Chapter 1

Introduction

1.1 Historical and Theoretical Background of Phase Retrieval[1]

One form of the "phase problem", as posed by Wolfgang Pauli in his famous Handbuch article, asks whether it is possible to reconstruct a complex spatial wavefunction given both its modulus and the modulus of its Fourier transform (momentum-space representation) (Pauli, 1933). This motivates the related more general phase problem of means for determining a beamlike unbounded complex wavefunction given measurements of probability density over one or more two-dimensional planes (Yu et al., 2005). Such measurements can be obtained via an ensemble of identically prepared quantum systems, e.g. in the context of electron diffraction using a modern source (Bajt et al., 2000); or via a coherent optical system, e.g. in the context of coherent visible light optics (Barty et al., 1998) or coherent X-ray optics (Paganin, 2006).

The latter phase problem is sketched in Figure 1.1, which depicts a coherent mono-energetic spatial wave-field incident upon an elastically scattering potential (sample) A. For the sake of concreteness, this wave might correspond to a coherent electron wavefield in which spin may be neglected. A transmission electron microscope may be used to obtain a focal series of two images B and C, over each plane of which the probability density of the scattered-electron wavefunction is registered (Yu et al., 2005; Bajt et al., 2000). This probability-

[1] A part of this Chapter was taken from a joint paper with David Paganin, Timur Gureyev and Konstantin Pavlov (Schmalz et al., 2011)

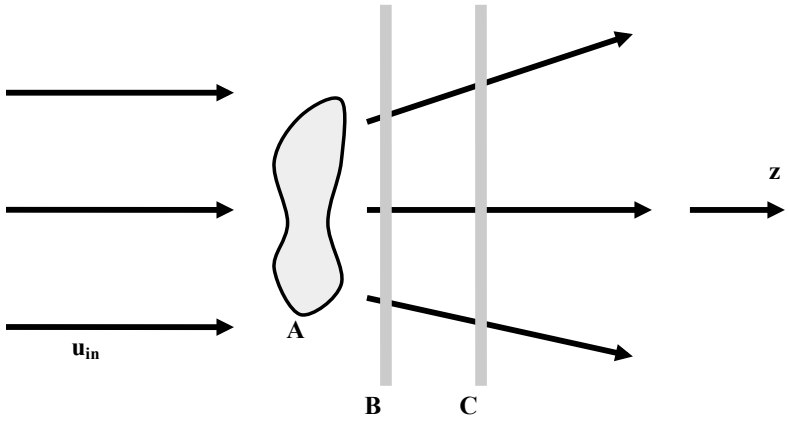

Figure 1.1: Generic phase-retrieval scenario using radiation or matter wavefields. Here A is a sample, B is an object plane, C is an image plane, u_{in} is a complex amplitude of the incident wave and z is the direction of the optical axis.

density map, hereafter termed intensity, is measured through the individual collapse of a large number of almost-identically-prepared electron wavefunctions emitted by a tungsten source or field emission gun (Bajt et al., 2000). If we assume that the scattering potential is sufficiently slowly varying in space, and that the incident field is a z-directed plane wave, then the elastically scattered field will be paraxial (i.e., all probability-current-density vectors, downstream of the object, will make a small angle with respect to the optical axis z) (Paganin, 2006, Ch. 1.4).

This indicative high-energy potential scattering scenario has direct analogs in the scattering of visible light or hard X-rays by a spatially-slowly-varying refractive-index distribution (Wilkins et al., 1996), scattering of coherent visible light by cold atoms (Turner et al., 2004) and Bose-Einstein condensates (Sanner et al., 2011), or in the scattering of neutrons from macroscopic samples (Allman et al., 2000). In all of the above cases, "slowly varying" refers to the scatterer being slowly varying over length scales comparable to the wavelength of the

incident matter or radiation wavefield (Gureyev and Wilkins, 1998).

The phase-reconstruction problem, mentioned in the first paragraph, has ramifications for the problem of structure determination using the methods described in the second paragraph. Specifically, one seeks to reconstruct information regarding the scattering potential A (see Figure 1.1) given measurements of wave-field moduli over planes B, C etc. downstream of the sample A. Often, one has a transparent or weakly-absorbing scatterer A, which implies that the phase but not the magnitude, of the wave-field over the plane B, is significantly affected by the sample (Cowley, 1995, Ch. 3.4). Examples of such phase objects are well known in the context of imaging using visible light (Barty et al., 1998; Turner et al., 2004; Sanner et al., 2011; Bremmer, 1952), electrons (Bajt et al., 2000), (Spence, 2003, Ch. 2), neutrons (Allman et al., 2000) and hard X-rays (Paganin, 2006, Ch. 4.4).

We elaborate on three examples of this common situation regarding phase objects:

(i) Electrons - Medium-resolution bright-field transmission electron microscopy of thin non-crystalline samples often yields an in-focus image with negligible contrast (Spence, 2003, Ch. 2). Under the projection approximation, the phase of the wavefunction over the exit-surface B (see Figure 1.1) is given by the famous Aharonov-Bohm phase factor (Aharonov and Bohm, 1961; Yu et al., 2011), with the probability density over the sample plane being approximately constant.

(ii) X-rays - Materials with similar X-ray attenuation coefficients (e.g., different types of biological soft tissues) are often difficult to distinguish using conventional attenuation-based X-ray imaging or Computed Tomography (CT). In such cases, a suitable form of phase-contrast imaging (PCI) can be helpful (Fitzgerald, 2000; Lewis, 2004; Zhou and Brahme, 2008). For example, propagation-based phase contrast (Gureyev et al., 2009; Nugent, 2009), in which the act of free-space propagation can convert transverse phase variations (over plane B in Figure 1.1) into transverse intensity variations (over plane C), can help to significantly improve the contrast and signal-to-noise ratio in X-ray images of samples consisting predominantly of low-Z materials.

(iii) Neutrons - When performing propagation-based phase-contrast imaging using neutrons (Allman et al., 2000), one often has the case where elements

in the sample of interest (e.g. small organic samples, lead samples, etc.) are transparent, or the closely related case where distinct materials (e.g. titanium and steel) have very similar absorptive properties, but impart different phase shifts on the neutron beam traversing the sample.

To obtain quantitative information from phase-contrast images, as may be required e.g. in X-ray (Momose et al., 1995; Bonse and Busch, 1996; Raven et al., 1996; Cloetens et al., 1997b; Mayo et al., 2003; Pfeiffer et al., 2007; Shi and Anastasio, 2006; Laperle et al., 2007), electron or neutron CT (Yu et al., 2011; Lade et al., 2005a,b; Phatak et al., 2008, 2010), it is usually necessary to perform an image processing operation known as phase retrieval. Phase retrieval uses one or more images, collected with a position-sensitive detector, to produce a quantitative map of projected phase shifts in transverse sections of the beam transmitted through a sample (Paganin, 2006, Ch. 4.5). Image acquisition parameters that are varied between the multiple images used for phase retrieval, may include the sample-to-detector and/or source-to-sample distance, the energy or the spectrum of the illuminating field (Gureyev et al., 2009). In each case a unique algorithm is required to convert the intensity distributions in the collected images into the quantitative maps of the corresponding phase distribution.

One commonly used phase-retrieval method, generic to all of the previously cited modalities using radiation and matter wavefields, is based on the Transport of Intensity Equation (TIE). The TIE was proposed for electromagnetic radiation by Teague in 1983 (Teague, 1983), and later generalised (Paganin and Nugent, 2002) to the wider class of fields considered here. To solve this equation, Teague introduced an auxiliary function that transforms the TIE into a classical two-dimensional Poisson equation, which can be easily solved e.g. by using a Fourier-transform-based method (Teague, 1983; Paganin and Nugent, 1998a) or other approaches (Gureyev et al., 1999). The former method has enjoyed widespread use, with successful experimental applications to phase retrieval using visible light (Barty et al., 1998), neutrons (Allman et al., 2000), X-rays (Gureyev et al., 2009; Nugent, 2009) and electrons (Bajt et al., 2000; De Graef and Zhu, 2001; Petersen et al., 2007; Yu et al., 2010); and proposed experimental applications to Bose-Einstein condensates (Paganin and Nugent, 2002; Martin and Allen, 2007) and atom lasers (Paganin and Nugent, 2002).

However, these and related studies have been undertaken without a detailed analysis of the validity of phase retrieval using Teague's method, usually producing phase maps that are in good agreement with a priori known properties of the sample or as verified by alternative analytical techniques. It is timely that the foundations upon which the method is based be thoroughly examined. Such an examination is the core topic of Chapter 8.

1.2 Application of the Propagation-Based Phase-Retrieval Method to X-Ray Computed Tomography[2]

X-ray computed tomography (CT) is a well-established technique for non-destructive three-dimensional (3D) imaging of internal structure of optically opaque samples (Herman, 1980; Natterer, 2001), (Born and Wolf, 2005, Ch. 4.11). Conventional CT is based on the differential attenuation of transmitted X-rays by different components of the sample. This contrast mechanism is effective for distinguishing between components with significant differences in atomic number or density, e.g. between flesh and bones in the case of medical CT. However, the difference in X-ray attenuation by different types of soft tissues (e.g. healthy and malignant ones) is rather weak, which typically results in poor image contrast.

Propagation-based (or in-line) X-ray images of most samples usually display conventional attenuation contrast as well as phase contrast (Snigirev et al., 1995; Wilkins et al., 1996). The in-line phase contrast appears as the result of free-space propagation of the transmitted beam, which transforms phase variations in the plane located immediately after the sample into measurable intensity variations in the detector plane. The in-line phase contrast will appear provided the propagation distance is large enough for the given X-ray energy, imaging configuration and sample composition and structure, the beam is sufficiently spatially coherent, and the spatial resolution of the detector is sufficiently high. For hard X-rays (with the wavelength of around 1 Å or shorter), the complex refractive index, $n(\mathbf{r}) = 1 - \delta(\mathbf{r}) + i\beta(\mathbf{r})$, of most materials is very close to unity (e.g., the decrement, $\delta(\mathbf{r})$, and the imaginary

[2] A part of this Chapter was taken from a joint paper with Thomas Baillie, Konstantin Pavlov and Timur Gureyev (Baillie et al., 2012)

part, $\beta(\mathbf{r})$, are typically of the order of 10^{-6} and 10^{-8}, respectively), hence the refraction angles are small too and it is usually possible to approximate the X-ray trajectories through a sample by straight lines. Here we also do not consider such diffraction effects as small-angle and wide-angle scattering as it is done for instance in diffraction tomography (Born and Wolf, 2005, Ch 4.11), (Wolf, 1969; Harding et al., 1985; Pavlov et al., 2000). This simplifies the mathematical formalism of phase-contrast tomography (PCT) and leads to relatively simple linear reconstruction algorithms (Cloetens et al., 1997b; Gbur and Wolf, 2002; Bronnikov, 2002; Gureyev et al., 2006). The algorithms are particularly simple in the case of objects exhibiting negligible attenuation (Bronnikov, 1999) and, more generally, in the case of homogeneous objects (Gureyev et al., 2006).

Recently, considerable effort has been applied to developing successful practical strategies for PCT of "general" objects, i.e. the objects that are not homogeneous. X-ray absorption and refractive properties of such objects are both non-trivial and uncorrelated. It is intuitively clear that in order to reconstruct the 3D distribution of the complex refractive index in such objects, one needs to measure 2D intensity distributions in at least two X-ray projections at each rotational position of the sample, with different values of some essential parameter, such as the sample-to-detector distance or the X-ray energy (Teague, 1983; Gureyev and Wilkins, 1998). The main difficulty in the reconstruction of the complex refractive index distribution from such data has so far been presented by the spatial and temporal instability of X-ray beams produced by laboratory sources and synchrotron beamlines. This instability usually prevents one from accurate co-registration of different images collected (at different points of time) at a fixed view angle. Recently some techniques have been suggested to overcome this difficulty (Beltran et al., 2010; Langer et al., 2010), but a general practical solution is yet to be found. In this thesis (Chapter 9) we investigate the options for reconstruction of a newly introduced quantity, which represents a particular combination of the real and imaginary parts of the refractive index, from X-ray PCT data. This quantity is theoretically derived from the phase-retrieval approach which was proposed by Teague (Section 8.3). We show that in the case of objects with negligible attenuation the new quantity coincides with the decrement of the refractive index of the object. In the general

case, the quantity is affected by the imaginary part of the refractive index as well, with the dependence displaying a relatively simple and predictable character, as demonstrated by our numerical experiments. The advantage of the proposed approach is in the significantly simplified form of the PCT algorithm that can be applied for the reconstruction of the introduced quantity. This method resembles the single-step PCT algorithm originally proposed in (Bronnikov, 1999).

1.3 Structure of this Thesis

The literature review, which provides the background needed for the main results, is integrated in Chapters 2 to 6.

Chapter 2 is dedicated to the Helmholtz equation, which describes the interaction of X-rays with matter. We derive the Helmholtz equation from Maxwell's equations, and then present the paraxial form of the Helmholtz equation, on which subsequent material in this thesis depends.

Chapter 3 explains the physical nature of the attenuation index and the refractive index decrement. We provide formulæ showing the connection between the two-dimensional intensity distributions in the image plane, the three-dimensional distribution of the attenuation index and the three-dimensional distribution of the refractive index decrement. We also discuss direct and inverse problems.

Chapter 4 discusses various phase-retrieval methods, such as X-ray interferometry, analyser-based phase retrieval, grating-based phase-retrieval, and propagation-based phase retrieval. One of the propagation-based phase-retrieval methods, based on the TIE, is central to this thesis.

Chapter 5 is devoted to the phase-retrieval method, based on the TIE. Different approaches to the TIE are presented to allow application and interpretation in different contexts. Methods of solving the TIE are discussed.

Chapter 6 provides a brief history of the development of tomography and explains its mathematical basis. A formula is introduced for the reconstruction of the distribution of the absorption index and the problem of three-dimensional refractive index decrement retrieval is discussed.

The subsequent Chapters are based on the author's original work.

Chapter 7 is based on (Schmalz et al., 2010). A method is presented for finding the Green's function for the Helmholtz equation in free space, subject to Sommerfeld radiation conditions. In contrast to the methods used in much of the literature, the method presented in Chapter 7 allows one to find all possible Green's functions before choosing the one that satisfies the boundary conditions.

Chapter 8 is based on (Schmalz et al., 2011) and examines the validity of the widely used solution method of the TIE originally suggested by Teague (Teague, 1983). We prove that Teague's method is accurate for both the phase and homogeneous objects. In the general case we deduce a sufficient condition for the solution obtained by Teague's method to be close to the real solution. While these conditions are generally satisfied in practice, this is not always the case. We construct a counter-example, where the actual solution and the solution obtained by Teague's method differ considerably.

Chapter 9 applies the method proposed by Teague to tomography (Baillie et al., 2012), with the aim of simplifying the reconstruction of the refractive index decrement distribution δ. A new function, introduced to achieve this, is a combination of the real and imaginary parts of the refractive index n. We describe a single-step method to calculate this function directly from the intensity distributions. Then we study the connection between the newly introduced function and the refraction index decrement. This connection is supported by numerical experiments.

For ease of reading, some proofs and calculations are relegated to the Appendix.

Chapter 2

Helmholtz Equation

2.1 Introduction

The Helmholtz equation (Sommerfeld, 1949), (Born and Wolf, 2005, Ch. 1.1) represents the interaction of monochromatic X-ray wave field with matter.

$$(\nabla^2 + k^2 n^2(x, y, z))\psi_\omega(x, y, z) = 0, \tag{2.1}$$

where $\psi_\omega(x, y, z)$ is the spatial part of the wave function $\psi_\omega(x, y, z)e^{-i\omega t}$, $\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$ is the 3D Laplacian, $k = \frac{2\pi}{\lambda}$, λ is the wavelength, and n is the complex refractive index (see Chapter 1.2). The Helmholtz equation plays a central role in the theory of scattering and radiation of electromagnetic waves (Merzbacher, 1961, Ch. 11), (Sommerfeld, 1949, Ch. 2), (Davydov, 1973, Ch. 14), (Born and Wolf, 2005, Ch. 1.1), (Morse and Feshbach, 1953, Ch. 2.1), (Jackson, 1998, Ch. 1), (Messiah, 1999, Ch. 1), (Paganin, 2006, Ch. 2). This Chapter derives the Helmholtz equation from Maxwell's equations, and introduces the paraxial form of the Helmholtz equation.

2.2 Derivation of the Helmholtz Equation from Maxwell's Equations

Maxwell's equations are four partial differential equations (Jackson, 1998, Ch. 1), (Paganin, 2006, Ch. 1). The first one is

$$\nabla \cdot \mathbf{D}(x, y, z, t) = \rho(x, y, z, t). \tag{2.2}$$

Here $\mathbf{D}(x, y, z, t)$ is the electric displacement, and $\rho(x, y, z, t)$ is the free electric charge density, and $\nabla = (\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z})$. The second Maxwell's equation is Gauss's Law for magnetism. It states that magnetic monopoles do not exist:

$$\nabla \cdot \mathbf{B}(x, y, z, t) = 0. \tag{2.3}$$

Here $\mathbf{B}(x, y, z, t)$ is the magnetic induction. The next equation is Faraday's Law of induction:

$$\nabla \times \mathbf{E}(x, y, z, t) + \frac{\partial}{\partial t}\mathbf{B}(x, y, z, t) = \mathbf{0}, \tag{2.4}$$

where $\mathbf{E}(x, y, z, t)$ is the electric field. The last one is Ampere's circuit Law modified by Maxwell:

$$\nabla \times \mathbf{H}(x, y, z, t) - \frac{\partial}{\partial t}\mathbf{D}(x, y, z, t) = \mathbf{J}(x, y, z, t). \tag{2.5}$$

Here $\mathbf{H}(x, y, z, t)$ is the magnetic field and $\mathbf{J}(x, y, z, t)$ is the free electric current density.

We derive the Helmholtz equation under the following assumptions. We assume that the materials are linear and isotropic, i.e. $\mathbf{D} = \epsilon\mathbf{E}$ and $\mathbf{B} = \mu\mathbf{H}$, where ϵ is the electrical permittivity and μ is the magnetic permeability. We assume also that ϵ does not depend on time, and that the materials are non-magnetic, i.e. the magnetic permeability μ is equal to the magnetic permeability in the free space μ_0. For magnetic materials we would need to take into account the bound current generated by this material in magnetic field. Further, we take $\rho = 0$ and $\mathbf{J} = \mathbf{0}$ as free charge density and free current density are not present. Then Maxwell's equations can be rewritten as

$$\nabla \cdot (\epsilon(x, y, z)\mathbf{E}(x, y, z, t)) = 0, \tag{2.6}$$

$$\nabla \cdot (\mathbf{H}(x, y, z, t)) = 0, \tag{2.7}$$

$$\nabla \times \mathbf{E}(x, y, z, t) + \frac{\partial}{\partial t}\left(\mu_0\mathbf{H}(x, y, z, t)\right) = \mathbf{0}, \tag{2.8}$$

$$\nabla \times \mathbf{H}(x, y, z, t) - \frac{\partial}{\partial t}\left(\epsilon(x, y, z)\mathbf{E}(x, y, z, t)\right) = \mathbf{0}. \tag{2.9}$$

Apply the curl to Equation (2.8) and get

$$\nabla(\nabla \cdot \mathbf{E}) - \nabla^2 \mathbf{E} + \frac{\partial}{\partial t}(\mu_0 \nabla \times \mathbf{H}) = 0. \tag{2.10}$$

Now substitute the expression for $\nabla \times \mathbf{H}$ from Equation (2.9):

$$\nabla(\nabla \cdot \mathbf{E}) - \nabla^2 \mathbf{E} + \mu_0 \epsilon(x, y, z) \frac{\partial^2}{\partial t^2} \mathbf{E} = 0, \tag{2.11}$$

or

$$\left(\mu_0 \epsilon(x, y, z) \frac{\partial^2}{\partial t^2} - \nabla^2\right) \mathbf{E} = -\nabla(\nabla \cdot \mathbf{E}). \tag{2.12}$$

The electric displacement is equal to $\mathbf{D} = \epsilon_0 \epsilon_r(x, y, z)\mathbf{E}$, where ϵ_0 is the permittivity in the free space, and ϵ_r is the relative permittivity $\epsilon_r = 1 - \frac{r_e N \lambda^2}{\pi}$ (Pinsker, 1974, Eq. 2.27), (Authier, 2005, Ch. 2). As the classical electron radius r_e is approximately equal to 2.8×10^{-15}м (Authier, 2005, Eq. 2.32), X-rays' wavelength λ is of the order 10^{-10}м and the number of electrons in a unit volume N is estimated to be $10^{29} - 10^{31}$м$^{-3}$ (Pinsker, 1974, Ch. 2.1), we conclude that the term $\frac{r_e N \lambda^2}{\pi}$ is of the order $10^{-6} - 10^{-4}$, and therefore $\mathbf{D} \approx \epsilon_0 \mathbf{E}$. From Equation (2.2) and the assumption that $\rho = 0$ we conclude that $\nabla \cdot \mathbf{E} \approx 0$. Then the right-hand side of Equation (2.12) can be neglected:

$$\left(\mu_0 \epsilon(x, y, z) \frac{\partial^2}{\partial t^2} - \nabla^2\right) \mathbf{E} = 0. \tag{2.13}$$

Analogously we apply the curl to Equation (2.9), substitute the expression for the curl $\mathbf{E}$ (from Equation (2.8)) and obtain

$$\left(\mu_0 \epsilon(x, y, z) \frac{\partial^2}{\partial t^2} - \nabla^2\right) \mathbf{H} = -\nabla(\nabla \cdot \mathbf{H}) = 0, \tag{2.14}$$

as follows from Equation (2.7). Equations (2.13) and (2.14) are called wave equations, each describes the three-dimensional wave function. Consequently the Cartesian components of both vector functions $\mathbf{E}$ and $\mathbf{H}$ satisfy the wave equation. A scalar complex function $\Psi(x, y, z, t)$, which corresponds to the Cartesian components of the wave field, is introduced. It satisfies the wave equation

$$\left(\epsilon(x, y, z)\mu_0 \frac{\partial^2}{\partial t^2} - \nabla^2\right) \Psi(x, y, z, t) = 0. \tag{2.15}$$

We write the wave function as the Fourier integral (see Appendix A) over angular frequencies ω:

$$\Psi(x, y, z, t) = \int_0^\infty \psi_\omega(x, y, z) e^{-2\pi i \omega t} d\omega. \tag{2.16}$$

Substitution of this decomposition into Equation (2.15) gives

$$\int_0^\infty \left(\nabla^2 + \epsilon(x, y, z)\mu_0\omega^2\right) \psi_\omega(x, y, z) e^{-2\pi i \omega t} d\omega = 0. \tag{2.17}$$

As the Fourier transform is invertible (Property A.8), its null space is trivial, and therefore

$$\left(\nabla^2 + \epsilon(x, y, z)\mu_0\omega^2\right) \psi_\omega(x, y, z) = 0. \tag{2.18}$$

Using the following formulæ $\omega = ck$ (where c is the speed of light in vacuum) and $n^2(x, y, z) = c^2\epsilon(x, y, z)\mu_0$ (Born and Wolf, 2005, Ch. 1.2) we get the Helmholtz equation in matter

$$(\nabla^2 + k^2 n^2(x, y, z))\psi_\omega(x, y, z) = 0. \tag{2.19}$$

We devote Chapter 7 (Schmalz et al., 2010) to a rigorous mathematical solution to the Helmholtz equation in vacuum.

2.3 Paraxial Helmholtz Equation in the Presence of Matter

In this section we derive a paraxial variant of the Helmholtz equation. Assume the paraxial approximation, it means that the directions of X-rays after their interaction with matter are close to the previous illumination direction, which is the positive z-direction. The perturbed plane wave $\psi_\omega(x, y, z) = \psi(x, y, z)e^{ikz}$ is a solution of the Helmholtz equation (Equation (2.1)). For paraxial approximation the following conditions

$$\left|\frac{\partial\psi}{\partial z}\right| << |k\psi| \tag{2.20}$$

and

$$\left|\frac{\partial^2\psi}{\partial z^2}\right| << \left|k\frac{\partial\psi}{\partial z}\right| \tag{2.21}$$

are satisfied (Nieto-Vesperinas, 2006, Ch. 3.7). We substitute the expression for perturbed plane wave into the Helmholtz Equation (2.1):

$$\nabla^2(\psi(x,y,z)e^{ikz}) + k^2n^2\psi(x,y,z)e^{ikz} = 0. \tag{2.22}$$

After some algebra, using $\nabla^2(fg) = g\nabla^2 f + f\nabla^2 g + 2\nabla f \nabla g$ and neglecting the second derivative $\frac{\partial^2\psi}{\partial z^2}$ in comparison with $k\frac{\partial\psi}{\partial z}$ (Equation (2.21)), we obtain

$$\left(\nabla_\perp^2\psi(x,y,z) - \psi(x,y,z)k^2 + 2\frac{\partial}{\partial z}\psi(x,y,z)ik + k^2n^2\psi(x,y,z)\right)e^{ikz} = 0, \tag{2.23}$$

where $\nabla_\perp^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$. We arrive at the paraxial Helmholtz equation in the presence of matter:

$$\left(2ik\frac{\partial}{\partial z} + \nabla_\perp^2 + k^2(n^2(x,y,z) - 1)\right)\psi(x,y,z) = 0. \tag{2.24}$$

This equation is the starting point for the derivation of the Transport of Intensity Equation in Chapter 5.

Chapter 3

X-ray Imaging

3.1 Introduction

Consider an object illuminated by an X-ray plane wave. The two-dimensional intensity distribution registered by the detector characterises the internal structure of the object on the basis of attenuation or/and scattering properties of materials contained in the object.

The equation for a monochromatic plane wave in vacuum is $\Psi_0(\mathbf{r}, t) = Ae^{i(\mathbf{k}_0\mathbf{r}-\omega t)}$. Here the vector $\mathbf{r} = (x, y, z)$ gives the coordinates of a point in Cartesian coordinates. The wave-vector $\mathbf{k}_0 = (k_{0x}, k_{0y}, k_{0z})$ gives the direction of the propagation of the plane wave in the same Cartesian coordinates. The magnitude of this vector $|\mathbf{k}_0| = k$. The complex amplitude of the plane wave is denoted by A, the angular frequency ω is equal to $\frac{2\pi}{T}$, where T is the period of the wave $T = \frac{\lambda}{c}$. We denote the spatial part of the wave function as $\psi_\omega(\mathbf{r}) = Ae^{i\mathbf{k}_0\mathbf{r}}$. Suppose the plane wave propagates along the z-direction, i.e. $\mathbf{k}_0 = (0, 0, k)$.

After interaction with the material of the object, the spatial part of the perturbed wave is described by the equation $\psi_\omega = \tilde{\psi}(\mathbf{r})e^{i\mathbf{k}\mathbf{r}}$. The intensity distribution over the object plane $z = z_0$ is equal to $I(x, y, z = z_0) = |\psi_\omega|^2(x, y, z = z_0)$. An object plane is a plane which is perpendicular to the direction of illumination and located immediately downstream the object along the direction of propagation of the X-ray wave (Fig. 3.1). The goal of the absorption contrast imaging is to characterise the internal structure of the object on the basis of the measured intensity on the object plane. Absorption contrast imaging method is described in Section 3.3.

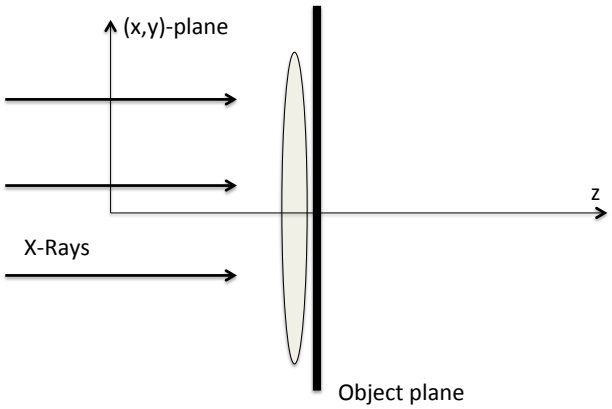

Figure 3.1: An object is illuminated by an X-ray plane wave. The object plane is a plane which is perpendicular to the direction of illumination and located immediately downstream the object along the direction of propagation of the X-ray wave.

The scattering effect is described in Section 3.4.

3.2 Refractive index

The interaction of X-rays with matter is described by the refractive index $n(x, y, z) = 1 - \delta(x, y, z) + i\beta(x, y, z)$, where δ and β are called the refractive index decrement and the absorption index, respectively. The wave vector $\mathbf{k}_0$ (in vacuum) will change into $n\mathbf{k}_0$ while the wave interacts with medium. We assume that the conditions for the projection approximation are satisfied, i.e. the scatterers are sufficiently weak so that the ray paths through scatterers are close to straight lines (Paganin, 2006, Ch. 2.2). Then the perturbed wave function in the object plane is

$$\psi_\omega(x, y, z = z_0) = Ae^{-k\int \beta(x,y,z)\,dz} \times e^{ik\int (1-\delta(x,y,z))\,dz}. \tag{3.1}$$

The amplitude of the wave changes by the factor $e^{-k\int\beta(x,y,z)dz}$. The expression $\Delta\phi = -k\int\delta(x,y,z)dz$ corresponds to the phase shift due to interaction with matter. The physical nature of the absorption index and refractive index decrement will be discussed in the next sections.

3.3 Absorption Cross Section, Attenuation Coefficient, Absorption Index

The effects of attenuation can be described in terms of the interaction of X-ray photons with atoms of the material (Als-Nielsen and McMorrow, 2001, Ch. 1), (Guinier, 1963, Ch. 1). Some photons do not reach the detector placed after the object, because they are absorbed or scattered by the atoms.

The effect when a photon is absorbed by an atom is called photoelectric absorption. The energy of the X-ray photon is transferred to an electron of the atom, the electron is ejected. The vacancy in the shell is filled by another electron from the outer shell. The energy released by the electron moving to the lower energy orbit can be used to emit a photon (fluorescent X-ray emission) or to expel another electron (Auger electron emission).

A photon can be scattered, i.e. deviate from a straight trajectory after the interaction with medium. If the energy of the photon does not change, the scattering is called elastic. A photon can be scattered from a bound electron (Rayleigh scattering) or from a free electron (Thomson scattering). Inelastic scattering or Compton scattering is the scattering with a loss of energy (Guinier, 1963, Ch. 1.2.3). All three types of scattering contribute to the attenuation coefficient.

The initial intensity I is the number of photons registered in the beam per unit area of the cross section which is perpendicular to the illumination direction and per second. Let us consider first the photoelectric absorption effect. The loss of intensity ΔI is the number of photons absorbed when passing through the thin layer with the thickness Δz per unit area, and per second. Then ΔI is proportional to I and the number of atoms per unit area in the layer $\tilde{\rho}(\mathbf{r})\Delta z$, where $\tilde{\rho}(\mathbf{r})$ is the number of atoms per unit volume. The coefficient of

proportionality is called the absorption cross section and is denoted by $\sigma(\mathbf{r})$:

$$\Delta I = -\sigma(\mathbf{r})\, I\, \tilde{\rho}(\mathbf{r})\, \Delta z. \tag{3.2}$$

The absorption cross section depends also on the energy of photons $\mathcal{E}$ and the atomic number of the absorber Z. It increases with approximately the fourth power of the atomic number of the absorber for light elements, and with approximately the third power of the atomic number of the absorber for elements with high Z. The absorption cross section decreases with approximately the third power of the energy of photons (Als-Nielsen and McMorrow, 2001, Ch. 6).

We denote the product $\sigma(\mathbf{r})\,\tilde{\rho}(\mathbf{r})$ by $\mu_{abs}(\mathbf{r})$. Then we may rewrite the Equation (3.2) as

$$\Delta I = -\mu_{abs}(\mathbf{r})\, I\, \Delta z. \tag{3.3}$$

Here $\mu_{abs}(\mathbf{r})$ is the absorption coefficient and can be determined experimentally. Note that scattering effects, considered in the subsequent Section, also diminish the number of photons transmitted through an object. Therefore the linear attenuation coefficient, μ, takes into account both the absorption, described by $\mu_{abs}(\mathbf{r})$, and scattering effects. For every small distance Δz, the approximate loss of intensity in the matter is $\frac{\Delta I}{I} = -\mu(x, y, z)\Delta z$. Letting Δz approach zero, we get the differential equation $\frac{dI}{I} = -\mu(x, y, z)dz$ with the initial condition $I(x, y, z = 0) = I_{in}$. The solution to this differential equation is $I(x, y, z = z_0) = I_{in} e^{-\int \mu(x,y,z)\,dz}$. We note that the direct problem has a unique solution. If the distribution of the linear attenuation coefficient in the object is known, this formula gives the distribution of the intensity $I(x, y, z = z_0)$ over the object plane. The inverse problem we are interested in, does not have a unique solution. If the intensity distribution over the object plane $I(x, y, z = z_0)$ is measured, we can reconstruct the projection of the linear attenuation coefficient:

$$\int \mu(x, y, z)dz = -\ln \frac{I(x, y, z = z_0)}{I_{in}}, \tag{3.4}$$

but not the linear attenuation coefficient itself. The projection integral contains the information about the linear attenuation coefficient and the thickness of an object. The linear attenuation coefficient μ and the absorption index β are related by the equation $\mu = 2k\beta$. To reconstruct the 3D distribution of the

absorption index in the object, we need to take some images of the object from different viewing angles. This is what tomography does and is discussed in Chapter 6.

3.4 Scattering, Refractive Index Decrement, Phase Contrast

The previous section described the effect of absorption of photons by the atoms of matter. Here we discuss the scattering effect. An X-ray beam with initial intensity I_{in} interacts with matter. Let I_{sc} be the number of photons scattered per second into a solid angle $\Delta\Omega$. The number of scattered photons I_{sc} is proportional to I_{in}, to the angle $\Delta\Omega$, and to the number of atoms $\tilde{N}$ in the sample per unit area seen along the incident beam direction. The coefficient of proportionality is called the differential cross section and is denoted as $\frac{d\sigma}{d\Omega}$ (Als-Nielsen and McMorrow, 2001, Ch. 1.2, Appendix A). The differential cross section characterises how efficiently the matter scatters the photons.

$$I_{sc} = \frac{d\sigma}{d\Omega} I_{in} \tilde{N} \Delta\Omega \tag{3.5}$$

Both elastic and inelastic scattering effects are included in this formula.

The angle of scattering is proportional to the gradient of the phase shift. The refractive index decrement δ is connected with the electron density ρ in the scatterer, the classical electron radius r_e, and the wave length $\lambda = \frac{2\pi}{k}$ by the equation (Als-Nielsen and McMorrow, 2001, Ch. 3.1)

$$\delta = \frac{2\pi\rho r_e}{k^2}, \tag{3.6}$$

where the classical electron radius r_e is calculated by the formula (Als-Nielsen and McMorrow, 2001, Ch. 1.2)

$$r_e = \frac{e^2}{4\pi\epsilon_0 m_e c^2}. \tag{3.7}$$

Here e is the electric charge of the electron, m_e is the mass of the electron.

We recall that the refractive index decrement δ characterises the additional phase shift: $\Delta\phi = -k\int \delta(x, y, z)dz$. However the intensity measured on the object plane, $|\psi_\omega(x, y, z = z_0)|^2 = |Ae^{-k\int \beta(x,y,z)dz}|^2$, gives no information about the phase shift. Methods for phase shift retrieval are discussed in Chapter 4.

3.5 Summary

This Chapter describes the physical nature of the linear attenuation coefficient μ and the refractive index decrement δ. Two formulæ, which are essential for the subsequent Chapters, are introduced. One of these formulæ, $\int \mu(x, y, z)dz = -\ln \frac{I(x,y,z=z_0)}{I_{in}}$, describes the relationship between the projection along the illumination direction of the linear attenuation coefficient and the change of X-ray intensity resulting from the interaction with matter. The other, $\int \delta(x, y, z)dz = -\frac{1}{k}\Delta\phi$, gives the relationship between the projection of the refractive index decrement and the phase shift of X-rays, caused by the interaction with matter.

Chapter 4

Phase-Retrieval Methods

4.1 Introduction

The aim of phase-contrast tomography is to reconstruct the distribution of the refractive index decrement δ in objects, which can be retrieved from the phase distribution over the object plane. As detectors measure intensity, but not the phase, the problem arises of phase retrieval from the intensity distribution. This Chapter discusses phase-retrieval methods such as X-ray interferometry in Section 4.2, analyser-based phase-retrieval method in Section 4.3, grating-based phase-retrieval method in Section 4.4 and propagation-based phase-retrieval method in Section 4.5.

4.2 X-Ray Interferometry

The principle of an interferometer is explained on the example of Bonse-Hart interferometer (Bonse and Hart, 1965), see Figure 4.1. The initial monochromatic plane wave has the form $Ae^{i\phi}$. The first crystal (beam splitter) of the interferometer splits this plane wave into two waves, the second crystal (mirror) changes the direction of propagation of these two waves and brings them together, the third crystal (analyser) is used for the visualisation of the Moire image. The analyser splits both plane waves. The propagated plane wave from one of the arms of the interferometer is combined with the reflected plane wave from another arm of the interferometer. A phase shifter creates a phase shift $\Delta\phi$ for one of these waves: $Ae^{i(\phi-\Delta\phi)}$. The waves $Ae^{i\phi}$ and $Ae^{i(\phi-\Delta\phi)}$ overlap to create an interferometric picture on the detector

plane in absence of the object. The object of interest is placed in one of

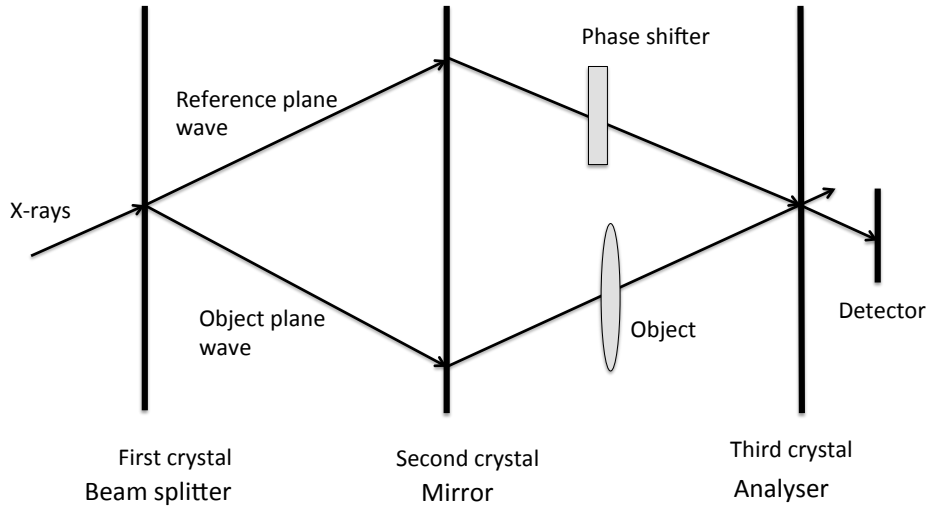

Figure 4.1: Bonse-Hart X-ray interferometer.

the arms of the interferometer, its presence changes the interference picture. Suppose that the wave disturbed as the result of the interaction with the object, is described by the equation $B(x,y)e^{i(\phi+\Phi(x,y))}$. Here $B(x,y)$ is an amplitude, $\Phi(x,y)$ is the phase shift created by the interaction with the sample $\Phi(x,y) = -\frac{2\pi}{\lambda}\int \delta(x,y,z)dz$. The intensity of the interference picture of the waves $Ae^{i(\phi-\Delta\phi)}$ and $B(x,y)e^{i(\phi+\Phi(x,y))}$ is described by the following equation

$$\begin{aligned} I(x,y) = & \\ & \left(Ae^{i(\phi-\Delta\phi)} + B(x,y)e^{i(\Phi(x,y)+\phi)}\right)\left(Ae^{i(\phi-\Delta\phi)} + B(x,y)e^{i(\Phi(x,y)+\phi)}\right)^* \\ & = A^2 + B^2(x,y) + 2AB(x,y)\cos(\Phi(x,y)+\Delta\phi), \quad (4.1) \end{aligned}$$

here the asterisk denotes the complex conjugate. We denote the average intensity $A^2 + B^2(x,y)$ by $a(x,y)$, and the fringe amplitude $2AB(x,y)$ by $b(x,y)$. Then the last equation can be rewritten as

$$I(x,y) = a(x,y) + b(x,y)\cos(\Phi(x,y)+\Delta\phi). \quad (4.2)$$

Two methods of phase retrieval from Equation (4.2) are discussed in this section.

The first method is called the fringe-scanning method or phase-shifting interferometry (Bruning et al., 1974). Some number $M > 3$ of intensity distribution measurements are made for phase shifts $\Delta\phi + \frac{2\pi p}{M}$, which are created by the phase shifter. The sum for all p from 1 to M from the corresponding intensities with the weight $e^{-\frac{2\pi ip}{M}}$ is taken. The sum gives

$$S(x,y) = \sum_{p=1}^{M} \left(a(x,y) + b(x,y)\cos(\Phi(x,y) + \Delta\phi + \frac{2\pi p}{M}) \right) e^{-\frac{2\pi ip}{M}}$$
$$= \frac{1}{2} M b(x,y) e^{i(\Phi(x,y)+\Delta\phi)}. \quad (4.3)$$

It follows that

$$\Phi(x,y) = \arg(S(x,y)) - \Delta\phi \ (\mathrm{mod} 2\pi). \quad (4.4)$$

The phase shift $\Delta\phi$ can be determined in the absence of an object. An indispensable feature of this method is that more than $M > 3$ measurements must be taken.

The second method of phase retrieval is called the Fourier transform method (Takeda et al., 1982). Carrier fringes of the frequency f_0 in the x-axis direction are added to the interferometric picture using a wedge phase shifter. The intensity is given by

$$I(x,y) = a(x,y) + b(x,y)\cos(\Phi(x,y) + \Delta\phi + 2\pi f_0 x). \quad (4.5)$$

We can rewrite this formula as

$$I(x,y) = a(x,y) + c(x,y)e^{2\pi i f_0 x} + c^*(x,y)e^{-2\pi i f_0 x}, \quad (4.6)$$

where

$$c(x,y) = \frac{1}{2} b(x,y) e^{i(\Phi(x,y)+\Delta\phi)}. \quad (4.7)$$

Take the Fourier transform (Equation (A.1)) of Equation (4.6) with respect to x.

$$\mathcal{F}_1[I](\xi,y) = \mathcal{F}_1(a)(\xi,y) + \mathcal{F}_1(c)(\xi - f_0, y) + \mathcal{F}_1(c^*)(\xi + f_0, y). \quad (4.8)$$

Each summand in this equation corresponds to a peak in Fourier space (Figure (4.2)).

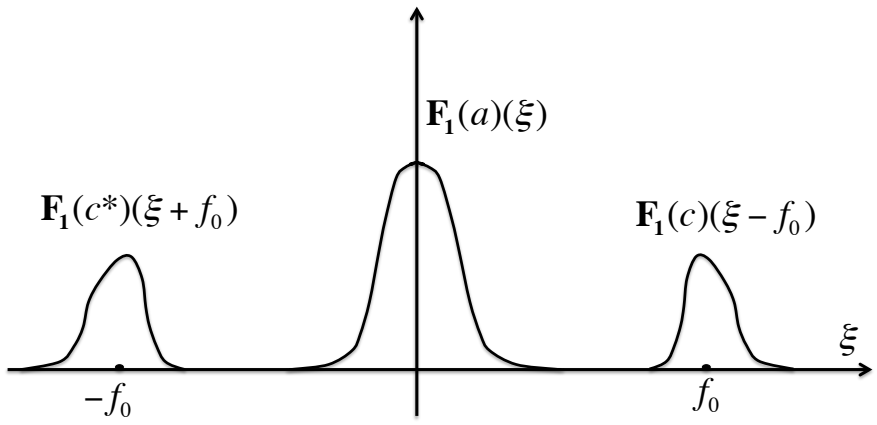

Figure 4.2: Peaks in Fourier space, corresponding to summands of Equation (4.8).

These peaks do not intersect if their widths are small in comparison with f_0, which defines the distance between two neighbour peaks. This is true if $a(x, y)$, $b(x, y)$ and $\Phi(x, y)$ are slowly varying functions. We take, for example, the peak corresponding to the second term $\mathbf{F_1}(c)(\xi - f_0, y)$ and transpose it to the origin $\mathbf{F_1}(c)(\xi, y)$. Taking the inverse Fourier transform (Equation (A.2)), we find $c(x, y)$. From Equation (4.7) we conclude that

$$\Phi(x, y) = \arg(c(x, y)) - \Delta\phi \pmod{2\pi}. \tag{4.9}$$

The advantage of the Fourier transform method is that only one measurement of the intensity distribution is needed for phase retrieval. The disadvantage is that it can be applied only if $a(x, y)$, $b(x, y)$ and $\Phi(x, y)$ are slowly varying functions.

It can be seen from Equations (4.4) and (4.9) that the problem of phase retrieval has a unique solution if the phase shift $\Phi(x, y)$ is smaller than 2π, which is true only for thin objects. For example, for hard X-rays with wavelength of the order $\lambda \approx 10^{-10}\,m$ and materials with light carbon-containing compounds

($\delta \approx 10^{-6}$), the phase shift caused by the object, is $\frac{2\pi\delta l}{\lambda} \approx 2\pi l \times 10^4$, where l is the length (in metres) of the path through the object. In this case, the phase shift is smaller than 2π only for the objects thinner than $l \approx 100\,\mu m$. Figure (4.3) shows how to retrieve the phase for objects creating a phase shift larger than 2π in the one-dimensional case. The graph of wrapped phase is shown

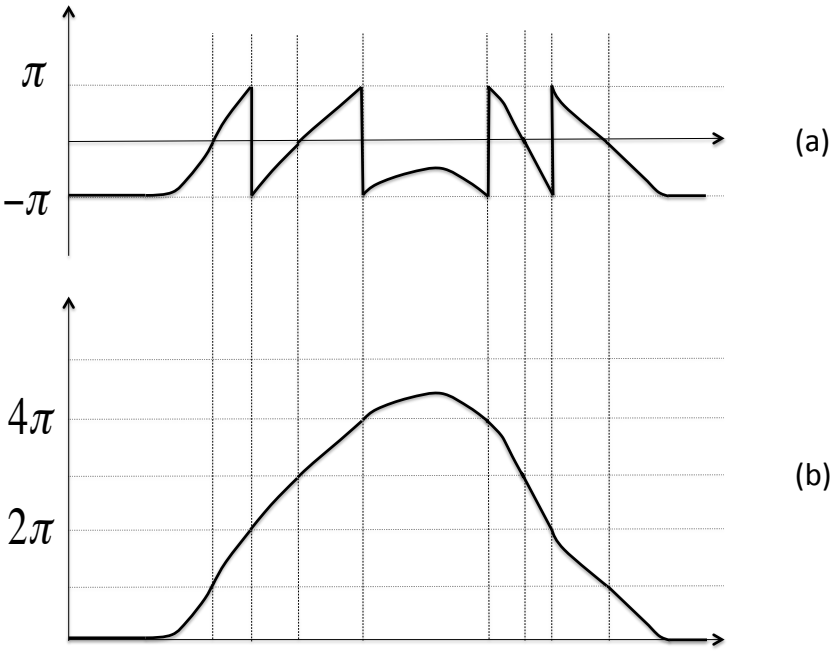

Figure 4.3: (a) Wrapped phase, (b) Unwrapped phase.

in Figure (4.3 (a)). If the phase change is continuous, the phase graph can be reconstructed as shown in Figure (4.3 (b)). Problems can arise in phase unwrapping because of noise from devices, because of a phase jump between adjoined pixels, or because of holes in the object (Judge and Bryanston-Cross, 1994). There are several methods for phase unwrapping in one-dimensional and two-dimensional cases, including cellular automata method (Ghiglia et al., 1987), cut methods of phase unwrapping (Goldstein et al., 1988), phase unwrapping by regions (Gierloff, 1987) and temporal phase unwrapping (Huntley and Saldner, 1993).

4.3 Analyser-Based Phase Retrieval

Analyser-based phase retrieval uses a strong reflection by a perfect crystal of X-rays which are incident on the family of atomic planes at an angle θ_B (and close to this angle), satisfying Bragg's law, i.e. $\lambda = 2d \sin\theta_B$. Here d is the spacing between the two adjacent atomic planes in the crystal. An experimental scheme of the analyser-based phase-contrast imaging is shown on Figure 4.4 (Bushuev et al., 1998; Bravin, 2003), (Paganin, 2006, Ch. 4.4.3). X-rays change their propagation direction as they pass through the sample. This angular deviation is proportional to the gradient of the phase of the wave. Suppose a monochromatic plane wave with unit amplitude is incident on the

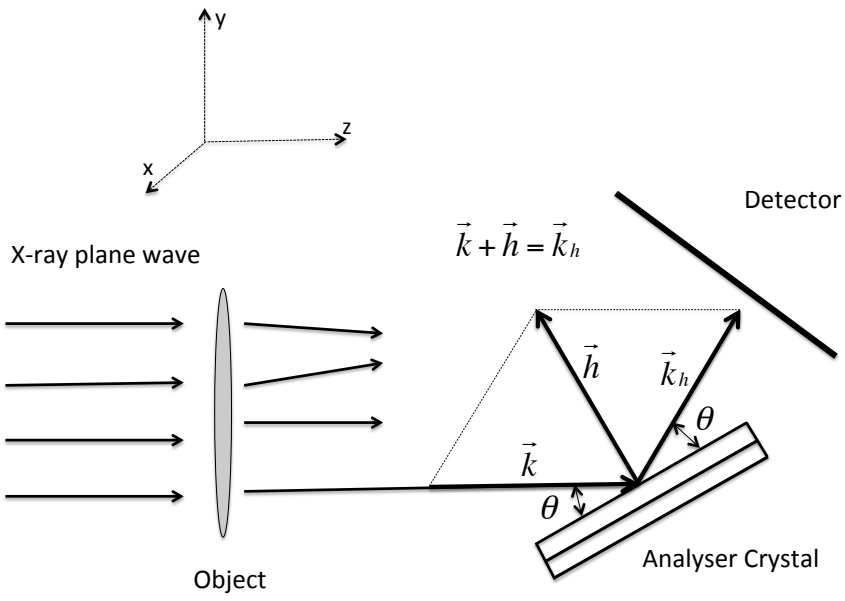

Figure 4.4: Experimental scheme of analyser-based phase-contrast imaging. Wave vector of the incident wave is denoted by $\vec{k}$, $\vec{k_h}$ is a wave vector for the scattered wave, and $\vec{h}$ ($|\vec{h}| = \frac{2\pi}{d}$) is perpendicular to the scattering plane.

family of crystal's atomic planes at the angle θ. The graph of the function (for non-absorbing crystals) describing the dependence of the intensity of the reflected radiation on the incident angle, is called the rocking curve (Figure

(4.5)).

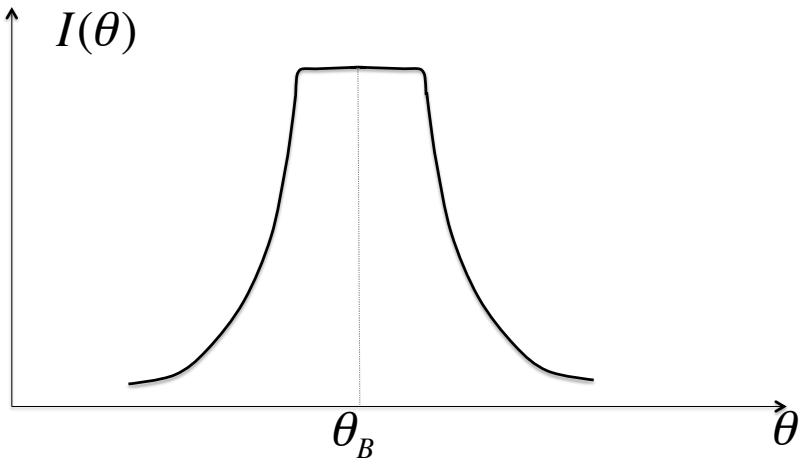

Figure 4.5: Rocking curve of the analyser for non-absorbing crystals (Pinsker, 1974). Note that the effect of refraction in the analyser is neglected.

If the absorption in the crystal is negligible, the rocking curve is symmetrical with respect to an angle equal to the Bragg angle $\theta = \theta_B$ if one neglects the effect of refraction in the analyser crystal. For an absorbing crystal the shape of the rocking curve becomes asymmetrical (Pinsker, 1974, Ch. 11.3). A typical width of the rocking curve for hard X-rays is of the order of a few micro-radians, which makes the experimental scheme sensitive to small angular deviations of the X-rays propagation direction.

We denote by $\psi(x, y)$ the complex amplitude of the wave, incident on the analyser crystal, and by $\psi_{out}(x, y)$ the complex amplitude of the scattered wave. The point spread function $G(y)$ is a function satisfying the following property (Paganin, 2006, Ch. 4.4.3):

$$\psi_{out}(x, y) = \psi(x, y) * G(y), \tag{4.10}$$

where $*$ is the convolution with respect to y. The function $G(y)$ depends only on y, because the crystal is sensitive only to the variation of the wave vector $\mathbf{k}$ lying in the yz-plane (the diffraction plane) (Figure 4.4). Take the one-dimensional Fourier transform (Equation (A.1)) of the left and right hand side of Equation (4.10) with respect to y:

$$\mathbf{F_1}\left[\psi_{out}\right](x, k_y) = \mathbf{F_1}\left[\psi\right](x, k_y) \cdot \mathbf{F_1}\left[G\right](k_y). \tag{4.11}$$

The Fourier transform of the point spread function, $\mathbf{F_1}\left[G\right](k_y)$, is a transfer function of the analyser. The variable k_y is conjugate to y, and is proportional to the angle of deviation from the Bragg angle. We denote the transfer function by $\mathbf{F_1}[G](k_y) = r(\theta)$, the modulus squared of this function $|r(\theta)|^2$ is the rocking curve. The aim of the analyser-based phase-retrieval method is to find the function $\psi(x, y)$.

The problem of finding the function $\psi(x, y)$, which characterises the distribution of the refractive index in the object, is a nontrivial inverse problem. There are methods to solve it in some cases, based on different approximations (Diemoz et al., 2010).

The most frequently used approximation is the geometrical optics approximation (GO). This approximation is valid if the phase of the wave incident on an analyser crystal is a slowly varying function over the length scale of the crystal's extinction length, which is inversely proportional to the rocking curve width (Gureyev and Wilkins, 1997; Pavlov et al., 2004). The extinction length is the depth of a crystal, in which the intensity of the initial beam is reduced by e times. (Als-Nielsen and McMorrow, 2001, Ch. 5.4.2).

One of the GO approximation methods is the diffraction-enhanced imaging (DEI) algorithm (Chapman et al., 1997). It is based on the assumption that the rocking curve can be locally approximated by its Taylor series up to the first order term. Only two images are required for the DEI algorithm (one image at each slope of the rocking curve), which is an advantage compared with other methods mentioned below (e.g. generalised diffraction-enhanced imaging algorithm (Rigon et al., 2007), or pixel-by-pixel Gaussian curve fitting algorithm (Huang et al., 2007)). But if the refraction angles are of the order of the rocking curve width or more, the approximation leads to incorrect results (Diemoz et al., 2010).

The extended diffraction-enhanced imaging algorithm also requires only two images, but the rocking curve itself is used instead of its linearisation (Maksimenko, 2007). Another method (Hu et al., 2008) is based on the approximation of the rocking curve by a Gaussian curve.

The generalised diffraction-enhanced imaging algorithm (Rigon et al., 2007) is based on the local second-order Taylor series approximation for the rocking curve. It is more accurate for refraction angles of the order of the rocking curve width, but needs three images, recorded at different angular positions of the analyser crystal.

Statistical methods for reconstructing the rocking curve include the multiple image radiography algorithm (Wernick et al., 2003), and the pixel-by-pixel Gaussian curve fitting algorithm (Huang et al., 2007). The object rocking curve and the reference rocking curve are retrieved from the pixel-by-pixel measurements of the intensity for several positions of the analyser (more than three), in the presence and in the absence of the object respectively. These methods require a series of images for different angular positions of the analyser, but they are stable with respect to noise.

Other approximations for solving Equation (4.10) include weak object approximation and linear transfer function approximation. The former (Nesterets et al., 2004) is based on the assumption that the modulus of the differences ($\Delta\mu$ and $\Delta\phi$) between the attenuation factor, $-k\int\beta\,dz$, and the phase factor, $-k\int\delta\,dz$, and their averages over (x, y), $\bar{\mu}$ and $\bar{\phi}$, respectively, are much smaller than one. Then the amplitude can be approximated as $\psi(x, y) \approx \bar{\psi}(1+\Delta\psi(x, y))$, where $\bar{\psi} = e^{i\bar{\phi}-\bar{\mu}}$, $\Delta\psi(x, y) = i\Delta\phi(x, y) - \Delta\mu(x, y)$. Linear transfer function approximation is based on the approximation of the transfer function $r(\theta)$ by a linear function in the Fourier space (Paganin et al., 2004b).

4.4 Grating-Based Method

A typical scheme of the X-ray grating interferometer is shown on Figure (4.6). The first grating is placed perpendicular to z, the direction of propagation of the incident wave. The lines of this grating are parallel to the y-axis (see Figure (4.6)), they do not absorb X-rays but change their phase. The incident plane X-ray wave is split by the phase grating in the diffracted beams, from

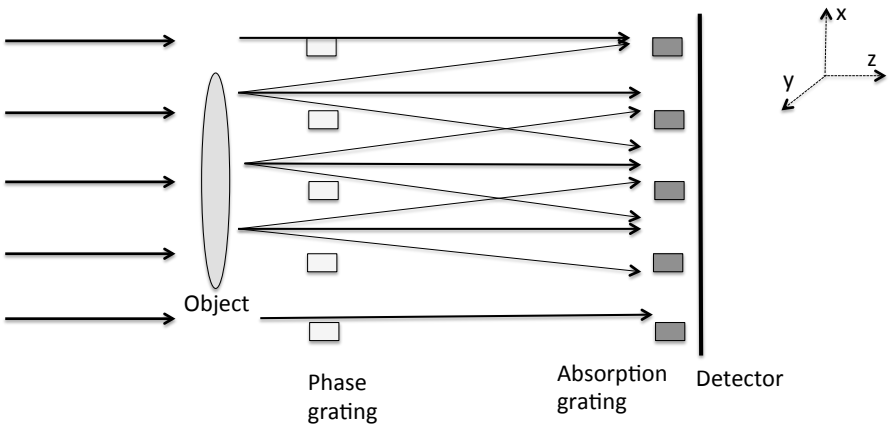

Figure 4.6: X-ray grating interferometer.

which two first-order diffraction beams make the most prominent contribution (Paganin, 2006). The divided waves interfere and create an interference pattern downstream from the phase grating in the planes perpendicular to the z-axis. As the wavelength $\lambda \approx 10^{-10}\text{m}$ of the incident wave is much smaller than the grating period $a \approx 10^{-6}\text{m}$ (Weitkamp et al., 2005), the angle between two first-order diffracted waves ($2 \arctan \frac{\lambda}{\sqrt{a^2-\lambda^2}}$ (Paganin, 2006, Ch. 3.2)) is of the order 10^{-4} radians. When the object is placed before the phase grating, the interference pattern changes.

The second grating with absorbing lines is placed downstream from the phase grating immediately before the detector. The absorbing grating has the same pitch and the same orientation as the phase grating. The distance between two gratings is chosen in such a way that it is an integer multiple of the Talbot distance (Cloetens et al., 1997a). The Talbot effect (Talbot, 1836) is the effect of self-imaging of the object with periodical structure (such as a grating, with the period a) illuminated with the monochromatic plane wave.

Self-imaging occurs periodically at distances which are integer multiple of the Talbot distance $T = \frac{2a^2}{\lambda}$, where λ is the wavelength. The intensity measured by the detector contains information about the phase gradient of the object (Weitkamp et al., 2005).

In some experiments the absorbing grating has a pitch different from the phase grating pitch, and the absorbing grating is not parallel to the phase grating. For example in the approach of (Weitkamp et al., 2005; Pfeiffer et al., 2006) the absorbing grating has a pitch which is 1/2 of the phase grating, since the interference patterns are observed at a fractional Talbot distance. Momose and collaborators (Momose et al., 2003) put the analyser grating at a slightly misaligned angle with respect to the phase grating, thus highlighting a Moire pattern.

The phase stepping method is applied for phase retrieval (Weitkamp et al., 2005; Pfeiffer et al., 2006). The phase grating is scanned along the x-direction by the shift distance x_g, which is smaller than the distance between adjacent grating lines. The intensity $I(x, y, x_g)$ is detected for each pixel with coordinates (x, y). The function $I(x, y, x_g)$ is an oscillating function depending on x_g and has period a. The relation between the phase $\Phi(x, y)$, i.e. the argument of the oscillating function $I(x, y, x_g)$, and the derivative of the phase $\phi(x, y)$ with respect to x is expressed by the formula (Born and Wolf, 2005, Ch. 8.6), (Weitkamp et al., 2005)

$$\Phi = \frac{\lambda d}{a} \frac{\partial \phi}{\partial x}, \tag{4.12}$$

where d is the distance between the two gratings. The phase ϕ is found by the integrating Φ with respect to x. At least three steps are needed for extraction of ϕ (Weitkamp et al., 2005). The phase stepping method can be used for the dark-field imaging (i.e. imaging based on scattered X-rays) as well (Pfeiffer et al., 2008).

4.5 Propagation-Based Phase-Retrieval Methods

A schematic representation of propagation-based (also called in-line) phase-contrast imaging methods is shown in the Figure 4.7. A plane wave, propagating in the z-direction, passes through an object. As a result of the interaction with

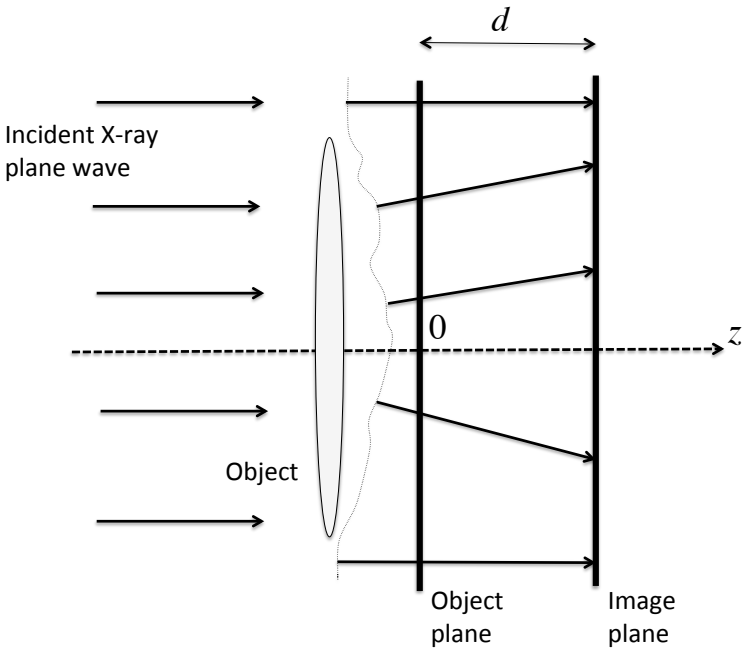

Figure 4.7: Propagation-based phase-contrast imaging setup.

the object, the direction of propagation of X-rays changes. We assume that the angular deviations from the original direction of propagation are small, thus the paraxial approximation can be is used (Born and Wolf, 2005, Ch. 4.9). Propagation-based phase-retrieval methods require intensity measurements over one or more planes, placed parallel to the object plane downstream of the object. Some propagation-based phase-contrast reconstruction techniques are presented here.

The multi-plane intensity measurements method is based on the weak object approximation. The detector is placed at different distances $\mathtt{d}_m$ from the object plane (m is the index of the corresponding image plane $z = \mathtt{d}_m$), and the phase is retrieved from the intensities measured over these planes (Guigay et al., 2007; Gureyev et al., 2004b; Cloetens et al., 1999). Another approach is the multiple-energy approach, when the intensity measurements are performed in one image plane, but at different X-ray energies (Gureyev et al., 2001, 2004b).

The iterative algorithm for the phase retrieval suggested by Gerchberg and

Saxton (Gerchberg and Saxton, 1972) uses the intensity distributions on the planes $z = 0$ and $z = d$, where the distance between the object and image planes d satisfies the condition $N_F << 1$, where $N_F = \frac{b^2}{\lambda d}$ is the Fresnel number, b is the aperture diameter (see e.g. (Goodman, 2005, Ch. 4.5), (Paganin, 2006, Ch. 1.5)). Then the connection between the wave fields $\psi(x', y', 0)$ and $\psi(x, y, d)$ can be described by the Fraunhofer diffraction integral (Goodman, 2005, Ch. 4.3), (Paganin, 2006, Ch. 1.5):

$$\psi(x, y, d) = -\frac{i}{\lambda d} e^{ikd} e^{\frac{ik}{2d}(x^2+y^2)} \int\int \psi(x', y', z = 0) e^{-\frac{ki}{d}(xx'+yy')} dx' dy', \quad (4.13)$$

which we rewrite as

$$\psi(x, y, d) = -\frac{i}{\lambda d} e^{ikd} e^{\frac{ik}{2d}(x^2+y^2)} \left(\mathbf{F_2}[\psi]\right)\left(\xi = \frac{x}{\lambda d}, \eta = \frac{y}{\lambda d}, z = 0\right). \quad (4.14)$$

Here, we used the Fourier transform (Equation (A.3)) in two-dimensional space $\mathbb{R}^2$, where (ξ, η) are the coordinates in the Fourier space conjugated to (x, y).

We can see from Equation (4.14) that the modulus of the wave complex amplitude $|\psi(x, y, d)|$ in the plane $z = d$ is proportional to the modulus of the Fourier transform of the function $\psi(x', y', 0)$: $|\mathbf{F_2}[\psi(x', y', 0)]|$. The Gerchberg and Saxton approach is as follows. Take a first approximation of the phase distribution in the object plane, $\phi_{est}(x', y', 0)$ (it could be a set of random numbers from the interval from $-\pi$ to π, or any constant from the same interval). Then the first iteration of the wave function is $\psi_1(x', y', 0) = |\psi(x', y', 0)| e^{i\phi_{est}}$. In the Fourier transform of this function $\mathbf{F_2}[\psi_1(x', y', 0)]$ we replace its absolute value $|\mathbf{F_2}[\psi_1(x', y', 0)]|$ by a function proportional to $|\psi(x, y, d)|$. Now we take the inverse Fourier transform of this function (which is proportional to $|\psi(x, y, d)|$) and get a wave function, which amplitude we replace by $|\psi(x', y', 0)|$, i.e. the square root of the intensity distribution on the plane $z = 0$. The resulting function, $\psi_2(x', y', 0)$ is the second iteration of the wave function in this iteration process. The same steps are applied to $\psi_2(x', y', 0)$, resulting the third iteration of the wave function $\psi_3(x', y', 0)$. The solution exists if this sequence of functions converges. Then the phase is defined up to a constant. It is not unique if both intensity distributions in the object plane and in the image planes are centrally symmetric (Gerchberg and Saxton, 1972). Generalisations of the Gerchberg and Saxton method can be found in (Fienup, 1993).

The method based on the Transport of Intensity Equation (TIE) allows one to find the distribution of the phase $\phi(x, y, 0)$ from the measurements of the intensities $I(x, y, 0)$ and $I(x, y, d)$ on the object plane and the image ($z = d$) planes, respectively. As the TIE plays the central role in this thesis, Chapter 5 is devoted to an analysis of it, including its derivation, solution methods and conditions for validity.

4.6 Summary

In this Chapter we introduced several phase-retrieval methods applicable to X-ray interferometry, analyser-based imaging, grating-based imaging and propagation-based imaging. The next Chapter is devoted to one of the propagation-based phase-retrieval methods, which is based on the Transport of Intensity Equation (TIE).

Chapter 5

The Transport of Intensity Equation

5.1 Introduction

Conventional X-ray detectors only register the intensity distribution. In-line phase-contrast imaging allows one to register the phase-contrast effects caused by the real part of the refractive index, and consequently, to reconstruct the distribution of the real part of refractive index in the sample. One of the in-line phase-retrieval techniques (see Chapter 4 for other phase-retrieval methods) is based on the Transport of Intensity Equation (TIE) (Teague, 1983):

$$\nabla_{\perp} \cdot (I(x, y, 0) \nabla_{\perp} \phi(x, y, 0)) = -k \partial_z I(x, y, 0), \tag{5.1}$$

where $I(x, y, 0)$ is the intensity distribution and $\phi(x, y, 0)$ is the phase distributions over the object plane $z = 0$. The distribution of the derivative of the intensity with respect to z over the object plane is denoted by $\partial_z I(x, y, 0)$. In this Chapter we discuss some derivations of the TIE (Section 5.2), and solution methods (Section 5.3).

5.2 Derivations of the Transport of Intensity Equation

5.2.1 Derivation of the TIE from the Fresnel diffraction integral

The connection between the wave functions $\psi_\omega(x, y, 0)$ and $\psi_\omega(x, y, d)$ is given by the diffraction operator (Paganin, 2006, Ch. 1.4):

$$\psi_\omega(x, y, d) = e^{ikd}\mathbf{F_2}^{-1}\left[e^{-\lambda\pi id(\xi^2+\eta^2)}\mathbf{F_2}[\psi_\omega(x, y, 0)]\right]. \tag{5.2}$$

If the distance d between the object and image planes is small enough, i.e.

$$\left|\lambda\pi d(\xi^2 + \eta^2)\right| << 1, \tag{5.3}$$

the propagator $e^{-\lambda\pi id(\xi^2+\eta^2)}$ can by approximated by the expression $1 - \lambda\pi id(\xi^2 + \eta^2) = 1 - \frac{2\pi^2}{k}id(\xi^2 + \eta^2)$. Substituting this approximation into Equation (5.2) we get

$$\begin{aligned}\psi_\omega(x, y, d) &= e^{ikd}\mathbf{F_2}^{-1}\left[\left(1 - \frac{2\pi^2}{k}id(\xi^2 + \eta^2)\right)\mathbf{F_2}[\psi_\omega(x, y, 0)]\right] \\ &= e^{ikd}\mathbf{F_2}^{-1}\left[(\mathbf{F_2}[\psi_\omega(x, y, 0)]] + \frac{id}{2k}\mathbf{F_2}\nabla^2\psi_\omega(x, y, 0)\right] \\ &= e^{ikd}\left(1 + \frac{id\nabla_\perp^2}{2k}\right)\psi_\omega(x, y, 0). \end{aligned}\tag{5.4}$$

Taking the square of the modulus of the left and the right hand side of this equation and ignoring the term containing d^2 (which is negligible when $|\nabla_\perp^2\psi_\omega| << \frac{2k}{d}|\psi_\omega|$), we obtain the following expression for the intensity distribution over the image plane:

$$\begin{aligned} I(x, y, d) &= \left(e^{ikd}\left(1 + \frac{id\nabla_\perp^2}{2k}\right)\psi_\omega(x, y, 0)\right)\left(e^{-ikd}\left(1 - \frac{id\nabla_\perp^2}{2k}\right)\psi_\omega^*(x, y, 0)\right) \\ &\approx I(x, y, 0) + \frac{id}{2k}(\psi_\omega^*\nabla_\perp^2\psi_\omega - \psi_\omega\nabla_\perp^2\psi_\omega^*) \\ &= I(x, y, 0) - \frac{d}{k}Im(\psi_\omega^* \cdot \nabla_\perp^2\psi_\omega). \end{aligned}\tag{5.5}$$

On the other hand

$$\begin{aligned} I(x, y, 0)\nabla_\perp\phi = \psi_\omega\psi_\omega^*\nabla_\perp\phi = Im(\psi_\omega\psi_\omega^*\nabla_\perp \ln\psi_\omega) = \\ Im(\psi_\omega\psi_\omega^*\frac{\nabla_\perp\psi_\omega}{\psi_\omega}) = Im(\psi_\omega^*\nabla_\perp\psi_\omega). \end{aligned}\tag{5.6}$$

Applying the divergence operator to the last equation yields

$$\nabla_\perp \cdot (I(x,y,0)\nabla_\perp \phi(x,y,0)) = \mathrm{Im}(\nabla_\perp \cdot (\psi_\omega^* \nabla_\perp \psi_\omega)) = \mathrm{Im}\,(\psi_\omega^* \cdot \nabla_\perp^2 \psi_\omega), \quad (5.7)$$

because $\mathrm{Im}(\nabla_\perp \psi_\omega^* \cdot \nabla_\perp \psi_\omega) = 0$. Combining Equations (5.5) and (5.7), we get

$$I(x,y,d) \approx I(x,y,0) - \frac{d}{k}\left(\nabla_\perp \cdot (I(x,y,0)\nabla_\perp \phi(x,y,0))\right). \quad (5.8)$$

Equation (5.8) is a finite-difference form of the TIE (Equation (5.1)) and allows us to reconstruct the phase $\phi(x,y,0)$ from the intensity distributions in the image and object planes.

5.2.2 Teague's Method

The Transport of Intensity Equation can also be derived from the homogeneous paraxial Helmholtz equation Equation (2.24), which describes the propagation of a wave in vacuum (n=1) parallel to the z-axis

$$(2ik\frac{\partial}{\partial z} + \nabla_\perp^2)\psi(x,y,z) = 0. \quad (5.9)$$

We substitute the expression for the complex amplitude of the wave function $\psi(x,y,z) = \sqrt{I(x,y,z)}e^{i\phi(x,y,z)}$ into the paraxial equation (5.9) and take the imaginary part of the resulting equation to obtain the TIE:

$$\nabla_\perp \cdot (I(x,y,z)\nabla_\perp \phi(x,y,z)) = -k\partial_z I(x,y,z). \quad (5.10)$$

This method to derive the TIE was suggested by Teague (Teague, 1983).

5.2.3 Guigay's method

The TIE can also be derived from the following equation, which describes the Fourier transform (Equation (A.3)) of the intensity distribution in the near-field regime (Guigay, 1977; Guigay et al., 2007):

$$\mathbf{F}_2[I(x,y,d)](\xi,\eta) = \mathbf{F}_2\left[T\left(x - \frac{\lambda d\xi}{2}, y - \frac{\lambda d\eta}{2}, 0\right) T^*\left(x + \frac{\lambda d\xi}{2}, y + \frac{\lambda d\eta}{2}, 0\right)\right](\xi,\eta). \quad (5.11)$$

The transmittance function $T(x,y,0) = a(x,y,0)e^{i\phi(x,y,0)}$ describes the interaction of the X-ray wave with an object, where $a(x,y,0)$ is the attenuation term and $\phi(x,y,0)$ is the phase shift induced by the object.

If $\lambda d\sqrt{\xi^2+\eta^2} << 1$, the transmittance function $T(x,y,0)$ and its conjugate $T^*(x,y,0)$ can be approximated by their first-order Taylor expansion

$$T\left(x-\frac{\lambda d\xi}{2}, y-\frac{\lambda d\eta}{2}, 0\right) = T(x,y,0) - \frac{\lambda d}{2}\langle\xi,\eta\rangle\cdot\nabla_\perp T(x,y,0), \quad (5.12)$$

$$T^*\left(x+\frac{\lambda d\xi}{2}, y+\frac{\lambda d\eta}{2}, 0\right) = T^*(x,y,0) + \frac{\lambda d}{2}\langle\xi,\eta\rangle\cdot\nabla_\perp T^*(x,y,0). \quad (5.13)$$

We substitute Equations (5.12), (5.13) and $T(x,y,0) = a(x,y,0)e^{i\phi(x,y,0)}$ into Equation (5.11) and neglect the term containing d^2 to obtain

$$\mathbf{F_2}[I(x,y,d)](\xi,\eta) = \mathbf{F_2}[I(x,y,0)](\xi,\eta) - i\lambda d\langle\xi,\eta\rangle\cdot\mathbf{F_2}\left[I(x,y,0)\nabla_\perp\phi(x,y,0)\right](\xi,\eta), \quad (5.14)$$

where $I(x,y,0) = |a(x,y,0)|^2$ is the intensity distribution over the object plane. The Fourier transform of the derivative (Equation A.9) (Vladimirov, 1971, Ch. 2.9), (Kak and Slaney, 1999, Ch.2.2) yields

$$2\pi i\langle\xi,\eta\rangle\cdot\mathbf{F_2}\left[I(x,y,0)\nabla\phi(x,y,0)\right](\xi,\eta) = \mathbf{F_2}\left[\nabla\cdot\left(I(x,y,0)\nabla\phi(x,y,0)\right)\right](\xi,\eta). \quad (5.15)$$

Combining this formula with Equation (5.14) and applying the inverse Fourier transform (Equation (A.4)), we obtain

$$I(x,y,d) = I(x,y,0) - \frac{d\lambda}{2\pi}\nabla\cdot\left(I(x,y,0)\nabla\phi(x,y,0)\right). \quad (5.16)$$

5.3 Methods for Solving the TIE

There are different methods for solving the TIE. One of them is based on the series expansion of the function $\phi(x,y)$ in the circle lying in the object plane as a series of Zernike's polynomials (Zernike, 1934). Then the partial differential equation (5.1) leads to a system of linear equations. This method was first suggested for the case of uniform intensity $I(x,y,z=\text{const})$, i.e. for the case when the intensity is constant in a circular domain, and is zero outside of this domain (Gureyev et al., 1995). In the next paper (Gureyev and Nugent, 1996) the solution method was presented for the case when the intensity $I(x,y,z=\text{const})$ is a continuous function, which is positive inside of the circular domain and zero outside of the domain.

Teague (Teague, 1983) proposed to solve the TIE using an auxiliary function $f(x, y, z = \text{const})$, such that

$$\nabla_\perp f(x, y, z = \text{const}) = I(x, y, z = \text{const}) \nabla_\perp \phi(x, y, z = \text{const}). \tag{5.17}$$

This method and its validity conditions are discussed in Chapter 8. Teague's method was developed by (Paganin and Nugent, 1998b) and applied in several publications, e.g. (Yu et al., 2005; Bajt et al., 2000; Barty et al., 1998; Allman et al., 2000; Yu et al., 2011; Lade et al., 2005a,b; De Graef and Zhu, 2001; Petersen et al., 2007; Yu et al., 2010; Frank et al., 2010; Langer et al., 2008; McMahon et al., 2003), (Paganin, 2006, Ch. 4.5.2). The validity of Teague's assumption, i.e. the existence of the auxiliary function $f(x, y, z = \text{const})$ required in Equation (5.17), has been thoroughly examined in (Schmalz et al., 2010).

We mention also that the solution to the TIE can be used as a first estimation in an iterative procedure (for example for the Gerchberg – Saxton – Fienup's algorithm) (Gureyev, 2003). Then the iteration will be more likely to converge than if we would use a random function or a constant as a first estimate.

5.4 Summary

This Chapter presents derivations of, and techniques for solving, the TIE. Different derivations provide different perspectives on the nature of the TIE. Teague's technique for solving the TIE is of interest for our research. Chapter 8 investigates conditions needed for Teague's technique. We apply Teague's method to tomography in Chapter 9.

Chapter 6

X-ray Computed Tomography

6.1 History of the Development of Tomography

The problem of the reconstruction of an object from its projections for the two-dimensional case was solved theoretically first by Radon in 1917 (Radon, 1917). Another theoretical approach to reconstructing one plane of interest inside objects was developed in 1932 by Ziedses des Plantes (des Plantes, 1932). But the practical use of tomography only began in 1960s - 1970s when electronic devices and computer technology became available. In 1979 a Nobel Prize was awarded to Allan M Cormack and Godfrey N Hounsfield for their contributions toward the development of computed tomography.

The first devices were built on the basis of the method suggested by Ziedses des Plantes (Garrison et al., 1969). This method is called X-ray classical or conventional geometric tomography (Dobbins and Godfrey, 2003). The idea of the classical tomography is as follows An X-ray source and a detector move synchronously with respect to an object placed between them, the initial intensity is constant before the object, and the intensity measured at the detector is $I(\phi_k)$ for several projection angles $\phi_k < \pi$ (Figure 6.1). The corresponding intensities $I(\phi_k)$ are measured and an average of logarithms of the intensities is calculated: $1/n \sum_1^n \ln(I(\phi_k))$. The values of the average give a sharp image over the plane of interest (called a focal plane), other planes are blurred and give an intensity background.

It was shown (des Plantes, 1932) that it is possible to reconstruct not only the focal plane, but other parallel planes from one set of measurements (Dobbins and Godfrey, 2003). This procedure of generating some slice images from one

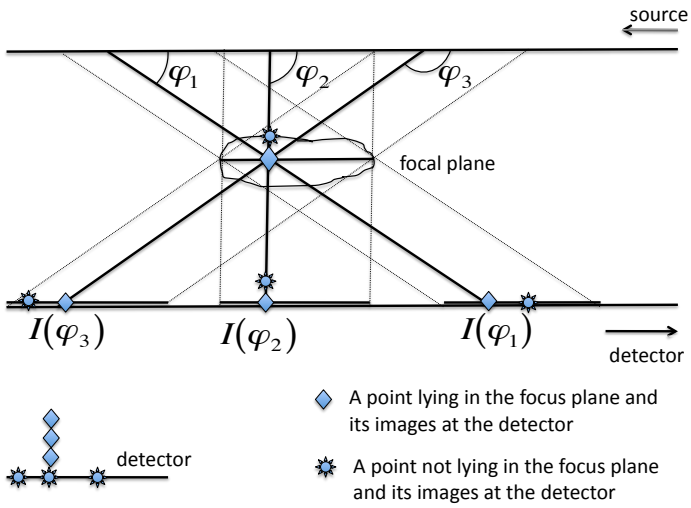

Figure 6.1: Images of a point in a focal plane of the object are on the same spot of the detector for every inclination angle. Images of a point not lying in the focal plane are distributed over the detector creating an intensity background (Dobbins and Godfrey, 2003)

.

set of projections is called tomosynthesis or photographic laminography (Miller et al., 1971; Helfen et al., 2005, 2007, 2009).

Both classical tomography and tomosynthesis suffer from the blurring of the object outside the plane of interest. Some approaches for removing blurring are described in (Dobbins and Godfrey, 2003).

6.2 Mathematical Basis for Computed Tomography (CT)

Let the function $f(x, y, z)$ describe a physical property of a three-dimensional object. We discuss in this section how to reconstruct $f(x, y, z)$ from its two-dimensional projections.

The projection is a function, defined on the plane perpendicular to the direction of propagation of X-rays, given by line integrals of the function $f(x, y, z)$ along the direction of propagation (Figure 6.2). For an exact definition

a coordinate system (x', y, z') is introduced, which is obtained by rotating the coordinate system (x, y, z) through the angle $\theta \in [0, \pi[$ about the axis 0y, so that

$$x' = x\cos\theta - z\sin\theta, \quad y = y, \quad z' = x\sin\theta + z\cos\theta. \tag{6.1}$$

The direction of illumination is along the z' axis, the angle θ determines an illumination angle. The projection (or X-ray transform) of the function $f(x, y, z)$ in the direction $0z'$ given by the unit vector $\langle \sin\theta, 0, \cos\theta \rangle$ in (x, y, z) coordinates, is defined (Natterer, 2001, Ch. II.1) as a line integral

$$(P_\theta f)(x', y) = \int_{-\infty}^{\infty} f(x', y, z')\, dz' \tag{6.2}$$

(Figure 6.2). A projection $P_\theta(f)$ maps a function $f(x, y, z)$ of the three-dimensional

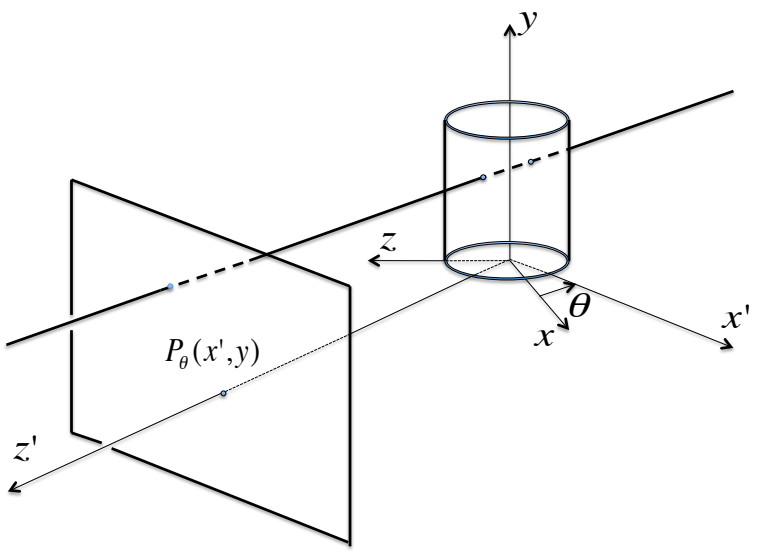

Figure 6.2: A projection $P_\theta(f)$ maps a function $f(x, y, z)$ of the three-dimensional space $\mathbb{R}^3$ into the set of its line integrals $(P_\theta f)(x', y)$.

space $\mathbb{R}^3$ into the set of its line integrals $(P_\theta f)(x', y)$, $(x', y) \in \mathbb{R}^2$. If the function $f(x, y, z)$ is known, the projections for every illumination angle $\theta \in [0, \pi[$ can

be found. This is an example of a direct problem. We consider the following inverse problem. Reconstruct the function $f(x, y, z)$ from the set of projections $(P_\theta f)(x', y), (x', y) \in \mathbb{R}^2$ corresponding to illumination angles θ. The reconstruction of the function from its line integrals is called computed tomography (CT) (Natterer, 2001, Ch. I). The solution to this inverse problem will be unique if all projections $(P_\theta f)(x', y)$, $(x', y) \in \mathbb{R}^2$, for every $\theta \in [0, \pi[$ are given.

The coordinates (ξ, η, ζ) (see Appendix A) and the coordinates (ξ', η', ζ') in the Fourier space are related by

$$\xi' = \xi \cos\theta - \zeta \sin\theta, \quad \eta = \eta, \quad \zeta' = \xi \sin\theta + \zeta \cos\theta. \tag{6.3}$$

(see Equation 6.1). From Equation (6.3) we get

$$\xi = \xi' \cos\theta + \zeta' \sin\theta, \quad \eta = \eta, \quad \zeta = -\xi' \sin\theta + \zeta' \cos\theta. \tag{6.4}$$

The slice theorem states that the two-dimensional Fourier transform (Equation (A.3)) of the projection coincides with the corresponding cross section of the three-dimensional Fourier transform (Equation (A.6)) of the function itself (Kak and Slaney, 1999, Ch. 3.2), (Natterer, 2001, Ch. 2.2):

$$\mathbf{F_2}[P_\theta f](\xi', \eta) = \mathbf{F_3}[f](\xi' \cos\theta, \eta, -\xi' \sin\theta). \tag{6.5}$$

Here the same plane is given first by $\zeta' = 0$ in the coordinate system (ξ', η, ζ') on the left-hand side of (Equation (6.5)), and then by equations $\xi = \xi' \cos\theta, \eta = \eta, \zeta = -\xi' \sin\theta$ on the right-hand side of (Equation (6.5)). This can be seen by substituting $\zeta' = 0$ into (Equation (6.4)).

If the projections of the function for each of the directions $0z'$, $\theta \in [0, \pi[$: $(P_\theta f)(x', y)$ are known, the three-dimensional function $f(x, y, z)$ can be reconstructed by the following steps (Kak and Slaney, 1999, Ch. 3.2):

1. For each illumination angle $\theta \in [0, \pi[$ (Figure (6.2)) calculate the two-dimensional Fourier transform of the projection $\mathbf{F_2}[P_\theta f](\xi', \eta)$, which is defined on the plane $\zeta' = 0$ in Fourier space (Figure (6.3)). These planes are perpendicular to directions $0\zeta'$ and pass through the origin. They fill three-dimensional Fourier space.

2. Applying the slice theorem (Equation (6.5)), find the values of the three-dimensional Fourier transform $\mathbf{F_3}[f](\xi, \eta, \zeta)$ of the function f at each point of the three-dimensional Fourier space.

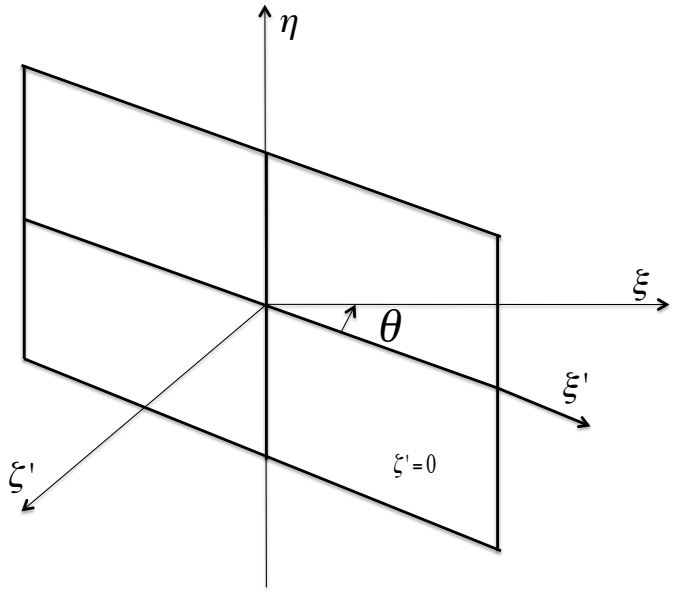

Figure 6.3: The two-dimensional Fourier transform of the projection $\mathbf{F_2}[P_\theta f](\xi', \eta)$ is defined on the plane $\zeta' = 0$ in three-dimensional Fourier space. The planes with $\theta \in [0, \pi[$ fill the three-dimensional Fourier space.

3. Calculate the three-dimensional function $f(x, y, z)$ taking the inverse Fourier transform (Equation (A.6)) and using (Equation (A.8)):

$$f(x, y, z) = \mathbf{F_3^{-1}} \left[\mathbf{F_2}[P_\theta f]\right]. \tag{6.6}$$

It is natural to describe a point in Fourier space by giving its coordinates (τ, η) in the plane $\zeta' = 0$ and an angle between the planes $\zeta = 0$ and $\zeta' = 0$. The corresponding coordinates in Fourier space are:

$$\xi = \tau \cos\theta, \quad \eta = \eta, \quad \zeta = -\tau \sin\theta, \tag{6.7}$$

where $\tau \in \mathbb{R}$ and $\theta \in [0, \pi[$. Then the formula (Equation (6.6)) becomes

$$f(x, y, z) = \int_0^\pi \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{2\pi i(\tau(x\cos\theta - z\sin\theta) + y\eta)} \hat{f}(\tau, \eta, \theta) |\tau| \, d\tau d\eta d\theta. \tag{6.8}$$

In Equation (6.8) $|\tau|$ gives a weight to $\hat{f}(\tau, \theta, \eta)$ in frequency (reciprocal or Fourier) space (Kak and Slaney, 1999, Ch. 3.3). Theoretically, the integration

with respect to τ in Equation (6.8) is from $-\infty$ to ∞, but practically $|\tau|$ is limited, because the resolution of the optical system is always limited. This method of reconstructing a three-dimensional function from a set of its projections is called filtered back projection (Kak and Slaney, 1999, Ch. 3.3).

Of course, it is not possible to know projections for all directions given by $\theta \in [0, \pi[$, only a finite number of the projections can be known in practice. Moreover, the data for any projection can be sampled only at a finite number of points in the plane. The sampling theorem (Natterer, 2001, Ch. III), (Kak and Slaney, 1999, Ch. 2.1), (Goodman, 2005, Ch. 2.4) answers the question as to how close to each other these sample points should be in order to reconstruct the 3D function uniquely.

6.3 Reconstruction of the Absorption Index β

We seek the reconstruction of the internal structure of an object on the basis of attenuation contrast. Measuring the difference of logarithms of intensities before (I_{in}) and after (I_θ) the object, we find the projection of the absorption index β (compare with Equation (3.4)): $(P_\theta\beta)(x', y) = \int_{-\infty}^{\infty} \beta(x', y, z')dz' = -\frac{1}{2k} \ln \frac{I_\theta(x', y)}{I_{in}}$ along the illumination directions z' which makes an angle θ with the z-axis (Figure 6.2). Combining the projections for different angles $\theta \in [0, \pi[$ and applying the slice theorem (Equation (6.6)) we reconstruct the 3D distribution of the absorption index in the object: $\beta(x, y, z) = \mathbf{F_3^{-1}}\,[\mathbf{F_2}[P_\theta\beta]]$. Thus we use the last formula for reconstruction of the object if the absorption indexes of materials containing in this object differ significantly:

$$\beta(x, y, z) = -\frac{1}{2k} \int_0^\pi \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{2\pi i(\tau(x\cos\theta - z\sin\theta) + y\eta)}\, \mathbf{F_2}\left[\ln\left(\frac{I_\theta}{I_{in}}\right)\right](\tau, \eta)|\tau|\, d\tau d\eta d\theta. \tag{6.9}$$

If the attenuation contrast between different components of the object is not significant, X-ray phase-contrast computed tomography may be applied.

6.4 Phase-Contrast X-Ray CT

It has been suggested that X-ray phase contrast can be utilised for improvement of the contrast in transmission images of samples consisting predominantly of light chemical elements (Snigirev et al., 1995; Momose et al., 1995; Wilkins et al., 1996).

Subsequently, phase-contrast X-ray CT (PCT) has been implemented in several forms, including crystal-based and grating-based X-ray interferometry PCT(Momose et al., 1995; Weitkamp et al., 2006; Momose et al., 2006), analyser-based PCT (Dilmanian et al., 2000; Pavlov et al., 2001) and in-line phase contrast PCT (Raven et al., 1996; Cloetens et al., 1997b; Barty et al., 2000; Mayo et al., 2003; Anastasio et al., 2004).

Chapter 9 discusses phase-contrast X-ray CT in detail. There we apply the imaging method developed in Chapter 8 to the 3D reconstruction of the refractive index.

Chapter 7

Derivation of a Green's Function for the Helmholtz Equation Using Generalised Functions[1]

7.1 Introduction

In this Chapter we consider the classical problem of finding the free-space Green's function for the Helmholtz equation (Chapter 2) subject to Sommerfeld radiation conditions, provide a rigorous derivation of the Green's functions for the Helmholtz equation, and then find the unique solution subject to the Sommerfeld conditions. This Chapter presents the result obtained by the candidate under the supervision of Timur Gureyev and Konstantin Pavlov. The scientific question was raised by the candidate herself, the method was suggested by Gerd Schmalz.

The Helmholtz equation can be obtained from Maxwell's equations (Section 2.2). The (homogeneous) Helmholtz equation describes the behavior of the spatial component of the wave function $\psi_\omega(\mathbf{r})e^{-i\omega t}$, corresponding to a monochromatic component of a wave field in free space:

$$(\nabla^2 + k^2)\psi_\omega(x, y, z) = 0. \tag{7.1}$$

Fundamental solutions of the Helmholtz equation by definition satisfy the equation

$$(\nabla^2 + k^2)G(x, y, z) = -4\pi\delta(x, y, z), \tag{7.2}$$

[1]This Chapter is based on a joint paper with Gerd Schmalz, Timur Gureyev and Konstantin Pavlov (Schmalz et al., 2010)

where $\delta(x, y, z)$ is the Dirac delta-function (Merzbacher, 1961, Ch. 11.3). Equation (7.2) corresponds to a single point scatterer or a point source located at the origin (the constant -4π is introduced by convention). If $G(\mathbf{r})$ is a solution to Equation (7.2), it is easy to see that $(\nabla^2_{\mathbf{r}} + k^2)G(\mathbf{r} - \mathbf{r_0}) = -4\pi\delta(\mathbf{r} - \mathbf{r_0})$; that is, $G(\mathbf{r} - \mathbf{r_0})$ is by definition the Green's function for the Helmholtz equation (7.1). Therefore we will use the terms "fundamental solution" and "Green's function" for $G(x, y, z)$ interchangeably.

To find a unique solution to Equation (7.2) we need to consider suitable boundary conditions (Sommerfeld, 1949, Anhang II). The usual boundary condition for Equation (7.2) is the Sommerfeld radiation condition:

$$\lim_{r\to\infty} r\left(\frac{\partial G(r)}{\partial r} - ikG(r)\right) = 0, \tag{7.3}$$

where $r = |\mathbf{r}|$. Equation ((7.3)) means that the fundamental solution must behave at infinity like an outgoing spherical wave, which is a physically meaningful choice given that we are considering only one point scatterer (or source) at the origin.

The Green's functions of the Helmholtz equation have been known for a long time (Merzbacher, 1961, Ch. 11.3), (Sommerfeld, 1949, Ch. 2), (Davydov, 1973, Ch. 14), (Born and Wolf, 2005, Ch. 13.1), (Morse and Feshbach, 1953, Ch. 7.2), (Jackson, 1998, Ch. 6.4), (Messiah, 1999, Appendix B. IV.11), (Paganin, 2006, Ch. 1.6), (Huang and Nevels, 1992). We will obtain a general formula for the Green's functions of the Helmholtz equation using Fourier transforms of generalised functions.

7.2 Generalised Functions

Generalised functions are associated with the name of Paul Dirac (for example, Dirac's delta-function). A mathematically rigorous theory of generalised functions was developed by Sobolev (Sobolev, 1936) and Schwartz (Schwartz, 1951) and was extended by many others (Vladimirov, 1979; Hörmander, 1958). Generalised functions have numerous applications in physics (Vladimirov, 1979, Ch. 2), (Reed and Simon, 1980, Ch. 5.3), (Vladimirov, 1971, Introduction). To give an idea what a generalised function is, it is instructive to take a different view of classical functions. A classical function f can be identified by the values

of all possible integrals of the form $\int f\varphi \, dx$, where φ is an arbitrary smooth function (such that the integral exists). In this sense f can be interpreted as a linear functional on the space of test functions φ:

$$\langle f, \varphi \rangle = \int f\varphi \, dx. \tag{7.4}$$

That is, f linearly maps the space of all test functions into the space of real or complex numbers. As a natural extension of this interpretation of conventional functions, a generalised function F is defined as a linear continuous functional on the space of test functions. Note that not every linear functional defined on the space of smooth test functions can be represented in the form of Equation (7.4); that is, not every such functional F can be identified with a conventional function f. A well-known example of a generalised function that cannot be identified with a conventional one is the Dirac δ-function, defined as $\langle \delta, \varphi \rangle = \varphi(0)$. Therefore, the integral notation in Equation (7.4) is only symbolic when used with generalised functions. Generalised functions can always be added together, subtracted, multiplied by a number or differentiated any number of times. However, multiplication of generalised functions is not always a "safe" operation, for example, the square of the Dirac δ-function is not defined (Vladimirov, 1979, Ch. 1), (Reed and Simon, 1980, Ch. 1.4), (Vladimirov, 1971, Ch. 2).

There are different spaces of generalised functions that depend on the choice of space of the test functions. In this thesis we consider the space $\mathcal{S}$ of test functions consisting of all infinitely differentiable functions φ, which absolute values decrease when $|x| \to \infty$ faster than any power of $|x|^{-1}$, and so do all their derivatives (Vladimirov, 1979, Ch. 1), (Reed and Simon, 1980, Ch. 5.3), (Vladimirov, 1971, Ch. 2.5). Generalised functions (or tempered distributions), defined as continuous linear functionals on the space $\mathcal{S}$, constitute the space $\mathcal{S}'$ called the Schwartz space. The continuity of a functional F in this context means that if a sequence φ_n of test functions from $\mathcal{S}$, together with all the derivatives $\partial^\alpha \varphi_n$ and the products $x^m \partial^\alpha \varphi_n$ for any fixed m,uniformly converges to some limit function as $n \to \infty$, then the sequence $\langle F, \varphi_n \rangle$ must also converge to a limit as $n \to \infty$. The topology of the Schwartz space $\mathcal{S}'$ is defined via so-called "weak" convergence; that is, by definition $F_n \to F (n \to \infty)$ if and only if $\langle F_n, \varphi \rangle \to \langle F, \varphi \rangle \, (n \to \infty)$ for every φ from $\mathcal{S}$ (Reed and Simon,

1980). In particular, the Fourier transform, defined as $\langle \mathcal{F}[F], \varphi \rangle = \langle F, \mathcal{F}[\varphi] \rangle$, is an invertible linear continuous operator in the space $\mathcal{S}'$ (Vladimirov, 1979, Ch. 1), (Reed and Simon, 1980, Ch. 5.3), (Vladimirov, 1971, Ch. 2).

In the following we first convert Equation (7.2) to Fourier space and solve the resulting equation in terms of generalised functions (Section 7.3). Then we obtain the general solution for spherically symmetric boundary conditions, convert the result back to real space, and find the unique solution corresponding to the Sommerfeld boundary condition (Section 7.4). In Section 7.5 we consider the standard "$i\varepsilon$-technique" often used in the physics literature in similar contexts and compare it with our method. Rigorous proofs for some more technical mathematical statements can be found in the Appendix.

7.3 General Solution for Green's Functions for the Helmholtz Equation in Fourier Space

We express the Green's function $G(x, y, z)$ and the Dirac function $\delta(x, y, z)$ considered as tempered distributions (see chapter 7.2), (Vladimirov, 1971, Ch. 2.8) via their three-dimensional Fourier transforms:

$$G(x, y, z) = \iiint e^{2\pi i(\xi x + \eta y + \zeta z)} \, \hat{G}(\xi, \eta, \zeta) \, d\xi d\eta d\zeta, \tag{7.5}$$

$$\delta(x, y, z) = \iiint e^{2\pi i(\xi x + \eta y + \zeta z)} \, d\xi d\eta d\zeta, \tag{7.6}$$

where

$$\hat{G}(\xi, \eta, \zeta) = \iiint e^{-2\pi i(\xi x + \eta y + \zeta z)} \, G(x, y, z) \, dx dy dz \tag{7.7}$$

is the Fourier transform (Equation (A.5)) of $G(x, y, z)$. The integral notation for the Fourier transform of the tempered distributions in Equations (7.5) and (7.6) is symbolic. We treat all the equations in this Chapter as equations for generalised functions.

After substituting Equations (7.5) and (7.6) into Equation (7.2) and introducing $\rho = k/2\pi$, we obtain

$$\iiint \left[(\rho^2 - \xi^2 - \eta^2 - \zeta^2) \hat{G}(\xi, \eta, \zeta) + 1/\pi \right] e^{2\pi i(\xi x + \eta y + \zeta z)} \, d\xi d\eta d\zeta = 0, \tag{7.8}$$

which represents an inverse Fourier transform (Equation (A.6)) of the expression in brackets. Because the inverse Fourier transform, being invertible (Equation (A.8)), has no null space, the expression in brackets must vanish. In other words, we need to find all solutions of the equation:

$$(\xi^2 + \eta^2 + \zeta^2 - \rho^2)\hat{G}(\xi, \eta, \zeta) = 1/\pi. \tag{7.9}$$

Solving Equation (7.9) effectively requires us to divide a constant by a polynomial, which is a nontrivial problem, given that polynomials may have real zeros (as is the case with Equation (7.9)). As a consequence, the result of such a division may have non-integrable singularities, and may not be defined as a conventional function. Therefore, it is often necessary to use the formalism of generalised functions when working with problems of this type. Hörmander (Hörmander, 1958) showed that the problem of division of a constant by a polynomial always has a solution in the class of tempered distributions (such a solution is typically not unique). This fundamental result can be compared to the well-known theorem stating that any algebraic Nth order equation with complex coefficients has N roots in the complex plane. Just as the extension of the real axis to the complex plane is necessary to ensure that any algebraic equation has roots, so it is also necessary to extend the space of conventional functions to that of generalised functions (distributions) to ensure that any partial differential equation with constant coefficients has a fundamental solution (and, consequently, has a solution for an arbitrary function on the right-hand side of that differential equation).

We know (Lang, 1987, Ch. XIII) that any solution to a linear inhomogeneous equation such as Equation (7.9) can be represented as a sum of a particular solution and a solution to the corresponding homogeneous equation, that is,

$$\hat{G} = \hat{G}_{\text{part}} + \hat{G}_{\text{hom}}. \tag{7.10}$$

Here $\hat{G}_{\text{part}}$ is a particular solution of Equation (7.9) and $\hat{G}_{\text{hom}}$ is a solution of the corresponding homogeneous equation; that is, it corresponds to the case with no sources or scatterers

$$(\xi^2 + \eta^2 + \zeta^2 - \rho^2)\hat{G}_{\text{hom}} = 0. \tag{7.11}$$

As a particular solution we can take, for example, the principal value of the fraction

$$\hat{G}_{\text{part}} = \text{v.p.}\frac{1}{\pi(\xi^2+\eta^2+\zeta^2-\rho^2)}, \tag{7.12}$$

which is a generalised function defined as:

$$\langle \hat{G}_{\text{part}}, \varphi\rangle = \text{v.p.}\iiint \frac{1}{\pi(\xi^2+\eta^2+\zeta^2-\rho^2)}\varphi\, d\xi d\eta d\zeta \tag{7.13a}$$

$$= \lim_{\varepsilon\to 0}\iiint_{|\xi^2+\eta^2+\zeta^2-\rho^2|>\varepsilon} \frac{1}{\pi(\xi^2+\eta^2+\zeta^2-\rho^2)}\varphi\, d\xi d\eta d\zeta, \tag{7.13b}$$

where v.p. denotes the principal value integral in the sense of Cauchy (Shabat, 1992, Ch. 7.26).

Equation (7.12) is not a unique solution of Equation (7.9) because the homogeneous Equation (7.11) has nontrivial solutions in the class of generalised functions. This choice of a particular solution of Equation (7.9) seems the most simple and natural mathematically. Although this choice might not appear equally natural from the point of view of the relevant physics, we emphasise that this choice is inconsequential in our approach. Any solution of Equation (7.9) can be obtained from a single arbitrary (partial) solution by adding an appropriate solution of the homogeneous equation (7.11). (See Section 7.5 regarding other possible choices of the initial partial solution.)

It is clear that solutions of Equation (7.11) may have nonzero values only on the (Ewald) sphere S, $S = \{(\xi,\eta,\zeta)\in\mathbb{R}^3\colon \xi^2+\eta^2+\zeta^2=\rho^2\}$. We may divide both sides of Equation (7.11) by $\xi^2+\eta^2+\zeta^2-\rho^2$ if this expression is different from 0. Then we obtain $\hat{G}_{\text{hom}} = 0$ at all the points except for the points on the sphere.

To solve Equation (7.11), we introduce the generalised function $\mu\delta_S$ acting on the test functions φ according to the rule:

$$\langle \mu\delta_S, \varphi\rangle = \iint_S \mu(\xi,\eta,\zeta)\varphi(\xi,\eta,\zeta)\, dS. \tag{7.14}$$

Here μ is a generalised function on the sphere S in the Fourier space. The surface integral in this definition is only symbolic. The generalised function $\mu\delta_S$ is a generalisation of the Dirac function and is called a simple layer over S with density μ (see Ref. (Vladimirov, 1971, p.75). In other words, the simple

layer assigns to any test function φ a numerical value that depends only on the restriction of φ to the sphere S.

We claim that any solution $\hat{G}_{\text{hom}}$ of Equation (7.11) can be expressed as a simple layer $\mu\delta_S$. In Appendix (Section B) we provide a rigorous proof of this statement. Consequently, we can give the expression for all the Green's functions of the Helmholtz equation in Fourier space:

$$\hat{G} = \nu.p.\frac{1}{\pi(\xi^2+\eta^2+\zeta^2-\rho^2)} + \mu\delta_S. \tag{7.15}$$

Equation (7.15) is a general expression for all solutions of Equation (7.9). At this stage, we do not specify any boundary conditions for the Green's function. The first term on the right-hand side of Equation (7.15) describes the effect of a point source/scatterer located at the origin. The second term contains an arbitrary weight function $\mu(\xi,\eta,\zeta)$, which can be "tuned" to satisfy the appropriate boundary conditions for the Green's function.

7.4 Green's Function for the Helmholtz Equation Subject to the Sommerfeld Boundary Conditions

To obtain a solution of the Helmholtz equation subject to the Sommerfeld radiation condition we use a spherically symmetric ansatz. Searching for a spherically symmetric solution is justified by the symmetry of Equations (7.2) and (7.3). Because the solution can be shown to be unique (Vladimirov, 1971, Ch. 5), finding a spherically symmetric solution will prove that the solution of Equation (7.2) with boundary condition Equation (7.3) is spherically symmetric.

Next we note that the Fourier transform of a function is spherically symmetric if and only if the function is spherically symmetric. This property of the Fourier transform is well known (see for example, Ref. (Kak and Slaney, 1999, Ch. 2.2) or, for a proof, see Appendix (Section C)).

It is intuitively clear that any spherically symmetric simple layer over a sphere must have a constant density; that is, in this case

$$\langle \mu\delta_S, \varphi\rangle = C\iint_S \varphi(\xi,\eta,\zeta)\,dS \tag{7.16}$$

for some constant C. Indeed, if a classical function is defined on a sphere and does not change under any rotations of the sphere, its values at any two points

on the sphere must be the same. A proof of this fact for generalised functions is nontrivial and can be found in Appendix (Section D).

An arbitrary solution of Equation (7.9) with the spherically symmetric ansatz can now be written in accordance with Equation (7.15) as

$$\hat{G} = \nu.p.\frac{1}{\pi(\xi^2+\eta^2+\zeta^2-\rho^2)} + C\delta_S, \tag{7.17}$$

where C is some constant. Equation ((7.17)) describes all fundamental solutions of the Helmholtz equation satisfying spherically symmetric boundary conditions.

We can now find the Green's function by evaluating the inverse Fourier transform of Equation (7.17). First we find the inverse Fourier transform of Equation (7.12):

$$G_{\text{part}}(x,y,z) = \frac{1}{\pi}\,\nu.p.\iiint \frac{e^{2\pi i(\xi x+\eta y+\zeta z)}}{\xi^2+\eta^2+\zeta^2-\rho^2}\,d\xi d\eta d\zeta. \tag{7.18}$$

To calculate the integral in Equation (7.18) we rewrite it in spherical coordinates (R,ϕ,θ) in Fourier space. As the integral 7.18 is rotation invariant, the value of this integral will not change if we rotate the coordinate system. In order to simplify the calculations, we choose the coordinate system so that $\mathbf{r} = (0,0,r)$. Then

$$G_{\text{part}}(r) = \frac{1}{\pi}\,\nu.p.\int_0^{\pi}\int_0^{2\pi}\int_0^{\infty}\frac{e^{2\pi iRr\cos\theta}}{R^2-\rho^2}R^2\sin\theta\,dR d\phi d\theta =$$

$$\frac{1}{\pi ir}\,\nu.p.\int_0^{\infty}\frac{2i\sin(2\pi Rr)}{R^2-\rho^2}R\,dR = -\frac{1}{2\pi^2 r}\frac{d}{dr}\,\nu.p.\int_0^{\infty}\frac{2\cos(2\pi Rr)}{R^2-\rho^2}\,dR. \tag{7.19}$$

Since $\cos(2\pi Rr)$ is an even function, and $\sin(2\pi Rr)$ is an odd function, we can rewrite the last expression as

$$G_{\text{part}}(r) = -\frac{1}{2\pi^2 r}\frac{d}{dr}\,\nu.p.\int_{-\infty}^{\infty}\frac{e^{2\pi iRr}}{R^2-\rho^2}\,dR. \tag{7.20}$$

The principle value of the integral is

$$\nu.p.\int_{-\infty}^{\infty}\frac{e^{2\pi iRr}}{(R-\rho)(R+\rho)}dR = -\frac{\pi\sin(2\pi r\rho)}{\rho}, \tag{7.21}$$

which yields

$$G_{\text{part}}(r) = \frac{\cos(2\pi\rho r)}{r}. \tag{7.22}$$

This particular solution corresponds to the standing wave $\frac{\cos(kr)}{r}$.

We next evaluate the inverse Fourier transform of the generalised function $C\delta_S$. We introduce spherical coordinates in the expression

$$G_{hom} = C\iiint e^{2\pi i(\xi x+\eta y+\zeta z)}\,\delta_S\,d\xi d\eta d\zeta, \tag{7.23a}$$

and obtain

$$G_{hom}(r) = C\int_0^{\pi}\int_0^{2\pi} e^{2\pi i\rho r\cos\theta}\,\rho^2\sin\theta\,d\phi d\theta \tag{7.23b}$$

$$= -2\pi C\rho^2\int_0^{\pi} e^{2\pi i\rho r\cos\theta}\;d(\cos\theta) \tag{7.23c}$$

$$= 2C\rho\,\frac{\sin(2\pi\rho r)}{r}. \tag{7.23d}$$

As we denoted $\rho = k/2\pi$, we rewrite

$$G_{hom}(r) = \frac{Ck}{\pi}\,\frac{\sin(kr)}{r} \tag{7.24}$$

We can combine $G_{part}(x,y,z)$ and $G_{hom}(x,y,z)$ to find a constant C such that their sum satisfies the Sommerfeld radiation condition Equation (7.3). It is easy to see that for $C = \pi i/k$, we obtain

$$G(r) = G_{part}(r) + G_{hom}(r) = \frac{\cos(kr)}{r} + \frac{i\sin(kr)}{r} = \frac{e^{ikr}}{r}, \tag{7.25}$$

which is the well-known fundamental solution to the Helmholtz equation satisfying the radiation condition.

7.5 Comparison with the iε-Technique

The iε-technique of the derivation of the Green's function for the Helmholtz equation subject to the Sommerfeld conditions is widely used (Merzbacher, 1961, Ch. 11.3), (Davydov, 1973, Ch. 14), (Paganin, 2006, Ch. 2.3). It differs from our method starting after Equation (7.9). The solution of Equation (7.9) is written as

$$\hat{G}_{\pm} = \frac{1}{\pi}\,\frac{1}{\xi^2+\eta^2+\zeta^2-k^2\pm i0}, \tag{7.26}$$

which is defined as

$$G_{\pm}(x,y,z) = \frac{1}{\pi}\lim_{\varepsilon\to 0}\iiint \frac{e^{2\pi i(\xi x+\eta y+\zeta z)}}{\xi^2+\eta^2+\zeta^2-\rho^2\pm i\varepsilon}\,d\xi d\eta d\zeta. \tag{7.27}$$

Adding or subtracting a small imaginary value $\pm i\varepsilon$ in the denominator allows us to remove the singularities (poles) from the path of integration. It can be shown that the limits at $\pm\varepsilon \to 0$ exist and define tempered distributions (which are denoted by Equation (7.26)).

The inverse Fourier integral of Equation (7.12) gives a standing wave $\frac{\cos(kr)}{r}$; Equation (7.27) gives the Green's function for the outgoing wave $\frac{e^{ikr}}{r}$ for $-i\varepsilon$, and the incoming wave $\frac{e^{-ikr}}{r}$ for $+ i\varepsilon$.

The $i\varepsilon$-technique does not give the general expression for all the Green's functions of the Helmholtz equation (compare Equation (7.15)), but only some particular solutions; one of them solves Equation (7.2) subject to the boundary condition Equation (7.3).

7.6 Summary

We have derived the Green's function for the Helmholtz equation and obtained the well-known result. This research was initiated by the author's observation that some popular physics textbooks do not explain the corresponding derivation correctly. In fact, some of the published derivations of the Green's function for the Helmholtz equation can be quite misleading. Unlike the $i\varepsilon$ or other techniques used for similar purposes in most textbooks, we applied a more direct and general approach which, however, requires some skills in manipulating generalised functions. A potential advantage of the latter approach is that it might allow one to find a general form for the Green's functions of a given partial differential equation with constant coefficients. The imposition of suitable boundary conditions allows one to then fix the values of relevant free constants and to find a unique solution. This approach is based on the fact that a general solution of an inhomogeneous equation can always be represented as a sum of a single particular solution and a general solution of the associated homogeneous equation.