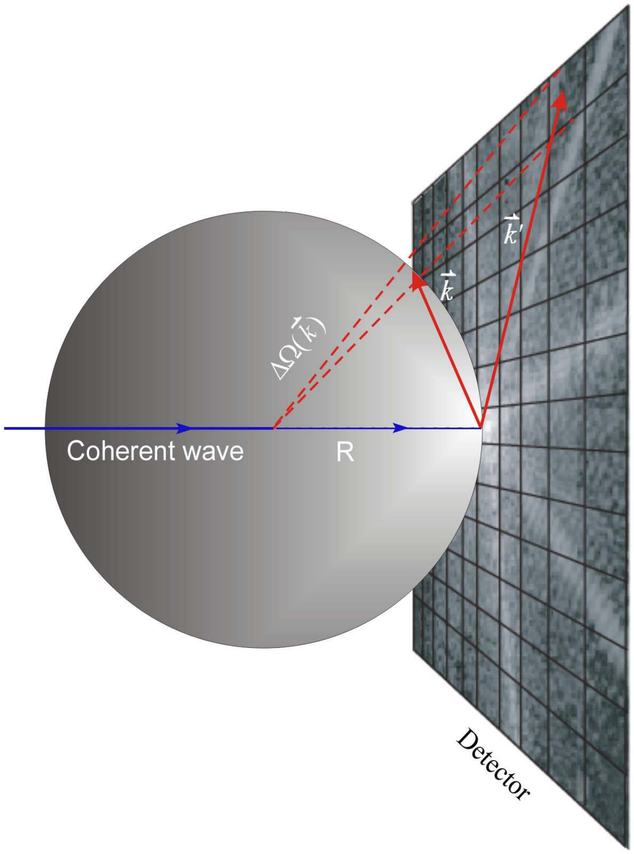

Figure 4.2: A sketch for a high numerical aperture experimental setup, using a planar detector. $\vec{k}'$ represents a vector in the detector plane. $\vec{k}$ is the corresponding location on the Ewald sphere.

4.3.3 Constraints

For a single diffraction pattern, the data is incomplete; as a result, this is an ill posed problem. Such problems are hard to solve, unless we incorporate additional information. Here we introduce two general constraints: smoothness and phase extension.

Smoothness constraint is a form of regularization, which can be implemented inside and outside the support region. In the outside support region, we apply the uniformity constraint, using a box filter, in conformity with the prior knowledge of zero scattering. Inside the density region, we apply the continuity constraint, using a Gaussian filter. Recently, this constraint was shown to outperform various forms of denoising strategies used in phase retrieval [62].

Amplitude extension is a method that gradually incorporates information from reconstructions obtained at lower resolutions. Phase retrieval with small sized data ($\leq 17^3$), quickly converges. To obtain lower resolution reconstructions, we truncate the measurement to a suitable resolution, by excluding high angular measurements. After obtaining such a small sized reconstruction, we use the corresponding Fourier transform amplitudes as constraints, and hence the higher resolution reconstruction will be more constrained by the data. By incrementally increasing the resolution of the structure, we minimize the amount of missing data, and as a result the phase retrieval process becomes easier for larger arrays.

4.4 Simulations

To demonstrate the potential of single shot imaging using ankylography, we simulated the experiment in two cases: using a material science and biological samples. In the following subsections, we detail the simulations and the reconstructions.

4.4.1 Glass structure

The peculiar characteristic of the glass structure is its long range aperiodic form. Although a solid material, the glass structure resembles a liquid arrangement, where the structural order is correlated at long scales prohibiting structure determination with crystallographic techniques.

The glass model used in this experiment is $25Na_2O$–$75SiO_2$, (figure 4.3). These sodium silicate glasses are the most typical glass form used in a variety of applications, from glasswares to photonics devices. Visualizing these types of structures at atomic resolution will provide an insight into the mysterious formation of these long range molecular morphologies, and hence enable better understanding of their formation and subsequently improve their usage.

The $3D$ glass structure was simulated by computationally annealing the glass using molecular dynamics. The final geometry of the core and valence electrons was obtained by optimization using *ab-initio* density functional theory calculations. The molecular dynamics simulations include a series of melt and quench with partial charge potentials as described in [63]. The sample used contains 204 atoms, with volume of $14 \times 14 \times 14$ $\mathring{A}^3$ and density of 2.43 g/cc. A simulated incident beam generated by an Energy Recovery Linac, with $10^{14}/s$ coherent photons and a 2 $\mathring{A}$ wavelength, focused down to a 100 nm spot [64], scattered off the glass particle. Data was assumed to be collected on a spherical detector, where the signal is acquired up to $2\theta = 90°$, with an exposure time of 5.1 s as shown in Fig 4.3. Poisson noise was added to the diffraction pattern which is measured on a shell embedded in a volume of $64 \times 64 \times 64$, corresponding to an oversampling degree $O_d = 2.7$.

Phase recovery is carried through the iterative projection algorithm HIO. The initial density estimate is a random distribution. The constraints used are the

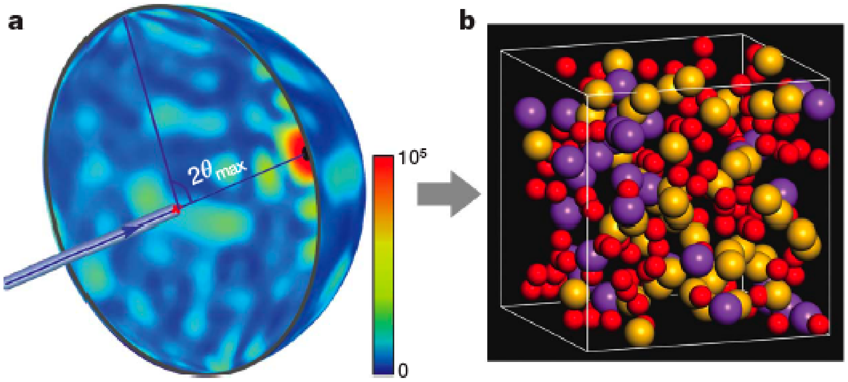

Figure 4.3: (a) Schematic layout for the high numerical aperture diffraction experiment, where a coherent beam scatters off the sample, with a maximum diffraction angle $2\theta = 90°$ (The colorbar shows the dynamic range in the data). (b) The simulated glass structure with volume $14 \times 14 \times 14$ $\AA^3$ obtained by molecular dynamics simulations. The red, purple and yellow correspond to oxygen, sodium, and silicon atom's positions, respectively.

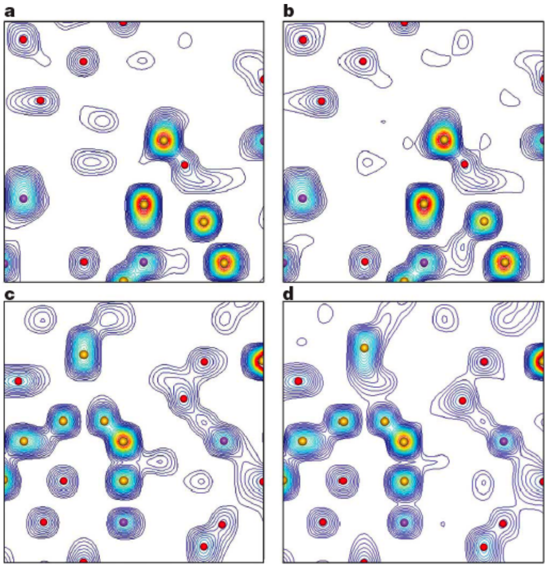

Figure 4.4: Left panel: (a,c) are two reconstructed slices along xy and xz. Right panel: (b,d) are the original electron density distributions obtained through molecular dynamics simulations. By measuring the distance between the two binding atoms, we confirm the resolution of 2 $\AA$, which corresponds to $\lambda = 2$ $\AA$ in agreement with 90° scattering angle. (a, c) are reconstructed from the measured spherical shell though phase retrieval, along the xy, yz planes respectively. The slight discrepancy between the reconstructed and the simulated density, is due to noise in the measurement.

modulus constraint enforcing the measured points on the shell, and positivity, boundedness and smoothness. Smoothness constraint inside the support is implemented periodically to prevent stagnation, while the uniformity outside the support is applied at each iteration to reduce the dimensionality of the problem. In both cases, the corresponding convolution kernel is a $3 \times 3 \times 3$ voxels, a Gaussian smooth edge curve inside the support, and a box filter outside the support. To monitor the convergence of the retrieval process, the error metric ϵ_m on the Ewald shell is computed. As the initial seed dictates the evolution of the algorithm, fifty different random starts were conducted, and the smallest ϵ_m is selected as the optimal solution.

Fig 4.4 (a, c) are the two slices obtained along the xy and xz directions. The distance between the nearest neighbour atoms demonstrate that the resolution obtained is indeed 2 $\mathring{A}$ in agreement with the resolution computed for a $\lambda = 2$ $\mathring{A}$ and a scattering angle of 90°. Fig 4.4 (b, d) are the corresponding molecular dynamics simulations. The agreement between the reconstructed and the simulated structure is high; the slight discrepancy is attributed to the noise level present in the measured data.

4.4.2 Polio virus

The Polio virus is an excellent model system for studying the cell entry mechanism of simple non-enveloped viruses. How this kind of non-envelop virus releases its viral RNA into the host cell is not known. This virus contains a single-stranded RNA genome within an icosahedral capsid that is delivered into a host cell via interaction of virus coat proteins with a poliovirus receptor. During the process, the virus undergoes an irreversible conformational change that results in an increased affinity for the poliovirus receptor, producing what is know as the $135S$ particle. The particle was obtained by cryo-electron microscopy where 8244 particles were averaged [65]. The particle was downloaded from the Electron Microscopy Data

Bank at http://www.ebi.ac.uk/pdbe/emdb/entry1136.

Here we simulate an XFEL pulse at wavelength $\lambda = 1.77\ nm$ with 10^{13} photons, focused down to a 100 nm spot, scattering off the $135S$ particle. Measurements with added Poisson noise were calculated on the Ewald sphere up to 62.6° angle, corresponding to 2.0 and 3.3 nm transverse and depth resolutions. To simulate the beam stop, the central $7 \times 7 \times 7$ pixels were removed. Fig 4.5, left panel shows the front view of the spherical diffraction pattern used for this simulation, with a dynamic range from 0 to $\sim 10^6$ photons. The right panel shows the reconstructed poliovirus. This iso-surface rendering is in excellent agreement with the model, where the reconstructed capsid form the five fold symmetry resolved by the star-shaped mesa (red dots), and the propeller tip (blue dots) forming the three fold symmetry. Fig 4.5 a, b also show the canyon (arrow) which is the expected binding site of the poliovirus receptor. Phase retrieval was performed through a similar process as for the glass structure reconstruction. However, due to the larger array size in this case, in addition to the smoothness constraint, we used the amplitude extension constraint, as detailed in 4.3.3; table 4.1 shows the incremental steps used. The total oversampling degree O_t refers to the total reciprocal space points divided by the number of unknowns in real space. The points in the reciprocal space include the Ewald sphere as well as the points acquired by amplitude extension. Diffraction angle refers to the maximum scattering angle used for a particular low resolution simulation. $\epsilon_{m_{ewald}}$ is the data fidelity error as measured on the Ewald sphere. As this is a simulated experiment, we have access to the entire domain; $\epsilon_{m_{entire}}$ is error acquired in the entire reciprocal space volume.

Due to the discrete nature of the model, it is impossible to exactly simulate the continuous diffraction intensities on the Ewald sphere. To resolve this issue, we first interpolate the structure into a $10 \times 10 \times 2$ voxels (table 4.1), and calculate its oversampled noisy Fourier magnitudes A_1 embedded in the Ewald sphere with a $40 \times 40 \times 6$ voxels. By using the iterative algorithm, we reconstruct the low

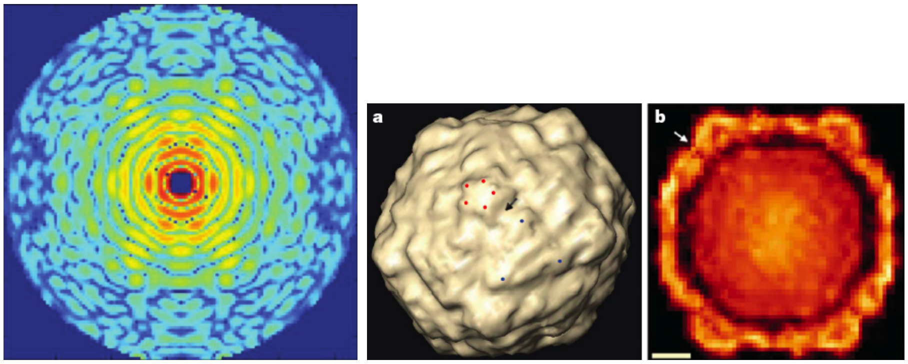

Figure 4.5: Left panel: Front view of the noisy spherical diffraction pattern (dynamic range $[0, 1.4 \times 10^6]$ photons) used for the reconstruction of the Poliovirus, where the intensities at the central 7×7 region were removed. The pattern was embedded into a volume of $128 \times 128 \times 78$ voxels. Right panel. Reconstructed structure of an individual poliovirus from a single simulated X-FEL pulse. (a) Iso-surface rendering of the reconstructed viral capsid structure, showing a five-fold mesa (red dots) and the tips of a three-fold propeller (blue dots). The canyon between the mesa and the propeller structure (arrowed) is the expected binding site to its receptor the volume occupied is $32 \times 32 \times 20$. (b) A 1.65 nm thick central slice of the reconstructed 3D virus structure across the five-fold mesa and the propeller structure. The arrow indicates the receptor binding site. Scale bar, 5 nm. The two panels (diffraction pattern and reconstructed structure) are not shown to the correct relative scale.

R-space Array	K-space Array	O_t	Diffraction Angle	$\epsilon_{m_{ewald}}$	$\epsilon_{m_{entire}}$
$10 \times 10 \times 2$	$40 \times 40 \times 6$	7.29	16.1°	0.0004	0.002
$16 \times 16 \times 4$	$64 \times 64 \times 16$	7.34	26.3°	0.0007	0.010
$20 \times 20 \times 6$	$80 \times 80 \times 24$	14.58	33.7°	0.002	0.015
$24 \times 24 \times 10$	$96 \times 96 \times 38$	13.99	41.7°	0.002	0.021
$32 \times 32 \times 20$	$128 \times 128 \times 78$	9.07	62.6°	0.003	0.14

Table 4.1: Progressive amplitude extension used in Polio virus.

resolution virus structure as well as the entire Fourier magnitudes, $|F^{A_1}_{Recon}(\vec{k})|$. We then interpolate the reconstructed structure into a $16 \times 16 \times 4$ voxels, and also calculate its Fourier magnitudes on the Ewald sphere with Poisson noise, $|F^{A_2}_{Ewald}(\vec{k})|$, where A_2 represents a $64 \times 64 \times 16$ voxels. To implement amplitude extension, we calculate the hybrid Fourier magnitudes by:

$$|F^{A_2}_{hybrid}(\vec{k})| = \begin{cases} \omega \cdot |F^{A_2}_{True}(\vec{k})| + (1-\omega) \cdot \frac{|F^{A_1}_{Recon}(\vec{k})|}{|F^{A_1}_{True}(\vec{k})|} \cdot |F^{A_2}_{True}(\vec{k})| & \vec{k} \in A_1 \\ |F^{A_2}_{Ewald}(\vec{k})| \qquad \vec{k} \in A_2 \end{cases} \tag{4.7}$$

Where $|F^{A_1}_{True}(\vec{k})|$ and $|F^{A_2}_{True}(\vec{k})|$ represent the entire Fourier magnitudes calculated from the $10 \times 10 \times 2$ and $16 \times 16 \times 4$ voxels, and ω is a parameter between 0 and 1. We adjusted ω such that:

$$\frac{\sum ||F^{A_2}_{Hybrid}(\vec{k})| - |F^{A_2}_{True}(\vec{k})||^2}{\sum |F^{A_2}_{True}(\vec{k})|^2} = \frac{\sum ||F^{A_1}_{Recon}(\vec{k})| - |F^{A_1}_{True}(\vec{k})||^2}{\sum |F^{A_1}_{True}(\vec{k})|^2} \qquad \vec{k} \in A_1 \tag{4.8}$$

The assembled $|F^{A_2}_{Hybrid}(\vec{k})|$ is then used to reconstruct the $16 \times 16 \times 4$ voxels. We repeat the amplitude extension procedure in table 4.1 until the full-size reconstruction of the poliovirus structure is obtained.

CHAPTER 5

Experimental Methods and Reconstructions

In this chapter, we detail three experiments conducted with hard x-ray, soft x-ray and optical microscopes. The first experiment was performed at Spring-8 in Japan, where we used a mitochondria as a sample. The data quality was not optimal for a reconstruction, nonetheless, we gained important experience regarding the experimental setup §5.1. The second experiment was conducted at Jila in Colorado as a proof of concept for table top diffraction imaging [66]. In this case we also used the data to demonstrate ankylography with soft x-ray data [35], §5.2. Lastly, the optical experiment was fully conducted at UCLA where we obtained clean measurements as well as an excellent reconstruction, demonstrating the feasibility of ankylography in the optical regime [36], §5.3.

5.1 The coherent X-ray microscope

Hard x-rays, between ~ 0.1 and 2 $\AA$ wavelengths, diffracting on the atomic scale can provide high resolution images of the scattering electron densities. Unlike table top light sources, access to coherent synchrotron radiation requires collaboration between institutions. Synchrotron facilities offer advantages such as small beam divergence, continuous energy modulation, and high beam flux and thus greatly facilitate the acquisition of high resolution data. Such beam characteristics are due to the underlying sophisticated ring and beamline design. To provide an overview of the coherent x-ray diffraction microscope (CXDM), we describe the beamline BL29XUL at the Super Photon Ring 8 GeV (SPring-8) and detail

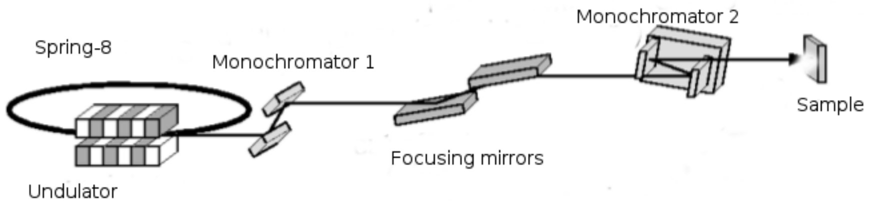

Figure 5.1: A sketch for the x-ray optics showing the beam circulating in the ring, entering the undulator, subsequently, passing through the first monochromator to enhance the beam spectral coherence. The focussing mirrors and the additional monochromators downstream further improve spectral coherence. This sketch was obtained from [68].

the experimental setup.

BL29XUL is the third beamline constructed at Spring-8 with three major parts: front-end, optics hutch, and three experimental hutches (EH1, EH2, EH3) [67]. At the front end, a standard in vacuum undulator is located, with tunable 140 periodic magnets resulting in an optimized x-ray flux emission ($\sim 10^{13}$ photons/s) and covering a spectral range of 4.5 to 18.7 KeV. At the optics hutch, there is a liquid nitrogen cooled Si Double Crystal Monochromator (DCM) and a pair of reflecting mirrors used to control the spectral-flux modulation and beam collimation (Fig 5.1).

The experimental hutch (EH1) houses the customized diffraction microscope, which is located ~ 52 m from the x-ray source. Fig 5.2 shows the schematic layout of the diffraction microscope. The incident beam delivered to EH1 is 1.3 mm in width and 0.7 mm in height; a relatively large area in comparison with a typical sample size of around $\sim 1 - 10$ μm. Here we describe the experimental hutch by dividing it into three regions: upstream, sample and downstream region.

To enhance spatial coherence, a 20 μm-diameter pinhole aperture is installed about 1 m upstream from the sample [69]. Downstream of the pinhole, two thick

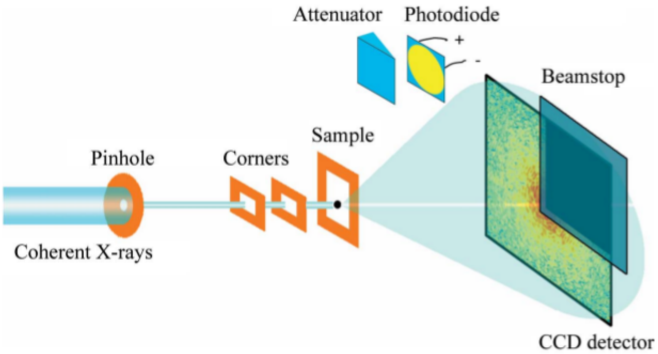

Figure 5.2: A sketch of the CXDM microscope; figure is obtained from [43].

silicon windows with beveled edges are introduced inside the sample chamber as L-shaped guard corners, where the lower-right corners are used to minimize the scattering from the pinhole edges. The combination of pinhole and corners produce clean diffraction signal in three adjacent quadrants on the detector.

The sample is mounted on a thin, 30 nm thickness, silicon-nitride membrane framed by 200 microns thick Si. The thin membranes are transparent to the x-ray energies, and hence they form a non-interfering sample holder. One imperative requirement for CXDM is that the sample be well isolated on the membrane and that the membrane be free from films, residue, dust and condensate. Non-uniformities in the region surrounding the sample result in noisy interference measured in the diffraction pattern which would hinder the image reconstruction, as specifying the support which includes the noisy regions would be a harder problem. To reduce the background signal, the CXDM is operated in vacuum ($\sim 10^{-4}$ Pa). In-vacuum piezo-actuator coupled motion stages (Newport CMA-25) with a resolution of 1 μm/step are used to manipulate guard corners and sample positions. Motion stages are controlled using LabView. Prior to mounting the sample inside the chamber, the sample's relative position referenced by the membrane edges is mapped using a high resolution optical microscope.

A movable attenuator, positioned downstream of the sample, permits a direct beam measurement, and facilitate the alignment of optical components. The location of the incident beam provides a rough estimate of the central pixel, corresponding to the missing center in the measurements; this location is further determined in the subsequent data analysis. At the downstream of the attenuator, a movable photodiode measures the beam strength. As a cautionary step, the attenuator is used, when measuring the intensity of the incident beam, as preventive measure shielding the CCD from the intense direct beam. The diffraction patterns are measured by a deep-depletion and liquid-nitrogen-cooled CCD camera with 1340×1300 pixels and a pixel area of $20 \times 20\ \mu m^2$ (PI-LCX1300). The distance between the sample and the CCD camera is adjustable in order to fulfill the oversampling requirement [70], which is a function of the x-ray wavelength, the sample size and the detector pixel size [71]. A large beamstop, mounted just in front of the CCD detector, is used to block the fourth noisy quadrant as well as the direct beam. As the membrane edges are the only landmark for the sample location, the sample stage is scanned to find the designated edges. This step is performed using the photodiode as the intensity counter. At the moment the incident beam hits the sample on the membrane, the scattering is quite noticeable as seen on the detector. Once the edge position is known, the sample can be located using the offset coordinates obtained from the optical microscope mapping. Fine adjustment of the sample position is carried out by maximizing the counts recorded on a CCD detector. A 2D scanning of the sample is performed in the plane perpendicular to the beam direction. A low-resolution diffraction pattern is measured by the CCD for each scanning position of the sample. The optimal sample/beam relative position corresponds to the position that yields maximum diffraction intensity. In order to measure a 3D data set, the sample stage is mounted on a rotary stage. The sample is usually restricted to rotation angles between $\sim \pm 70^\circ$ - 80°, due to the obstruction with the silicon-nitride-membrane edges [43].

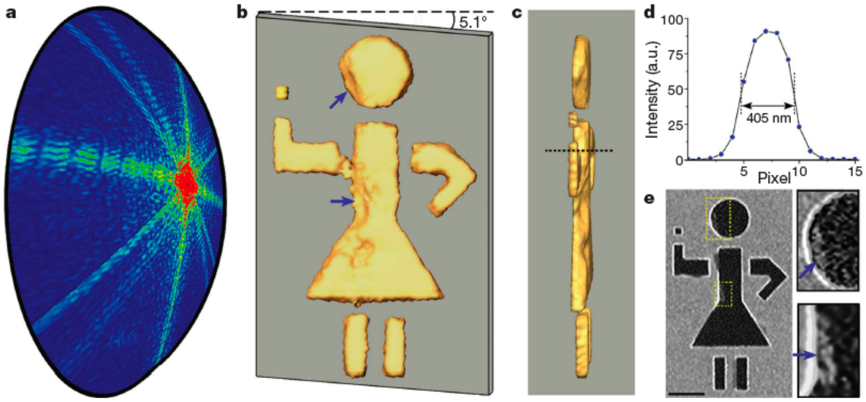

Figure 5.3: (a) Oversampled 2D diffraction pattern on the Ewald sphere. (b,c) Iso-surface renderings of the 3D reconstructed image in the transverse, and the longitudinal planes respectively. (d) Line scan along the dashed line in (c) indicates the width of the tilted sample's projection onto the z-axis to be 405 nm. (e) SEM image of the sample (scale bar 1μ m). Insets (right) show two structure defects (blue arrows) in the sample which are spatially resolved in the 3D reconstructed image (blue arrows in b).

5.2 Table top soft X-ray data

Although, I have not participated in the experimental setup, we re-used the data to demonstrate Ankylography. Here, we detail the 3D reconstruction and briefly describe the experimental setup. For more detailed description of the experiment please see [66]. The diffraction data was obtained with a soft x-ray table top light source, with a wavelength $\lambda = 47$ nm, and a temporal coherence $\frac{\lambda}{\Delta\lambda} \approx 10^4$ where a 2D reconstruction was performed [66]. The sample is a thick aperture of a waving stick figure. The test pattern was etched on a substrate made of silicon nitride membrane with thickness ≈ 100 nm. The substrate is opaque to the soft x-ray, except at the etched location. The test pattern had a slant along the longitudinal direction of the beam, and thus provided a $3D$ depth to the sample. A CCD detector was placed 14.5 mm from the sample, where the 2D diffraction pattern was measured. The CCD detector is an Andor with 2048×2048 pixels, operating at the extreme ultraviolet regime, with 13.5×13.5 μm^2 per pixel area.

For the case where the experimental data has a high oversampling, it is beneficial to bin the data to enhance the signal to noise ratio in the measurements. Hence, here we integrated the intensity over a 3×3 pixel area. Subsequently, we interpolated the data on the Ewald sphere as described in §4.3. The spherical diffraction pattern with an angle $2\theta = 35.9^\circ$ was embedded into a $3D$ array of size $420 \times 420 \times 240$ voxels with an oversampling degree $O_d = 2.6$.

Similar to the simulation objects in §4.4, phase retrieval is implemented through the iterative projection algorithm using HIO, where positivity, smoothness inside the support, and uniformity outside the support were enforced. Fig 5.3 b,c show the iso-surface rendering of the 3D reconstruction in the transverse as well as the longitudinal plane. The array size of the reconstructed image is embedded in a $100 \times 170 \times 7$ voxels. In Fig 5.3 b, the blue arrows correspond to dust particles present in the scanning electron microscope (SEM) image (Fig 5.3 e). The spatial resolutions of the reconstruction along the xy and xz planes were estimated to be 80, 14 nm, respectively. We also measured the tilt angle of the sample relative to beam to be 5.1° (Fig 5.3 b). Figure 5.3 d shows a line scan along the z axis of the reconstructed image. From the reconstruction analysis, the sample's depth is 405 nm which is in close agreement with the expected value of 389 nm calculated from the sample geometry and tilt angle [35].

5.3 Table top optical microscope

A naive description of the optical setup makes it sound simple: a planar monochromatic wavefront weakly scatters off an isolated sample; the scattered signal is measured on a planar detector. However, in practice, the minutae of the sensitive details, play the major role in the success of this optical diffraction microscope. Here we describe the four major experimental steps, in addition to the algorithmic step which facilitate the success of this optical microscope: sample preparation, beam quality, direct beam suppression, data collection and lastly image analysis

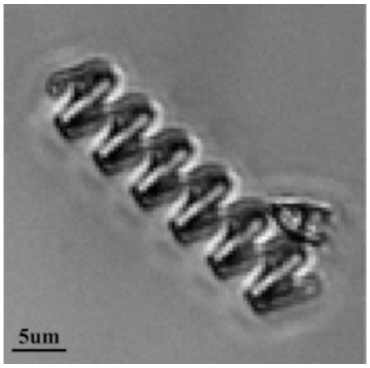

Figure 5.4: DIC microscope image of the raftlike thin object with volume $\sim 4 \times 7 \times 1\ \mu m^3$.

and reconstruction.

The Sample preparation is a very delicate step which requires the isolation of $7 - 12\ \mu m$ sized particles on a 30 nm transparent membrane of SiN4. Such membrane is completely transparent to the beam as its thickness falls below the incident beam's wavelength ($0.543\ \mu m$). Any dust particle ($> 0.4 \mu m$) in the beam adds an undesirable noise to the interference pattern. To prevent noise due to airborne dust particles, the experiment was mounted on an optical bench, with clean air blower. The sample is an optically transparent particle, a dielectric phase pattern made of non-absorbing SU-8 epoxy photoresist that had been cross linked by using an Ultratech XLS stepper [72, 73]. The differential-interference-contrast (DIC) image shows a raft like arrangement of letters (WWWA) with size $\sim 4 \times 7 \times 1\ \mu m^3$ (Fig. 5.4).

Beam quality is a function of the source generator and the auxiliary optics used to reduce its size. The finite atomic transitions of the HeNe Gas laser render these kind of sources with an exceedingly narrow frequency bandwidth, suitable for a diffraction experiments. Such lasers can operate with transverse mode TEM_{00} of a Gaussian beam profile. This profile is desirable for two main reasons: it is the closest approximation for the spatially confined plane wave and preserves its

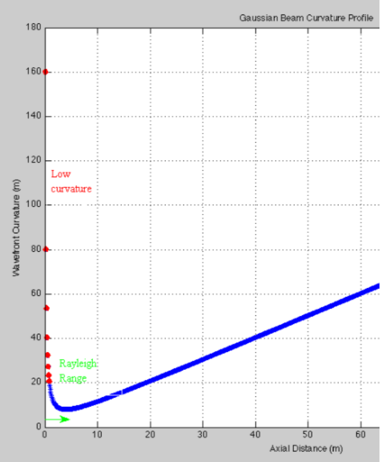

Figure 5.5: Wavefront curvature as a function of propagation distance.

shape as it passes through a collimation system.

Due to the difficulty of isolating the sample to within the beam cross sectional area of $\sim 1350\ \mu$ m, we reduce the beam size to $\sim 200\ \mu$ m and hence drastically minimize the probability of scattering from a parasitic object in close vicinity to the sample. Ideally, a top hat beam profile of cross section slightly larger than the sample's cross section is desirable to avoid any defects that may be present on the membrane. However, the inverse relation between beam size and beam divergence places a lower limit on how small the beam size can be. To reduce the beam size, we used a combination of two lenses. The distance between the lenses and the respective positions from the sample and the laser were optimized following the theoretical wave propagation of the beam through the lenses, and fine tuned experimentally. The transmission of a Gaussian beam through the optical lenses alters the beam waist W_0, the radius of curvature $\mathcal{R}$ as well as the depth of focus z_0, however the beam profile is preserved. For laser cavity with a confocal resonator, the beam waist is located at the cavity's center. To bring the focal waist in front of the cavity, we place a 500 mm plano-convex lens, at a close distance from the laser, such that the beam entering the first lens is approximately planar, as the curvature of the beam is infinite at the waist, but drastically decreases, to a minimum at the Rayleigh's distance $z_0 = 4.5$ m. However, by placing the first lens at $z << z_0$ such that the curvature is still low (Fig. 5.5), we find that the focused location of the beam is within the ray optics calculation. Hence by adding the second lens at one focal distance from the first focal point, we reduce the beam by a factor of $\frac{f_2}{f_1}$. The propagated waves quickly acquire a curvature, once the beam size is reduced, as the new depth of focus z_R is further reduced by $(\frac{f_2}{f_1})^2$. Therefore it is important to operate very near the beam waist, where $\mathcal{R}$ is large:

$$\mathcal{R}(z) = \begin{cases} \infty & \text{for} \quad z << z_R \\ z & \text{for} \quad z >> z_R \end{cases} \tag{5.1}$$

Internal reflections at the lenses' surfaces add additional challenges to the set up. By slightly tilting the lenses while keeping the beam within the central lens region, we offset the reflected beam. Furthermore, the laser's lasing medium, the red halo, is also transmitted along the laser beam. To eliminate this parasitic light, we place a pinhole at the first focal point. The pinhole acts as a spatial filter, where the opening is painstakingly adjusted not to vignette the direct beam, but narrow enough to block the multiple reflections of both the lasing medium and the reflected direct beam. Figure 5.6 shows a sketch of the optical set up. The sample holder was rotated at 45° to create additional depth as the sample's thickness was below detection.

The Beam stop protects the CCD from intense direct radiation that will destroy pixels hit by direct exposure. In figure 5.6, the faint shadow is the beam stop seen through the pattern; the dark region on the pattern is the projection of the beam stop. The direct beam amplitude is many orders of magnitudes larger than the scattered beam, and is beyond the detector's dynamic range. Hence, it is crucial to completely block the direct beam once it has passed through the sample. To collect high numerical aperture diffraction data, the CCD was placed very close from the sample. The compactness of the setup required a very delicate

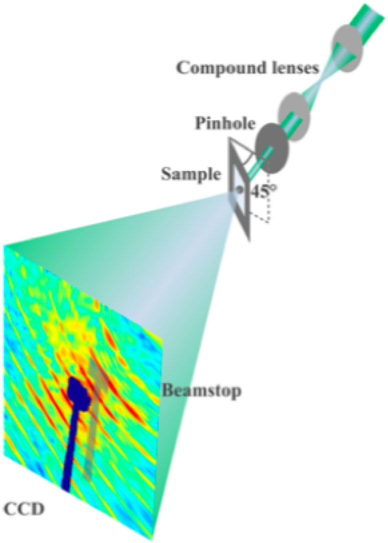

Figure 5.6: A sketch of the experimental set up.

construction of the beam stop which was made of highly reflective SiO_2 material. In order to minimize the reflection, we also blackened the surface of the beamstop. The reflection direction was optimized such that the reflected beam does not intercept any other reflective surfaces which would have added an extra background to the detector. In this case the reflection from the beam stop was aimed on the wall covered by a black cloth and towards the blind side of the detector.

Data collection The intensity measurements were recorded on a liquid Nitrogen cooled CCD camera. To minimize dark current noise, the CCD temperature was held low during the experiment. With 1340×1300 pixels and 20×20 μm^2 per pixel area, we collect a thousand exposures for each high and low resolution data (fig 5.7). The high resolution exposures were taken with a shorter time length due to the faster saturation of the detector, as the signal is stronger at a location close to the CCD. The reason we acquired data at low resolution was to ensure that the missing center is confined within the centrospeckle [26]. Prior to recording the scattering intensities, we collect background exposures with the same acquisition length. For each CCD position, we optimize the lenses to suppress any additional

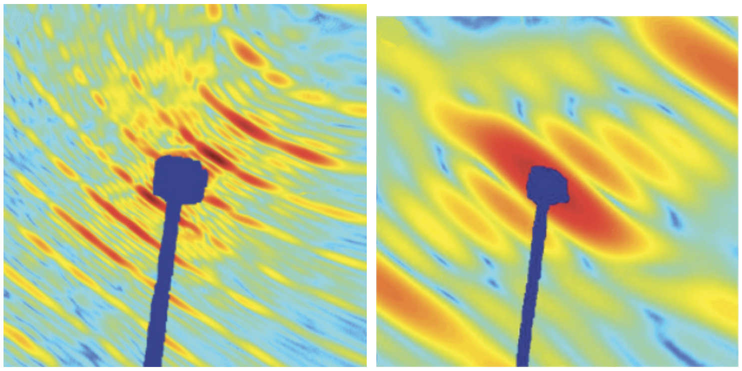

Figure 5.7: High and low resolution diffraction patterns collected at distances of $31, 108$ mm respectively from the sample.

scattering, where we obtain a clean background. At the high resolution position, we acquired four patterns. To obtain highest possible angles, we translated the detector both horizontally and vertically. The span of these translation was limited by the camera's mechanical rail and not by the signal's strength.

Data analysis As mentioned above, four diffraction patterns were acquired at high resolution, and one at low resolution. After background subtraction, the four diffraction patterns were seamlessly merged by aligning and normalizing the overlapping regions. Prior to aligning with the high resolution data and to account for the larger oversampling, the low resolution data were binned. Subsequently, we performed image pre-processing which includes flux normalization and interpolation (§4.3 for more details) before the image reconstruction step, where the pattern was embedded in the respective 3D volume (fig. 5.8).

Image Reconstruction of the WWWA sample. HIO was used to perform the phase retrieval. The transverse support was first estimated from the reconstruction of the 2D pattern, where the reconstructed image shows defocusing effects related to the curvature of the data, nonetheless it provides a rough estimate of the cross sectional support area. We project the 2D support along the longitudinal direction where the thickness can be roughly estimated from the longitudinal reso-

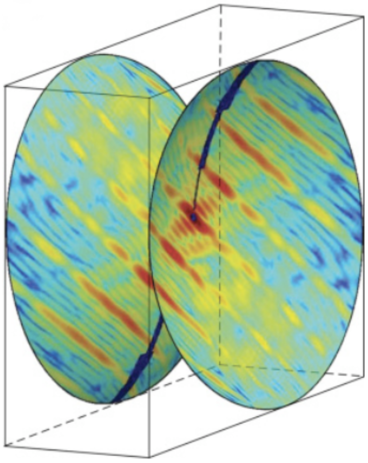

Figure 5.8: Patterns used for the Ankylographic reconstruction, post flux normalization and spherical shell interpolation, the corresponding diffraction angle is 32.3°.

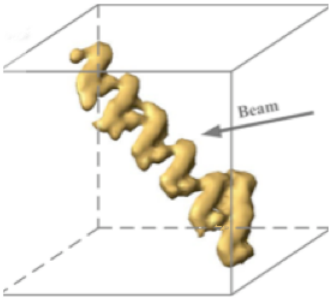

Figure 5.9: Surface rendering of the WWWA reconstruction.

lution and the transverse support. Subsequently, the 3D support was refined using the 3D reconstruction algorithm, where we applied the smoothness constraint to minimize the effects of stagnation, as discussed in §4.3. The convergence of the algorithm was monitored through the ϵ_m metric. Unlike a 2D reconstruction, 3D requires a large number of iterations (~ 5000). Figure 5.9 shows a surface rendering of the reconstruction, where the tilt angle, with respect to the incident beam, is accurately recovered. The transverse and longitudinal resolutions obtained were 1.0 μm and 3.5 μm.

CHAPTER 6

Experimental and Computational Challenges

Noisy, incomplete data can be directly attributed to the experimental challenges which subsequently cause the computational challenge. With the availability of new high flux sources, radiation damage is the main obstacle to higher resolution imaging. In this chapter, we first review the radiation damage problem §6.1.1, then we list current experimental methods aimed at overcoming this obstacle §6.1.2 6.1.3. We then show that the computational difficulty of noisy incomplete data leads to degenerate, non-unique solutions §6.2. To gain further insight into the computational problem, we examine a low dimensional system §6.2.1, we also offer a quick preview to the adaptive phase retrieval algorithm §6.2.1.1, which is fully described in chapter 8.

6.1 Experimental challenges

Higher resolution structure determination using x-ray diffraction, requires shorter wavelengths and larger numerical aperture. While the high frequency radiation increases the potential of specimen damage, collecting signal at high diffraction angle requires longer exposure time or higher flux, each of which also cause sample deterioration §6.1.1. Various experimental methods were developed to overcome this challenge: femtosecond imaging §6.1.2, generating reproducible samples and using statistical averaging to improve the signal to noise ratio §6.1.3.

6.1.1 Radiation damage

The scattered intensity is proportional to the fourth power of the resolution [74, 75], as a result, high resolution imaging requires a tremendous increase in the scattering power, which requires higher incident flux. With a stronger flux, the additional energy deposited into the sample can cause structural changes and higher order scattering events. In CDI, we are mainly interested in single elastic scattering, where the relation between the measurement and the structure is well defined. Furthermore, the energy of high frequency radiation deposited onto the sample destroys the electrostatic bonding between the molecules, and also lead to sample deterioration [76, 75, 77]. While cryo-protection has been used to increase the maximum tolerable dose for certain bio-particles [78], with femtosecond imaging it is possible to obtain data prior to the occurrence of radiation damage §6.1.2.

6.1.2 Diffract and destroy

With the intense short x-ray pulses of X-FEL (< 10 fs, 10^{12-13} photons), a billion times more powerful than the third generation synchrotron radiation light source, it is possible to collect diffraction data, before the coulomb explosion of the specimen takes place [79, 80]. Reconstruction from single pulse was recently demonstrated with the aerosolization of mimivirus, and the live cell solution experiments [81, 82]. For three dimensional structure determination, and in the absence of prior knowledge, a single diffraction pattern might not contain sufficient information; as a result, diffraction from multiple copies is further explored §6.1.3.

6.1.3 Multiple copies

To collect multiple diffraction patterns of particles at random orientations, delicate methods of sample preparation are being developed. Container free sample injection alleviates the need for a manual sample replacement and takes advantage of the fast pulse rate of X-Fel sources. Various methods of particle streaming have already been successfully demonstrated, for example the gas phase injector as used in the mimivirus data [81] and the liquid phase injector used for the nano-crystal data [78]. Two competent strategies for recovering a single bio-molecule structure from correlation data are being explored. a) Using a single particle diffraction and by maximizing the correlations among multiple snapshots of identical copies, structural information can be recovered, as proposed by Fung et al where they predicted the assembly of three dimensional information from faintly scattering objects, based on the common-line method where the curvature of the Ewald sphere is manifested in a line intersection among the various orientations [83][84]. b) The concept of multiple particles in solution using correlated x-ray scattering measurements, is being re-visited now that the high intensity sources are coming online; for an excellent review please see [85].

6.2 Computational challenges

Because the energy landscape ϵ_m^2 is a non-convex function, projection based iterative algorithms, thereafter referred to as IPPRA, are prone to getting stuck in local minima. This problem is magnified as the amount of missing data increases. Here we simulate reconstructions from diffraction data with increasing level of missing points as plotted in figure 6.1. With ϵ_m on the x-axis and ϵ_r (image quality) on the y-axis, each of these plots contains a batch of reconstructions $\mathcal{N}_b = 256$. The top left plot correspond to measurements with high signal to noise ratio; the correspondence between ϵ_k, ϵ_r is one to one, as can be inferred from the slope of the linear trend. As the noise level increases (going clockwise across the panels), the slope approaches ∞, and hence it becomes insufficient to measure the accuracy of a reconstruction using only ϵ_m. To summarize, the non convexity in the energy landscape yields local minima which can create degeneracy in case of very high level of missing data. To further understand this transition between a well defined local minima and degenerate states, let us examine the energy landscape for a three-pixel system §6.2.1.

6.2.1 Low dimensional study

A qualitative description of the energy landscape is best understood using a graphical representation; however, with high dimensional images ($\sim 10^6$ pixels) no such visual representation exists. While low dimensional versions of high dimensional problems can often build intuition about their high dimensional counterparts, they cannot be taken too seriously since qualitatively different behaviour may emerge in the high dimensions. Nonetheless, using a simple model of a three-pixel system can provide an insight into the missing data problem. Here, we present figure 6.2 as simple illustration of the basic problem associated with missing and noisy data. Given a simple model $\rho(\vec{x})$ of a three-pixel system, with the corresponding measurement matrix $I(\vec{k})$, i.e:

$$\rho(\vec{x}) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 7 & 5 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad \text{and} \quad I(\vec{k}) = |\mathcal{F}\rho(\vec{x})|^2 \tag{6.1}$$

we compute the energy landscape ϵ_m^2 as:

$$\epsilon_m^2(x_1, x_2) = \frac{||\sqrt{I(\vec{k})} - |\hat{\rho}(\vec{k})||_2^2}{||I(\vec{k})||_2} \tag{6.2}$$

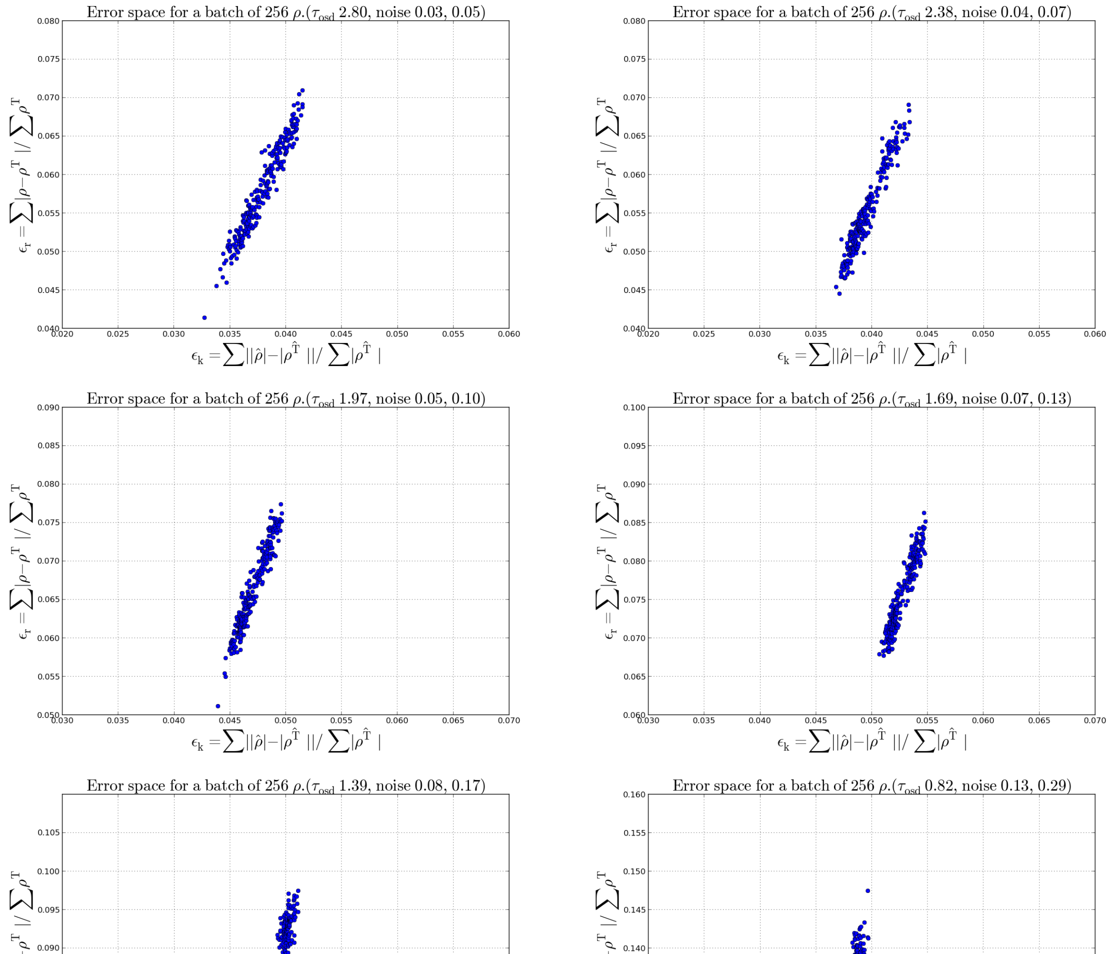

Figure 6.1: ϵ_m, ϵ_r as a function of noise, for a batch of 256 reconstructions. From top left to bottom right, the amount of missing data gradually increases.

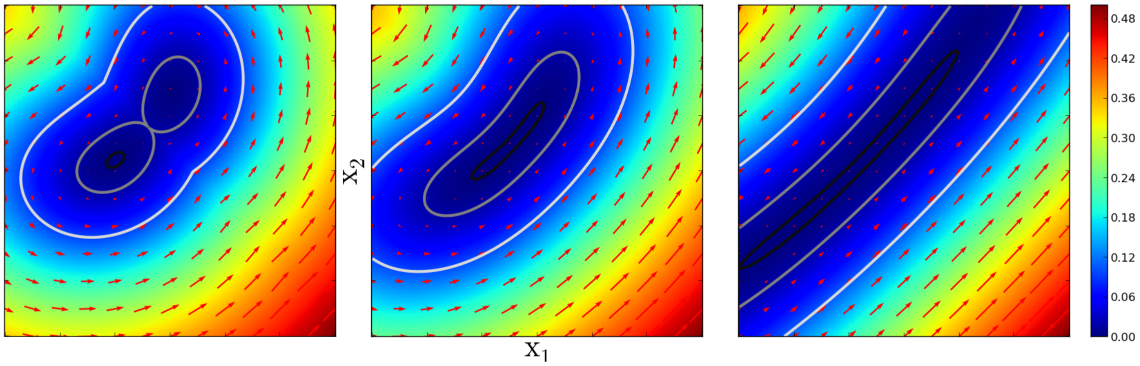

Figure 6.2: Energy landscape (ϵ_m^2) for a two dimensional system (x_1, x_2), where the missing measurements increase from $0, 20$ to 50% (left to right). The red arrows display the gradient. The solution lies within the black contour, local minima within the gray contour (far left subplot). As the missing data increases, the location of the global minimum becomes undefined, as the black contour stretches over a shallower landscape.

By plotting ϵ_m^2 as a function of the real space variables x_1, x_2, for different amount of missing data (Fig 6.2), one observes qualitative differences. Note x_3 can be computed through normalization. The left most sub-plot is for complete data where the location of the global minimum (black contour) is conspicuous, as a result, an IPPRA successfully reaches the global solution. The middle and far right sub-plots for incomplete measurements, where the high resolution measurements are missing, we observe smoother landscape, as the error vector in Eq. 6.2 has less components. The solution is somewhere within the shallow (black) contour. In such case, the IPPRA stagnates wandering around with no clear direction. To solve this problem, additional metrics should be included, as we demonstrate in chapter 8.

6.2.1.1 Adaptive phase retrieval (Preview)

Here we give a quick preview of the performance of adaptive phase retrieval (APR) algorithm, we show the low dimensional study, in comparison with IPPRA. In particular, structure determination for a three pixel system is implemented using three different iterative algorithms, namely alternating projection (AP), difference map (DM) and adaptive phase retrieval (APR). All maps are listed below for easy reference:

$$\rho^{(n+1)} = \mathbf{P_S P_m} \rho^{(n)} \qquad \text{AP}$$

$$\rho^{(n+1)} = \rho^{(n)} + (\mathbf{P_S}(2\mathbf{P_m} - \mathbf{I}) - \mathbf{P_m})\rho^{(n)} \qquad \text{DM}$$

$$\rho^{(n+1)} = \rho^{(n)} + (\mathbf{P_S}(2\mathbf{P_m} - \mathbf{I}) - \mathbf{P_m})\rho^{(n)} + \beta(\mathbf{P_{prior}} - \mathbf{I})\rho^{(n)} \qquad \text{APR}$$

The underlying image whose energy landscape is shown in fig. 6.3, fig. 6.4 is nine dimensional $\rho = [a, b, c, 0, 0, 0, 0, 0, 0]$, oversampled by three. As $c = \sum \rho - a - b$, this problem is a two dimensional phase retrieval problem. Note, that the twin image occurs at $\rho = [c, b, a, 0, 0, 0, 0, 0, 0]$. Hence, we expect global minima at (a, b) and (c, b). The background of the energy landscape ϵ_m^2 is computed using Eq.6.2. The global solutions are marked by the cyan squares, which are also the points of total intersections between all ellipses. Each ellipse corresponds to a single measurement point at k. We have nine measurements, but because of the symmetry in the Fourier transform (real density), we show only five sets (for $k = 0$, we have the whole plane). The sets are generated using Eq 3.4.

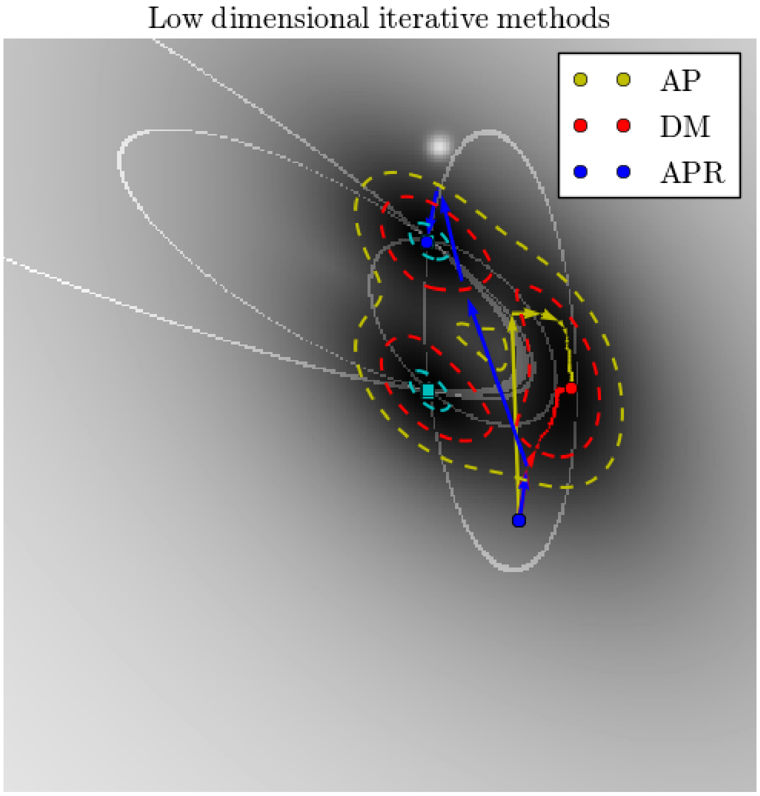

Figure 6.3: Background is the energy landscape ϵ_m^2. The white ellipses correspond to $I(\vec{k})$, for $k = 1, 2, 3, 4$ ($k = 0$ is the whole plane, $k = 5, 6, 7, 8$ are equivalent to $1, 2, 3, 4$, respectively). The global minima (twin) are marked by the cyan squares (or within the cyan contour). Within the red contour, a local minimum is located. The Gaussian located at the upper part of the outskirt of the yellow energy contour corresponds to the prior knowledge. Notice how the AP (yellow) and DM (red) iterations are stuck at the local minimum, while the APR (blue) easily emerges out of the stagnation point and into the global solution, despite the uncertainty in the prior knowledge. The beginning and the end of the iterative map is marked by a circle (the red dot is where DM ends, a local min.)

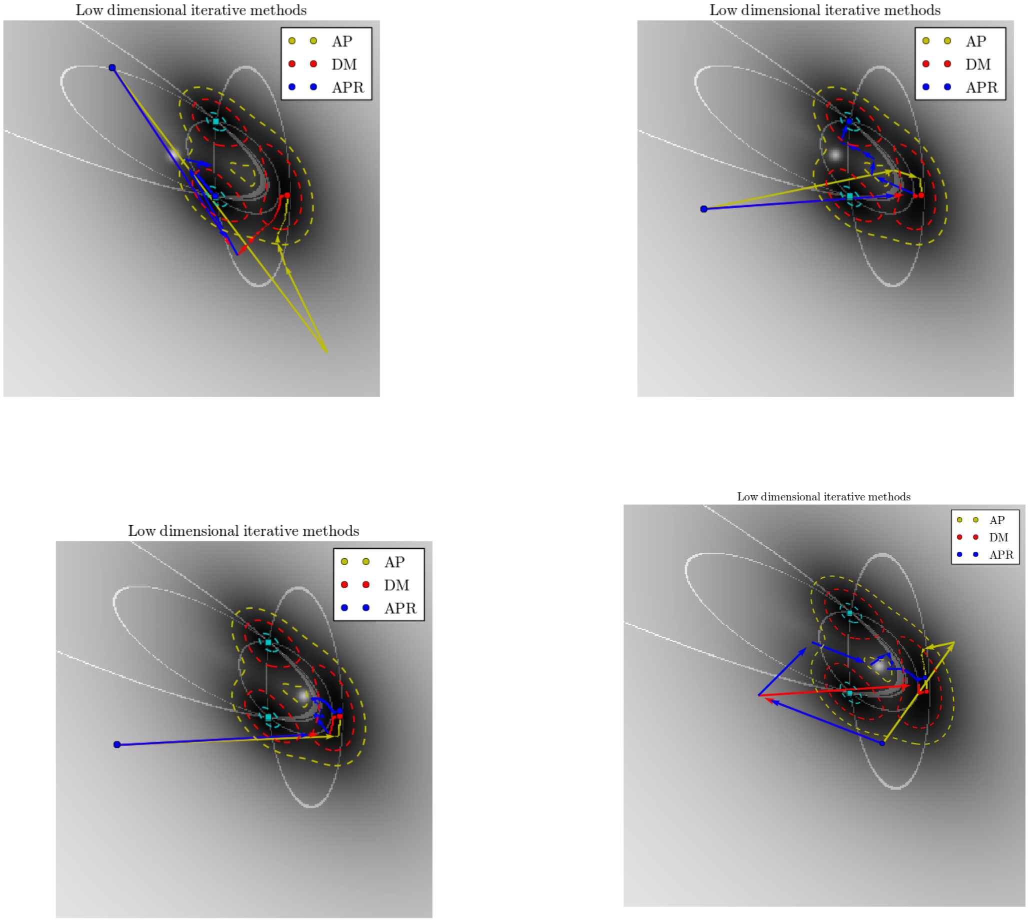

Figure 6.4: Top row: The performance of the algorithms using different starting points and prior location; APR successfully identifies the location of the global minima. Bottom row: Cases where APR might fail; when the prior is closer to a local minima; a case of an inaccurate prior.

CHAPTER 7

Bayesian Methods and Prior Knowledge

In this chapter, we give a brief review of Bayes'rule and its history in solving inference problems §7.1. We subsequently introduce Bayesian methods into phase retrieval for the case of noisy incomplete data §7.2.

7.1 Background

Bayesian methods provide a mathematical framework for calculating inferences. First documented by Thomas Bayes, an amateur mathematician (in the 17^{th}), as a method for solving inverse probabilities. Bayes' rule re-emerged in its current elegant form, due to the work of Pierre Simon Laplace who re-discovered Bayes's rule while estimating the mass of Saturn (early 19^{th} century), yielding a value with a one percent uncertainty. After many years of accumulated data, today the estimation of the mass of Saturn lies within the estimate calculated by Laplace with an improved uncertainty of only 0.63%. In her book *the theory that would not die*, Sharon McGrayne highlights the history of Bayes' rule, from obscurity to time of tremendous success [86]. The main argument against the Bayesian method came from statisticians, in particular frequentists who adhered to the concept that the only way to measure probability is through the frequency of occurrence.

Although there are certain experiments, where physicists learn about the physical world through measuring the frequency of occurrence, as in classical statistical mechanics, the scope of these experiments is limited. Frequentists define probability as a measurable description of the physical world and not as a measure of our ignorance about certain events. Although the frequency designation of probability is familiar: by counting the occurrence of heads and tails in a coin toss experiment, one can predict the coin bias in the forward sense. This approach, however, does not take into account relevant prior information. For example, the maximum likelihood solution in the case of a single head flip, corresponds to the hypothesis that the coin has heads on both sides. Using the most simple prior, such a uniform distribution which reflects our ignorance, one can obtain a more plausible outcome. Thus making sense of the data, using the inverse probability has a wider range of applications.

To understand inverse probability in its simplest form, take the urn experiment as formulated by Edwin Jaynes [87]. Given a box with two different balls of colors white or black, let A designate the first draw out of the urn and B the second draw. The probability of A being white is $P(A = w|I_1) = \frac{1}{2}$ where I_1 is the background information that there are only two balls in the urn with different colors. Imagine then the following scenario, we pick A without examining the color, we thus have $P(A = w|I_1) = P(A = b|I_1) = \frac{1}{2}$. We pick the second ball and find out it is white, as a result $P(B = w|I_1) = 1$. What is $P(A = w|I_1, I_2)$ now? where I_2 is the knowledge gained from the second draw. A moment of thought will yield $P(A = w|I_1, I_2) = 0$.

This last example showed that logical inference rather than causal inference leads to a knowledge update. Not a measure of the state of a physical world (white or black ball), probability defines our ignorance about the ball's identity. It is not because B is a white ball caused A to be a black ball, but rather because we *knew* the color of B, we were able to update our knowledge about the color of A.

By formally stating the prior information, we are objectively evaluating the subjectivity in a given assumption. Two people with the same knowledge will come to the same conclusion. As a result, a standardized framework of appraisal

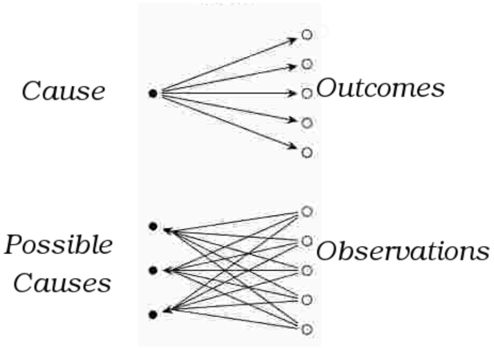

Figure 7.1: Inductive versus deductive logic diagrams. Figure obtained from [88].

is possible [87]. Before we go into the details of this method, let us review the two kinds of logic we are familiar with, so as to highlight the vastness of scientific inquiries where the second form of logic is prevalent.

7.1.1 Deductive versus inductive logic

Let A and B be two propositions. For A to imply B, the outcome is certain and is referred to as *deductive logic*, as depicted by figure 7.1 (top). For example, if we have the crystal structure, we can compute the diffracted intensities, and it will agree with the experimental results. Another example, if A is the proposition of a rainy weather and B is the proposition of cloudy skies, then the conclusion that rain implies clouds is certain. On the other hand, the existence of clouds does not imply rain, only renders the *possibility* of rain more plausible. This is a form of a bottom up reasoning (fig. 7.1, bottom) where the observation does not define the cause or outcome. Most scientific problems are reasoning of the second type, where *inductive logic* plays a role in assessing the state of knowledge by numerically evaluating the given propositions using inverse probabilities, as described in §7.1.2.

7.1.2 Fundamental rules of probability

To reason the best we can with the incomplete information we have is the goal of a systematic Bayesian inference, as stated by Jaynes [89]. To illustrate this idea, and following Jaynes [87] Tribus [90] explanation, we discuss the four essential pillars of the inference strategy: unambiguity, universality, consistency, and candor. Given these desired properties, Bayes' rule will naturally follow:

Unambiguity refers to the description of certain propositions. For example, if A is the proposition of rain in the afternoon, the proposition does not contain other implicit assumptions. Propositions are given within a context, i.e A given I or $A|I$, where I is the additional information explicitly indicated. The contradiction of A is referred to as $\bar{A}$.

Universality associates real numbers to the probabilities, such that various events can be compared on an absolute scale.

Consistency among various ways of reasoning is essential. If we have the same state of knowledge about certain propositions, we need to assign the same numerical values in order to be consistent. Richard T. Cox (1946) formally proved that the two laws of probability listed below are to be satisfied for consistency to be preserved:

$$P(AB|C) = P(A|BC)P(B|C) \qquad \text{product rule} \tag{7.1}$$

$$P(A+\bar{A}) = P(A) + P(\bar{A}) \qquad \text{sum rule} \tag{7.2}$$

Candor is stating explicitly the unknown information. Maximum entropy, which was the contribution of Jaynes [91], signifies the concept of candor. The maximum entropy approach will become obvious as we include an uncertainty measure into the prior information as shown in §7.2.1.

7.1.2.1 Bayes Rule

Bayes' rule can be trivially derived using the joint probability $P(A, B)$, and the product rule (Eq.7.1), as well as the symmetry in $P(AB) = P(BA)$:

$$
\begin{aligned}
P(AB|C) &= P(B|C)P(A|BC) \\
P(BA|C) &= P(A|C)P(B|AC) \\
P(A|BC) &\propto P(A|C)P(B|AC)
\end{aligned}
$$

7.1.3 One dimensional example of Bayesian inference

To further understand the importance Bayes' rule plays in a variety of disciplines, let us work out a simple example related to disease diagnosis, following the derivation in [92]. Let θ be the hypothesis to be tested; for example, θ can take on two values θ_1 chicken pox or θ_2 shingles. Let x be the random variable signifying the symptom, i.e a rash. A patient comes to a doctor, with a rash outbreak, wondering whether it is due to chicken pox or shingles. With the likelihood function relating the percentage of cases where chicken pox versus shingles causing a rash, we have $P(x|\theta_1) = 0.8$ and $P(x|\theta_2) = 0.5$. Thus, a diagnosis that doesn't involve any other information will mistakengly attribute the symptoms to chicken pox. However, if we take into account the prior knowledge, available through the public health diagnosis database, the diagnosis might vary. For example, if the percentage of the population diagnosed with chicken pox over the past ten years is 3%, in other words $P(\theta_1) = 0.03$, while shingles patients are more common with $P(\theta_2) = 0.8$, the posterior probability is updated to:

$$P(\theta_1|x) \propto P(x|\theta_1) \times P(\theta_1) = 0.8 \times 0.03 = 0.24$$
$$P(\theta_2|x) \propto P(x|\theta_2) \times P(\theta_2) = 0.5 \times 0.8 = 0.4$$

Hence, using the posterior probability, a doctor can more accurately predict that the underlying cause of the rash is shingles. In this simple example, one can see how the initial characterization of a certain disease based on prior knowledge obtained from public health records, improves the diagnosis accuracy.

7.2 Bayes rule in image reconstruction

Unlike the one parameter hypothesis testing of §7.1.3, the phase retrieval problem is testing a hypothesis with mega dimensions; however, we can still formulate the problem in a Bayesian framework. Let $\rho \in \mathbf{R}^n$ be the multi-dimensional structure in analogy with the hypothesis θ above, and $\sqrt{I}$ represents the measurements, in correspondence with the random variable x above. Using a certain model $\mathcal{M}$ which specifies the prior information, the goal is to maximize $P(\rho|\sqrt{I}\mathcal{M})$. Bayes' rule yields:

$$\begin{aligned} P(\rho|\sqrt{I}\mathcal{M}) &\propto P(\rho|\mathcal{M})P(\sqrt{I}|\rho\mathcal{M}) \\ &\propto P(\rho|\mathcal{M})P(\sqrt{I}|\rho) \end{aligned} \tag{7.3}$$

$P(\sqrt{I}|\rho\mathcal{M}) = P(\sqrt{I}|\rho)$ as once we define ρ, the informational content of $\mathcal{M}$ is irrelevant. Eq. 7.3 is a multi-dimensional probability distribution which is hard to visualize, but we can use optimization methods to solve this problem (chapter 8). Let us first examine the prior and likelihood distributions.

7.2.1 Prior knowledge

In order to formulate the prior distribution, we introduce the notion of entropy as defined in information theory by Shannon. Subsequently, by maximizing the entropy subject to some constraints we derive the corresponding probability distribution.

Shannon's Entropy Let x be a random variable, $p(x)$ its probability distribution and $h(x)$ the information measure describing the uncertainty in the measurements. There are two concepts that define this measure: monotonicity and statistical independence:

- If x is a predictable event, we expect the informational content to be very low. On the other hand, if x is highly improbable, than the informational content is high. Thus $h(x)$ should be monotonically increasing function of $p(x)$.
- For two statistically independent events x and y, the informational measure adds up; $h(x, y) = h(x) + h(y)$. Using the product rule, we deduce $h(x)$:

  $$h(x) = -\ln(p(x)) \tag{7.4}$$

  As $0 < p(x) < 1$, the minus sign prevents negative value for the informational content. For a high probability event $p(x) \approx 1$, the informational content approaches zero. While for a low probability event $p(x) \approx 0$, the informational content is very high.
- For a distribution of x_i, the sum of the expectation values of $h(x_i)$ defines the entropy:

$$H(x) = -\sum_i p(x_i)\ln(p(x_i)) \tag{7.5}$$

The probability distribution As we will show in chapter 8, the noise encountered in a batch of reconstructions obtained through phase retrieval, yields a specific standard deviation σ with an average μ. The global noise in the batch reflects the noise level of the measurements. Hence the prior distribution should satisfy the following constraints:

$$\begin{aligned}
\int_{-\infty}^{\infty} p(x)dx = 1 &\qquad \text{Normalization constraint} \\
\int_{-\infty}^{\infty} xp(x)dx = \mu &\qquad \text{Noise specific constraints} \\
\int_{-\infty}^{\infty} (x-\mu)^2 p(x)dx = \sigma^2 &
\end{aligned}$$

To solve this constrained optimization problem, we use the Lagrange multiplier to maximize the following functional, in addition to the entropy:

$$\begin{aligned}
J(p(x)) \equiv -\int_{-\infty}^{\infty} p(x)\ln p(x)dx + \lambda_1\Big(\int_{-\infty}^{\infty} p(x)dx - 1\Big) \\
+\lambda_2\Big(\int_{-\infty}^{\infty} xp(x)dx - \mu\Big) + \lambda_3\Big(\int_{-\infty}^{\infty} (x-\mu)^2 p(x)dx - \sigma^2\Big)
\end{aligned} \tag{7.6}$$

By differentiating in respect to $p(x)$, we have:

$$\frac{\partial J(p(x))}{\partial p(x)} = \int_{-\infty}^{\infty} (-\ln p(x) - 1 + \lambda_1 + \lambda_2 x + \lambda_3 (x-\mu)^2)dx \tag{7.7}$$

At the maximum, we have $\frac{\partial J(p(x))}{\partial p(x)} = 0$, thus:

$$-\ln p(x) - 1 + \lambda_1 + \lambda_2 x + \lambda_3 (x-\mu)^2 = 0$$

or

$$p(x) = \exp^{-1+\lambda_1+\lambda_2 x+\lambda_3(x-\mu)^2} \tag{7.8}$$

Thus the prior distribution is a member of the exponential family. To solve for $\lambda_1, \lambda_2, \lambda_3$, we use our knowledge of the Gaussian integral and substitute $p(x)$ back in the first constraint equations:

$$\int_{-\infty}^{\infty} \frac{\exp^{\lambda_2 x+\lambda_3(x-\mu)^2}}{\exp^{1-\lambda_1}} dx = 1 \tag{7.9}$$

Hence, we deduce that $\lambda_1 = 1 - \frac{\ln 2\pi\sigma^2}{2}, \lambda_2 = 0, \lambda_3 = \frac{-1}{2\sigma^2}$. Hence the prior is a Gaussian distribution, dominated by the error term $\lambda_3(x-\mu)^2$. By assuming a prior knowledge model $\mathcal{M}$, where $\mathbf{P}_{\mathcal{M}}$ is the corresponding projection operator (further discussed in chapter 8), with an associated error metric $\epsilon_{\mathcal{M}}$, such that $\epsilon^2_{\mathcal{M}}(\rho) = ||(\mathbf{I} - \mathbf{P}_{\mathcal{M}})\rho||_2^2$, we deduce $P(\rho|\mathcal{M}) \propto \exp -\alpha\epsilon^2_{\mathcal{M}}$.

7.2.2 Likelihood

The likelihood function relates the measurement to the hypothesis in the forward manner. For each measured point, assume a specific standard deviation $1/\sqrt{\beta_i}$, with an average value $\sqrt{I_i}$, then the probability distribution is $P(I_i) \propto \exp -\beta_i(|\hat{\rho}_i| - \sqrt{I_i})^2$, where $|\hat{\rho}_i|, \sqrt{I_i}$ represent the reconstructed and measured intensities, respectively. Assuming that the intensity measurements are statistically independent data points, the likelihood follows:

$$P(\sqrt{I}|\rho) \propto \prod_{i=0}^{n} \exp -\beta_i||\hat{\rho}_i| - \sqrt{I_i}|^2 \approx \exp -\beta\epsilon_m^2$$

Note here, that the likelihood function is not a probability distribution, hence the fact that it does not necessarily add to unity should be of no concern.

7.2.3 Posterior distribution

Unlike the likelihood, the posterior function encapsulates more information about the structure ρ, as it is related to the joint probability distribution $P(\mathcal{M}, \rho)$. This distribution does not provide a solution but offers a platform for risk analysis. The most common solution obtained from this probability distribution is the Maximum a posteriori (MAP). In the case of image reconstruction, where the multi-dimensional probability distribution is not convex, to seek a solution we implement a hybrid gradient descent with feedback mechanism, as further discussed in chapter 8.

CHAPTER 8

Adaptive Phase Retrieval

Limited by flux and radiation damage, the extent to which coherent diffraction imaging (CDI) requires less exact measurements, the more powerful tool it becomes. From nanomaterials to single proteins, structure recovery from incomplete diffraction data is at the forefront of advancement in biological and material sciences. As the diffracted signal scales with sample size [74], the amount of missing data is aggravated at the nanoscale. We have seen in previous chapters, how iterative projection phase retrieval algorithms, denoted hereafter by IPPRA, play a fundamental role in CDI. We will refer to these maps (i.e difference map, RAAR, particularly those covered in [54]) with the compact notation $\boldsymbol{\mu}$:

$$\Delta\rho^{(\nu)} = \boldsymbol{\mu}(\rho^{(\nu)}) \tag{8.1}$$

Although phase retrieval in the case of high SNR data had been successful, the case of noisy incomplete data is a bottleneck in the field. Here we focus on phase retrieval from incomplete noisy data through the incorporation of prior knowledge using Bayesian learning. In §8.1, we give a short overview of the algorithm; more detailed explanations follow from §8.2.

8.1 Introduction

The phase problem is to solve for n unknown variables from a large number m of non-linear, non-convex equations. The classical phase retrieval algorithms require

$m > 2n$ [70]. Motivated by our previous progress in determining the structure of simple 3D objects from limited diffraction data [35], and by the successes in compressed sensing where the phases are assumed to be known, we seek a method for recovering images from diffraction patterns that are not sufficiently oversampled. In what follows, we develop a general method for incorporating prior knowledge to compensate for fewer measurements or for restrictive measurement geometries. We find that we can robustly recover structural information from $m \in [\alpha n, 2n]$ coefficients, where typically $\alpha \sim 1.5$ for natural images; and $\alpha \sim 0.9$ for simple images. Here, we present a framework that incorporates priors as projection operators into the existing IPPRA; we show that these operators also have a probabilistic interpretation §8.2.1. We introduce three novel constraints: symmetry, total and local variation §8.2.2. We demonstrate the numerical success of this method in §8.3. We discuss the extensibility of this framework to general types of prior knowledge §8.4.

8.2 Methods

There are two sources of uncertainty in IPPRA using incomplete data. (*a*) The lack of convergence guarantee, as missing data deforms the energy landscape, and hence stagnation becomes a problem. In §8.2.1, we introduce a general map that incorporates prior knowledge, alleviating the stagnation problem. (*b*) Independent reconstructions result in distinct and inconsistent outputs, a consequence of the non-convexity in $\epsilon_m^2(\rho)$. In §8.2.2, we use these outputs to infer constraint information. In §8.2.3, we summarize the algorithm.

8.2.1 Reconstruction map

The adaptive phase retrieval algorithm (APR) is a general algorithm for incorporating new information into the phase recovery problem. There are two ways to formulate APR: using $\boldsymbol{\mu}$ as a starting map or optimizing a probabilistic model, heuristically.

Based on $\boldsymbol{\mu}$, we begin with the observation that conventional algorithms typically fail because their maps do not contain enough information to move out of local minima, long valleys, or plateaus of the energy landscape as in figure 6.2, where missing data generally flattens the landscape. We introduce a feedback condition through a set of constraint functions $\epsilon^2_{\mathcal{P}_i}$, preferably convex, where a gradient descent method is sufficient to minimize these metrics. Thus, we simply extend $\boldsymbol{\mu}(\rho)$ to:

$$(\boldsymbol{\mu} - \sum_{i=1}^{h} \frac{\beta_i}{2} \nabla \epsilon^2_{\mathcal{S}_i})\rho \tag{8.2}$$

Where $\beta_i^{(\nu)}$ is iteration and case dependent parameter, which can be optimized empirically; h is the number of the various types of additional information we include. When the iterate is near the constraint solution, $\nabla \epsilon^2_{\mathcal{S}_i} \rho^{(\nu)} \approx 0$. If $\epsilon^2_m(\rho)$ has sufficient structure to move the iterate towards the solution, the original map $\boldsymbol{\mu}$ will dominate the sum. By contrast, if the energy landscape is flat, far from the solution, and we are also far from satisfying the constraints, then $\nabla \epsilon^2_m << \nabla \epsilon^2_{\mathcal{S}_i}$. This creates the feedback condition between the constraints and the modulus error function.

The Probabilistic interpretation follows from Bayes'rule: The probability of having ρ as the density, given $\sqrt{I}$ for the measurements and a background knowledge X, can be expressed as the posterior equation:

$$P(\rho|\sqrt{I}X) \propto P(\sqrt{I}|\rho X)P(\rho|X) \tag{8.3}$$

For the first term, we treat the intensity points as nearly independent [1], and by using the central limit theorem, we can express the likelihood term as $\propto \exp(-\beta\epsilon_m^2(\rho))$. Hence in the absence of further information (assuming oversampling and positivity), the conventional methods tend to minimize the negative log (NL) likelihood (i.e Eq. 8.1). The second term represents the prior functional, designating a global feature of ρ. Given a set of constraints with associated projection operators $\mathbf{P}_{\mathcal{P}_i}$ on convex sets, we can define the corresponding $\epsilon_{\mathcal{P}_i}^2(\rho)$ as $||(\mathbf{I} - \mathbf{P}_{\mathcal{P}_i})\rho||^2$, where we defer further discussion of $\mathbf{P}_{\mathcal{P}_i}$ till §8.2.2. Hence the prior is $\propto \exp(-\sum_{i=1}^{h}\beta_i\epsilon_{\mathcal{P}_i}^2(\rho))$, and thus we seek to minimize the NL posterior equation: $\mathrm{argmin}_\rho\{(\beta\epsilon_m^2 + \sum_{i=1}^{h}\beta_i{\epsilon_i}^2)\rho\}$. To restate, the map $\boldsymbol{\mu}$ minimizes the first term, while gradient descent minimizes the second term, granted $\epsilon_{\mathcal{P}_i}^2(\rho)$ are convex functions.

8.2.2 Constraints

Using a given $\boldsymbol{\mu}(\rho)$, we obtain a *batch* of reconstructions, $\{\rho^z\}$ where $z \in [1, \mathcal{N}_b]$ and $\mathcal{N}_b \approx \sqrt{n}$. As the reconstructed batch represents all we can know about the data through $\boldsymbol{\mu}$, we can interchange $\sqrt{I}$ and $\{\rho^z\}$ in Eq.8.3. Furthermore, assuming independence among the individual reconstructions, the generalized likelihood distribution can be expressed as:

$$P(\{\rho^z\}_{z=1}^{\mathcal{N}_b}|\rho X) \propto \exp(-\sum_{z=1}^{\mathcal{N}_b} w^z {\epsilon_m^z}^2) \tag{8.4}$$

---

[1] Even though they are related through the scattering object, where the correlation among the data points is intractable

Where ϵ_m^z is the error at the z^{th} reconstruction, and w^z is the weight associated with reconstruction ρ^z, based on the ranking of the batch elements (For more details on w^z please see §8.6.1).

There are two types of consistent priors that we introduce here: informative and weakly informative; the symmetry informative prior is inferred directly from the data, while the total and local variation priors simply incorporate empirical facts about natural images.

8.2.2.1 Symmetry

Natural images can be approximated by a set of coefficients $v \in \mathbf{R}^k$ such that $k << n$. Independent systematic studies (by the authors) have shown that phase retrieval is significantly facilitated if the iterate $\rho^{(\nu)}$ is confined to a sparser representation $\mathcal{M}$ such that:

$$\rho \approx \mathcal{M} = \sum_{j=1}^{k} \mathcal{S}_j v_j \quad \mathcal{S}_j \text{ is the } j^{th} \text{ segment} \tag{8.5}$$

The robustness to this approximate representation to noise, in retrieving the phases, can be explained in terms of phase error further discussed in §8.6.3. The symmetry constraint [2] consists of two operations: inferring $\mathcal{M}$ and refining $\rho^{(\nu)}$.

To infer $\mathcal{M}$, we first extract a template $\rho^\star$ from the batch, using the Bayesian update equation:

$$P(\rho^\star|\{\rho^z\}_{z=1}^{\mathcal{N}_b}X) \propto P(\{\rho^k\}_{z=1}^{\mathcal{N}_b}|\rho^\star X)P(\rho^\star|X) \tag{8.6}$$

---

[2] the name will become obvious by the end of this section.

Where the likelihood can be obtained as in Eq.8.4. Since the spatial complexity is proportional to the change points, we seek a prior that minimizes the isotropic total variation semi norm of $\rho^\star$ i.e $\propto \exp^{-TV(\rho^\star)}$. As a result, the convex optimization problem can be expressed as:

$$\underset{\rho^\star}{\operatorname{argmin}} \sum_{z=1}^{\mathcal{N}_b} w^z ||\rho^z - \rho^\star||_{\mathcal{F}}^2 + 2\lambda TV(\rho^\star) \tag{8.7}$$

$\rho^\star$ is obtained using the Fast Gradient Algorithm (FGP) of Beck and Teboulle[93]. Subsequently, $\mathcal{M}$ is solved by k means segmentation of $\rho^\star$:

$$\underset{\mathcal{S}}{\operatorname{argmin}} \sum_{j=1}^{k} \sum_{i \in \mathcal{S}_j} ||\rho_i^\star - \mu_j||^2 \quad \text{where} \quad \mu_j =< \rho^\star >_{i \in \mathcal{S}_j} \tag{8.8}$$

Where $i \in [1, n]$. The value μ_j is of little concern to us, as we are mainly interested in how the pixels $\{i\}$ are associated. As we will see in the subsequent sections, the modulus constraint, in conjunction with the estimator will enhance accurate value convergence within the segments $\mathcal{S}_j$.

To estimate $\rho^{(\nu)}$, we model $\rho_i^{(\nu)} \in \mathcal{S}_j$ as linearly corrupted with Gaussian iid noise (Please see §8.6.2 for more details):

$$\rho_i^{(\nu)} \approx v_j + \xi_i^{(\nu)} \qquad \forall \ i \in \mathcal{S}_j \tag{8.9}$$

ξ accounts for the unknown error in the piece-wise approximation of ρ. In this case, the maximum likelihood estimate is given by $< \rho_i >_{i \in \mathcal{S}_j}$. Hence the symmetry projection operator $\mathbf{P}_{\mathcal{S}}$ on the set $\{\mathcal{S}_j\}$ yields:

$$\mathbf{P}_{\mathcal{S}} \rho =< \rho_i >_{i \in \mathcal{S}_j} \tag{8.10}$$

One can verify the idempotency of $\mathbf{P}_{\mathcal{S}}$, by the simple observation that the average of the average is the average. Note that $\mathbf{P}_{\mathcal{S}}$ applies equally well inside the support as well as outside. In fact, we have previously introduced smoothing outside the support as a heuristic operation in [35]. However, this operation can be derived more formally by projecting ρ onto the most likely prior knowledge $\mathcal{M}$, where $\mathcal{M}$ includes the oversampled region, as an additional segment with a smooth noise function.

$\mathcal{M}$ gradually improves over the course of the algorithm (§8.2.3); and thus at the onset, edge pixels remain unconstrained by $\mathbf{P}_{\mathcal{S}}$, due to their uncertain classification. Given the fact that most natural images are smooth, we enforce a smooth transition across the segments' edges; having previously obtained $\rho^{\star}$, we can further constraint its isotropic TV semi norm, i.e $c \approx TV(\rho^{\star})$. Thus we introduce the following inner optimization problem:

$$\min_{\tilde{\rho}} ||\rho^{(\nu)} - \tilde{\rho}||^2_{\mathcal{F}} \quad \text{s.t} \quad TV(\tilde{\rho}) \leq c \tag{8.11}$$

Where we associate a projection operator with the solution $\tilde{\rho}$ (Please see §8.6.4 for details on solving Eq.8.11):

$$\mathbf{P}_{\mathcal{TV}}\rho^{(\nu)} = \tilde{\rho} \tag{8.12}$$

8.2.2.2 Local variation

The smooth edge constraint can also be enforced on hot pixels. At the onset of the iterations, the values at each pixel are dynamically varying. To detect a hot pixel i, we not only check the pixel value but also its gradient. Hence i will satisfy the following tests:

$$\begin{aligned} \rho_i &> 2\nabla\rho_i \\ |\rho_i - \mu| &> 4\sigma \end{aligned}$$

μ and σ are the average and the standard deviation of the iterate $\rho^{(\nu)}$; computed at each iteration. Here also we associate a projection operator on the hot pixel:

$$\mathbf{P}_{\mathcal{LV}}\rho_i^{(\nu)} =< \rho_j^{(\nu)} >_{j\in\mathcal{B}} \tag{8.13}$$

where $\mathcal{B}$ are the neighboring pixels .

8.2.3 The APR algorithm

APR takes advantage of the non-uniqueness and degeneracy observed in IPPRA for the noisy incomplete measurements, and extract statistical information which are then incorporated into the projection based algorithm as a form of an empirical Bayes prior. This bayesian prior is obtained from the reconstruction batch using machine learning as described in §8.6.1. This approach is also generalized to various domains, as we will see in §8.4.

The overall algorithm can be summarized in two steps:

- Repeat until convergence ($\min \epsilon_m^2$):
  1. Fix constraints, and reconstruct using the map in Eq. 8.2
  2. Infer constraints, as described in §8.2.2

With each iteration, $\mathcal{M}$ and c are refined and as a result ϵ_m^2 decreases. An increase in ϵ_m^2 signals a re-adjustment for the number of segments k or amount of edge pixels; hence, the adaptivity of the algorithm is anchored by ϵ_m^2. Figure 8.1

(ϵ_k 0.0576,ϵ_r 0.3010) (ϵ_k 0.0500,ϵ_r 0.1745) (ϵ_k 0.0449,ϵ_r 0.1512) (ϵ_k 0.0407,ϵ_r 0.1313)

(ϵ_k 0.0352,ϵ_r 0.1056) (ϵ_k 0.0297,ϵ_r 0.0757) (ϵ_k 0.0272,ϵ_r 0.0587) (ϵ_k 0.0261,ϵ_r 0.0448)

(ϵ_k 0.0261,ϵ_r 0.0429) (ϵ_k 0.0245,ϵ_r 0.0254) (ϵ_k 0.0245,ϵ_r 0.0254) (ϵ_k 0.0245,ϵ_r 0.0255)

(ϵ_k 0.0246,ϵ_r 0.0255) (ϵ_k 0.0245,ϵ_r 0.0238) (ϵ_k 0.0245,ϵ_r 0.0237) (ϵ_k 0.0245,ϵ_r 0.0237)

Figure 8.1: This panel shows the evolution of the reconstructions as the algorithm evolves. The top left figure corresponds to the best HIO reconstruction, the remaining figures correspond to reconstructions using prior knowledge that is updated through the refinement of $\mathcal{M}$.

shows a typical performance of the algorithm, where at each cycle the improved $\mathcal{M}$ causes the reconstruction to become more refined.

8.3 Numerical results

The geometrical distribution of the diffraction data, i.e polar or random distribution (Please see figures 8.2, 8.3), impacts the reconstruction quality using IPPRA, whereas APR achieves consistent results in both cases. Data in polar geometry contains more information, as the crucial low resolution data are preserved, unlike in a random distribution where missing data is extended uniformly. The purpose

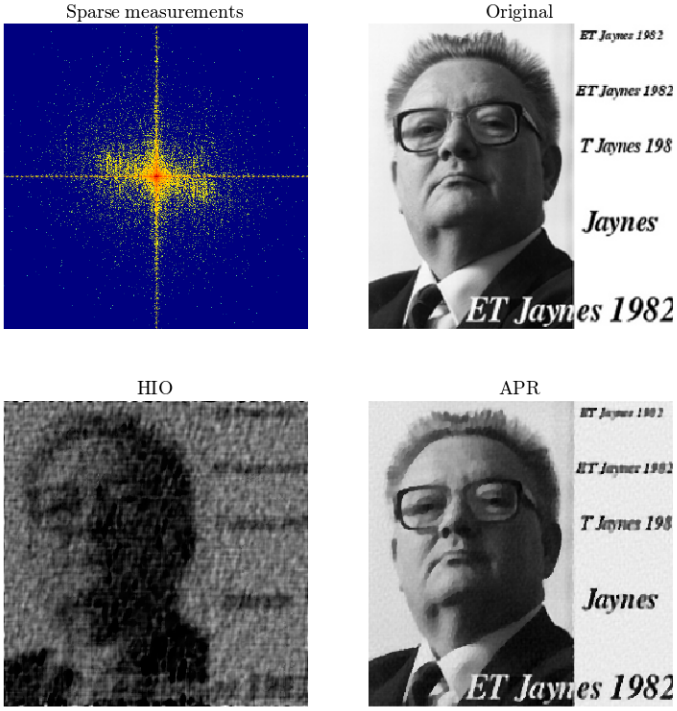

Figure 8.2: Using the sparse measurements, (top left figure), as the simulated data with an oversampling degree of 0.23, we obtain the HIO output (bottom left), while the output of the APR algorithm (bottom right) is in striking agreement with the original image (top right).

for exploring random geometry, is to analyse extreme undersampling case, beneficial for the $3D$ case. With APR, coherence among the reconstruction becomes stronger, as the addition of prior knowledge, restrict the search space, as can be seen in figure 8.2.

8.4 Discussion

APR framework is fairly general and works for more constraints that the two we introduce above. For example, we have successfully implemented these methods using wavelet and atomicity constraints. We only show the wavelets results for

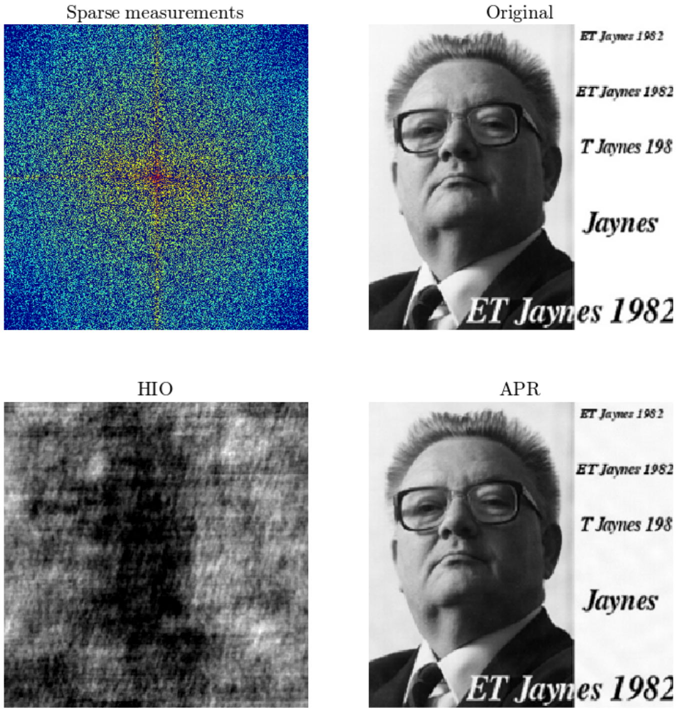

Figure 8.3: Here the measurements are randomly distributed, with an oversampling degree of 1.49. The HIO output is chaotic, while the APR output is almost undistinguishable from the original image.

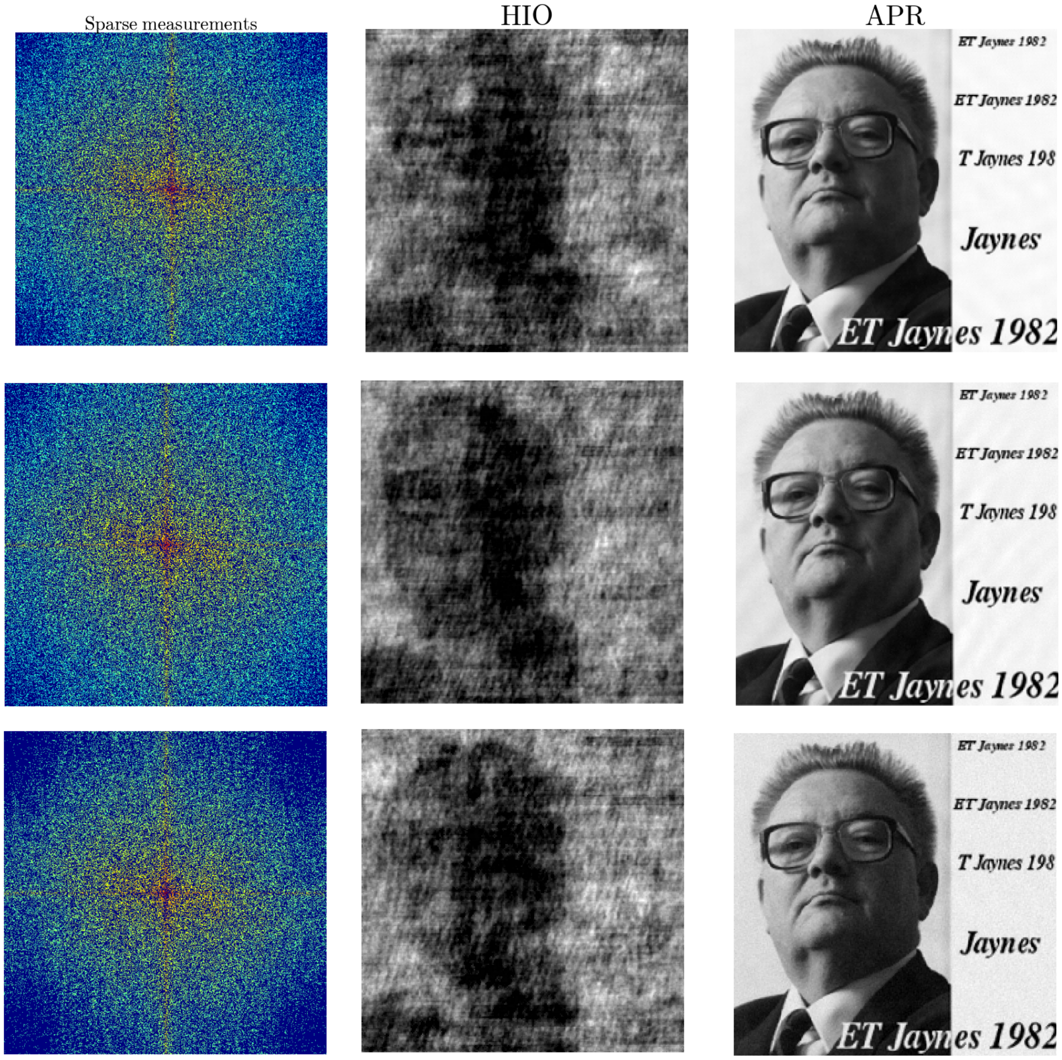

Figure 8.4: Here we test the robustness of APR in the case of noisy data. Top row shows oversampling degree of 1.49 noise less. Middle row shows data with 1% noise. Bottom row with 6% noise. In all these case HIO output is chaotic, while the APR output is robust and consistent.

brevity. The systematic numerical results shows the robustness of this new constraint in high noise, under sampled data. The numerical studies are performed on 2D objects, while eventually we will apply these methods to three dimensional (3D) data, which encounters noisy, incomplete data [37, 38, 94, 40], primarily due to radiation sensitivity, finite flux, and practical experimental concerns. While the extreme undersampled 2D cases shown in this manuscript form a good testing ground, the generalization to 3D will be the topic of subsequent research.

8.4.1 Generalization to other domains

Once we obtain $\mathcal{M}$, we can use this information in other domains. For example, we can compute the corresponding wavelets coefficients in the wavelet domain, and extract another map in that domain as explained below:

8.4.1.1 Wavelet domain

In general, the sparsity of an image is not restricted to the object domain. Here we explore the advantage of wavelet domain sparsity which can be implemented simply based on $\mathcal{M}$. As $\mathcal{M}$ represents an improved reconstruction, we can extract the region where the wavelet coefficients are small, but first, it is important to choose a wavelet that is orthogonal so as to prevent leakage.

$$\rho_{\mathcal{W}} = \mathcal{W}\rho \tag{8.14}$$

$$\mathcal{W}^T\mathcal{W} = \mathbf{I} \tag{8.15}$$

$\tilde{\mathbf{P}}_{\mathcal{W}}, \mathbf{P}_{\mathcal{W}}$ projection operators in the wavelet and the object domains.

$$\mathbf{P}_{\mathcal{W}} = \mathcal{W}^T\tilde{\mathbf{P}}_{\mathcal{W}}\mathcal{W}$$

$$\tilde{\mathbf{P}}_{\mathcal{W}}\rho_{\mathcal{W}_{ij}} = \begin{cases} \rho_{\mathcal{W}_{ij}} & i,j \in \gamma_{\mathcal{W}} \\ 0 & \text{otherwise} \end{cases}$$

Where $\gamma_{\mathcal{W}}$ refers to the spatial region of coefficients of interest. Here we also introduce the corresponding error metric:

$$\epsilon^2_{\mathcal{W}}(\rho) = ||(\mathbf{I} - \mathbf{P}_{\mathcal{W}})\rho||_2^2 \tag{8.16}$$

8.5 Summary

By using Bayesian inference we inferred information from conventional IPPRA outputs, and subsequently applied this information as priors within the phasing algorithm. We were able to determine structures from significantly fewer measurements than by standard methods. We anticipate the framework presented here will have a significant impact on solving the phase problem in the case of noisy incomplete data.

8.6 Appendix

8.6.1 The batch filter

Some reconstructed images from Fourier modulus ($\sqrt{I}$) exhibit the twin image characteristic. This feature is an entangled state between a local solution and its flipped image, where the orientation dominance of the two possible states fluctuates. While visually detecting these images is possible, the associated error metric (ϵ_m) does not always indicate their presence. Here we implement an automatic guide that removes twin images, and failed reconstructions from the reconstruction batch, and subsequently rank the individual reconstructions in terms of their deviation from a reference (ρ^{ref}).

Given the batch, we obtain an aligned set by minimizing the deviation from ρ^{ref} whose error metric:

$$\epsilon_m^{\rho^{ref}} = \min\{\epsilon_m^{\rho^j}\}_{j\in\mathcal{N}_b} \tag{8.17}$$

To find a better reference, we segment the batch elements using k-means segmentation with k = 2 centroids [3]. Then choose a newer reference, which yields the lowest ϵ_m, after segmentation. This method of updating the reference is more robust than simply choosing the reconstruction with the smallest ϵ_m. To select the set of reconstructions $\{\rho^j\}_{j\in\mathcal{N}_{b'}}$ that are within the neighbourhood of ρ^{ref}, we compute the mean square error deviation MSE_j of the discrete wavelet transform of $\{\rho^j\}_{j\in B'}$ from ρ_{ref}, as:

$$MSE_j = \sum_{wl=0}^{n_{wl}} |\mathcal{W}\rho^{ref} - \mathcal{W}\rho^j|_{wl}^2 \tag{8.18}$$

The wavelet level l is the decomposition that exhibits the largest standard deviation across the batch, which designates the scale level at which the reconstructions diverge most, and hence it is the scale of choice when differentiating among reconstructions. n_{wl} is the number of coefficients at l. We disregard reconstructions $\{\rho^j\}$ such that $MSE_j > \sigma_{MSE}$, from further analysis.

With the reconstructions ranked from ρ^{ref}, the weight ω_j are assigned as follows:

$$\omega_j = \frac{1}{\mathcal{N}_b'} \sum_{j=1}^{\mathcal{N}_{b'}} \frac{1}{j} \tag{8.19}$$

Figures in 8.5 show a typical performance of the batch filter. Here we only show 16 elements from the batch.

---

[3] The more centroids the better the outcome, as long as the uncertainty in these distributions is low.

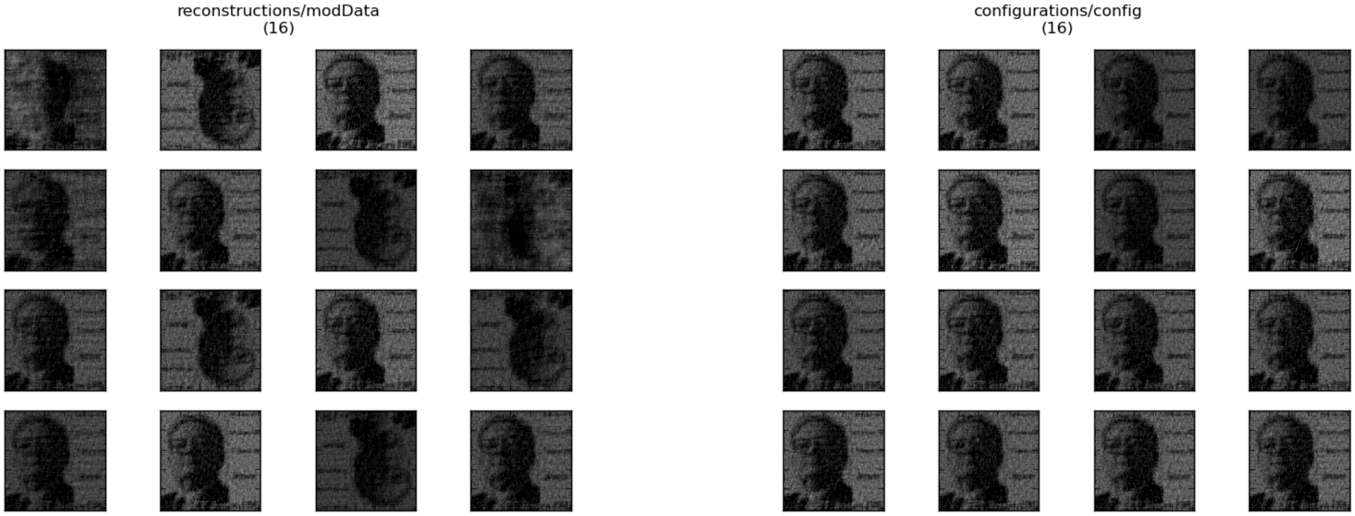

Figure 8.5: Left: The first 16 elements of a batch. Right: The first 16 elements of the filtered batch. Notice how the batch filter eliminated the twin image problem. In particular, the first figure at top row and the last figure in the second row are both disregarded from further analysis.

8.6.2 Noise in the batch

Figures in 8.6 plot the noise distribution for random pixels chosen from the image, across the batch. The histogram distribution is an exponential distribution, that is approximately Gaussian. The blue dot is the location of the zero noise which is off-center and hence a simple averaging of the batch elements is not a sufficient method for denoising. The yellow dot is the location of the reference reconstruction.

Figure 8.7 shows the histogram for the noise distribution within a reconstructed image $\tilde{\rho}$. This plot shows a zero mean Gaussian distribution which encourages the use of simple averaging (Eq. 8.2.2.1), as a form of denoising.

8.6.3 Phase error

The propagated error η due to an inaccurate $\mathcal{M}$, can be evaluated by examining the phase and modulus errors: $\Delta\phi$, Δm. Let us assume η is an additive noise in

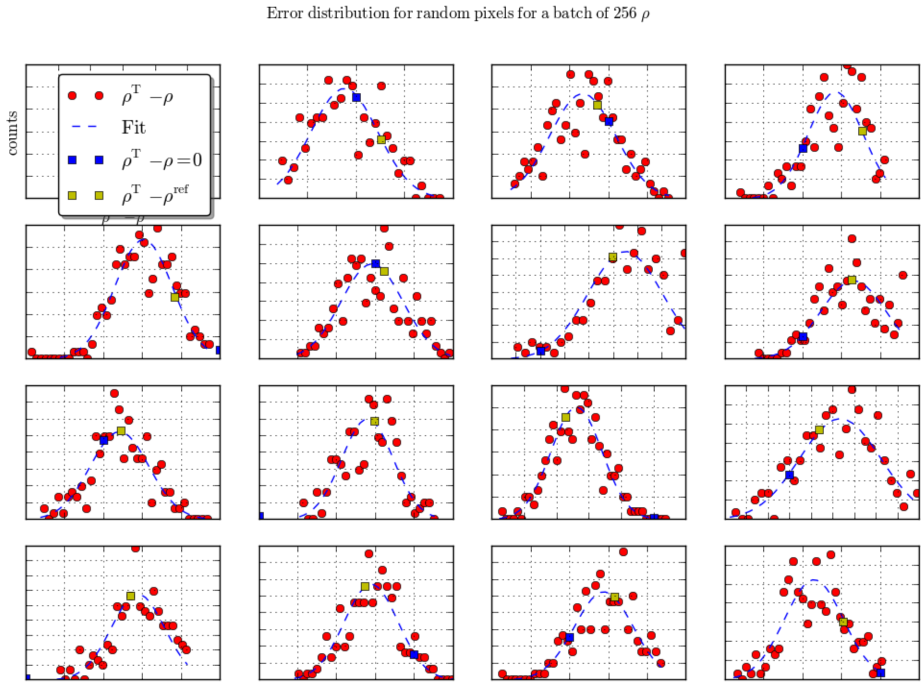

Figure 8.6: Histogram of randomly chosen pixels across multiple reconstructions (256). Blue dot is the location of the zero mean. The dashed line is the Guassian fit. The yellow dot is the location of the reference reconstruction.

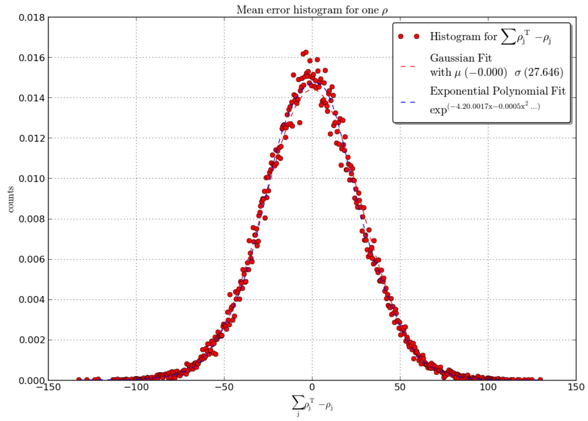

Figure 8.7: Error histogram of $\rho - \tilde{\rho}$. A zero mean Gaussian noise distribution.

$\mathcal{M}$,then:

$$\mathcal{M} = \rho^T + \eta \tag{8.20}$$

The phase at position k is:

$$\begin{aligned}
\phi_k &= \tan^{-1} \frac{\sum_{i=1}^n \mathcal{M}_i \sin\theta_{ik}}{\sum_{i=1}^n \mathcal{M}_i \cos\theta_{ik}} \qquad \text{where } \theta_{ik} = \frac{2\pi i k}{n} \\
&= \tan^{-1} \frac{\sum_{i=1}^n \rho_{T_i} \sin\theta_{ik} + \sum_{i=1}^n \eta_i \sin\theta_{ik}}{\sum_{i=1}^n \rho_{T_i} \cos\theta_{ik} + \sum_{i=1}^n \eta_i \cos\theta_{ik}} \\
&= \tan^{-1}[\tan(\phi_{T_k}) \frac{1+\hat{\epsilon}_\eta}{1+\hat{\epsilon}'_\eta}] \\
\text{where } \hat{\epsilon}_\eta &= \frac{\sum_i \eta_i \sin\theta_{ik}}{\sum_i \rho_{T_i} \sin\theta_{ik}},\ \hat{\epsilon}'_\eta = \frac{\sum_i \eta_i \cos\theta_{ik}}{\sum_i \rho_{T_i} \cos\theta_{ik}}
\end{aligned}$$

Thus, we have:

$$\begin{aligned}
\phi_k \approx \phi_{T_k} \quad & \text{for } \epsilon_\eta, \epsilon'_\eta << 1 \text{ or } \epsilon_\eta = \epsilon'_\eta \\
\text{and} \quad \Delta\phi = \phi_k - \phi_{T_k} \approx 0 \quad & \text{as long as } \eta << \rho_T
\end{aligned} \tag{8.21}$$

And the corresponding modulus error from:

$$\begin{aligned}
|\hat{\mathcal{M}}|^2 &= |\hat{\rho}_T|^2 + |\hat{\eta}|^2 + 2\hat{\rho}_T \star \hat{\eta} \\
\Delta m &= |\hat{\eta}|^2 + 2\hat{\rho}_T \star \hat{\eta}
\end{aligned} \tag{8.22}$$

Unlike $\Delta\phi$, the modulus error is magnified by the signal's strength $\hat{\rho}_T$. However, by using the modulus constraint $\mathbf{P}_m$, Δm is eliminated entirely. In summary, both $\Delta\phi$ and Δm, for the case of approximate symmetry are negligible if:

$$||\rho_T - \mathcal{M}|| << \rho_T \tag{8.23}$$

8.6.4 TV Projection Operator

To solve Eq.8.11, we follow the steps outlined in [93], where we preserve the geometrical dimensionality of the image, i.e: $\rho \in \mathbf{R}^{m\times m}$ $(m^2 = n)$, and re-express the TV operation as:

$$TV(\tilde{\rho}) = \max_{p,q\in\mathcal{P}} \mathrm{Tr}(\mathcal{L}(p,q)^T\tilde{\rho}) \tag{8.24}$$

$$TV(\tilde{\rho}) = \max_{p,q\in\mathcal{P}} \mathrm{Tr}(\mathcal{L}(p,q)^T\tilde{\rho}) \tag{8.25}$$

where

$$\mathcal{P} = \{p,q\}\ s.t\ \ (p_{ij}^2 + q_{ij}^2 \leq 1\ ,\ |p_{im}|, |q_{mj}| \leq 1)$$

$$\mathcal{L}_{ij} = p_{ij} + q_{ij} - p_{i-1j} - q_{ij-1}$$

$$p_{ij} = \tilde{\rho}_{ij} - \tilde{\rho}_{i+1j} \qquad q_{ij} = \tilde{\rho}_{ij} - \tilde{\rho}_{ij+1}$$

Hence our goal is to solve the following objective $\mathcal{H}$:

$$\max_{p,q\in\mathcal{P}} \max_{\lambda\geq 0} \min_{\tilde{\rho}\in\mathrm{R}^{m\times m}} (||\rho^{(\nu)} - \tilde{\rho}||_{\mathcal{F}}^2 + \lambda(\mathrm{Tr}(\mathcal{L}(p,q)^T\tilde{\rho}) - c)) \tag{8.26}$$

By differentiate in respect to $\tilde{\rho}$, we find:

$$\tilde{\rho} = \rho^{(\nu)} - \frac{\lambda}{2}\mathcal{L}(Q), \quad Q = (p,q) \tag{8.27}$$

And the reduced objective can be written as:

$$\mathcal{H}(Q,\lambda) = \frac{-||\mathcal{L}(Q)||_{\mathcal{F}}^2}{4} + \lambda(\mathrm{Tr}(\mathcal{L}^T(Q)\rho^{(\nu)}) - c) \tag{8.28}$$

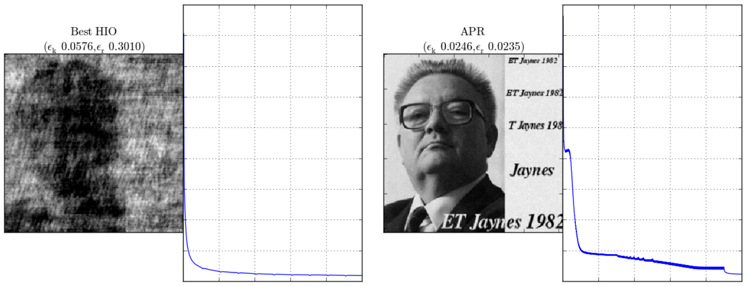

Figure 8.8: Error convergence. (left) HIO. (right) APR. Note the improvement in (b) and the non-stagnating error metric due to the addition of prior knowledge.

Here we compute $\frac{\partial \mathcal{H}}{\partial Q}$ and $\frac{\partial \mathcal{H}}{\partial \lambda}$, to iteratively update Q, λ using gradient ascent with variable steps α_1, α_2, and thus optimizing the corresponding $\mathcal{H}$:

$$\begin{aligned} \lambda &:= \lambda + \alpha_1 \frac{\partial \mathcal{H}}{\partial \lambda} \\ (p, q) &:= (p, q) + \alpha_2 \frac{\partial \mathcal{H}}{\partial Q} \end{aligned} \tag{8.29}$$

From which, $\mathcal{L}$ is updated, and using Eq. 8.27, we find $\tilde{\rho}$.

8.6.5 Error convergence

With the incorporation of prior knowledge, ARP enhances the error convergence, as shown in the right figure of 8.8. The non-stagnation of the reconstruction error, from noisy and incomplete data, demonstrates the success of APR.

Figure 8.9 shows the radial error in the reconstructed images using both HIO and APR.

8.6.6 Additional test objects

Figures 8.10 to 8.21 show a variety of test objects, where we implemented HIO and APR reconstruction algorithms. For each test object, we show two different measurement distributions: Polar and random coordinate. The improvement

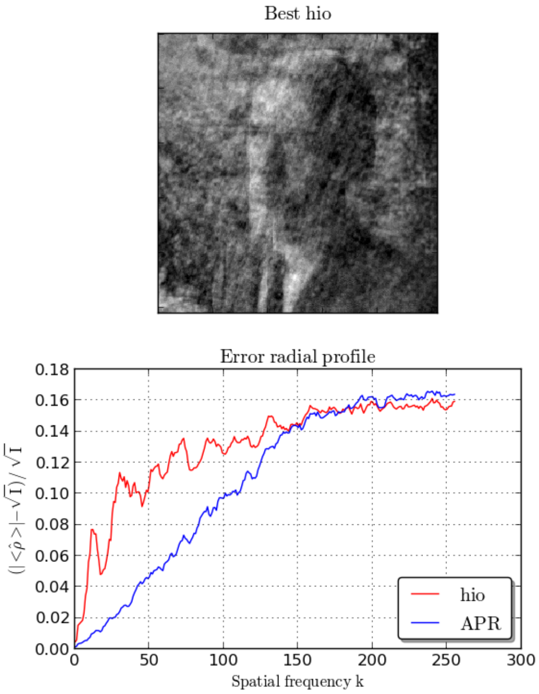

Figure 8.9: Blue HIO, red APR. The radial error profile shows a remarkable improvement at the lower resolution region.

using the APR algorithm is consistent in all cases.

8.6.7 Effect of Image complexity on error tolerance

Using the Schelling segregation model [95], we generated various structures where the complexity is measured by the spatial information within each structure, as shown in the sub-figures of 8.22. To study these structures in low dimension, we confined the images to three values. With the use of the map $\mathcal{M}$ transforming the high dimensional image to low dimension, we were able to generate the corresponding low dimensional energy landscape figure 8.23 (b). By varying the complexity level (a) and the measurement noise (c), we obtain different energy landscapes (b) which can be quantified by one value, ϵ_{ls}^2. ϵ_{ls}^2 corresponds to the error in the energy landscape due to the noise in the measurements, and is computed as follows:

$$\epsilon_{ls}^2 = \frac{\sum |\epsilon_{m0}^2 - \epsilon_m^2|}{\sum \epsilon_{m0}^2} \tag{8.30}$$

where

$$\epsilon_{m0}^2 = \frac{\sum ||\hat{\rho}| - I_{nl}|^2}{\sum I_{nl}^2} \qquad \epsilon_m^2 = \frac{\sum ||\hat{\rho}| - I_n|^2}{\sum I_n^2} \tag{8.31}$$

where I_{nl}, I_n correspond to the noiseless and noisy data, respectively. $|\hat{\rho}|$ is the absolute value of the Fourier transform of a hypothetical density, where the two independent values are varied through the plane (the third value is confined once the two values are chosen). Figure 8.24 shows the condensed information related to the various complexities and noise levels. The pixel values in figure 8.24, across and down the graph correspond to an increase in the measurement noise and a decrease in the image complexity, respectively. This phase diagram confirms that images with high complexity are less tolerant to noise, as the high error on the upper right corner shows.

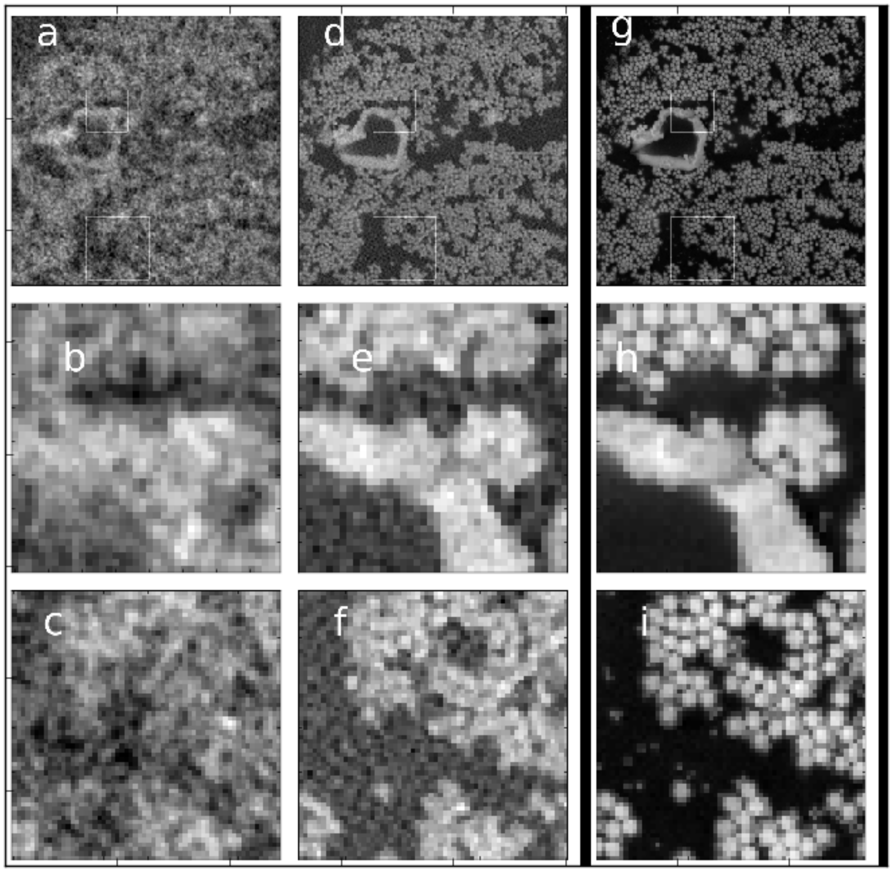

Figure 8.10: Diffraction data collected randomly in cartesian space. a, b, c) HIO reconstruction with two zoom windows d,e,f) The APR reconstruction g, h, i) The original image. The total oversampling degree τ_{osd} is 2.8, with 7% poisson noise.

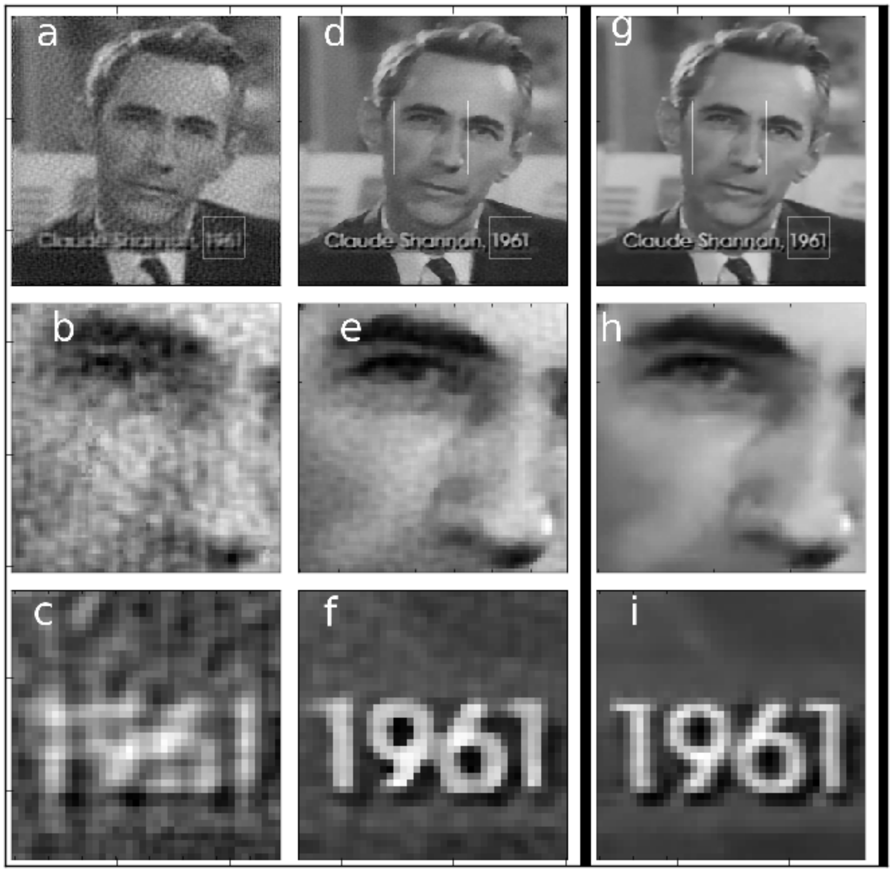

Figure 8.11: Diffraction data collected in polar coordinates with noise level of 4%. a, b, c) HIO reconstruction and two zoom in windows d,e,f) The corresponding APR reconstruction g, h, i) The original image. The total oversampling degree τ_{osd} is 0.97

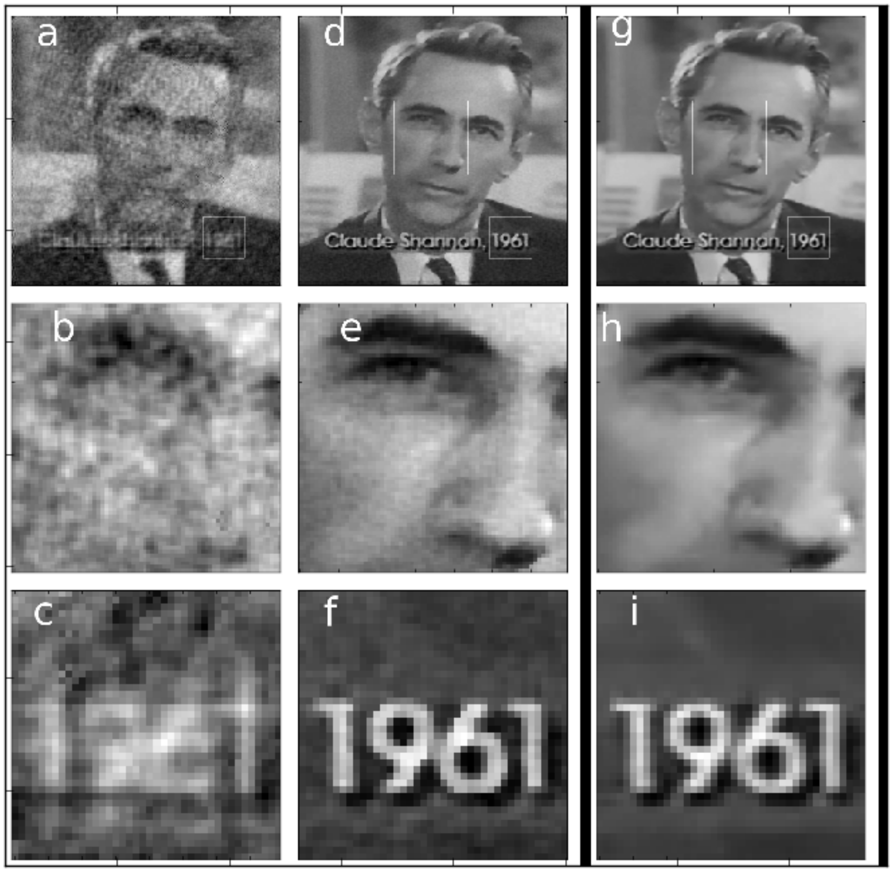

Figure 8.12: Diffraction data collected randomly in cartesian coordinates with 7% noise. a, b, c) HIO reconstruction and zoom in windows. d,e,f) APR reconstruction g, h, i) The original image. The total oversampling degree τ_{osd} is 1.25

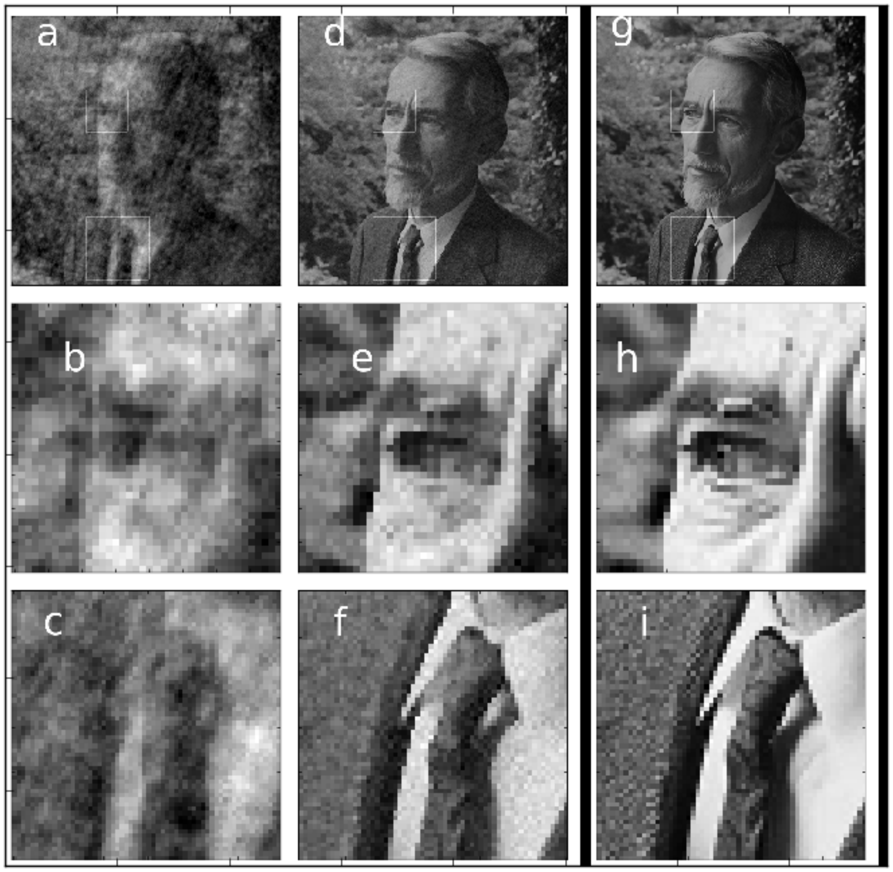

Figure 8.13: Diffraction data collected randomly in cartesian coordinate with 7% noise. a, b, c) HIO reconstruction and two zoom windows. d,e,f) APR reconstruction g, h, i) the original image. The total oversampling degree τ_{osd} is 2.48

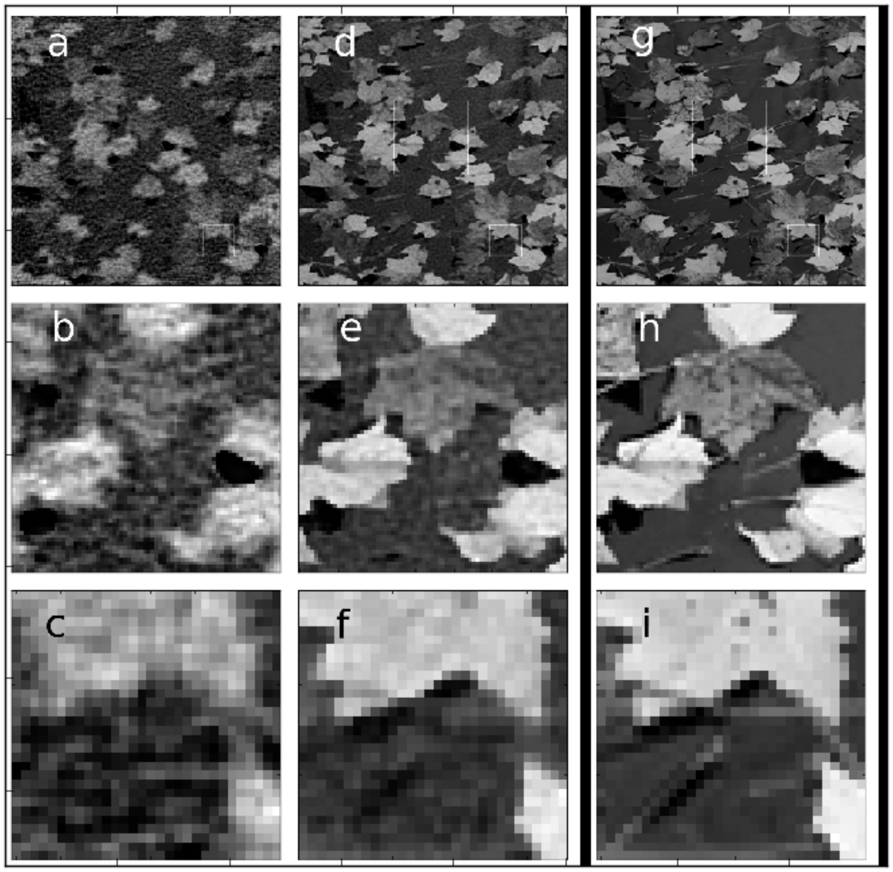

Figure 8.14: Diffraction data collected in polar coordinates with 2% noise. a, b, c) HIO reconstructions and two zoom windows. d,e,f) APR reconstruction g, h, i) the original image. The total oversampling degree τ_{osd} is 1.0

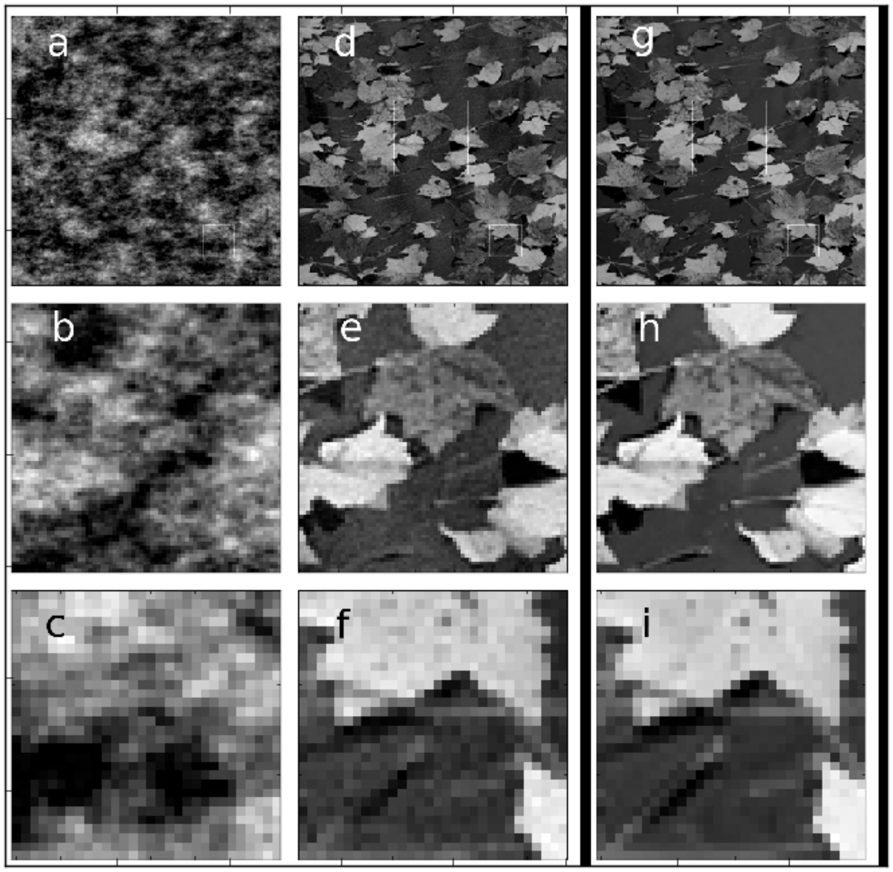

Figure 8.15: Diffraction data collected randomly in cartesian coordinate with 5% noise. a, b, c) HIO reconstruction with two different zooms d,e,f) APR reconstruction g, h, i) the original image. The total oversampling degree τ_{osd} is 1.83

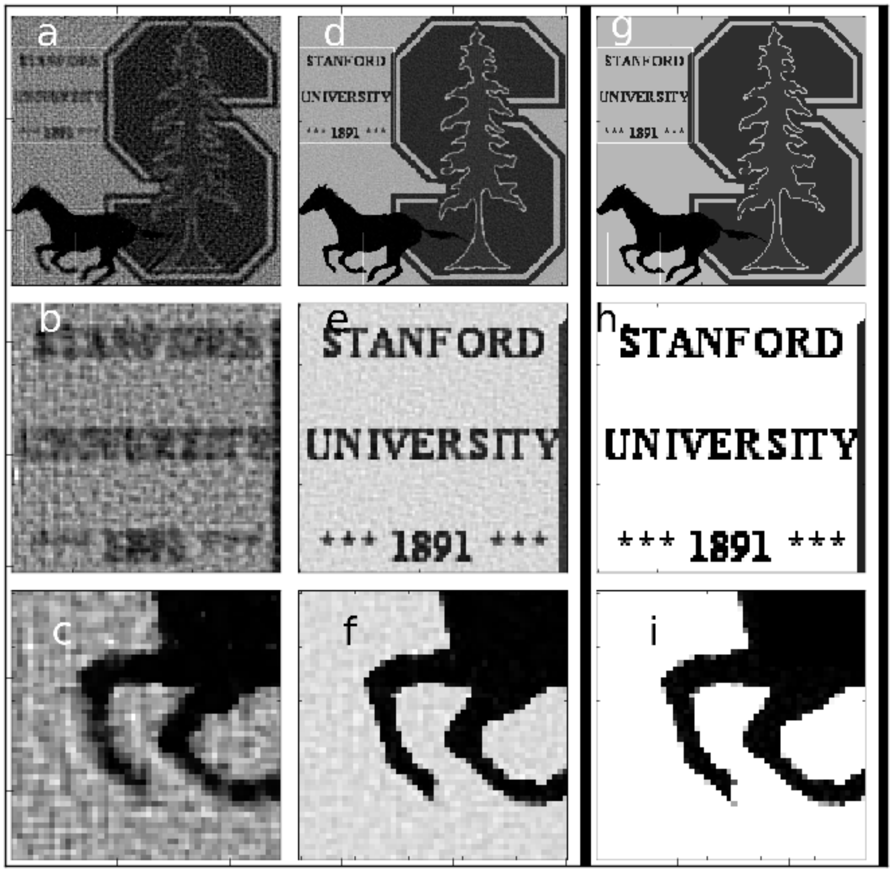

Figure 8.16: Diffraction data collected in polar coordinates with 3% noise. a, b, c) HIO reconstruction with two different zooms. d,e,f) APR reconstruction g, h, i) the original image. The total oversampling degree τ_{osd} is 1.0

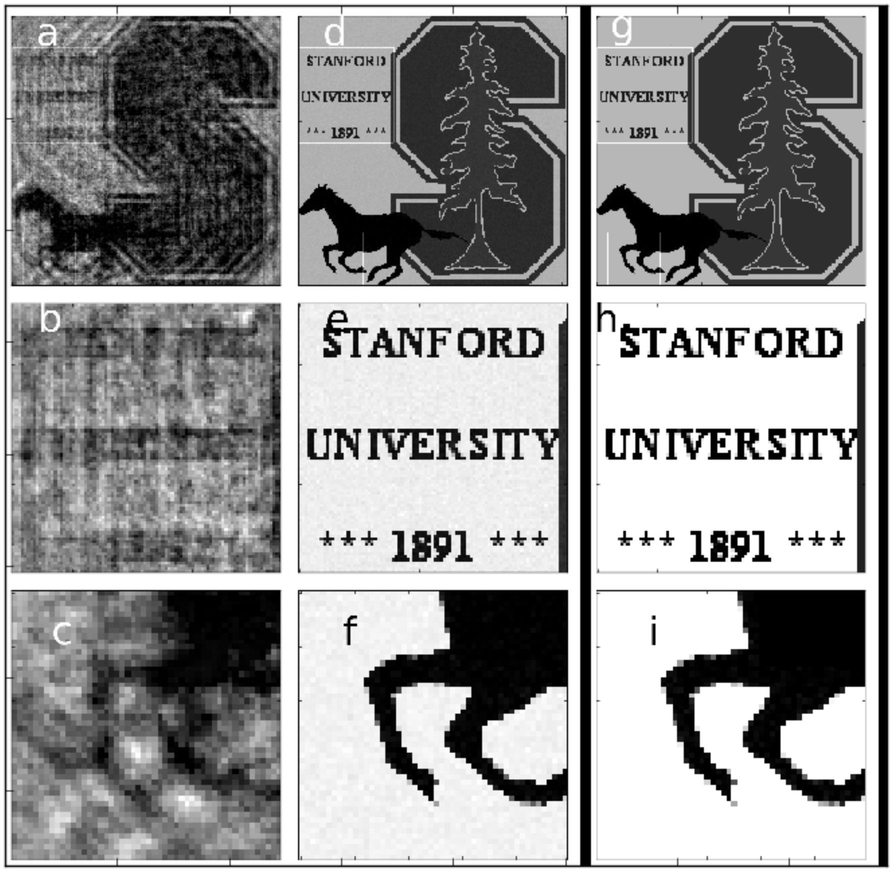

Figure 8.17: Diffraction data collected randomly in cartesian coordinate with 5% noise. a, b, c) HIO reconstruction with two different zoom. d,e,f) APR reconstructions g, h, i) the original image. The total oversampling degree τ_{osd} is 2.32

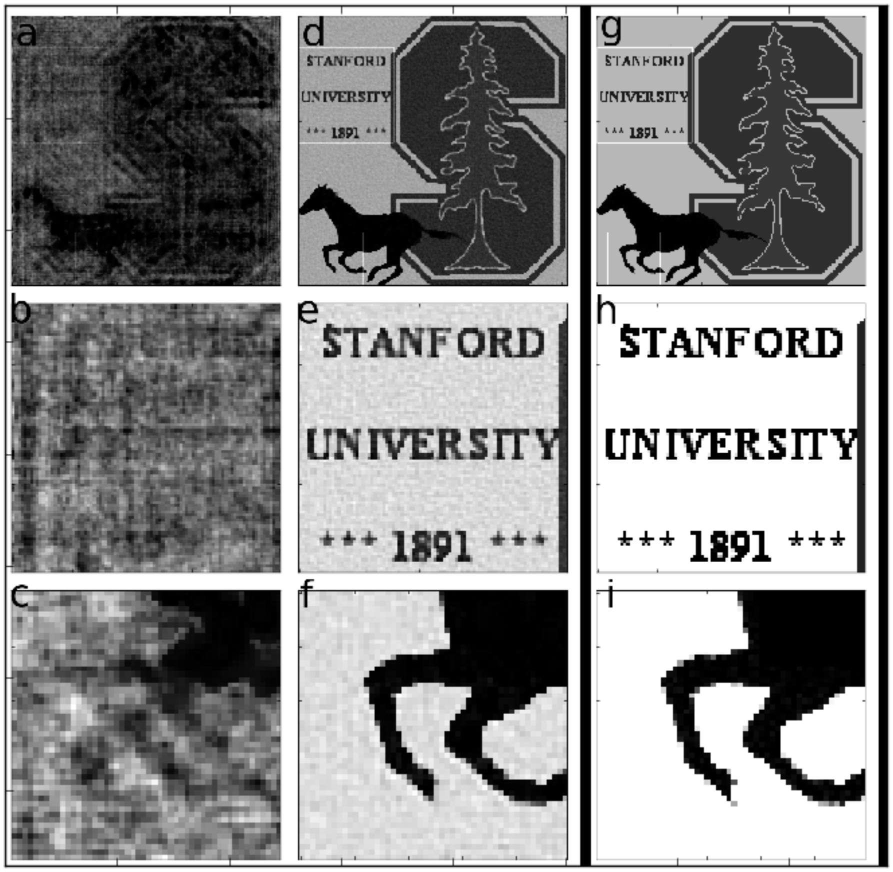

Figure 8.18: Diffraction data collected in polar coordinates with 2% noise. a, b, c) HIO reconstruction with two zoom windows. d,e,f) APR reconstruction. g, h, i) the original image. The total oversampling degree τ_{osd} is 0.74. The linear oversampling is 1.5

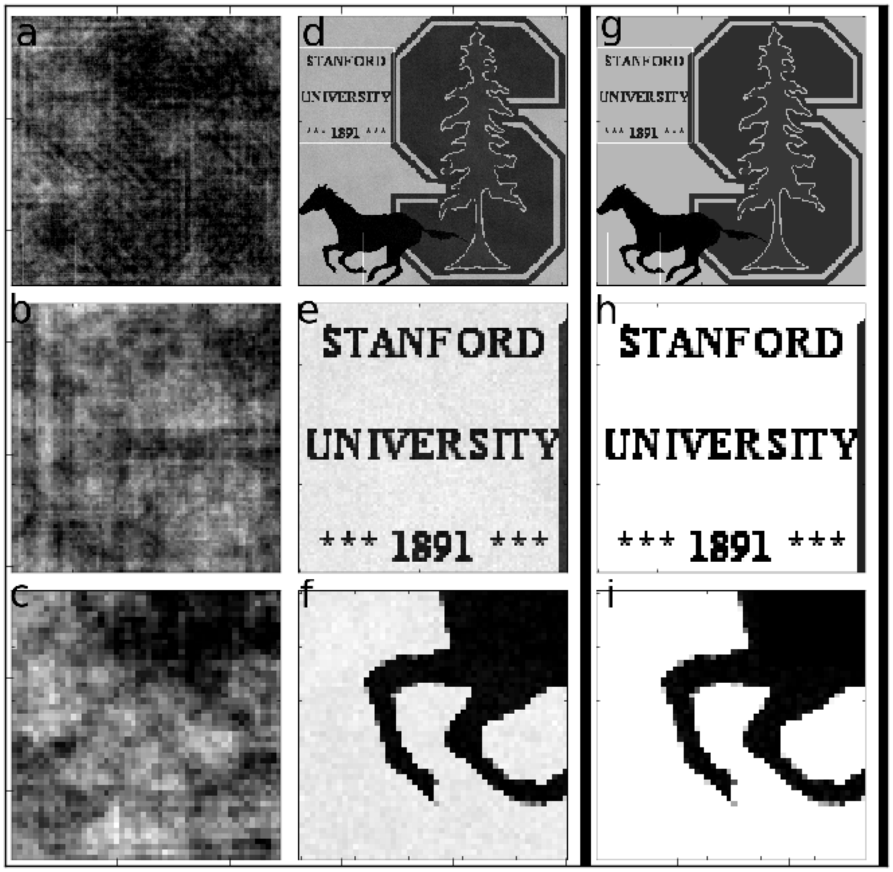

Figure 8.19: Diffraction data collected randomly in cartesian coordinate with 4% noise. a, b, c) HIO reconstruction and zoom in windows. d,e,f) APR reconstruction g, h, i) the original image. The total oversampling degree τ_{osd} is 1.32. The linear oversampling is 1.5.

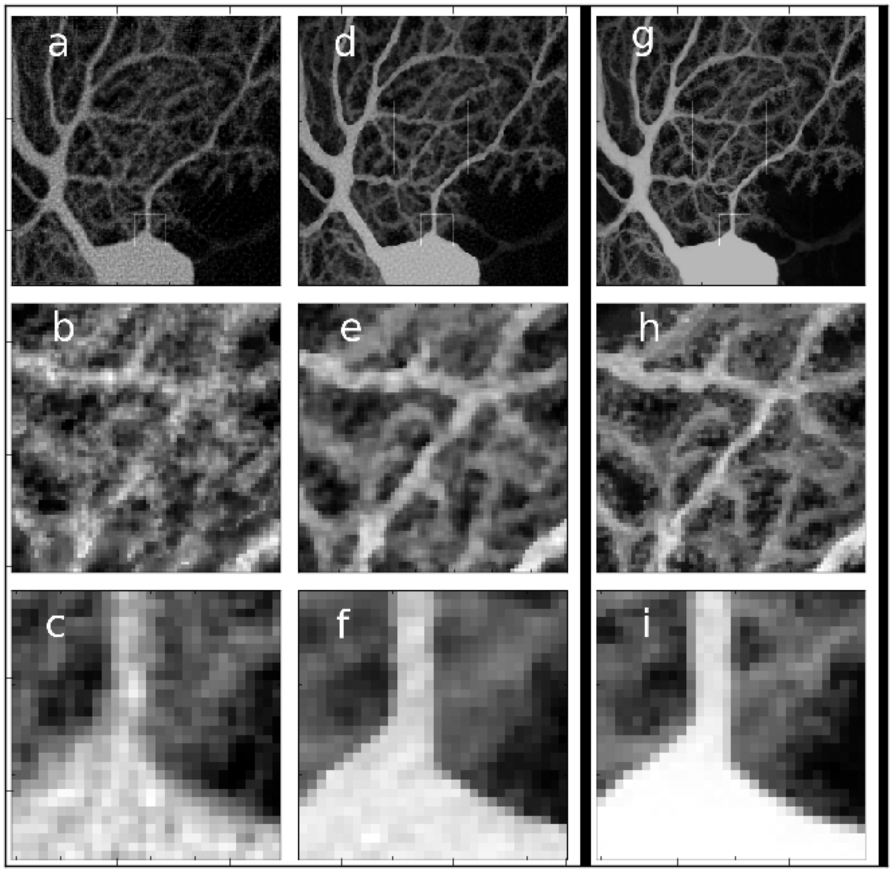

Figure 8.20: Diffraction data collected in polar coordinate with 2% noise. a, b, c) HIO reconstruction with zoom windows. d,e,f) APR reconstructions g, h, i) the original image. The total oversampling degree τ_{osd} is 1.0.

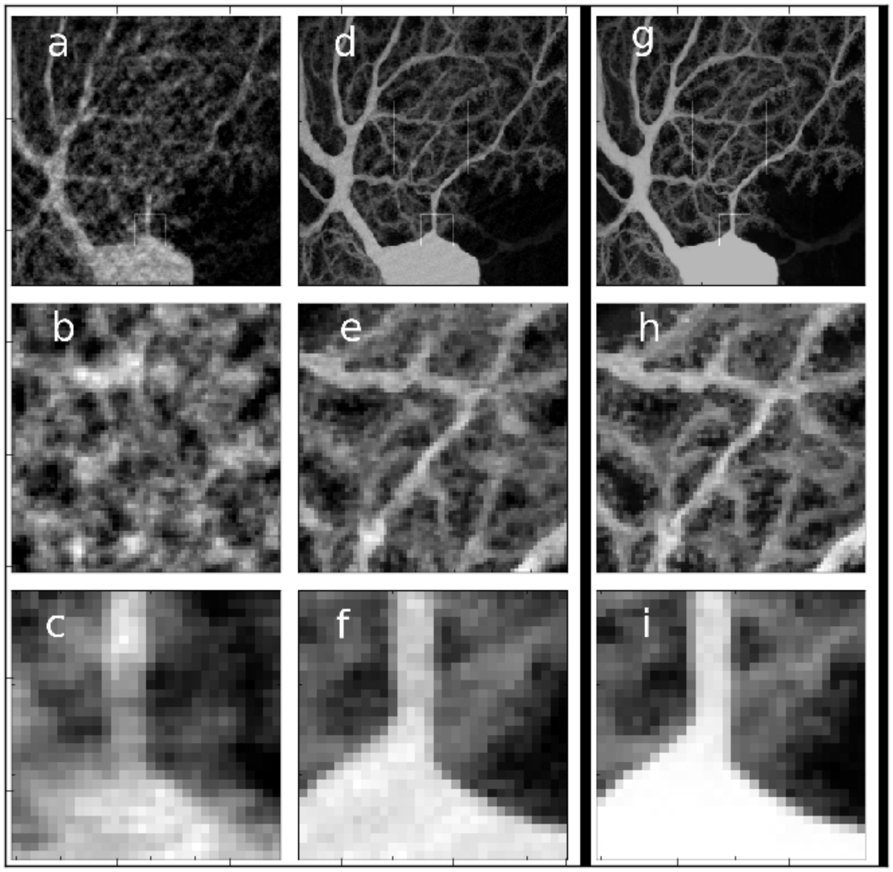

Figure 8.21: Diffraction data collected randomly in cartesian coordinate with 5% noise. a, b, c) HIO reconstruction with zoom windows. d,e,f) APR reconstructions g, h, i) the original image. The total oversampling degree τ_{osd} is 2.64.

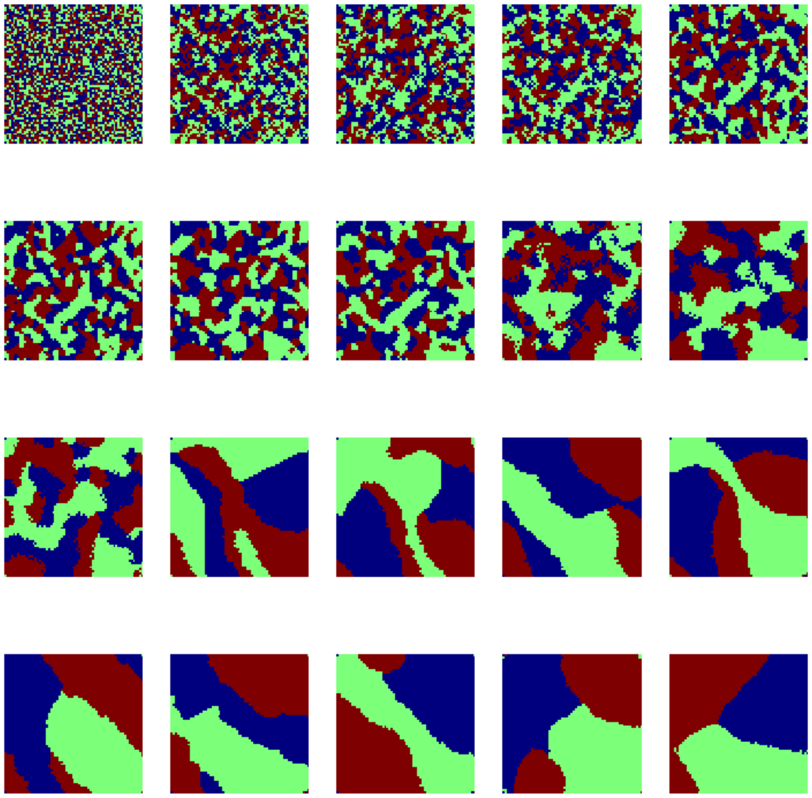

Figure 8.22: Using the Schelling model for segregation, we generated various images with decreasing complexities: Top left is a random partition into three values, bottom right maximum seggregation. The complexity can be assessed visually, or by measuring the change in the sobel gradient. By increasing the size of the neighborhood, the structure becomes simpler.

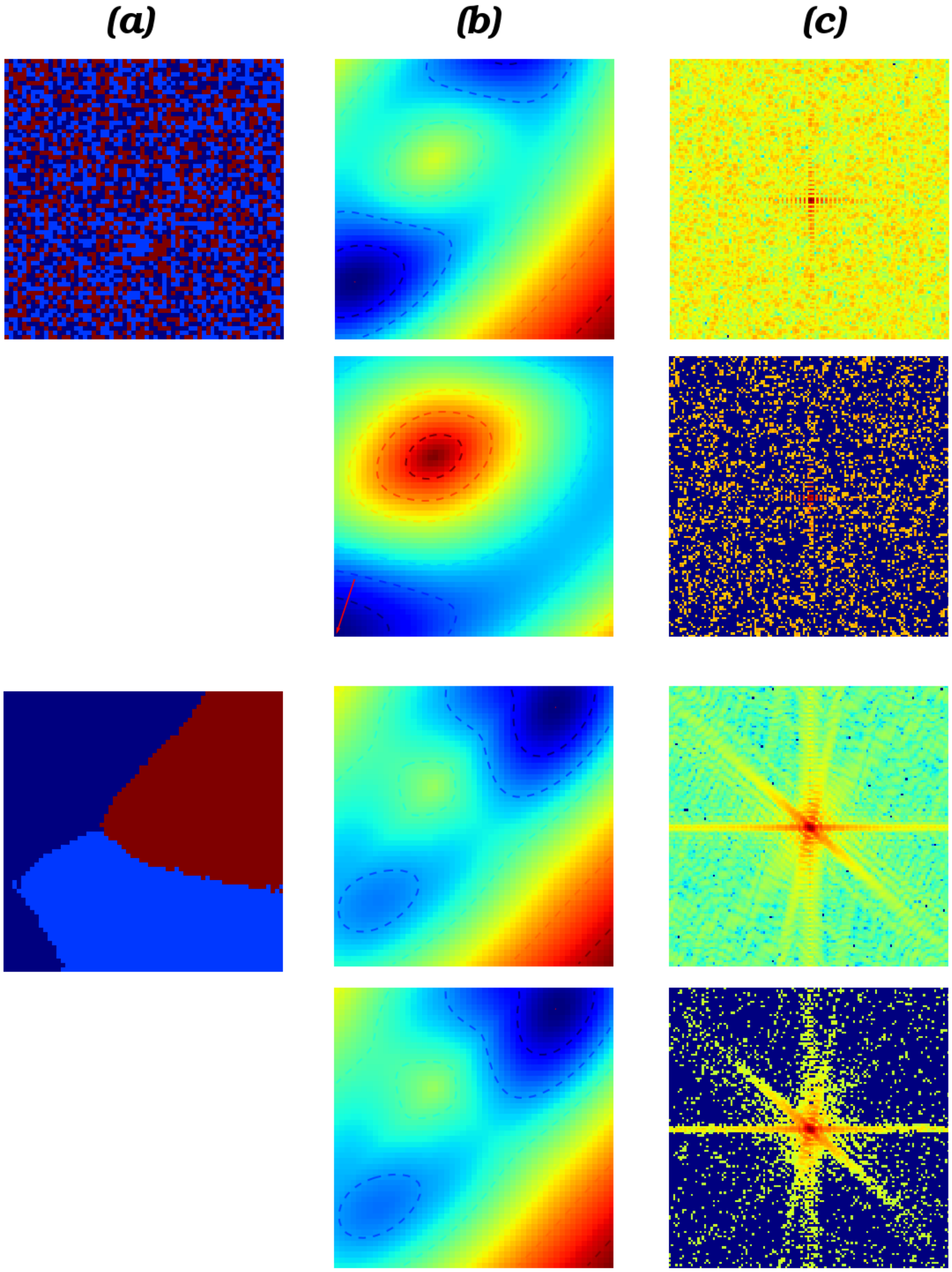

Figure 8.23: In column (a) we have two stuctures, high and lower complexities. The energy landscape for each case (noise and noisless data) (b). The corresponding measurements with varying noise levels (c). In (b) the red arrow designates the shift in the global minimum between noiseless and noisy measurements. Conclusion: For simple structures, the tolerance (as a function of noise level) in the energy landscape is higher than the case of a complex structure.

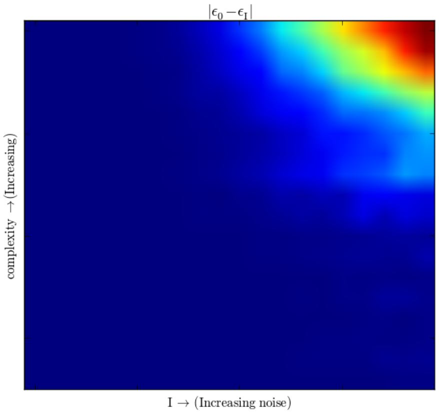

Figure 8.24: Error in the energy landscape ϵ_{ls} (defined in the text), as a function of the measurement noise (horizontal axis) and image complexity (vertical axis). Conclusion: Less complex images are more robust to noise

CHAPTER 9

Conclusion

Coherent diffraction imaging (CDI) by X-rays at third and fourth generation synchrotron radiation facilities has opened a new horizon for three dimensional quantitative high resolution structure determination, with applications ranging from materials science to biology [70, 22, 39, 40]. Limited by the scattering angle and the incident beam energy, the quality of the reconstructed structure is directly affected by the signal to noise ratio of the collected data and the phase retrieval algorithm. To improve the resolution in CDI, we offered two contributions: Ankylography (chapter 4 and 5) and the adaptive phase retrieval algorithm (chapter 8).

Although high numerical aperture (NA) imaging offers higher resolution, the conventional phase retrieval methods suffered from the effects of geometrical aberrations [66]; the data is collected on planar detector, while the diffracted waves lie on a spherical geometry (figure 4.2). We demonstrate Ankylography as a novel imaging modality for high NA data; this new method increases the resolution of the reconstructed structure in three dimensions and offers many new advantages [35, 36].

At the inception of Ankylography, the goal was to enhance the quality of reconstructions obtained from a table top high NA soft x-ray source [66], by adapting to the natural curvature of the data following the kinematic theory of scattering under the Born approximation [41]. By conforming to the geometry of the measurement, the three dimensional structural information is also extracted [35].

Hence, the idea of single shot imaging was immediately conceived as a method to extract three dimensional information from a single diffraction data. With a single shot imaging, the amount of tilt angles is greatly reduced and hence improving the possibility of a single shot diffraction prior to a destruction (Figure 9.1) [32]. To further test this new idea, we conducted systematic simulations where we used simulated pulse from an FEL and an energy recovery linac to reconstruct a sodium silicate glass particle as well as a polio virus structure, from high NA single diffraction pattern. With the excitement of the possibility of three dimensional imaging from a two dimensional diffraction pattern, a buzz in the scientific community emerged. This concept of three dimensional imaging was challenged on two grounds: the experimental data used to demonstrate Ankylography is a tilted aperture, hence not a true 3D object; furthermore, the simulations are obtained on relatively small sized objects, most interesting biological structures occupy a larger volume [96, 97]. To vindicate the first criticism, we performed an optical experiment, where we successfully reconstructed a phase object [36]. To address the dimensionality curse, we re-examined phase retrieval in the case of noisy incomplete two dimensional data.

As the main source of noise in diffraction data is radiation damage, a robust phase retrieval method in the case of noisy, incomplete data plays an important role in the advancement of CDI's range of applications. Hence the second theme of this thesis is phase retrieval in the case of noisy and undersampled data. To this end, we developed novel algorithmic tools that are robust for this case of interest. For extremely undersampled diffraction data, conventional phase retrieval algorithms suffer from a stagnation problem, due to the non-uniqueness of the search space. This stagnation is a direct effect of insufficient information. The approach we offer, directly incorporates prior knowledge according to the fully self consistent theory of Bayesian inference. We developed a framework that in-

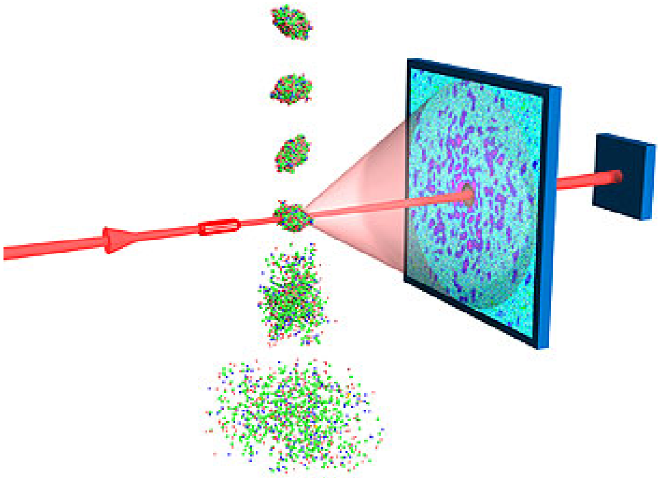

Figure 9.1: Multiple molecules fall through the bright x-ray source. With the short pulse on the order of femtoseconds, it is possible to collect diffraction data before the molecule is destroyed by the intense source. Figure courtesy of Lawrence Livermore National Laboratory.

fers prior knowledge from the data, and subsequently refines the phase retrieval by adding additional navigation metrics. We introduced four navigation metrics: symmetry, wavelet, bound on the total variation and local variation. The symmetry metric was inferred from the data, using convex optimization on the batch of reconstructions obtained with conventional phase retrieval algorithms, as well as machine learning. Using the fast gradient descent algorithm developed by Beck and Teboulle [93], we optimize the output from the reconstruction batch. We obtain a simplified model of the structure, by clustering the output into segments using unsupervised learning algorithm (k-means). The key finding here was that even *one single bit* of information per pixel inside the support is sufficient to obtain accurate reconstruction from extremely undersampled diffraction data. In addition, this symmetry metric is convex, hence a gradient descent optimization method is sufficient to find a solution. The nomenclature of *symmetry* constraint is justified under the approximation of exchange symmetry among pixels within each segment. Hence this constraint has the effect of drastically reducing the dimensionality of the phase problem, which justifies its seemingly

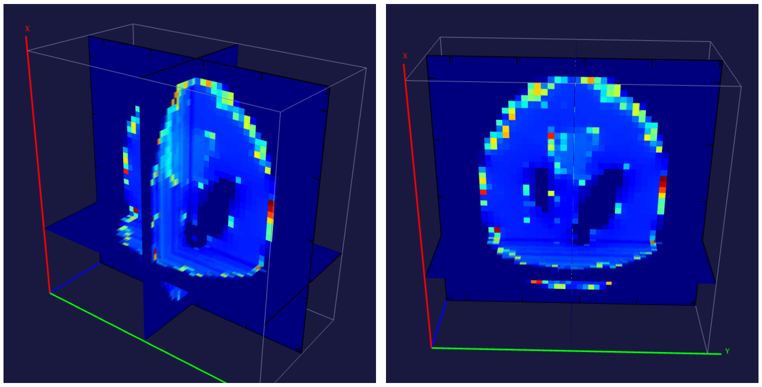

Figure 9.2: 3D reconstruction of sparse shepp Logan, using one curved diffraction pattern at 90°.

unrealistic success. The wavelet constraint is complementary to the symmetry constraint, this method can be implemented in the case of lower resolution image obtained through other imaging modality. While the total variation and local variation constraint are general constraints, and the corresponding functions are convex as well. Hence the adaptive phase retrieval algorithm incorporates these metrics into the conventional algorithm using the gradient terms as additional navigation directions. With APR, we demonstrate a tremendous improvement in the quality of the reconstructions from noisy incomplete data (chapter 8).

Lastly, with the additional complexity that the iterative projection algorithm became engulfed in, an improved computational platform was needed. With the large computational time required at each iteration, due to the addition of the new constraints and the large number of reconstructions performed, it became necessary to optimize the performance of the iterative algorithm and re-write the total variation function in $C++$. Hence the third back stage theme in this thesis, is the incorporation of object oriented programming into the phase retrieval algorithm, allowing the modularization of constraints and hence providing an easier platform for future function updates. We optimized the total variation constraint using ideas from the fast gradient projection of Beck [93]. Although, the algorithmic codes are not included in this thesis, we are in the process of uploading a user friendly version as on open source on the *git* repository. For more information or inquiries about the software, please contact the authors [98].

Due to time and resource constraints, our simulations were only tested on two dimensional data. Although, we obtained preliminary 3D reconstruction of the shepp logan, from a single diffraction pattern Fig. 9.2, fully developed three dimensional algorithmic tool will be a future endeavour. The three dimensional case entails designing three dimensional wavelet and total variation constraints.

By incorporating informative prior into phase retrieval, we re-formulated this *ill posed* inverse problem, where the consistency of our results became a trademark of a *better defined* problem. We anticipate that the adaptive phase retrieval algorithm and in conjunction with the high numerical aperture of Ankylography, will have an impact on single particle imaging; the promise of X-FEL [27, 32].

REFERENCES

[1] M. J. Allen, V. C. Tung, and R. B. Kaner, "Honeycomb carbon: a review of graphene," *Chemical reviews*, vol. 110, no. 1, pp. 132–145, 2009.

[2] E. Abbe, "A contribution to the theory of the microscope and the nature of microscopic vision," *Proceedings of the Bristol Naturalists' Society*, pp. 200–261, 1874.

[3] D. Gabor, "A new microscopic principle," *NATURE*, no. 4098, pp. 777–8, 1948.

[4] D. Gabor, "Microscopy by reconstructed wavefronts," *Proceedings of the Royal Society of London*, vol. 197, no. 1051, pp. 454–487, 1949.

[5] G. Faigel and M. Tegze, "X-ray holography," *Rep. Prog. Phys.*, no. 62, pp. 355–393, 1999.

[6] M. Tegze and G. Faigel, "X-ray holography with atomic resolution," *Nature*, vol. 380, 1996.

[7] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379–423, 623–656, July, October 1948.

[8] C. E. Shannon, "Communication in the Presence of Noise," *Proceedings of the IRE*, vol. 37, pp. 10–21, Jan. 1949.

[9] D. Sayre, "Some implications of a theorem due to Shannon," *Acta Crystallographica*, vol. 5, pp. 843+, Nov. 1952.

[10] A. Snigirev, I. Snigireva, V. Kohn, S. Kuznetsov, and I. Schelokov, "On the possibilities of x-ray phase contrast microimaging by coherent high-energy synchrotron radiation," *Review of Scientific Instruments*, vol. 66, pp. 5486–5492, Dec 1995.

[11] Z. Huang and K.-J. Kim, "Review of x-ray free-electron laser theory," *Phys. Rev. ST Accel. Beams*, vol. 10, p. 034801, Mar 2007.

[12] B. Katherine, "Tabletop x-rays light up," *Nature*, vol. 486, Jun 2012.

[13] R. Gerchberg and W. Saxton, "A practical algorithm for the determination of the phase from image and diffraction plane pictures," *Optik*, vol. 35, no. 237, 1972.

[14] J. Fienup, "Phase retrieval algorithms: a personal tour [invited]," *Applied Optics*, vol. 52, pp. 45–56, 2013.

[15] J. R. Fienup, "Reconstruction of an object from the modulus of its Fourier transform.," *Optics Letters*, vol. 3, no. 1, pp. 27–29, 1978.

[16] J. Miao, P. Charalambous, J. Kirz, and D. Sayre, "Extending the methodology of x-ray crystallography to allow imaging of micrometre-sized noncrystalline specimens," *Nature*, vol. 400, pp. 342–344, 1999.

[17] J. Miao, K. O. Hodgson, T. Ishikawa, C. A. Larabell, M. A. LeGros, and Y. Nishino, "Imaging whole escherichia coli bacteria by using single-particle x-ray diffraction," *Proceedings of the National Academy of Sciences*, vol. 100, no. 1, pp. 110–112, 2003.

[18] D. Shapiro, P. Thibault, T. Beetz, V. Elser, M. Howells, C. Jacobsen, J. Kirz, E. Lima, H. Miao, A. M. Neiman, and D. Sayre, "Biological imaging by soft x-ray diffraction microscopy," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 43, pp. 15343–15346, 2005.

[19] Y. Takayama and M. Nakasako, "Humidity-controlled preparation of frozen-hydrated biological samples for cryogenic coherent x-ray diffraction microscopy," *Review of Scientific Instruments*, vol. 83, no. 5, pp. –, 2012.

[20] Y. Takahashi, N. Zettsu, Y. Nishino, R. Tsutsumi, E. Matsubara, T. Ishikawa, and K. Yamauchi, "Three-Dimensional Electron Density Mapping of Shape-Controlled Nanoparticle by Focused Hard X-ray Diffraction Microscopy," *Nano Letters*, vol. 10, pp. 1922–1926, May 2010.

[21] M. Pfeifer, W. Garth, I. Vartanyants, R. Harder, and I. Robinson, "Three-dimensional mapping of a deformation field inside a nanocrystal," *Nature*, no. 7098, p. 6366, 2006.

[22] J. Miao, C.-C. Chen, C. Song, Y. Nishino, Y. Kohmura, T. Ishikawa, D. Ramunno-Johnson, T.-K. Lee, and S. H. Risbud, "Three-dimensional $gan - ga_2o_3$ core shell structure revealed by x-ray diffraction microscopy," *Phys. Rev. Lett.*, vol. 97, p. 215503, Nov 2006.

[23] B. Abbey, G. J. Williams, M. A. Pfeifer, J. N. Clark, C. T. Putkunz, A. Torrance, I. McNulty, T. M. Levin, A. G. Peele, and K. A. Nugent, "Quantitative coherent diffractive imaging of an integrated circuit at a spatial resolution of 20 nm," *Applied Physics Letters*, vol. 93, no. 21, pp. –, 2008.

[24] A. Barty, S. Marchesini, H. N. Chapman, C. Cui, M. R. Howells, D. A. Shapiro, A. M. Minor, J. C. H. Spence, U. Weierstall, J. Ilavsky, A. Noy, S. P. Hau-Riege, A. B. Artyukhin, T. Baumann, T. Willey, J. Stolken, T. van Buuren, and J. H. Kinney, "Three-dimensional coherent x-ray diffraction imaging of a ceramic nanofoam: Determination of structural deformation mechanisms," *Phys. Rev. Lett.*, vol. 101, p. 055501, Jul 2008.

[25] N. Marcus, L. Steven, H. Ross, and R. Ian, “Three-dimensional imaging of strain in a single zno nanorod,” *Nature Materials*, no. 2, p. 120124, 2010.

[26] J. Miao, K. O. Hodgson, and D. Sayre, “An approach to three-dimensional structures of biomolecules by using single-molecule diffraction images,” *Proceedings of the National Academy of Sciences*, vol. 98, no. 12, pp. 6641–6645, 2001.

[27] J. C. H. Spence and R. B. Doak, “Single molecule diffraction,” *Phys. Rev. Lett.*, vol. 92, p. 198102, May 2004.

[28] M. Howells, T. Beetz, H. Chapman, C. Cui, J. Holton, C. Jacobsen, J. Kirz, E. Lima, S. Marchesini, H. Miao, D. Sayre, D. Shapiro, J. Spence, and D. Starodub, “An assessment of the resolution limitation due to radiation-damage in x-ray diffraction microscopy,” *Journal of Electron Spectroscopy and Related Phenomena*, vol. 170, no. 13, pp. 4 – 12, 2009. Radiation Damage.

[29] X. Huang, J. Nelson, J. Kirz, E. Lima, S. Marchesini, H. Miao, A. M. Neiman, D. Shapiro, J. Steinbrener, A. Stewart, J. J. Turner, and C. Jacobsen, “Soft x-ray diffraction microscopy of a frozen hydrated yeast cell,” *Phys. Rev. Lett.*, vol. 103, p. 198101, Nov 2009.

[30] M. Bogan, H. Benner, S. Boutet, U. Rohner, M. Frank, A. Barty, M. Seibert, F. Maia, S. Marchesini, S. Bajt, *et al.*, “Single particle x-ray diffractive imaging,” *Nano letters*, vol. 8, no. 1, pp. 310–316, 2008.

[31] H. Chapman and et al, “Femtosecond diffractive imaging with a soft-x-ray free-electron laser,” *Nature physics*, vol. 2, pp. 839–843, 2006.

[32] J. Spence, “X-ray imaging: Ultrafast diffract-and-destroy movies,” *Nat Photon*, vol. 2, pp. 390–391, 2008.

[33] M. Matthew, R. Justin, and B. Richard, “Compressive phase retrieval,” vol. 6701, pp. 670120–670120–11, Sept. 2007.

[34] A. Szameit, Y. Shechtman, E. Osherovich, E. Bullkich, P. Sidorenko, H. Dana, S. Steiner, E. B. Kley, S. Gazit, T. Cohen-Hyams, S. Shoham, M. Zibulevsky, I. Yavneh, Y. C. Eldar, O. Cohen, and M. Segev, “Sparsity-based single-shot subwavelength coherent diffractive imaging.,” *Nature materials*, vol. 11, pp. 455–9, May 2012.

[35] K. S. Raines, S. Salha, R. L. Sandberg, H. Jiang, J. A. Rodríguez, B. P. Fahimian, H. C. Kapteyn, J. Du, and J. Miao, “Three-dimensional structure determination from a single view,” *Nature*, vol. 463, no. 7278, p. 30, 2010.

[36] C.-C. Chen, H. Jiang, L. Rong, S. Salha, R. Xu, T. G. Mason, and J. Miao, "Three-dimensional imaging of a phase object from a single sample orientation using an optical laser," *Physical Review B*, vol. 84, no. 22, p. 224104, 2011.

[37] J. Miao, T. Ishikawa, B. Johnson, E. H. Anderson, B. Lai, and K. O. Hodgson, "High resolution 3D x-ray diffraction microscopy.," *Physical Review Letters*, vol. 89, no. 8, p. 088303, 2002.

[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," *Journal of the Optical Society of America A*, vol. 23, no. 5, pp. 1179–1200, 2006.

[39] Y. Nishino, Y. Takahashi, N. Imamoto, T. Ishikawa, and K. Maeshima, "Three-dimensional visualization of a human chromosome using coherent X-ray diffraction.," *Physical Review Letters*, vol. 102, no. 1, p. 018101, 2009.

[40] H. Jiang, C. Song, C.-C. Chen, R. Xu, K. S. Raines, B. P. Fahimian, C.-H. Lu, T.-K. Lee, A. Nakashima, J. Urano, T. Ishikawa, F. Tamanoi, and J. Miao, "Quantitative 3D imaging of whole, unstained cells by using X-ray diffraction microscopy," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, no. 25, pp. 11234–11239, 2010.

[41] M. Born and E. Wolf, *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light.* Cambridge university press, 2002.

[42] B. Henke, E. Gullikson, and J. Davis, "X-ray interactions: Photoabsorption, scattering, transmission, and reflection at e = 50-30,000 ev, z = 1-92," *Atomic Data and Nuclear Data Tables*, vol. 54, no. 2, pp. 181 – 342, 1993.

[43] R. Xu, S. Salha, K. S. Raines, H. Jiang, C.-C. Chen, Y. Takahashi, Y. Kohmura, Y. Nishino, C. Song, T. Ishikawa, and J. Miao, "Coherent diffraction microscopy at SPring-8: instrumentation, data acquisition and data analysis," *Journal of Synchrotron Radiation*, vol. 18, pp. 293–298, Mar 2011.

[44] P. R. Ribic and G. Margaritondo, "Status and prospects of x-ray free-electron lasers (x-fels): a simple presentation," *Journal of Physics D: Applied Physics*, vol. 45, no. 21, p. 213001, 2012.

[45] C. Pellegrini and J. Stöhr, "X-ray free-electron lasersprinciples, properties and applications," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 500, no. 1, pp. 33–40, 2003.

[46] T. Popmintchev, M.-C. Chen, D. Popmintchev, P. Arpin, S. Brown, S. Ališauskas, G. Andriukaitis, T. Baliunas, O. D. Mcke, A. Pugzlys, A. Baltuka, B. Shim, S. E. Schrauth, A. Gaeta, C. Hernndez-Garca, L. Plaja, A. Becker, A. Jaron-Becker, M. M. Murnane, and H. C. Kapteyn, "Bright coherent ultrahigh harmonics in the kev x-ray regime from mid-infrared femtosecond lasers," *Science*, vol. 336, no. 6086, pp. 1287–1291, 2012.

[47] E. Hecht, *Optics*. Pearson, 2014.

[48] B. Saleh and M. Teich, *Fundamentals of photonics*. Wiley, 2014.

[49] D. H. Bilderback, P. Elleaume, and E. Weckert, "Review of third and next generation synchrotron light sources," *Journal of Physics B: Atomic, Molecular and Optical Physics*, vol. 38, no. 9, p. S773, 2005.

[50] D. Attwood, "Spatial and temporal coherence," *Lecture notes*, 2007.

[51] E. Wolf, "Solution of the phase problem in the theory of structure determination of crystals from x-ray diffraction experiments," *Phys. Rev. Lett.*, vol. 103, p. 075501, Aug 2009.

[52] J. R. Fienup, "Phase retrieval algorithms: a comparison.," *Applied Optics*, vol. 21, no. 15, pp. 2758–2769, 1982.

[53] J. Goodman, *Introduction to Fourier Optics*. Ben Roberts, 2005.

[54] S. Marchesini, "Invited article: A unified evaluation of iterative projection algorithms for phase retrieval," *Review of Scientific Instruments*, vol. 78, no. 1, pp. –, 2007.

[55] D. R. Luke, J. V. Burke, and R. G. Lyon, "Reconstruction : Theory and Numerical Methods ," vol. 44, no. 2, pp. 169–224, 2002.

[56] J. R. Fienup, "Iterative method applied to image reconstruction and to computer-generated holograms," *Optical Engineering*, vol. 19, no. 3, pp. 193297–193297–, 1980.

[57] V. Elser, "Phase retrieval by iterated projections," *JOSA A*, vol. 20, no. 1, pp. 40–55, 2003.

[58] D. R. Luke, "Relaxed averaged alternating reflections for diffraction imaging," *Inverse Problems*, vol. 21, no. 1, p. 37, 2005.

[59] I. C. Rankenburg and V. Elser, "Protein structure prediction by an iterative search method," *arXiv preprint arXiv:0706.1754*, 2007.

[60] L. S. Martin, C.-C. Chen, M. D. Seaberg, D. E. Adams, and J. Miao, "Multiple-shell ankylography," in *CLEO: Science and Innovations*, pp. CF3C–6, Optical Society of America, 2012.

[61] C. Song, D. Ramunno-Johnson, Y. Nishino, Y. Kohmura, T. Ishikawa, C.-C. Chen, T.-K. Lee, and J. Miao, “Phase retrieval from exactly oversampled diffraction intensity through deconvolution,” *Physical Review B*, vol. 75, no. 1, p. 012102, 2007.

[62] J. A. Rodriguez, R. Xu, C.-C. Chen, Y. Zou, and J. Miao, “Oversampling smoothness: an effective algorithm for phase retrieval of noisy diffraction intensities,” *Journal of applied crystallography*, vol. 46, no. 2, pp. 312–318, 2013.

[63] J. Du and L. R. Corrales, “Compositional dependence of the first sharp diffraction peaks in alkali silicate glasses: A molecular dynamics study,” *Journal of non-crystalline solids*, vol. 352, no. 30, pp. 3255–3269, 2006.

[64] D. H. Bilderback, P. Elleaume, and E. Weckert, “Review of third and next generation synchrotron light sources,” *Journal of Physics B: Atomic, molecular and optical physics*, vol. 38, no. 9, p. S773, 2005.

[65] D. Bubeck, D. J. Filman, N. Cheng, A. C. Steven, J. M. Hogle, and D. M. Belnap, “The structure of the poliovirus 135s cell entry intermediate at 10-angstrom resolution reveals the location of an externalized polypeptide that binds to membranes,” *Journal of virology*, vol. 79, no. 12, pp. 7745–7755, 2005.

[66] R. L. Sandberg, C. Song, P. W. Wachulak, D. A. Raymondson, A. Paul, B. Amirbekian, E. Lee, A. E. Sakdinawat, L.-O. Chan, M. C. Marconi, *et al.*, “High numerical aperture tabletop soft x-ray diffraction microscopy with 70-nm resolution,” *Proceedings of the National Academy of Sciences*, vol. 105, no. 1, pp. 24–27, 2008.

[67] K. Tamasaku, Y. Tanaka, M. Yabashi, H. Yamazaki, N. Kawamura, M. Suzuki, and T. Ishikawa, “Spring-8 riken beamline iii for coherent x-ray optics,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 467, pp. 686–689, 2001.

[68] Y. Takata, K. Tamasaku, T. Tokushima, D. Miwa, S. Shin, T. Ishikawa, M. Yabashi, K. Kobayashi, J. Kim, T. Yao, T. Yamamoto, M. Arita, H. Namatame, and M. Taniguchi, “A probe of intrinsic valence band electronic structure: Hard x-ray photoemission,” *Applied Physics Letters*, vol. 84, pp. 4310–4312, May 2004.

[69] Y. Kohmura, Y. Nishino, T. Ishikawa, and J. Miao, “Effect of distorted illumination waves on coherent diffraction microscopy,” *Journal of applied physics*, vol. 98, no. 12, p. 123105, 2005.

[70] J. Miao, D. Sayre, and H. Chapman, "Phase retrieval from the magnitude of the fourier transforms of nonperiodic objects," *JOSA A*, vol. 15, no. 6, pp. 1662–1669, 1998.

[71] J. Miao, T. Ishikawa, E. H. Anderson, and K. O. Hodgson, "Phase retrieval of diffraction patterns from noncrystalline samples using the oversampling method," *Physical Review B*, vol. 67, no. 17, p. 174104, 2003.

[72] C. J. Hernandez and T. G. Mason, "Colloidal alphabet soup: monodisperse dispersions of shape-designed lithoparticles," *The Journal of Physical Chemistry C*, vol. 111, no. 12, pp. 4477–4480, 2007.

[73] J. Wilking and T. Mason, "Multiple trapped states and angular kramers hopping of complex dielectric shapes in a simple optical trap," *EPL (Europhysics Letters)*, vol. 81, no. 5, p. 58005, 2008.

[74] Q. Shen, I. Bazarov, and P. Thibault, "Diffractive imaging of nonperiodic materials with future coherent x-ray sources," *Journal of synchrotron radiation*, vol. 11, no. 5, pp. 432–438, 2004.

[75] M. R. Howells, T. Beetz, H. N. Chapman, C. Cui, J. Holton, C. Jacobsen, J. Kirz, E. Lima, S. Marchesini, H. Miao, *et al.*, "An assessment of the resolution limitation due to radiation-damage in x-ray diffraction microscopy," *Journal of Electron Spectroscopy and Related Phenomena*, vol. 170, no. 1, pp. 4–12, 2009.

[76] C. Jacobsen, J. Kirz, and S. Williams, "Resolution in soft x-ray microscopes," *Ultramicroscopy*, vol. 47, no. 1, pp. 55–79, 1992.

[77] R. Henderson, "The potential and limitations of neutrons, electrons and x-rays for atomic resolution microscopy of unstained biological molecules," *Quarterly reviews of biophysics*, vol. 28, no. 02, pp. 171–193, 1995.

[78] H. N. Chapman, P. Fromme, A. Barty, T. A. White, R. A. Kirian, A. Aquila, M. S. Hunter, J. Schulz, D. P. DePonte, U. Weierstall, *et al.*, "Femtosecond x-ray protein nanocrystallography," *Nature*, vol. 470, no. 7332, pp. 73–77, 2011.

[79] H. N. Chapman, A. Barty, M. J. Bogan, S. Boutet, M. Frank, S. P. Hau-Riege, S. Marchesini, B. W. Woods, S. Bajt, W. H. Benner, *et al.*, "Femtosecond diffractive imaging with a soft-x-ray free-electron laser," *Nature Physics*, vol. 2, no. 12, pp. 839–843, 2006.

[80] R. Neutze, R. Wouts, D. van der Spoel, E. Weckert, and J. Hajdu, "Potential for biomolecular imaging with femtosecond x-ray pulses," *Nature*, vol. 406, no. 6797, pp. 752–757, 2000.

[81] M. M. Seibert, T. Ekeberg, F. R. Maia, M. Svenda, J. Andreasson, O. Jönsson, D. Odić, B. Iwan, A. Rocker, D. Westphal, *et al.*, “Single mimivirus particles intercepted and imaged with an x-ray laser,” *Nature*, vol. 470, no. 7332, pp. 78–81, 2011.

[82] T. Kimura, Y. Joti, A. Shibuya, C. Song, S. Kim, K. Tono, M. Yabashi, M. Tamakoshi, T. Moriya, T. Oshima, *et al.*, “Imaging live cell in micro-liquid enclosure by x-ray laser diffraction,” *Nature communications*, vol. 5, 2014.

[83] R. Fung, V. Shneerson, D. K. Saldin, and A. Ourmazd, “Structure from fleeting illumination of faint spinning objects in flight,” *Nature Physics*, vol. 5, no. 1, pp. 64–67, 2009.

[84] G. Huldt, A. Szőke, and J. Hajdu, “Diffraction imaging of single particles and biomolecules,” *Journal of structural biology*, vol. 144, no. 1, pp. 219–227, 2003.

[85] R. A. Kirian, “Structure determination through correlated fluctuations in x-ray scattering,” *Journal of Physics B: Atomic, Molecular and Optical Physics*, vol. 45, no. 22, p. 223001, 2012.

[86] S. B. McGrayne, *The theory that would not die: how Bayes’ rule cracked the enigma code, hunted down Russian submarines, & emerged triumphant from two centuries of controversy.* Yale University Press, 2011.

[87] E. T. Jaynes, *Probability theory: the logic of science.* Cambridge university press, 2003.

[88] D. S. Sivia, *Data analysis: a Bayesian tutorial.* Oxford university press, 1996.

[89] E. T. Jaynes, “Bayesian methods: General background,” 1986.

[90] G. J. Erickson, *Maximum-entropy and Bayesian methods in science and engineering*, vol. 32. Springer, 1988.

[91] E. T. Jaynes, “Information theory and statistical mechanics,” *Physical review*, vol. 106, no. 4, p. 620, 1957.

[92] J. V. Stone, *Bayes’ Rule: A Tutorial Introduction to Bayesian Analysis.* JV Stone, 2013.

[93] A. Beck and M. Teboulle, “Fast Gradient-Based Algorithms for Constrained Total Variation Image Denoising and Deblurring Problems,” *IEEE Trans, Image Proc.*, vol. 18, no. 11, pp. 2419–2434, 2009.

[94] Y. Nishino, Y. Takahashi, N. Imamoto, T. Ishikawa, and K. Maeshima, “Three-dimensional visualization of a human chromosome using coherent x-ray diffraction,” *Physical review letters*, vol. 102, no. 1, p. 018101, 2009.

[95] T. C. Schelling, "Dynamic models of segregation," *Journal of mathematical sociology*, vol. 1, no. 2, pp. 143–186, 1971.

[96] E. S. Reich, "Three-dimensional technique on trial," *Nature*, vol. 480, no. 7377, pp. 303–303, 2011.

[97] J. Miao, C.-C. Chen, Y. Mao, L. S. Martin, and H. C. Kapteyn, "Potential and challenge of ankylography," *arXiv preprint arXiv:1112.4459*, 2011.

[98] S. Salha and K. Raines, "Adaptive phase retrieval for noisy and incomplete data," *Manuscript in preparation*, 2014.