STUDIES IN PHASE AND INVERSION PROBLEMS FOR DYNAMICAL ELECTRON DIFFRACTION

Helen Mary Louise Faulkner

Submitted in total fulfilment of the requirements of the degree of Doctor of Philosophy

School of Physics
The University of Melbourne
Australia

13th March 2003
Abstract

This thesis examines problems in electron diffraction and related areas of theoretical optics. It begins with a study of the phase of a quantum mechanical wavefunction and the behaviour of phase vortices and vortex cores. Several rules for vortex core evolution are given and simulated vortex trajectories are studied. These simulations show that in electron microscopy at atomic resolution and in other similar situations, vortices occur in the wavefunctions very frequently. This means any image processing methods which deal with the wavefunction phase must permit vortices to occur.

In this context a number of methods of phase retrieval are compared and evaluated. The criteria of evaluation are the accuracy of the phase retrieval, its ability to cope with vortices, its numerical stability and its required computational resources. The best method is found to be an iterative algorithm similar in approach to the Gerchberg-Saxton method, but based on a through focal series of images.

Using this phase retrieval method as an essential tool, the thesis continues with a study of inverse problems in electron optics. The first problem considered is that of using a set of images taken to characterise the coherent aberrations present in a general imaging system. This problem occurs in many areas of optics and is studied here with a focus on transmission electron microscopy. A method of using software to simultaneously determine aberrations and subsequently remove them is presented and tested in simulation. This method is found to have a high level of accuracy in aberration determination.

The second inverse problem studied in this thesis is the inversion problem in dynamical electron diffraction. This problem is solved for a periodic object, giving an accurate and unique solution for the projected potential in the multiple scattering case. An extension of this solution to objects which are non-periodic in the direction of the incident wave is investigated. Finally a model computation solving the general inversion problem for dynamical diffraction in an aberrated transmission electron microscope is performed, illustrating this and previous material and summing up the advances presented in this work.
Declaration

This is to certify that

1. the thesis comprises only my original work,

2. due acknowledgment has been made in the text to all other material used,

3. the thesis is of less than 100,000 words in length, exclusive of tables, figures, bibliographies, appendices and footnotes.

Helen Faulkner.
Acknowledgments

My thanks are due first of all to my supervisor, Dr Les Allen, for patient guidance and challenging inspiration throughout our time working together. I am also grateful to Dr Allen for advice in writing up and presenting my work, and for being supportive as I managed all of this in the context of my life over the last few years.

I am grateful to Mark Oxley for ongoing help with my research and with many technical details. For help with solving mathematical problems, improving algorithms and finding coding errors, my thanks to Ben Burton.

The material presented in Chapter 3 is the result of collaboration with Les Allen, Mark Oxley, David Paganin and Keith Nugent. Chapter 4 contains the result of collaboration with Les Allen, Mark Oxley and David Paganin, and Chapter 5 the result of collaboration with Les Allen. My contribution to each part of this collaborative work was substantial and direct.

Thanks to John Spence for making me welcome when visiting Arizona State University, and for valuable discussions during that time. I would also like to thank Christoph Koch, Scott Findlay and Nicole O’Leary for useful discussions during the course of my research. Thank you to Sandra Brownscombe, David Brownscombe and Deanna Wang for proofreading my thesis draft.

Finally, thanks to Rosie Faulkner for reminding me to value the small achievements and for adding much happiness and laughter to my life over the last few years.
# List of abbreviations used

The following abbreviations are used in this work:

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CBED</td>
<td>Convergent Beam Electron Diffraction</td>
</tr>
<tr>
<td>FEG ST</td>
<td>Field Emission Gun Scanning Transmission</td>
</tr>
<tr>
<td>FFT</td>
<td>Fast Fourier Transform</td>
</tr>
<tr>
<td>GS</td>
<td>Gerchberg-Saxton (method)</td>
</tr>
<tr>
<td>SSE</td>
<td>Sum Squared Error</td>
</tr>
<tr>
<td>SVD</td>
<td>Singular Value Decomposition</td>
</tr>
<tr>
<td>TDS</td>
<td>Thermal Diffuse Scattering</td>
</tr>
<tr>
<td>TEM</td>
<td>Transmission Electron Microscopy</td>
</tr>
<tr>
<td>TFS</td>
<td>Through Focal Series (method)</td>
</tr>
<tr>
<td>TIE</td>
<td>Transport of Intensity Equation</td>
</tr>
<tr>
<td>TPN</td>
<td>Teague, Paganin and Nugent (method)</td>
</tr>
<tr>
<td>HRTEM</td>
<td>High Resolution Transmission Electron Microscopy</td>
</tr>
<tr>
<td>VDC</td>
<td>Van Dyck and Coene (method)</td>
</tr>
</tbody>
</table>
# Contents

1 Introduction .................................................. 11  
1.1 Background motivation ................................... 11  
1.2 Definition of thesis problem .............................. 12  
1.3 Aim of research ........................................... 14  
1.4 Research approach ....................................... 14  

2 The phase problem and phase singularities ............... 17  
2.1 Basic properties of phase ................................. 17  
2.2 Motivation for phase retrieval ........................... 19  
2.3 A simple scattering problem .............................. 19  
2.4 Introduction to vortices .................................. 22  
2.5 Properties of vortices and vortex cores .................. 25  
2.6 Tracking vortex trajectories .............................. 30  
2.7 Vortex trajectory simulation .............................. 32  
2.8 Summary .................................................. 35  

3 Methods of phase retrieval ................................... 37  
3.1 Transport of Intensity Equation ......................... 38  
3.2 Teague, Paganin and Nugent method ..................... 40  
3.3 Van Dyck and Coene method .............................. 42  
3.4 Gerchberg-Saxton method ................................. 44  
3.5 Fienup method ............................................ 46  
3.6 Through Focal Series method .............................. 47  
3.7 Analysis of iterative algorithms ......................... 49  
3.8 Phase retrieval results for periodic wavefunctions ....... 53  
3.9 Noise analysis of iterative algorithms .................... 59  
3.10 Phase retrieval for non-periodic wavefunctions ........ 60  
3.11 Summary .................................................. 64
# List of Figures

1.1 Parts of the research problem addressed in this thesis. 13

2.1 Wave travelling through a potential barrier. 20

2.2 Behaviour of the phase about a vortex core. 23

2.3 Greyscale map of phase behaviour around vortex cores with topological charges +1 and +2. 24

2.4 Two vortices connected by a branch cut. 24

2.5 Change of route of branch cuts between vortices. 25

2.6 Vortex trajectory with plane of observation. 27

2.7 Vortex trajectory with two planes of observation. 28

2.8 Illustration of Freund’s Sign Rule for vortex trajectories. 29

2.9 Sign rule for junctions. 30

2.10 Evolution of vortex trajectories in a wavefunction propagating out of a GaAs crystal. 33

2.11 Evolution of a single vortex trajectory with a simple loop structure. 33

2.12 Vortex trajectories within an $xy$ “unit”. 34

2.13 Vortex trajectories from Fig. 2.12 looking down the $z$ axis. 34

3.1 Gerchberg-Saxton algorithm. 45

3.2 Fienup algorithm. 46

3.3 TFS algorithm for three planes. 48

3.4 Mixing of initial intensity and phase as the wavefunction is propagated to different defocus values. 49

3.5 Behaviour of iterative methods with convex and non-convex sets of constraints. 51

3.6 Test data for phase retrievals. 54

3.7 Input data for TPN and VDC methods, first test case - $C_s = 1.23$ mm, $\Delta f = 641$ Å, $t = 90$ Å. 55
3.8 Input data for TPN and VDC methods, second test case - $C_s = 0.05$ mm, $\Delta f = 140$ Å, $t = 90$ Å. ................................. 55
3.9 Input data for TFS method, first test case - $C_s = 1.23$ mm, $\Delta f = 641$ Å, $t = 90$ Å. .................................................. 56
3.10 Input data for TFS method, second test case - $C_s = 0.05$ mm, $\Delta f = 140$ Å, $t = 90$ Å. .................................................. 57
3.11 Diffraction pattern data required by GS algorithm. ............... 57
3.12 Phase retrieval comparison for periodic images. .................... 58
3.13 Phase retrieval for noisy data - retrieved and aberration corrected phases. ................................................................. 59
3.14 Input model wavefunction. ........................................... 61
3.15 Series of defocused images and phase maps. ......................... 62
3.16 Vortex trajectory corresponding to arrows in Fig. 3.15. .......... 63
3.17 Image and recovered phase map for $\Delta f = 500$ Å and propagated back to $\Delta f = 0$ Å. ........................................... 63

4.1 The first 21 Zernike polynomials. ................................. 71
4.2 Setup for characterisation and removal of the aberrations of a given imaging system. .................................................... 74
4.3 Original and retrieved phase factors. ............................... 77
4.4 Image and phase of the aberrated wave function in the principal image plane. ................................................................. 78
4.5 Through focal series used to retrieve phase and the corresponding phase maps in each plane. ........................................... 79
4.6 Effect of aberration correction on measured wavefunction image and phase. ................................................................. 80

5.1 Schematic diagram of electron diffraction. ......................... 85
5.2 Projected potential and images for single and multiple scattering at 400keV. ................................................................. 90
5.3 Indexing of the structure matrix. ..................................... 94
5.4 Real and absorptive parts of the optical potential for the [110] zone axis in GaAs. ................................................................. 98
5.5 Number of symmetries leading to equations of the type in Eq. 5.30 as a function of $N$. ....................................................... 100
5.6 Two crystals sandwiched together in the $z$ direction. .......... 101
5.7 Focussed images of silicon crystals at 100keV that are identical except for translation of the unit cell. 101
5.8 Projected potentials and images for a silicon crystal with a stacking fault at 100keV. 103

6.1 Schematic of optical systems in TEM. 108
6.2 A point source incident onto the post-specimen optics. 108
6.3 Input phase factor $W(q)$ in the post-specimen optics and phase factor retrieved by aberration characterisation. 109
6.4 Through focal series for aberration characterisation. 109
6.5 Through focal series for model problem at (000) tilt with no noise included. 110
6.6 Some of the input images for model problem at different tilts, no noise. 111
6.7 Input images at different tilts, 5% noise level. 112
6.8 Input images at different tilts, 10% noise level. 112
6.9 Original and recovered elastic and absorptive potentials for model problem. 113
Chapter 1

Introduction

1.1 Background motivation

Inverse problems have challenged physicists and mathematicians in a variety of fields for many years [1–12]. They occur whenever the output of a physical process can be measured, but the exact details of the process are unknown. The problem is to use the known theories of the physical situation to reconstruct the exact process which produced that output. An example of an inverse problem is use of the measured results of a medical scan, such as an X-ray or ultrasound, to discover the structure of the body (arrangement of bones, soft tissues, etc.) which would produce those results [10, 11]. Another example occurs when astrophysicists use measurements of various types of radiation incident onto Earth to deduce the location and properties of stars, galaxies and other objects in space [3].

One area in which many inverse problems occur is microscopy [12–16]. This thesis focuses on problems which arise in electron microscopy, in particular as a result of dynamical electron diffraction. In microscopy, an important problem is that of precisely determining the physical characteristics of the imaging system in use [17]. It is often difficult to directly measure such characteristics as the spherical aberration or axial astigmatism of a lens. However without such knowledge, it is impossible to precisely interpret the information contained in an image. Being able to solve an inverse problem in order to accurately characterise the aberrations present in a system is an important goal. However current methods of aberration determination are inexact and difficult to implement. Improvement in this area is greatly desirable and the issue is addressed in this thesis.

Another inverse problem in microscopy arises when one or more images have been taken of a sample and mathematical techniques are used to reconstruct
physical properties such as the arrangement and types of atoms implied by its projected potential [18–24]. It is desirable to be able to solve inversion problems of this type because such solutions allow us to understand the structure of matter at the atomic level. This challenging, theoretical task is also addressed in this thesis.

The benefits and potential applications of solutions to these inverse problems in electron microscopy are great. Given successful solutions, it will be possible in the future to use such techniques to image the structure of any object at the atomic level. This will lead to greater understanding of how the atoms combine to affect the macroscopic properties of the material. The results of this study could be used to great advantage in many areas. These include materials technology, where methods of creating and analysing the vast numbers of specialised materials needed in modern industry may be improved, as well as pure scientific research, to further understanding of the structure and properties of matter.

The research presented in this thesis focuses on electron microscopy as a specific application. However, since many areas of optics have the same mathematical foundations, much of this work is not specific to the electron microscopy case. Thus parts of the research have broad implications for optics in general as well as many applications in different kinds of microscopy and optical work.

1.2 Definition of thesis problem

Research into phase and inversion problems in dynamical electron diffraction can be divided into three parts, shown in Fig. 1.1. The first arises because in order to solve an inverse problem in quantum mechanics, it is necessary to know the entire complex wavefunction, i.e. intensity and phase, produced by the system studied. The phase information contained in the wave can be deduced from measurements of the intensity in one or more planes [25–35]. This process is known as phase retrieval or solving the phase problem. The problem exists because it is impossible to directly measure the relative phase of a quantum mechanical wavefunction. Therefore indirect methods are required to deduce the phase from measurements that can be taken.

It is generally necessary to use measurements of the wavefunction in multiple planes to find the phase. These measurements may be the diffraction pattern and the focused image, or images at other defoci. There are numerous methods of phasing the wavefunction, each with different requirements for input data.
Accurately phasing a wavefunction is a challenging problem.

The second and third parts of the overall problem both involve using a known complex wavefunction to determine information about the physical system which has produced it. These are inverse problems. The second part of the problem is to characterise the properties of the imaging system used to produce the measured wavefunction in terms of the standard types of aberrations. Since accurate aberration determination and correction is difficult in many areas of microscopy [13, 15, 17, 36–39], the ability to do this effectively is a major step forward in the development of the science.

The third part in the overall problem assumes completion of the first and second parts. Given measurements taken with a known microscope system, the problem involves using the aberration corrected exit surface wavefunction to find the structure in the form of the projected potential of the scattering object. Previous research has been able to achieve this in a limited way by making the approximation that only single scattering occurs [40–42]. This approximation is valid for many cases in scattering theory when the interaction between the scattering wave and the object is weak. However it is often inaccurate in electron diffraction because electrons interact more strongly with matter than many other particles, and thus are more likely to result in multiple or dynamical scattering [18, 19]. It is therefore important in electron microscopy to develop methods of solving inversion problems which allow multiple scattering to occur [20, 21].
1.3 Aim of research

The central aim of this research is to find improved methods of solving phase and inversion problems in dynamical electron diffraction. In doing so it is important to consider the ability of proposed methods to cope with key features of the experimental situation, such as noise, lens aberrations or the presence of singularities. An improved method might be accurate in a wider range of physical situations while still remaining numerically reliable and computationally practical.

Having successfully identified improved methods of solving these problems, the next aim of this research is to apply such methods in simulation to a variety of important problems in electron diffraction. These include characterising and removing the aberrations in an imperfect imaging system and finding the projected potential of a sample. Achieving these aims potentially allows wider use of microscope technology given the limitations posed by currently available experimental equipment.

1.4 Research approach

The first step in achieving the aim involves exploring the theoretical aspects of the problems in order to obtain greater understanding of potentially useful methods of solution. Chapter 2 introduces a number of properties of the phase of a wavefunction and discusses the motivation for phase retrieval after the wavefunction has interacted with a scattering object. The importance of knowing the phase in order to solve inversion problems is demonstrated with analysis of a simple scattering problem.

Chapter 2 goes on to examine the phase singularities known as vortices. This discussion begins by describing the ubiquitous nature of vortices and explains some key aspects of the behaviour of vortex cores. The results of a simulated investigation into the evolution of vortex cores are analysed and the implications of this for phase retrieval in general are discussed.

The next step is to computationally implement some potentially useful algorithms. Chapter 3 is a comparative review of several topical methods of retrieving the phase of a wavefunction. The mathematical backgrounds of the algorithms used are explained. In addition, the required input information and practical issues involved in measuring this experimentally are discussed. Special consideration is paid to the need to find phase retrieval techniques which are successful
when vortices are present in the phase.

Using these techniques, solutions to the phase problem are implemented and tested. Some methods are only applicable to certain situations while others are more generally useful. Potential computational methods are tested thoroughly with data sets covering a range of simulated experimental situations, including effects such as noise and lens aberrations. The results of comparing these techniques for phase retrieval in the case of periodic objects are given. Chapter 3 also goes on to discuss the non-periodic case, detailing the mathematical implications of loss of periodicity and discussing the effectiveness of alternate methods for phase retrieval of non-periodic wavefunctions. The material in Chapter 3 leads to an acceptable solution to the first part of the research problem (the phase problem).

The solution to the second part of the research problem is discussed in Chapter 4. This is the determination and correction of aberrations in an imaging system such as an electron microscope. As mentioned, the phase retrieval method found in Chapter 3 is an essential tool for the solution of this problem. The use of Zernike polynomials as a basis for the characterisation of aberrations is explained, and a general algorithm for aberration determination is proposed and tested. This algorithm is found to successfully retrieve coherent aberrations to high accuracy.

Chapter 5 introduces part three of the research problem, which involves retrieving structure by solving the inversion problem for dynamical electron diffraction from periodic objects. This also utilises the phase retrieval techniques studied in Chapter 3. It begins by examining the theoretical mathematical formulation of the direct problem, from which the inversion methods proceed. It then outlines the problems involved in attempting to invert from a measured wavefunction to find the projected potential of a periodic object. A general solution to the dynamical inversion problem is given. The effect of absorption is included in the solution and simulated inversion results are given for several different periodic objects of varying complexity. Retrieved projected potentials are shown. The possible applications of this work, as well as extensions to the non-periodic case are discussed. Chapter 5 finishes with a brief discussion of the problem of finding the projected potential of an object which is not periodic in one dimension, as found in the situation when a crystal has a stacking fault.

As a summary of all the preceding work, Chapter 6 presents the solution of an example problem where an unknown periodic sample is analysed by a trans-
mission electron microscope with unknown coherent aberrations in the objective lens system. First the aberrations are characterised, using the phase retrieval algorithms from Chapter 3 and the aberration control techniques from Chapter 4. Next the sample is added and its structure determined using both phase retrieval and the inversion techniques for periodic structures from Chapter 5. The success of the model calculation in the presence of noisy data is investigated.

The thesis concludes in Chapter 7 with a summary of what has been achieved. This is followed by a discussion of the implications and future applications of this research.
Chapter 2

The phase problem and phase singularities

The importance of phase retrieval as a precursor to solving inverse problems involving a complex wavefunction is briefly mentioned in Chapter 1. The reasons for this are now explored in more detail. We start by examining the basic properties of that part of a quantum mechanical wave known as its phase. This is followed in later sections with the reasons phase retrieval is an essential tool for the structure retrieval work contained in the second part of this thesis.

The second part of this chapter is about the phase singularities known as vortices. We begin with a description of what vortices are and where they occur, followed by a discussion of the importance of considering the effects of vortices when studying dynamical electron diffraction. Next we explore a number of mathematical rules which restrict the behaviour of vortices as they travel through \( N \) dimensional space [26, 43, 44]. Finally, a simple technique for computationally locating vortices in a simulated wavefunction is given. Some results found by using this method to study the evolution of vortex cores are shown and discussed.

2.1 Basic properties of phase

A complex wave can be described mathematically as follows

\[
\psi(r) = |\psi(r)|e^{i\theta(r)}.
\]  

Here \(|\psi(r)|\) is known as the amplitude or magnitude of the wave, \( r \) is the position and \( \theta(r) \) is the phase of the wave. The phase describes the state of evolution of
the wave at any point in space, or in other words the location of its peaks and troughs. Note that when referring to the phase of the wavefunction, we refer to the relative phases of different parts of the wave. There is always some unknown overall phase shift, which it is impossible to measure. This is an expression of the principle of gauge invariance which in no way affects the physical behaviour of the system under consideration [45].

The complex exponential has the well known property that \( e^{2\pi n} = 1 \), where \( n \) is any integer. It can be seen from this that phases which differ by an integer multiple of \( 2\pi \) produce the same physical wavefunction. Thus when considering phase it sometimes makes sense to be interested in the value of the phase modulo \( 2\pi \), since this is the physically important quantity. However in other situations, especially when investigating singularities of the phase, it is important to understand that mathematically it can have any real value. A phase which is restricted to the region \([0, 2\pi)\), or any other range of \( 2\pi \), is known as a wrapped phase. Phase unwrapping is the process of reversing the modulo function and finding the original values of the phase before the wrapping took place. Since the modulo function is not mathematically one to one, it can be difficult to accurately unwrap the phase of a wavefunction, especially if there are singularities present. There exist sophisticated algorithms which attempt to unwrap phase by using information such as the continuity of the phase in known regions, the locations of singularities in the phase or other known facts [46,47].

If the wavefunction is broken into its real and imaginary parts, \( \text{Re}[\psi(r)] \) and \( \text{Im}[\psi(r)] \), the phase may be found using the relation.

\[
\theta(r) = \arctan\left\{ \frac{\text{Im}[\psi(r)]}{\text{Re}[\psi(r)]} \right\}
\]

This description makes clear some features of the phase behaviour of a wavefunction. Firstly it is easy to see that if \( \text{Im}[\psi(r)] = 0 \), then \( \theta(r) = n\pi \) with \( n \) an integer, and the wavefunction is entirely real. If \( \text{Re}[\psi(r)] = 0 \), then \( \theta(r) = \frac{\pi}{2} + n\pi \), and the wavefunction is entirely complex. The most interesting fact is that if both the real and imaginary parts of the wavefunction are zero at a given point in space, in other words, if the wave has zero magnitude, the phase is not well defined, since the argument of \( \arctan \) is an undefined quantity. Thus it is possible for singularities of the phase to exist at points where the magnitude is zero [48]. This point is crucial to the discussion on vortices in Sections 2.4 to 2.7.
2.2 Motivation for phase retrieval

Although the phase of a wavefunction contains important information, it can not be directly measured. This is due to the well known fact that the only directly measurable quantities are the magnitudes squared, or intensities,

$$\psi(r)\psi^*(r) = |\psi(r)|^2,$$

(2.3)
of the wavefunction $\psi(r)$. The expression for the intensity of the wave is entirely real. This makes it impossible to directly deduce the phase of the wavefunction from the measurable quantity (i.e. a single image), since the separate real and imaginary parts of the wavefunction are not known.

Essentially what this means is that only part of the information contained in any complex scalar wavefunction can be directly measured. However it is often necessary to know the phase of an observed wavefunction, since many problems require knowledge of the entire complex wavefunction. Applications such as crystallographic structure determination, speech recognition and image reconstruction often involve solving inverse problems which depend on an accurate knowledge of the phase of a wavefunction [49].

In electron diffraction, it is often essential to know the phase of the wavefunction exiting the scattering object, whether that “object” is a sample of material under study or part of the microscope imaging system. This is because the scattering process alters the relative phases of different parts of the original wavefunction, as well as changing the magnitudes. It is therefore impossible to reconstruct the scattering potential unless the information contained in the phase of the wavefunction is known. In other words, to solve an inversion problem in electron diffraction, it is necessary to know the entire complex wavefunction exiting the diffracting system.\(^1\) This point is demonstrated in Section 2.3, where we examine the solution of a simple problem in scattering theory.

2.3 A simple scattering problem

Consider the case of a plane wave travelling through a square potential barrier (of constant potential), as shown in Fig. 2.1. Materials 1 and 2 are different, resulting

\(^1\)Actually in dynamical electron diffraction it is usually necessary to know the entire phased wavefunction at multiple orientations of the incident wavefunction, with a fixed energy and sample thickness. This is discussed in detail in Chapter 5
in the potential difference indicated by the red box. For example material 1 could be air and material 2 some other medium through which we wish to pass the wave, such as a crystal. We assume for simplicity that there is forward scattering only (i.e. the wave does not reflect back at either potential change), that both waves have unit intensity and that there is elastic scattering only (i.e. no absorption). The incident wave, travelling in material 1, is given by

\[ \psi_1(r) = e^{kr}, \quad (2.4) \]

where \( r \) is the position vector in three dimensional space, and \( k \) is the incident wave vector. The wave inside material 2 is given by

\[ \psi_2(r) = e^{Kr}, \quad (2.5) \]

where \( K \) is the wave vector inside the potential.

![Figure 2.1: Wave travelling through a potential barrier.](image)

We assume that the incident wave is normal to the potential barrier and that it has zero relative phase at the start of the potential barrier. If the barrier has thickness \( t \), \( k = |k| \) and \( K = |K| \), then the wave travelling through the barrier is given by

\[ \psi_2(t) = e^{Kt}, \quad (2.6) \]

at the exit surface.

If the wave does not pass through the potential barrier it is given by

\[ \psi_1(t) = e^{kt}. \quad (2.7) \]
There is a phase difference between these two waves of

$$\Delta \theta = (k - K)t.$$  \hfill (2.8)

The phase difference exists because the two wave vectors are different, i.e. the wave travels at a different speed inside the different materials. This means that even in such a simple case, the phase of the wave is altered after passing through a scattering object.

The problem we wish to solve is that of finding the potential difference between the two materials. It is assumed that the wavelength of the incident wave $k$ and the thickness $t$ of material 2 are known. The constant scattering potential is unknown, however it is possible to find it by determining the phase of the exit wave as compared to that expected from an unaltered incident wave. The wave vector in material 2 is found by rearranging Eq. 2.8 to give

$$K = k - \frac{\Delta \theta}{t}.$$ \hfill (2.9)

Then using the relation

$$K^2 = k^2 + U,$$ \hfill (2.10)

where $U$ is the potential difference between material 1 and material 2 [21], we find that the potential is given by

$$U = \left(\frac{\Delta \theta}{t}\right)^2 - 2k\frac{\Delta \theta}{t}.$$ \hfill (2.11)

In a more complex scattering situation the phase of the exit wavefunction is changed in a more complicated way. Since any scattering event requires a change in the potential experienced by the incoming electron wavefunction, any such event alters the phase of the incident wave. We conclude that it is necessary to know this altered phase in order to deduce the structure of the scattering potential.

The above discussion is based on the properties of a plane wave. However since any more complicated wave can be described as a sum of plane waves [50], this conclusion is accurate for any complex, scalar wavefunction.
2.4 Introduction to vortices

An important consideration when dealing with the phase of a wavefunction in a situation like electron microscopy is whether phase singularities are present. This is because any mathematical techniques chosen to retrieve phase from a wavefunction that contains phase singularities must be able to handle them. A closely related question is whether the intensity of the wavefunction is ever zero, since a non-zero wavefunction allows the phase to be undefined. It is in this situation that singularities such as vortices may occur.

Vortices occur commonly in many natural situations [26, 48, 51–54]. They are a natural feature of many dynamical systems and are therefore a topic of interest in a number of academic fields. If a vortex is broadly considered to be a rotational singularity, common macroscopic examples include vortices in a rising plume of smoke, in a whirlpool or cyclone and at the centre of a spiral galaxy. Quantum mechanical vortices appear in the fluid dynamical behaviour of superfluids and superconductors. They also occur in Bose Einstein condensates and angular momentum eigenstates of hydrogen. In the study of optics, vortices arise in phenomena such as diffraction free vortex beams, vortex solitons and holographically produced screw dislocations [26].

In the context of dynamical electron diffraction, a vortex is a singularity of the phase of the wavefunction. Vortices occur frequently in electron waves and there are many situations where it is difficult if not impossible to avoid them. There are instances where the electron wavefunction exiting a crystal does not contain any vortices. However vortices almost always appear quickly as the wavefunction propagates through free space. When multiple measurements of images are required at different defoci or tilts of the incident beam, it is unlikely that vortices will be absent in all measurements taken. These singularities are a fundamental feature of the fine structure of waves, as is discussed by Nye [54]. Indeed, the existence of spontaneously generated wave field vortices in the focal region of a lens has been known for some time [55]. Vortices have also been observed experimentally in the optical region [56]. The ubiquitous nature of the vortex is yet to be fully appreciated in the electron diffraction community. Thus careful analysis of the behaviour of phase retrieval methods in the presence of vortices has not previously been done. If a phase retrieval method is to be successful in the case of dynamical electron diffraction (or indeed in many other physical situations), it must be able to accurately retrieve the phase of a wavefunction containing one or more vortices.
A vortex is a singularity in the phase of a wavefunction of the type often known as a screw dislocation (a vortex may actually be a combination of screw and edge dislocations, depending on the angle of observation). Both the wavefunction and its intensity are well defined, well behaved, single valued objects. However when the intensity of the wavefunction is zero, the phase is not well defined, and thus may be discontinuous and multivalued. Since phase is defined modulo $2\pi$, the circulation, or integral around a closed loop, may in general be equal to any integer multiple of $2\pi$. This is expressed in the following equation:

$$\oint_{\Gamma} \nabla \theta(r) \cdot d\hat{n} = 2\pi m.$$  \hspace{1cm} (2.12)

Here, $\hat{n}$ is the unit vector tangential to the closed path $\Gamma$ around which the loop integral is taken, $\Gamma$ is any closed loop in the parameter space $r$ over which the modulus of the wave function is non-zero, and $m$ is an integer. If $m$ is not zero, a vortex-type singularity is present. The integer $m$ is often termed the topological charge of the singularity. The behaviour of the (multi-valued) phase about a vortex core is shown in Fig. 2.2 and is dominated by the “spiral staircase” structure characteristic of vortex-type features.

Although for the purposes of this discussion we assume the wavefunction exists in a three dimensional coordinate space $r$, which corresponds to the usual three spatial directions, this can easily be generalised to higher dimensions by including other parameters. These parameters could for example be aberrations such as spherical aberration or astigmatism or a changing medium traversed by the wave. All of the following arguments hold for a parameter space with more than three dimensions.

Fig. 2.3(a) shows a greyscale map of the region around a vortex with topo-
Figure 2.3: Greyscale map of phase behaviour around vortex cores with topological charges +1 and +2.

logical charge 1. An example of a closed path $\Gamma$ around the vortex is marked. In this mapping, black corresponds to a phase of zero and white corresponds to phase $2\pi$. The figure shows how the phase increases smoothly around the centre, or vortex core, until it reaches $2\pi$ and then crosses a branch cut to start from zero again. Calculating the circulation of the phase along the path $\Gamma$, according to Eq. 2.12, produces the expected value of $2\pi$. In a vortex of higher order, the wrapping process occurs multiple times in one circulation about the vortex core. For example, a vortex with topological charge 2 is shown in Fig. 2.3(b). In both pictures, it is of interest to note the one or more branch cuts coming from each vortex core. These join up with a second vortex core with the opposite direction of rotation, which corresponds to the negative of the first topological charge. This is illustrated in Fig. 2.4, which shows a vortex with topological charge $m = 1$ (left) joined by a branch cut to a second vortex with charge $m = -1$ (right).

Figure 2.4: Two vortices connected by a branch cut.

The branch cut line indicates a jump of $2\pi$ in the phase. However it is not fixed in location as the vortices are. Since an overall constant phase may be added to the entire wavefunction without changing its physical properties, the branch line may take many different routes, although the vortex cores remain in the same locations. This effect is shown in Fig. 2.5, which shows the phase map
of a wavefunction resulting from the diffraction of a 400keV electron beam by a 1000 Å thick GaAs crystal. In Fig. 2.5 (a) four vortex core locations are marked by arrows. One branch cut runs between the top pair of vortices and another connects the bottom pair. In Fig. 2.5(b), which shows the same wavefunction with an added phase shift, the branch cuts have changed to run between the left and right pairs of vortices instead, although the vortex core locations have not changed. In Figs. 2.5(c) and (d) the route of the branch cuts change, compared to (b), with each additional overall phase shift added.

![Image](image.png)

(a) Overall phase shift of 0 radian
(b) Overall phase shift of $\frac{\pi}{3}$ radian.
(c) Overall phase shift of $\frac{2\pi}{3}$ radian
(d) Overall phase shift of $\pi$ radian.

Figure 2.5: Change of route of branch cuts between vortices.

### 2.5 Properties of vortices and vortex cores

Vortices have a number of interesting properties, which have been discussed by Allen et al. [26], Freund [44], Berry [48], Nye and Berry [53] and others. We begin this discussion with the statement, proof and discussion of three theorems concerning topological structures in the generic wave function $\psi(r)$. Although simple, these three properties of complex scalar wavefunctions give considerable insight into the behaviour of vortices in such wavefunctions.
**Theorem 1** The same circulation is possessed by all closed paths $\Gamma$, in the parameter space of coordinates $r$, which can be continuously deformed into one another without encountering zeros of the field.

**Proof 1** Suppose the initial path $\Gamma$ is infinitesimally perturbed to a new path $\Gamma + \delta \Gamma$. The modulus squared of the wave function is assumed to be strictly positive along every point $r$ of both paths. Since the wave function is assumed to be a continuous function of $r$, and the intensity is non-zero at every point on $\Gamma + \delta \Gamma$, the phase is defined and continuous at every point on $\Gamma + \delta \Gamma$. Thus the circulation over $\Gamma + \delta \Gamma$ can be at most infinitesimally different to the circulation over $\Gamma$. However due to the rule stated in Eq. 2.12, the circulation must take on discrete values $2\pi m$, and so the circulation is unchanged under the infinitesimal transformation of $\Gamma$ to $\Gamma + \delta \Gamma$. Since any finite deformation of $\Gamma$ is a sum of infinitesimal deformations, under each of which the circulation is unchanged, the proof is complete.

**Theorem 2** If a path $\Gamma$ has non-zero circulation, it may not be continuously contracted to a point without encountering a point $r$ of parameter space where $\psi(r) = 0$. Thus a screw dislocation region or dislocation line is always associated with zero intensity.

**Proof 2** Assume the converse, namely that a path $\Gamma$ of non-zero circulation may be continuously contracted to a point without encountering a zero of the field. Having collapsed $\Gamma$ down to a point, the circulation must be zero, which contradicts the initial assumption and therefore proves the theorem.

**Theorem 3** The screw dislocation region, in the parameter space of coordinates $r$, over which $\psi(r) = 0$ encompasses either an infinite line or a closed loop (with the path $\Gamma$ travelling through the centre of the loop) [57].

**Proof 3** This is a consequence of Theorem 2. Because Theorem 2 operates in all dimensions of the parameter space, the screw dislocation region must have a shape such that it is impossible to perturb a path $\Gamma$ in any way so as to escape the screw dislocation region. The only shapes which have this feature are infinite lines (i.e the path $\Gamma$ can never reach the end of the line) or closed loops (so that $\Gamma$ can never “unstring” itself from the loop).
Of course, Theorem 3 permits the dislocation region to have a more complicated topology than that of a simple loop or line, allowing such features as knots and branches. Sections of the region may lie in the plane. Where this occurs, we observe a line of zero intensity across which the phase jumps by $\pi$ radians. This is known as an edge dislocation.

![Figure 2.6: Vortex trajectory with plane of observation.](image)

**Berry’s Paradox**  An interesting feature arising from these theorems is that a vortex in a propagating wave may be described exactly by tracing the location of its centre, or core, in three dimensional space, assuming the topological charge of the vortex is known at one point. This general idea may be extended to a higher dimensional parameter space in which the vortex behaviour is being examined. Fig. 2.6 shows the movement of a vortex (and its counter-propagating partner) through space, as defined by the path traced out by its core.\(^2\) One possible plane of observation is marked. The vortex core intersects the plane of observation at two points only, corresponding to a pair of counter-propagating vortices, which are marked by crosses. As the location of the plane of observation is varied from left to right these counter-propagating vortices meet and annihilate. An interesting point to note is that the direction of rotation of the vortex depends on the point of view of the observer. An observer from the right of the plane sees a vortex rotating in one direction, whereas an observer from the left sees a vortex rotating in the opposite direction. Thus the sign of the topological charge is not

\(^2\)In this and following figures a green circle indicates a clockwise rotating vortex and a red circle an anti-clockwise rotating vortex, from the point of view of an observer on the right.
an intrinsic quality of a vortex, but depends on the point of view of the observer.

A vortex core may travel in any direction in a three dimensional space. Thus it may double back and return towards the direction it came from. Fig. 2.7 shows the trajectory of a vortex core which changes direction as it travels. Two possible planes of observation are shown. An observer at plane A sees two vortices with opposite directions of rotation. An observer at plane B sees four vortices. Movement of the observation plane from left to right shows an initial two vortices, joined by the spontaneous formation of an additional pair of vortices, followed by two separate annihilation events resulting in all the vortices disappearing. However all this complex behaviour depends on the direction of movement of the observer. If the movement is from top to bottom, a different sequence of events is seen, with creation and annihilation events occurring at different places and times.

![Vortex trajectory with two planes of observation.](image)

This demonstrates what is known as Berry’s Paradox [44, 51]. The Paradox is that vortex creation and annihilation events occur at many different and contradictory locations on the vortex core path, depending only on the position and movement of the observer. Since these events have no physical existence that is independent of the observer we must abandon the idea of creation and annihilation of vortices. It is more useful to consider the vortex core path or trajectory as an entire, single object, rather than a collection of individual vortices connected by creation and annihilation events.

Berry’s Paradox has implications for the common classification of dislocations.
into edge and screw types. We briefly note here that there appears to be some confusion in the literature over the connection between edge and screw dislocations and how they may combine in terms of vortex core direction. A vortex core travelling in the plane of observation produces an edge dislocation, which takes the form of a discrete jump of $\pi$ in the phase. A core with a direction of travel that is normal to the plane of observation produces a screw dislocation, observed as a vortex. Edge and screw dislocations may be combined, resulting in cores which are travelling at some other angle to the plane of observation. These also result in an observed vortex. Since this classification of dislocations depends entirely on the particular plane of observation, it is not useful in the context of this thesis and is not considered further.

**Freund’s Sign Rule** Freund [43, 44], Berry [48, 51], Nye [53] and others have examined mathematically some of the more complex rules for the behaviour of vortex core trajectories. One of particular interest is the Sign Rule as discussed by Freund [43]. This lays down very simply a key rule for the branching and joining of vortex core trajectories. The rule states that if two trajectories lie next to each other in the same manifold, they must have opposite directions of rotation. This is illustrated in Fig. 2.8, where the circle arrows indicate the direction of rotation of the vortices. Trajectories A and B are permitted to coexist by the Sign Rule. However trajectory C can not exist directly alongside trajectory B because they both have the same direction of rotation.

![Figure 2.8: Illustration of Freund's Sign Rule for vortex trajectories.](image)

An implication of the Sign Rule is that although a single vortex core trajectory is permitted to branch off into multiple trajectories, in practice the ways it can do this are restricted. For example, the Sign Rule forbids a three way junction of trajectories. As shown in Fig. 2.9(a), although two of the paths (A and B)

---

3An $N$-manifold is a set of points with the property that a small neighbourhood of each point behaves like $N$ dimensional space.
leading to the junction can have opposite signs, the third (C) must have the same sign as one of the others, breaking the Sign Rule. What this implies for vortex evolution is that it is impossible for a single vortex trajectory to break into two, because that would produce a three way junction. A four way junction may exist however, so it is permitted for two trajectories (A and B) to come together, touch and branch off into two new trajectories (C and D), as shown in Fig. 2.9(b).

The Sign Rule for vortex trajectories has the following interesting result. For a group of first order trajectories \( m = \pm 1 \) coexisting in the same manifold, it is possible to deduce the sign of the topological charge of all of them if just one topological charge is known. This means that there are situations where it is very easy to learn the topological charge of observed vortices as there are only a small number of possible combinations of charges if all trajectories are to obey the Sign Rule. Freund \[43\] discusses these and other rules for vortex trajectory evolution.

### 2.6 Tracking vortex trajectories

Simulations were performed to study the behaviour of vortices and vortex trajectories. The structure of specific simulated trajectories has not previously been analysed, although vortices are frequently observed in two dimensional wavefunctions and mathematical analyses of the general behaviour of vortex trajectories have been conducted by a number of people \[43, 44, 48, 51, 53, 57\]. This type of behaviour is difficult to measure experimentally due to the large number of phase retrievals that are required. However, using appropriate models of propagation of a wavefunction, we can accurately simulate the behaviour of vortex trajectories.
The model wavefunction studied is that resulting from a scattering interaction between an electron beam and a GaAs crystal. This is done using a Bloch wave model [58] calculation. Absorption is included in the calculations assuming that it is due to thermal diffuse scattering (TDS) based on an Einstein model [18,59,60]. TDS is the main absorptive mechanism leading to loss of diffraction contrast.

As previously mentioned, it is simpler to study the behaviour of the vortex trajectories in a three dimensional space rather than one with higher dimensions. Thus for the two dimensional wavefunction in a plane, only the amount of defocus is changed. This is equivalent to observing the behaviour of vortices in a wavefunction travelling through free space with no other parameter changing. It is, of course, possible to observe the behaviour when other parameters, such as spherical aberration or input electron beam energy are gradually changed. However, it seems most useful to consider the more natural situation of a wavefunction travelling through free space, since this corresponds to a common physical situation. If different parameters were used, the results would be similar, since the vortices must still obey the rules described in the preceding sections.

The following algorithm is used to obtain data for study. Note that it is applicable to any wavefunction which can be calculated at a known defocus, not just one emerging from a crystal, as in this example.

1. Calculate the wavefunction for a known defocus in space.

2. Find the zeros of intensity of the wavefunction. In practice this means numerically finding all points with intensity lower than a cutoff value $\delta$. If $\delta$ is low, potential vortex locations may be missed because of pixelation errors. If it is high, the time required for computation is increased because a greater number of possible zeros are identified.

3. At each possible zero point, numerically perform the path integral from Eq. 2.12. This is done by taking a closed loop path around the zero point and adding the differences in phase from pixel to pixel together. Ideally the result is an integer multiple of $2\pi$. In practice, it is necessary to accept a range of values close to integer multiples of $2\pi$. For example we can decide that any value in the range $\left[\frac{7\pi}{4}, \frac{9\pi}{4}\right]$ will be accepted as indicating the presence of a vortex with topological charge 1.

4. For zero locations with a phase path integral indicating a non-zero topological charge, designate these points as vortices.
5. Repeat this process at a range of defoci, resulting in knowledge of the location of all vortex core points in a three dimensional space. The data is then plotted and features of the vortex core behaviour observed.

Step 3 is technically difficult due to the unavoidable pixelation of data. It is necessary to choose a path for integration close enough to the zero of intensity to ensure that the path can not surround two vortices which are close together. However the path must be far enough from the zero to clearly identify the phase jump of $\pm 2\pi$ which indicates a branch cut. This problem can be reduced, at the cost of greater computational time, by using a larger number of pixels in the $xy$ plane of the simulated wavefunction. The computational resources available dictate at what level of detail it is possible to observe the vortex trajectory.

### 2.7 Vortex trajectory simulation

Fig. 2.10 shows the evolution of numerous vortex trajectories in a simulated wavefunction. The wavefunction is that produced by a 400keV electron beam exiting a 1000 Å thick crystal of GaAs, seen along the [110] zone axis. One repeating “unit” only of the periodic wavefunction (in the $xy$ direction) is shown. The defocus values are relative to the in-focus image. On inspection, the spontaneous creation and annihilation of vortices occurs frequently as the wave travels through space. This example is typical of the behaviour of vortex trajectories in electron diffraction. It is interesting to note that all the vortex trajectories seem to form single loops. Although more complex structures are permitted by the mathematical rules of vortex behaviour, they do not occur in our simulations. All the trajectories observed here have topological charge $m = \pm 1$.

Fig. 2.11 shows the evolution of a single vortex trajectory from the previous example. It clearly has the expected loop structure as discussed in earlier sections. Since it travels forward and backwards several times in the $z$ or defocus direction, the vortex behaviour associated with this trajectory is quite complex, involving a number of separate vortices. GaAs has a simple crystal structure and this trajectory is one of four similar ones contained in a repeating “unit” as shown in Fig. 2.12. The trajectories have the same shape, subject to translation and reflection, and do not interact with each other directly. This can be seen by looking at the trajectories down the defocus ($z$) axis as in Fig. 2.13. Note that the apparent gaps in the trajectories are a result of the pixelation of the data and do not represent actual breaks in the trajectory.
Figure 2.10: Evolution of vortex trajectories in a wavefunction propagating out of a GaAs crystal.

Figure 2.11: Evolution of a single vortex trajectory with a simple loop structure.
Figure 2.12: Vortex trajectories within an $xy$ “unit”.

Figure 2.13: Vortex trajectories from Fig. 2.12 looking down the $z$ axis.
2.8 Summary

This chapter starts with a discussion of some of the important properties of the phase of a quantum mechanical wavefunction. In particular we see that knowledge of the phase is essential for the solution of inverse problems involving such waves. Although this is a simple point, it is an extremely important idea in this thesis, since phase retrieval and inverse problems are the major themes of this work.

The chapter goes on to examine properties of the phase more closely, with particular attention being paid to vortices. Rules for their behaviour are presented and simulated trajectories given. It can be seen from the simulation results presented here, that it is normal for a number of vortices to coexist in a wavefunction such as that generated by passing an electron beam through a crystal. The trajectories tend to form simple loops although they may have many changes of direction. Junctions of four or more paths are not observed in these simulations, although technically they are possible. Vortex trajectories with topological charge greater than 1 do not occur in any simulation.

In Chapter 3 we go on to discuss phase retrieval, with emphasis on methods which can accurately retrieve phase when vortices are present. Having observed the frequency with which vortices can occur in the propagating wave field resulting from passing an electron beam through a crystal, it is clear that ability to cope with vortices is a major consideration for phase retrieval in dynamical electron diffraction. Thus a key requirement for a useful method of phase retrieval in this context is that it can successfully deal with vortices.
Chapter 3

Methods of phase retrieval

It is shown in Chapter 2 that knowledge of the phase of a quantum mechanical wavefunction is an important, often essential, requirement for understanding the physical situation that produces it. Many different methods exist for retrieving the phase of an observed wavefunction. They vary greatly in the types of measurement and the computational resources required. During the last two decades, considerable effort has been devoted to finding ways to retrieve not only the amplitude but also the phase of the complex wave function at the exit surface of a crystal for high-energy transmission electron diffraction. Holography [61–63] may allow retrieval of the phase information, as can the information in a through focal series [33–35,61,64–68], or a tilt-azimuth series [69] of images. Ptychography (from the Greek πτυξ, meaning fold), involves determining the phase of the wave function via the interference patterns in overlapping convergent beam electron diffraction (CBED) disks obtained using a coherent incident beam [22,70–72].

The work in this thesis is concerned with non-interferometric phase retrieval methods. These place greatly reduced coherence requirements on the source of radiation compared to their interferometric counterparts [30]. This is a useful property since it allows phase recovery in a situation where interferometric methods are largely ineffective [25]. The coherence of an electron source can be measured using methods such as that of James and Rodenburg [73].

This chapter begins with a description of two phase retrieval methods that are based on the Transport of Intensity Equation (TIE), which describes conservation of flux. These methods solve equations directly to retrieve the phase of the wavefunction. Next, three algorithms are discussed which use iteration to progressively reduce the error in a guessed wavefunction until it possesses the correct phase. Simulated results using all of these methods are given and analysed.
Since, as discussed in Chapter 2, the ability to cope with vortices is a critical requirement of a phase retrieval method used on wavefunctions where vortices are expected to be present, this is an important criterion used in evaluating the success of the methods. Other criteria include the speed, numerical stability and computational resources required for the algorithm, as well as the practicality of taking the needed measurements in an experimental situation.

3.1 Transport of Intensity Equation

The TIE describes continuity of the flux of a wavefunction as it travels through free space. It is a second order partial differential equation in the phase of a wave and its solution can therefore be used to gain complete knowledge of a wavefunction when the intensity is already known. In the absence of zeros of intensity, the TIE can be solved to find a unique solution of the phase problem [25]. The TIE is derived from the Schrödinger equation as follows.

In the absence of a potential, the Schrödinger equation for propagation of a wavefunction through space is written as

\[ (\nabla^2 + k^2)\psi(r) = 0, \] (3.1)

where \( k \) is the magnitude of the wave vector and the wavefunction can be written as \( \psi(r) = |\psi(r)|e^{i\theta(r)} \) with \( |\psi(r)| \) the magnitude of the wavefunction and \( \theta(r) \) its phase. Let us multiply Eq. 3.1 by the complex conjugate \( \psi^*(r) \) of the wavefunction to obtain

\[ \psi^*(r)\nabla^2\psi(r) + k^2\psi^*(r)\psi(r) = 0. \] (3.2)

Taking the complex conjugate of this equation we have also that

\[ \psi(r)\nabla^2\psi^*(r) + k^2\psi(r)\psi^*(r) = 0. \] (3.3)

Subtracting Eq. 3.3 from Eq. 3.2 gives

\[
0 = \psi^*(r)\nabla^2\psi(r) - \psi(r)\nabla^2\psi^*(r) \\
= \psi^*(r)\nabla \cdot (\nabla \psi(r)) + \nabla \psi(r) \cdot \nabla \psi^*(r) - \psi(r)\nabla \cdot (\nabla \psi^*(r)) - \nabla \psi^*(r) \cdot \nabla \psi(r) \\
= \nabla \cdot (\psi^*(r)\nabla \psi(r)) - \nabla \cdot (\psi(r)\nabla \psi^*(r)) \\
= \nabla \cdot (\psi^*(r)\nabla \psi(r) - \psi(r)\nabla \psi^*(r)). \] (3.4)
To evaluate the term in brackets we note from the definition of $\psi$ that

$$
\nabla \psi(r) = |\psi(r)| \nabla e^{i\theta(r)} + e^{i\theta(r)} \nabla |\psi(r)|
$$

$$
= |\psi(r)| e^{i\theta(r)} \nabla (i \psi(r)) + e^{i\theta(r)} \nabla |\psi(r)|
$$

$$
= e^{i\theta(r)} (i |\psi(r)| \nabla \theta(r) + \nabla |\psi(r)|) .
$$

(3.5)

Similarly,

$$
\nabla \psi^*(r) = e^{-i\theta(r)} (-i |\psi(r)| \nabla \theta(r) + \nabla |\psi(r)|) .
$$

(3.6)

Therefore

$$
0 = \nabla \cdot (\psi^*(r) \nabla \psi(r) - \psi(r) \nabla \psi^*(r))
$$

$$
= \nabla \cdot (|\psi(r)| e^{-i\theta(r)} e^{i\theta(r)} (i |\psi(r)| \nabla \theta(r) + \nabla |\psi(r)|))
$$

$$
- \nabla \cdot (e^{i\theta(r)} e^{-i\theta(r)} (-i |\psi(r)| \nabla \theta(r) + \nabla |\psi(r)|))
$$

$$
= \nabla \cdot (i |\psi(r)|^2 \nabla \theta(r) + |\psi(r)| \nabla |\psi(r)| + i |\psi(r)|^2 \nabla \theta - |\psi(r)| \nabla |\psi(r)|)
$$

$$
= \nabla \cdot (2i |\psi(r)|^2 \nabla \theta(r)) .
$$

(3.7)

From Eq. 3.7 we find the general TIE,

$$
\nabla \cdot (|\psi(r)|^2 \nabla \theta(r)) = 0 .
$$

(3.8)

Breaking Eq. 3.8 into its components parallel ($z$) and perpendicular ($\perp$) to the direction of travel of the wave, defined by $k$ we obtain

$$
\nabla \perp \cdot (|\psi(r)|^2 \nabla \theta(r)) + \partial_z (|\psi(r)|^2 \partial_z \theta(r)) = 0 ,
$$

(3.9)

and thus

$$
\nabla \perp \cdot (|\psi(r)|^2 \nabla \theta(r)) + \partial_z |\psi(r)|^2 \partial_z \theta(r) + |\psi(r)|^2 \partial_z^2 \theta(r) = 0 .
$$

(3.10)

For the applications considered, we make the paraxial approximation for propagation of a wave. This assumes that $\partial_z \theta(r) = 2\pi/\lambda = k$ and $\partial_z^2 \theta(r) = 0$. We therefore finally attain the paraxial form of the TIE as

$$
\nabla \perp \cdot (|\psi(r)|^2 \nabla \theta(r)) = -k \partial_z |\psi(r)|^2 .
$$

(3.11)

The paraxial TIE may be solved uniquely, provided that the intensity is strictly positive everywhere in the region of solution. This implies that the solu-
tion for the phase is continuous. Since the paraxial TIE contains the derivative of the wavefunction with respect to the direction of propagation \( z \), it is necessary for the methods based on this equation to assume knowledge of the derivative of the intensity of the wavefunction. In order to accurately estimate this, two or more measurements of intensity are needed at planes of propagation which are relatively close together. These methods are only suitable for experimental situations where this is practical. This requirement is in contrast to the iterative methods we later examine, which need measurement in distant planes, with relatively different intensities, for greater accuracy.

### 3.2 Teague, Paganin and Nugent method

This method, henceforth referred to as the TPN method, is based on the paraxial TIE. It extends a method of solution proposed by Teague [32] which has been further developed by Paganin and Nugent [25,30] using pseudo differential operators. The solution given here uses the standard Fourier transform representation of partial derivatives and has been published by Allen and Oxley [74].

Teague’s solution is based on the Helmholtz theorem [75]. This states that a vector field can be decomposed into a continuous scalar field \( \phi(\mathbf{r}) \) and a vector potential \( \mathbf{A}(\mathbf{r}) \) as follows:

\[
|\psi(\mathbf{r})|^2 \nabla_{\perp} \theta(\mathbf{r}) = \nabla_{\perp} \phi(\mathbf{r}) + [\nabla \times \mathbf{A}(\mathbf{r})]_{\perp} \tag{3.12}
\]

The Teague approximation assumes that the rotational term \( [\nabla \times \mathbf{A}(\mathbf{r})]_{\perp} = 0 \), allowing Eq. 3.12 to be rewritten as

\[
\nabla_{\perp} \theta(\mathbf{r}) = |\psi(\mathbf{r})|^{-2} \nabla_{\perp} \phi(\mathbf{r}) \quad (3.13)
\]

Thus the TIE becomes a Poisson equation for the auxiliary function \( \phi(\mathbf{r}_{\perp}, z) \), where \( \mathbf{r}_{\perp} \) is the real space coordinate in the \( xy \) plane.

\[
\nabla_{\perp}^2 \phi(\mathbf{r}_{\perp}, z) = -k \partial_z |\psi(\mathbf{r}_{\perp}, z)|^2. \tag{3.14}
\]

To solve this equation, we start with the standard Fourier transform representation for the \( n \)-th partial derivative with respect to \( x \) of the function \( f(\mathbf{r}_{\perp}) \). [76],

\[
\mathcal{F} \left[ \partial_x^{(n)} f(\mathbf{r}_{\perp}) \right] = i^n q_x^n \mathcal{F} \left[ f(\mathbf{r}_{\perp}) \right]. \tag{3.15}
\]
In this equation $x$ and $q_x$ are conjugate variables in real and reciprocal space respectively and $\mathcal{F}$ denotes the Fourier transform. Using this relation we write

$$\nabla_{\perp} f(r_\perp) = \mathcal{F}^{-1}\{i\hat{x}q_x\mathcal{F}[f(r_\perp)] + i\hat{y}q_y\mathcal{F}[f(r_\perp)]\}$$  \hspace{1cm} (3.16)

where $\hat{x}$ and $\hat{y}$ are unit vectors in the $x$ and $y$ directions. It follows that

$$\nabla_{\perp}^2 f(r_\perp) = -\mathcal{F}^{-1}\{q_x^2 \mathcal{F}[f(r_\perp)]\} = -\mathcal{F}^{-1}\{q_y^2 \mathcal{F}[f(r_\perp)]\}.$$  \hspace{1cm} (3.17)

Using this, Eq. 3.14 becomes

$$-\mathcal{F}^{-1}\{q_{\perp}^2 \mathcal{F}[\phi(r_\perp, z)]\} = -k \partial_z |\psi(r_\perp, z)|^2.$$  \hspace{1cm} (3.18)

Therefore we have the solution of Eq. 3.14:

$$\phi(r_\perp, z) = \mathcal{F}^{-1}\{q_{\perp}^{-2} \mathcal{F}[k \partial_z |\psi(r_\perp, z)|^2]\}.$$  \hspace{1cm} (3.19)

Taking the derivative of Eq. 3.13, we have

$$\nabla_{\perp}^2 \theta(r_\perp, z) = \nabla_{\perp} \cdot \left[ |\psi(r_\perp, z)|^{-2} \nabla_{\perp} \phi(r_\perp, z) \right].$$  \hspace{1cm} (3.20)

The quantity $\phi(r_\perp, z)$ can be calculated from Eq. 3.19. By applying Eq. 3.16 (twice) to Eq. 3.20, we can therefore compute the right hand side of this Poisson equation for the phase, $\theta(r_\perp, z)$.

Eq. 3.20 can be therefore now be solved using Eq. 3.17 to find the phase $\theta(r_\perp, z)$.

$$\theta(r_\perp, z) = -\mathcal{F}^{-1}\{q_{\perp}^{-2} \mathcal{F}[\nabla_{\perp} \cdot (|\psi(r_\perp, z)|^{-2} \nabla_{\perp} \phi(r_\perp, z))]\}.$$  \hspace{1cm} (3.21)

It is clear from this solution for the phase, that because the quantity $|\psi(r_\perp, z)|^{-2}$ is used, it is assumed the intensity of the wavefunction never falls to zero. This is a requirement for the method to find a unique solution. Therefore it is immediately obvious that singularities such as vortices and edge dislocations, which require regions of zero intensity to exist, are not retrieved accurately using this method. The method is also unable to retrieve a non-zero vector potential in the phase, due to the approximation made by Teague. The presence of a vortex core produces a rotational vector potential in the phase that may be significant over a relatively large area of the surrounding wavefunction, and a non-zero vector potential may also exist without any singularities present. This means that the
TPN method produces inaccurate results in such regions, which do not involve actual singularities of the phase [74].

3.3 Van Dyck and Coene method

The VDC method, developed by Van Dyck and Coene [33] is based on a Fourier analysis of the TIE as given in Eq. 3.11. Like the previous method, it assumes knowledge of the intensity of the wavefunction $|\psi|^2$ and its derivative $\partial \partial_z |\psi|^2$.

Eq. 3.11 expands to give

\[(\nabla_\perp |\psi(r)|^2)(\nabla_\perp \theta(r)) + |\psi(r)|^2 \nabla_\perp^2 \theta(r) = -k \partial_z |\psi(r)|^2 . \tag{3.22}\]

Differentiating the $|\psi(r)|^2$ terms gives

\[(2|\psi(r)|^2 \nabla_\perp |\psi(r)|) \cdot (\nabla_\perp \theta(r)) + |\psi(r)|^2 \nabla_\perp^2 \theta(r) = -2k |\psi(r)| \partial_z |\psi(r)| . \tag{3.23}\]

Assuming the previously discussed requirement that the intensity of the wavefunction is always greater than zero, cancelling $|\psi(r)|$ results in

\[2(\nabla_\perp |\psi(r)|) \cdot (\nabla_\perp \theta(r)) + |\psi(r)| \nabla_\perp^2 \theta(r) = -2k \partial_z |\psi(r)| . \tag{3.24}\]

For convenience we rewrite the last line as

\[\partial_z |\psi(r)| = -\frac{1}{2k} \left[ 2(\nabla_\perp |\psi(r)|) \cdot (\nabla_\perp \theta(r)) + |\psi(r)| \nabla_\perp^2 \theta(r) \right] . \tag{3.25}\]

The amplitude of the wavefunction can be expanded as a Fourier series as follows, where $r_\perp$ is the real space coordinate in the image plane, and $g$ the corresponding reciprocal space coordinate.

\[|\psi(r)| = \sum_g \Psi_g(z) e^{ig \cdot r_\perp} \tag{3.26}\]

Clearly, $\Psi_g$ are the Fourier coefficients of the amplitude. The derivative of the amplitude may be written as

\[\partial_z |\psi(r)| = \sum_g \frac{d\Psi_g(z)}{dz} e^{ig \cdot r_\perp} . \tag{3.27}\]
We complete the Fourier representation of Eq. 3.25 by expressing the phase of the wavefunction with the Fourier expansion

\[ |\theta(r)| = \sum_g \Theta_g(z) e^{ig \cdot r}. \quad (3.28) \]

After substituting Eqs. 3.26 – 3.28 into Eq. 3.25 and equating corresponding Fourier components, the following is obtained:

\[ \sum_{g'} \left[ g^2 - (g - g')^2 \right] \Psi_{g-g'}(z) \Theta_{g'}(z) = 2k \frac{d\Psi_g(z)}{dz}. \quad (3.29) \]

This system of coupled, linear equations can be solved to find the Fourier coefficients of the phase \( \Theta_g(z) \). There are \( n^2 \) equations, where \( n \) is the number of beams used in the Bloch wave representation of the periodic wavefunction. This is because \( n \) Bloch wave beams result in up to \( n^2 \) different Fourier frequencies due to the frequency cross-terms produced by subtracting the beams \( g \) and \( g' \). The result of this is that the computational resources required to solve these equations scale as \( n^4 \), creating a significant problem with this method. Since its demand for computational resources is far greater than other methods studied, the available level of accuracy from the VDC method can be lower when such resources are limited. It is important to make sure that there are no significant Fourier coefficients left out of the approximation, as this leads to an inaccurate determination of the phase.

Singular value decomposition (SVD) is used to solve the linear equations [77]. This algorithm is an appropriate and robust method for solving large systems of linear equations and examples of code which implement this method are widely available. Once the Fourier coefficients of the phase are found, it is simple to sum them together to recreate the original phase of the wavefunction.

An important and interesting observation to make about the VDC method is that it makes no assumptions about the wavefunction other than the assumption, implicit in the TIE, that the intensity of the wavefunction does not go to zero at any point. Thus this method allows the phase of the wavefunction to take any other form, as long as it can be represented by a Fourier series. While this excludes some particularly pathological functions, it does allow the vector potential contribution from Eq. 3.12 to be significant, unlike the TPN method.

Therefore, in a situation where the vector potential contribution is significant, we expect the VDC method to be more accurate than the TPN method. However
if the Fourier series of the wavefunction includes many high order terms, the VDC method requires a large amount of computational resources to obtain accurate solutions. In that case the TPN method may be more accurate. Neither method is accurate in a situation where the intensity is zero, so neither method will retrieve the phase when vortices or edge dislocations are present. In this situation, a continuous, non-unique solution that is physically inaccurate will be retrieved.

### 3.4 Gerchberg-Saxton method

This method is the first of a number of similar, iterative algorithms for phase retrieval, which are based on a set of measurements of the intensity of images or diffraction patterns. A description of the algorithm was first published by Gerchberg and Saxton in 1972 [78] with more analysis appearing in later papers [31]. Recently, there has been a great deal of interest in the use of the Gerchberg-Saxton (GS) algorithm and its variants for phase retrieval in general situations [26,27,79].

A general outline of the GS algorithm is shown in Fig. 3.1. The following steps are involved:

1. Start with the focused image. The amplitude of the wavefunction at this point, $|\psi(r)|$ is known. The phase is unknown, so we begin the algorithm by guessing a phase for the image.

2. Computationally propagate the entire wavefunction to the diffraction plane, where the intensities of the spots in the diffraction pattern are known. This propagation can be done using a Fast Fourier Transform (FFT), since the image plane and diffraction plane are related by the Fourier transform.\(^1\) Use of FFT algorithms greatly reduces the computation time required for this process.

3. Assuming the initial guess of the phase was wrong, the propagated wave now has incorrect intensities.\(^2\) However the correct intensities of the diffraction pattern are known from the data measurements. Correct the magnitude of

---

\(^1\)The diffraction pattern is actually a stretched version of the Fourier transform of the image. This is easily seen in the case of a periodic image since the diffraction pattern is a collection of points representing the Fourier frequencies in momentum space.

\(^2\)If the initial guess was lucky and the wavefunction phase is now correct the algorithm terminates.
the wavefunction to the measured intensities. Leave the incorrect propagated phase.

4. Take the inverse FFT to return to the image plane.

5. Again correct the wavefunction intensities. The error between the propagated and actual intensities is reduced. This point is discussed in Section 3.7 and proven mathematically in Appendix A.

6. Repeat steps 2 to 5, moving between the image and diffraction pattern, correcting the intensity each time. The error continues to decrease until the propagation produces the correct intensities (to within a suitably small margin) with each propagation step. At this point, since the wave is propagating correctly, it is known that the phase of the wave must also be correct [80–82]. The entire phased wavefunction is now known at each measurement location.
### 3.5 Fienup method

The Fienup algorithm is currently very popular in phase retrieval, having been used in experiments by Weierstall et al. [83], Miao et al. [84–86], and a number of others. It is similar in approach to the GS method. However instead of needing the intensity pattern in both the image and diffraction planes, it is only necessary to know the intensities in the diffraction plane. In the image plane, all that must be known is the shape of the support. This is the region of the image over which the intensity is non-zero. There exist applications for which the support is clearly known, or easily determined from other information known about the scattering object or wavefunction under examination. Weierstall et al [83] have explored other methods of finding the shape of the support, based on knowing the autocorrelation of the diffraction pattern, as well as limits for the size of the non-zero image compared to the surrounding region.

![Fienup Algorithm Diagram](image)

**Figure 3.2:** Fienup algorithm.

The Fienup method is appealing for experimental High Resolution Transmis-
sion Electron Microscopy (HRTEM) because it is relatively easy to measure the
diffraction pattern to high resolution, and this method does not require the more
difficult measurement of the focused image, or images at other defocus planes.
However the method uses less information than the GS method to retrieve the
phase, since it is only known that over a certain region the image intensity is zero, rather than knowing the exact intensity of the entire image. It is therefore likely
to be slower to converge than the GS method, and it is necessary to make sure
that there is enough information contained in the diffraction pattern and support
to actually find the phase solution. It turns out on analysis of the Nyquist rate
for sampling [87] that it is necessary for the support to cover at least half of
the image intensity, so at least half the image is entirely known in its intensity
and phase (the phase is assumed to be zero for the zero intensity regions). Even
when this amount of information is known, it appears that the Fienup method
does not retrieve the phase to the same level of accuracy as the other iterative
methods. Thus while it is useful in some contexts, and is certainly an area of
current research, it is not useful for this study compared to the GS method. The
Fienup method is not discussed further in this thesis.

3.6 Through Focal Series method

The Through Focal Series (TFS) method [26, 27], is illustrated schematically in
Fig. 3.3. It uses a similar algorithm to the GS method. However instead of using
images in the focused plane and the diffraction plane, the TFS method uses two
or more different images in a through focal series.\(^3\) This renders it potentially less
practical to use in an experimental situation, since it is more difficult to measure
defocused images than diffraction patterns. However if the wavefunction being
examined is non-periodic, the TFS method is easier to implement than the GS
method. This is because in a non-periodic situation it is difficult to find the
origin in the diffraction pattern plane which corresponds accurately to that in
the image plane. In other words the area of the diffraction pattern must match
the area of view in the image accurately [88]. The choice of origin in the image
is not crucial for periodic objects.

\(^3\)A derivation of the free space propagator used to propagate the wavefunction is given in
Appendix B. Since this free space propagator produces a wavefunction that repeats itself as
it travels, it is possible to have multiple images at different defoci which are in fact the same
and contain the same information. In the discussion that follows it is assumed that the images
used are different.
A number of other methods based on through focal series of images have been proposed. It is important to note that our method differs from these, being essentially based on the GS algorithm, rather than other methods such as the three dimensional paraboloid method [34, 35, 68] or the maximum likelihood method [66]. These alternate methods are not examined in this thesis.

A useful feature of the TFS method is the ability to include more than two images in the focal series. Including more information in the form of more images usually greatly reduces the running time of the algorithm.

![TFS Algorithm Diagram](image)

Figure 3.3: TFS algorithm for three planes.
3.7 Analysis of iterative algorithms.

The iterative algorithms are based around the fact that the iterative process progressively reduces the error in the propagated image, until the images in different planes are propagating correctly to and from each other. This means that the phase is also correct, since an incorrect phase would not produce a correctly propagated image. In special cases, such as when the object has a centre of symmetry, a well understood two-fold ambiguity arises. Criteria have been formulated by Huise [81,82] and others which allow one to check, after the phase retrieval calculations, whether the solution obtained is unique. The effect of noise on uniqueness can also be accounted for. In general, knowledge of the image and its diffraction pattern, as well as the implicit fact that the wave functions are analytic, is sufficient to phase both uniquely [80].

![Image](image.png)

Figure 3.4: Mixing of initial intensity and phase as the wavefunction is propagated to different defocus values.

It is easy to understand intuitively why these algorithms work by considering the fundamental fact that propagating a complex wave through space, whether to the diffraction plane or simply to another image plane, has the effect of mixing the phase and image information from the original plane. This can be seen in Fig. 3.4 where an intensity and a phase are combined to create a wavefunction which is then propagated through space. In the original image plane the phase and intensity are clearly separate and have very different features. However wavefunctions at further defoci show a mixing of the information contained in the original phase and intensity, so aspects of each picture can be seen combined.
In particular, the original phase information can clearly be seen mixed into the propagated images. This mixing means that errors in the phase at one defocus cause errors in the intensity at the next defocus. Correcting these errors in the intensity produces an increasingly better approximation to the phase and propagated intensity as the algorithm progresses.

In all the iterated algorithms, a Sum Square Error (SSE) is used to measure the success of the algorithm. The SSE at the nth iteration of the algorithm is defined as

$$SSE_n = \frac{1}{\sum |\psi_n|^2} \sum \left| |\psi_n| - |\psi_0| \right|^2,$$  \hspace{1cm} (3.30)

where $|\psi_n|$ is the amplitude of the guessed wavefunction at the nth iteration and $|\psi_0|$ is the amplitude of the correct wavefunction. The scaling factor $\sum |\psi_n|^2$ is the same at all planes, due to energy conservation, and allows comparison of cases with different energies contained in the wave. This error decreases at each iteration of the algorithm [31, 78, 89]. A proof of this is given in Appendix A.

Another way of examining the theory behind the iterative algorithms is to consider the mathematics of convex sets. A set of points in a space is mathematically convex if it is possible to draw a straight line joining any two points in the set, where the line does not leave the set. An example of a convex set is the set of points $(x, y)$, in a cartesian plane, contained inside a circle. We can draw a line between any two points inside the circle without the line ever going outside the circle. An example of a non convex two dimensional set is a star shape, since a line drawn between two extremes of the star may leave the set.

Fig. 3.5 shows diagrammatically the operation of the iterative algorithms for different pairs of sets. For simplicity only two sets are shown in each diagram, however this can be generalised to multiple sets. In this application the sets are the constraints on the intensities of the wavefunctions in the Fourier and image domains. The intersection of the sets is the set of solutions which satisfy both constraints and which therefore form physical solutions to the phase problem. There may be multiple solutions, depending on whether the sets overlap or just meet at one point.

If the two sets do not intersect there are no solutions. This situation is shown in Fig. 3.5(a). The algorithm moves from one set to the other, each time moving to the closest possible point in the other set. If it reaches a point where the sets intersect, a solution that satisfies both constraints has been found. If it can not reach an intersection point it will find a local minimum of the distance between the two sets, as is shown in the diagram.
Convex sets, no solution - iteration reaches local minimum.

Convex sets with solutions - iteration reaches nearest solution.

Non-convex sets - iteration reaches local minimum which is not a solution.

Non-convex sets - with a different starting point, iteration reaches solution.

Figure 3.5: Behaviour of iterative methods with convex and non-convex sets of constraints.

Fig. 3.5(b) shows the situation where the sets do intersect and both sets are convex. It can be seen that for convex sets, the starting point of the algorithm, which corresponds to the initial guess of phases does not matter because any starting point leads to a correct solution.

Compare this with the diagrams in Fig. 3.5(c) and Fig. 3.5(d), where the sets are not convex. Depending on the starting point, it is possible for the algorithm to find a local minimum which does not lead to a solution of the problem, as in Fig. 3.5(c). It will be clear from the large but stagnating SSE values that a solution has not been reached. In this situation it is necessary to restart the iterative process and hope that a true solution will be found eventually, such as is shown in Fig. 3.5(d).

It is clear that the convex set situation is very desirable, since the correct solution is reached on the first attempt. It is worth noting that knowledge of the phase of the wavefunction and knowledge of the shape of the support required for the Fienup algorithm are both convex constraints. However in practice it

51
may be difficult to find a set of constraints which are convex and practical for the algorithm in use. It is possible in some cases to convexify a non-convex constraint by mathematical manipulation, although this can be complicated and difficult to achieve in practice. These and other mathematical foundations of this material are discussed by Combettes [90].

The GS, Fienup and TFS algorithms all use the constraint that the intensity of the wavefunction in one or more planes is known and must be exactly satisfied. This is not a convex constraint. This can be seen by considering the constraint at one point in the $xy$ plane of the wavefunction. Requiring that the intensity, or modulus squared, of a wavefunction be exactly equal to a specific real number is like requiring the complex number to lie on a circle of radius $|\psi|$, where $|\psi|^2$ is the intensity. It is easy to understand that a set of points on the circumference of a circle is not a convex set, since a straight line between any point and any other point is not contained within the set. Thus the iterative methods discussed here do not have the ideal property of using convex constraints. We therefore note that it is possible for the iterative methods used to fall into local minima which are not solutions to the phase problem. However it is easy to identify a local minimum as opposed to a solution because the minimum does not produce a correctly propagating wavefunction, whereas a correct solution does. When a local minimum is found and the iteration does not converge to a sufficiently low error measure, the algorithm is restarted from a different starting position, and this is repeated until a solution is found. The theoretical problems posed by the lack of convex constraints do not pose practical problems, other than in the increased computational time required to find a solution.

An important observation to make about all the iterative methods discussed is that they do not make any assumptions about the form of the wavefunction other than that it is a complex scalar wave. These methods have their own potential problems, as discussed above, in that they may become caught in a local minimum. However they do not require the assumption that the intensity of the wavefunction is never zero as do the TPN or VDC methods. This makes a crucial difference in the ability of the methods to accurately retrieve phase when singularities such as vortices are present.

The iterative methods in general require measurements in planes that are relatively widely spaced along the axis of propagation, whereas the TIE requires measurements in planes close enough to allow accurate determination of the derivative. This is because it is important, for fast convergence of the algorithm,
to have different information contained in each image. The actual distance required depends on how fast the wavefunction is changing, but it is important to understand this difference in the required experimental setup for the two different types of method. This feature of the methods means that the TIE methods in general retrieve low frequency information accurately because the lower frequency components of the wave allow more accurate determination of the derivative. In contrast, the iterative methods retrieve high frequency detail effectively.

Having reached understanding of the several different methods of phase retrieval, we now go on to examine their behaviour in simulations. The next sections present the results of these methods for periodic and non-periodic situations, and come to conclusions about the most effective way to retrieve the phase in situations such as dynamical electron diffraction.

### 3.8 Phase retrieval results for periodic wavefunctions

We now calculate, compare and discuss some phase retrieval results for periodic wavefunctions. In particular, we examine simulations of the wavefunctions produced by passing electron beams through crystals, using Bloch wave theory as a tool to produce appropriate approximations of the wavefunctions. Bloch theory is discussed in more detail in Chapter 5 where the inverse problem in dynamical electron diffraction is considered. For the current purpose it should be noted that a periodic wavefunction may be approximated well with a finite number of Fourier coefficients and a discrete diffraction pattern made up of bright spots at locations representing the Fourier frequencies in diffraction space.

Phase retrieval using simulated images and diffraction patterns is demonstrated for 200 keV electrons incident on a crystalline slab of GaAs viewed along the exact [110] direction. Structure factors for elastic scattering are calculated using the atomic scattering factors of Waasmaier and Kirfel [91]. It is assumed that the crystal is at 300 K and absorption, in the form of thermal effects using temperature factors $B_{\text{Ga}} = 0.6476 \text{ Å}^2$ and $B_{\text{As}} = 0.6965 \text{ Å}^2$ [92], is incorporated. We include the effect of thermal diffuse scattering (TDS) in our calculations via an Einstein model [59, 93]. The mean free path for TDS is approximately 1700 Å. This choice of parameters in the model problem allows comparison with the experimental work of Haider et al. [39, 94, 95]. Bloch wave calculations are carried out in an $N = 27$ beam approximation. This is because calculations with
55 beams show small changes of detail compared to those with 27 beams but make the VDC method intractable since, as mentioned in Section 3.3, its need for computational resources scales as $N^4$. Including cross-terms, this means that 95 Fourier coefficients are included in the calculation of the projected optical potential.

![Unaberrated data - $C_s=0$ mm, $\Delta f = 0$ Å, $t=90$ Å.](image1)

![First test case - $C_s=1.23$ mm, $\Delta f = 641$ Å, $t=90$ Å.](image2)

![Second test case - $C_s=0.05$ mm, $\Delta f = 140$ Å, $t=90$ Å.](image3)

The unaberrated exit surface image of a unit cell and the corresponding phase map for $t = 90$ Å are shown in Figs. 3.6(a) and (b) respectively (128×91 pixels). The maximum value of the intensity in (a) is 4.58 and the minimum is $3.40 \times 10^{-5}$ (to be compared with unit intensity for the incoming plane wave). The arrows in the phase map in Fig. 3.6(b) show two vortex cores joined by a branch cut. Three other such pairs of vortices are visible.

Simulated images and phase maps for $t = 90$ Å with two different sets of values of the spherical aberration parameter $C_s$ and defocus $\Delta f$ are shown in Figs. 3.6(c)-(f). The two sets of data make good test cases for the phase retrieval algorithms. The first case, where $C_s = 1.23$ mm, corresponds to the value of $C_s$ of the Philips CM 200 FEG ST microscope used by Haider et al. [39,94,95]. We assume that $\Delta f = 641$ Å, the (extended) Scherzer defocus. The simulated image is shown in Fig. 3.6(c) and has a maximum intensity of 2.61 and a minimum of $2.18 \times 10^{-2}$. No vortices are evident in the corresponding phase map in Fig. 3.6(d).

![Unaberrated data - $C_s=0$ mm, $\Delta f = 0$ Å, $t=90$ Å.](image4)

![First test case - $C_s=1.23$ mm, $\Delta f = 641$ Å, $t=90$ Å.](image5)

![Second test case - $C_s=0.05$ mm, $\Delta f = 140$ Å, $t=90$ Å.](image6)

Figure 3.6: Test data for phase retrievals.

\[\text{Note that aberration determination and correction are discussed in Chapter 4 and for now simply consider } C_s \text{ and } \Delta f \text{ to be parameters which may easily be varied to produce different wavefunctions.}\]
This case serves to benchmark the five phase retrieval methods we test in the presence of vortices with the second case.

Using their aberration corrected microscope, Haider et al. [39, 94, 95] nevertheless worked with a residual spherical aberration with $C_s = 0.05$ mm and a defocus $\Delta f = 140 \, \text{Å}$, to enhance phase contrast. This set of parameters forms the second test case. The simulated image is shown in Fig. 3.6(e) and has maximum intensity 4.81 and minimum intensity $5.45 \times 10^{-4}$. The phase map, shown in Fig. 3.6(f), has four pairs of vortices, one of which is indicated by arrows.

![Image](image.png)

Figure 3.7: Input data for TPN and VDC methods, first test case - $C_s = 1.23$ mm, $\Delta f = 641 \, \text{Å}$, $t = 90 \, \text{Å}$.

![Image](image.png)

Figure 3.8: Input data for TPN and VDC methods, second test case - $C_s = 0.05$ mm, $\Delta f = 140 \, \text{Å}$, $t = 90 \, \text{Å}$.

The input data required by the TPN and VDC methods for the two cases is shown in Figs. 3.7 and 3.8. Each set of data consists of a through focal series.
of images with 20 Å separation between each image in the series. The defocus values in the figures are shown relative to the centre image in each series. So, for example, the overall defocus values for the first case are $\Delta f = 601, 621, 641, 661$ and 681 Å. These images are used to calculate the derivative of the intensity, using a five point formula at each pixel. The phases are also shown, although these of course do not form part of the input information. It is important to note that none of the phases in Fig. 3.7 contain vortices, whereas the phases in Figs. 3.8 do, as indicated by the arrows.

Figure 3.9: Input data for TFS method, first test case - $C_s = 1.23$ mm, $\Delta f = 641$ Å, $t = 90$ Å.

The TFS algorithm requires as input data images in a through focal series with greater step sizes than the TPN or VDC methods. The through focal series used for the TFS algorithm in the first test case is shown in Fig. 3.9. The step size for this series is 100 Å, resulting in a set of much more varied images (and phases) than those in Figs. 3.7 and 3.8. The series used for the TFS algorithms in the second case is shown in Fig. 3.10. There are vortices (examples marked with arrows) in every image in this series.

The GS method requires as input the focal plane image and the diffraction pattern produced by the wavefunction. The diffraction pattern for 200 keV electrons incident at the exact zone axis orientation is shown in Fig. 3.11. The focal plane images for the two cases are shown in Fig. 3.6 (c) and (e).

---

5It is important to note that this relatively small defocus variation is used in simulation assuming that there is no limitation in temporal coherence. The presence of focal spread could require the use of greater defocus steps, since the defocus variation for each image could be larger than the defocus step to the next image. We note that Van Dyck and Coene used an even smaller defocus step of 5 Å in their work [33].

56
Figure 3.10: Input data for TFS method, second test case - $C_s = 0.05$ mm, $\Delta f = 140$ Å, $t = 90$ Å.

Figure 3.11: Diffraction pattern data required by GS algorithm.

The original and retrieved phases for the two cases are shown in Fig. 3.12. Using the TPN method to retrieve the phase for the first case results in the phase map shown in Fig. 3.12(a). The input phase map is given in Fig. 3.6(d), which is reproduced in Fig. 3.12(e) for ease of comparison. The phases are shown wrapped into the interval between $-\pi$ and $\pi$ and, for each parameter set, the phase maps are normalised so that the grey scale level of the pixel in the top left hand corner is constant. The differences between the input and retrieved phases are due to the neglect of the vector potential in this method of phase retrieval.

Fig. 3.12(b) shows the result using the VDC method for the first case. It reproduces the input model phases to within a few percent, a clear improvement on the TPN method for this case. This greater accuracy is because the VDC method allows the vector potential to be non-zero.

In Fig. 3.12(c) we show the result of using the TFS algorithm for the first case. The input through focal series in this case is that shown in Fig. 3.9. Convergence using the centre three of these images is reasonably fast but five images give improved rates of convergence similar to the GS method. An SSE less than $10^{-15}$ is obtained after 83 iterations, with five images, starting from an initial guess of a constant phase across the unit cell. Excellent agreement with the model phase
map is obtained.

For both the TPN and TFS methods we need to know the distance between the images and their alignment. This information is not needed for periodic objects if the GS algorithm is used. In Fig. 3.12(d) we show the phase obtained using the GS method, starting from the image and diffraction pattern in Figs. 3.6(c) and 3.11 respectively. An SSE of $10^{-15}$ is achieved after 150 iterations (once again starting from a constant phase).

![Figure 3.12: Phase retrieval comparison for periodic images.](image)

We now consider the second phase retrieval case. As expected given the presence of vortices, retrieval of the phase using the TPN method fails as can be seen from the retrieved phase in Fig. 3.12(f). The VDC method is also unable to retrieve phase in the presence of vortices, although an incorrect continuous phase is reproduced.

The TFS method, based on the series of five images 100 Å apart shown in Figs. 3.10, retrieves the phase correctly, even though there are vortices in the phase maps for all these defocus values. Convergence to an SSE of less than $10^{-15}$ is obtained after 104 iterations. The GS algorithm, using the image and diffraction pattern in Figs. 3.6(e) and 3.11 respectively, also recovers the phase excellently, as seen in Fig. 3.12(i). Convergence is obtained after 127 iterations to an SSE of again less than $10^{-15}$.

In practice both the TFS and GS methods must utilise images with resolution constrained by the information limit of the system. The diffraction data may extend to a higher resolution. In the GS method the accuracy of the phase map retrieved in the image plane can be improved by a knowledge of the diffraction pattern to higher orders (the correct relative normalisation of image and
diffraction pattern needs to be considered). An important consideration in the phase retrieval and subsequent aberration correction described here is the fact that contrast in experimental images is often considerably less than in simulated images. The possible causes for this discrepancy have been discussed in detail by Boothroyd [97]. The lack of contrast is, to a good approximation, a constant factor and a correction needs to be made to the images. However in the aberration-corrected instrument of Haider et al. [39, 94, 95] images show “a remarkable suppression of artifacts and a strong increase in contrast apart from the improved resolution”. Therefore the phase retrieval using the TFS or GS algorithms, followed by correction of the known aberrations\(^6\), and used in tandem with their experimental approach will deliver the best outcomes, correcting other aberrations at the same time.

### 3.9 Noise analysis of iterative algorithms

![Phase retrieval for noisy data - retrieved and aberration corrected phases.](image)

Figure 3.13: Phase retrieval for noisy data - retrieved and aberration corrected phases.

Simulations are also run with noise in images and diffraction patterns for both the TFS and the GS methods. We assign a number of counts to each pixel assuming that an intensity of unity in either an image or diffraction pattern (the incident plane wave has unit intensity) corresponds firstly to 100 counts and then to 10000 counts. This corresponds to noise at the 10% and 1% level respectively.

\(^6\)This is discussed in Chapter 4.
The statistical errors at each pixel are assigned using a random deviate drawn from a Poisson distribution with mean corresponding to the noise-free number of counts for a given pixel [77]. Correct (albeit noisy) results are once again obtained, as can be seen in Fig. 3.13. This shows the retrieved phases, in (a)-(e), starting from the noisy images and diffraction patterns calculated with $C_s = 0.05$ mm, $\Delta f = 140$ Å and $t = 90$ Å and the phases after subsequent correction for spherical aberration and defocus [36], in (f) to (i). It is interesting to note that Fig. 3.13(b) contains an additional pair of vortices (one of which is indicated by the arrow) not present in the model phase map. These spurious vortices arise as a consequence of the noisy input data. However on correction for the spherical aberration and defocus, the spurious vortices are no longer present, as seen in Fig. 3.13(g).

As was pointed out by Haider et al., an energy of 200 keV avoids radiation damage since the energy of the electrons is less than the threshold for knock on processes. Therefore, given stability of an experimental system, long exposure times should be possible which would reduce statistical noise. Smoothing to eliminate noise in the experimental data, using for example the method of generalised cross-validation [98], would further enhance the speed of convergence and also minimise the possibility that noise would lead to a spurious solution.

### 3.10 Phase retrieval for non-periodic wavefunctions

Phase retrieval for non-periodic wavefunctions has been a topic of much interest recently [27, 38, 79, 83–85]. With that in mind, we now consider the non-periodic problem. When phase retrieval is performed on non-periodic wavefunctions, it is important to remember that the GS method is not so useful because of problems associated with locating the origin of the diffraction pattern. Since we are interested in exploring phase retrieval for a general non-periodic wavefunction, vortices may be present. Therefore neither the TPN method or the VDC method are likely to be accurate. In this case, the only practical algorithm of those considered is the TFS method. This is the algorithm used in the following investigation.

The specific application of phase retrieval methods to non-periodic wavefunc-

---

7This aberration correction is discussed in Chapter 4
tions containing vortices has not previously been studied in depth and is therefore both interesting and important to investigate. The method is applicable to a wide range of scenarios such as the imaging of imperfect crystals, quantitative determination of the strength of vortex filaments in high-temperature superconductors and X-ray and electron holography [26,27,99].

We generate wave fronts with vortices in the phase from an input model wave function without vortices which has input intensity (Mark Twain) and phase map (Louis-Victor de Broglie) as shown in Fig. 3.14. The intensity varies between 0 and 1 and the phase varies between $-\pi$ and $\pi$. The arbitrary nature of this input data serves to highlight the fact, mentioned earlier, that almost any choice of images for the probability density and phase of the wave function will lead to phase vortices after the wave field has been allowed to propagate a sufficient distance through free space.

Pictures originally containing $128 \times 128$ pixels are padded out to $160 \times 160$ pixels - by ones in the intensity and zeros in the phase. This padding is necessary because of the mathematical properties of the Fourier transforms used to perform the iterative calculations. Firstly, the FFT algorithm implicitly assumes it is dealing with a periodic object, and requires periodic boundary conditions. Secondly, padding the initial image and phase allows the wavefunction to expand as it propagates through space, as it would naturally do.

Our example can be considered to apply to two cases. Firstly assume that the intensity and phase map in Fig. 3.14 (each $160 \times 160$ pixels) have dimensions $75 \, \text{Å} \times 75 \, \text{Å}$ and are formed by a beam of electrons of energy 100 keV (wavelength $0.037 \, \text{Å}$. This case is equivalent to a situation where we can consider the image in Fig. 3.14 to have dimensions $1.28 \, \text{mm} \times 1.28 \, \text{mm}$ and the incident radiation to be laser light of wavelength $6328 \, \text{Å}$ (HeNe laser). Then the image and phase maps in Fig. 3.15 correspond to the defocus values indicated in the square brackets.
Propagation of the wave function in free space yields the series of phase maps shown in Fig. 3.15. Noise is added at the 5% level. Vortices are present in all these phase maps. One such vortex is indicated by an arrow in each phase map. The vortices lie on trajectories which may be tracked, as is discussed in Chapter 2. The vortex trajectory found for the vortices shown by the arrows in Fig. 3.15 is shown in Fig. 3.16.\footnote{The lack of resolution for this trajectory compared to those given in Figs. 2.10-2.13 is because the trajectory is much smaller. It is difficult, given the available computational resources, to simulate and track a small trajectory with a large enough number of pixels in the image to produce better resolution. However the general shape can be seen, and in principle there is no reason why a more detailed examination of this trajectory could not be made.}

As mentioned, we use the TFS method to perform the phase retrieval. We discuss our approach to the phase retrieval in terms of the electron optics case but it is equally applicable to the laser optics case just mentioned (or any other equivalent situation). An SSE of $6.48 \times 10^{-4}$ is obtained after 939 iterations. Although this SSE is much larger than those achieved in simulations when noise was not included, it does represent a converged (noisy) solution. The retrieved wave function at $\Delta f = 500$ Å is shown in Fig. 3.17. Also shown is the wave function which is then obtained by back propagation to zero defocus, which can be compared with the wave function in Fig. 3.14. The effect of noise in the focal series in Fig. 3.15 is mainly evident in the intensity rather than the phase. Using
Figure 3.16: Vortex trajectory corresponding to arrows in Fig. 3.15.

Reproduced wavefunction, \( \Delta f = 500 \text{ Å} \) [8.55 mm]

Reproduced wavefunction at original defocus, \( \Delta f = 0 \text{ Å} \) [0 mm]

Figure 3.17: Image and recovered phase map for \( \Delta f = 500 \text{ Å} \) and propagated back to \( \Delta f = 0 \text{ Å} \).
five images 250 Å apart, convergence is obtained after only 230 iterations.

Tested on numerous other focal series of three or more images the method is always successful. It is clear that the TFS algorithm is equally applicable to periodic and non-periodic objects. A practical consideration for non-periodic applications is that the images need to be carefully aligned so that they cover the equivalent regions in the plane of the propagating wavefunction.

3.11 Summary

In this chapter several possible algorithms for phase retrieval are discussed and compared. Particular attention is paid to the case where vortices occur in the phase of the wave function at the image plane. Methods based on conservation of flux, which assume that phase is continuous in an image plane, are not robust in the presence of vortices. In addition, the TPN method does not retrieve the phase when a vector potential is present. Both the TFS and GS approaches are efficient methods of phase retrieval in the presence of vortices when the wavefunction is periodic. The TFS method has the advantage that it is also applicable to non-periodic objects while the GS algorithm has the advantage that the alignment and spacing of images is not an issue.

When the wavefunction in question is not periodic and it may contain vortices, the only practical method of phase retrieval, (of those considered) is the TFS method. Since it can not be proven that this iterative method will find a solution of the problem, issues of finding a correct solution arise. Fortunately, a local minimum that is not a solution may be easily identified, since the error will stagnate and fail to become small enough to indicate that a correct solution has been found.

There are many anticipated applications of the method of phase retrieval demonstrated. Immediate applications may be made to the quantitative imaging of crystal defects using HRTEM, where phase vortices are ubiquitous [26]. The method may also be readily applied to the imaging of high-temperature superconductors containing a network of vortex filaments [99]. In the context of transmission electron microscopy on such samples, the correct phasing of the exit surface wavefunction is easily related to the strength of the vortex filaments in the superconductor and thereby to their desirable technological properties. Lastly, the methods presented can be applied to electron and X-ray holographic microscopy.
In conclusion, there exist many scenarios where the presence of vortex-like phase singularities demands a phase retrieval algorithm which is able to cope with their existence. The ideas discussed here comprise a simple, robust and rapid method for the phase retrieval of both quantum mechanical and classical wave fields in the presence of vortices. This method is an essential tool required to implement much of the work presented in the remainder of this thesis.
Chapter 4

Aberration determination and correction

Having identified a method of phase retrieval that is useful in the context of electron optics, we now consider the solution of some inverse problems which require accurate phase retrieval as a tool. The first of these is the issue of aberration control in optical systems. Although the material presented here is generally applicable to many areas of microscopy, we concentrate on electron microscopy in particular as the application of interest.

Aberration determination and correction is a significant issue in electron microscopy. Many researchers must put a great deal of effort into first characterising and then removing aberrations from their imaging equipment, in order to obtain the best possible data. This chapter presents and tests a practical method of computationally determining the characteristic aberrations of a generalised imaging system and then removing these aberrations using software, if that is desired. The method is easily applicable to any system where a point source is incident onto a lens system, and a through focal series of three or more images can be taken in order to perform phase retrieval. By using software instead of hardware to control the aberrations, significant resources can be saved and corrections can be made which are not possible with existing hardware.

4.1 Background to aberration control

In an ideal microscope, light, an electron beam, or some other wavefunction can be processed and modified without unintentionally changing any of its characteristics. In practice all microscopes have some aberrations due to imperfections
in the lenses and other details of the experimental setup. This may be desirable, since aberrations can be used to increase contrast in an image, as in phase contrast imaging [100, 101], or in other ways aid the understanding and interpretation of the image information. However, aberrations may cause problems because they reduce the maximum resolution obtainable by the imaging system and they distort the information present. It is useful to be able to characterise the aberrations present in a quantitative way.

For example, in HRTEM, aberrations of the objective lens corrupt spatial information over length scales smaller than the point resolution (the least resolvable distance in an image). This corruption of information reduces the resolution that can in practice be obtained, by orders of magnitude, leading to much less detailed images. An electron microscope with a coefficient of spherical aberration $C_s$ has a point resolution given by $d \approx 0.66(C_s \lambda^3)^{1/4}$, where $\lambda$ is the wavelength of the incident electrons [102]. Decreasing $d$ by decreasing $\lambda$, via an increase in the energy of the incident electrons, has the disadvantage of increasing radiation damage in the sample. Therefore the most effective way to obtain higher resolution images is to remove or reduce the aberrations.

The question of aberration determination and correction has been studied theoretically and experimentally in a variety of fields. Various experimental methods of reducing spherical and other aberrations have been found and are regularly used [37, 39, 94, 95, 103–106]. Some such methods do not in general allow characterisation of the aberrations, so the actual degree of aberration present is unknown. In addition, many methods only deal with spherical aberration and defocus, whereas it is becoming increasingly important to take higher order aberrations into account [38].

In recent years, progress has been made in aberration characterisation, albeit in limited situations. Kirkland et al. have achieved aberration compensation and image reconstruction using tile-azimuth data [69]. Saxton [13, 107] has described a method for characterising aberrations using diffractogram orientations. Meyer et al. have recently found a method for characterising symmetric aberrations that is based on analysis of the phase correlation function between pairs of images [108]. Lehman [14] has used a heuristic algorithm to characterise aberrations in off-axis electron holography to an accuracy of about 5%. While successful, this algorithm requires initial knowledge of the approximate values of the aberrations in question. Steinecker and Mader [15] have achieved aberration determination to a similar accuracy using image displacements in a beam tilt series in HRTEM.
Wang et al. have implemented a method for characterisation and correction of aberrations based on an amplitude contrast criterion for a holographic image \[109\]. Although these studies represent improvements on previous methods, they do not characterise all aberrations present in a system accurately. It is desirable to be able to do this, as well as to find increasingly accurate methods of aberration determination which make no initial assumptions about the aberration values.

The method presented here computationally determines all coherent aberrations present in a shift-invariant linear imaging system such as an electron microscope\(^1\), and can be extended to also characterise incoherent aberrations. If desired, the aberrations in the image data may then be corrected using software (rather than optical hardware) using an approach proposed by Allen et al. \[26,36\]. Unlike some previous methods, the algorithm presented here simultaneously characterises all the aberrations present. This is a practical improvement on techniques which determine each aberration with a separate measurement, since only one set of experimental data needs to be taken. The method, which is readily automated, finds aberration coefficients to within 0.5\% of their actual value for the example considered here.

### 4.2 Aberration representation

Image formation in HRTEM is usually described by Abbe’s theory \[102,111\]. The effect of aberrations on the image can be described in momentum space by the contrast transfer function \[102,111\]

\[
T(q) = A(q) \exp[iW(q)].
\]

(4.1)

Here \(A(q)\) is a damping function which can be obtained as a product of envelope functions relating to chromatic aberration, the shape dependence of the source, a sample drift envelope, a specimen vibration envelope, the shape of the objective aperture and a detector envelope function \[112\]. This damping function characterises the incoherent aberrations in the system. The function \(W(q)\) is a phase factor characterizing the coherent aberrations in the system, with \(q\) a vector in the reciprocal space. It is a sum of the phase factors for whatever coherent aberrations are present in the imaging system.

\(^1\)A shift-invariant linear imaging system is one in which a linear combination of inputs yields the same linear combination of outputs and a transverse shift in input leads to a shift proportional to this in the output \[110\].
While the method presented here is able to simultaneously recover all orders of aberration, in simulations we consider only defocus, spherical aberration and axial astigmatism, since in practice these are often the three most important aberrations. For these aberrations only, we have [16]

\[
W(q) = W(q, \phi) = \pi(-\Delta f \lambda q^2 + \frac{1}{2} C_s \lambda^3 q^4 + \frac{1}{2} C_a \lambda q^2 \sin 2\phi), \tag{4.2}
\]

where \(\Delta f\) is the defocus (assumed positive for underfocus), \(C_s\) is the spherical aberration coefficient, \(C_a\) is the coefficient of axial astigmatism, \(q = |q|\) and \(\phi\) is the polar angle describing the direction of \(q\).

In order to correct for aberrations we need to know the real space wave function in the (principal) image plane, \(\psi_a = |\psi_a| \exp(i\theta_a)\), where \(\theta_a\) is the aberrated phase and \(|\psi_a|^2\) the aberrated intensity. The wave function corrected for aberrations (in the focal plane) is given by \(\psi_c = |\psi_c| \exp(i\theta_c)\). In momentum space the corresponding wave functions are \(\Psi_a\) and \(\Psi_c\). Assuming that we know the phase \(\theta_a\) of \(\psi_a\), then \(\Psi_a\) in momentum space is found from

\[
\Psi_a = \mathcal{F}[\psi_a], \tag{4.3}
\]

where \(\mathcal{F}\) denotes a Fourier transform. Since \(\Psi_a = A(q) \exp[iW(q)]\Psi_c\) it follows that

\[
\Psi_c = A^{-1}(q) \exp(-iW(q))\Psi_a, \tag{4.4}
\]

where it is assumed that we are considering only values of \(q\) for which the shape function \(A(q)\) is not zero. Therefore, knowing \(A(q)\) and \(W(q)\), we can construct the wave function corrected for these aberrations in the momentum space. The corrected wave function in the focal plane is then obtained by the inverse Fourier transformation

\[
\psi_c = \mathcal{F}^{-1}[\Psi_c], \tag{4.5}
\]

from which the aberration-corrected image \(|\psi_c|^2\) and aberration-corrected phase \(\theta_c\) can be obtained.

A convenient way of expressing \(W(q)\), especially when a number of different aberrations are of interest, is to use the Zernike polynomials [113, 114], the first 21 of which are shown in Fig. 4.1. The Zernike polynomials form an orthonormal basis in two dimensions inside the unit disk [115]. They are useful because a disk is a convenient shape for many optical situations involving lenses, and because the polynomials are closely related in form to many of the optical aberrations.
Figure 4.1: The first 21 Zernike polynomials.

For example, the 5th Zernike polynomial describes axial astigmatism and the 11th polynomial describes a combination of spherical aberration and defocus. Table 4.1 gives the forms of the first 21 Zernike polynomials in terms of the polar angle $\phi$ and the radius $\rho \equiv q/q_{\text{max}}$, where $q_{\text{max}}$ is defined to be the radius of the disk outside of which the damping function $A(q)$ is negligibly small. Expressing $W(q)$ in terms of the Zernike polynomials we have

$$W(q) = \sum_i C_i Z_i(\rho, \phi) = \sum_i C_i Z_i(q/q_{\text{max}}, \phi) \equiv \sum_i C_i Z_i(q),$$

where $Z_i$ is the $i$-th Zernike polynomial and $C_i$ is the dimensionless coefficient of the $i$-th Zernike polynomial. Since the Zernike polynomials form an orthonormal basis, we can decompose $W(q)$ into individual polynomials by finding the
projection of $W(q)$ onto each polynomial $Z_i(q)$ as follows:

$$C_i = \int W(q) Z_i(q) d\mathbf{q}.$$  \hspace{1cm} (4.7)

where $C_i$ is the coefficient of the $i$-th Zernike polynomial. The $C_i$ can then be related back to the coefficients of the aberrations of interest to complete the characterization of the aberrations. Specifically for the three aberrations of interest, we observe that the phase factor given in Eq. 4.2 uses the polynomials $Z_1$, $Z_4$, $Z_5$ and $Z_{11}$, given in Table 4.1.

In terms of $\rho$, Eq. 4.2 becomes

$$W(\rho, \phi) = \pi \left( -\Delta f \lambda \rho^2 q_{\text{max}}^2 + \frac{1}{2} C_s \lambda^3 \rho^4 q_{\text{max}}^4 + \frac{1}{2} C_a \lambda \rho^2 q_{\text{max}}^2 \sin 2\phi \right)$$  \hspace{1cm} (4.8)

Equating coefficients of $\rho$ gives the following equations, where $C_1$, $C_4$, $C_5$ and $C_{11}$ are the coefficients of the relevant Zernike polynomials:

$$0 = C_1 - \sqrt{3} C_4 + \sqrt{5} C_{11}$$  \hspace{1cm} (4.9)

$$-\pi \Delta f \lambda q_{\text{max}}^2 = 2 \sqrt{\frac{3}{\pi}} C_4 - 6 \sqrt{\frac{5}{\pi}} C_{11}$$  \hspace{1cm} (4.10)
Eqs. 4.9 to 4.12 give the standard aberration coefficients in terms of the coefficients of the Zernike equations. To find the Zernike coefficients in terms of the coefficients of the standard aberrations we must solve the equations as follows. From Eq. 4.12 we have that

\[ C_5 = \sqrt{\frac{\pi}{6}} \frac{C_a \lambda q_{max}^2}{\sqrt{6}} = \frac{\pi^{\frac{3}{2}} C_a \lambda q_{max}^2}{2\sqrt{6}}. \] (4.13)

From Eq. 4.11 we have that

\[ C_{11} = \sqrt{\frac{\pi}{5}} \frac{C_s \lambda^3 q_{max}^4}{12\sqrt{5}} = \frac{\pi^{\frac{3}{2}} C_s \lambda^3 q_{max}^4}{12\sqrt{5}}. \] (4.14)

Substituting Eq. 4.14 into Eq. 4.10 we have that

\[
C_4 = \frac{1}{2} \sqrt{\frac{\pi}{3}} \left( 6 \sqrt{\frac{5}{\pi}} C_{11} - \pi \Delta f \lambda q_{max}^2 \right) = 3 \sqrt{\frac{5}{3}} C_{11} - \frac{\pi^{\frac{3}{2}} \Delta f \lambda q_{max}^2}{2\sqrt{3}} = \sqrt{15} C_{11} - \frac{\pi^{\frac{3}{2}} \Delta f \lambda q_{max}^2}{2\sqrt{3}} = \frac{\pi^{\frac{3}{2}} C_s \lambda^3 q_{max}^4}{4\sqrt{3}} - \frac{\pi^{\frac{3}{2}} \Delta f \lambda q_{max}^2}{2\sqrt{3}}. \] (4.15)
Substituting Eqs. 4.14 and 4.15 into Eq. 4.9 we have that

\[ C_1 = \sqrt{3} C_4 - \sqrt{5} C_{11} \]

\[ = \sqrt{3} \left( \frac{\pi \frac{3}{2} C_s \lambda^3 q_{max}^4}{4 \sqrt{3}} - \frac{\pi \frac{3}{2} \Delta f \lambda q_{max}^2}{2 \sqrt{3}} \right) - \sqrt{5} \frac{\pi \frac{3}{2} C_s \lambda^3 q_{max}^4}{12 \sqrt{5}} \]

\[ = \frac{\pi \frac{3}{2} C_s \lambda^3 q_{max}^4}{4} - \frac{\pi \frac{3}{2} \Delta f \lambda q_{max}^2}{2} - \frac{\pi \frac{3}{2} C_s \lambda^3 q_{max}^4}{12} \]

\[ = \frac{\pi \frac{3}{2} C_s \lambda^3 q_{max}^4}{6} - \frac{\pi \frac{3}{2} \Delta f \lambda q_{max}^2}{2}. \]  

(4.16)

Eqs. 4.13 to 4.16 allow characterisation of the three aberrations of interest in this case in terms of the relevant Zernike polynomials.

**4.3 Aberration determination**

Consider the schematic depiction of a shift-invariant linear imaging system which is shown in Fig. 4.2(a). Here we have the imaging system illuminated by a point source in the absence of a sample and are considering a through focal series of three or more two dimensional intensity measurements over planes \( \Pi_1, \Pi_2, \Pi_3 \cdots \), which are perpendicular to the nominal optic axis. These images are modified, relative to those produced by a perfect system, by the aberrations that are present. Our method will take such a through focal series as data and obtain the aberrations affecting the images.

![Figure 4.2: Setup for characterisation and removal of the aberrations of a given imaging system.](image_url)
The algorithm for aberration determination is as follows:

1. Perform a computational phase retrieval using the TFS method.

2. Having determined the aberrated wave functions over the image planes, select one, which we refer to as the principal plane. (The wave function for a plane towards the middle of the through focal series is usually more accurately determined.) Transform this wave function into momentum space using the FFT.

3. Now decompose, using Eq. 4.7, the phase of the aberrated wave function in momentum space into its component Zernike polynomials. The coefficients of the Zernike polynomials describe the coherent aberrations present in the imaging system. As many coefficients as desired may be examined to determine aberrations up to the order of interest. We can then solve a small set of simultaneous equations, such as Eq. 4.9 to 4.12, to find the usual coefficients of the coherent aberrations. The incoherent aberrations can also be determined from the modulus of the aberrated wave function in momentum space.

### 4.4 Aberration correction

Consider the scenario sketched in Fig. 4.2(b), where we have introduced a sample of interest into the imaging system, which has now been characterised using the procedure outlined in the previous section. It is crucial to realise that the value of $\Delta f$ obtained in the aberration determination, described in the previous section, is specific to a point source at a specific location and without any sample between the source and the imaging system. For an incident plane wave or with a sample in place, the focal plane changes. However the rest of the contrast transfer function, namely $T'(q) = T(q) \exp(i\pi \Delta f \lambda q^2) = A(q) \exp[iW'(q)]$, remains unchanged. It is assumed that the imaging system can create distorted images of the sample in a series of planes $\Lambda_1, \Lambda_2, \cdots$ separated by a known distance. From this series of aberrated images we can obtain the phase of the aberrated wave function in these planes, using the TFS method.

We now proceed as follows:

1. Take the FFT of the real space wave function to find the aberrated momentum space wave function, $\Psi_a$. 

2. Multiply this in momentum space by \( T'^{-1}(q) = A^{-1}(q) \exp[-iW'(q)] \) to correct for aberrations other than defocus. If this contrast transfer function has zeros, then a regularised form [116] of this procedure involves multiplication by the Tikhonov-regularised filter \( T'^*(q)/(|T'(q)|^2 + \alpha) \) [117], where the regularization parameter \( \alpha \) is a small positive number.

3. If the defocus is known, correct for this by multiplying in momentum space by the free space propagator \( \exp(-i\pi \Delta f \lambda q^2) \). Otherwise generate a through focal series of images with the aim of identifying that with \( \Delta f = 0 \) or utilizing those which are most useful in interpreting the structure of the sample.

4. Take the inverse FFT of the momentum space wave function(s) to find the aberration-corrected wave function(s) in real space.

Of course it is not always desirable to remove all of the aberrations known to be present, since some experimental applications require a small amount of aberration. In these cases, only the desired corrections are made. By using the two steps of aberration determination and correction, control of the aberrations present in the imaging system is achieved using software techniques.

4.5 Model example

We investigate a simple model example in which the aberrations as described by Zernike polynomials are first characterised and then removed from the simulated wave function. The incident wave function is assumed to be formed by a point source emitting electrons with wavelength \( \lambda = 0.025 \text{ Å} \). This corresponds to a 200 keV electron beam. The image size is 200 Å, and the aperture has a radius of 0.5 Å\(^{-1} \) in momentum space. The envelope function \( A(q) = 1 \) inside the aperture and is zero elsewhere. Our simple example is that of an imaging system with spherical aberration \( C_s = 1 \text{ mm} \), defocus \( \Delta f = 1000 \text{ Å} \) and an axial astigmatism coefficient \( C_a = 500 \text{ Å} \). It is assumed that in all other respects the imaging system is ideal. The Zernike representation for these input aberrations is given in Table 4.2. In this case only \( C_1, C_4, C_5 \) and \( C_{11} \) are nonzero. The values for these coefficients are calculated using Eqs. 4.13 to 4.16.

This combination of Zernike coefficients creates the phase of the aberrated wave function in the momentum space shown in Fig. 4.3(a). Inside the momentum space aperture, the aberrated wave function intensity is a constant, unit intensity.
Table 4.2: Original and retrieved Zernike Coefficients $C_i$.

<table>
<thead>
<tr>
<th>$i$</th>
<th>Input</th>
<th>Retrieved</th>
<th>% Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-8.338</td>
<td>-7.340</td>
<td>11.97</td>
</tr>
<tr>
<td>2</td>
<td>0.000</td>
<td>8.518E-004</td>
<td>–</td>
</tr>
<tr>
<td>4</td>
<td>-2.198</td>
<td>-2.189</td>
<td>0.41</td>
</tr>
<tr>
<td>6</td>
<td>0.000</td>
<td>7.792E-004</td>
<td>–</td>
</tr>
<tr>
<td>8</td>
<td>0.000</td>
<td>7.668E-004</td>
<td>–</td>
</tr>
<tr>
<td>10</td>
<td>0.000</td>
<td>1.128E-003</td>
<td>–</td>
</tr>
<tr>
<td>12</td>
<td>0.000</td>
<td>9.946E-003</td>
<td>–</td>
</tr>
<tr>
<td>14</td>
<td>0.000</td>
<td>1.760E-002</td>
<td>–</td>
</tr>
<tr>
<td>16</td>
<td>0.000</td>
<td>-1.063E-003</td>
<td>–</td>
</tr>
<tr>
<td>18</td>
<td>0.000</td>
<td>1.113E-003</td>
<td>–</td>
</tr>
<tr>
<td>20</td>
<td>0.000</td>
<td>1.3685E-003</td>
<td>–</td>
</tr>
</tbody>
</table>

This is the Fourier transform of the original point source in real space. This wave function is transformed into real space, creating the aberrated intensity and phase in the principal plane, as shown in Figs. 4.4(a) and (b). Note that for an unaberrated image we would expect to see the familiar, circular Airy pattern in the centre [118]. In the phase map shown in Fig. 4.4(b) the large amount of wrapping of the aberrated phase in the image plane is a prominent feature. In Figs. 4.4(c) and (d) the central part of the image and phase respectively are shown enlarged. The deviation from the simple Airy pattern is obvious. In particular the image and phase are no longer circular due to the presence of astigmatism. Vortices are present in the phase map and one, which is connected to another by a branch line, is indicated by the arrow.

The aberrated wave function in the principal image plane is computationally propagated using the Fourier space representation of the free space Fresnel propa-
Note that this method is also applicable when other propagators such as the Kirchhoff propagator [119, 120] or a propagator in aberration space [36] need to be used. Five images are created with a propagation distance between each image of 1000 Å, as shown in Fig. 4.5. Noise, at the level of 0.25% on the brightest pixel of the central image, is included, with greater noise on all other pixels. This exceeds the noise level expected in most experimental situations.

A phase retrieval is performed using the TFS method. This process produces an aberrated wave function with a phase wrapped between $\pi$ and $-\pi$. It is relatively simple to computationally unwrap the two dimensional phase, since there are no discontinuities in the aberrated momentum space phase produced by the sum of Zernike polynomials. The unwrapped phase, shown in Fig. 4.3(b), is decomposed using Eq. 4.7 to produce the retrieved coefficients of the Zernike polynomials, shown in Table 4.2.

As can be seen, the coefficients $C_4$, $C_5$ and $C_{11}$ are retrieved to within 0.5% of their original values. Most of the remaining Zernike coefficients retrieved are found to be zero to a high accuracy, as expected. However not all the supposedly zero coefficients are accurately retrieved. The notable exceptions are $C_1$, $C_{13}$ and $C_{14}$. Looking at the form of each of these inaccurately retrieved Zernike coefficients, it is possible to see why the inaccuracies exist. $Z_1$ is the constant polynomial. Its presence is a result of the fact that there is always an overall

\[2\text{See Appendix B for the derivation of this propagator}\]

\[\begin{array}{c}
\text{Aberrated Image} \\
\text{Aberrated Phase}
\end{array}\]

\[\begin{array}{c}
\text{Aberrated Image} \\
\text{Aberrated Phase}
\end{array}\]

Figure 4.4: Image and phase of the aberrated wave function in the principal image plane.
unmeasurable phase shift which does not affect the physics of the situation. Thus the $C_1$ value is arbitrary and can be ignored.

The reason for the discrepancies in the $C_{13}$ and $C_{14}$ values can be seen by looking at the form of these Zernike polynomials, as given in Table 4.1. $Z_{13}$ contains in its form the expression $\rho^2 \sin 2\phi$ which is the expression for axial astigmatism given in $Z_5$. The similarity of the polynomials means that where there are slight inaccuracies in the process of applying Eq. 4.7 to find the projection of the total aberration onto $Z_{13}$, such as those introduced by numerical integration techniques, the presence of $Z_5$ in the aberrations results in inaccuracy in the $C_{13}$ result. This effect is a result of the pixelation and discretisation of the data.

The inaccuracy in $C_{14}$ is because $Z_{14}$ has high intensity areas concentrated at the horizontal and vertical extremes of the disk (i.e. the north, south, east and west points). Therefore projections onto this polynomial greatly emphasise information in these regions of the disk. Looking at the disk shape produced by the aperture of the lens, one can see that pixelation errors are most severe at these endpoints of the disk, because of the horizontal/vertical orientation of the pixel grid. In effect, the inaccuracy of the $C_{14}$ coefficient is a result of the disk being not strictly circular due to pixelation effects.

Fortunately, it is easy to identify coefficients which may be affected by errors of this type. If it is known that aberrations such as $Z_{13}$ and $Z_{14}$ are not actually present, or if they are not of interest, these inaccurate terms may simply be ignored in the final aberration correction step. This does not affect

Figure 4.5: Through focal series used to retrieve phase and the corresponding phase maps in each plane.
the accuracy of determination of the aberrations of interest, since each projection
calculation is independent of the others. Thus these errors do not pose major dif-
ficulties in this method of aberration determination. In addition, using a greater
number of pixels in the calculation results in great reduction of this type of error.
The errors reduce by roughly an order of magnitude with each doubling of the
number of pixels along the image edge, which results in a halving of the momentum
space pixel separation. So if higher accuracy is desired, it can be achieved
with suitable experimental data.

Using the retrieved coefficients of interest, namely those which make up the
spherical aberration, defocus and axial astigmatism known to be present in the
original system, the aberration coefficients are recovered, using Eqs. 4.10 to 4.12,
as $C_s = 1.004$ mm (input as 1.000 mm), $\Delta f = 1002$ Å (input as 1000 Å) and
$C_a = 501.9$ Å (input as 500.0 Å). These values are correct to better than 0.5%,
demonstrating the high level of accuracy of the aberration characterization. It is
important to recall that, unlike $C_s$ and $C_a$, the defocus $\Delta f$ depends on the fact
that the point source is at a given position relative to the imaging system.

Figure 4.6: Effect of aberration correction on measured wavefunction image and
phase.
Having accurately characterised the aberration coefficients of the model system using the through focal series of images in Fig. 4.5, we are ready to computationally remove the effect of those aberrations on the images taken of a sample of interest. In this context, we use a picture of Mark Twain for the unaberrated wave function image and a picture of Louis-Victor de Broglie for the unaberrated phase, as shown in Figs. 4.6(a) and 4.6(b). The image and phase are 50 Å ×50 Å square and are represented by 128 × 128 pixels. They are padded by zeros to 512 × 512 pixels before spherical aberration with $C_s = 1$ mm, astigmatism with $C_a = 500$ Å and a defocus $\Delta f = 577.4$ Å (i.e. Scherzer defocus, assumed to be known independently of the aberration characterization step) are added. The resulting wave function is shown in Figs. 4.6(c) and (d). Clearly, the image is distorted beyond recognition. It would be very difficult to find accurate information about a sample if its exit surface image had been so changed. While the phase retains some features it too is greatly altered. Generating a through focal series of images about this principal plane, we retrieve the phase and then correct for aberrations using $C_s = 1.004$ mm, $C_a = 501.9$ Å (as determined in the aberration determination step) and our known value $\Delta f = 577.4$ Å. We obtain the wave function shown in Figs. 4.6(e) and (f). This is a greatly improved result, which demonstrates the suitability of this method for aberration determination and subsequent computational aberration correction for a shift-invariant linear imaging system.

4.6 Summary

In this chapter, we have proposed a computational method for characterizing and subsequently computationally correcting aberrations, in an arbitrary shift-invariant linear imaging system, such as might be used in an electron microscope. Unlike existing methods of aberration determination, the method developed here is quite general and can, without modification, determine all coherent (and in principle incoherent) aberrations of a given linear imaging system. The method uses the Zernike polynomials as a convenient basis for describing coherent aberrations, and retrieves existing Zernike coefficients with an accuracy of better than 0.5% for the model case considered here. The method proposed has many potential applications in image processing and microscopy.

The ability to identify and subsequently remove aberrations in software, rather than relying on hardware, is an attractive feature of the method, allowing us to
reach the theoretical information limit of a given imperfect imaging system [36]. This advance is likely to be of particular importance in contexts such as HRTEM, and to a lesser extent X-ray microscopy [121] and neutron microscopy [122]. In all of these contexts the absence of aberration free optical elements can be compensated for using software rather than hardware, according to the methods outlined in this chapter. Our particular focus is in the field of HRTEM, where this method of aberration control may facilitate improved interpretation of HRTEM images.

An advantage of this method flows from the fact that it is able to recover both the amplitude and phase of the radiation which is being used in the given imaging system. As such, it confers total knowledge of the information encoded in the wavefunction after passage through the object under study. This total knowledge of the information contained in a given experiment gives us the power to emulate the action of any other imaging system, a fact which has already been demonstrated in at least two different contexts [123,124]. As pointed out earlier, we do not necessarily desire to negate all of the aberrations present in a given imaging system, and may instead opt to use the retrieved unaberrated intensity and phase to computationally emulate the action of such “aberrated” imaging systems as those which yield Zernike phase contrast [100], differential interference contrast [101] etc. The ability to emulate such a suite of imaging systems given images obtained using a single aberrated imaging system is a powerful motivation for this work.
Chapter 5

Inversion of dynamical electron diffraction

This chapter examines the second inverse problem considered in this thesis, which is a problem in the field of scattering theory. Scattering theory is the study of systems in which a wave, such as light or an electron beam, is incident onto a scatterer, such as a single atom, a crystal or another structure. The input wave interacts with the scatterer on an atomic level. If the structure of the scattering object is known with sufficient accuracy, the resulting complex wavefunction can be predicted theoretically. Scattering theory has many applications in areas such as optics, crystallography, tomography and any situation where an image is formed by measurement of a scattered beam.

In this context, an inverse problem arises when the structure of the scattering object is not known but the intensity of the exiting wavefunction can be measured. The problem is to use intensity measurements to deduce the atomic structure of the scattering object in terms of its projected or two dimensional potential.

We specifically consider the situation known as dynamical electron diffraction, where the scattering wave is an electron beam and the scatterer is a periodic crystal with the property that the electrons interact strongly with the scattering material. This is the case when an electron beam is incident onto a crystal that is not extremely thin. In this case, known as the dynamical or multiple scattering case, each electron is assumed to interact with the crystal atoms many times during its passage through the material.\footnote{We note that of course dynamical scattering occurs in non-periodic objects too.} This results in a complex scattering situation that, although well understood, is difficult to invert to find the original scattering potential.

\footnote{We note that of course dynamical scattering occurs in non-periodic objects too.}
This chapter starts with a discussion of the direct problem in dynamical electron diffraction. The difficulty of solving the inverse problem in the dynamical scattering case as opposed to the single scattering case is examined. A method of solving the inversion problem, including the effects of absorption, is then given and model calculations shown. The success of the method in the presence of noise is examined. Finally, possible extension of this method to a more complex situation, where the crystal has a stacking fault, is discussed.

5.1 Direct problem

The direct problem in dynamical electron diffraction is the situation in which complete knowledge of the incoming wave and the crystal structure are used to find the amplitude and phase of the wavefunction in the crystal or at its exit surface. This research deals with the inverse problem, where the structure or potential is determined from scattering data taken at several orientations of the incident electron beam. However, it is instructive to consider the forward problem first.

The Schrödinger equation for the wavefunction of a fast electron scattered by the periodic crystal potential \( V(r) \) is given by [58]

\[
\nabla^2 \psi(r) + \frac{2m|e|}{\hbar^2} [E + V(r)] \psi(r) = 0 .
\]

(5.1)

\( V(r) \) can be described by the Fourier series

\[
V(r) = \frac{\hbar^2}{2m} \sum_g W_g \exp(ig \cdot r) = \sum_g V_g \exp(ig \cdot r) .
\]

(5.2)

The values \( W_g \) are the Fourier coefficients of the optical potential for the scattering of electrons from the crystal. They can be broken up as follows

\[
W_g = U_g + iU'_g ,
\]

(5.3)

where \( U_g \) is the Fourier coefficient associated with elastic scattering processes, and \( U'_g \) is the coefficient associated with absorption. Therefore, the real part of the periodic crystal potential (associated mainly with elastic Coulomb scattering) is written as

\[
V(r) = \frac{\hbar^2}{2m} \sum_g U_g \exp(ig \cdot r) .
\]

(5.4)
The effect of absorptive (inelastic) scattering processes is taken into account by a component \( iV'(r) \) in the optical potential, where \( V'(r) \) may be written as

\[
V'(r) = \frac{\hbar^2}{2m} \sum_g U'_g \exp(i\mathbf{g} \cdot \mathbf{r}).
\] (5.5)

The vectors \( \mathbf{g} \) are reciprocal lattice vectors. It is assumed that if \( \mathbf{g} \) is included in the sum, \( -\mathbf{g} \) is also included. The wavefunction \( \psi(r) \) can be expanded in a sum of Bloch states as

\[
\psi(r) = \sum_i \alpha_i \phi^i(r) = \sum_i \alpha^i \sum_g C^i_g \exp[i(k^i + g) \cdot r].
\] (5.6)

The physical meaning of Eq. 5.6 is shown schematically in Fig. 5.1. The wavefunction inside the crystal consists of a set of superimposed waves, \( \phi(r) \), known as Bloch states, which consist themselves of sums of exponential terms with coefficients \( C^i_g \). The coefficients \( \alpha^i \) are obtained from the boundary conditions at the entrance surface of the crystal, which require that the amplitude of the directly transmitted beam is one and the amplitudes of the diffracted beams are zero.

Each Bloch state \( \phi^i(r) \) is characterised by an intrinsic wave vector \( k^i \), which depends on both the energy of the incident beam and the crystal structure, and can be obtained from the solution of the Schrödinger equation. The wave vectors \( k^i \) can be expressed in the form

\[
k^i = K + \lambda^i \hat{n},
\] (5.7)
where \( \mathbf{K} \) is the wave vector of the incoming plane wave in the crystal and the complex eigenvalue \( \lambda^i = \gamma^i + i\eta^i \) \([18,58]\). The unit vector \( \mathbf{\hat{n}} \) is a surface normal directed into the entrance crystal surface, the values \( \gamma^i \) are the \textit{anpassung} and the values \( \eta^i \) are the absorption coefficients.

Using the expansions given in Eqs. 5.2 and 5.6, the Schrödinger equation can be recast into the form of an eigenvalue problem as \([23,58]\)

\[
\mathcal{A}\mathbf{C} = 2\mathbf{K}[\lambda^i]_D .
\]

Here \( \mathcal{A} \) is the structure (Bethe) matrix and has the form

\[
\mathcal{A} = \begin{pmatrix}
\vdots & \vdots & \vdots & \vdots & \vdots \\
\cdots & -(k_t + h)^2 + iU'_0 & W_{h-g} & W_h & W_{h+g} & W_{2h} & \cdots \\
\cdots & W_{g-h} & -(k_t + g)^2 + iU'_0 & W_g & W_{2g} & W_{g+h} & \cdots \\
\cdots & W_{-h} & W_{-g} & -k_t^2 + iU'_0 & W_{-g} & W_{-h} & \cdots \\
\cdots & W_{-g-h} & W_{-2g} & W_{-g} & -(k_t - g)^2 + iU'_0 & W_{-g+h} & \cdots \\
\cdots & W_{-2h} & W_{-h-g} & W_{-h} & W_{h+g} & -(k_t - h)^2 + iU'_0 & \cdots \\
\cdots & \vdots & \vdots & \vdots & \vdots & \vdots & \cdots 
\end{pmatrix}
\]

The non diagonal elements are the Fourier coefficients of the crystal potential. The diagonal elements are found from the known reciprocal lattice vectors \((g, h, \ldots)\) and the tangential components of the incident wave vector \(k_t\).

The \( \mathcal{A} \) matrix possesses an important symmetry across the anti-diagonal which can be expressed as

\[
\mathcal{A}_{m,n} = \mathcal{A}_{N+1-n,N+1-m} , \quad \text{with } m \neq n \text{ if } k_t \neq 0 .
\]

\( \mathbf{C} \) is the matrix of eigenvectors of \( \mathcal{A} \), and contains elements as follows

\[
\mathbf{C} = \begin{pmatrix}
\vdots & \vdots & \vdots & \vdots \\
C^i_{1h} & C^i_{2h} & \cdots & C^i_{Nh} \\
C^i_{1g} & C^i_{2g} & \cdots & C^i_{Ng} \\
C^i_{10} & C^i_{20} & \cdots & C^i_{N0} \\
C^i_{1-g} & C^i_{2-g} & \cdots & C^i_{N-g} \\
C^i_{-h} & C^i_{-2h} & \cdots & C^i_{-Nh} \\
\vdots & \vdots & \vdots & \vdots 
\end{pmatrix} .
\]

It has been found \([18,125]\) that \( \alpha^i = [\mathbf{C}^{-1}]_{i,(N+1)/2} \), i.e. it is the \( i \)th element in the column of \( \mathbf{C}^{-1} \) which corresponds to \( g = 0 \). \([\lambda]_D \) is a diagonal matrix containing
the eigenvalues of $\mathcal{A}$.

Since the potentials $V(\mathbf{r})$ and $V'(\mathbf{r})$ given by Eqs. 5.4 and 5.5 respectively are real, $U_\mathbf{g} = U^*_\mathbf{g}$ and $U'_\mathbf{g} = U'^*_\mathbf{g}$. If the potential is assumed to be entirely elastic, $\mathcal{A}$ is Hermitian and the matrix of eigenvectors given by Eq. 5.11 is therefore unitary. This simplifies the inversion problem mathematically for situations where absorption is not a large effect, since the mathematical properties of unitary and Hermitian matrices can be used as information to help solve the inverse problem. However, in the presence of absorption, $(U_\mathbf{g} + iU'_\mathbf{g}) \neq (U^-_\mathbf{g} + iU'^-_\mathbf{g})^*$. In these situations $\mathcal{A}$ is not Hermitian and $\mathcal{C}$ is not unitary.

Let us assume an $N$-beam approximation ($\mathcal{A}$ becomes $N \times N$). Furthermore let us relabel the elements of the corresponding $N \times N$ eigenvector matrix $\mathcal{C}$ as follows:

$$
\mathcal{C} = \begin{pmatrix}
C_{11} & C_{12} & \ldots & C_{1i} & \ldots & C_{1N} \\
C_{21} & C_{22} & \ldots & C_{2i} & \ldots & C_{2N} \\
\vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
C_{N1} & C_{N2} & \ldots & C_{Ni} & \ldots & C_{NN}
\end{pmatrix}
$$

(5.12)

This relabelling facilitates writing down the representation for the elements of $\mathcal{A}$, which follows from Eq. 5.8. Then we can write

$$
\mathcal{A}_{m,n} = 2K \sum_i C_{mi} \lambda_i [C^{-1}]_{in},
$$

(5.13)

where the sum extends over $N$ terms. $[C^{-1}]_{in}$ is the element in the $i$th row and $n$th column of the inverse of $\mathcal{C}$.

Knowledge of elements of the matrix $\mathcal{C}$, the eigenvalues $\lambda_i$ from Eq. 5.8 and the incoming wavevector $K$ allow the construction of the wave function of the fast electron in the crystal as a sum of Bloch states, according to Eq. 5.6.

At the exit surface of the crystal the Bloch waves decouple into plane waves, as indicated in Fig. 5.1. At this transition the tangential components remain unchanged and therefore the amplitude of the beam $g$ at the exit surface of a crystal of thickness $t$ is obtained from Eq. 5.6 as

$$
v_g(t) = \sum_i \alpha^i C^i_g \exp(i\lambda^i t).
$$

(5.14)

---

2This is a standard property of the discrete Fourier transform [126].
Introducing the vector $v = (v_g)$ we find the exit surface wavefunction using
\[ v = Su , \] (5.15)

where the vector $u = (\delta_{g0})$ characterises the incident beam (a plane wave) and
\[ S = \exp \left( \frac{it}{2K} A \right) = C[\exp(i\lambda't)]_D C^{-1} \equiv C[A^i]_D C^{-1} \] (5.16)
is the scattering matrix (again $[\ ]_D$ denotes a diagonal matrix). The scattering matrix relates the incident electron wave at the entrance surface of the crystal to the diffracted wave at the exit surface of the crystal of thickness $t$ [58]. We can represent $S$ as
\[
S = \begin{pmatrix}
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\ldots & S_{h,h} & S_{h,g} & S_{h,0} & S_{h,-g} & S_{h,-h} & \ldots \\
\ldots & S_{g,h} & S_{g,g} & S_{g,0} & S_{g,-g} & S_{g,-h} & \ldots \\
\ldots & S_{0,h} & S_{0,g} & S_{0,0} & S_{0,-g} & S_{0,-h} & \ldots \\
\ldots & S_{-g,h} & S_{-g,g} & S_{-g,0} & S_{-g,-g} & S_{-g,-h} & \ldots \\
\ldots & S_{-h,h} & S_{-h,g} & S_{-h,0} & S_{-h,-g} & S_{-h,-h} & \ldots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots 
\end{pmatrix} . \] (5.17)

For an $N$-beam approximation we have the representation
\[ S_{g_n,g_m} = \sum_i C_{ni} \exp(i\lambda't)[C^{-1}]_{im} \] (5.18)
for the $S$ matrix, where $i$, $m$ and $n$ range over all $N$ beams.

We note that if $k_t = 0$, the $S$ matrix possesses the same symmetry as the $A$ matrix, namely that
\[ S_{m,n} = S_{N+1-n,N+1-m} , \quad \text{if } k_t = 0 . \] (5.19)

This is true because $S$ may be expressed in a Taylor series as a sum of multiples of $A$,
\[ S = I + \frac{it}{2K} A + \left( \frac{it}{2K} \right)^2 \frac{A^2}{2!} + \left( \frac{it}{2K} \right)^3 \frac{A^3}{3!} + \ldots . \] (5.20)

If two matrices commute, and both possess the anti-diagonal symmetry described for every element in the matrix, then their product also has that symmetry. $A$
commutes with itself and possesses the symmetry, in every element when $k_t = 0$. Thus every term in Eq. 5.20, and therefore $S$, also possesses the symmetry. However if $k_t \neq 0$ then $A$ does not possess the symmetry along its diagonal and thus $S$ does not have the symmetry at all.

Having reached an understanding of the direct problem in dynamical electron diffraction, we now go on to consider ways of inverting this process to find the potential of the scattering object. This is known as solving the inversion problem in dynamical electron diffraction.

### 5.2 Inverse problem

Attempts have been made to solve the inversion problem in dynamical electron diffraction for decades. Recent advances were made by Allen et al. [19, 20] and Spence [22, 23] using general inversion schemes which unambiguously obtain the projected potential for scattering data obtained from any perfect crystalline slab. Rez [24] has suggested schemes for determining the crystal potential by varying the incident energy and Sinkler and Marks [127] have discussed the use of minimum relative entropy approaches for samples with a thickness of up to 200 Å. Another method, more effective for multiple scattering, is to use a fixed orientation of the crystal and to take measurements for different crystal thicknesses [128].

This work has made solution of this problem possible for an increasingly wide number of situations and different scattering objects. The overall aim is to solve this problem in a way that is generally applicable and practical to implement.

The solution of the problem involves using measurements of the intensity of a wavefunction exiting the scattering crystal to determine its projected potential. It applies in situations where the structure is unknown, but the exiting images or diffraction patterns can be measured. If the scattering object is thin enough, the inverse problem is relatively simple. This is because in thin crystals it is reasonable to assume that the electrons interact only once with the crystal, or in other words that single scattering only is taking place. The relative simplicity of the single scattering situation can be seen from Figs. 5.2 (a) and (b) which show the projected potential of a GaAs crystal down the [110] zone axis and the corresponding image for a 400 keV electron beam incident onto a 90 Å crystal. Single scattering is an accurate approximation for this situation. For simplicity, the effects of absorption are neglected in this figure. It is clear that the image is very similar in structure to the potential, and information about the atomic
Figure 5.2: Projected potential and images for single and multiple scattering at 400keV.

positions can easily be deduced from the image.

If multiple scattering occurs the problem becomes much more complex. This can be seen in Fig. 5.2 (c) which shows the image for a crystal of 180 Å thickness. Using only this image to deduce the atomic positions or other information contained in the projected potential will produce erroneous results. In this case, more information than is contained in one image is needed to find the potential. Multiple scattering may be a significant factor in samples as thin as 100 Å, and it is desirable to use samples which are much thicker than this, to reduce the difficulties associated with sample preparation and radiation damage.

The inverse problem for dynamical electron diffraction consists of two distinct parts. The first part is to find the complex \( S \) matrix using available measurements. One issue with this is that, as discussed in Chapter 2, the structure of a general scattering object is encoded in both the amplitude and phase of the wavefunction. Therefore, it is necessary to determine the entire complex exit surface wavefunction, using the TFS phase retrieval method described in Chapter 3.

The other difficulty with finding the \( S \) matrix is that, as mentioned, it is impossible to find enough information to deduce the entire matrix from one measurement only of the complex wavefunction. Multiple measurements at different orientations of the incident beam are required. In the \( N \) beam case described previously, it is necessary to measure images at the principal orientation \((k_t = 0)\) and \( N - 1 \) other well defined orientations \((k_t \neq 0)\) in order to find all \( N \) columns of \( S \). This corresponds to knowing both the phase and magnitude of all the elements of the scattering matrix \( S \).

Once the entire complex \( S \) matrix is known, the second part of the problem is to obtain \( A \) via an inversion from \( S \). Eq. 5.16 leads to the following relationship
between $\mathcal{S}$ and $\mathcal{A}$:
\[
\mathcal{A} = \frac{2K}{it} \ln(\mathcal{S}) = \frac{2K}{it} C[i\lambda^i t] D C^\dagger.
\] (5.21)

The eigenvectors $C^i$ can be found by diagonalising $\mathcal{S}$. However the complex logarithm is not a unique function, and therefore an infinite number of possible values for $\lambda$ exist:
\[
i\lambda^i t = i(\theta^i + 2n^i \pi), \quad n^i = 0, \pm 1, \pm 2, \ldots,
\] (5.22)

where $\theta$ is the principal value. This ambiguity must be resolved in order to find $\mathcal{A}$ uniquely, using Eq. 5.13. Once this is done, the Fourier coefficients of the potential are known, and the crystal potential can be determined using the Fourier expansion given in Eq. 5.2.

### 5.3 Inversion of single scattering

Attempts have been made to implement an inversion of the dynamical scattering process and recover the projected potential from the exit surface wave function for a single orientation of the incident beam [40, 41, 129–134], assuming that the space group of the lattice is known (for example by the method proposed by [42]).

Using Eq. 5.21, it is necessary to know all the elements of $\mathcal{S}$ in order to find $\mathcal{A}$. However methods of inversion do exist for crystals thin enough to result in single scattering, which do not require a full knowledge of $\mathcal{S}$. In this situation, i.e. when $\frac{it}{2K}$ is small, the relationship between $\mathcal{S}$ and $\mathcal{A}$ given in Eq. 5.20 can be described by the lowest order term in the exponential. Thus,
\[
\mathcal{S} \approx I + \frac{it}{2K} \mathcal{A}.
\] (5.23)

In situations where Eq. 5.23 is accurate, it is clear that the Fourier coefficients in the central column of the $\mathcal{A}$ matrix can be described in terms of the centre column of $\mathcal{S}$ as follows:
\[
\mathcal{A}_{i,c} = \frac{2K}{it} (S_{i,c} - \delta_{i,c}) ,
\] (5.24)

where $S_{i,c}$ and $\mathcal{A}_{i,c}$ are the elements in the $i$th row and centre column of $\mathcal{S}$ and $\mathcal{A}$ respectively. This linear relationship means that the Fourier coefficients of the exit surface wavefunction (measured at the principal orientation), contained in the central column of $\mathcal{S}$, can be used to directly obtain the Fourier coefficients of the projected potential, contained in the central column of $\mathcal{A}$. This is a simple
problem to solve.

As soon as a thicker crystal is considered, the higher order terms of the exponential are necessary to calculate the potential accurately. This means that there is no longer a linear relationship between a single element of \( A \) and any single element of \( S \). In these situations it is necessary to know all of \( S \) in order to determine any element of \( A \), producing a much more complex inverse problem, namely the inversion problem in dynamical electron diffraction.

### 5.4 Inversion of dynamical scattering

To resolve the ambiguities expressed in Eq. 5.22 we proceed in a manner analogous to that outlined by Allen et al. [20] for the case of elastic scattering only. In this case we solve the problem for symmetric (\( k_t = 0 \)) orientations of the incident beam since we can use the known symmetry of \( S \) expressed in Eq. 5.19 to phase each column of \( S \) relative to the others. With the inclusion of absorption, the elements on the diagonal of the \( A \)-matrix are given by

\[
A_{n,n} = -(k_t + g_n)^2 + iU' \tag{5.25}
\]

where the additional term \( U' \) gives the mean value of the absorption potential. Using Eq. 5.13, this leads to a set of \( N \) linear equations in the \( N \) unknown \( \lambda_i \)'s:

\[
\sum_i C_{ni}\left[C^{-1}\right]_{in}\lambda_i = \frac{-(k_i + g_n)^2 + iU'}{2K} \tag{5.26}
\]

The use of these equations in the dynamical inversion implies knowledge of both \( U_0 \) and \( U'_0 \) (the former implicit in \( K \), since \( K^2 = k^2 + U_0 \)). If \( U_0 \) is not known, the approximation \( U_0 = 0 \) \((K \approx k)\) is a good one for high energy electrons. We can also neglect the term \( U'_0 \). We investigate these approximations in our model calculations in the next section.

Further linear equations involving the \( \lambda_i \)'s can be obtained using the fact that

\[
A_{k,l} = U_{g_k - g_l} = U_{g_m - g_n} = A_{m,n} \tag{5.27}
\]

whenever \( g_k - g_l = g_m - g_n \), for \( k \neq l \) and \( m \neq n \). Using Eq. 5.13 these
symmetries lead to the following set of homogeneous linear equations:

$$
\sum_i (C_{ki}[C^{-1}]_{il} - C_{mi}[C^{-1}]_{in})\lambda^i = 0 \quad \text{whenever} \quad g_k - g_l = g_m - g_n, \\
\quad \text{for} \quad k \neq l \quad \text{and} \quad m \neq n. \quad (5.28)
$$

Here $g_j$ refers to the reciprocal lattice vector defining the $j$th row of $C$ given in Eq. 5.12. A subset of these equations which are always present in the $A$ matrix are the symmetries expressed in Eq. 5.10. Using these we can express Eq. 5.28 as follows:

$$
\sum_i (C_{ki}[C^{-1}]_{il} - C_{N+1-l,i}[C^{-1}]_{i,N+1-k})\lambda^i = 0, \quad \text{with} \quad k \neq l \quad \text{if} \quad k_t \neq 0 \quad \text{and} \quad k + l \leq N. \quad (5.29)
$$

The remaining constraints of the type given by Eq. 5.28 and independent of those given by Eq. 5.29 are found by confining ourselves to elements on and above the anti-diagonal of $A$ to give

$$
\sum_i (C_{ki}[C^{-1}]_{il} - C_{mi}[C^{-1}]_{in})\lambda^i = 0, \quad \text{with} \quad g_k - g_l = g_m - g_n, \\
\quad k + l \leq N + 1 \quad \text{and} \quad m + n \leq N + 1. \quad (5.30)
$$

Fig. 5.3 shows the indexing of the $A$ matrix for the [110] zone-axis in a face-centred cubic system (such as GaAs) in a seven beam approximation. The central column shows the seven reciprocal lattice vectors. The symmetries across the anti-diagonal evident in Eq. 5.9 and leading to Eq. 5.29 are explicitly seen in Fig. 5.3 for this case. For example one of these symmetries (there are 18 in total) is indicated by the two circled indices that each have the value (113). There are six symmetries leading to constraints of the type given by Eq. 5.30, as indicated in Fig. 5.3. We investigate the use of these relations in uniquely retrieving the eigenvalues $2K\lambda^i$ of $A$ in the next section. For a zone axis case, the symmetries given by Eq. 5.30 are obtained by inspection of the indexing of the $A$ matrix for the particular space group and zone axis considered. However, for a systematic row orientation, all the reciprocal lattice vectors are multiples of one vector, and
we observe that the $A$ matrix becomes a band structured matrix of the form

$$
A = \begin{pmatrix}
\vdots & \vdots & \vdots & \vdots & \vdots & \\
\vdots & - (k_t + 2g)^2 + iU_0' & W_g & W_{2g} & W_{3g} & W_{4g} & \vdots \\
\vdots & -(k_t + g)^2 + iU_0' & W_g & W_{2g} & W_{3g} & W_{4g} & \vdots \\
\vdots & W_{-g} & -(k_t + g)^2 + iU_0' & W_g & W_{2g} & W_{3g} & \vdots \\
\vdots & W_{-2g} & W_{-g} & -(k_t + g)^2 + iU_0' & W_g & W_{2g} & \vdots \\
\vdots & W_{-3g} & W_{-2g} & W_{-g} & -(k_t + g)^2 + iU_0' & W_g & \vdots \\
\vdots & W_{-4g} & W_{-3g} & W_{-2g} & W_{-g} & -(k_t + 2g)^2 + iU_0' & \vdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & 
\end{pmatrix}
$$

(5.31)

This means the symmetries given by Eq. 5.30 are generally evident, due to the band structure of $A$ which occurs irrespective of the space group or systematic row. The problem of structure retrieval in the systematic row case is further discussed in [135].

If $k_t = 0$ then Eq. 5.29 also holds for $k = l$. Thus we have a matrix for which the given symmetry is true for all elements on or above the anti-diagonal. The coefficients of $\lambda^i$ in Eq. 5.29 are then all zero as is now shown.

From Eq. 5.8 we can express the eigenvalues of $A$ as

$$
2K\lambda^i = \sum_k \sum_l [C^{-1}]_{ik}A_{kl}C_{li}.
$$

(5.32)

Assume that $k_t = 0$. Then $A_{k,l} = A_{N+1-k,N+1-l}$ for all $k$ and $l$. It then follows
that $\mathcal{A}_{k,l} = A_{N+1-k, N+1-l}$, where $t$ denotes an element in the transpose. Using the fact that $[2K\lambda^i]_D$ is a diagonal matrix, and therefore is its own transpose, we can write

$$2K\lambda^i = \sum_k \sum_l [C]_{ik}^t A_{kl}^t [C^{-1}]_{li}^t = \sum_k \sum_l [C]_{ki} A_{N+1-k, N+1-l} [C^{-1}]_{il} .$$

(5.33)

Making the change of variables $m = N + 1 - k$ and $n = N + 1 - l$ we obtain

$$2K\lambda^i = \sum_m \sum_n C_{N+1-m, i} A_{mn} [C^{-1}]_{i, N+1-n} .$$

(5.34)

From Eqs. 5.32 and 5.34 it follows that

$$\sum_k \sum_l (C_{ki}[C^{-1}]_{il} - C_{N+1-l,i} [C^{-1}]_{i, N+1-k}) A_{lk} = 0 .$$

(5.35)

This can only be true for any choice of the matrix elements $A_{lk}$ if the coefficients in the brackets are all zero.

When $k_t = 0$ the equations given by Eq. 5.26 are no longer linearly independent and yield at most $(N + 1)/2$ linearly independent (complex) equations (one for each $|g_i|^2$), while we must find $N$ complex parameters $\lambda^i$. This is similar to the case of no absorption [20]. We then use Eqs. 5.26 and 5.30 to uniquely obtain the $\lambda^i$ and hence $\mathcal{A}$. In this situation, $N$ needs to be large enough for Eqs. 5.26 and 5.30 to have a coefficient matrix of rank $N$. For example, $N = 3$ will not work because there are not enough elements to produce three separate symmetries which generate three independent equations. This is not an important restriction in practice, since it is desirable to use a high value of $N$ anyway, to obtain an accurate representation of the projected potential. Furthermore, the rank of the set of linear equations used to obtain the $\lambda^i$, and thus the uniqueness of the solution, is easily checked and there is no danger of unwittingly obtaining a spurious solution to the inversion problem.

Having solved the inversion step $\mathcal{S} \rightarrow \mathcal{A}$ we can calculate the Fourier coefficients $U_g$ and $U'_g$ in Eqs. 5.4 and 5.5 from [23]

$$U_g = \frac{W_g + W^*_g}{2} \quad \text{and} \quad U'_g = \frac{W_g - W^*_g}{2i} .$$

(5.36)
The potential associated with elastic scattering and that due to absorption can thus each be recovered separately.

5.5 Model solutions of the inversion problem

We consider the [110] zone axis in GaAs as an example and calculate the $S$ matrix, which is then used as input in testing our inversion procedure. The calculations are for an incident electron energy of 400 keV and a crystalline slab of thickness 1000 Å. However the method works for arbitrary incident energy and sample thickness. The thickness does not need to be known explicitly. We continue with $N = 7$ for the detailed example, chosen for simplicity of illustration, but emphasise that the method works just as well for larger values of $N$. The Fourier coefficients for the elastic potential are calculated using the electron scattering factors provided in [91] and thermal effects are incorporated via a Debye-Waller factor. A temperature factor $B = 0.6$ Å$^2$ is used for both Ga and As [19, 131]. Absorption due to TDS is included in the calculations. The real part of the projected potential $V(r)$ is shown in Fig. 5.4(a) and the potential $V'(r)$ representing absorption due to TDS is shown in Fig. 5.4(b). These representations of the potential (each with 19 Fourier coefficients) are of course not converged, as can be seen by comparison with Fig. 5.3 in [136], where many more Fourier coefficients were used to plot the potentials. It is worth noting that, while $V(r)$ is intrinsic to the crystal, depending only on the temperature, $V'(r)$ (for TDS) depends on the energy of the incident electron beam.

<table>
<thead>
<tr>
<th>Case No.</th>
<th>$k_t$</th>
<th>Eq. 5.26</th>
<th>Eq. 5.26 &amp; Eq. 5.29</th>
<th>Eq. 5.26 &amp; Eq. 5.30</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(000)</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>2</td>
<td>(001)</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>3</td>
<td>(110)</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>4</td>
<td>(111)</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>5</td>
<td>(113)</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Table 5.1: Results of inversion for different orientations and sets of equations.
The $S$-matrix has been calculated and then diagonalised to find its eigenvectors for each of the principal orientations shown in Table 5.1. The outcomes of retrieving the set $\lambda^i$ correctly (uniquely) using Eq. 5.26 alone, Eqs. 5.26 and 5.29 and then Eqs. 5.26 and 5.30 to determine the parameters $\lambda^i$ are noted.

We solve the set of linear equations to obtain the $\lambda^i$ using SVD [77]. For the first case, when $k_t = 0$, there is insufficient information to determine the $\lambda^i$ in the symmetric orientation using Eqs. 5.26 and 5.29. This is as expected, and we must therefore use Eqs. 5.26 and 5.30.

For the [110] zone in a face centred cubic structure, the equations given by Eq. 5.30 arise from symmetries 1 to 6 in Fig. 5.3. For $k_t = 0$, Eq. 5.26 yields only three linearly independent equations for the $\lambda^i$ out of $(N+1)/2 = 4$ possible linearly independent equations. This is because the equations relating to the $\langle 110 \rangle$ reflections are the same. Symmetries 1 to 4 in Fig. 5.3 yield the remaining four equations which are required to determine the $\lambda^i$ uniquely. It should be noted that symmetries 5 and 6 yield equivalent equations to 3 and 4 respectively. It is also worth pointing out that while the equations obtained from symmetries 1 and 2 are complex conjugates, complex conjugation is not a linear operation and thus the equations are linearly independent.

For the second case, when $k_t = (00\bar{T})$, (i.e. the reflection (002) is in the exact Bragg orientation) there is still sufficient symmetry in the system to render the equations given by Eq. 5.26 linearly dependent. However addition of the symmetry constraints in Eq. 5.29 determines $\lambda^i$ uniquely. Tilting along the $(\bar{1}10)$ direction (the third case) gives similar results. Tilting so that the reflection $(1\bar{1}0)$ is in the exact Bragg orientation (the fourth case) removes symmetries in the system so that now Eq. 5.26 alone is already sufficient to determine the $\lambda^i$ uniquely. The arbitrary orientation given in case 5 also breaks symmetries in such a way that the orientation constraints are sufficient to solve for the $\lambda^i$.

We assume that the mean potentials are known in these tests and in all cases the potentials shown in Figs. 5.4(a) and 5.4(b) are retrieved to high precision (at least six figure accuracy), as shown in Fig. 5.4 (c) and (d). The results are not significantly changed if we assume that $K \approx k$ since, at 400 keV, $K = 60.8326 \text{ Å}^{-1}$ and $k = 60.8311 \text{ Å}^{-1}$. For the $k_t = 0$ case, assuming both $V_0 = 0$ and $V'_0 = 0$, we obtain by inversion the potentials shown in Figs. 5.4(e) and (f). The real and absorptive potentials obtained from the exact and approximate inversions differ exactly by the mean potentials $V_0 = 14.68 \text{ eV}$ and $V'_0 = 0.3164 \text{ eV}$, and are consistent with the level of agreement between $K$ and $k$. It is not surprising that
Figure 5.4: Real and absorptive parts of the optical potential for the [110] zone axis in GaAs.
the approximate inversion is accurate up to these shifts, since $V'(r)$ is an order of magnitude smaller than $V(r)$ and the absorptive scattering can be treated perturbatively [18,58].

<table>
<thead>
<tr>
<th>Crystal</th>
<th>sys. row/zone axis</th>
<th>$N$</th>
<th>Crystal</th>
<th>sys. row/zone axis</th>
<th>$N$</th>
</tr>
</thead>
<tbody>
<tr>
<td>GaAs</td>
<td>[110] zone axis</td>
<td>7</td>
<td>Si</td>
<td>[110] zone axis</td>
<td>9</td>
</tr>
<tr>
<td>GaAs</td>
<td>[100] zone axis</td>
<td>13</td>
<td>Si</td>
<td>[100] zone axis</td>
<td>13</td>
</tr>
<tr>
<td>GaAs</td>
<td>[111] zone axis</td>
<td>19</td>
<td>Si</td>
<td>[111] zone axis</td>
<td>19</td>
</tr>
<tr>
<td>GaAs</td>
<td>[332] zone axis</td>
<td>7</td>
<td>Si</td>
<td>[332] zone axis</td>
<td>7</td>
</tr>
<tr>
<td>GaAs</td>
<td>{100} sys. row</td>
<td>5</td>
<td>Si</td>
<td>{100} sys. row</td>
<td>5</td>
</tr>
<tr>
<td>GaAs</td>
<td>{111} sys. row</td>
<td>5</td>
<td>Si</td>
<td>{111} sys. row</td>
<td>5</td>
</tr>
<tr>
<td>GaAs</td>
<td>[110] zone axis</td>
<td>129</td>
<td>Bi$_2$Sr$_2$CaCu$_2$O$_8$</td>
<td>[001] zone axis</td>
<td>69</td>
</tr>
</tbody>
</table>

Table 5.2: Test cases for several crystals for a number of exact symmetric (systematic row) or zone axis orientations which provide a unique solution to the dynamical inversion problem using Eqs. 5.26 and 5.30.

Next we test the use of Eqs. 5.26 and 5.30 to solve the dynamical inversion problem for $k_t = 0$ for different crystals, using both systematic rows and zone axis orientations. Provided that the number of beams is large enough (typically of the order of ten), then a unique reconstruction of $\mathcal{A}$ from $\mathcal{S}$ is always achieved. We note that, in contrast, when using Eqs. 5.26 and 5.29 for $k_t \neq (000)$ unique solutions to the inverse scattering problem can be found for any odd $N \geq 3$. These results are shown in Table 5.2 and are independent of incident energy and crystal temperature and thickness.

For all the results shown, with the exception of those in the last line, the number of beams indicated is the minimum number for which the inversion procedure works, and not excluding any beams with the same magnitude (scattering angle) as those already in the set. For example, for the [111] zone axis in GaAs or Si, 15 beams is sufficient but this excludes some reciprocal lattice vectors of the same magnitude as some already included. All systematic row cases require only five beams. Notice that for the [110] zone axis case in Si, which is centrosymmetric,
nine beams are required for a unique solution, unlike GaAs (noncentrosymmetric) which required only seven. The centrosymmetry of silicon means that the origin of the unit cell can be chosen so that the structure factors are real. The equations which are obtained from pairs of symmetries, like symmetries 1 and 2 in Fig. 5.3, which are related by complex conjugation then become identical.

In practice it is likely that a unique solution to the inversion problem for the symmetric orientation using Eqs. 5.26 and 5.30 will always be possible for a values of $N$ well below those giving a reasonable representation of the dynamical scattering. The number of symmetries leading to equations of the type in Eqs. 5.30, as a function of $N$ for the [110] zone axis in a face-centred cubic structure, is shown in Fig. 5.5. A rapid increase as a function of $N$ is evident. The results in the last line of Table 5.2, for GaAs with 129 beams and the superconductor $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_8$ with 69 beams, show that the inversion method works for large numbers of beams and for more complex structures. Clearly this method is successful in simulation.

### 5.6 Inversion for a crystal with a stacking fault.

Having solved the inversion problem in dynamical electron diffraction for a periodic crystal, the next question is whether it is possible to generalise this method in order to apply it to a situation where the scattering object is not periodic. In this section, we attempt to apply the inversion method to a very simple situation where two crystals are sandwiched together in the $z$ direction, as shown in
Fig. 5.6. This is a simple type of aperiodicity that may be introduced while still maintaining as much symmetry as possible in the problem. For simplicity it is assumed that the crystals have the same sized unit cell. The case examined is that of two crystals of the same material which are shifted relative to one another, as shown in Fig. 5.7. This is a common defect found naturally in crystals, and is known as a stacking fault. The calculations for this and subsequent figures in this section are done assuming an electron beam energy of 100keV incident onto silicon.

The method developed for retrieving the potential of two or more crystals joined together is an extension of that previously mentioned. Passing the input wave through multiple crystals results in an overall $S$ matrix describing the effect of all the crystals on the wave. This is the product of the individual $S$ matrices for each of the crystals, since the exit surface wavefunction for the first crystal is the input wave to the second crystal.

Finding an effective $\mathcal{A}$ matrix containing Fourier coefficients for the total combined potential of the crystals is more difficult. The mathematical description
of this situation is shown below.

\[
S = S_2 S_1 = \exp \left( \frac{i t_2}{2K} A_2 \right) \exp \left( \frac{i t_1}{2K} A_1 \right) = \exp \left( \frac{i (t_1 + t_2)}{2K} A \right) \quad (5.37)
\]

Here \( S \) and \( A \) are the matrices describing the overall interaction with the original incoming wave, and \( S_1, S_2, A_1, A_2 \) are the matrices describing the effect of crystals 1 and 2 individually.

The problem exists because the product of two exponentials of matrices cannot in general be defined in a closed form. This means that it is impossible to calculate the exact structure of the overall \( A \) and \( S \) matrices, and therefore that it is impossible to know whether they possess the same symmetries as the individual \( A_1, A_2, S_1 \) and \( S_2 \) matrices (i.e. those symmetries given by Eqs. 5.10 and 5.19). More insight may be gained by calculating the form of the overall \( S \) matrix, using the standard exponential expansion

\[
S = \left( I + \frac{it_2}{2K} A_2 + \ldots \right) \left( I + \frac{it_1}{2K} A_1 + \ldots \right) = \left( I + \frac{it_2}{2K} A_2 + \frac{it_2}{2K} A_1 - \frac{t_2 t_1}{4K^2} A_2 A_1 + \ldots \right) \quad (5.39)
\]

Unless the two matrices \( A_1 \) and \( A_2 \) are identical, they do not in general commute. Thus the symmetry rules for the \( S \) matrix are not obeyed by the product \( A_2 A_1 \). Hence the overall \( S \) matrix does not obey the usual rules for the \( S \) matrix of a single crystal. It follows that the overall \( A \) matrix does not obey the symmetry rules on which the potential retrieval method is based, and thus the exact potential is not retrieved.

However the method is correct to first order, since to first order, the expression for \( S \) in Eq. 5.39 does obey the expected rules for an \( S \) matrix. Hence in the first order, or single scattering approximation (i.e. for sufficiently thin crystals), this method of retrieving the projected potential is useful.

Potentials are calculated for the case of a silicon crystal, looking down the [110] zone axis. The crystal has a stacking fault, and the problem is treated as that of having two identical crystals sandwiched together with a translation difference between their unit cells. The total thickness of the two crystals is 40 Å and different ratios of thickness between the first crystal and the second crystal are considered. Images, as found from the exit surface wavefunction are also calculated. The results are shown in Fig. 5.8.
In these results it is seen that the potentials retrieved, shown in Fig. 5.8 (a), (c) and (e), show a clear superposition of the unit cells of both crystals. The faint interference effects observed are a result of the failure of the single scattering approximation, even for crystals of these small thicknesses. However the atom locations for each crystal can clearly be seen. As the ratio of thickness between the crystals changes, the relative brightness of the different unit cells varies. In each case, the brighter atoms correspond to the thicker crystal. This suggests that a successful inversion can be used not only to deduce the crystal structures, but also to find the relative thickness of the two crystals.

In contrast, the images calculated, shown in Fig. 5.8 (b), (d) and (f) are much more greatly affected by the first crystal encountered by the incoming plane wave. Even when the second crystal is as thick as the first, as in Fig. 5.8 (d), the image is dominated by the structure of the first crystal. This means that the images can be misleading, as seen in Fig. 5.8 (b), where the existence of the second crystal is hard to observe from the image.

Both the images and the retrieved potential become increasingly misleading for thicker crystals. This is due to the increasingly great failure of the single
scattering approximation.

It is clear from the results that where the electron beam is passed through more than one crystal, the physical situation is complicated to the point that images taken of the exit surface wavefunction can be very misleading. Calculating the projected potential of the crystal, by use of the inversion techniques developed, is more useful and leads to a much more accurate understanding of the crystal structure. Information such as the projected potentials of each separate crystal involved, and the relative thickness of each crystal can be found by examining the calculated potential. However mathematical limitations of this inversion technique limit its usefulness to situations where the single scattering approximation is valid. It is hoped that in the future, techniques will be refined in order to calculate accurate potentials for much thicker combinations of crystals.

5.7 Summary

In this chapter, the forward and inversion problems in dynamical electron diffraction are examined. The relative complexity of the dynamical case is explained and a solution presented which correctly finds the projected potential of a periodic object in this case.

The main steps in the inversion of the dynamical scattering, taking absorptive scattering into account, can be described as follows. We assume that all the complex elements of the scattering matrix $S$, for a given energy and thickness, are determined (up to an arbitrary overall phase) using a through-tilt series of measurements at the principal and well defined secondary orientations of the incident beam. $S$ is then diagonalised to obtain its eigenvectors, giving the eigenvalues of $A$, since the eigenvectors of $S$ and $A$ are the same. The eigenvalues of $A$ are then obtained uniquely using Eq. 5.26 (following from knowledge about the diagonal elements of $A$, which are determined by the principal orientation, $U_0$ and $U'_0$) and Eq. 5.29 (derived from general symmetries of $A$ across its anti-diagonal) or Eq. 5.30. For an exact zone axis orientation of the incident beam we must use Eq. 5.30. In all cases there is the requirement that $N$ needs to be large enough. This constraint is likely to be always satisfied when modelling the dynamical scattering with a number of beams large enough to produce a converged representation of the crystal potential. The matrix $A$ is then constructed uniquely from its eigenvectors and eigenvalues using the spectral representation of $A$ given by Eq. 5.13. Both the projected potentials associated with elastic scattering and
that associated with absorption are then obtained. The inversion of the dynamical scattering process is finished by the solution of sets of linear equations. This yields a unique solution and is numerically stable.

The chapter is completed with a brief examination of the problem of finding the projected potential of a crystal with a stacking fault. Although a solution in the multiple scattering case is not found, due to the loss of periodicity in the $z$ plane, it is instructive to examine this case anyway and to hope for future solutions to this and more complex problems.

The main challenge in the practical implementation of the methods discussed here is likely to be the accurate determination of the entire scattering matrix $S$. The amplitude and phase of all elements needs to be known. This requires the determination of the exit surface wave function at the principal and secondary orientations of the incident beam. The large number of experimental measurements needed to use this method are a practical difficulty to be overcome. It is to be hoped that modern microscopes will soon make it easier to take series of measurements such as those required by this method, allowing accurate experimental determination of the scattering potential via the inversion techniques described in this chapter.
Chapter 6

Model problem

In the preceding chapters we have discussed methods of phase retrieval for wavefunctions containing vortices, aberration determination and correction, and solving the inverse problem in dynamical electron diffraction. As a final demonstration, drawing together all the aspects of this thesis, a model calculation is presented which demonstrates the use of all these techniques. The problem is to first characterise the aberrations in a transmission electron microscope and then to retrieve the structure of the projected potential of a crystal by solving the inversion problem.

6.1 Aberration control for model system

We simulate the optical situation found in transmission electron microscopy (TEM) [111, 137], as shown schematically in Fig. 6.1, with a specimen included in the diagram. The electron beam produced by the emission gun is processed before being projected onto the specimen. These pre-specimen optical processors include condenser lenses and the condenser aperture. The post-specimen processors include the objective aperture, objective lens, intermediate lenses and projector. Finally the output beam is detected by a CCD camera.

It is assumed that the pre-specimen optics are able to accurately produce an incident plane wave. This is because the solution to the inverse problem in dynamical electron diffraction assumes a plane wave to be incident onto the specimen. It is further assumed that the pre-specimen optics are able to produce a point source incident onto the post-specimen optics. This could be achieved by switching off the pre-specimen lenses to allow the external point source through unaltered, as illustrated in Fig. 6.2. This facility is required to accurately determine aberra-
tions present in the post-specimen optics, which can be both characterised and removed computationally using the algorithms in Chapter 4.

In this model calculation it is assumed that the only aberrations present in the microscope are coherent aberrations of the post-specimen optics. A system with spherical aberration $C_s = 0.05$ mm and defocus $\Delta f = 120$ Å is simulated. These are values that could be produced by a system with some existing aberration corrections present in the hardware, and are similar to those used by Haider et al. [39, 94, 95]. The aperture is assumed to be $1$ Å$^{-1}$, and the electron beam wavelength $0.025$ Å, which corresponds to an incident energy of 200 keV. The CCD camera is assumed to have a $1024 \times 1024$ pixel imaging capacity.

These parameters produce the phase factor $W(q)$ in momentum space shown in Fig. 6.3(a). The through focal series generated by these aberrations in the
Figure 6.3: Input phase factor $W(q)$ in the post-specimen optics and phase factor retrieved by aberration characterisation.

Image space is shown in Fig. 6.4, with defocus values measured relative to the central image. A noise level of 0.1% on the brightest pixel in the central image in Fig. 6.4 is assumed. Since the central image contains a small, bright point which has much higher intensity than the surrounding region, the noise is considerably higher on other pixels. The phase of the central image shows a similar level of complexity (in particular, vortex structure) as the example given in Fig 4.4, requiring the use of the TFS method for the phase retrieval process.

After phase retrieval and unwrapping, the phase map shown in Fig. 6.3(b) is retrieved. This is decomposed to find the aberration values $C_s = 0.05005$ (input as 0.05) and $\Delta f = 120.2$ (input as 120). In the following model inversion calculation, aberrations corresponding to the original values are added to the images. We correct for aberrations corresponding to the retrieved values.
6.2 Model inversion calculations

The model inversion calculation is done for a GaAs crystal of thickness 300 Å in the [110] zone axis orientation. Although the thickness of the crystal does not need to be known to perform the inversion, it is needed for the calculations for the forward scattering simulation. A 27 beam Bloch wave approximation is used, since this provides an acceptably good representation of the dynamical scattering processes occurring. Absorption due to TDS is included as in previous calculations (see Chapters 4 and 5).

![Image](image.png)

Figure 6.5: Through focal series for model problem at (000) tilt with no noise included.

The input data takes the form of a through focal series of five images at each of the 27 orientations of the incident beam. Each image is aberrated using the original $C_s = 0.05$ mm and defocused to the Scherzer defocus (129.5 Å), which can be experimentally identified by examining the contrast of the images [138]. An example of an aberrated through focal series, for the (000) beam, is shown in Fig. 6.5. Only one unit cell of the periodic wavefunction is shown in each image, since this allows the calculation to proceed with a greater amount of detail included given the image size of 128 × 128 pixels.

There are vortices in the wavefunction at each plane in the series which, as discussed in Chapter 2, is typical of the type of wavefunction encountered in dynamical electron diffraction. This necessitates the use of the TFS method of phase retrieval, as discussed in Chapter 3. Comparing the central image to the original projected potential (shown in Fig. 6.9) shows clearly that multiple scattering is occurring, since the image bears little resemblance to the potential.

The central images in the through focal series for 14 of the 27 tilts are shown in Fig. 6.6. To perform the model inversion, the phase of the wavefunction at each
central image is found, again using the TFS method. For the examples in Fig. 6.6 the phase is retrieved to an accuracy of $10^{-15}$ or better for every orientation.

Once the entire complex wavefunction is known at each orientation, the aberrations are corrected using the characterised spherical aberration value of $C_s = 0.05005$ mm and the Scherzer defocus value of $\Delta f = 129.5$ Å. These aberration corrected wavefunctions are then Fourier transformed to give the corresponding elements of the $S$ matrix. The inversion problem is then solved, using the method of solution given in Chapter 5, to find the $A$ matrix, which contains the Fourier coefficients, $W_g$, of the potential. These are summed together, as in Eq. 5.2, to give the projected potential of the crystal.

![Figure 6.6: Some of the input images for model problem at different tilts, no noise.](image)

The phase retrieval and inversion are repeated with noise levels of 5% and 10% noise on the brightest pixels in the central images. Example of the resulting noisy images are shown in Figs. 6.7 and 6.8. As seen in the images, the quality of the input data is degraded by the addition of noise, especially at the 10% level. This results in less accurate phase retrieval. For the 5% case the phase retrieval SSE is around $10^{-2}$ and for the 10% case it is around $10^{-1}$. These uncertainties are much greater than those in the situation where no noise is added.
Figure 6.7: Input images at different tilts, 5% noise level.

Figure 6.8: Input images at different tilts, 10% noise level.
Figure 6.9: Original and recovered elastic and absorptive potentials for model problem.

The original and recovered projected potentials of the GaAs crystal are shown in greyscale form in Fig. 6.9. The values for the elastic potentials range from $-20$ eV (black) to 140 eV (white). The absorptive potentials range from $-3$ eV to 12 eV. Thus the elastic potential is the most significant recovered result. When no noise is present, the potential is recovered to high accuracy, being indistinguishable by eye from the original in both the elastic and absorptive parts, and having a SSE of $2.17 \times 10^{-5}$ for the elastic potential and $1.25 \times 10^{-2}$ for the absorptive potential. The structure of two different atoms in the [110] orientation of a face centred cubic crystal is clearly seen in the elastic potential with the slightly brighter arsenic atom seen to the right of the gallium in each pair. The absorptive potential shows similar structure although the range of values is smaller because the absorptive scattering is a smaller effect than the elastic scattering.

In both the 5% and 10% noise cases, the elastic potential is recovered well, with SSE values of $2.61 \times 10^{-4}$ and $2.67 \times 10^{-3}$ respectively. The increased sizes of the SSE values are of course due to the increased noise levels. It is more difficult to identify the brighter atom in the 10% noise case, since both atoms appear to be of similar brightness.
In the absorptive potential, the SSE values for the 5% and 10% noise cases are $2.85 \times 10^{-1}$ and $4.10 \times 10^{-1}$. Here, the noise in the input data results in the appearance of structure which is not present in the original potential. This is particularly clear in the 10% case, where the locations of the atoms in the absorptive potential are more obscured. However, on the whole, the inversion to find the projected potential appears very robust in the presence of noisy data, even given the slightly inaccurate correction of the wavefunction aberrations produced by the simulated system.

6.3 Discussion of experimental practicalities

The model problem examined in this chapter is successfully solved. However there are practical difficulties to be overcome before these simulations can become experimental reality. The most obvious of these is the large number of measurements required to perform the inversion calculation. For the simulation performed in this chapter, 135 measurements of the image at different tilts and defoci are required, assuming the aberration characterisation has already been performed.

If it is important to reduce the number of measurements, one strategy would be to perform phase retrievals with through focal series of only three images. This will make the phase retrieval algorithm slower, but will still produce an accurate result. As an example, phase retrieval for the through focal series in Fig. 6.5 achieves an SSE of $7.8 \times 10^{-17}$ in 11 iterations with all five images. If the central three images only are used, the SSE achieved is $5.4 \times 10^{-16}$ after 40 iterations. Depending on the number of pixels used, the difference in running time varies from a few seconds (with $128 \times 128$ pixels as used here) to several minutes (with $1024 \times 1024$ or larger arrays).

Another possible strategy would be to take measurements at fewer tilts of the incident beam. This would result in the recovery of fewer Fourier coefficients of the potential and a less accurate result, as discussed in [135]. However that may be all that is required in some situations, rendering the experimental problem more practical. The greatest difficulty will arise when a complex structure is examined, since a higher number of Fourier coefficients are needed for accurate structure retrieval.
6.4 Summary

This chapter demonstrates the solution of a model problem that illustrates all the major points of this thesis. The wavefunctions created for the calculations display the vortices and other phase behaviour discussed in Chapter 2. The TFS method of phase retrieval examined in Chapter 3 is used throughout as an essential tool for the solution of the two inverse problems simulated. Firstly we use the material in Chapter 4 to characterise the aberrations in the post-specimen optics in the microscope. Secondly we use the solution to the inverse problem in dynamical electron diffraction given in Chapter 5 to find the projected potential of a GaAs crystal.

The simulation incorporates features of an actual experimental situation such as noise, absorption and the presence of aberrations in the imaging equipment. The aberrations present are characterised to an accuracy of around 1% in this case, and the inversion process handles the presence of the residual aberrations without difficulty, still accurately retrieving the projected potential. The inversion is also stable in the presence of noise and works extremely well with a noise level of at least 10%, retrieving an accurate elastic potential. The recovery of the absorptive potential is less effective with noisy data, but a recognisable structure is still present. Altogether, the solution of this model problem is a success and it is to be hoped that actual experiments of this type can be used to perform structure retrieval in the future.
Chapter 7

Conclusion

This thesis examines a number of related problems in dynamical electron diffraction and phase retrieval. Although much of the work is applicable to many other optical situations, e.g. when the wave field is produced by electromagnetic radiation or other particles, we concentrate in particular on situations which simulate those found in modern experimental electron microscopy. The aim is to find better methods of processing electron microscope data, which allow a greater range of information to be obtained with present day hardware technology.

The study begins in Chapter 2 by examining some simple properties of phase and phase singularities, with particular emphasis on vortices and vortex trajectories. A number of mathematical rules governing vortex behaviour are given and discussed and simulated results showing the evolution of vortex trajectories in space are shown. Such detailed simulations of the behaviour of vortex trajectories have not been done previously, and prove to be extremely interesting, as the vortex cores turn out to possess rich and varying structure. The results are consistent with theory and provide evidence of the ubiquity of vortices in wave-functions such as those resulting from dynamical electron diffraction. It is clear from these results that any processing or handling of the phase of such wavefunctions, especially phase retrieval, must be able to successfully handle vortices and vortex trajectories.

With this in mind, Chapter 3 contains a comparative analysis of several different methods of non-interferometric phase retrieval. The methods examined are either based on the Transport of Intensity Equation or are iterative methods similar in style to the Gerchberg-Saxton algorithm. These methods are compared using simulated model problems. It is found that in situations where vortices or similar phase singularities are present, the iterative methods are much more
successful than the TIE methods. This is because the iterative methods make no assumptions about the wavefunction whereas the TIE methods must assume that the intensity of the wavefunction is never zero and therefore that vortices are not present. Clearly for the purposes of this thesis, the iterative methods are the appropriate choice. In particular, the method based on a through focal series of images is found to be most applicable, because of its ability to handle non-periodic data as well as vortices. This method is examined for a non-periodic case with emphasis on its ability to cope with noise and its general stability. It is found that the algorithm is robust in the presence of noise and that although in theory it may not immediately return a solution, it is a simple matter to identify stagnation and repeat the iterative process until a solution is found.

Using the TFS method of phase retrieval as a tool, Chapter 4 presents an algorithm for computational determination of the coherent (and in principle incoherent) aberrations present in an imaging system such as those used in HRTEM. This method allows accurate determination of all the aberrations present in the system. Once the aberrations are characterised it is easy to remove them, if this is desired, using a previously found computational method. This method of computational aberration control is potentially extremely useful because it increases the amount of accurate information available to the experimental researcher without requiring lengthy adjustments or complex hardware.

Chapter 5 presents a solution to the inversion problem for periodic objects in dynamical electron diffraction. This problem was unsolved for many years and its solution is a significant step forward. The method implicitly assumes use of the previous step of aberration characterisation and removal, and uses the TFS method of phase retrieval as an essential tool. The issue of solving the inversion problem for a crystal with a stacking fault is also investigated.

The thesis concludes with a final model problem, presented in Chapter 6. This demonstrates all of the previous material, from characterising and removing the coherent aberrations present in a TEM system, to solving the inversion problem to find the projected potential of a sample crystal, using phase retrieval as a part of each process.

In summary, this thesis successfully finds improved methods of processing optical information in the areas of phase retrieval and inversion of electron microscope data. There is much work still to be done in these fields. In phase retrieval the aim will be to find algorithms which are computationally more reliable, faster and thus more useful. Real-time solution of the phase problem is a
goal which is yet to be reached in situations where phase singularities are known to exist. In part, this will be increasingly well achieved with the rapid improvement in computer resources at our disposal. However improved algorithms will make such progress easier.

In the area of inversion of dynamical scattering processes, such as that found in electron diffraction, future work will include finding ways to solve the inversion problem for non-periodic objects. Such methods will also be extended to finding three dimensional potentials, eventually allowing us to image any object at the atomic level in three dimensions. The usefulness of such technology in a wide variety of fields can not be underestimated. There will be applications of this work to fields of research in chemistry, medicine, biology, materials technology and innumerable other areas.
Bibliography


Appendix A

Proof that the Gerchberg-Saxton algorithm converges

In this section we prove that the sum squared error in the Gerchberg-Saxton algorithm decreases or stays the same with each iteration of the algorithm [78,89]. To start with, we recall that the sum squared error at iteration $n$ of the algorithm is defined in Eq. 3.30 as

$$\text{SSE}_n = \frac{1}{\sum |\psi_0|^2} \sum ||\psi_n| - |\psi_0||^2,$$  \hspace{1cm} (A.1)

where the sum is over all pixels in the image, $\psi_n$ is the estimated wavefunction in real space at iteration $n$, and $\psi_0$ is the actual wavefunction. $\sum |\psi_0|^2$ is a scaling factor only, and has the same value at every plane, due to energy conservation in the wavefunction. We neglect this term and write

$$\text{SSE}_n' = \sum ||\psi_n| - |\psi_0||^2.$$

(A.2)

Note that since in practice we only know the intensities of the actual wavefunction, the error is defined in terms of these only. It is a measure of the correction required to fulfil the intensity criteria in real space. When the SSE reduces to zero, the wavefunction is propagating correctly and the phase is retrieved.

Since we correct only the moduli of the wavefunction at each step, the difference between the estimated and actual moduli in Eq. A.2 has the same magnitude as the difference between the estimated and corrected wavefunction, which both, by definition, have the same phase. We define $\psi_{0,n}$ to be the corrected wavefunction at iteration $n$. It has the correct intensity and the same (probably incorrect)
phase as $\psi_n$. Note that at each pixel $|\psi_n - \psi_{0,n}| = ||\psi_n| - |\psi_0||$, because the functions on the left have the same phase. We can therefore rewrite Eq. A.2 as

$$SSE'_n = \sum |\psi_n - \psi_{0,n}|^2 .$$  \hspace{1cm} (A.3)

We make use of Parseval’s theorem [75], which states that sum of squares of the moduli of a function is equal to the sum of squares of the moduli of its Fourier transform. This is effectively a statement of energy conservation between the image and diffraction planes. Thus taking the Fourier transform of Eq. A.3 we can state that

$$SSE'_n = \sum |\mathcal{F}[\psi_n] - \mathcal{F}[\psi_{0,n}]|^2 ,$$  \hspace{1cm} (A.4)

Recalling that the corrected wavefunction in the diffraction plane is inverse Fourier transformed to create the new guessed wavefunction in the image plane in the next step of the algorithm, we have that

$$\mathcal{F}[\psi_n] = \Psi_{0,n} .$$  \hspace{1cm} (A.5)

Since the corrected wavefunction in the image plane is forward Fourier transformed to create the new guessed wavefunction in the diffraction plane in the next step of the algorithm, we also have that

$$\mathcal{F}[\psi_{0,n}] = \Psi_{n+1} .$$  \hspace{1cm} (A.6)

We therefore see that

$$SSE'_n = \sum |\Psi_{0,n} - \Psi_{n+1}|^2 .$$  \hspace{1cm} (A.7)

We now use the fact that each intensity correction step performs the minimal necessary change to the guessed wavefunction to make it satisfy the intensity criteria in that plane. The corrected wavefunction in the diffraction plane, $\Psi_{0,n+1}$, is therefore the closest function to the guessed wavefunction, $\Psi_{n+1}$, which satisfies the intensity criteria. Thus at each pixel of the image,

$$|\Psi_{n+1} - \Psi_{0,n+1}| \leq |\Psi_{0,n} - \Psi_{n+1}| .$$  \hspace{1cm} (A.8)

We can therefore see, using Eq. A.7 that

$$\sum |\Psi_{n+1} - \Psi_{0,n+1}|^2 \leq SSE'_n .$$  \hspace{1cm} (A.9)
The sum squared error in the diffraction plane step is less than that in the preceding image plane step. Using a similar argument we can return to the image plane and find that the error in the image plane for iteration $n + 1$ is now less than that in the diffraction plane. Thus we see that the SSE' in the GS algorithm decreases (or at least does not increase) at each step, and therefore the original SSE does not increase either.

This proof may be generalised to any similar iterative error reduction algorithms [89], that use two images in different planes including the TFS method for two planes and the Fienup algorithm. However the proof does not generalise to a situation where there are three or more images. This is because it relies on it being possible to go “backwards” through the algorithm, as is done to find Eq. A.5 as well as “forwards”, as done to find Eq. A.6, and reach the same image plane in both directions. This can only be done when the algorithm is turning from forward to backward propagation, or vice versa. In a through focal series of five images, as used in the TFS algorithm here, this proof holds for the error reducing as the algorithm moves from the fourth image to the fifth and back to the fourth again, and similarly between the second and first images. In the intermediate stages, moving between the middle three images, the proof does not hold.

It is clear, however, that the TFS method does work. This can be understood by noting that the error decreases at each end of the through focal series, as described. In the middle stages, the only alteration to the wavefunction as it propagates is to correct it to a more accurate version, i.e. with the right intensities. Thus the middle stages in practice do not cause the SSE to increase, since they are adding more correct information to the wavefunction at each step.
Appendix B

Derivation of free space propagator

In this section we derive the free space propagator used throughout this thesis to propagate a wavefunction in space. The Wave Equation (in the absence of a potential) can be written as [139]

\[(2ik\partial_z + \nabla_\perp^2)\psi(r_\perp, z) = 0,\]  \hspace{1cm} (B.1)

where \(\psi\) is the complex wavefunction and \(k = \frac{2\pi}{\lambda}\). A solution to Eq. B.1 is

\[\psi(r_\perp, z) = e^{\frac{ik}{2\pi}\nabla_\perp^2} \psi(r_\perp, z = 0).\]  \hspace{1cm} (B.2)

To check this,

\[2ik\partial_z \psi(r_\perp, z) = 2ik\partial_z e^{\frac{ik}{2\pi}\nabla_\perp^2} \psi(r_\perp, z = 0) = 2ik \left( \frac{i\nabla_\perp^2}{2k} \right) e^{\frac{ik}{2\pi}\nabla_\perp^2} \psi(r_\perp, z = 0) = -\nabla_\perp^2 e^{\frac{ik}{2\pi}\nabla_\perp^2} \psi(r_\perp, z = 0) = -\nabla_\perp^2 \psi(r_\perp, z).\]  \hspace{1cm} (B.3)

Therefore \(\psi(r_\perp, z) = e^{\frac{ik}{2\pi}\nabla_\perp^2} \psi(r_\perp, z = 0)\) is a solution to the wave equation in free space. This is the free space propagator in the paraxial approximation.

Let us rewrite \(\psi(r_\perp)\) as

\[\psi(r_\perp) = \mathcal{F}^{-1}\mathcal{F} \left[ e^{\frac{ik}{2\pi}\nabla_\perp^2} \psi(r_\perp, z = 0) \right].\]  \hspace{1cm} (B.4)
Where \( \mathcal{F} \) represents the Fourier transform, defined as follows in two dimensions.

\[
\mathcal{F}[\psi(\mathbf{r}_\perp)] = \Psi(k_\perp) = \int_{r_x} \int_{r_y} \psi(\mathbf{r}_\perp)e^{-i\mathbf{k}_\perp \cdot \mathbf{r}_\perp} \, d\mathbf{r}_\perp 
\]  
(B.5)

\[
\mathcal{F}^{-1}[\Psi(k_\perp)] = \psi(\mathbf{r}_\perp) = \frac{1}{4\pi^2} \int_{k_x} \int_{k_y} \Psi(k_\perp)e^{i\mathbf{k}_\perp \cdot \mathbf{r}_\perp} \, dk_\perp, 
\]  
(B.6)

where \( k_\perp = \frac{4\pi}{r_\perp} \) and the integrals range from \(-\infty\) to \(+\infty\). Taking the Fourier transform of Eq. B.4 we have that

\[
\mathcal{F}[\psi(r_\perp, z)] = \mathcal{F} \left[ e^{ik_\perp \nabla_\perp^2 \psi(r_\perp, z = 0)} \right]. 
\]  
(B.7)

Using Eq. B.6, we see that

\[
\nabla_\perp^2 \psi(r_\perp) = \frac{1}{4\pi^2} \int \int (-k_\perp^2) \Psi(k_\perp)e^{i\mathbf{k}_\perp \cdot \mathbf{r}_\perp} \, dk_\perp. 
\]  
(B.8)

Therefore,

\[
f(\nabla_\perp^2)\psi(r_\perp) = \frac{1}{4\pi^2} \int \int f(-k_\perp^2) \Psi(k_\perp)e^{i\mathbf{k}_\perp \cdot \mathbf{r}_\perp} \, dk_\perp. 
\]  
(B.9)

In other words, we have the Fourier transform pair

\[
f(\nabla_\perp^2)\psi(r_\perp) \leftrightarrow f(-k_\perp^2)\Psi(k_\perp). 
\]  
(B.10)

Thus we find that

\[
\mathcal{F}[\psi(r_\perp, z)] = e^{-\frac{i\pi}{4\pi^2}k_\perp^2} \mathcal{F}[\psi(r_\perp, z = 0)]. 
\]  
(B.11)

Or, taking the inverse Fourier transform

\[
\psi(r_\perp, z) = \mathcal{F}^{-1} \left[ e^{-\frac{i\pi}{4\pi^2}k_\perp^2} \mathcal{F}[\psi(r_\perp, z = 0)] \right]. 
\]  
(B.12)

This is the form of the free space propagator used in the implementation of the TFS method and elsewhere in this thesis.
Appendix C

List of publications


