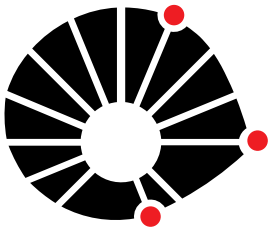


UNICAMP

UNIVERSIDADE ESTADUAL DE CAMPINAS

Instituto de Física "Gleb Wataghin"

YURI ROSSI TONIN

Coherent X-Ray Diffraction Imaging: Image reconstruction via a matrix model of the inhomogenous Helmholtz equation

Imagem por Difração Coerente de Raios-X: Reconstrução de imagem via um modelo matricial da equação de Helmholtz inomogênea

Campinas

2022

Yuri Rossi Tonin

Coherent X-Ray Diffraction Imaging: Image reconstruction via a matrix model of the inhomogenous Helmholtz equation

Imagem por Difração Coerente de Raios-X: Reconstrução de imagem via um modelo matricial da equação de Helmholtz inomogênea

Dissertação apresentada ao Instituto de Física "Gleb Wataghin" da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Mestre em Física, na área de Física Aplicada.

Dissertation presented to the "Gleb Wataghin" Institute of Physics of the University of Campinas in partial fulfillment of the requirements for the degree of Master in Physics, in the field of Applied Physics.

Supervisor: Jean Rinkel

Este exemplar corresponde à versão final da Dissertação defendida pelo aluno Yuri Rossi Tonin e orientada pelo Prof. Dr. Jean Rinkel.

Campinas

2022

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Física Gleb Wataghin
Lucimeire de Oliveira Silva da Rocha - CRB 8/9174

Tonin, Yuri Rossi, 1994-
T614c Coherent X-ray diffraction imaging : image reconstruction via a matrix model of the inhomogenous Helmholtz equation / Yuri Rossi Tonin. – Campinas, SP : [s.n.], 2022.

Orientador: Jean Rinkel.
Dissertação (mestrado) – Universidade Estadual de Campinas, Instituto de Física Gleb Wataghin.

1. Raios X - Difração. 2. Imagem de raio X. 3. Equação de Helmholtz. 4. Tomografia. 5. Luz síncrotron. I. Rinkel, Jean, 1980-. II. Universidade Estadual de Campinas. Instituto de Física Gleb Wataghin. III. Título.

Informações para Biblioteca Digital

Título em outro idioma: Imagem por difração coerente de raios-X : reconstrução de imagem via um modelo matricial da equação de Helmholtz inomogênea
Palavras-chave em inglês:
X-rays - Diffraction
X-ray imaging
Helmholtz equation
Tomography
Synchrotron light
Área de concentração: Física Aplicada
Titulação: Mestre em Física
Banca examinadora:
Luiz Fernando Zagonel
Helio Cesar Nogueira Tolentino
Maicon Ribeiro Correa
Data de defesa: 21-07-2022
Programa de Pós-Graduação: Física

Identificação e informações acadêmicas do(a) aluno(a)
- ORCID do autor: https://orcid.org/0000-0003-1403-815X
- Currículo Lattes do autor: http://lattes.cnpq.br/4106573850104828

MEMBROS DA COMISSÃO JULGADORA DA DISSERTAÇÃO DE MESTRADO DO ALUNO YURI ROSSI TONIN - RA 263646 APRESENTADA E APROVADA AO INSTITUTO DE FÍSICA “GLEB WATAGHIN”, DA UNIVERSIDADE ESTADUAL DE CAMPINAS, EM 21/07/2022.

COMISSÃO JULGADORA:

- Prof. Dr. Luiz Fernando Zagonel - Presidente (IFGW/UNICAMP)

- Prof. Dr. Helio Cesar Nogueira Tolentino (Centro Nacional de Pesquisa em Energiaa e Materiais)

- Prof. Dr. Maicon Ribeiro Correa (IMECC/ UNICAMP)

OBS.: Ata da defesa com as respectivas assinaturas dos membros encontra-se no SIGA/Sistema de Fluxo de Dissertação/Tese e na Secretaria do Programa da Unidade.

CAMPINAS

2022

Aos meus pais, que não pouparam esforços para que eu tivesse acesso a uma boa educação.

Acknowledgements

First of all, I would like to thank my parents, Adalto and Angela, and my sister, Yane, for all the support they have given me throughout all my life as a student. Having you by my side has given me strength to persevere even in the hardest moments, and for that I am forever grateful.

My special thanks goes to my advisor, Jean Rinkel, with whom I have had great pleasure to work with since 2019. Having you as an advisor really made this Masters a more pleasant endeavour. In our meetings, you always provided me with great useful insights and knew how to point me in the right direction. At the same time, you knew how to praise my results, which helped me to keep motivated, at the same that you knew how to provide constructive criticism when needed. I feel I grew a lot as a researcher during these years and I hope I have learned to think a bit more like you do when solving challenging and interesting problems. Even though we could meet in person only a few times during these years because of the pandemic, I really consider you as a friend and I hope we can continue to collaborate in the future. Thank you!

A big thanks goes to our main contributor, Eduardo Miqueles, for having participated in meetings and for really helpful suggestions that helped us achieve better results. Your mathematical thinking and the ease with which you handle mathematical tools has really kindled once again my interest in Mathematics. I hope to use this revived curiosity to become more proficient in the field and thereupon a better scientist.

I cannot forget here my advisors from undergraduate years: Emanuel Henn, Christian Illiadis and Fernando Paiva. You were amazing mentors and taught me how to think like a scientist when solving problems. I also need to thank some professors from those years, namely Paulo Liboni, Celso Villas-Boas, Adenilson Chiquito and Vivaldo Leiria. The depth of your knowledge and quality of your lessons brought excitement to my studies, a vital ingredient if you want to overcome all the challenges when learning and answering scientific questions.

I also want to thank all my friends who have in some way helped me during this Masters, especially: Thainá Lollo, Humberto Rigamonti, Sergio Lordano, Francesco Lena, Murilo Barbosa, Alain Peixinho, Francisco Figueiredo and Flavio Borges.

At last, I thank the Brazilian Synchrotron Light Laboratory, for the flexibility of allowing me to take graduate courses during the work time, without which I would not be able to finish this project in due time.

Resumo

Este trabalho propõe um novo modelo de reconstrução de imagem tridimensional para Difração Coerente de Raios-X. O método proposto evita que seja realizada a Aproximação da Projeção, a qual é tradicionalmente aplicada à equação que modela a interação radiação-matéria: a equação de Helmholtz inomogênea (IHE).

Imagem por Difração Coerente de Raios-X pode ser realizada em duas etapas: recuperação de fase e tomografia. A primeira consiste na recuperação da informação de fase que é perdida quando realizamos um medida de intensidade de um padrão de difração por um detector de área. Quando a amostragem do padrão de difração é feita de forma adequada, podemos, em princípio, recuperar a informação de fase utilizando métodos computacionais.

Num experimento, coleta-se múltiplos padrões de difração para diferentes posições da amostra. Para cada padrão, aplica-se a etapa de recuperação de fase. Em seguida, algoritmos de tomografia são aplicados para se obter uma imagem. É nesta etapa em que propomos um novo método. Demonstramos que a interação radiação-matéria pode ser modelada por uma matriz-sistema, a qual denominamos "Matriz-M". A relação linear entre a matriz da amostra e a Matriz-M permitiu que adaptássemos a Técnica de Reconstrução Algébrica ao nosso modelo, o qual nos referimos neste trabalho como algoritmo "M-ART".

O principal objetivo deste trabalho foi avaliar se o algoritmo M-ART poderia melhorar a qualidade da reconstrução de imagem. Em particular, técnicas tradicionais baseadas na Aproximação da Projeção apresentam um limite inerente para a resolução que pode ser alcançada quando se reconstrói uma amostra de certa espessura. Como M-ART dispensa a Aproximação da Projeção, fizemos a hipótese de que poderíamos superar esse limite de resolução. Se correta, a hipótese permitirá melhor aproveitar o potencial das novas fontes de luz coerente ao redor do globo, as quais possibilitam o estudo de amostras de tamanho micrométrico com resolução nanométrica.

Implementamos e validamos os algoritmos de recuperação de fase e M-ART separadamente realizando simulações em linguagem Python. No entanto, uma validação completa do método combinando ambas as etapas provou-se mais difícil do que antecipado Isso é consequência direta do custo computacional do algoritmo iterativo, o qual aprensenta complexidade $\mathcal{O}(n^3)$. Isso impossibilitou simulações de amostras maiores, necessárias para uma avaliação completa da reconstrução combinando ambas as etapas.

A simulação obtida do algoritmo M-ART indica sua superioridade comparada a Aproximação da Projeção. M-ART não piora fortemente a qualidade da reconstrução conforme

aumenta-se a espessura da amostra. Porém, ainda é possível que tal resultado possua um viés de “crime inverso”, isto é, ao fato de sintetizarmos o dado de simulação utilizando o mesmo modelo que resolve o problema inverso posteriormente.

Como perspectiva futura, este trabalho pode otimizar a velocidade do algoritmo M-ART por meio de estratégias de computação de alto-desempenho como: computação paralela, cálculos em GPU, e converter partes críticas do código para a linguagem C.

Palavras-chave: Imagem por Difração Coerente. Tomografia. Reconstrução de Imagem.

Abstract

This work proposes a new three-dimensional image reconstruction model for Coherent X-ray Diffraction. The method we propose innovates by circumventing the Projection Approximation (PA) that is traditionally applied to the equation that models the radiation-sample interaction, namely, the inhomogenous Helmholtz equation (IHE).

Coherent X-ray Diffraction image (CDI) reconstruction may be performed in two-steps: phase-retrieval and tomography. The first consists in the recovery of the phase information that is lost when an intensity measurement of a diffraction pattern is performed by an area detector. When properly sampled, the diffraction pattern in principle allows the retrieval of the lost phase information using iterative computational methods.

In an experiment, one collects multiple diffraction patterns for different positions of the sample. For each pattern, we employ the phase-retrieval step. Afterwards, tomography algorithms can be employed to reconstruct the three-dimensional image. This is the step where we propose a novel approach. We demonstrate that the radiation-sample interaction can be modelled by a system-matrix, which we call the "M-matrix". The linearity between sample matrix and the M-Matrix allowed us to adapt the Algebraic Reconstruction Technique to our model, which we refer in this work as the "M-ART" algorithm.

The main goal of this project was to evaluate if M-ART could improve the quality of image reconstruction. In particular, traditional reconstruction techniques based on the Projection Approximation present an inherent limit to the resolution that can be achieved when reconstructing a sample of a certain thickness. Since we managed to dispense with the Projection Approximation, we hypothesized we could dispose of such resolution limitation. If correct, the hypothesis will allow us to better explore the potential of the novel Coherent light sources around the globe, which in principle make it possible to study samples of micro-metric size with nano-resolution.

We managed to implement and validate both the phase-retrieval and M-ART algorithms separately performing simulations in Python language. Nonetheless, the complete validation of the method combining the two steps proved to be more difficult than anticipated. This is a direct consequence of the computational cost of the iterative algorithm, which presents $\mathcal{O}(n^3)$ complexity. This hindered the use of the bigger samples we needed to successfully perform the full CDI simulation combining the two reconstruction steps.

The simulation of the M-ART algorithm alone indicates its superiority compared to the PA. M-ART does not strongly deteriorate reconstruction quality as one increases the thickness

of the sample. Nonetheless, it is still possible that this result is being influenced by an "inverse crime" bias, that is, the act of synthesizing the simulation data using the same model that solves the inverse problem afterwards.

A future perspective for this work is to improve the speed of the M-ART algorithm, by employing high-performance computing strategies such as GPU and parallel computing, as well as converting the critical parts of the code to the faster C language.

Keywords: Coherent Diffraction Imaging. Tomography. Image Reconstruction.

List of Figures

List of Tables

Contents

1 Introduction

1.1 Imaging and X-rays

The evolution of science in the past few centuries has been, to a large extent, linked to the development of imaging. From the early telescopes and microscopes used to observe stars and microorganisms, humans have used images to properly analyze and characterize nature. The advances in imaging techniques continue up to this day and, thanks to the great advances in mechanical and control engineering, optics, computational hardware, software, among others, we are now able to obtain images of distant galaxies as well of the atomic structure of matter.

In particular, material science has evolved rapidly during the past few decades due to developments in microscopy and other imaging techniques, such as electron micrsocopy, crystallography and coherent diffractive imaging (CDI). Each imaging technique makes use of a distinct mechanism, presenting a different domain of application together with particular advantages. Naturally, the most common and well-known technique is visible light microscopy. The downside is that the resolution Δx that can be achieved with such microscopes is limited by the Rayleigh resolution, also known as the diffraction-limit:

$$\Delta x = 0.61 \frac{\lambda}{\mathrm{NA}} \tag{1.1}$$

where λ is the wavelength light and NA the numerical aperture of the lens. Therefore, due to the wavelength of visible light ranging from 400 - 700 nm, the achieved resolution is typically limited to a few hundred nanometers.

X-ray and Electron microscopes circumvent this limitation due to their lower wavelengths. The former can in principle be used to achieve resolutions of a few nanometers, while modern electron microscopes reach sub-nanometer resolutions [2]. However, such increase in resolution comes with a cost. First, the higher energy of the beams may cause radiation damage to the sample. In particular, this becomes a bigger issue for biological samples and partly justifies the success of light microscopy in this field, since visible light will not cause such damage, allowing to probe biological phenomena over longer periods of time.

1.2 Radiation Damage and Penetration Distance

Although X-rays achieve lower resolutions when compared to electron microscopy, it presents a clear advantage: how far it penetrates in matter. This is essential if

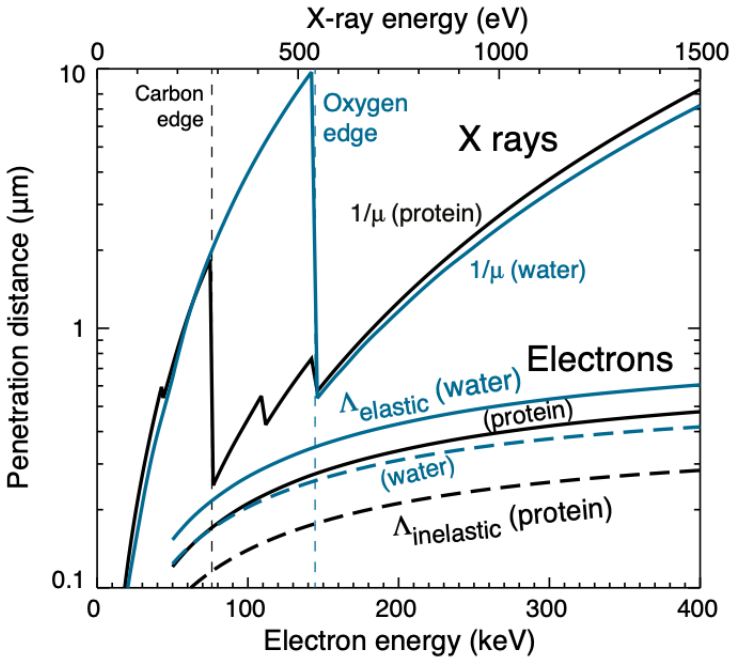


Figure 1 – Penetration distance of X-rays and mean free path of electrons as a function of energy for water (blue) and a protein (black). Figure from [1].

one wishes to obtain information about the internal structure of matter. Indeed, X-ray imaging allows us to obtain tridimensional images of matter, with modern techniques being able to reach, in principle, resolutions of a few nanometers.

Naturally, each material will interact differently with the radiation. The linear attenuation coefficient μ quantifies how strongly it gets attenuated as it propagates through matter. Considering a material of thickness t and an incident radiation of intensity I_0, Beer's law gives us that the intensity is exponentially decaying

$$I = I_0 \exp(-\mu t) \tag{1.2}$$

One can rewrite this as

$$I = I_0 \exp\left(-\frac{\mu}{\rho} x\right) \tag{1.3}$$

where $x = \rho t$ is called the mass thickness, defined as mass per unit area. The quantity μ/ρ is called the *mass attenuation coefficient*. It is is commonly used to quantify the interaction of the material with the radiation, as we shall see in the next section.

Figure 1 shows the penetration distances as a function of energy in the soft x-ray region[1]. One can see that the distance for X-rays is greater over almost all the energy range shown. The curves show data for water and a protein (representing carbon). One can see a great difference in the penetration distance of X-rays in the energy region between 290 eV and 540 eV, the K shell absorption edges of carbon and oxygen in water. This is

[1] We consider soft x-rays to be those with energy from roughly 250 eV to about 10keV. Above this energy, we shall refer to them as hard x-rays.

the so-called "water windows", which does not occur for electrons. This energy region is suitable, for instance, for obtaining great contrast for biological imaging, since organic matter will not be as transparent as water, presenting yet another advantage of X-rays over electrons.

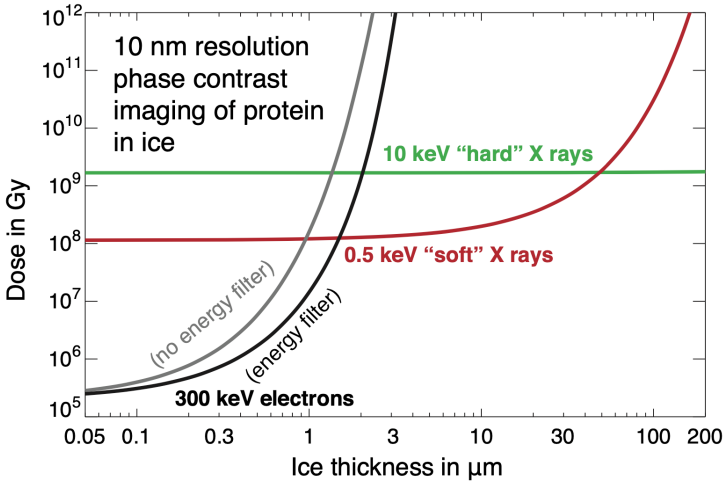


Figure 2 – Radiation dose imparted by X-rays and electrons for proteins in ice as a function of ice thickness. Figure from [1].

Figure 2 shows yet that radiation damage also limits what can be achieved with electron microscopy, especially when dealing with biological samples. We have the dose imparted to a sample of proteins in ice as a function of the ice thickness for achieving a resolution of 10 nanometers using phase contrast imaging. We see that only for a thickness up to a micron the radiation dose of electrons is lower than that of X-rays. Consequently, one needs X-rays to image bigger samples such as entire eukaryotic cells, not only because of the higher penetration of X-rays (allowing to probe the sample tridimensionally), but also because only then the radiation damage would be small enough to allow the image acquisition process to occur.

1.3 Interaction of X-rays with Matter

Different phenomena may happen when radiation interacts with matter. We will here restrict ourselves to those that happen within the energy range of interest to X-rays, namely: photoelectric absorption, elastic (or Rayleigh) scattering and inelastic (or Compton) scattering. Other effects such as pair production, triplet production and photonuclear interaction [3] are not treated here, since these are restricted to higher photon energies outside our scope.

The interaction of radiation with matter is quantified using the *cross-section* σ, which can be thought of as an effective area for intercepting an incoming photon. The

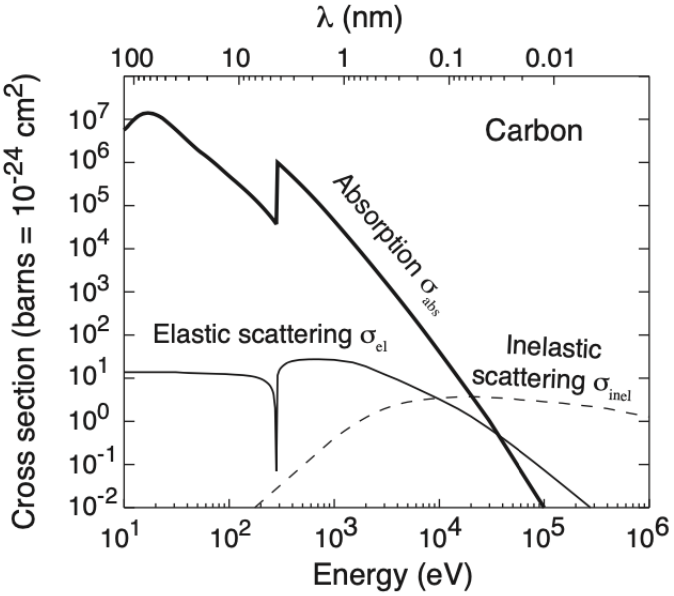


Figure 3 – X-ray cross-sections of Carbon for absorption, elastic scattering and inelastic scattering as a function of energy. Figure from [1].

cross-section is typically described in units of barns ($= 10^{-24} cm^2$). The total cross-section σ_t is a direct sum of the respective cross-sections of each phenomena

$$\sigma_t = \sigma_a + \sigma_e + \sigma_i \tag{1.4}$$

where σ_a, σ_e and σ_i are the cross-sections for photoelectric absorption, elastic scattering and inelastic scattering, respectively. The cross-section directly relates to the mass attenuation coefficient according to [4]:

$$\frac{\mu}{\rho} = \frac{\sigma_t}{uA} = \frac{(\sigma_a + \sigma_e + \sigma_i)}{uA} \tag{1.5}$$

where A is the relative atomic mass of the element and u the atomic mass unit. Tabulated values are usually given in mass attenuation coefficient units.

Figure 3 shows the cross-section of Carbon for energies between 10 eV to 1 MeV. Notice that for energies up to about 10 keV, absorption dominates over scattering. This indicates that multiple scattering events are rare at such energies, since after scattering a first time, the photon will likely be subsequently absorbed. As seen in Figures 1 and 2, this energy range comprises the water window and also can be an appropriate region for experimenting without causing too much radiation dose to the sample. Consequently, many synchrotron beamlines operate at such energies. In that case, the majority of incident photons will be absorbed instead of elastically scattered. Therefore, scattering techniques require a high flux of photons so that the detector collects a sufficient number of scattered photons in reasonable time for satisfactory levels of photonics noise.

1.4 Synchrotron Light Sources

Synchrotron light sources are arguably one of the most versatile machines for studying materials. Electrons are accelerated in a controlled manner such that they emit photons in a vast spectrum and in vast quantity. Figure 4 illustrates the electromagnetic spectrum energy and wavelength values. The shaded region in pink corresponds to those energies covered by lasers. Synchrotrons also cover this range, as well as orders of magnitude above it in the X-ray spectrum, as shown by the shaded yellow region. Roughly, the synchrotron light spectrum covers the range of 0.1 eV up to 1×10^6 eV [5].

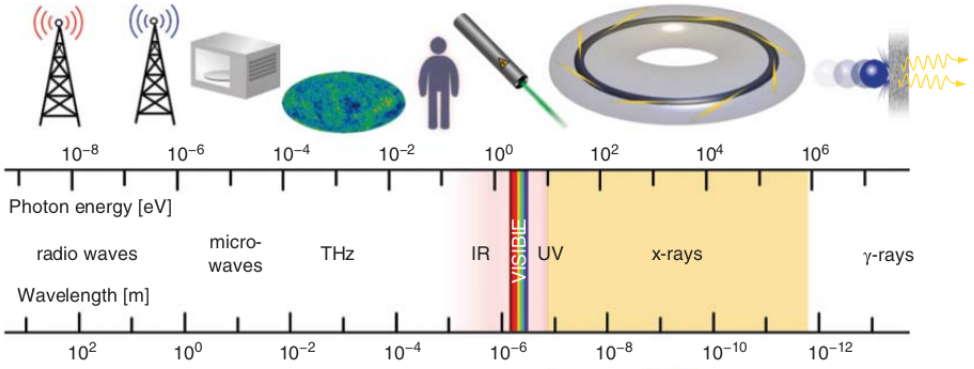


Figure 4 – Electromagnetic spectrum energy and corresponding wavelengths. The shaded regions in pink show values typically covered by lasers and synchrotron light sources. The shaded region in yellow represents the X-ray frequencies also generated by the synchrotron. Adapted from [5].

Furthermore, the huge number of emitted photons makes synchrotron light sources the only practical option for many scientific cases of interest, since an experiment that would take weeks in a common laboratory X-ray source, would take only minutes with the synchrotron. The quality of the synchrotron is usually measured by its *brilliance*, defined as

$$\text{Brilliance} = \frac{\text{photons/second}}{(\text{source size [mm}^2\text{]}) \quad (\text{source divergence [mrad}^2\text{]}) \quad (0.1\% \text{ bandwidth})} \tag{1.6}$$

Hence, for high brilliance one needs high flux (number of photons per second) and a small beam with low divergence. Usually, one quantifies beam size B and divergence B' together by the *emittance* ϵ, defined as (Figure 5):

$$\epsilon_{xy} = B_{xy} B'_{xy} \tag{1.7}$$

The new Brazilian Synchrotron, Sirius, is the second fourth-generation synchrotron light source to start operations, the first experiments being performed in late

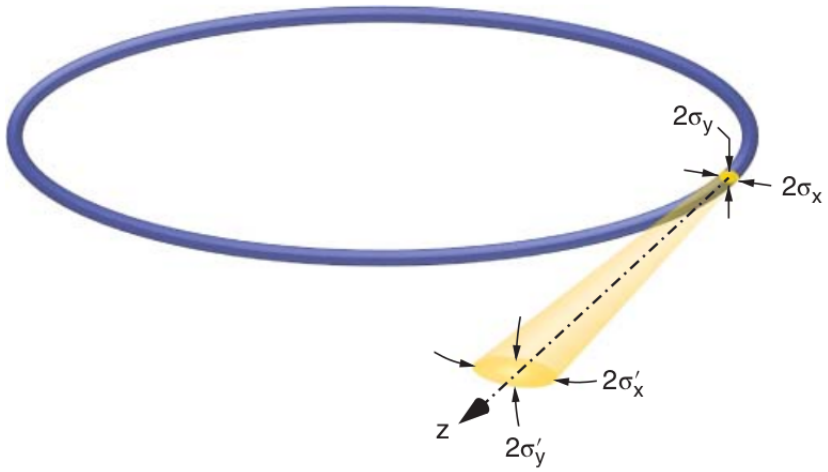


Figure 5 – Illustration of the parameters that define the emittance of the synchrotron light beam. Figure from [5].

2020. It is a 3 GeV storage ring with 518 m circumference, presenting a low emittance of 0.25 nm rad [6].

Of particular interest for this work is the Cateretê beamline of Sirius, which started accepting external users in April 2022. This beamline is dedicated for the techniques of Coherent Diffractive Imaging (CDI), X-ray Photon Correlation Spectroscopy (XPCS), Small Angle and Ultra-Small Angle X-ray Scattering (SAXS and USAXS). Cateretê works in the energy range of 3 to 23 keV and has a fully-coherent beam with variable focus of $1 \times 1\ \mu m^2$ to $30 \times 30\ \mu m^2$ full-width at half-maximum (FWHM). An area detector with 3072×3072 pixels (each with $55 \times 55\ \mu m^2$ size) is placed in vacuum, with adjustable distance from 1 to 28 meters. Coherent flux is 10^{11} ph/s at 6 keV, 10^{10}ph/s at 9 keV [7]. We further discuss CDI in the next section, as this is the technique of main interest for this work.

1.5 Coherent Diffractive Imaging

Traditionally, imaging systems make use of lenses to recover an image. Lenses can be thought of a device that inverse-Fourier transforms the wavefield that has undergone a Fourier transform due to free-space propagation. This is illustrated in Figure 7 (left). Nonetheless, several issues arise at X-ray wavelengths when using lenses. Since the refractive index is close to unity, one needs to use several lenses together (*compound refractive lenses*) in order to focus the radiation. This is not only complex to fabricate, but also strongly attenuates the radiation. Another option is the use of Fresnel Zone Plates, which focus radiation by diffracting the X-ray beams. However, the focal length of such component depends on the wavelength, which deteriorates the focus and, consequently, the spatial

resolution. Furthermore, if one wishes to reach a resolution Δx of a few nanometers, fabrication of such components becomes extremely complicated because the width Δw_N of the outermost ring relates to Δx as [5]:

$$\Delta x = 1.22 \Delta w_N \tag{1.8}$$

At last, the power efficiency of X-ray optics is usually quite low (below 60% [1, Chapter 10]), making it impractical in many situations.

An alternative to lenses is the use of *lensless imaging* techniques, where the resolution is no longer limited by the lens, but rather by three other parameters: the wavelength λ of the incident radiation, the detector-sample distance L, and the detector size D:

$$\Delta x = 4\lambda \frac{L}{D} \tag{1.9}$$

This allows one to achieve, in principle, a resolution of a few nanometers, as shown in Figure 6.

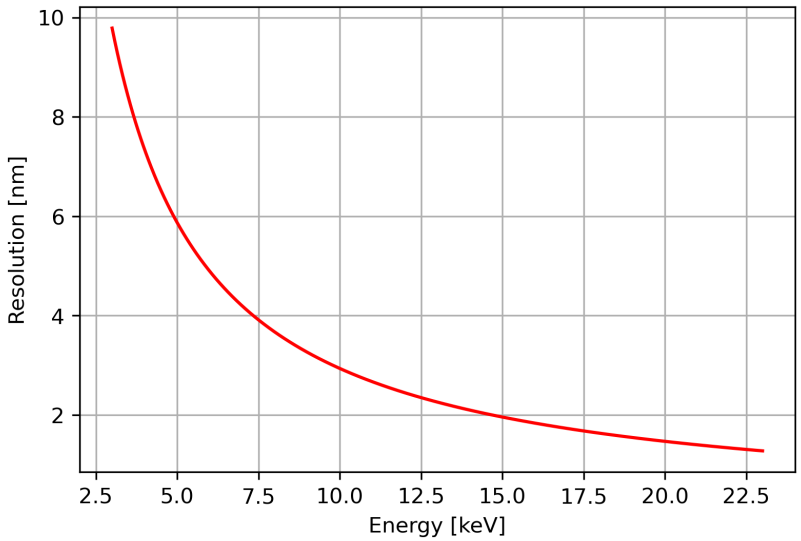


Figure 6 – Spatial resolution, computed from equation 1.9 for the parameters of Cateretê ($D = 55 \times 3072\,\mu\text{m}$ and $L = 1\,\text{m}$) as a function of the energy range covered by the beamline. In principle, a resolution of only a few nanometers may be achieved via lensless imaging techniques.

In these diffraction experiments, a detector will measure the scattering pattern of the sample. This consists in an intensity measurement and brings significant difficulty to the imaging problem. The intensity measurement in the Fraunhofer regime consists in

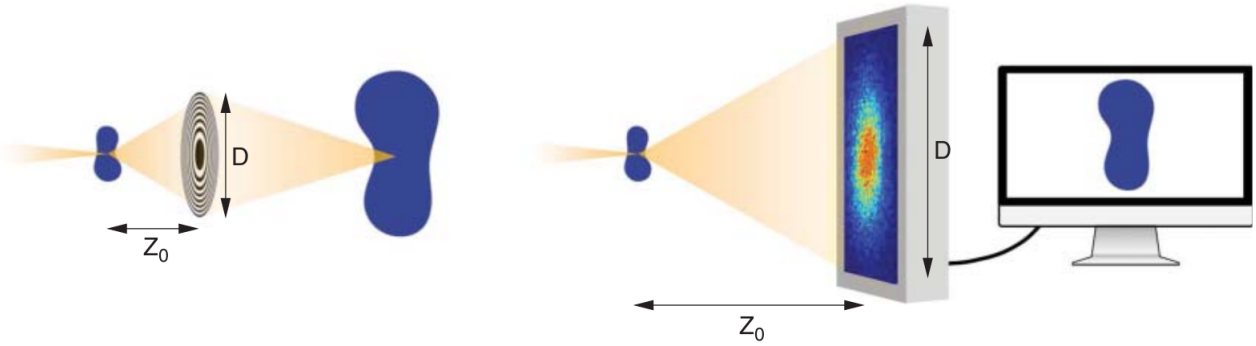


Figure 7 – A typical imaging experiment uses an optical component (such as a lens of Fresnel Zone plate) to recover and image of the object (left). On the other hand, Coherent Diffractive Imaging relies on measuring the reciprocal space of the object and then recovering the image with the aid of computational methods (right). Figure from [5].

the absolute squared value of the Fourier Transform (FT) of the electron-density[2]:

$$I = |\mathcal{F}(\rho(x))|^2 \tag{1.10}$$

An intensity measurement, therefore, is a measurement of the object's reciprocal space which looses all phase information about the scattered wave. Consequently, to recover the three-dimensional structure of the object, one cannot simply perform an inverse Fourier Transform to the measured signal, since part of the information is missing. Instead, one must rely on numerical computational methods to recover the lost phase information and then obtain an image. Figure 8 qualitatively illustrates the relevance of the information content carried by the phase. The original image on the left is Fourier transformed and an inverse Fourier transform is applied separately to the magnitude (central image) and phase content (right image) of reciprocal-space. It is evident that the information recovered only from the magnitudes does not resemble the original signal, while the one recovered from the phases does. This is a strong indicator of the necessity to properly retrieve the phase for a reliable image reconstruction.

We note that the intensity measurement being directly related to the Fourier Transform brings yet another advantage to the lensless techniques. Due to the properties of the FT, a shift of $\rho(\mathbf{x})$ in space results only in a phase-shift of its Fourier Transform. This means the technique is robust against mechanical vibrations of the sample during the experiment, since the recorded diffraction pattern over time will not become blurred due to integration in the detector.

Coherent Diffractive Imaging is broadly used to refer to different techniques,

[2] As we shall detail in section 1.3, this is the case strictly when working at high energies far from absorption edges.

Figure 8 – Qualitative example of the information contained in the magnitude and in the phase of a signal. The original image is Fourier Transformed (FT). The central image is then recovered by performing the inverse Fourier transform (IFT) using only the magnitudes of the FT, whereas the image on the right is recovered via the IFT of the phases. One can clearly see that the recovered image using the phases resembles the original image, but the same cannot be said about the image from the magnitudes. Figure from [1].

the main ones being Plane Wave Coherent Diffractive Imaging and Ptychography [3]. In this work, we use the CDI abbreviation to refer strictly to the Plane-Wave technique. Although the novel development of this work applies to both techniques, we chose to focus on Plane Wave CDI for two main reasons: first, due to its simplicity compared to Ptychography; second, because the existence of analytical forms of the diffraction pattern for some object geometries (section 3.2.5) would in principle help us validate the newly proposed method.

CDI reconstruction may be performed in two different ways, which we call single and two-step reconstruction. Although experimentally similar, the two methods differ in the approach used to recover the three-dimensional structure of the sample. In single-step reconstruction [9], one performs multiple measurements to probe the 3D reciprocal space of the object. After phase-retrieval, one has complete information of the object's 3D reciprocal space and can recover the image via a direct 3D inverse Fourier transform. On the other hand, the more common two-step reconstruction relies on obtaining 2D projection images from multiple measurements, and then using tomographic reconstruction algorithms to recover the three-dimensional structure.

Traditionally, the tomographic part of the two-step reconstruction relies on the Projection Approximation (PA) [10]. In this scenario, the Radon transform applies and traditional tomography algorithms such as Filtered Backprojection (FBP) may be applied to recover the 3D image [11]. Unfortunately, under this approximation there is a limit to

[3] Ptychography combines Scanning X-ray microscopy and plane-wave CDI. It relies on performing multiple measurements while scanning the sample using a small beam. The overlapping of the illuminated regions between each diffraction measurement provides us with redundant information so that one obtains a well-posed phase-retrieval problem. We do not focus on Ptychography in this text and refer the reader to other texts for details [1, 5, 8].

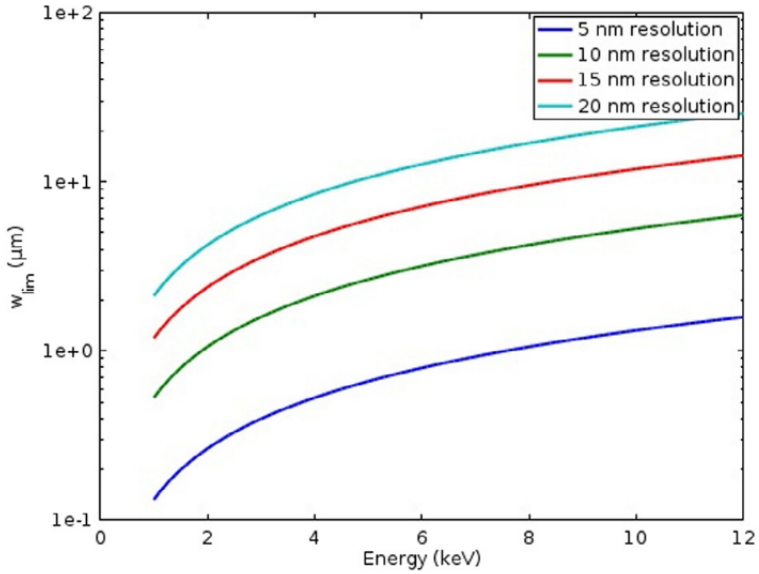


Figure 9 – Limit sample thickness as a function of energy for different reconstruction resolutions. Figure from [12].

the reconstruction resolution Δr that can be achieved for a sample of thickness t being imaged under radiation of wavelength λ:

$$t = \frac{4}{0.61}\frac{\Delta r^2}{\lambda} \tag{1.11}$$

This equation[4] is plotted in Figure 9 for different resolution values. From the 5 nm curve, we see that we need to limit ourselves to samples only 1 micron thick, even for X-ray energies as high as 12 keV. Such thickness is small compared to the size of the $30 \times 30\,\mu m^2$ FWHM coherent beam one has access to in a fourth-generation beamline such as Cateretê. Limiting ourselves to a sample thickness of only 1 micron would impede us, for instance, to reliably study in ultra-high resolution biological samples such as bacteria and eukaryotic cells, greatly limiting the potential of fourth-generation Synchrotron light sources for scientific discovery.

In this work, we were able to develop a numerical reconstruction method to recover the 3D sample without the need to make the Projection Approximation. Then, we attempted to answer the following question: by avoiding the PA, are we able to go beyond the reconstruction resolution limit posed by equation 1.11? As we shall see in section 4, we have encountered several challenges when trying to answer this question, especially due to the computational cost of the newly proposed reconstruction model. Nonetheless, the obtained simulation results indicate that, indeed, going beyond the Projection Approximation helps us to obtain better reconstructions.

[4] There is some variation in the literature regarding the constant factor [1] in equation 1.11, depending on how one defines resolution.

2 Objectives

This Masters' project main objective was the development and simulation of a new three-dimensional image reconstruction model for X-Ray Coherent Diffractive Imaging (CDI). The most widely used 3D reconstruction method in CDI relies on two-steps: phase-retrieval and tomography.

The first consists in performing multiple scattering experiments rotating a sample, in which a diffraction pattern is measured by an area detector for each sample position. The acquisition consists in an intensity measurement, which lacks information about the phase of the measured wavefield. Nonetheless, given the right experimental conditions (more specifically an appropriate combination of detector distance, detector pixel-size and radiation energy), one may use computational techniques to retrieve the phase information that is encoded in the diffraction pattern – also referred to as *speckle pattern* in the case of coherent radiation. The goal of the first step is to apply such techniques in order to successfully retrieve a two-dimensional projection of the sample. A first objective of this work was then to implement Phase-Retrieval algorithms that would allow us to eventually have a complete simulation of the CDI reconstruction process.

By the end of the first step, one has several 2D projections, each acquired for a different incidence angle of the radiation with respect to the sample. These projections then allow us to move to the second-step: the tomographic reconstruction. This is the step for which we propose a new approach called M-ART. Traditionally, tomography relies on an approximation of the inhomogenous Helmholtz equation (IHE) called the *Projection Approximation* (PA), under which the Radon transform applies. In this case, one can use the fast tomography *Filtered Backprojection* (FBP) algorithm to obtain a 3D sample from the multiple 2D projections obtained in the first step. However, within the PA, there is an inherent limit to the resolution Δr that can be achieved for a sample of thickness t.

The second main objective was the development of a discrete matrix model of the IHE without the PA, that allows us to relate the interaction of the incident radiation and a sample of arbitrary size. Throughout this work, we refer to the matrix as the "*M-Matrix*". Under this condition, the FBP algorithm can no longer be used. Fortunately, the linear relationship between M-Matrix and sample allowed us to use well established Algebraic Reconstruction Techniques (ART) to perform the tomographic part of the reconstruction. The next main goal was to adapt the ART algorithm to the proposed matrix method and simulate the second-step to validate it.

ART algorithms are known for being slow due to their iterative nature. In our case, the matrix model complicates the problem even further, since it necessarily

causes the algorithm to have $\mathcal{O}(n^3)$ complexity instead of $\mathcal{O}(n)$, as in the PA scenario. The computational cost for performing the reconstruction indeed turned out to be huge, which made it difficult and prevented us from achieving some of the secondary goals of this project, namely: the full-simulation of a 3D CDI reconstruction using the proposed method, and the validation of the method using real experimental data from a beamline.

3 Theory

3.1 Wave propagation

In this initial section, we give an overview of the models of interaction of propagating electromagnetic radiation with matter, more specifically for the case of interest: X-rays. We particularly find to be lacking in the literature a unified, yet simple review of the theoretical topics needed to understand Coherent Diffraction Imaging with all its prerequisites. This section aims at providing such simplified overview of all the required topics and hopefully to serve as an useful reference for the future.

The text herein presented is compiled from multiple references [1, 5, 10, 13–20]. For the sake of comprehensibility, some cumbersome demonstrations are not included and we refer the reader to the references for more details.

3.1.1 Vacuum propagation

We start by reviewing the propagation of electromagnetic waves in vacuum. In a synchrotron experiment, the propagation of the wave from the sample to the detector may occur in air or in vacuum, depending on the experiment's purpose. In the case of CDI at Sirius' Cateretê beamline, a low vacuum environment is created with approximately 1×10^{-3} mbar. It is useful to see how the free-space propagation works. Starting from Maxwell's equations in vacuum [21]

$$\begin{aligned} \nabla \cdot \vec{E} &= 0 & \nabla \cdot \vec{B} &= 0 \\ \nabla \times \vec{E} &= -\frac{\partial \vec{B}}{\partial t} & \nabla \times \vec{B} &= \mu_0 \varepsilon_0 \frac{\partial \vec{E}}{\partial t} \end{aligned} \tag{3.1}$$

with ε_0 and μ_0 being the electric permittivity and magnetic permeability of vacuum, respectively. Using vector calculus identities, we can deduce the wave-equations both for the electric $\mathbf{E}$ and magnetic $\mathbf{B}$ field vectors

$$\left[\varepsilon_0 \mu_0 \frac{\partial^2}{\partial t^2} - \nabla^2\right] \mathbf{E}(x, y, z, t) = \mathbf{0} \tag{3.2}$$

$$\left[\varepsilon_0 \mu_0 \frac{\partial^2}{\partial t^2} - \nabla^2\right] \mathbf{B}(x, y, z, t) = \mathbf{0} \tag{3.3}$$

It is convenient to switch the mathematical treatment to the scalar wave-equation[1] [10], such that

$$\left[\varepsilon_0\mu_0\frac{\partial^2}{\partial t^2} - \nabla^2\right]\Psi(x,y,z,t) = 0 \tag{3.4}$$

where $\Psi(x,y,z,t)$ is a scalar wavefunction representing any component of the electromagnetic field. The complex wavefunction may not be monochromatic, that is, it may be a wave composed of multiple frequencies. We can then decompose Ψ as a superposition of monochromatic components ψ_ω

$$\Psi(x,y,z,t) = \frac{1}{\sqrt{2\pi}}\int_0^\infty \psi_\omega(x,y,z)\exp(-i\omega t)d\omega \tag{3.5}$$

Combining the above equations, we obtain the *Helmholtz equation* for each monochromatic component:

$$[\nabla^2 + k^2]\psi_\omega(x,y,z) = 0, \qquad k = \omega/c \tag{3.6}$$

Therefore, the problem of solving the wave-equation is substituted by solving the simpler Helmholtz equation for each monochromatic component, which is solely a function of space, and then performing the integral in equation 3.5 for recovering Ψ. For simplicity, we refer herein to the monochromatic ψ_ω simply as "complex wavefunction".

Intensity

We first point out that the electromagnetic field is in reality a real quantity. The complex wavefunction ψ_ω of a monochromatic wave is therefore used for mathematical convenience. One can simply recover the real wavefunction by taking the real part $\psi_R = \Re(\psi_\omega)$. The intensity can then be calculated as a function of both quantities. It is defined as the optical power per unit area, which is the average of the squared real wavefunction or, conveniently, as the absolute square of the complex wavefunction:[2]

$$I(\mathbf{r}) = 2 < \psi_R^2(x,y,z,t) >= |\psi_\omega(x,y,z)|^2 \tag{3.7}$$

The plane and spherical waves

Two solutions admitted by the Helmholtz equations are the plane and spherical waves. A plane wave is a solution of the form

$$\psi_\omega(x,y,z) = \psi(x,y,z)\exp\left(-i\mathbf{k}\cdot\mathbf{r}\right) = \psi(x,y,z)\exp\left[-i(k_x x + k_y y + k_z z)\right] \tag{3.8}$$

[1] As pointed out by [10], it is not immediate to see why can make a transition to a scalar field to describe **E** and **B** in free space. We refer the reader to the references therein for such explanation.

[2] We assume here the average to be performed for a time long compared to the frequency of the wave.

where $\mathbf{r} = (x, y, z)$ and

$$k^2 = k_x^2 + k_y^2 + k_z^2 \tag{3.9}$$

$\psi(x, y, z)$ is the called the *complex envelope* of the wave. To understand why the plane wave is named like so, note that the phase of the complex wavefunction is

$$\arg \psi_\omega(x, y, z) = \arg \psi(x, y, z) - \mathbf{k} \cdot \mathbf{r} \tag{3.10}$$

Equating this to a constant phase value of $2\pi q$ ($q \in \mathbb{Z}$) we obtain the equation for the surfaces of constant phase (also called *wavefronts*):

$$\mathbf{k} \cdot \mathbf{r} = k_x x + k_y y + k_z z = 2\pi q + \arg\left(\psi(x, y, z)\right) \tag{3.11}$$

In this case, this is precisely the equation of planes perpendicular to the wavevector $\mathbf{k}$, each plane being separated by a distance equal to the wavelength $\lambda = 2\pi/k$. Hence the name *plane wave*. The spherical wave solution, on the other hand, writes

$$\psi_\omega(r) = \frac{\psi}{r} \exp\left(-ikr\right) \tag{3.12}$$

which, following the same procedure as for the plane wave, can be shown to be precisely concentric spherical wavefronts separated by a radial distance λ [14].

Fourier Decomposition

Analogous to the temporal frequency decomposition of the complex wavefunction into monochromatic components (equation 3.5), we can decompose each ψ_ω for different spatial frequencies as a sum of plane waves

$$\psi_\omega(x, y, z) = \frac{1}{\sqrt{2\pi}} \int\int \hat{\psi_\omega}(x, y, z) e^{-i(k_x x + k_y y + k_z)} dk_x dk_y dk_z \tag{3.13}$$

The above equation turns out to be precisely a tridimensional Fourier Transform with respect to the spatial frequencies k_i. $\hat{\psi_\omega}$ is called the Fourier Transform of ψ_ω. Decomposing the wave into planar components also greatly simplifies how we treat wave propagation, as we shall see ahead.

Paraxial waves

In ray optics, a *paraxial ray* is said to be those rays that make a small angle with respect to the optical axis of the system. In the context of wave optics, these can be thought to be the normal vector to the wavefronts. Considering the optical axis to coincide with the z-direction, a paraxial wave is described as

$$\psi_\omega(\mathbf{r}) = \psi(\mathbf{r}) \exp\left(-ikz\right) \tag{3.14}$$

where the complex amplitude $\psi(\mathbf{r})$ is a a slowly varying function of position[3]. From equation 3.14, the real wavefunction in ψ_R writes

$$\psi_R(\mathbf{r}, t) = |\psi(x, y, z)| \cos(\omega t - kz + \arg \psi(x, y, z)) \tag{3.15}$$

As illustrated in Figure 10, the above equation indicates that, if $\psi(r)$ varies slowly, the amplitude $|\psi(r)|$ is a slow varying envelope of the sinusoidal wave (Figure 10a). Furthermore, at a specific time t, the surfaces of constant phase will change little due to the $\arg \psi(x, y, z)$ factor, such that $kz + \arg \psi(x, y, z) \approx kz = 2\pi q$, meaning the wavevectors bend only slightly as the wavefront propagates (Figure 10b), maintaining the paraxial nature of the wave.

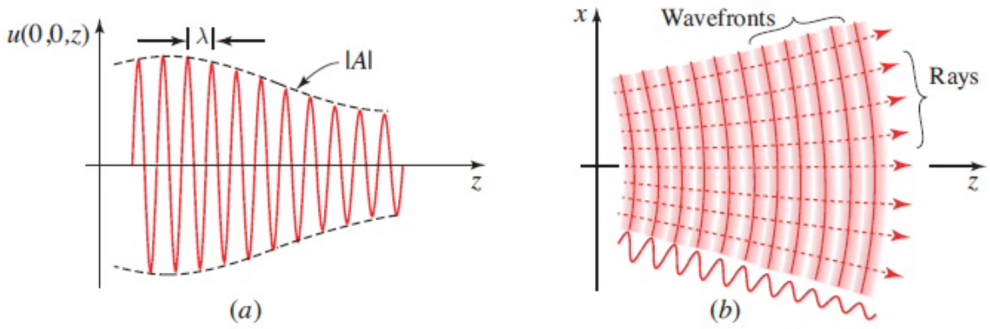


Figure 10 – On the left, a "side-view' of the wave, showing the amplitude $|\psi_R(0, 0, z)|$ for the real wavefunction. The amplitude of the complex envelope $|\psi|$ (here denoted by $|A|$) varies slowly in comparison to the wavelength. On the right, we see the surfaces of constant phase (wavefronts) "from above", where the rays bend slightly with z because of the slow variation of the phase $\arg(\psi_R(0, 0, z))$. Figure from [14].

3.1.2 Fresnel and Fraunhoffer Regimes

We now briefly turn our attention back to the spherical wave solution. Consider the wave to originate at the origin, as is in fact the case for equation 3.12. Consider also we are dealing with points in an output plane at z such that $z >> \sqrt{(x^2 + y^2)} = a$.[4]. In that case, 3.12 can be approximated by:

$$\psi_\omega(x, y, z) \approx \frac{\psi}{z} \exp(-ikz) \exp\left(-ik\frac{x^2 + y^2}{2z}\right) \tag{3.16}$$

[3] Meaning both the amplitude and its derivative vary little within the distance of one wavelength.

[4] If the rays reaching this circle of radius a in the output plane were originated at the origin, one could consider these to be paraxial rays.

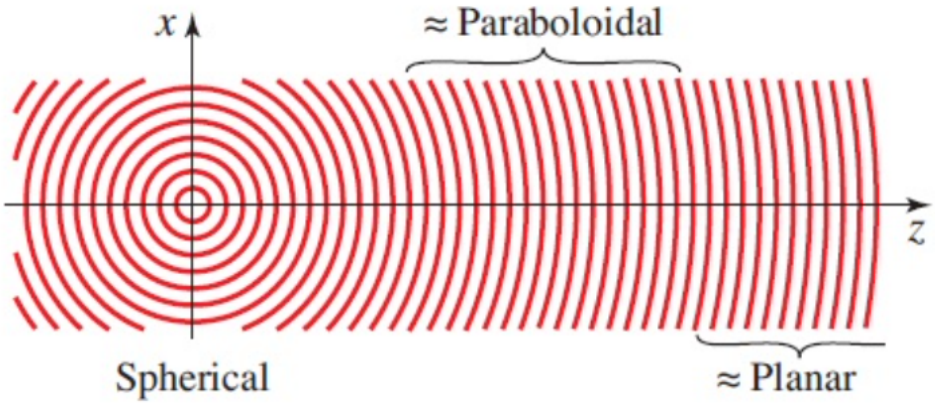


Figure 11 – Illustration of the a spherical wave wavefront originating at (x,z) = (0,0). The scale of the drawings show that for medium distances, the wavefronts become apparently paraboloidal, whereas for the last values of z the curvature is barely noticeable, indicating the plane wave approximation. Figure from [14]

which is known as the Fresnel regime of the spherical wave. Equation 3.16 can be interpreted as a plane wave $\psi \exp(-ikz)$ modulated by the factor $\exp\left(-ik\frac{x^2+y^2}{2z}\right)/z$. In that case, the surfaces of constant phase amount to be the paraboloid of revolution

$$\frac{x^2+y^2}{z} = 2\pi q, \qquad q \in \mathbb{Z} \tag{3.17}$$

If we consider yet bigger value of z (larger propagation distances), the quadratic exponential term tends to zero, whereas the exponential dependence on z dominates over $1/z$, and we recover a plane wave $\psi_\omega(x,y,z) = \psi(x,y,z)\exp(-ikz)$. This evolution of the wavefront curvature is illustrated in Figure 11 and is referred to as the *Fraunhoffer regime*. Additionally, when we mention large values of z, it is important to have mind: large with respect to what? Besides the radius a of the region of interest in the output plane indicating one relevant dimension to our problem, the wavelength is another important one. As a rule of thumb, the Fresnel approximation is valid when

$$\frac{N_f}{4}\frac{a^2}{z^2} << 1 \tag{3.18}$$

where

$$N_f = \frac{a^2}{\lambda z} \tag{3.19}$$

and N_f is called the *Fresnel number*. Since $a << z$ implies that $a^2/z^2 << a^2/z$, it is usually sufficient to guarantee that $N_f << 1$. On the other hand, besides the condition of equation 3.18, the Fraunhoffer approximation requires that

$$N_F = \frac{b^2}{\lambda z} << 1 \tag{3.20}$$

where b is the radius of the region of interest in the input-plane – for instance, the radius of a circular aperture through which the wave propagates. These approximations are convenient to help us simplify the mathematics of wave propagation. Consider a plane wave

$$\psi_\omega(x, y, z) = \exp\left[-i(k_x x + k_y y + k_z z)\right] \tag{3.21}$$

Using equation 3.9, one can rewrite it as

$$\begin{aligned}\psi_\omega(x, y, z) &= \exp\left(-i(k_x x + k_y y + k_z z)\right) \\ &= \exp\left(-i(k_x x + k_y y)\right) \exp\left(-iz\sqrt{k^2 - k_x^2 - k_y^2}\right) \\ &= \exp\left(-iz\sqrt{k^2 - k_x^2 - k_y^2}\right)\psi_\omega(x, y, z) \\ &= T(k_x, k_y, z)\psi_\omega(x, y, z = 0)\end{aligned} \tag{3.22}$$

That is, in order to obtain the wave at a position $z > 0$ from a position $z = 0$, one has to simply multiply the former wave by $T(k_x, k_y, z) = \exp\left(-iz\sqrt{k^2 - k_x^2 - k_y^2}\right)$, called the *free-space propagator* or the *transfer function of free space*. One interesting aspect is worth mentioning when we rewrite T as a function of wavelength. Note that:

$$T(k_x, k_y, z) = \exp\left(-i2\pi z\sqrt{\lambda^{-2} - \nu_x^2 - \nu_y^2}\right), \quad k_{xy} = 2\pi/\nu_{xy} \tag{3.23}$$

For $\nu_x^2 + \nu_y^2 \leqslant \lambda^{-2}$, the exponential becomes a negative real quantity. In other words, the transfer function becomes an attenuation factor. Consequently, one can show that, for propagation distances $z >> \lambda$, any plane wave of spatial frequencies above the cutoff frequency λ^{-1} will quickly be attenuated. In some sense, one might say that free-space propagation acts a low-pass filter. This is a profound realization that closely relates to the diffraction limit of light: any spatial frequency of the wave that is finer than a wavelength (information finer than λ) will not be transmitted by the wave over distances $z > \lambda$. Consequently, the resolution of imaging techniques is intrinsically limited by the wavelength of the radiation.

Fresnel and Fraunhoffer Diffraction

From equation 3.13, we have seen that an arbitrary wave may be decomposed as a sum of plane waves via the Fourier Transform $\mathcal{F}$. Consequently, the wave at a position $z = L$ may be obtained from its configuration at $z = 0$ simply by decomposing it into its plane wave components, propagating these components to the position $z = L$ with the transfer function $T(k_x, k_y, z = L)$, and then finally composing to total wave again using an Inverse Fourier Transform $\mathcal{F}^{-1}$. In operator notation, we have $\mathcal{D}_L$

$$\mathcal{D}_L = \mathcal{F}^{-1} \exp\left(-iL\sqrt{k^2 - k_x^2 - k_y^2}\right)\mathcal{F} \tag{3.24}$$

such that

$$\psi_\omega(x, y, z = L) = \mathcal{D}_L \psi_\omega(x, y, z = 0) \tag{3.25}$$

where the subscript L denotes the propagation distance.

It is important to emphasize here that the wavefunction $\psi_\omega(x, y, z = 0)$ at the beginning is arbitrary. Therefore, this may be for instance the wave conformation after it has interacted with matter, where $z = 0$ would be the output plane of a sample. As we shall see in section 3.1.3 ahead, the interaction of an incident wave with matter will produce such an arbitrary output wave with amplitude and phase modulations with respect to the incident one. In the particular case where the incident wave is a perfect plane wave propagating in the z direction, the operator $\mathcal{D}_L$ may be directly applied to the transfer function of the sample. This may be a complex sample, causing both attenuation and phase-shits, as well as a simple slit that blocks part of the incident radiation. This interaction is actually a diffraction phenomena, and therefore $\mathcal{D}_L$ is also referred to as a *diffraction operator*.

The Fresnel approximation may be once again applied here to the square-root term in equation 3.24, such that we get the Fresnel approximation $\mathcal{D}^f$ of the diffraction operator when $N_f << 1$:

$$\begin{aligned}\psi_\omega(x, y, z = L) &\approx \mathcal{D}_L^f \times \psi_\omega(x, y, z = 0) \\ &= \exp(ikL)\mathcal{F}^{-1} \exp\left[\frac{-iL(k_x^2 + k_y^2)}{2k}\right]\mathcal{F} \times \psi_\omega(x, y, z = 0)\end{aligned} \tag{3.26}$$

as well as $\mathcal{D}_L^F$ for the Fraunhoffer approximation when $N_F << 1$:

$$\begin{aligned}\psi_\omega(x, y, z = L) &\approx \mathcal{D}_L^F \times \psi_\omega(x, y, z = 0) \\ &= -\frac{ik \exp(ikL)}{L} \exp\left[\frac{ik}{2L}(x^2 + y^2)\right] \times \hat{\psi}_\omega\left(k_x = \frac{kx}{L}, k_y = \frac{ky}{L}, z = 0\right)\end{aligned} \tag{3.27}$$

$\hat{\psi}_\omega$ being the Fourier Transform of ψ_ω. Note that the Fourier Transform is calculated at frequency points $k_i = i/L$, , $i = \{x, y, z\}$[5]. Therefore, equation 3.27 tells us that,

[5] Rigorously speaking, $\sin(\theta_i) = k_i/k$, with $k^2 = k_x^2 + k_y^2 + k_z^2$ and θ_i being the angle the component k_i makes with the propagation direction. Furthermore, simple trigonometry gives us $\tan(\theta_i) = i/L$. Combining these equations we obtain

$$k_i = k \sin(\arctan(i/L)) \approx \frac{ki}{L} = \frac{2\pi i}{\lambda L} \tag{3.28}$$

the approximation being valid when we have $\theta_i << 1$ (paraxial approximation).

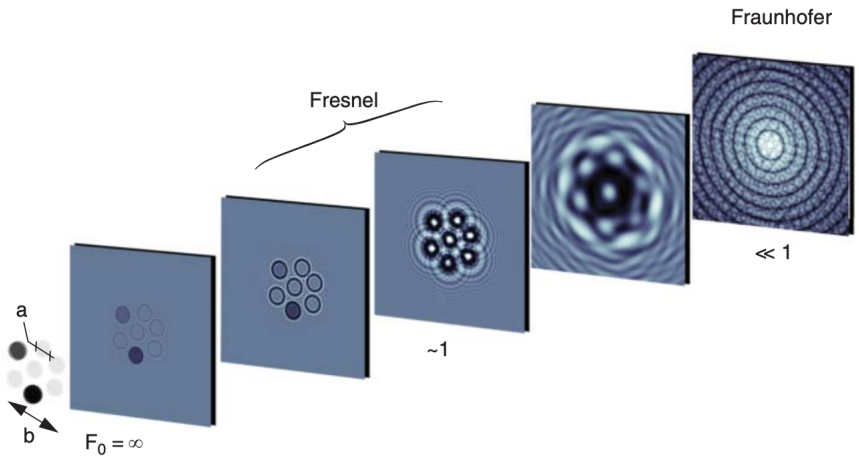


Figure 12 – The different diffraction regimes for an object composed of absorptive (dark) and phase-shifting (light) discs. Figure from [5].

for $N_F << 1$ (usually meaning long propagation distances), the wave will be a direct measurement of the frequency-space of the input complex amplitude, each point in the output plane representing a specific frequency of the plane waves composing the signal.

To summarize, we illustrate in Figure 12 the different diffraction regimes for an object composed of absorptive and phase-shifting discs. For small propagation distances (big N_F), the wave conformation resembles a projection of sample in the so called *contact regime*. As propagation distances increases such that $N_F \approx 1$, one enters the Fresnel regime where interference between different plane wave components start to happen, causing the appearance of fringes around the object. At big distances ($N_F << 1$), each plane wave component gets completely separated and one measures the diffraction pattern.

3.1.3 Waves in matter and the Projection Approximation

We now proceed to the modelling of a wave propagating through a material medium. The process is analogous to the one presented for the vacuum equations, where we reach a more refined version of the Helmholtz equation 3.6. Under both the paraxial approximation already discussed and the important Projection Approximation (PA), this equation presents a simple solution. Nonetheless, the validity of such approximation is questionable under certain circumstances, which is the central problem this work has tried to improve upon.

To arrive at the Projection Approximation, one starts from Maxwell's equations in material media and makes a series of assumptions about our sample [10]:

- Linear isotropic material ($\vec{D} = \varepsilon\vec{E}$ and $\vec{B} = \mu\vec{H}$);
- Static material (ε and μ independent of time);

- Non-magnetic material ($\mu(x, y, z) = \mu_0$);
- Neither current nor charge densities are present ($\rho(x, y, z, t) = \vec{J}(x, y, z, t) = 0$);
- Scatterers are sufficiently slowly varying over length scales comparable to the wavelength of the X-ray radiation.

These assumptions allows once again to describe the electromagnetic field by the scalar theory, where the complex scalar field $\Psi(x, y, z, t)$ now obeys a slightly modified version of equation 3.4:

$$\left[\varepsilon(x, y, z)\mu_0 \frac{\partial^2}{\partial t^2} - \nabla^2\right] \Psi(x, y, z, t) = 0 \tag{3.29}$$

The complex wavefunction can be decomposed into monochromatic components (as in equation 3.5), which when substituted into 3.29 provides us (after some algebra) with the "inhomogenous"[6] Helmholtz equation

$$[\nabla^2 + k^2 n_\omega^2(x, y, z)]\psi_\omega(x, y, z) = 0 \tag{3.30}$$

The above equation differs from 3.6 only by the position and frequency dependent $n_\omega^2(x, y, z)$ refractive index term that indicates the presence of the material medium.

Now, suppose we have an incident plane wave from the left reaching a scattering volume contained in the region where $0 < z < z_0$, as shown in Figure 13. We propose the following ansatz to the above equation:

$$\psi_\omega(x, y, z) = \psi(x, y, z) \exp(ikz) \tag{3.31}$$

That is, once again an unscattered plane wave $\exp(ikz)$ modulated by an envelope $\psi(x, y, z)$. Substituting it into equation 3.30, we arrive at

$$\left[2ik\frac{\partial}{\partial z} + \nabla_\perp + \frac{\partial^2}{\partial z^2} + k^2[n^2(x, y, z) - 1]\right] \psi(x, y, z) = 0 \tag{3.32}$$

where we have split the Laplacian $\nabla = \nabla_\perp + \dfrac{\partial^2}{\partial z^2}$ into its transversal and longitudinal components. Within the paraxial approximation mentioned in section 3.1.1, we can neglect the second derivative in z to arrive at the *inhomogeneous paraxial Helmholtz* equation:

$$\left[2ik\frac{\partial}{\partial z} + \nabla_\perp + k^2[n^2(x, y, z) - 1]\right] \psi(x, y, z) = 0 \tag{3.33}$$

[6] This is said to be inhomogeneous in the sense that we are dealing with an inhomogeneous medium, not in the traditional mathematical sense of inhomogenous equation.

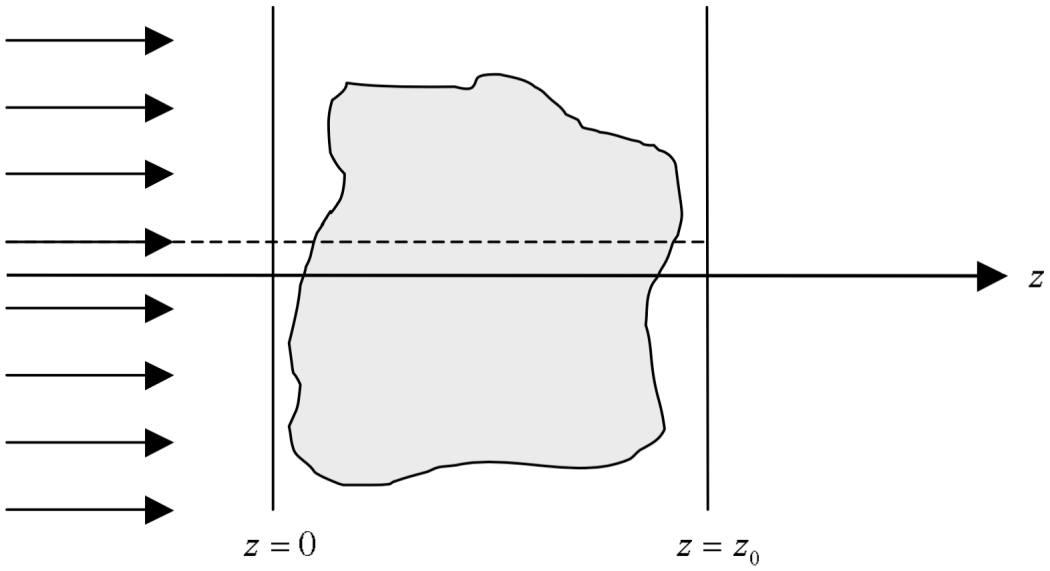


Figure 13 – Incident plane wave reaching a scattering object from the left. Only a "projected ray-path" entering the sample at $(x, y, 0)$ and leaving at (x, y, z_0) is responsible for influencing phase and absorption changes to the complex amplitude of the wave. In other words, the only regions of the sample that influence the ray are those that the wave would have travelled in vacuumm, had the object been absent. Figure from [10].

In vacuum, $n = 0$ and 3.33 reduces to the classic *paraxial Helmholtz* equation, for which the Gaussian beam is one of its solutions.

We subsequently perform the Projection Approximation. It consists in assuming that scattering is sufficiently weak so that the wavefield at the exit surface ($z = z_0$) will be determined only by amplitude and phase shifts accumulated along the "projected" ray-paths between the input and output planes. In other words, a ray entering the sample at $(x, y, 0)$ and leaving at (x, y, z_0) will be influenced only by those points through which the wave would have travelled in vacuumm, were the sample absent [10]. In the above equation, this means neglecting the transverse Laplacian, which is the term that couples neighboring trajectories. We are then left to solve a simple partial differential equation

$$\frac{\partial}{\partial z}\psi(x, y, z) \approx \frac{k}{2i}[1 - n^2(x, y, z)]\psi(x, y, z) \tag{3.34}$$

Therefore, the wave-field envelope at the exit surface is

$$\psi(x, y, z = z_0) \approx \exp\left(\frac{k}{2i}\int_{z=0}^{z=z_0}[1 - n^2(x, y, z)]dz\right)\psi(x, y, z = 0) \tag{3.35}$$

For frequencies in the X-ray range, the refractive is slightly lower than unity. It is then convenient to rewrite it as

$$n = 1 - \delta + i\beta \tag{3.36}$$

with both δ and β being real values. Kepping only first-order terms in δ and β we obtain

$$1 - n^2 \approx 2(\delta - i\beta) \tag{3.37}$$

which allows to rewrite the wavefield at the exit plane $z = z_0$ as

$$\begin{aligned} \psi(x, y, z = z_0) &\approx \exp\left(-ik \int_{z=0}^{z=z_0} (\delta(x, y, z) - i\beta(x, y, z))dz\right) \psi(x, y, z = 0) \\ &= T_{PA}(\delta, \beta, z_0) \times \psi(x, y, z = 0) \end{aligned} \tag{3.38}$$

T_{PA} works in analogous manner to the transmission function of equation 3.23. Such output wavefunction may then be further propagated using the diffraction operators 3.24, 3.26 and 3.27. Note that the argument ϕ of the exponential term is

$$\begin{aligned} \phi &= -ik \int_{z=0}^{z=z_0} (\delta(x, y, z) - i\beta(x, y, z))dz \\ &= -ik \int_{z=0}^{z=z_0} \delta(x, y, z)dz - k \int_{z=0}^{z=z_0} \beta(x, y, z)dz \end{aligned} \tag{3.39}$$

Therefore, equation 3.38 may be rewritten as

$$\psi(x, y, z = z_0) \approx \exp\left(-ik \int_{z=0}^{z=z_0} \delta(x, y, z)dz\right) \exp\left(-k \int_{z=0}^{z=z_0} \beta(x, y, z)dz\right) \psi(x, y, z = 0) \tag{3.40}$$

Note that the first exponential in the above equation is imaginary, while the second is real. Therefore, we see that the real part δ of the refractive index relates to a phase shift (i.e. refraction) of the wave as it propagates through the object, whereas the imaginary part β relates to the attenuation. We will further discuss this relationship in section 3.2.7.

3.2 Scattering of Radiation

In this section, we provide a more detailed description on how radiation gets scattered by matter. This will show us how the different physical quantities are related to each other, allowing us to conclude what information we try to recover in imaging experiments with X-rays. In other words, we will link how an intensity measurements from the experiments can be related to the refractive index of the sample and, in the particular case where absorption is absent, to its electron density.

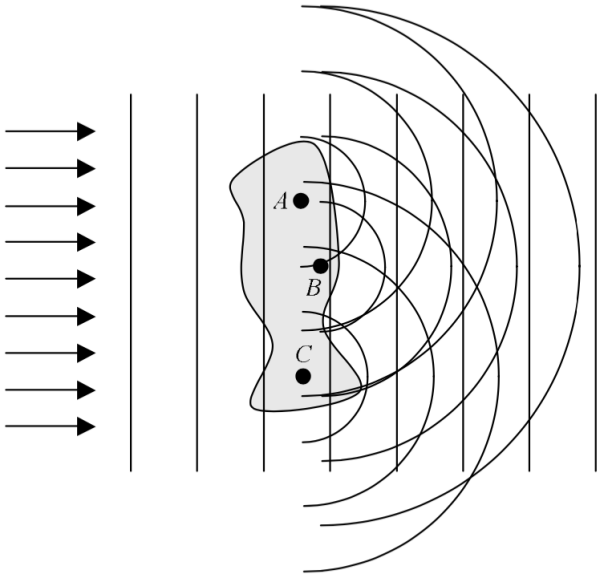


Figure 14 – Scheme illustrating the scattering of secondary spherical wavelets within the First Born Approximation. Figure from [10].

3.2.1 First Born Approximation

Our model works within the first Born Approximation. In essence, it states that we have *weak scattering*. As illustrated in Figure 14, given an incident plane wave, the total scattered wave will be the result from the interference of the incident wave ψ_{inc} and the secondary spherical wavelets originating from each scattering source. The weak scattering relies on assuming that no secondary scattering events will happen and that the incident wave barely changes upon interaction. In other words, a scattered wave exiting the sample will have interacted with only a single scattering source. Naturally, as the size of the sample increases, the wave is also more likely to interact multiple times with the sample and, therefore, the validity of the approximation is put under question.

This is indeed a reasonable assumption. As we have previously shown (Figure 3), the scattering cross section in an X-ray experiment is orders of magnitude lower than the absorption cross section for most of the energy range of interest. This indicates that the vast majority of photons that interact a second time with the sample will be absorbed instead of being scattered once again. Photons that indeed managed to escape the sample are more likely to have been scattered only once.

Quantum mechanics shows us how the first Born approximation gives a simple yet powerful result when modelling the interaction of the wave with matter [19, 20]. Starting from the time dependent Schrodinger's equation, with $V(r)$ representing a potential of finite range caused by the distribution of electronic charge density (in the case of X-rays), we have:

$$\left[\frac{\hbar^2}{2m}\nabla^2 + V(r)\right]\psi(\vec{r}) = E\psi(\vec{r}) \tag{3.41}$$

A following solution of the form exists, consisting in the combination of an incident plane wave with scattered spherical wave (Huygens principle [14]):

$$\psi(\vec{r}) = \psi_{\text{inc}} + \psi_{\text{scattered}} = e^{i\vec{k}\vec{r}} + f(\theta,\phi)\frac{e^{ikr}}{r} \tag{3.42}$$

where $f(\theta,\phi)$ is called the *scattering factor*. If the scattering were isotropic, we would expect a spherically symmetrical outgoing wave in which this factor would not depend on the angles. The scattering factor can then be thought of as a measure of the anisotropy of the scattering event. Using Green's function formalism [21] and assuming the First Born approximation, this term is found to be

$$f(\theta,\phi) = \frac{-2m}{\hbar^2}\frac{1}{4\pi}\mathcal{F}\{V(r)\} \tag{3.43}$$

where $\mathcal{F}$ denotes the Fourier Transform[7].

In an experiment, the potential $V(r)$ is what contains the information about our sample. If the scattering amplitude is what tells us something about the potential, we must connect $f(\theta,\phi)$ to a measurable quantity. As we shall see ahead, this connections happens precisely through the *differential scattering cross-section*.

3.2.2 Scattering Factor and the Cross-section

As introduced in section 1.3, the total atomic cross-section σ_T is the effective area the material has to scatter the incident radiation. It is defined as the ratio between the average scattered radiation power $\overline{P}_{\text{scatt}}$ by the average incident power $\overline{P}_{\text{inc}}$:

$$\sigma_T = \frac{\overline{P}_{\text{scatt}}}{\overline{P}_{\text{inc}}} \tag{3.45}$$

In practice, one cannot measure the photons scattered to all directions, but instead uses a detector that covers a region in space. If we consider this region to be sub-intended by a solid angle $d\Omega$ and measure the number of photons that reach this region per unit time normalized by the incident flux (Φ_{inc}), we have the *differential scattering cross-section*, which may also be described in terms of the radiated and incident electric field as [17]:

$$\frac{d\sigma}{d\Omega} = \frac{I}{\Phi_{\text{inc}}\Delta\Omega} = \frac{|E_{\text{rad}}|^2R^2}{|E_{\text{inc}}|^2} \tag{3.46}$$

[7] Note that we have an angular dependence of θ and ϕ on the left hand side that does not appear in the right hand side. Actually, this Fourier Transform is written as

$$f(\theta,\phi) = \frac{-1}{4\pi}\int e^{-i\boldsymbol{Q}\cdot\boldsymbol{r}}V(\boldsymbol{r})d^3r \tag{3.44}$$

that is, the Fourier Transform is evaluated at the momentum transfer $\boldsymbol{Q}$ (or scattering vector). This is where the angles θ and ϕ are "hidden".

The total cross-section can be calculated by integrating over the entire sphere:

$$\sigma_T = \int_{4\pi} \left(\frac{d\sigma}{d\Omega} \right) d\Omega \tag{3.47}$$

Given an incident sinusoidal plane wave of the form

$$\mathbf{E}(\mathbf{r}, t) = \mathbf{E}_0 e^{-i(\omega t - \mathbf{k}\cdot\mathbf{r})} \tag{3.48}$$

we will show how different structures scatter the radiation, giving different results for the scattering cross-section. For such, we need first to remind ourselves that accelerated charges *radiate*. The electric field caused by an accelerated charge is given by

$$\mathbf{E}(\mathbf{r}, t) = \frac{e}{4\pi\varepsilon_0 c^2} \frac{\mathbf{a}_T(t - r/c)}{r} \tag{3.49}$$

where $\mathbf{a}_T$ is the acceleration component in the transverse direction with respect to the propagation direction of the electric field. It can be shown that the *total power radiated* by the accelerated electron will then be [16]

$$P = \frac{8\pi}{3} \left(\frac{e^2 |\mathbf{a}|^2}{16\pi\varepsilon_0 c^3} \right) \tag{3.50}$$

Therefore, by calculating the **acceleration** of the excited electron, one can obtain the scattering cross-section using equation 3.45. In the next sections we shall see how different structures scatter the incident radiation, namely: the *free electron*, a simple *two-electron* structure, an *atom* containing a cloud of free-electrons and, finally, the most realistic case of an atom containing *bound-electrons*.

3.2.3 Scattering: Free electron

Consider first a free-electron and the incident plane wave. The force acting on this electron will be given by the Lorentz force

$$\mathbf{F}_{\text{driving}} = e(\mathbf{E_i} + \mathbf{v} \times \mathbf{B_i}) \tag{3.51}$$

For non-relativistic velocities, the magnetic force is negligible in comparison with the electric. Hence, from Newton's second law, we get

$$\mathbf{a}(\mathbf{r}, t) = -\frac{e}{m} \mathbf{E_i}(\mathbf{r}, t) \tag{3.52}$$

Considering that the wave direction makes an angle Θ with respect to the electron acceleration, we have $a_T = a \sin(\Theta)$. Then, from equation 3.49:

$$\mathbf{E}(\mathbf{r},t) = -r_e E_i \frac{e^{-i\omega(t-r/c)}}{r} \sin(\Theta) \tag{3.53}$$

where r_e is the classical electron radius

$$r_e = \frac{e^2}{4\pi\varepsilon_0 m_e c^2} = 2.82 \times 10^{-5}\ \text{Å} \tag{3.54}$$

From the formula for the differential scattering cross-section (equation 3.46) we obtain

$$\frac{d\sigma}{d\Omega} = r_e^2 \sin^2(\Theta) \tag{3.55}$$

where the $\sin^2(\Theta)$ is called the polarization factor. On the other hand, equations 3.45 and 3.50[8] can be used to calculate the total scattering cross-section of the free-electron.

$$\sigma_T = \frac{8\pi}{3} r_e^2 = 0.665 \quad \text{barn} \tag{3.56}$$

For other electronic configurations with multiple electrons, the *scattering factor* $f(Q)$ consists in a term that relates the electric field scattered by the respective configuration in comparison with the free electron. In other words, the scattering factor indicates both the anisotropy of the scattering (due to its angular dependence) as well as the strength in comparison with the free-electron scenario.

3.2.4 Scattering: Two electrons

Since we assume the electron to be structureless, we start of with the simplest possible structure in order to arrive at the atomic scattering factor: a pair of electrons. Consider the case presented in Figure 15 (left), where one electron is at the origin and the other is at position $\mathbf{r}$. We consider the scattering to be elastic, where the scattered k' and incident k wavevectors possess the same magnitudes.

$$\Delta\phi(\mathbf{r}) = \mathbf{k}\cdot\mathbf{r} - \mathbf{k'}\cdot\mathbf{r} = (\mathbf{k}-\mathbf{k'})\cdot\mathbf{r} = \mathbf{Q}\cdot\mathbf{r} \tag{3.57}$$

$\mathbf{Q}$ is called the *scattering vector*.[9] Since scattering is elastic ($k = k'$) the scattering vector has magnitude

$$Q = 2k\sin\theta = \frac{4\pi}{\lambda}\sin\theta$$

[8] The power per unit area for the incident wave can be shown to be [16]:

$$\overline{S}_{\text{inc}} = \frac{1}{2}\sqrt{\frac{\varepsilon_0}{\mu_0}}|\mathbf{E_i}|^2\mathbf{k_0}$$

with k_0 the wavenumber of the incident field.

[9] We further assume that the measurements are perform at the far-field limit, i.e. we operate in the Fraunhoffer regime.

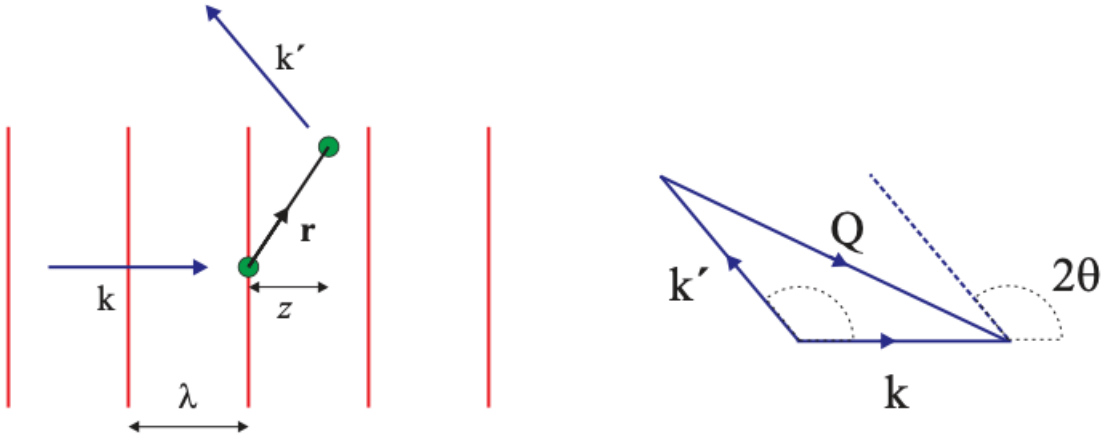


Figure 15 – Two-electron system consists in a base case for understanding the multiple electron atom. Figure from [17].

where 2θ is the angle between $\mathbf{k}$ and $\mathbf{k}'$, called the scattering angle. Assuming the polarization factor from the previous section to be unity, we have that the scattering amplitude for the two-electron system is

$$A(\mathbf{Q}) = -r_e - r_e e^{-i\mathbf{Q}\mathbf{r}} = -r_e(1 + e^{-i\mathbf{Q}\mathbf{r}})$$

where the exponential phase factor for the second electron arises because its position r with respect to the one at the origin (equation 3.57). Consequently, the intensity $I(\mathbf{Q}) = |A(\mathbf{Q})|^2$ equals

$$I(\mathbf{Q}) = A(\mathbf{Q})A^*(\mathbf{Q}) = 2r_e^2(1 + \cos(\mathbf{Q} \cdot \mathbf{r})) \tag{3.58}$$

The extension of the above amplitude for an ensemble of N electrons is straightforward:

$$A(\mathbf{Q}) = r_e \sum_{j=1}^{N} e^{-i\mathbf{Q}\cdot\mathbf{r}_j} \tag{3.59}$$

where $\mathbf{r}_j$ is the position of the j^{th} electron. To plot equation 3.58, one needs to specify the angles between the two vectors. For many systems, the position $\mathbf{r}$ of the scatterer is oriented randomly for a given $\mathbf{Q}$. In theses cases, an orientational averaging is performed by integrating the equation over all possible angles.

3.2.5 Scattering: Atom

For an atom, we consider a continuous electronic cloud of density $\rho_e(\mathbf{r})$ surrounding the nucleus such that, in a volume $d\mathbf{r}$, we have $\rho_e(\mathbf{r})d\mathbf{r}$ electrons. The integral over the whole volume should give the total number of electrons Z. The total scattering amplitude is then the continuous case of equation 3.59, where we consider that each point of the cloud at position $\mathbf{r}$ will contribute with a phase factor $e^{-i\mathbf{Q}\cdot\mathbf{r}}$ such that

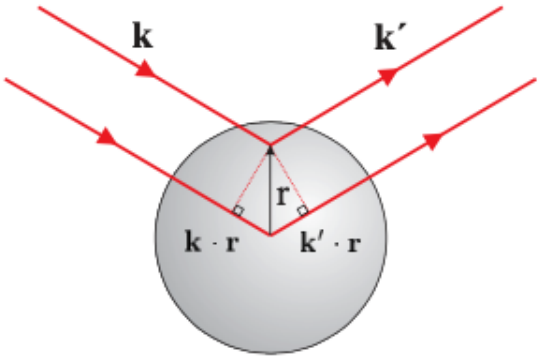


Figure 16 – For an atom, we consider a continuous electronic cloud where each point contributes to a phase shift $e^{-i\mathbf{Q}\cdot\mathbf{r}}$. Figure from [17].

$$f^0(\mathbf{Q}) = \int \rho(\mathbf{r}) e^{-i\mathbf{Q}\cdot\mathbf{r}} d\mathbf{r} \tag{3.60}$$

Note that we are missing the $-r_e$ factor[10], meaning we are writing $f^0(\mathbf{Q})$ in units of the free-electron scattering length. $f^0(\mathbf{Q})$ is also called the atomic *form factor*. Furthermore, note that

$$f^0(\mathbf{Q}) = \begin{cases} Z & \text{for } Q \to 0 \\ 0 & \text{for } Q \to \infty \end{cases}$$

The first case is a direct consequence of the definition of the electronic density: the integral of the density over the whole volume equals the number of electrons. The second case indicates that, as Q gets bigger (meaning the wavelength gets smaller), the phase factor gets large, such that any small variation in $\mathbf{r}$ will cause a big overall phase fluctuations. This amounts to a destructive interference happening for the waves scattered from the different electrons of the atom, nulling the form factor.

Spherical atomic form factor

We treat here a special case of a spherical electronic density, since this geometry is used as a study case for the simulations performed in this work. Consider the electronic density to be constant in a spherical volume of radius R. In that case, we can calculate the atomic form factor analytically through the integral in equation 3.60, obtaining

$$f^0(\mathbf{Q}) = 3\left[\frac{\sin(QR) - QR\cos(QR)}{(QR)^3}\right] \tag{3.61}$$

Figure 17 (left) shows the two dimensional plot of the above equation for spatial frequencies q_x, q_y, where $Q = \sqrt{q_x^2 + q_y^2}$. We consider a sphere of radius $R = 100\,\text{nm}$ with $3\,\text{keV}$ energy for a detector containing $N = 1024$ squared pixels of size $p = 55\,\mu\text{m}$ placed at a distance

[10] Or, more precisely, the $-r_e|\sin(\Theta)|^2$ factor

$L = 7.98\,\mathrm{m}$ from the sample. The spatial frequencies are calculated according to the frequencies values expected from the Fraunhoffer regime for small angles as

$$Q = \frac{2\pi}{\lambda}\frac{Npr}{L} \tag{3.62}$$

where r is the distance from the optical axis at the detector plane.

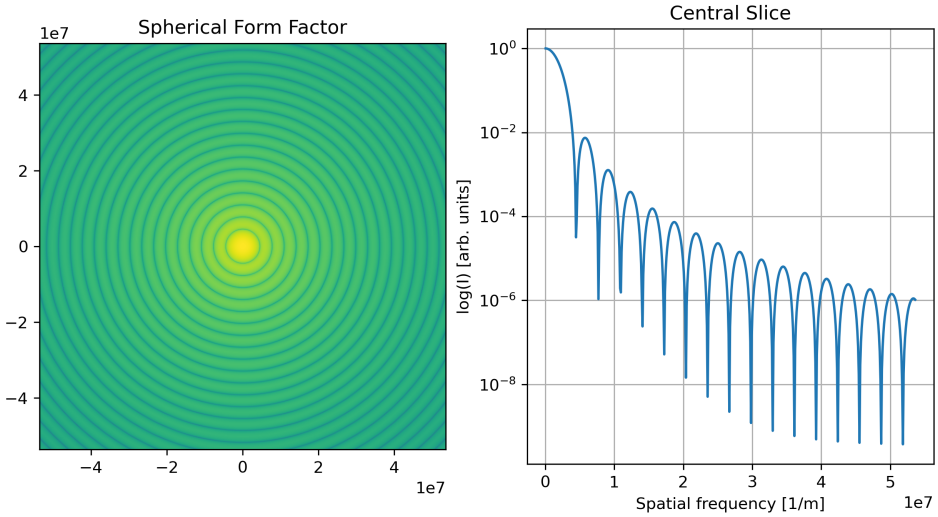


Figure 17 – Intensity I of the spherical form factor for a sphere of constant density (left) and the corresponding central slice in the horizontal direction showing the normalized intensity profile.

The analytical form factor above is precisely the diffraction pattern expected to be measured from a spherical sample. In principle, this allows us to perform a CDI experiment simulation starting from a "perfect" analytical diffraction pattern, without errors from numerical calculations.

3.2.6 Resonant Scattering: bounded-electrons

Close to absorption energies, we can no longer consider that the electrons will respond freely to the incident radiation. In that case, a useful approach is to model the electron-radiation interaction as a damped-driven harmonic oscillator [16, 18].[11] In this model, we consider that three forces act upon the electron. First, there is the binding force between the nucleus and the electron of the form

$$\mathbf{F}_{\text{binding}} = -k\mathbf{x} = -m\omega_0^2\mathbf{x} \tag{3.63}$$

that is, as a spring restoration force. Second, we have the damping force

[11] The quantum model is shown to give analogous, although more complete results. We refer the reader to [13] for such details.

$$\mathbf{F}_{\text{binding}} = -m\gamma\frac{d\mathbf{x}}{dt} \tag{3.64}$$

which in part can be attributed to the energy loss due to the radiation of the accelerated charge. At last, there is the driving Lorentz force from the incident field:

$$\mathbf{F}_{\text{driving}} = e(\mathbf{E_i} + \mathbf{v} \times \mathbf{B_i}) \tag{3.65}$$

Once again, the magnetic term of the Lorentz force is considered negligible for non-relativistic velocities. As a result of the oscillating electric field, $\mathbf{F}_{\text{driving}(t)} = e\mathbf{E_i}\exp(i\omega t)$ [12], and Newton's second law gives

$$m\frac{d^2x}{dt^2} = F_{\text{total}} = F_{\text{binding}} + F_{\text{damping}} + F_{\text{driving}} \tag{3.66}$$

In the steady-state, the system will oscillate together with the driving frequency: $X(x,t) = x\exp(i\omega t)$. Substituting the corresponding terms and applying the derivatives, we get

$$\mathbf{x} = \frac{1}{\omega^2 - \omega_0^2 + i\gamma\omega}\frac{e\mathbf{E_i}}{m} \tag{3.67}$$

which gives an acceleration

$$\mathbf{a} = \frac{-\omega^2}{\omega^2 - \omega_0^2 + i\gamma\omega}\frac{e\mathbf{E_i}}{m} \tag{3.68}$$

Following the same procedure as before for the free-electron, this corresponds to a scattering cross-section

$$\sigma_T(\omega) = \frac{8\pi}{3}r_e^2\frac{\omega^4}{(\omega^2 - \omega_0^2)^2 + (\gamma\omega)^2} \tag{3.69}$$

Different from the free-electron case, where the cross-section was independent of the radiation energy, the cross-section for the bound electron will depend on the frequency ω, presenting a maximum peak value for the resonant frequency $\omega = \omega_0$.

To generalize the above result for a multi-electron atom, the total electric field scattered from all the electrons is given by

$$\mathbf{E}(\mathbf{r},t) = \frac{e}{4\pi\varepsilon_0 c^2}\sum_{s=1}^{Z}\frac{\mathbf{a_{T,s}}(t - r_s/c)}{r_s} \tag{3.70}$$

where the subscript s denotes the individual acceleration and position of the s^{th} electron. The acceleration is analogous to the one presented in equation 3.68, but presents now an extra phase factor that accounts for the position $\Delta\mathbf{r}_s$ of each electron:

$$\mathbf{a_s} = \frac{-\omega^2}{\omega^2 - \omega_s^2 + i\gamma\omega}\frac{e}{m}\mathbf{E_i}e^{-i(\omega t - \mathbf{Q}_i\cdot\Delta\mathbf{r}_s)} \tag{3.71}$$

[12] We model the equation using the complex exponential, whose real part should be considered the result.

where ω_s is the resonant frequency of the s^{th} electron. Plugging this expression in equation 3.70 we obtain for the electric field:

$$E(\mathbf{r},t) = -\frac{r_e}{r} f(\mathbf{Q},\omega) E_i \sin\Theta e^{-i\omega(t-r/c)} \tag{3.72}$$

where

$$f(\mathbf{Q},\omega) = \sum_{s=1}^{Z} \frac{\omega^2 e^{-i\mathbf{Q}\cdot\mathbf{\Delta r_s}}}{\omega^2 - \omega_s^2 + i\gamma\omega} \tag{3.73}$$

is the *complex atomic scattering factor*. With the electric field, one can once again obtain the differential and total scattering cross-sections for the bounded-electrons case:

$$\frac{d\sigma(\omega)}{d\Omega} = r_e^2 \sin^2(\Theta) |f(\mathbf{Q},\omega)|^2 \tag{3.74}$$

$$\sigma_T(\omega) = \frac{8\pi}{3} r_e^2 |f(\mathbf{Q},\omega)|^2 \tag{3.75}$$

Comparing equation 3.55 with 3.74 and equation 3.56 with 3.75, we see that they differ only by a $|f|^2$ factor which depends on the frequency. The particular limiting case of scattering in the forward direction ($\theta = 0$) allows us to simplify equation 3.73 to

$$f^0(\omega) = \sum_{s=1}^{Z} \frac{\omega^2}{\omega^2 - \omega_s^2 + i\gamma\omega} \tag{3.76}$$

where the 0 superscript is used here to denote the forward scattering.

A more precise variation of the above equation introduces the *oscilattor strengths* g_s which account for the number of electrons associated with resonant frequency ω_s, where

$$\sum_{s}^{Z} g_s = Z \tag{3.77}$$

Therefore, the complex atomic scattering for the **forward scattering** case can be separated into a real f_1^0 and an imaginary f_2^0 part:

$$f^0(\omega) = \sum_{s=1}^{Z} \frac{g_s \omega^2}{\omega^2 - \omega_s^2 + i\gamma\omega} = f_1^0(\omega) - i f_2^0(\omega) \tag{3.78}$$

The real part, on the other hand, can be split into two others: the original f^0 of an unbound electron from equation 3.60 and another $f'(\omega)$ term that reduces the value of the scattering factor in comparison to the unbounded case. The total complex scattering factor then writes:

$$f(\mathbf{Q},\omega) = f_1^0(\omega) + if_2^0(\omega) = f^0(\mathbf{Q}) + f'(\omega) + if''(\omega) \tag{3.79}$$

As we shall see ahead, the above equation allows us to related the scattering factor to the real and imaginary parts of the refractive index.

3.2.7 Relationship to the Refractive Index

From Maxwell's equations in material media, we are able to deduce the following wave equation when we assume we have transverse waves and consider forward-scattering only [16]

$$\left[\frac{\partial^2}{\partial t^2} - \frac{c^2}{n^2(\omega)}\nabla^2\right]\mathbf{E_T}(\mathbf{r},t) = 0 \tag{3.80}$$

where $n(\omega)$ is recognized as a frequency dependent refractive index

$$\begin{aligned} n(\omega) &= \left[1 - \frac{e^2\rho_a}{\varepsilon_0 m}\sum_s \frac{g_s}{(\omega^2-\omega_s^2)+i\gamma\omega}\right]^{1/2} \\ &\approx 1 - \frac{1}{2}\frac{e^2\rho_a}{\varepsilon_0 m}\sum_s \frac{g_s}{(\omega^2-\omega_s^2)+i\gamma\omega} \end{aligned} \tag{3.81}$$

and ρ_a is the atomic number density. The approximation in equation 3.81 holds for X-ray frequencies, since in that case $\omega^2 >> \omega_s^2$ [1]. By rewriting it in terms of the electron radius r_e (equation 3.54) and using $\omega = 2\pi c/\lambda$, one recognizes the real and imaginary parts of the complex atomic scattering factor, obtaining

$$n(\omega) = 1 - \frac{r_e\rho_a\lambda^2}{2\pi}\left[f_1^0(\omega) - if_2^0(\omega)\right] \tag{3.82}$$

For X-ray frequencies, the refractive index is slightly lower than unity. It is then useful to rewrite it in the the complex form

$$n(\omega) = 1 - \delta + i\beta \tag{3.83}$$

By doing so, we can see how each term contributes to the interaction of the wave with a sample. Considering a plane wave propagating in a homogeneous medium:

$$E(\mathbf{r},t) = \mathbf{E_0}e^{-i(\omega t - \mathbf{k}\cdot\mathbf{r})} = \mathbf{E_0}e^{-i\omega(t-\frac{\mathbf{k}}{\omega}\cdot\mathbf{r})} \tag{3.84}$$

One can rewrite it in terms of the refractive index using $k/\omega = n/c = (1-\delta+i\beta)/c$ such that

$$E(\mathbf{r},t) = \mathbf{E_0}e^{-i\omega(t-r/c)}e^{-i2\pi\delta r/\lambda}e^{-2\pi\beta r/\lambda} \tag{3.85}$$

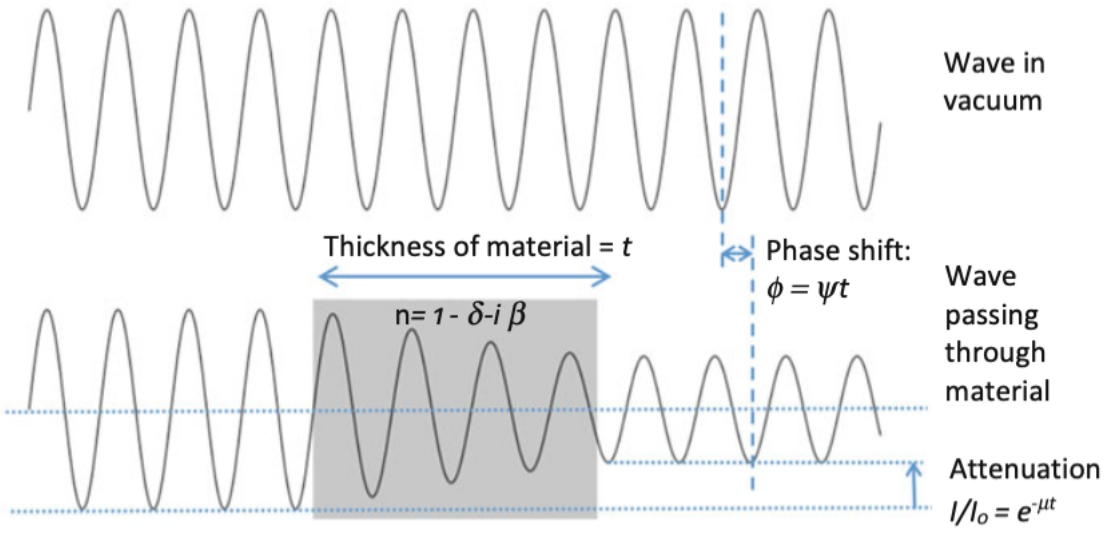


Figure 18 – Illustration of the phase-shift and absorption suffered by a wave travelling through a material. Figure from [22].

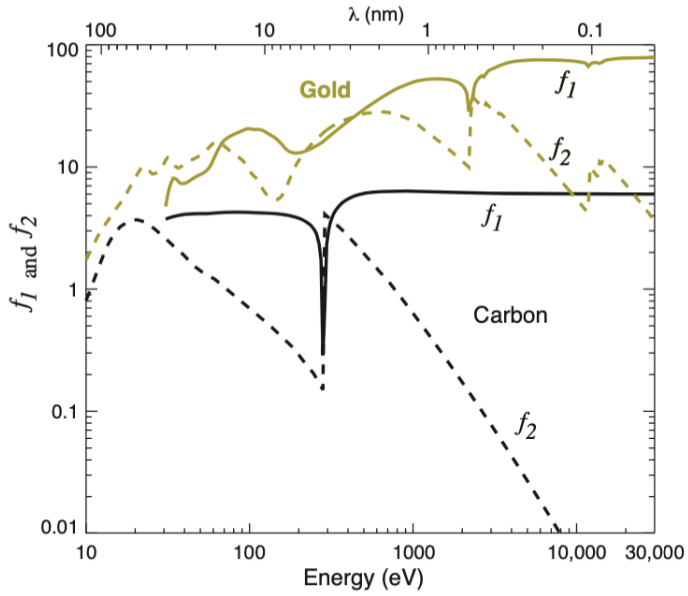


Figure 19 – Real and imaginary parts of the complex atomic scattering factor for Gold (yellow) and Carbon (black). For high-energies, that is $\beta \to 0$, we obtain $(f_1(\omega) - if_2(\omega)) \to Z$. Figure from [1].

The first exponential represents the standard propagation in vacuum. The second exponential is complex and indicates a phase shift caused by the medium, while the third (real) exponential indicates a decay. Consequently, the δ is associated to the refraction effects, while β to the attenuation in the medium, as illustrated in Figure 18.

If we now compare equations 3.82 and 3.83, it is immediate to recognize a direct relationship between the real and imaginary terms of the refractive index and of the complex atomic scattering factor:

$$\delta = \frac{r_e \rho_a \lambda^2}{2\pi} f_1^0 \qquad\qquad \beta = \frac{r_e \rho_a \lambda^2}{2\pi} f_2^0 \tag{3.86}$$

Consequently, f_1^0 and f_2^0 can also be associated with the refraction and absorption in the medium, respectively. Figure 19 shows values of f_1^0 and f_2^0 for Gold and Carbon as a function of energy. For some values, we see a sharp increases in f_2^0 (indicating absorption edges) while there is a corresponding sharp decrease in f_1^0. In particular, at high-energies $f_1(\omega)$ approaches the constant value of Z. In other words, $f'(\omega) \to 0$. At the same time, $f_2(\omega)$ decays with λ^2. In other words, electrons start to behave freely and we approach once again the behaviour described in equation 3.60. This behavior also indicates that phase-contrast imaging becomes favorable in comparison with absorption-contrast as we reach higher X-ray energies.

At last, it is worth noting that it has been experimentally verified that f_1^0 is usually positive, while f_2^0 is always positive for ordinary matter [23, 24]. This positivity has important consequences, since it serves as a constraint for the convergence of Phase-Retrieval algorithms, as shall be discussed in section 3.4.3.

3.2.8 Relationship to the Electron Density

The fact that at high-energies and away from absorption edges we have $(f_1(\omega) - if_2(\omega)) \to Z$ has one profound consequence. In this limit, equation 3.82 gives

$$n(\omega) = 1 - \frac{r_e \lambda^2 \rho_a}{2\pi} \left[f_1^0(\omega) - i f_2^0(\omega) \right] \to 1 - \frac{r_e \lambda^2}{2\pi} \rho_a Z \tag{3.87}$$

Since the electron density relates to the atomic density as $\rho_e = Z\rho_a$, we obtain a direct relationship between the electron density ρ_e and the real part of the refractive index δ:

$$\rho_e = \frac{2\pi}{\lambda^2 r_e} \delta \tag{3.88}$$

To conclude, we show in Figure 20 values of the "effective" electron density calculated using known values of δ in the above equation to recover ρ, and compare it with the real value of the electron density for the semiconductor GaAs. We see that both curves indeed agree well only within the expected energy values, that is, for those higher-energies above 1 keV and away from absorption edges.

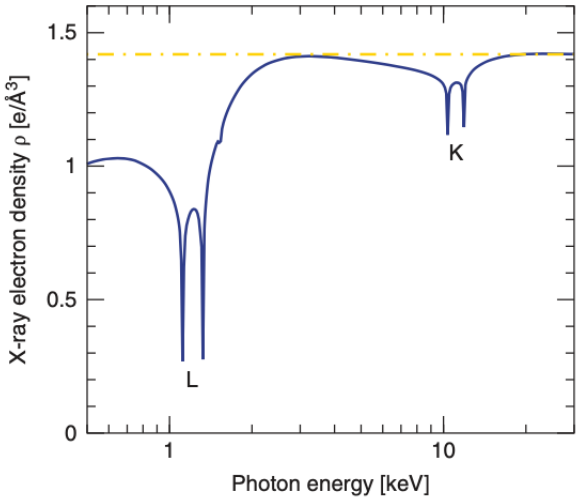


Figure 20 – Effective electron density of GaAs (black, continuous) as "felt" by the X-ray photons compared to the physical ρ_e (yellow, dashed). Figure from [5].

3.3 Imaging

3.3.1 Absorption and Phase-Contrast

In order to distinguish one object from another, imaging techniques rely on two contrast mechanisms: *absorption-contrast* and *phase-contrast*. As we have seen in equation 3.85, the real and imaginary parts of the refractive index quantify the phase-shift and attenuation of the travelling wave, respectively.

The absorption of X-rays has been present in both material science and medicine for over a hundred years now, since the discovery of X-ray radiation by Röntgen in 1895. On the other hand, X-ray phase-contrast took much longer to be discovered. As we have seen in Figure 19, the real part f_1 of the complex atomic scattering factor is greater than the complex part f_2 for most energies, especially above 1 keV. This is also shown in Figures 21 and 22, where we have the absorption and phase cross-sections both as a function of energy and atomic number. Although this strongly indicates the viability of phase imaging, this contrast mechanism only arose decades later. As pointed out by [1] in his book, page 156:

> "In hindsight it is obvious that one should exploit the phase-shifting part δ of the X-ray refractive index for high-contrast imaging, but the first clear statement of this came somewhat late in the history of the field via a conference presentation in August 1986 by Schmahl and Rudolph [Schmahl 1987], who discussed soft X-ray microscopy but also pointed towards the potential for using higher-energy X rays. (An earlier paper by Bonse and Hart on an X-ray crystal interferometer [Bonse 1965] mentioned the possibility of phase contrast X-ray

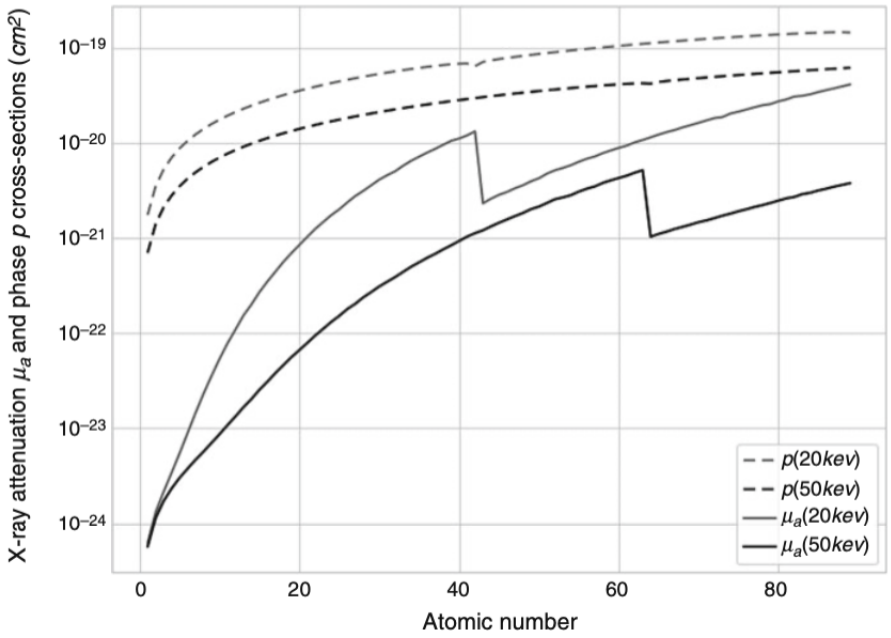


Figure 21 – Attenuation and phase cross-sections as a function of atomic number Z. Figure from [22].

> imaging with further brief comments appearing in subsequent reviews [Hart 1970b, Hart 1975], but Schmahl and Rudolph were the first to directly point out the potential for reduced radiation dose). In fact, it is truly remarkable that absorption contrast X-ray radiography has been used for over a century in medical imaging with nobody thinking of the potential of using phase contrast for lower radiation exposure – even though Einstein speculated on how $n = 1-\delta$ might produce grazing incidence reflection effects in medical imaging (Section 2.2) way back in 1918. Contemplate for a moment the collective blindness of so many X-ray scientists (myself included, in spite of work in X-ray holography around this time [Howells 1987]) for so long!"

From equations 1.2 and 3.85, one can show that the linear attenuation coefficient μ relates to the imaginary refractive index β as

$$\mu = \frac{4\pi\beta}{\lambda} \tag{3.89}$$

For a material composed of many elements, μ is calculated as a weighted sum of the mass attenuation coefficient μ/ρ of each element. Consequently, μ increases for higher-Z materials and we expect absorption contrast to be advantageous for such materials (Figure 21). As also indicated in Figure 22, working at lower energies can also be a good strategy for absorption contrast, since the phase (σ_p) and absorption (σ_a) cross-sections are similar at such energies. Furthermore, the optimized energy for transmission (where a sufficient number of photons, but not all, get absorbed) depends a lot on the material, which

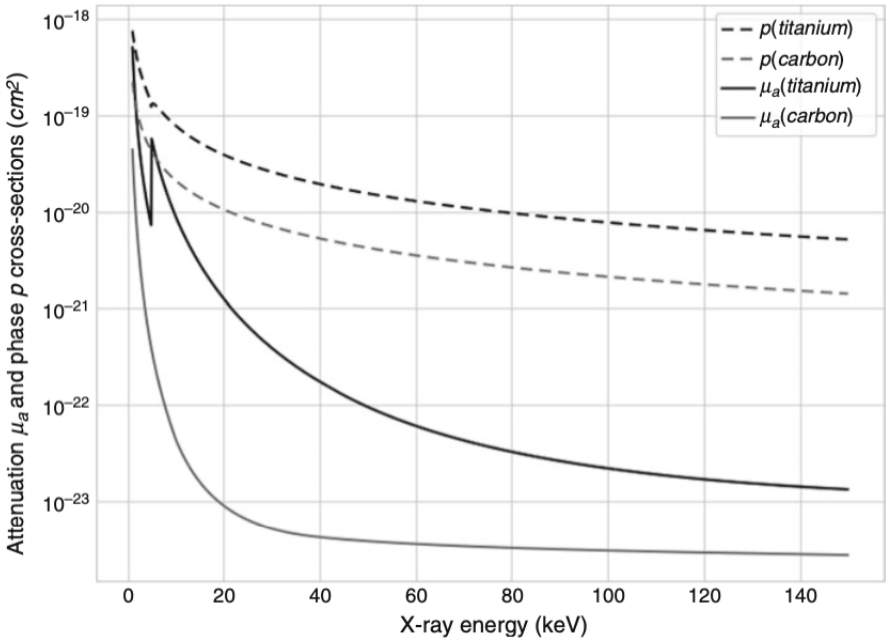


Fig. 3 Phase and attenuation cross sections for carbon and titanium versus X-ray energy

Figure 22 – Attenuation and phase cross-sections as a function of energy E for carbon and titanium. Figure from [22].

complicates imaging compound materials. Once again, phase-contrast imaging becomes the natural candidate. Furthermore, light materials present low fluorescence yield [1], essentially a measurement of how much fluorescence will occur for the element. Since both absorption and fluorescence are not advantageous for low-Z, phase-contrast is the natural candidate in this case. As shown by the plots, σ_p is not only higher than σ_a for practically all the energy range of interest, it also varies much less with both energy and atomic number, indicating the possible advantage to image a wider range of materials using phase-contrast. Added to the better spatial resolution achieved by such techniques – as is the case of CDI, it is clear that phase-contrast consists in a promising alternative to many investigations.

There are numerous imaging techniques based on phase-contrast: propagation based phase-contrast, grazing incidence imaging, Zernike phase-contrast, holography, differential phase-contrast imaging, to name a few.[13] Herein, we focus only in Coherent Diffractive Imaging, more specifically Plane Wave CDI, as it is the phase-contrast technique of interest for this work.

The equations presented in section 3.1.3 already gave us the required tools to discuss the contrast mechanisms of CDI. It is worth emphasizing that we deal with the recovery of a complex object, meaning we recover both phase and absorption information. As pointed-out with equation 3.38, within the projection approximation, the object's

[13] We refer to reader to the books of [17], [1] and [22] for a more detailed explanation of such techniques.

complex transmittance $T(x, y)$ is given by

$$T(x, y) = \exp\left(-ik \int_{z=0}^{z=z_0} (\delta(x, y, z) - i\beta(x, y, z))dz\right) \tag{3.90}$$

The presence of δ naturally indicates the existence of a phase-contrast mechanism for the technique. $T(x, y)$ can be thought of as a "projected thickness", that is, the total phase-shift $\delta_t(x, y)$ and absorption $\beta_t(x, y)$ imparted to the wave after it propagates through a material of thickness z_0 is equivalent to that of a material whose thickness were to be projected to single slice that would impart the same phase-shift and absorption, $\delta_t(x, y)$ and $\beta_t(x, y)$.

As we shall see later on, this work focuses on modelling the interaction between matter and radiation beyond the projection approximation for energies far from absorption lines, meaning and we have purely a phase-contrast mechanism due to negligible absorption. In that case, the idea of a projected thickness breaks down. Nonetheless, we will see that the phase-shift mechanism is expanded so that multiple regions beyond the projected ray (illustrated in Figure 13) influences the wave at a specific point (x, y, z_0) of the output plane.

3.3.2 X-ray Diffraction

X-ray diffraction has been used for recovering the structure of crystals for over a hundred years now and, although one does not directly recover an image of the sample, we get enough information to infer the crystal structure.

When an incident X-ray beam hits a crystalline structure, part of the radiation undergoes reflection by the crystalline planes. Although the efficiency of such reflection is low (weak scattering), the reflected waves from adjacent planes separated by a certain distance d will interfere constructively since the optical path difference of each wave equals an integer number of wavelengths λ, as illustrated in Figure 23. These planes will differ according to the incidence angle θ_B with respect to the crystalline planes. This gives rise to the famous "Bragg's law"

$$n\lambda = 2d \sin(\theta_B) \tag{3.91}$$

where n is an integer.

The constructive interference from the reflected waves will then be intense enough for one to measure sharp diffraction peaks at a detector. As it turns out, such model is analogous to the one presented in section 3.2.5. The peaks are also a measurement of reciprocal space, that is, the measured intensity is also the absolute square of the Fourier Transform of the electronic density:

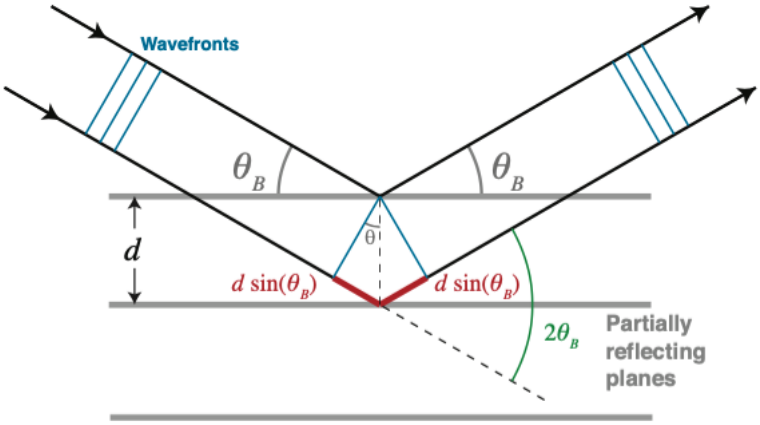


Figure 23 – Illustration of the optical path difference from rays reflected by adjacent crystalline plane. Those planes separated by a distance such that the path differs by an integer number of wavelengths will result in constructive. Figure from [1]

$$I = |\mathcal{F}[\rho(\mathbf{r})]|^2 \tag{3.92}$$

The position of such sharp peaks can then be used to retrieve the relative positions of atoms in the crystalline lattice. Note, however, that the intensity measurement is also a case of the phase problem already mentioned. Nonetheless, a solution to the phase problem in crystallography arose much earlier than for CDI, since the periodicity of the sample helps providing additional information to circumvent the phase problem.[14]

Nonetheless, crystallography relies on the existence of periodic structures. Simple molecules, virus capsids and even some macro-molecules can be crystallized and studied with such technique, Nonetheless, this applies only to a fraction of samples of interest in science [1]. In fact, biological samples tend to become more and more heterogeneous as physical dimensions increases. Such is the case for eukaryotic cells, for instance.

When the sample lacks regularity required for crystallography to be used, one must rely on the coherent diffraction of non-periodic objects. Given enough photon flux, a diffraction pattern may be obtained even for such structures, with the difference that it will not present sharp peaks as in the case of crystallography. Instead, the diffraction pattern becomes continuous. Figure 24 compares how the diffraction pattern evolves as we increase the number of repeated structures. Note from the insets on the bottom-left that the "lumpiness" of the diffraction pattern increases as we reduce the number of objects.

If one is able to obtain a diffraction pattern even for incoherent radiation, it

[14] We refer the reader to the book of [5] for explanations about phase recovery in crystallography.

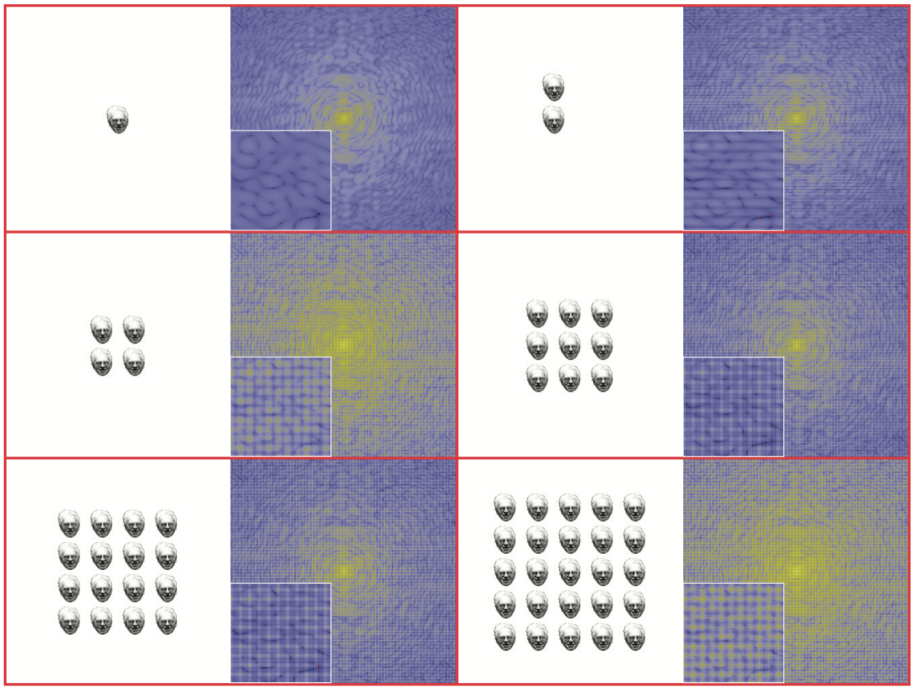


Figure 24 – A picture of David Sayre is repeated to form a periodic structure, with the corresponding diffraction patterns shown on the right of each sample. Figure from [1].

is natural to ponder: why the need for coherence? In fact, it has been pointed out that optical imaging under fully coherent illumination results in a lower spatial resolution than with incoherent illumination [25, Chapter 8]. The need for coherence arises because in such situation the diffraction pattern encodes the spatial structure of the object in a unique way [26] and, therefore, this structure can be uniquely recovered. That becomes the scope of Coherent Diffractive Imaging techniques, such as plane-wave CDI and Ptychography. In the next sections we define coherence and then explain how the wanted information is encoded in the diffraction pattern through the so-called *speckles*.

3.3.3 Ewald Sphere

We return now to the First Born approximation introduced in section 3.2.1. We saw that this approximation assumes that scattering is weak (no multiple scattering events for a single photon) and that the scattered wave is a spherical wavelet. In these conditions, we assume that the scattered amplitude is composed of the amplitude of the incident wave together with the amplitude of the scattered wavelets. Suppose that the incident plane wave wavevector $\mathbf{k_0}$ is aligned with the z-axis. Therefore, the resulting scattering amplitude ψ will be

$$\psi(\mathbf{r}) = \psi_{\text{inc}} + \psi_{\text{scattered}} = e^{i\mathbf{k_0 r}} + f(\mathbf{\Delta k})\frac{e^{ikr}}{r} \tag{3.93}$$

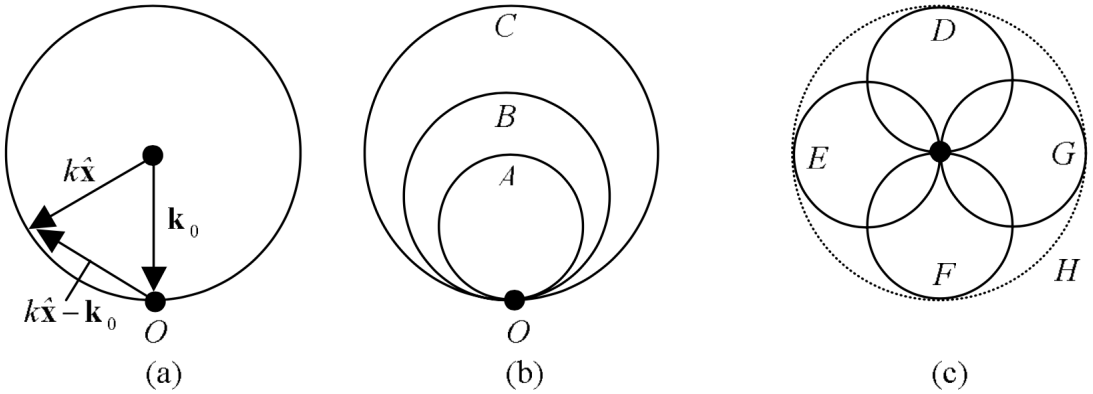


Figure 25 – (a) Scheme for visualizing the construction of the Ewald sphere in the simplified case of a plane. In this case, the direction of the scattered wavevector $\hat{\mathbf{r}}$ is shown as $\hat{\mathbf{x}}$. (b) Changing the energy means one changes the radius of the Ewald sphere, and one can access more points than a single surface in reciprocal space. (c) The same can be accomplished by changing the direction of the incident radiation with respect to the sample, since then the Ewald sphere will be rotated around the origin. Figure from [10].

where the scattering factor $f(\mathbf{\Delta k})$ can also be calculated as [10]

$$f(\mathbf{\Delta k}) = \frac{k^2}{4\pi} \int\int\int (n^2(r') - 1) e^{-i\mathbf{\Delta k} \cdot r'} d\mathbf{r} \tag{3.94}$$

with $r' = (x', y', z')$. The scattered wavevector is $\mathbf{k} = k\hat{\mathbf{r}}$ and

$$\mathbf{\Delta k} = \mathbf{k} - \mathbf{k_0} = (\Delta k_x, \Delta k_y, \Delta k_z) = (k_x, k_y, k_z - k_0) \tag{3.95}$$

where $|\mathbf{k_0}| = k_0$.

Equation 3.94 is a three-dimensional Fourier Transform evaluated at points $\mathbf{\Delta k}$. In principle, if one had access to all points in reciprocal space, a simple inverse Fourier Transform would allow us to access the values of $n(r')$. Nonetheless, for a particular experimental condition, that is not the case. Since the incident radiation is scattered to all directions, the scattered wavevector direction $\hat{\mathbf{r}}$ may assume any value. If we consider the incident wavevector $\mathbf{k_0}$ to be constant, we have that the set of values assumed by $\mathbf{\Delta k}$ are spanned by a spherical surface, called the *Ewald sphere*. Figure 25(a) shows a simplified two-dimensional scheme of Fourier space to help us understand how this sphere is formed. The incident wavevector $\mathbf{k_0}$ is fixed and points towards the origin O of Fourier space. Assuming elastic scattering, we have $k = k_0$, but the direction of the scattered wavevector $\hat{\mathbf{r}}$ may vary freely. Therefore, for all possible directions, the points $\mathbf{\Delta k}$ over which the Fourier Transform is calculated will span a spherical surface. Since $k_0 = 2\pi/\lambda$, the radius of the sphere will depend on the wavelength of the incident radiation.

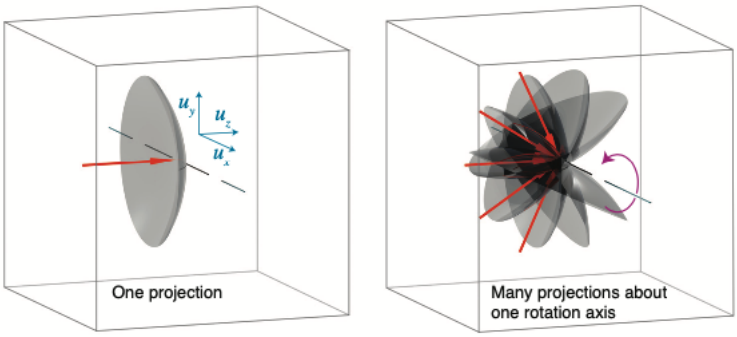


Figure 26 – In practice, one only measures part of the Ewald sphere due to the limited portion of the wavefield captured by the area detector. To gather three-dimensional information about the sample, one needs to perform the experiment rotating the sample to sweep out a volumetric region of F-space. Figure from [1].

Consider now that this Ewald sphere is located in Fourier space (F-space). Consider also that the 3D Fourier Transform $f(\mathbf{\Delta k})$ of $(n^2(r') - 1)$ (equation 3.94) is located in that same space. For a particular experimental condition (meaning a given λ and direction of the incident radiation with respect to the sample), one only has "access" to the points in reciprocal space spanned by the Ewald sphere. In other words, the wavefield for a particular experimental condition will contain information only of a two-dimensional region of reciprocal space. There are two approaches one can use to obtain more information of the reciprocal space. The first is to perform multiple measurements varying the wavelength of the incident radiation. In this way, one varies the radius of the Ewald sphere and access more points of F-space, as illustrated in Figure 25(b). Another option is to keep the energy constant, but to change the direction of the incident radiation with respect to the sample (for instance, rotating the sample while keeping the incident radiation constant). In this manner, one rotates the Ewald sphere around the origin of F-space (Figure 25(c)). With these strategies, one is able to "scan" a volume in reciprocal space, gathering more information.

In synchrotron imaging experiments, the most feasible and therefore common approach is to rotate the sample instead of varying the energy of the incident radiation. Nonetheless, it is important to emphasize that, although scattering happens in all directions, one does not measure the wavefield over the whole sphere around the sample. Instead, one places an area detector at a certain distance from the sample. This means that for each experiment one does not gather information over an entire spherical surface, but rather over a portion of it, i.e. a spherical cap, as shown in Figure 26 (left). If one rotates the sample, the spherical cap will rotate around the origin of reciprocal space, and one will then gather information over the volume spanned by the multiple spherical caps.

With the picture of the Ewald sphere and the different approaches one can

use to measure a volume over reciprocal space, we can summarize the different imaging modalities available, as illustrated in Figure 27. For an experiment using an incident plane wave (i.e. collimated illumination) of specific energy (monochromatic), the scattered wavefield will contain information over the Ewald sphere, i.e. a spherical region in reciprocal space (Figure 27(a)). Nonetheless, one measures in practice only part of this spherical surface because just a portion of the scattered wavefield is captured by the detector (Figure 27(b)). If instead one varies the energy of the incident radiation, the radius of the Ewald will change and a volume of reciprocal space will be scanned (gray region, Figure 27(c)). On the other hand, if the incident radiation is not a perfect plane wave (non-collimated), the different travelling directions of the plane wave components will cause the Ewald sphere to rotate around the origin, so again a volume of reciprocal space will be scanned (Figure 27(d)). The most common scenario is to use monochromatic radiation and rotate the sample, illuminating the sample through different directions. In that case, multiple spherical caps are collected throughout reciprocal space (Figure 27(e)).

The cases illustrated in Figures 27(f) and (g) bring us back to the Projection Approximation. As will be seen in section 3.5.2, when we assume the PA we are actually assuming that we collect not a spherical cap over Fourier Space, but rather a plane passing through the origin (Figure 27(f)). In that case, each plane in F-space is associated to a projection of the sample in real-space. Therefore, during an imaging experiment, the sample is rotated with respect to the beam and multiple planes are collected in F-space (Figure 27 (g)).

3.3.4 The First Born Approximation and the Far-field regime

We briefly return to equations 3.38

$$\psi(x,y,z=z_0) \approx \exp\left(-ik\int_{z=0}^{z=z_0}(\delta(x,y,z)-i\beta(x,y,z))dz\right)\psi(x,y,z=0)$$

and 3.94

$$f(\mathbf{\Delta k}) = \frac{k^2}{4\pi}\int\int\int(n^2(r')-1)e^{-i\mathbf{\Delta k}\cdot r'}d\mathbf{r}$$

to provide the link between the Born Approximation and Fraunhoffer diffraction measurement. Making use of equation 3.83, we can rewrite 3.94 as

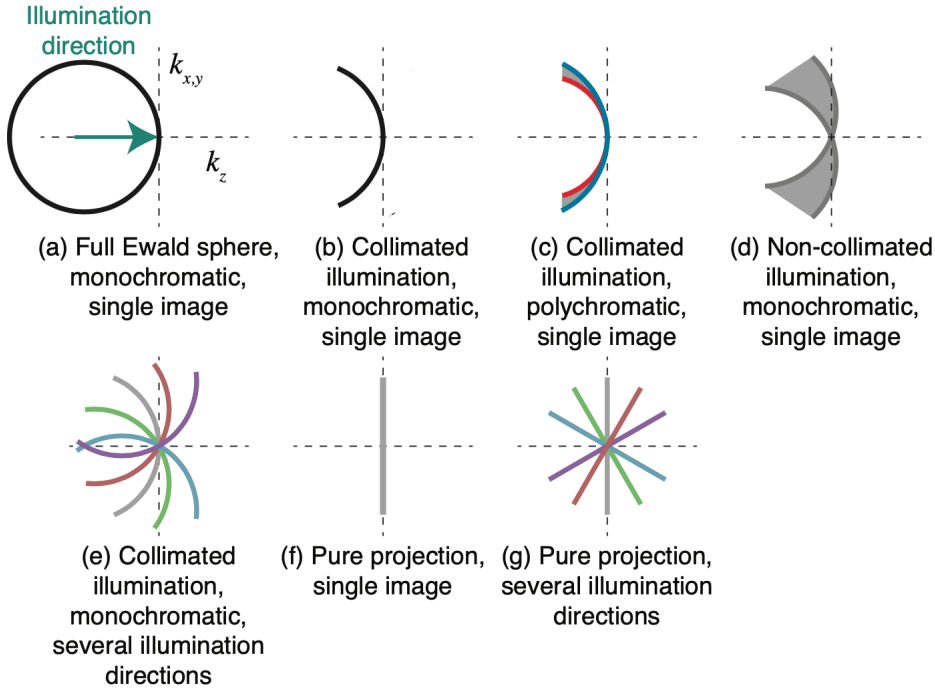


Figure 27 – Different possibilities of gathering information in reciprocal space. Figure from [1].

$$
\begin{aligned}
f(\Delta k_x, \Delta k_y, \Delta k_z) &= -\frac{k^2}{2\pi} \int_{x'=-\infty}^{+\infty} \int_{y'=-\infty}^{+\infty} e^{-i(\Delta k_x x' + \Delta k_y y')} \\
&\times \left\{ \int_{z'=0}^{z_0} (\delta(x', y', z') - i\beta(x', y', z')) e^{-i(\Delta k_z z')} dz' \right\} dx' dy' \\
&\approx -\frac{k^2}{2\pi} \int_{x'=-\infty}^{+\infty} \int_{y'=-\infty}^{+\infty} e^{-i(\Delta k_x x' + \Delta k_y y')} \\
&\times \left\{ \int_{z'=0}^{z_0} (\delta(x', y', z') - i\beta(x', y', z')) dz' \right\} dx' dy'
\end{aligned}
\tag{3.96}
$$

For the above approximation to hold, we assumed that the incident wave $\exp(ikz)$ travels in the z direction and that we are within the paraxial approximation, which means we have small scattering angles. Since scattering is elastic ($k = k_0$), we have small k_x and k_y, meaning $k_z \approx k_0$. Therefore, $\Delta k_z \approx 0$ and the exponential within the braces in equation 3.96 is replaced by 1. Furthermore, if we assume Born Approximation (weak scattering), we may linearize the exponent of the transmission function in 3.83:

$$
\exp\left(-ik \int_{z=0}^{z=z_0} (\delta(x, y, z) - i\beta(x, y, z)) dz\right) \approx 1 + \left(-ik \int_{z=0}^{z=z_0} (\delta(x, y, z) - i\beta(x, y, z)) dz\right)
$$

$$
\psi(x, y, z = z_0) \approx \psi(x, y, z = 0) + \left(ik \int_{z=0}^{z=z_0} (\delta(x, y, z) - i\beta(x, y, z)) dz\right) \psi(x, y, z = 0) \tag{3.97}
$$

Since the incident plane wave $\psi(x, y, z = 0) = \exp(ik0) = 1$ at the incidence plane, if we further discard the remaining unscattered component $\psi(x, y, z = 0)$ in the above equation, we get

$$\psi(x, y, z = z_0) \approx ik \int_{z=0}^{z=z_0} (\delta(x, y, z) - i\beta(x, y, z))dz \tag{3.98}$$

Now, the above integral can be substituted back in equation 3.96

$$\begin{aligned} f(\Delta k_x, \Delta k_y, \Delta k_z) &\approx -\frac{k^2}{2\pi} \int_{x'=-\infty}^{+\infty} \int_{y'=-\infty}^{+\infty} e^{-i(\Delta k_x x' + \Delta k_y y')} \left\{ \frac{\psi(x, y, z = z_0)}{ik} \right\} dx'dy' \\ &= \frac{ik}{2\pi} \int_{x'=-\infty}^{+\infty} \int_{y'=-\infty}^{+\infty} \psi(x, y, z = z_0) e^{-i(\Delta k_x x' + \Delta k_y y')} dx'dy' \end{aligned} \tag{3.99}$$

where the integral can be recognized as the Fourier Transform of $\psi(x, y, z = z_0)$:

$$f(\Delta k_x, \Delta k_y, \Delta k_z) \approx \frac{ik}{2\pi} \mathcal{F}\{\psi(x, y, z = z_0)\} \tag{3.100}$$

From equation 3.7, we saw that the intensity is calculated as the absolute squared of the complex wavefunction. If we now return to equation 3.93 and discard the unscattered component, we get:[15]

$$I = |\psi(\mathbf{r})|^2 \approx \left| \frac{ik}{2\pi} \frac{e^{ikr}}{r} \mathcal{F}\{\psi(x, y, z = z_0)\} \right|^2 = \frac{1}{r^2\lambda^2} |\mathcal{F}\{\psi(x, y, z = z_0)\}|^2 \tag{3.101}$$

Therefore, we once again see that, within the Born and Projection approximations, the measured intensity is proportional to the absolute squared Fourier Transform of exit wavefield $\psi(x, y, z = z_0)$. Note yet the inverse-squared dependence with the distance r, as expected for the decaying intensity due to isotropic propagation.

3.3.5 Resolution and the Validity of the Projection Approximation

Traditionally, imaging systems make use of lenses. As we have seen in section 3.1.2, free space propagation acts as a Fourier Transform to the wavefield. Lenses are nothing more than devices that perform an inverse Fourier-transform to the same wavefield, recovering the image (original wavefield) at the focal plane. If we think of the lens as a Fourier Transform device, we realize that a point-source will become a plane-wave after the lens. Inversely, an incident plane wave is focused to a point. Nonetheless, this assumes

[15] The unscattered beam is usually ignored experimentally by placing a beam stopper before the detector that block the direct beam. This is done because the high intensity of the direct beam may easily saturate the detector pixels and even damaged them, otherwise.

an infinite-extent lens, capable of gathering all the light from the wave. In reality, lenses have a finite aperture function P.

In the case of an infinitely small point source, the effect of this aperture is to "spread" the ideal point focus, also known as a *point-spread function*. One can use the particular case of a circular aperture to obtain an estimate to the theoretical resolution limit of an imaging system, the *diffraction limit*. The image of such infinitely small point-source at the focal plane becomes the Airy disk

$$A = \frac{2J_1(r)}{r} \tag{3.102}$$

where J_1 is the Bessel function of the first kind of order one. If one has two point sources being imaged, one has the intensity of two Airy disks imaged in the focal plane, as shown in Figure 28. The distance δ_r between the point sources allows us to quantify the resolution limit of the system. As this distance decreases, the Airy functions start to overlap, up to the point that the maximum of the first Airy intensity pattern reaches the first minimum of the other Airy pattern. This characterizes the *Rayleigh resolution* criterion given by

$$\delta_r = 0.61\frac{\lambda}{\mathrm{NA}} \tag{3.103}$$

where NA is the numerical aperture of the lens defined as

$$\mathrm{NA} = n\sin\theta \tag{3.104}$$

with θ being the maximum angle sub-intended by the aperture from the optical axis and n the refractive index of the medium. The overlap between the Airy functions is further illustrated in Figure 29. This configures a criteria for the limiting distance two objects can be apart from each other and still be distinguished by an imaging system.

Nonetheless, the criterion above makes use of an imaging system containing a focusing optics. How does one define such resolution limits when we are dealing with a *lensless* imaging system, such as the case of Coherent Diffractive Imaging? Typically this is done by associating the maximum detected scattering angle θ with the Numerical Aperture NA, and the thickness t of the object being image with the Depth of Field (DOF) of the imaging system[16]. If one considers an object of thickness t, one can show that [1, Chapter 4]

$$t = \frac{\lambda}{\theta^2} \Longleftrightarrow \mathrm{DOF} = \frac{\lambda}{\mathrm{NA}^2} \tag{3.105}$$

[16] Intuitively, the depth of field can be thought of as the depth in which objects can still be recognized as "in focus" by the imaging system. Therefore, if the whole object lies within the DOF, all the imaging planes through the object will in principle be in focus.

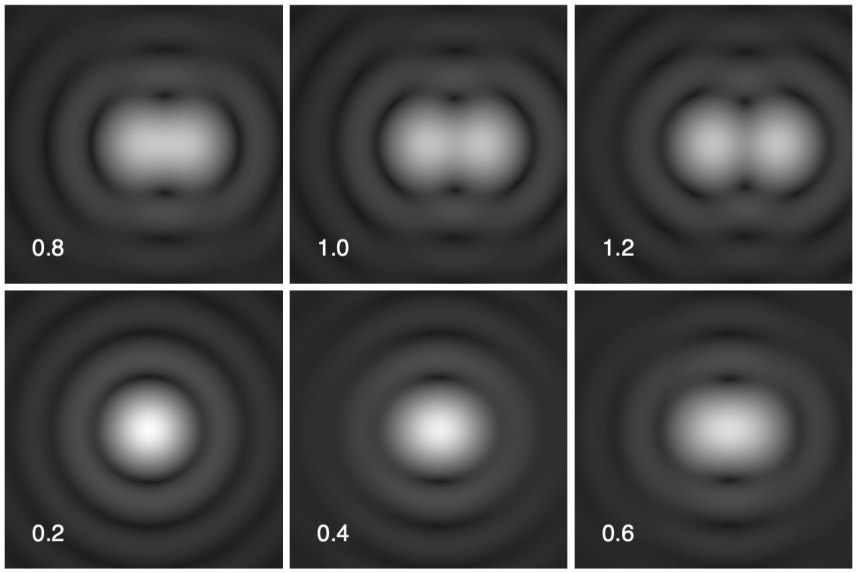


Figure 28 – Two Airy patterns plotted for different separation distances. For a certain distance, one can no longer distinguish the two patterns, which gives us a way of defining a resolution criterion for a certain wavelength. Figure from [1].

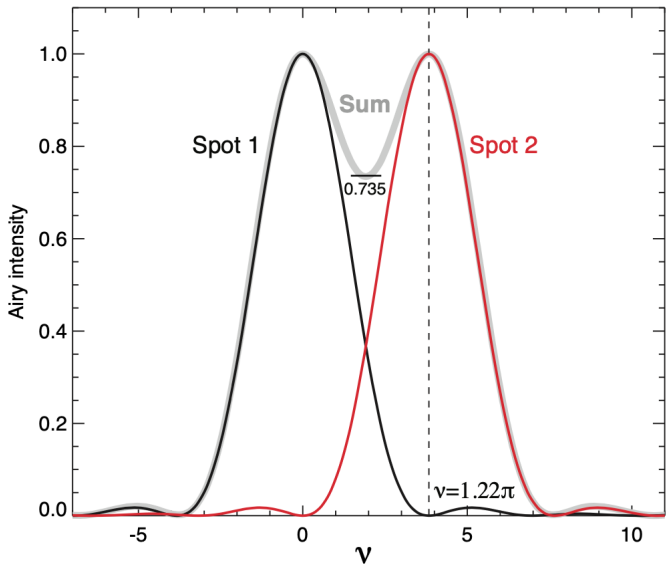


Figure 29 – The Rayleigh resolution is defined by considering the distance over which the maximum of one Airy pattern reaching the first minimum of the other Airy pattern. Figure from [1].

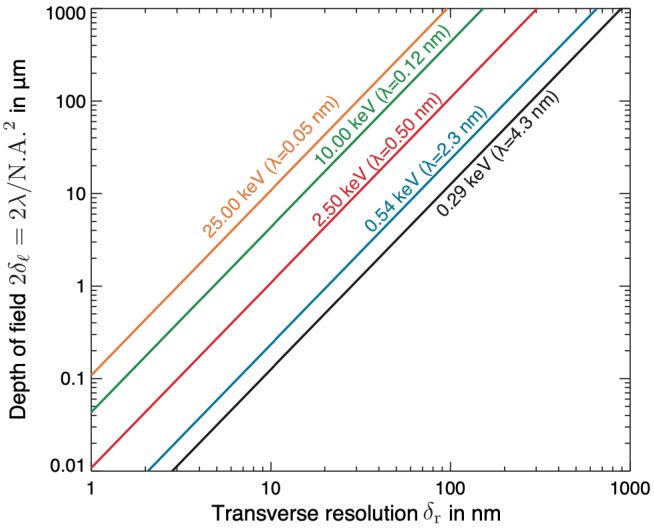


Figure 30 – Depth of Field (DOF) versus transverse resolution for multiple X-ray energies. Figure from [1].

$$\mathrm{DOF} = 4\frac{\Delta r^2}{\lambda} \tag{3.106}$$

Using the Rayleigh resolution criterion, one can also calculate the DOF for a lens imaging system [1], which turns out to be

$$\mathrm{DOF} = 5.4\frac{\Delta r^2}{\lambda} \tag{3.107}$$

indeed a close value to the lensless case. Of course, one must keep in mind that such criteria are somewhat arbitrary. Indeed, different expressions for the DOF of lensless imaging systems can be found in the literature (see [1], section 4.2.5 and references therein), but all with reasonable agreement with each other.

The DOF expression 3.106 is a core equation for this work. If we think that the DOF sets a limit to the thickness of the object being imaged, this equations shows us that, for a given energy, the thicker the object, the worst is the expected transverse resolution achieved. This is illustrated in the plot of Figure 30 for multiple X-ray energies. It is worth noting that the above equations are deduced within assumptions of the Projection Approximation. A core objective of this work was to evaluate if, by circumventing the Projection Approximation, the resolution limit given by equation 3.106 could also be surpassed.

At last, we summarize below the three important points that must be taken into consideration when we consider the limits of the Projection Approximation:

- **Ewald Sphere Curvature**: as extensively discussed in section 3.3.3, a particular experimental configuration maps a spherical cap in Fourier space instead of a plane,

as assumed by the PA.

- **Fresnel diffraction**: a propagating wave will naturally continue to evolve even within the sample. Therefore, for thicker samples one expects that Fresnel diffraction will eventually become relevant. This is problematic, since it means that intermediary planes within the sample will not see the same plane-wave incident field as the input plane.
- **Violation of the Born Approximation**: the weak-scattering assumption may also break down for thicker samples. As the wave needs to propagate longer distances within the sample, it becomes more likely that it will undergo a second scattering event.

3.3.6 Coherence

In many of the derivations previously made, such as in the section about Scattering, we have assumed that the incident radiation was a perfect plane-wave: that is, it consisted of a wave of unique temporal and spatial frequency, meaning single wavelength and propagation direction. Nonetheless, in practice a wave will not be perfectly plane. The two ways in which it can deviate from this perfect state are quantified by the coherence lengths. The first is the longitudinal (or temporal) coherence length and the second is the transversal (or spatial) coherence length, both described ahead.

Longitudinal Coherence

Consider two plane waves A and B travelling in the same direction but with slightly different wavelengths, λ and $(\lambda - \Delta\lambda)$, as depicted in Figure 31. At the beginning, the waves are in phase. The longitudinal coherence length L_l is the distance the waves travel until they are out of phase. Hence, the waves will be in phase again after travelling a distance $2L_l$. This will happen precisely when the wave of shorter wavelengths has had one full oscillation more than the other, meaning

$$2L_l = N\lambda = (N+1)(\lambda - \Delta\lambda) \tag{3.108}$$

The second equality implies $N \approx \lambda/\Delta\lambda$ for big N, which then gives through the first equality

$$L_l = \frac{1}{2}\frac{\lambda^2}{\Delta\lambda} \tag{3.109}$$

That is, the longitudinal coherence length will be inversely proportional to the bandwidth $\Delta\lambda$ of the radiation, as expected.

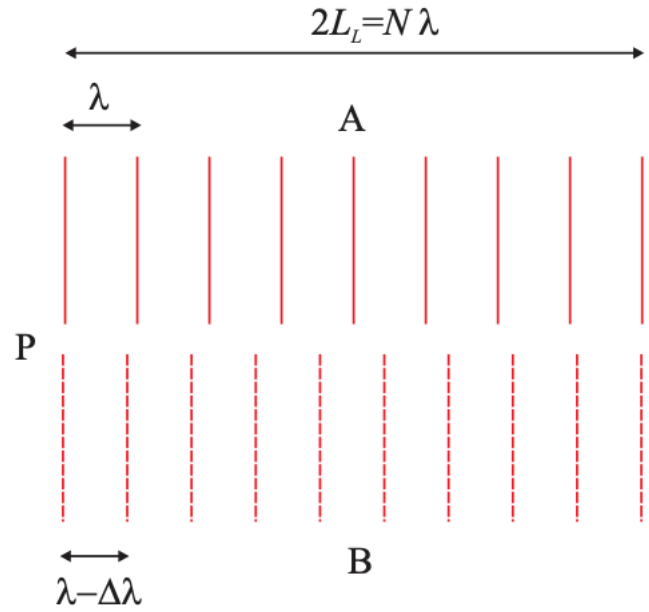


Figure 31 – Illustration of two plane-waves of different wavelengths traveling at the same direction. Figure from [17].

A condition for full longitudinal coherence is that the path length difference travelled by different waves due to the width w and thickness l of the sample is smaller than L_l. One can show that [26]

$$\frac{\lambda^2}{\Delta\lambda} = 4l\sin^2(\theta) \tag{3.110}$$

where 2θ is the scattering angle of the waves. Therefore, the longitudinal coherence sets a limit to the bandwidth of the radiation given a certain sample of interest.

Transversal Coherence

Consider now that waves A and B have the same wavelength, but slightly different propagation directions, as in Figure 32. After travelling a distance R from the source, their wavefronts meet at point P. The distance travelled along the wavefront A until it is out-of-phase with B defines the transversal coherence length. Considering that the difference in propagation direction arises due to a distance D between each source, simple trigonometry shows us that

$$L_t = \frac{\lambda}{2}\frac{R}{D} \tag{3.111}$$

Note that $\Delta\theta = D/R$ is the divergence of the beam. For scattering to be fully-coherent in the transverse direction, one needs the dimension a of the object to be smaller than L_t, meaning

$$\Delta\theta = \frac{D}{R} < \frac{\lambda}{a} \tag{3.112}$$

That is, the transversal coherence length sets a maximum desired divergence of the beam for a certain object size.

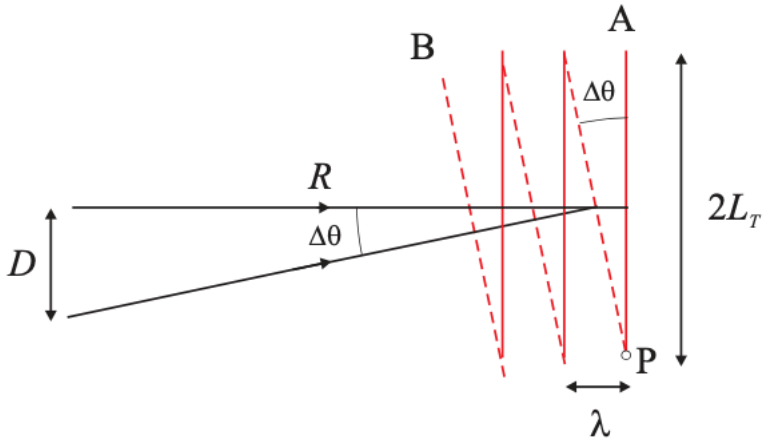


Figure 32 – Illustration of two plane waves with the same wavelength travelling in different directions. Figure from [17].

Speckles

For many decades X-ray imaging has been performed without the need for coherent radiation. Why then the need for coherence? The answer lies on the fact that, by using coherent radiation, the measurement will uniquely encode the position of all scattering sources of the sample. In other words, if one uses incoherent radiation, two different electronic configurations might yield the same intensity pattern. That does not happen when we deal with perfect monochromatic plane waves. As illustrated in Figure 33, the measured diffraction pattern will present small features named *speckles* when we use coherent radiation. In contrast, the use of incoherent radiation results in a more "blurred" diffraction pattern.

It is this unique speckle configuration that encodes the information needed to recover the phase of the wavefield. In order to retrieve the phase, one needs to properly sample the diffraction (or speckle) pattern. We shall further explain oversampling in detail in section 3.4. Figure 34 further illustrates how the diffraction pattern changes as one degrades the coherence of the incident wave, where a blurring of the diffraction pattern intensifies as coherence is lost.

3.4 Phase Retrieval

Throughout the text, we have mentioned the fact that CDI relies on an intensity measurement, such as equation 1.10:

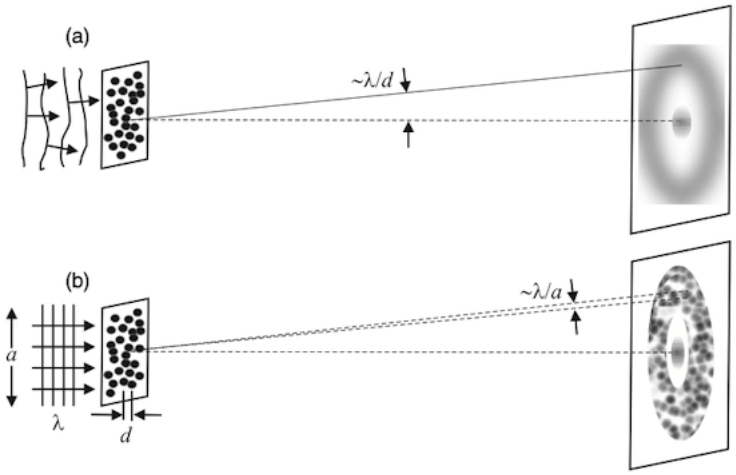


Figure 33 – The diffraction pattern when using incoherent radiation differs from that obtained in the presence of coherent radiation by the emergence of the so-called speckles. These small features uniquely encode the positions of the scattering centers in the sample. Figure from [26].

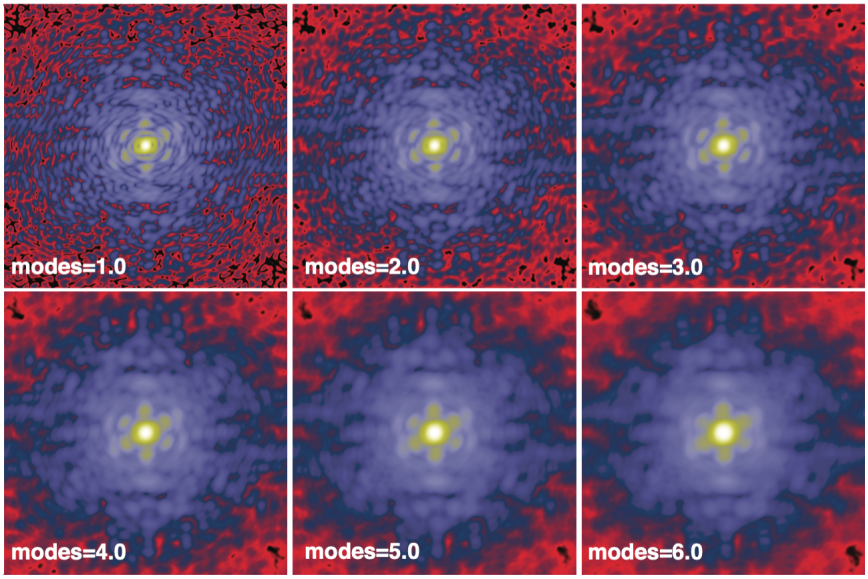


Figure 34 – Speckle pattern degrades as the incident wave looses its coherence, here quantified by the number of coherent-modes present in the wave. Figure from [1].

$$I = |\mathcal{F}(\rho(x))|^2$$

When we perform an intensity measurement, all phase information is lost. Therefore, one cannot recover an image from the intensity pattern via a simple inverse Fourier Transform.

It was suggested long ago [27] that, if one is able to sample the diffraction pattern finely enough (oversampling), the speckle pattern encodes the phase information and it can then be retrieved numerically using computational methods. The classic phase retrieval algorithm has its roots in electron microscopy [28], where one uses the magnitude information previously know for both real and reciprocal spaces. In the X-ray diffraction problem, one has only information about the latter, which requires the introduction of further *a priori* knowledge to successfully retrieve the phase. This was first proposed by [29], using *support* and *positivity* constraints, as we shall detail in sections 3.4.2 and 3.4.3. We detail ahead how the oversampling of the speckles allow us to retrieve the phase information.

3.4.1 Oversampling

Before moving forward to the algorithms, we shall further discuss the oversampling condition. It is important to emphasize that such condition comes directly from the mathematical relationship between real and reciprocal space. We start by defining the equations to reach such conclusion and later on give another physical interpretation to the oversampling condition.

We start by defining the Discrete Fourier Transform (DFT) of a signal $g(x_n)$, given by

$$G(f_k) = \sum_{0}^{N-1} g(x_n) \exp\left(-i2\pi f_k x_n\right) \tag{3.113}$$

where the real and reciprocal space variables x_n and f_k assume discrete values according to

$$\begin{aligned} x_n &= n\Delta x \quad , \quad n = 0, 1, ..., N-1 \\ f_k &= k\Delta f \quad , \quad k = 0, 1, ...N-1 \end{aligned}$$

Δx and Δf correspond to the sampling interval in real and reciprocal space, and N is the number of sampling points. The duality between both spaces is shown by the following equations

$$L_x = N\Delta x \quad , \quad L_f = N\Delta f \tag{3.114}$$

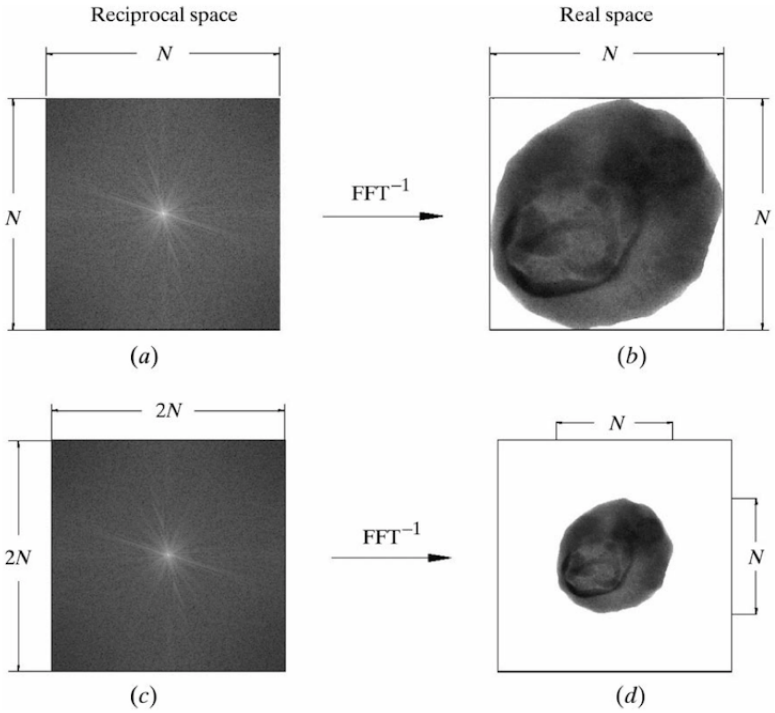


Figure 35 – Scheme illustrating the effect of increasing the number of points to better sample the reciprocal space. In real space, this action is equivalent to increasing the size of the signal. Figure from [30].

with

$$\Delta x = \frac{1}{L_f} \quad \text{and} \quad \Delta f = \frac{1}{L_x} \tag{3.115}$$

That is, there exists an inverse relationship between the dimension of real space with respect to the sampling of reciprocal space, and vice-versa.

The oversampling condition states that the diffraction pattern (i.e. a reciprocal space measurement) should be sampled finely enough to guarantee that one has enough information to recover the phase. From the above equations, we can see that sampling the diffraction pattern with a smaller Δf is equivalent to having a larger size of real space L_x. This idea is illustrated in Figure 35. If one doubles the number of points in real space (by zero padding the data matrix, for instance), the reciprocal space will have twice as many $(2N)$ points spanning the same frequency range.

That being said, we can now show why we need the oversampling condition to properly reconstruct the phases of our signal, which directly relates to the Nyquist sampling theorem. This theorem states that a signal that is non-null over a range a can be perfectly reconstructed as long as it is sampled with a frequency of at least $1/a$. Equivalently, one can say that if the highest frequency component in the signal is f, one needs to sample it at a frequency $f' \geqslant 2f$, known as the *Nyquist frequency*.

Now, we refer back to equation 1.10. If we inverse Fourier Transform our signal, we obtain

$$\begin{aligned}\mathcal{F}^{-1}|\mathcal{F}(\rho(x))|^2 &= \mathcal{F}^{-1}[\mathcal{F}(\rho(x))\mathcal{F}^*(\rho(x))] \\ &= \mathcal{F}^{-1}[\mathcal{F}(\rho(x))] \circledast \mathcal{F}^{-1}[\mathcal{F}^*(\rho(x))] \\ &= \rho(x) \circledast \rho(-x)\end{aligned} \tag{3.116}$$

where $\circledast$ is the convolution symbol. In the above equation, we have made use of the Fourier convolution theorem. The obtained result $\rho(x)\circledast\rho(-x)$ is the density *autocorrelation function*. If $\rho(x)$ is non-zero over a length a, this function will be non-zero over a length $2a$. Consequently, in order to reconstruct it,it has to be sampled at a frequency $1/2a$ according to Nyquist. This amounts to saying that we will sample reciprocal space with half the desired sampling frequency of $1/a$, i.e. we shall oversample the diffraction pattern. As previously shown in equation 3.4.1, this means having a real-space twice as big. This gives rise to the so called *support condition*, which is essential for the convergence of the phase-retrieval algorithms. By support we mean the region where the sample is located in space. In other words, zero-padding the sample to have twice the number of points is equivalent to oversampling the object's reciprocal space. Furthermore, by zero padding the sample, we are equivalently stating we know where the region of non-null electron-density is located.

Another approach to understanding the need for oversampling relies on the size of the speckles in the diffraction pattern. These consist in the small features in the diffraction pattern that arise due to the scattering of the coherent radiation. The speckles are uniquely determined by a certain electron-density configuration. They present an angular size of λ/a, where λ is the wavelength and a the size of the illuminated sample area [26]. If one considers a detector with pixel size p placed at a distance L from the sample, the angular size of the pixel will be p/L. The oversampling ratio σ can be analyzed in terms of the ratio between these dimensions [31]:

$$\sigma = \frac{\lambda/a}{p/L} \tag{3.117}$$

If we consider a hypothetical one-dimensional diffraction pattern, one must have $\sigma > 2$ for the oversampling condition to be satisfied. One can then in principle recover the phase information. This is equivalent to saying that the angular size of the pixel must be at least 2 times smaller than the size of the speckle, so that the speckle information is properly sampled according to the Nyquist sampling theorem. In fact, for higher dimensional spaces this condition is relaxed [31] and one can recover the phase information for

$$\sigma \geqslant \sqrt[N]{2} \tag{3.118}$$

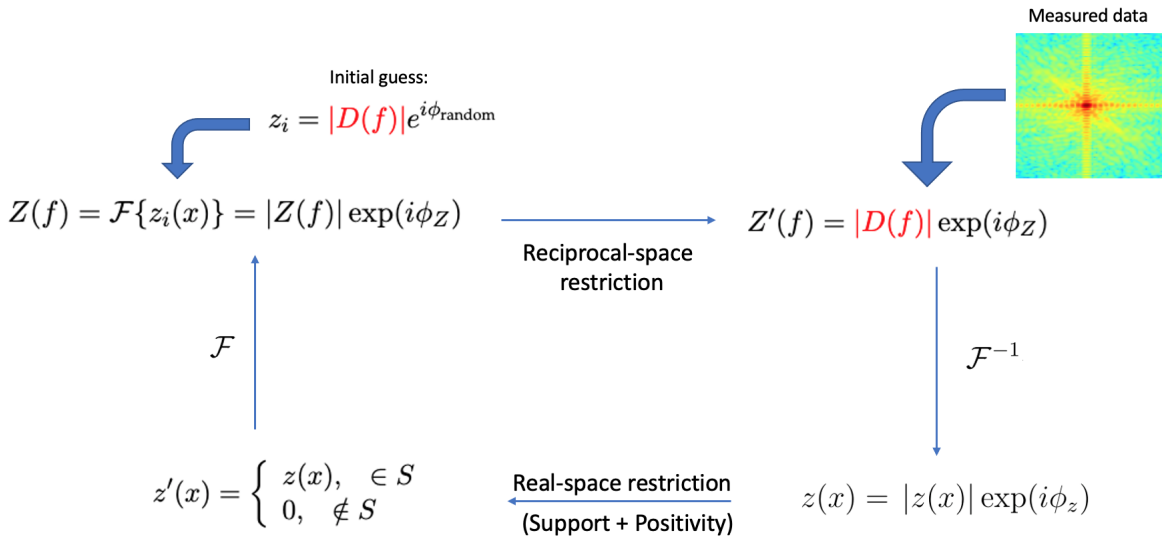


Figure 36 – Fluxogram of the Error Reduction algorithm.

where N is the number of dimensions of the sample. In order to have a well-posed problem and ease the convergence of the phase retrieval algorithms, one should have a value of σ as high as possible. Experimentally, this means having a detector with small pixel size p placed as far as possible from the sample. Notice also that, having a bigger coherently illuminated sample of size a will decrease the oversampling ratio. Therefore, it is vital to have flexibility to place the detector far away from the sample, to adjust σ depending on the sample to be imaged. This is one of the strengths of Cateretê beamline at Sirius, which allows the detector to be place up to 30 meters away under vacuum.

We detail ahead the Phase-Retrieval algorithms used in this work. These were originally proposed by [29], namely the *Error Reduction* and, its most used variant, *Hybrid Input-Output* algorithms.

3.4.2 Error Reduction (ER)

Assuming we have measured a diffraction pattern in oversampling condition, the Error Reduction algorithm iteratively alternates the image between real and reciprocal spaces, imposing constraints on the signal at each iteration. Figure 36 shows a fluxogram of the algorithm.

We start with an initial guess z_i for the complex image. One can compose this image with an amplitude and an initial random phase ϕ_{random}. The amplitude can be, for instance, a rough guess such as the amplitude of the reciprocal space $|D(f)|$, that is, the diffraction pattern intensity measurement itself.[17] We then apply the Fourier Transform to

[17] Although clearly a wrong guess, the idea is to at least use the same consistent initial guess for the amplitude an let only the phase be random when we start the algorithm. As an alternative, we may use a constant amplitude of ones, for instance.

this initial guess, obtaining a complex signal $Z(f)$ in reciprocal space that can be divided into an amplitude $|Z(f)|$ and phase components $\exp(i\phi_Z)$ as

$$Z(f) = \mathcal{F}\{z_i(x)\} = |Z(f)| \exp(i\phi_Z) \tag{3.119}$$

Afterwards, we impose the reciprocal-space restriction, which is to substitute the amplitude of the signal by its the correct value, which is our measurement $|D(f)|$ itself, obtaining an updated version of $Z'(f)$:

$$Z'(f) = |D(f)| \exp(i\phi_Z) \tag{3.120}$$

We then inverse-fourier transform $Z'(f)$ signal to obtain an updated version $z(x)$ of the real space signal, which can again be decomposed into amplitude and phase components:

$$z(x) = \mathcal{F}^{-1}\{Z'(k)\} = |z(x)| \exp(i\phi_z) \tag{3.121}$$

Finally, the last step of the iteration consists in applying *a priori* known real space information, the so-called *support* constraint. By support we mean the region of non-zero density, that is, the points of the image where the sample is located. At this step, we impose all the signal outside the support region to be null, as in equation 3.122:

$$z'(x) = \begin{cases} z(x), & \in S \\ 0, & \notin S \end{cases} \tag{3.122}$$

The updated signal $z'(x)$ is then Fourier Transformed again, and the algorithm repeats for a chosen number of iterations or until an error metric is met. The convergence error is tracked separately for both real and reciprocal spaces. For reciprocal space, we use the normalized absolute difference as

$$E_f = \frac{\sum_k ||Z(f)| - |D(f)||}{\sum_k ||D(f)||} \tag{3.123}$$

where the sum spans across all k pixels of the image. On the other hand, since we do not have a real-space measurement to compare with, the following error metric is typically used:

$$E_r = \frac{\sum\limits_{x \notin S} |z(x)|^2}{\sum\limits_{x \in S} |z(x)|^2} \tag{3.124}$$

We expect this error to go to zero, since the numerator – corresponding to the sum of values outside the support – will approach zero, while the denominator approaches a constant value.

3.4.3 Hybrid Input-Output (HIO)

The *Hybrid Input-Output* algorithm is a slight variation of the ER algorithm. The only difference occurs in the last step of the iteration, where it also applies the *positivity* constraint in real space. This constraint uses the *a priori* knowledge that the electron density is related to the complex atomic scattering factor $f_1 + if_2$ and that f_1 is usually positive, whereas f_2 is always positive [23, 24].

This constraint corresponds to keeping unaltered every positive pixel within the support. Nevertheless, if the pixel is located outside the support or if it is negative inside the support, it shall be corrected according to the following

$$z'(x) = \begin{cases} z(x), & \in S \text{ and } z(x) > 0 \\ z_i(x) - \beta z(x), & \notin S \text{ or } \in S \text{ with } z(x) < 0 \end{cases} \tag{3.125}$$

In this manner, the lower equation will gradually shifts negative values towards positive ones. This corresponds to a more subtle change of values compared to the ER equation 3.122 and helps the algorithm avoid being stuck at local minima at the first few iterations.

There are further variations of phase-retrieval iterative projection algorithms that have been compared in the literature [32]. These are usually more complex than HIO and ER, and tend to present better results. Nonetheless, we chose to focus on the more intuitive and widely-used ER and HIO algorithms since the focus of the work lies on the second step of the CDI reconstruction technique. We shall present the simulation results used the validate both ER and HIO in section 4.3.

3.5 Tomography

In this final theory section, we introduce tomography algorithms that can be used to perform the second step in CDI reconstruction. These are namely Filtered Back Projection (FBP) and the Algebraic Reconstruction Technique (ART). The former may be used within the Projection Approximation and we introduce the Radon Transform and the Central Slice Theorem in order to better understand the reciprocal-space picture already shown in Figure 27. To understand FBP, we anticipate some plots obtained from simulation results of this algorithm, which we implemented following the nearest-neighbor interpolation strategy for rotations. The equations herein are mainly due to [11].

3.5.1 Radon Transform

The Radon Transform consists in a linear operator $\mathbf{R} : \mathbf{U} \mapsto \mathbf{V}$ that maps the Cartesian into polar coordinates $(x, y) \to (t, \theta)$ through an integral operator between Hilbert spaces: the feature space $\mathbf{U}$ and the so-called sinogram space $\mathbf{V}$. We show the

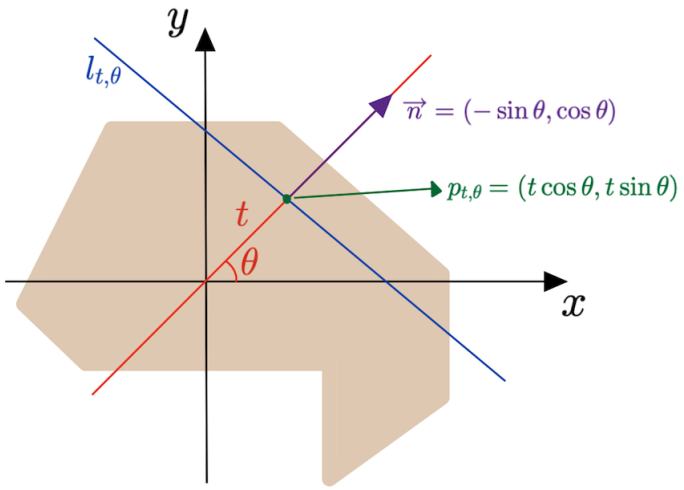


Figure 37 – Illustration of the xy-Cartesian plane with an object (brown). The line $l_{t,\theta}$ (blue) passes through point $p_{t,\theta}$ (green) and is normal to the $\hat{n}$ vector . An object is represented by the brown shape.

Cartesian plane in Figure 37, where an object is illustrated by the brown shade. First, we define the line $l_{t,\theta}$ that passes through point $p_{t,\theta} = (t\cos\theta, t\sin\theta)$ and is perpendicular to the unit vector $\vec{n} = (-\sin\theta, \cos\theta)$. The point $p_{t,\theta}$ makes an angle θ with the x-axis and is a distance t from the origin. Therefore, all possible lines $l_{t,\theta}$ is the set of values

$$\{l_{t,\theta} : t \geqslant 0, 0 \leqslant \theta \leqslant 2\pi\} \tag{3.126}$$

We note that $\vec{n}$ is perpendicular to $(\cos\theta, \sin\theta)$, and therefore every point on the line $l_{t,\theta}$ can be given as the vector sum

$$p_{t,\theta} = t(\cos\theta, \sin\theta) + s(-\sin\theta, \cos\theta) \tag{3.127}$$

with $-\infty < s < +\infty$. Therefore, any line can be parameterized by s as

$$l_{t,\theta} = \{(t\cos\theta - s\sin\theta, t\sin\theta + s\cos\theta) : -\infty < s < +\infty\} \tag{3.128}$$

With such parameterization, we define the Radon Transform $\mathcal{R}$ as the operator that integrates the function $f(x, y)$ over the line $l_{t,\theta}$ by varying the parameter s:

$$\mathcal{R}f(t,\theta) = \int_{l_{t,\theta}} f ds = \int_{s=-\infty}^{+\infty} f(t\cos\theta - s\sin\theta, t\sin\theta + s\cos\theta) ds \tag{3.129}$$

We note further that $\mathcal{R}$ is a linear transformation, that is

$$\mathcal{R}(\alpha f + \beta g) = \alpha\mathcal{R}f + \beta\mathcal{R}g \tag{3.130}$$

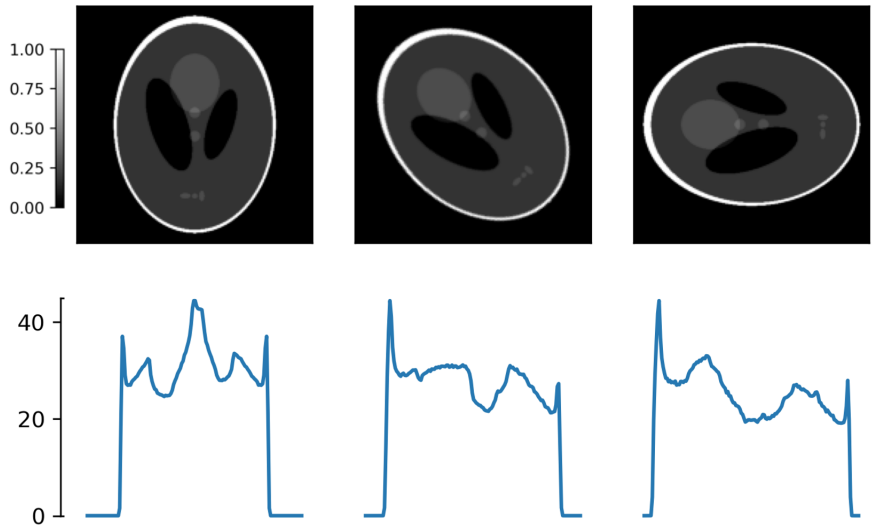


Figure 38 – Shepp-Logan phantom and the corresponding projection in the y-axis for rotations of 0^o, 45^o and 90^o.

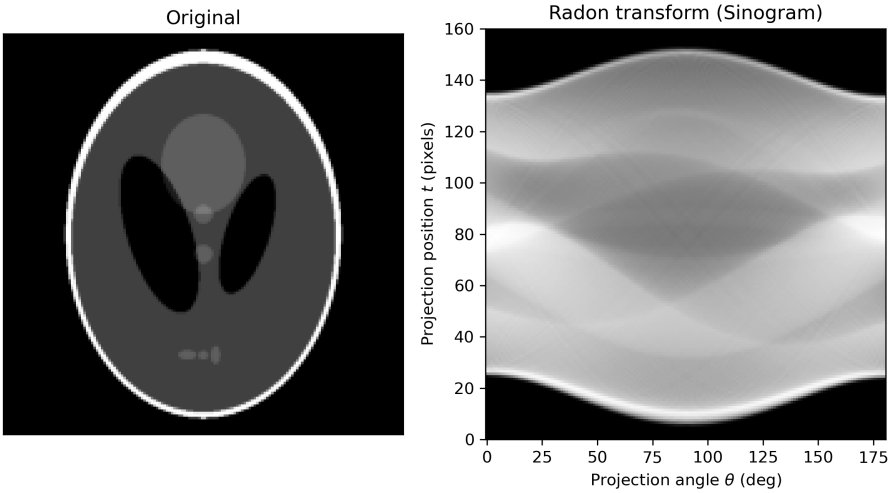


Figure 39 – The Shepp-Logan phantom in (x, y)-coordinates and its sinogram in (θ, t).

In the sense of equation 3.129, the Radon transform can be thought of as a "projection" of the values of the object for a given an angle θ. The famous Shepp-Logan phantom [33] is used in Figure 38 to illustrate this idea. The upper plots show the phantom rotated for angles of 0^o, -45^o and -90^o, whereas the lower plots show the value of their projections over the vertical direction. Rotating the sample by $-\theta$ is equivalent to varying the direction of the projection by θ. The Radon transform is popularly referred to as a *sinogram*. Figure 39 shows Shepp-Logan phantom in the xy-Cartesian plane together with its corresponding sinogram in (θ, t) coordinates. Each column of the sinogram corresponds to a projection for a certain angle, as shown in the lower plots of Figure 38.

In practice, the function f over which we apply the Radon transform can be thought of as the linear absorption coefficient of the sample at each point for absorption contrast, and as a phase-shift value for phase contrast.

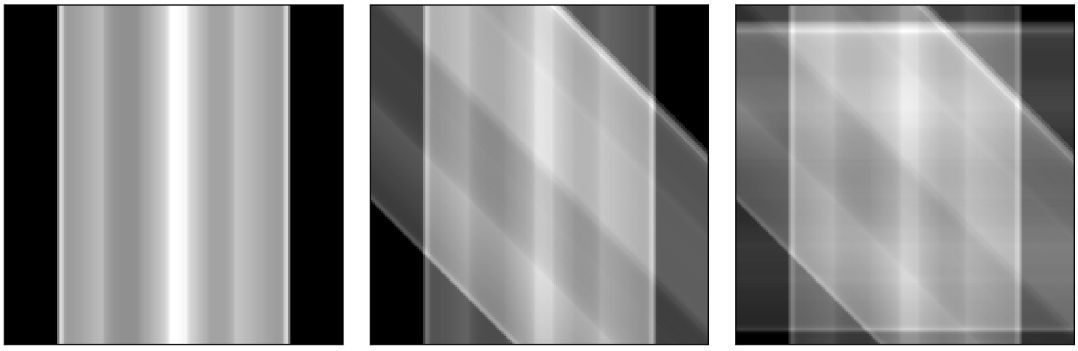


Figure 40 – Illustration of three backprojections of the Shepp-Logan phantom being performed for the angles of 0^o, 45^o and 90^o from left to right.

3.5.2 Image Reconstruction

How can one recover the function $f(x)$ given its sinogram? This is the main question to be answered in order to perform an image reconstruction. We first introduce the concept of a *Backprojection* (BP) and then its modified version, *Filtered Backprojection* (FBP), which gives a perfect reconstruction in an ideal-case.

Backprojection

Given point (x, y) in the plane, for each θ there is one value of t for which $l_{t,\theta}$ passes through the point (x, y): $t = x\cos\theta + y\sin\theta$. In other words, $l_{x\cos\theta+y\sin\theta,\theta}$ passes through (x, y). The Backprojection $\mathcal{B}$ of a function f is defined as

$$\mathcal{B}f(x,y) = \frac{1}{\pi}\int_{\theta=0}^{\pi} f(x\cos\theta + y\sin\theta, \theta)d\theta \tag{3.131}$$

and is nothing more than the average value of point (x, y) for all angles.

For the specific case of the Radon transform, we have $\mathcal{R}f(x\cos\theta + y\sin\theta, \theta)$ as the line integral over the line $l_{t,\theta}$ passing through (x, y). Therefore, the Backprojection of the Radon transform can be thought of as the average of all line integrals $\mathcal{R}f(x, y)$ through which the point (x, y) was accounted for.

$$\mathcal{B}\mathcal{R}f(x,y) = \frac{1}{\pi}\int_{\theta=0}^{\pi} \mathcal{R}f(x\cos\theta + y\sin\theta, \theta)d\theta \tag{3.132}$$

In this sense, the Backprojection $\mathcal{B}\mathcal{R}f(x, y)$ can be thought of as a first "averaged out" guess for the value of f at point (x, y). In Figure 40, we illustrate the Backprojection of the Shepp-Logan phantom for three angle values. At each line in the projection direction, the same value is attributed for all points. The process is repeated for each angle, where the value of the point is updated by summing the value from each projection. At the end, one obtains the smoothed-out image shown in Figure 42 (top-center).

Central Slice Theorem

The *Central Slice Theorem* is an important result that relates the Radon and Fourier Transforms of a function, and it is important when deriving the Filtered Backprojection formula in the next subsection. Although we do not show the demonstration here (see [11]), we show this result since it relates to the mapping of reciprocal space already mentioned in section 3.3.3. The theorem states that

$$\mathcal{F}_2\left\{f(t\cos\theta, t\sin\theta)\right\} = \mathcal{F}\left\{\mathcal{R}f(t,\theta)\right\} \tag{3.133}$$

where $\mathcal{F}$ and $\mathcal{F}_2$ represent the one-dimensional and two-dimensional Fourier Transforms, respectively. In simple words, it states that the Fourier Transform of the Radon transform of a function $\mathcal{F}(\mathcal{R}f)$ is contained in the two-dimensional Fourier Transform of that function $(\mathcal{F}_2 f)$. More precisely, it is a central slice of such 2D Fourier Transform whose direction depends on θ. We illustrate this idea in Figure 41.

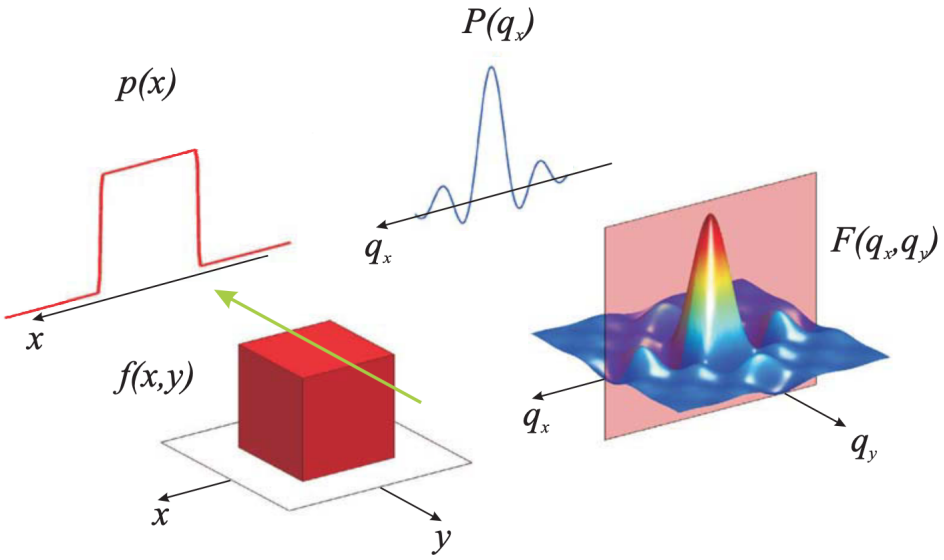


Figure 41 – Illustration of the Central Slice Theorem. The Radon Transform is calculated over object $f(x, y)$, resulting in a projection $p(x)$. The Fourier Transform of this projection is shown besides as $P(q_x)$ and corresponds to a central slice fo the 2D Fourier Transform $F(q_x, q_y)$ of the original object $f(x, y)$. This central slice goes through the origin and will be rotated around the q_z axis for the Radon transform of other θ values. Figure from [17].

In summary, we have that the Radon transform of a function f for certain angle θ gives a projection of f in that direction. For that reason, the Projection Approximation applies. As we saw earlier in Figure 13, the PA assumes that only those straight paths from the input and output planes would affect the wave travelling through the sample. In practice, the function f could represent a map of the linear attenuation coefficient for an absorption measurement, or the electron-density in a scattering experiment away from

absorption edges. Furthermore, looking from a reciprocal-space perspective, the Central Slice Theorem allows us to understand how the PA deviates from the correct result. As we already illustrated in Figure 27, the mapping of measured frequencies in reciprocal-space is given by a section of the Ewald sphere. Nonetheless, the PA assumes that we are actually mapping a plane passing through the origin. Consequently, one can say that the Projection Approximation better represents the true signal for the lower spatial frequencies where a section of the Ewald sphere may be approximated by a plane. In that sense, we expect the errors from the Projection Approximation to be greater for the higher frequencies of the reconstructed image.

Filtered Backprojection (FBP)

With the aid of the Central Slice Theorem, one can show that a slight modification to the Backprojection of the Radon transform actually is capable of perfectly reconstructing the function f, if one has access to all value of θ from 0 to 180^o. This result is know as the Filtered Backprojection of f, given by

$$f(x,y) = \frac{1}{2}\mathcal{B}\mathcal{F}^{-1}[|\nu|\mathcal{F}(\mathcal{R}f)(S,\theta)](x,y) \tag{3.134}$$

The absolute value function $|\nu|$ acts as a filter that multiplies the frequencies of the Radon transform in reciprocal-space. Note that if this function were absent, the Fourier Transforms cancel out, and we recover the Backprojection of the Radon transform from equation 3.132. It is important noticing that in practice one does note have access to all values of θ, but only to a discrete number of angles. The above equality will then only be an approximation, the better the most angles we use. In fact, one needs at least a number N_θ of angles

$$N_\theta = \frac{\pi}{2}N \tag{3.135}$$

to have sufficient information to reconstruct a sample of N^2 pixels, which is known as the Crowther criterion [1]. Figure 42 compares the reconstruction of the Shepp-Logan phantom using both the BP and FBP methods. For FBP, we used a modified version of the absolute value filter, known as the Shepp-Logan filter S_{SL}:

$$S_{SL} = \begin{cases} |\omega|\cos\left(\dfrac{\pi\omega}{2L}\right), & \text{if } |\omega| < L, \\ 0, & \text{otherwise.} \end{cases} \tag{3.136}$$

This filter is a modified version of the absolute-value filter, which differs from it by not attenuating high frequencies values as much, as shown in the lower-left plot of Figure 42. Below each reconstruction, we shown the reconstruction error with respect to the phantom.

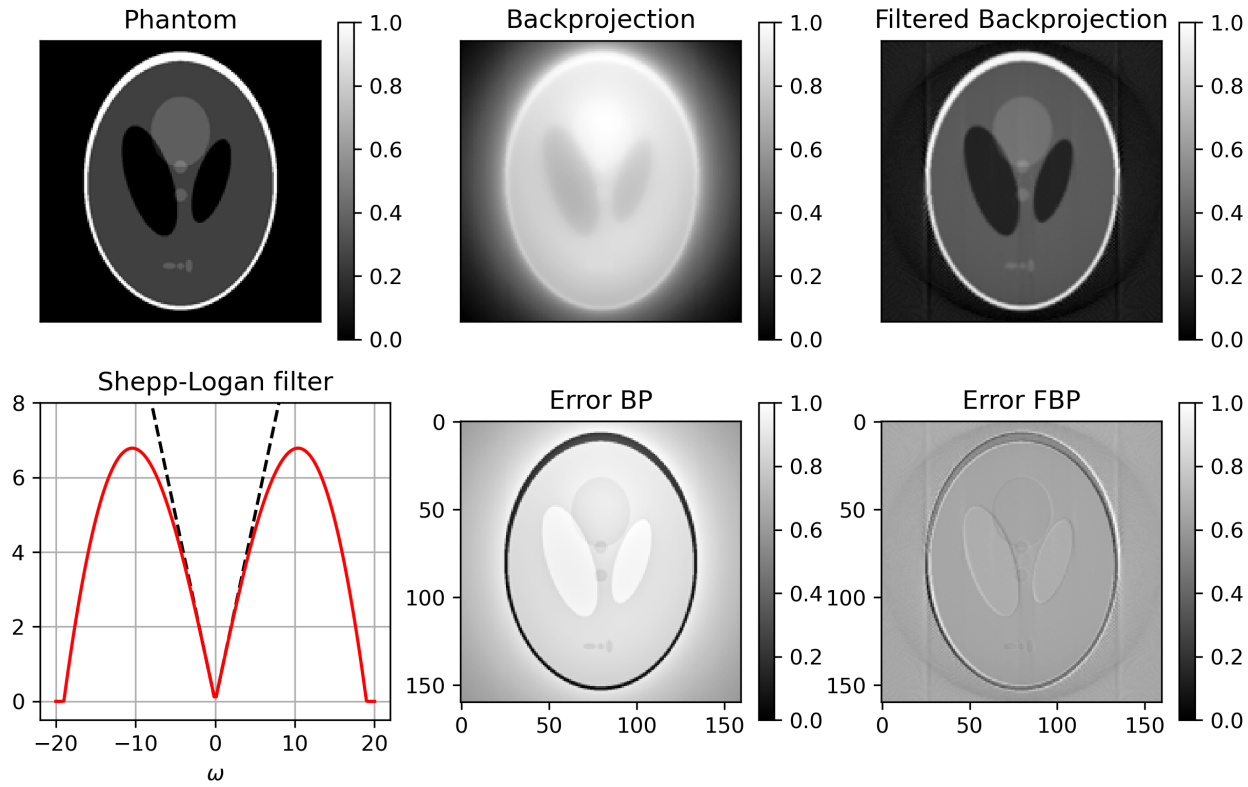


Figure 42 – Comparison between of a $N^2 = 160 \times 160$ pixels Shepp-Logan phantom with its reconstructions using Backprojection and Filtered Backprojection. The rightmost plot shows the error between the model and the FBP reconstruction. The values of θ were varies from 0^o to 180^o with steps of 0.5^o, totalling 360 angles. According to the Crowther criterion, one would need $N_\theta = 160\pi/2 \approx 252$ angles in this case to have sufficient information for reconstruction. The Shepp-Logan filter used in the FBP was calculated for $L = 19$ and $-20 < \omega < 20$ and is shown and the continuous-red curve in the lower left plot, together with the absolute value function (dashed-black line). Mean error value for reconstructions are $E_{\text{BP}} = 0.55$ and $E_{\text{FBP}} = 0.07$.

Table 1 – Maximum, mean and standard deviation values of the FBP reconstruction errors using the two filters.

Filter	**Max**	**Mean**	**StdDev**
Absolute value	0.67	0.040	0.077
Shepp-Logan	0.71	0.067	0.070

Note that all colorbars are in the same scale, and it becomes evident that the error of the BP reconstruction is much higher than the error of FBP.

We end this section by noting that the reconstruction of the FBP equation 3.134 is exact when using the absolute value filter, so we would in principle expect the reconstruction to be better using such filter. We compare the reconstruction using both filters in Figure 43. In this specific case, FBP with absolute-value filter does present a

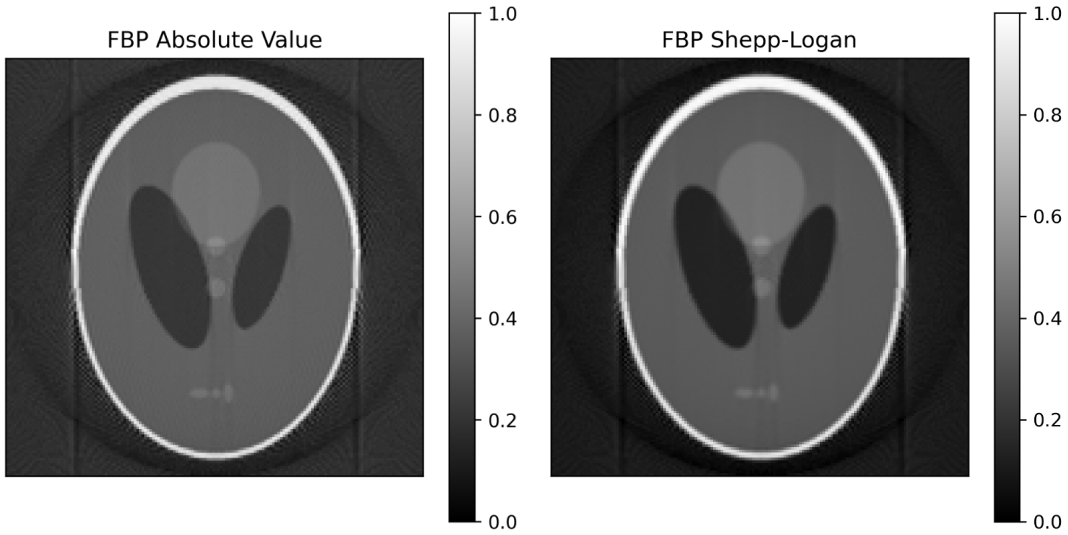


Figure 43 – Comparison between FBP reconstruction using the absolute-value and Shepp-Logan filters. The latter case present bigger maximum and mean errors, but presents better contrast with respect to the original. The original phantom was also normalized.

lower maximum and mean error value, as shown in Table 1. Nonetheless, the Shepp-Logan filter does not remove much of the high-values of the image, resulting in a contrast that better resembles the original images in a normalized-color plot.

3.5.3 Algebraic Reconstruction Techniques

The FBP algorithm has become one of the most used in Tomography reconstruction because of its speed. The fact that it is based on the Fourier Transform allows one to perform fast computations using the well established Fast-Fourier Transform (FFT) algorithm. Nonetheless, it relies on assuming the Projection Approximation, which is precisely what we wanted to avoid in this work. In that case, we need to rely on an older and more slow algorithm to perform the second step of the reconstruction. This is known as the Kaczmarz's method, popularly referred to as the *Algebraic Reconstruction Technique* (ART). In section 4.1, we will show that our proposed discrete model of the inhomogenous Helmholtz equation allows one to use ART to perform the tomographic reconstruction. To understand ART, we first refer to a few definitions and theorems ahead, as given by [11]. Given a linear system

$$A\mathbf{x} = \mathbf{p}, \tag{3.137}$$

where $\mathbf{x}$ and $\mathbf{p}$ are N-dimensional vectors and A is a $M \times N$ matrix, the ART algorithm can be used to find an approximate solution $\hat{\mathbf{p}}$ to such a system. By "approximate" we mean the vector $\hat{\mathbf{p}}$ that is the closest to $\mathbf{p}$ in the least-squares sense, that is, we look for

the vector $\hat{\mathbf{x}}$ such that

$$\|A\hat{\mathbf{x}} - p\| = \min_{\mathbf{x}\in\mathbb{R}^N} \|A\mathbf{x} - p\| \tag{3.138}$$

To understand ART, we need first to introduce the concepts of an *affine space* and of an *affine projection*:

Definition 1 (Affine space)**.** *For a fixed n-dimensional vector* $\mathbf{r}$ *and a number* p*, the affine space* $S_{\mathbf{r},p}$ *is defined as the set*

$$S_{\mathbf{r},p} = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{r} \cdot \mathbf{x} = p\} \tag{3.139}$$

The affine space differs from a vector subspace in the sense that the latter must contain the origin. An affine space will be a subspace of $\mathbb{R}^n$ if, and only if, $p = 0$. Therefore, the affine space can be thought of a copy of a subspace that has been shifted by a fixed vector. One important aspect that we will use ahead, is that the vector $\mathbf{r}$ is orthogonal to $S_{\mathbf{r},p}$ [11].

Definition 2 (Affine projection)**.** *Given a vector* $\mathbf{u}$ *and an affine space* $S_{\mathbf{r},p}$ *for some vector* $\mathbf{r}$ *and some number* p*, the affine projection of* $\mathbf{u}$ *in* $S_{\mathbf{r},p}$ *is the vector* $\mathbf{u}^*$ *in* $S_{\mathbf{r},p}$ *that is closest to* $\mathbf{u}$ *amongst all vectors in* $S_{\mathbf{r},p}$*.*

The idea behind an affine projection is illustrated in Figure 44. We wish to find the vector $\mathbf{u}^*$ that gives the closest point in $S_{\mathbf{r},p}$ that is the closest to $\mathbf{u}$. Therefore, we start at vector $\mathbf{u}$ and have to move in the direction orthogonal to $S_{\mathbf{r},p}$. Since, $\mathbf{r}$ is orthogonal to $S_{\mathbf{r},p}$, we can move in its direction. Therefore, for some real value λ:

$$\mathbf{u}^* = \mathbf{u} - \lambda\mathbf{r} \tag{3.140}$$

One can show that [11] λ is given by

$$\lambda = \frac{(\mathbf{r} \cdot \mathbf{u}) - p}{\mathbf{r} \cdot \mathbf{r}} \tag{3.141}$$

Therefore, equation 3.140 becomes

$$\mathbf{u}^* = \mathbf{u} - \frac{(\mathbf{r} \cdot \mathbf{u}) - p}{\mathbf{r} \cdot \mathbf{r}} \tag{3.142}$$

With these equations, we are able to state the iterative formula of the ART algorithm. Given the linear system $A\mathbf{x} = \mathbf{p}$, each equation $\mathbf{r}_i\mathbf{x} = p_i$ defines an affine space. Therefore, given an initial guess $\mathbf{x^0}$, Kaczmarz's method says

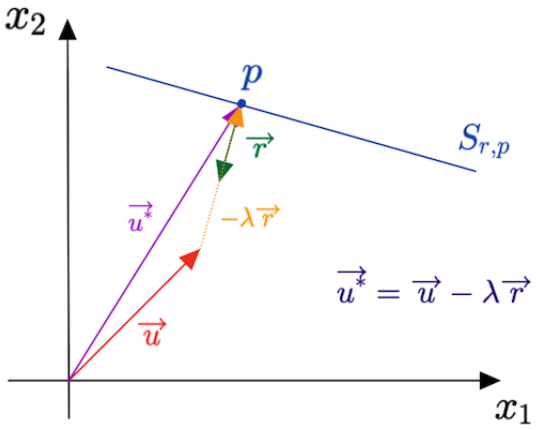
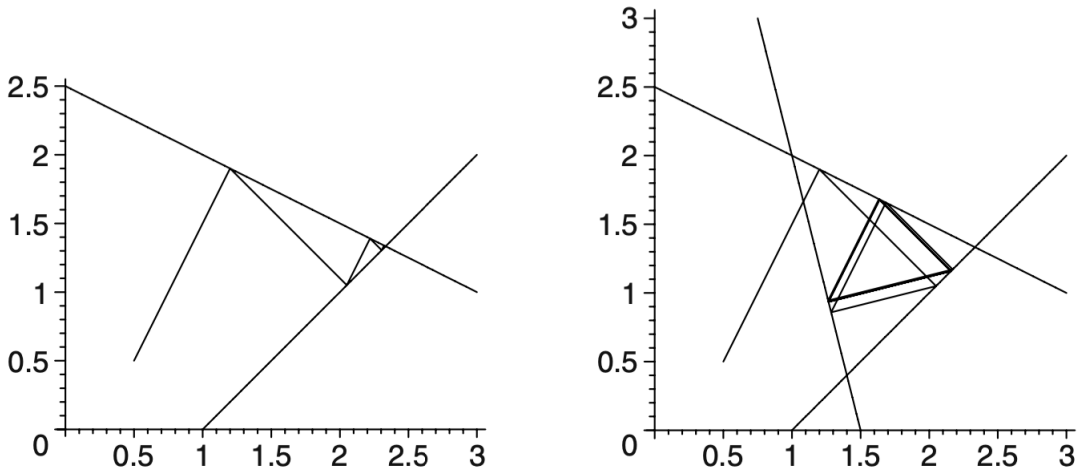


Figure 44 – Illustration of the affine projection in $\mathbb{R}^2$ into the affine space $S_{\mathbf{r},p}$



[!ht]

Figure 45 – Illustration of consecutive projection between two (left) and three (right) affine spaces. On the former, the solution quickly converges to a single points, whereas for the latter the solution gets stuck in a triangle. Figure from [11].

$$\mathbf{x^{i,e}} = \mathbf{u^{i,e-1}} - \frac{(\mathbf{r_i} \cdot \mathbf{u^{i,e-1}}) - p_i}{\mathbf{r_i} \cdot \mathbf{r_i}} \tag{3.143}$$

where the superscript e indicates the number M of equations of the linear system (i.e. the number of lines of matrix A), whereas the superscript i indicates the number of K iterations of the ART algorithm. Starting from an initial guess $\mathbf{x^{0,0}}$, one applies the above formula consecutively for $\mathbf{x^{0,1}}$, $\mathbf{x^{0,2}}$, ..., $\mathbf{x^{0,M}}$, ..., $\mathbf{x^{K,0}}$, ..., $\mathbf{x^{K,M}}$. By doing this, the following theorem guarantees the existence of a solution:

Theorem 1. *If the linear system $A\mathbf{x} = \mathbf{p}$ has at least one solution, then Kaczmarz's method converges to a solution of this system. Moreover, if $\mathbf{x^0}$ is in the range of A^T , then Kaczmarz's method converges to the solution of minimum norm.*

Figure 45 illustrates two scenarios for the convergence of the ART iterative

algorithm for $\mathbf{x} \in \mathbb{R}^2$, which help us get an intuitive notion of how the method works. The idea is that consecutive projections in each of the affine spaces will lead to the solution of the system. On the left, we have $M = 2$, meaning two affine spaces, and the consecutive affine projections quickly converge to the intersection between the affine spaces. On the right, for $M = 3$ we see that the method stagnates into a triangle pattern.

In summary, given a linear system, one can find a least-squares solution to it via the ART iterative method. Nonetheless, the fact that one needs to iterate through all equations of the system one by one indicates that this method can become slow for a large number of equations. Indeed, as we shall see, this became the main issue of the proposed M-matrix method and, consequently, the biggest and yet to be resolved challenge faced during the research.

4 Results

In this chapter we present the main results obtained throughout this Masters project. We start by presenting the core novelty of this work: the numerical modelling of the inhomogenous Helmholtz equation that culminates in a system-matrix form of this equation. We call this matrix the "M-matrix". Afterwards, we present the simulations we did to solve the Phase-Retrieval problem. We show implementations for the Error Reduction, Hybrid Input-Output and a mixture of both algorithms together. The third section shows simulations of the M-matrix model being used in conjunction with ART to solve the tomographic step of the reconstruction, which we call "M-ART". Although it indicates the efficacy of the ART algorithm using the M-matrix for obtaining a 3D image, we can not yet affirm it considerably improves the reconstruction quality in comparison to the Projection Approximation because of the occurrence of an "inverse-crime", as we shall discuss in the last section.

By "inverse-crime" we mean the act of synthesizing the input data for the reconstruction using the same model that does the reconstruction afterwards. In that scenario, the reconstruction may be successful and precise not necessarily because of the quality of the method itself. The idea of avoiding such an inverse-crime was a goal from the beginning of this project. It was one of the reasons we chose to focus on Plane-Wave CDI instead of Ptychography. By using Plane-Wave CDI, we would in principle be able to validate the model using analytical diffraction patterns for simple geometries, such as the spherical one presented in section 3.2.5. Therefore, one would be able to avoid the inverse-crime by starting the reconstruction from "the end": first performing phase-retrieval over the analytical diffraction pattern, and afterwards performing the M-matrix ART three-dimensional reconstruction. Nonetheless, the unanticipated huge computational cost of the proposed method introduced difficulties to this original plan, because we had to work with extremely small simulated samples ($N < 65$). The numerical errors due to such small resolution prevented us from validating M-ART in the absence of the inverse-crime.

4.1 The M-matrix model

We start by introducing the new proposed numerical model of the inhomogenous Helmholtz equation in three-dimensions. The method herein presented is the generalization of the initial work from the advisor and collaborators [12], which showed the M-matrix equation for the simpler two-dimensional case. In this work, we present for the first time the generalized three-dimensional case together with the proofs for the respective equations.

We start from the paraxial Helmholtz equation (3.33) and assume no absorption $\beta = 0$. Furthermore, we assume a perfect incident plane of the form

$$\psi(x, y, z) = \psi_0 \exp(ikz), \qquad \psi_0 = \text{const.} \tag{4.1}$$

In that case, equation 3.33 becomes

$$\left[i2k\frac{\partial}{\partial z} + \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} - 2k^2\delta(x, y, z) \right] \psi(x, y, z) = 0 \tag{4.2}$$

We then discretize the above equation, using the following forms of the first (forward-difference) and second derivatives:

$$\frac{\partial\psi(x, y, z)}{\partial z} \approx \frac{\psi(x, y, z+1) - \psi(x, y, z)}{\Delta z} \tag{4.3}$$

$$\frac{\partial^2\psi(x, y, z)}{\partial x^2} \approx \frac{\psi(x+1, y, z) + \psi(x-1, y, z) - 2\psi(x, y, z)}{\Delta x^2} \tag{4.4}$$

$$\frac{\partial^2\psi(x, y, z)}{\partial y^2} \approx \frac{\psi(x, y+1, z) + \psi(x, y-1, z) - 2\psi(x, y, z)}{\Delta y^2} \tag{4.5}$$

Figure 46 illustrates the discretization of our space in the xz directions, where the incident wave-field is ψ_0 and ψ_n represent the wavefield at the exit of voxel n. Note that the indices span $1 < x < X$, $1 < y < Y$ and $1 < z < z_0$. Substituting the numerical derivatives in the differential equation and considering $\Delta x = \Delta y = \Delta z = \Delta$, we obtain the following recursive relation:

$$\begin{aligned} \psi(x, y, z+1) = &\left(1 + \frac{2}{ik\Delta} - ik\Delta\delta(x, y, z+1)\right)\psi(x, y, z) + \\ &- \frac{1}{2ik\Delta}\left(\psi(x+1, y, z) + \psi(x-1, y, z) + \psi(x, y+1, z) + \psi(x, y-1, z)\right) \end{aligned} \tag{4.6}$$

Note that, in the left-hand side, the complex amplitude ψ is calculated for a point at the $z + 1$ plane, whereas the ψ terms in the right-hand side are all for plane z. In other words, one uses five values of the wave amplitude at a plane z to calculate a single point of the wave at the consecutive $z + 1$ plane.

Our goal, nonetheless, is to solve the inverse problem: we know the complex amplitude $\psi(x, y, z = z_0)$ at the exit plane of the sample for all x and y, but we do not know the values $\delta(x, y, z)$ (i.e. the sample). Hence, we cannot perform a propagation in the above manner.

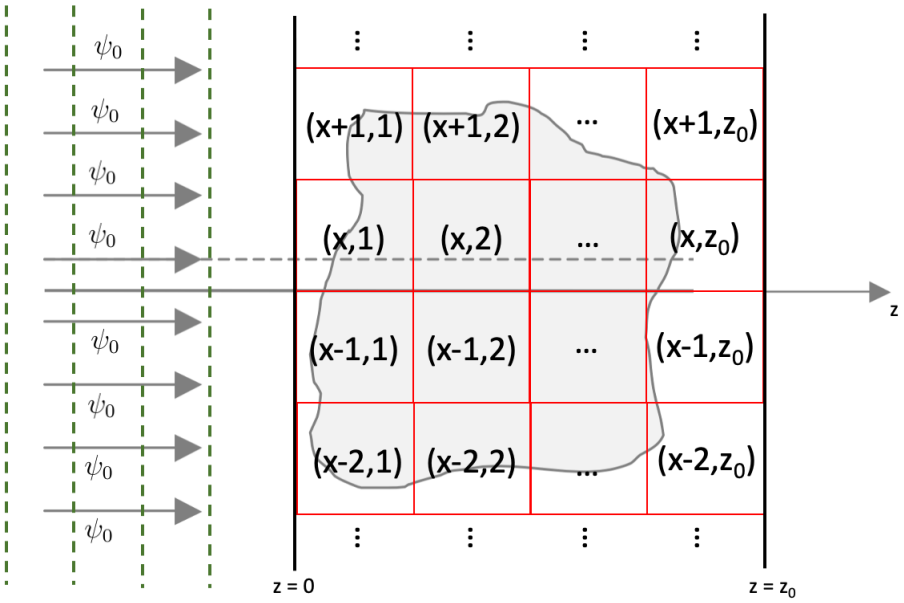


Figure 46 – Convention of indices for the discretization of 2D-space. Figure adapted from [10].

4.2 General Matrix form of the IHE

To solve the inverse-problem, we propose a matrix form of the IHE, modelling the radiation-sample interaction by a simple Hadamard product[1] of the system M-matrix and the sample δ matrix. We shall prove by mathematical induction that the wavefield at voxel $\psi(x_0, y_0, z_0)$ is given by the general formula

$$\psi(x_0, y_0, z_0) = \psi_0(1 - \underset{x_0,y_0,z_0}{M} \circ \delta_M) = \psi_0 \left(1 - \sum_{x=1}^{X} \sum_{y=1}^{Y} \sum_{z=1}^{z_0} \underset{x_0,y_0,z_0}{M}(x, y, z)\delta(x, y, z)\right) \tag{4.7}$$

where $\underset{x_0,y_0,z_0}{M}$ is the system M-matrix that generates a point of the wave-field $\psi(x_0, y_0, z_0)$ and δ_M is the matrix of the refractive index values at each voxel.

To ease the mental picture of what we are modelling, we anticipate an illustration of the M-matrix and compare it with the Projection Approximation scenario in Figure 47. The plots show the absolute value of the system-matrices for the PA and M scenarios. The system-matrix is essentially a matrix that weights the contribution of each voxel of the sample to the wave at a particular point in the output plane. One can think of the Projection Approximation as the case in which the system-matrix contains a single line of non-null voxels (Figure 47, left). On the other hand, our system M-matrix presents a pyramid shape (Figure 47, right), showing a whole region of the sample actually contributes to the wavefield, but gets ignored in the PA.

[1] The Hadamard product, here denoted by $\circ$, is the element-wise multiplication followed by sum of resulting matrix terms.

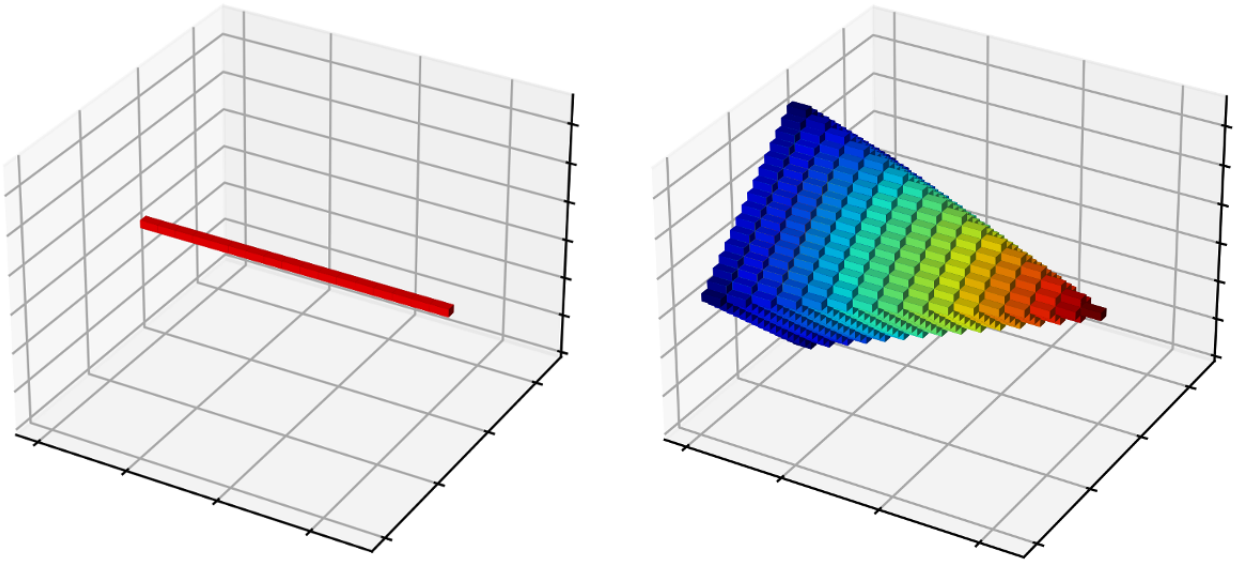


Figure 47 – The absolute value of the system matrices for a small discretized space with $N^3 = 17^3$ voxels. On the left, we have the system-matrix for the Projection Approximation case (PA-matrix), whereas on the right we have our system-matrix (M-matrix) obtained without the PA. The color code illustrates higher to lower values from warmer to colder colors in log-scale.

We emphasize that the above model gives **one matrix for each point** $\psi(x_0, y_0, z_0)$ in the output plane. That is, we perform an element-wise multiplication of the M-matrix $\underset{x_0,y_0,z_0}{M}(x, y, z)$ with the object voxel values $\delta(x, y, z)$, followed by the sum of all voxels of the resulting matrix, to generate the single point $\psi(x_0, y_0, z_0)$ in the output plane.

In the following, we demonstrate the general formula 4.7 through mathematical induction. In other words, we first prove that it obeys 4.7 for $z = 0$. Afterwards, assuming it works for z, we prove that it also works for $z + 1$.

- **Base case ($z = 0$):**

  For the base case, we start at plane z=0, which gives

$$\psi(x, y, 1) = \left(1 + \frac{2}{ik\Delta} - ik\Delta\delta(x, y, 1)\right)\psi_0 - \frac{1}{2ik\Delta}\left(\psi_0 + \psi_0 + \psi_0 + \psi_0\right) \tag{4.8}$$

hence,

$$\frac{\psi(x, y, 1)}{\psi_0} - 1 = -ik\Delta\delta(x, y, 1) \tag{4.9}$$

which indeed, is an equation of the form of the wanted general formula 4.7. For the consecutive planes, we use this result back in the recursive formula for $z = 1$, and so on, consecutively. For the next plane, for instance, one would get[2]:

[2] We neglect second-order terms in δ since $\delta^2 \approx 0$ for X-Rays.

$$\frac{\psi(x,y,2)}{\psi_0} - 1 = -(2 + ik\Delta)\delta(x,y,1) + \\ + \frac{1}{2}(\delta(x+1,y,1) + \delta(x-1,y,1) + \delta(x,y+1,1) + \delta(x,y-1,1)) - ik\Delta\delta(x,y,2) \quad (4.10)$$

- **Inductive case:** $(z \to z+1)$

Having shown that the general formula holds for $z = 0$, to finish our proof of the general formula we must show that, if the formula holds for z, so it must for $z+1$. For such, we substitute the ψ terms in the recursive formula 4.6 using 4.7:

$$\frac{\psi(x_0,y_0,z_0+1)}{\psi_0} = \left(1 + \frac{2}{ik\Delta} - ik\Delta\delta(x_0,y_0,z_0+1)\right)\left(1 - \sum_{x=1}^{X}\sum_{y=1}^{Y}\sum_{z=1}^{z_0} \underset{x_0,y_0,z_0}{M(x,y,z)}\delta(x,y,z)\right) \\ - \frac{1}{2ik\Delta}\left(4 - \sum_{x=1}^{X}\sum_{y=1}^{Y}\sum_{z=1}^{z_0}\left(\underset{x_0+1,y_0,z_0}{M(x,y,z)} + \underset{x_0-1,y_0,z_0}{M(x,y,z)} + \underset{x_0,y_0+1,z_0}{M(x,y,z)} + \underset{x_0,y_0-1,z_0}{M(x,y,z)}\right)\delta(x,y,z)\right) \quad (4.11)$$

Distributing the multiplications, we get

$$\frac{\psi(x_0,y_0,z_0+1)}{\psi_0} = 1 - \sum_{x=1}^{X}\sum_{y=1}^{Y}\sum_{z=1}^{z_0} \underset{x_0,y_0,z_0}{M(x,y,z)}\delta(x,y,z) + \frac{2}{ik\Delta} + \\ - \frac{2}{ik\Delta}\sum_{x=1}^{X}\sum_{y=1}^{Y}\sum_{z=1}^{z_0} \underset{x_0,y_0,z_0}{M(x,y,z)}\delta(x,y,z) - ik\Delta\delta(x_0,y_0,z_0+1) - \frac{2}{ik\Delta} + \\ + \frac{1}{2ik\Delta}\sum_{x=1}^{X}\sum_{y=1}^{Y}\sum_{z=1}^{z_0}\left(\underset{x_0+1,y_0,z_0}{M(x,y,z)} + \underset{x_0-1,y_0,z_0}{M(x,y,z)} + \underset{x_0,y_0+1,z_0}{M(x,y,z)} + \underset{x_0,y_0-1,z_0}{M(x,y,z)}\right)\delta(x,y,z) \quad (4.12)$$

Note that non-linear terms $\frac{2}{ik\Delta}$ cancel each other out. Therefore, we are left with

$$\frac{\psi(x_0,y_0,z_0+1)}{\psi_0} = 1 - \left(1 - \frac{2}{ik\Delta}\right)\sum_{x=1}^{X}\sum_{y=1}^{Y}\sum_{z=1}^{z_0} \underset{x_0,y_0,z_0}{M(x,y,z)}\delta(x,y,z) + \\ + \frac{1}{2ik\Delta}\sum_{x=1}^{X}\sum_{y=1}^{Y}\sum_{z=1}^{z_0}\left(\underset{x_0+1,y_0,z_0}{M(x,y,z)} + \underset{x_0-1,y_0,z_0}{M(x,y,z)} + \underset{x_0,y_0+1,z_0}{M(x,y,z)} + \underset{x_0,y_0-1,z_0}{M(x,y,z)}\right)\delta(x,y,z) + \\ - ik\Delta\delta(x_0,y_0,z_0+1) = 1 - \underset{x_0,y_0,z_0+1}{M} \circ \delta_M = 1 - \sum_{x=1}^{X}\sum_{y=1}^{Y}\sum_{z=1}^{z_0+1} \underset{x_0,y_0,z_0+1}{M(x,y,z)}\delta(x,y,z) \quad (4.13)$$

where the equality in the last line is once again the application of the 4.7, now for $\psi(x_0,y_0,z_0+1)$. If we separate the z_0+1 term in the last summation of the right-hand

side and group the terms within the summation on the left-hand side, we get the following equality, where the 1s have been canceled out:

$$\begin{aligned}
&\sum_{x=1}^{X}\sum_{y=1}^{Y}\sum_{z=1}^{z_0} -\left(1-\frac{2}{ik\Delta}\right)\underset{x_0,y_0,z_0}{M}(x,y,z)\delta(x,y,z)+\\
&\quad+\sum_{x=1}^{X}\sum_{y=1}^{Y}\sum_{z=1}^{z_0}\frac{1}{2ik\Delta}\left(\underset{x_0+1,y_0,z_0}{M}(x,y,z)+\underset{x_0-1,y_0,z_0}{M}(x,y,z)+\underset{x_0,y_0+1,z_0}{M}(x,y,z)+\underset{x_0,y_0-1,z_0}{M}(x,y,z)\right)\delta(x,y,z)\\
&-ik\Delta\delta(x_0,y_0,z_0{+}1)=\sum_{x=1}^{X}\sum_{y=1}^{Y}\sum_{z=1}^{z_0}-\underset{x_0,y_0,z_0+1}{M}(x,y,z)\delta(x,y,z)-\sum_{x=1}^{X}\sum_{y=1}^{Y}\underset{x_0,y_0,z_0+1}{M}(x,y,z_0+1)\delta(x,y,z_0{+}1)
\end{aligned} \tag{4.14}$$

For the above equality to hold, the coefficients that multiply the same $\delta(x,y,z)$ must be equal. Therefore, the inductive case is valid as long as we have:

- For $z = z_0 + 1$:

$$\underset{x_0,y_0,z_0+1}{M}(x,y,z_0+1)=\begin{cases} ik\Delta & \text{if } x=x_0 \text{ and } y=y_0\\ 0 & \text{, otherwise}\end{cases} \tag{4.15}$$

- For $z \leqslant z_0$:

$$\begin{aligned}
\underset{x_0,y_0,z_0+1}{M}(x,y,z)=&\left(1+\frac{2}{ik\Delta}\right)\underset{x_0,y_0,z_0}{M}(x,y,z)+\\
&-\frac{1}{2ik\Delta}\left(\underset{x_0+1,y_0,z_0}{M}(x,y,z)+\underset{x_0-1,y_0,z_0}{M}(x,y,z)+\underset{x_0,y_0+1,z_0}{M}(x,y,z)+\underset{x_0,y_0-1,z_0}{M}(x,y,z)\right)
\end{aligned} \tag{4.16}$$

Note that, due to the symmetry of the problem and the fact that the input wave ψ_0 is constant, the matrix $\underset{x_0\pm1,y_0\pm1,z_0}{M}(x,y,z)$ that generates a neighbor point to (x_0,y_0,z_0) is simply a shifted version of $\underset{x_0,y_0,z_0}{M}(x,y,z)$. That is, we have the following equality for the matrices of different points

$$\underset{x_0,y_0,z_0}{M}(x,y,z)=\underset{x_0\pm1,y_0,z_0}{M}(x\pm1,y,z)=\underset{x_0,y_0\pm1,z_0}{M}(x,y\pm1,z) \tag{4.17}$$

This simplifies equation (4.16) considerably, since we can write the matrix for ψ_{x_0,y_0,z_0+1} strictly as a function of the matrix for ψ_{x_0,y_0,z_0} at points $(x{+}1,y,z)$, $(x{-}1,y,z)$, $(x,y{+}1,z)$ and $(x,y-1,z)$:

$$\begin{aligned}
\underset{x_0,y_0,z_0+1}{M}(x,y,z)=&\left(1+\frac{2}{ik\Delta}\right)\underset{x_0,y_0,z_0}{M}(x,y,z)\\
&-\frac{1}{2ik\Delta}\left(\underset{x_0,y_0,z_0}{M}(x-1,y,z)+\underset{x_0,y_0,z_0}{M}(x+1,y,z)+\underset{x_0,y_0,z_0}{M}(x,y-1,z)+\underset{x_0,y_0,z_0}{M}(x,y+1,z)\right)
\end{aligned} \tag{4.18}$$

In summary, for the general formula to hold, equations (4.15) and (4.18) must be satisfied. Equation (4.18) states that five points of the M-matrix of thickness z can be used to obtain a point of the M-matrix of thickness $z+1$ (i.e. the matrix that models a thicker sample). By repeating the above formula for all (x, y) values in each plane, one can in principle construct an M-matrix of arbitrary thickness.

4.2.1 Recursive relation for M-matrix of arbitrary size

Although equation 4.18 allows us to obtain the M-matrix that models the radiation-sample interaction, it would be a very time consuming process since an entire system matrix of size $N \times N \times N$ would be needed to construct another of size $N \times N \times (N-1)$. This latter matrix, on the other hand, would required yet another matrix of size $N \times N \times (N-2)$, and so forth. Fortunately, the symmetry of the problem allows us to build a M-matrix of arbitrary size in a fast way. We illustrate this idea qualitatively through Figure 48. Equation 4.15 gives the last z-plane of the M-matrix. Therefore, the M-matrix $\underset{x_0,y_0,0}{M}$ containing a single plane will be given by such formula. This matrix for the central pixel (x_0, y_0) is illustrated on the top-left corner of Figure 48. This matrix is then used to build the next M-matrix $\underset{x_0,y_0,1}{M}$, containing two planes in the z-direction, as shown in the next column of the figure. Once again, the last line (now the plane $z = 1$) is given by equation 4.15. To obtain the plane $z = 0$, we then use equation 4.16.

If we continue this process indefinitely, we are able to build at each step a "thicker" M-matrix in the z-direction, as illustrated by the following lines in the figure. At each step, the process is the same: generate the last plane using equation 4.15, then use equation 4.16 to obtain the previous plane from the next. Note from Figure 48 that, from one M-matrix to another, the difference is that we are simply adding a new plane at $z = 0$. The remaining planes are identical to the planes of the previous matrix. Consequently, one can generate a matrix of arbitrary size simply by following the algorithm illustrated in Figure 49: start by defining an empty matrix of the arbitrary size wanted; define the last plane with equation 4.15; then, "back-propagate" the matrix with the aid of equation 4.16.

Therefore, the back-propagation strategy allows us to quickly calculate an M-matrix as needed. We reinforce here that the M-matrix is complex-valued and that the size in computer memory required to store it quickly increases with the number N of pixels, as shown in Table 2. This fact means we are not able to previously calculate and store all the M-matrices needed for the ART reconstruction. Instead, we store a single M-matrix in memory and perform operations on it (translations and rotations) to obtain the correct M-matrix at each iteration of the algorithm. As we shall see ahead, this necessity became problematic for the strict validation of the proposed model, due to the time it takes to operate sequentially with the M-matrix at each iteration.

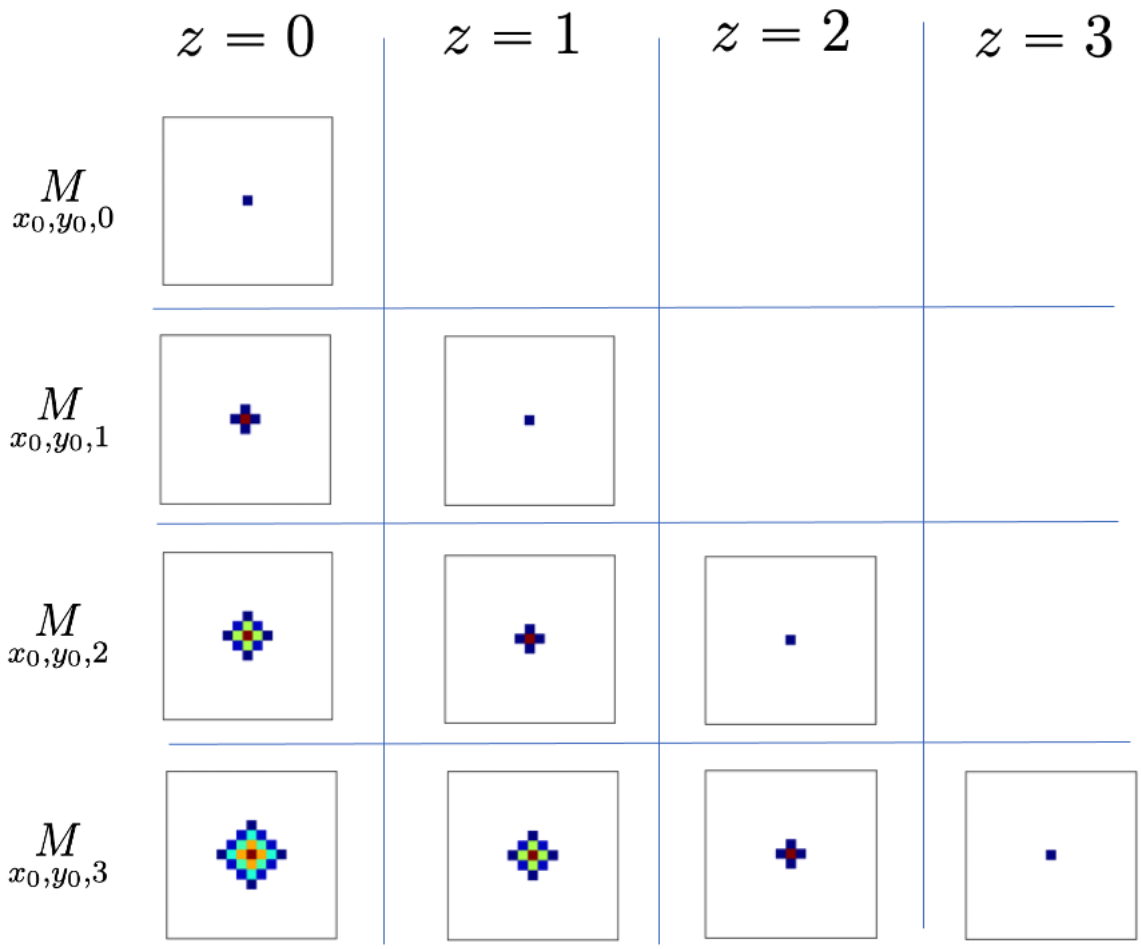


Figure 48 – Illustration of the z planes of each M-matrix of different sizes. Each new matrix differs from the previous one by the a new plane in the $z = 0$ column.

Table 2 – Size required in bytes to store a M-matrix of N^3 pixels. The value $N = 3072$ in the last column shows how big an M-matrix would in the worst case scenario for the Cateretê beamline which has a detector with 3072×3072 pixels[2].

N	**16**	**32**	**64**	**128**	**256**	**512**	**3072**
Size	0.0625 MBs	0.5 MBs	4.0 MBs	32.0 MBs	256 MBs	2 GBs	432 GBs

4.3 Phase Retrieval Simulations

The original idea of the project was to implement and validate the algorithms concerning the first and second-steps of the two-step reconstruction process separately, and only afterwards merge the whole process together. As we shall see, this full simulation also became problematic to perform. In this section we present our results for the Phase Retrieval simulations.

We first simulated the phase-retrieval of the real-valued image shown in Figure 50 using the HIO algorithm. The support was placed tightly around the image, and the initial values of the magnitude and phase were the rough guesses in Figure 50. The simulated diffraction pattern was obtained by first calculating the Fast-Fourier Transform

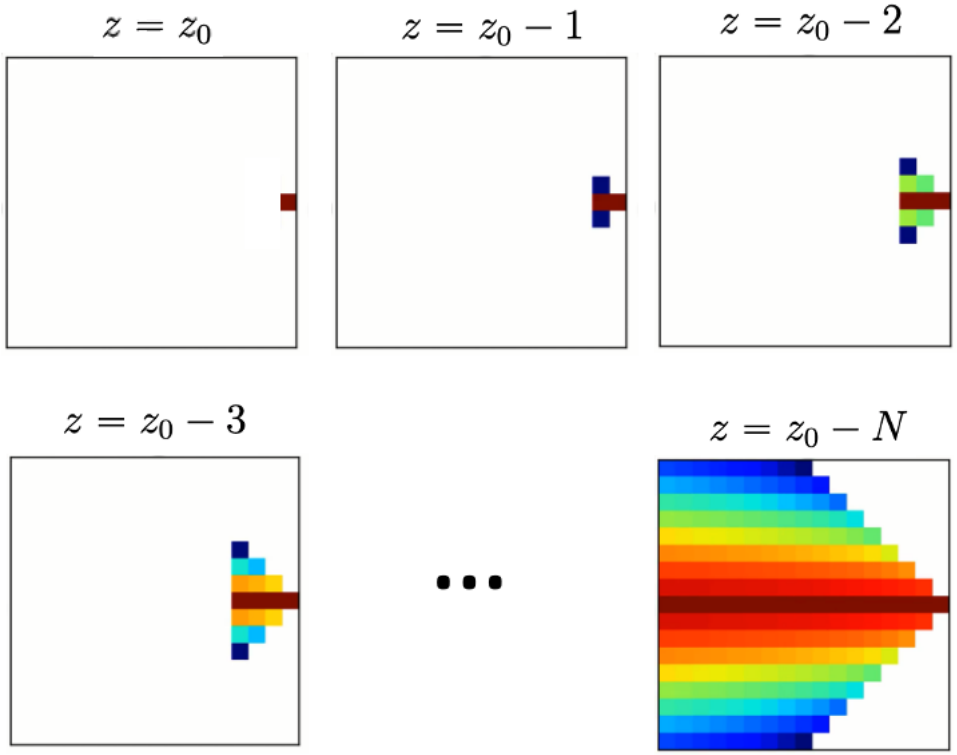


Figure 49 – Illustration of the "back-propagation" strategy to quickly create a M-matrix of arbitrary size.

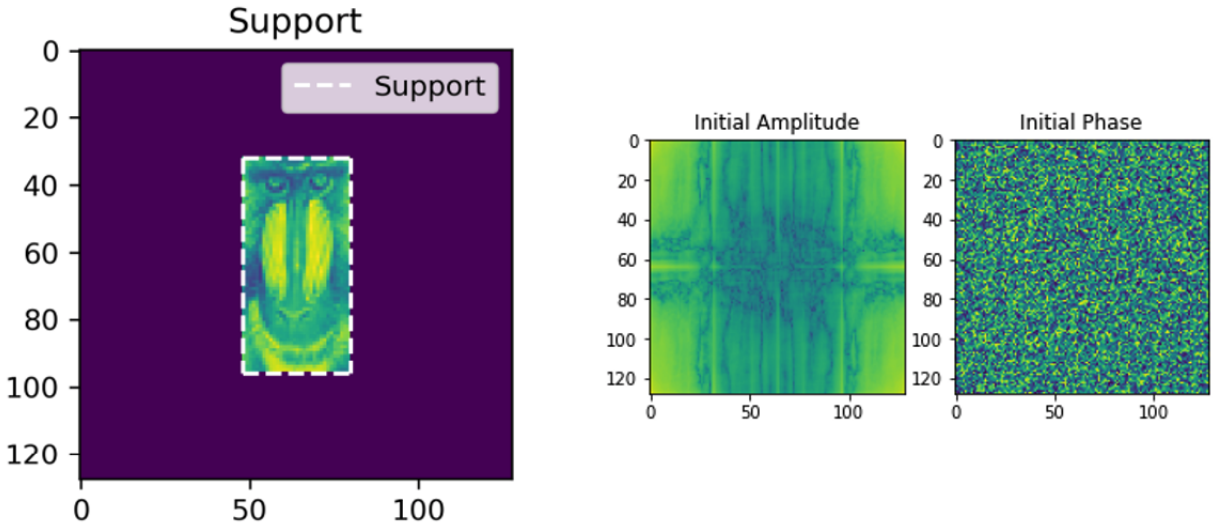


Figure 50 – Image to be reconstructed with the HIO algorithm is shown on the left, together with the used support region. The initial guesses for amplitude and phase are shown on the right.

of the original image and then obtaining the intensity from its absolute value squared (see equation 3.7).

Figure 51 shows the simulation results after 200 iteration of HIO. The right-most column shows the reconstruction error along the iterations. The frequency-space error E_f was calculated comparing the simulated diffraction pattern $|F(k)|$ with the reconstructed

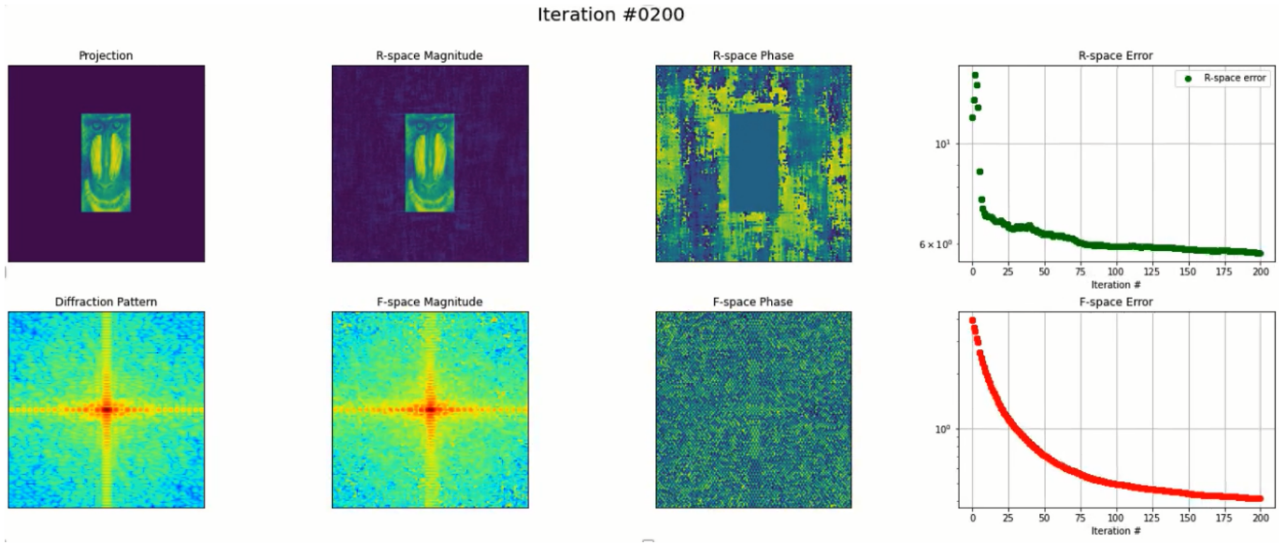


Figure 51 – Phase retrieval result after 200 iterations of the HIO algorithm. On the top row we have, from left to right: original image to be reconstructed, real-space magnitude of the reconstructed signal, real-space phase of the reconstructed signal, real-space reconstruction error along the 200 iterations. On the bottom from left to right: original simulated diffraction pattern, frequency-space magnitude of the reconstructed signal, frequency-space phase of the reconstructed signal, frequency-space reconstruction error along the 200 iterations.

signal diffraction pattern at each iteration according to

$$E_f = \sum_k ||Z(k)| - |F(k)||^2 \tag{4.19}$$

Since in an experiment we do not have access to the true signal in real-space, the real-space error E_R cannot be calculated in the same manner. Instead, we use the ratio of the signal outside of the support with respect to the signal inside the support, which we expect to approach zero as the algorithm converges, since the signal outside the support should approach that value:

$$E_r = \frac{\sum_{x \notin S} |z'(x)|^2}{\sum_{x \in S} |z'(x)|^2} \tag{4.20}$$

We see that the retrieved magnitudes in real and reciprocal-space (second column in the image) greatly resembles the original image and its simulated diffraction pattern. Some noise was still noticeable in the region outside the support of the reconstructed image. The error had almost stagnated by the 200^{th} iteration and, in fact, would not change much even if we left it running for as many as 1000 iterations.

We also simulated the Phase-Retrieval of a spherical sample using not the diffraction pattern calculated from the FFT of a simulated sphere, but instead the analytical

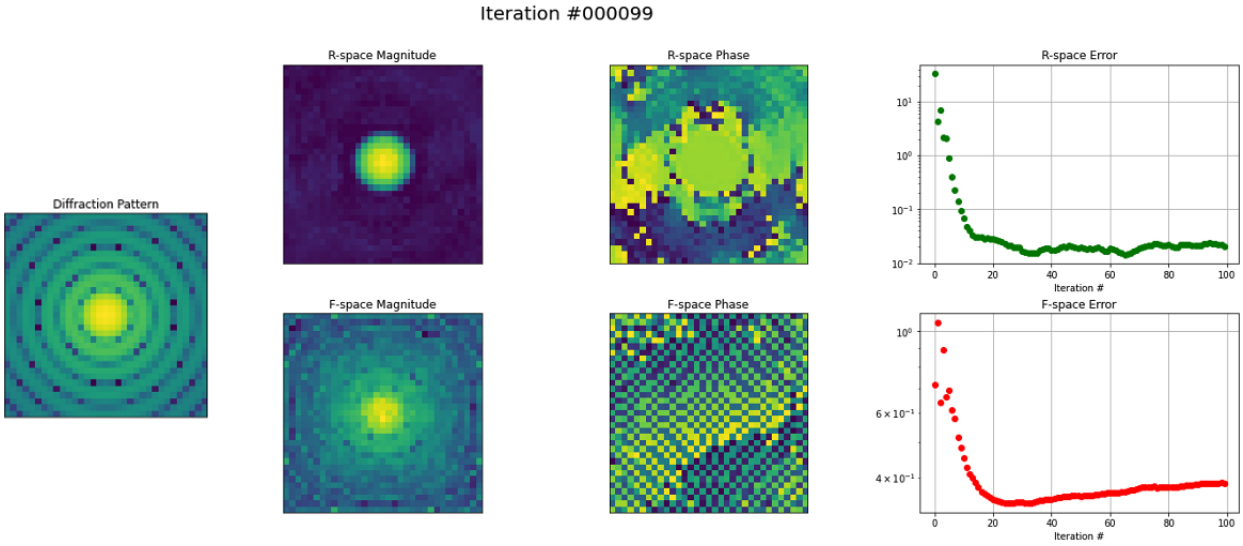


Figure 52 – Phase-Retrieval from the analytical diffraction pattern of the sphere using the HIO algorithm.

formula presented in section 17. A successful phase-retrieval in this case would be vital for us to validate the efficacy of the proposed M-ART reconstruction, since it would allow us to perform a complete reconstruction (phase-retrieval followed by ART, without committing an "inverse-crime").

The initial simulation results using the HIO algorithm are shown in Figure 52 for 100 HIO iterations. One can clearly see that the sphere projection seems to be recovered, but that the recovered F-space magnitude does not resemble the analytical diffraction pattern as before. In fact, we see that the F-space error seems to reach a minimum after about 30 iterations, and then starts to increase once again.

We tried implementing several strategies to improve the quality of the reconstruction in this case. Among them, making the support as tight as possible around the sphere, implementing positivity constraints with all three possibilities (only real, only imaginary and to both parts of the signal) as well as varying simulation parameters to increase the oversampling of the problem (equation 3.117). Although some variation was noticed in each trial, none of them were successful enough to provide a satisfactory recovery of the diffraction pattern. The relative simplicity and symmetry of the spherical sample might justify why such a bad diffraction pattern would still provide a real-space magnitude that resembles the expected spherical projection.

The only strategy that in the end worked to successfully recover the frequency-space magnitude was the combination of the HIO and ER algorithms, as shown in Figure 53. We first performed HIO for 150 iterations, followed by 50 iterations of ER. This strategy worked best than using only ER, only HIO or even ER followed by HIO. This can be intuitively understood as follows: the ER algorithm is quite "brute" since it forces

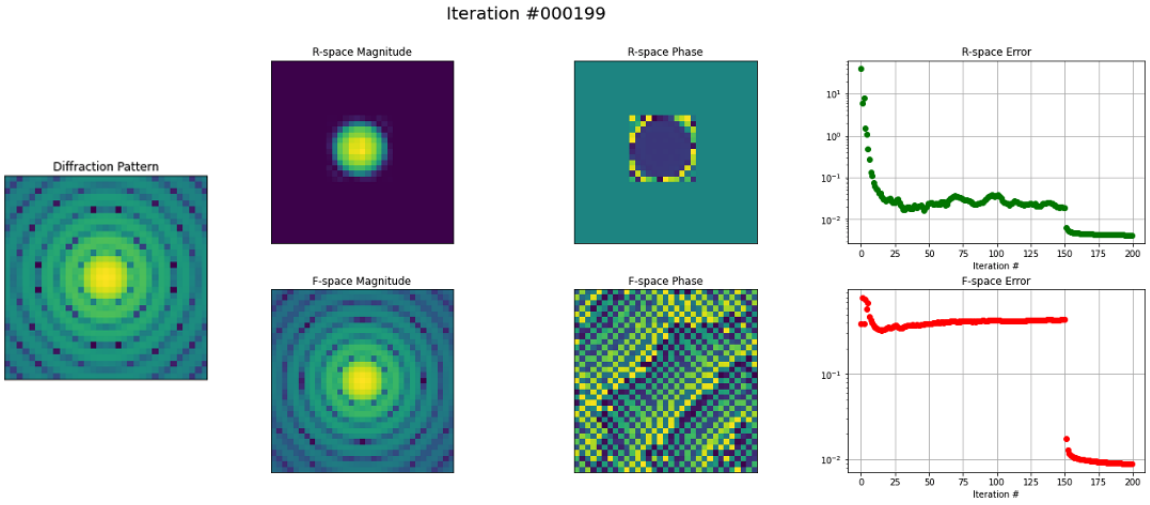


Figure 53 – Phase-Retrieval from the analytical diffraction pattern of the sphere using a combination of the HIO algorithm for 150 iterations followed by 50 iterations of the ER algorithm.

all values outside the support to zero. Using this algorithm immediately at the beginning usually causes it to get "stuck" in a local minimum solution. The HIO circumvents this problem by slowly taking the values outside the support towards zero. Nonetheless, in some cases it stagnates in a local minimum. We then use ER at the end to force it out of such local minimum to finally reach the desired solution. The error plots in Figure 53 show a sharp drop when the algorithm switches from HIO to ER.

4.4 Tomography Simulations

To validate the M-matrix reconstruction based on the ART algorithm (M-ART), we first needed to generate the simulated data. For that, we created a sinogram of a 3D model δ_M by multiplying it by a M-matrix $M_{\theta,p}$. The indices θ and p here indicate that we have a different M-matrix of each angle θ of the sample (angle with respect to the direction of the incident radiation) and each pixel p in the output plane of the sample. In the simulations, for a sample δ_M of size N^3 we have N^2 pixels in the output plane. Moreover, we considered $2N$ values of θ to sufficiently satisfy the Crowther criterion. This amounts to a total of $2N^3$ matrices. For each value of θ, the N^2 matrices for each output pixel are simply a translated version of each other. Because of the considerable size of the M-matrix above $N > 128$, as we showed in Table 2, and the huge number of matrices, it would be inpractical to store all of them in memory. Instead, we had to adopt the strategy of defining an "oversized" matrix with twice the dimensions of the sample in each direction (that is, $2N \times 2N \times 2N$) centered at the central pixel of the output plane (as illustrated in Figure 47), so that we could properly translate and rotate the matrix from -90^o to $+90^o$ and obtain a complete $M_{\theta,p}$ for all θ and p values. We refer to this as the "base M-matrix".

The linear relation between the the M-matrix and sample matrix δ_M given by the general formula 4.7 required us to adapt the iterative Kaczmarz equation to our case, such that

$$\delta_{k+1}(x,y,z) = \delta_k(x,y,z) - \frac{M_{\theta,p}(x,y,z)}{|M_{\theta,p}(x,y,z)|^2}(M_{\theta,p}(x,y,z) \circ \delta_k(x,y,z) - \psi_{\theta,p}) \tag{4.21}$$

where $\psi_{\theta,p}$ represents the value of pixel p of the sinogram for angle θ, and k runs from $0 \leqslant k \leqslant M \times 2N \times N^2 = 2MN^3$. In other words, the iteration process requires us to loop through three nested loops, namely:

- The innermost loop containing N^2 iterations for each pixel p in the output plane;
- The middle loop containing $2N$ iterations concerning each rotation angle θ;
- The outermost loop containing M iterations concerning the repetitions of the ART algorithm.

This corresponds to a problem the scales with $\mathcal{O}(N^3)$ complexity. Whatever calculations had to be performed, they would necessarily be repeated $2MN^3$ times. This huge number is a consequence of the linear relation between the system and sample matrices being and operation from $\mathbb{R}^3 \to \mathbb{R}^1$, that is, each M-matrix models a single pixel in the output plane. Within the Projection Approximation, the iterative method (PA-ART) becomes much faster, since a single matrix is capable of mapping the entire output plane at once, by simple "repeating" the projection ray (illustrated in Figure 47, left) for all (x, y) positions of the cube, totalling only $2MN$ operations ($\mathcal{O}(N)$ complexity).

After generating the whole sinogram for all $2N$ values of θ for the sample shown in Figure 54, we used equation 4.21 to reconstruct the sample. For each iteration, the base M-matrix had to first be translated to the pixel p of interest and then accordingly rotated by the angle θ before calculating the next ART iteration. Figure 55 shows the central slice of the object in each of the three directions on the bottom row, compared to the same slices after 5, 10 and 20 iterations of M-ART.

Having successfully implemented the M-ART algorithm, we moved forward to compare it against PA-ART. The PA reconstruction could in principle be performed quickly using the FBP algorithm. Nonetheless, in order to reliably evaluate the improvement of the M-matrix model over the PA, we chose to do both reconstructions using the PA and M system-matrices through the ART algorithm.

To compare PA-ART and M-ART, we used the spherical sample shown in Figure 56, with $N = 33$. The successful reconstruction of it using the ART method is

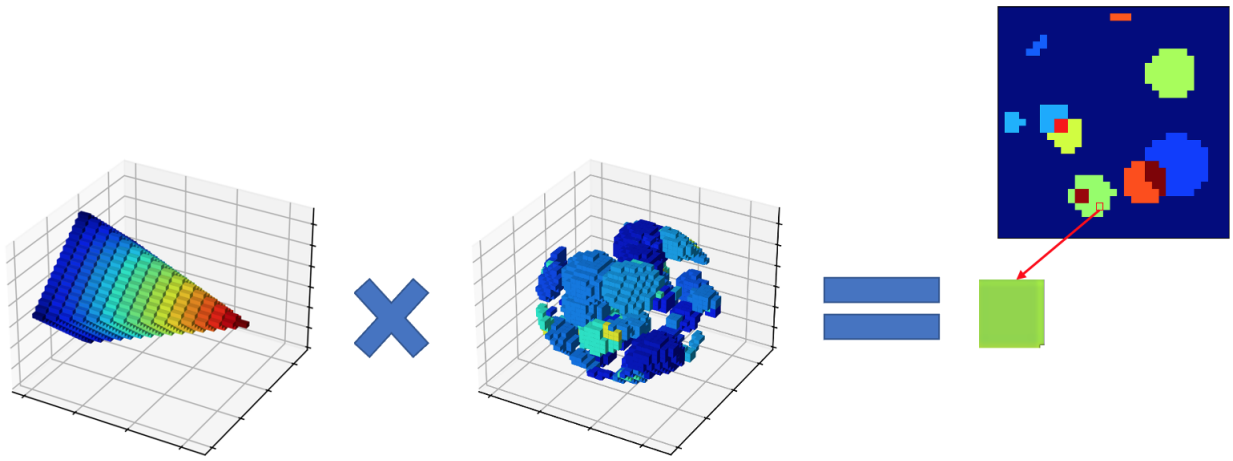


Figure 54 – Illustration of the sinogram generation process using the simulated sample and "base" M-matrix. The central plot in blue illustrates the three-dimensional sample used to simulate and validate the ART algorithm. One M-matrix generates a single pixel of the sinogram. Consecutive translations of the base matrix provide the matrix we use to generate all the pixels of a single sinogram frames. The base matrix is then rotated and the translation process is repeated to generate the next sinogram frame.

illustrated in Figure 57, where again we have the three central slices of both the model and of the reconstructed sample after 20 ART iterations. The reconstruction error along iterations for the above sample is plotted for both PA-ART and M-ART in Figure 58. This was a first indication of an improvement of the reconstruction quality of the M-ART over PA-ART. We see that the error of M-ART is not only lower, but also that it continues to drop more steeply compared to the PA-ART error. This error E_{ART} was calculated as the mean absolute error between model and reconstructed object

$$E_{\text{ART}} = \frac{\sum_{N^3} |\text{Object} - \text{Model}|}{N^3} \tag{4.22}$$

We finally proceeded to answer the main question posed by the project from the beginning: would a model that skips the Projection Approximation be able to reconstruct samples with thicknesses t above the theoretical limit imposed by equation 3.106 without deteriorating resolution?

$$t = \text{DOF} = 4\frac{\Delta r^2}{\lambda} \tag{3.106}$$

To answer that question, we performed multiple reconstructions of the spherical sample, each for 10 ART iterations, for both PA-ART and M-ART, and varying its diameter at each reconstruction. The final reconstruction error was calculated an plotted against sphere diameter in Figure 58. We also include in the plot the thickness limit from equation 3.106. One can see here that the reconstruction error for PA-ART indeed grows as we

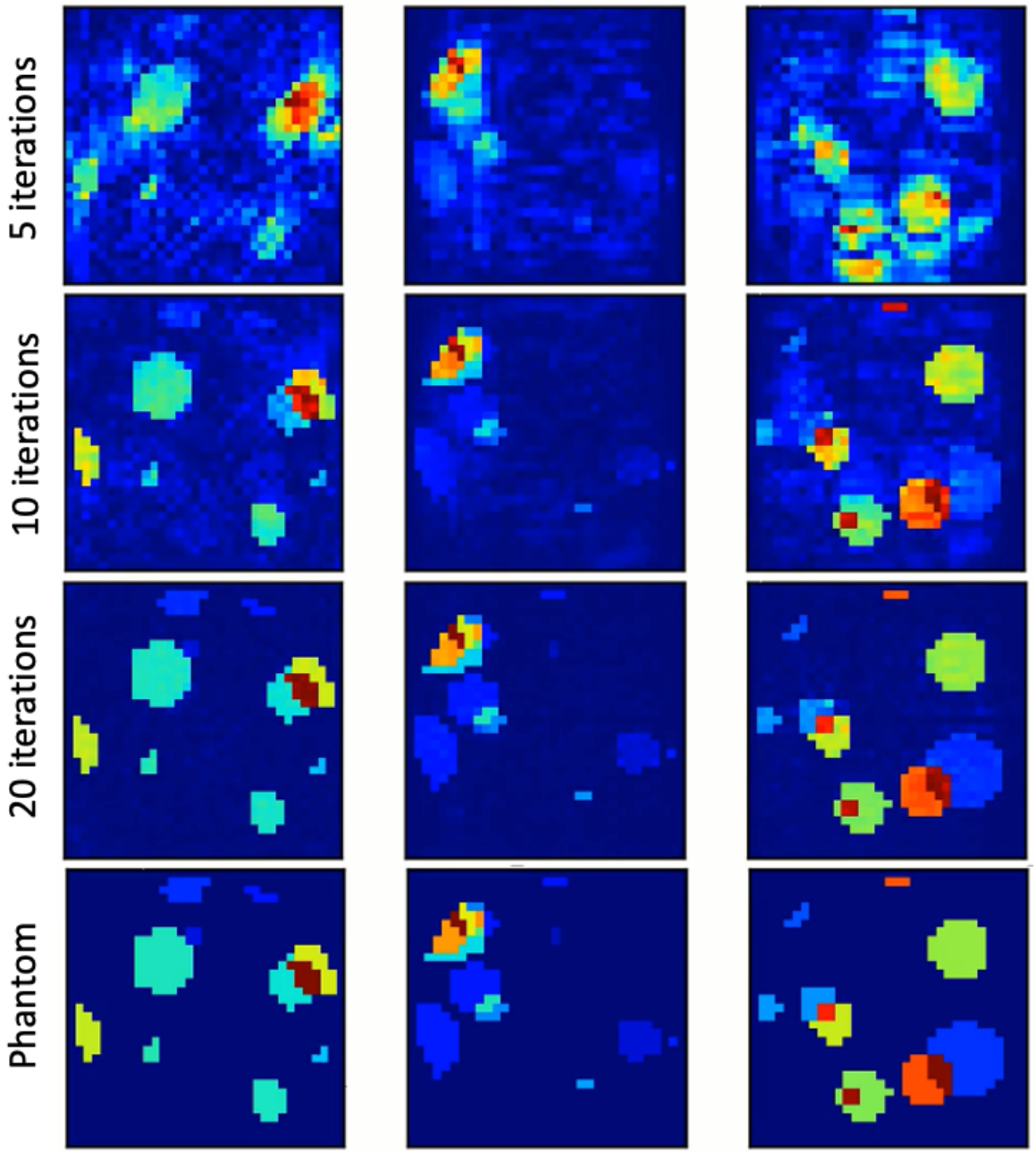


Figure 55 – Central slices in each of the three directions for the original sample (fourth row), reconstruction using the M-matrix after 5 iterations (first row), 10 iterations (second row) and 20 iterations (third row). All plots have their respective color map normalized by the maximum value of the image.

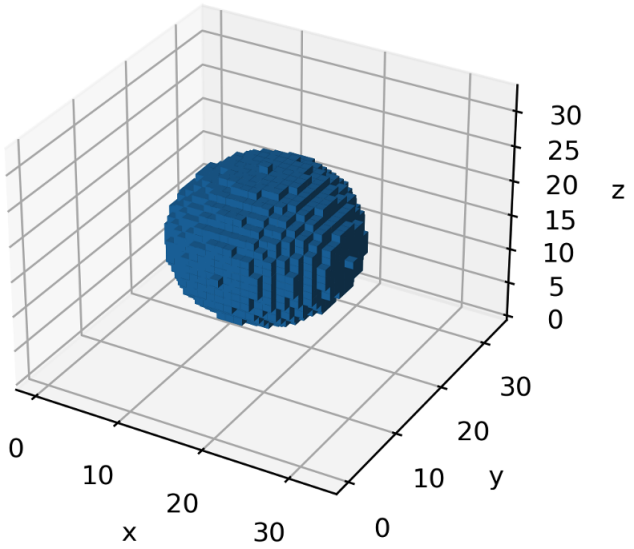


Figure 56 – Spherical sample used to further validate and test the M-matrix based ART reconstruction.

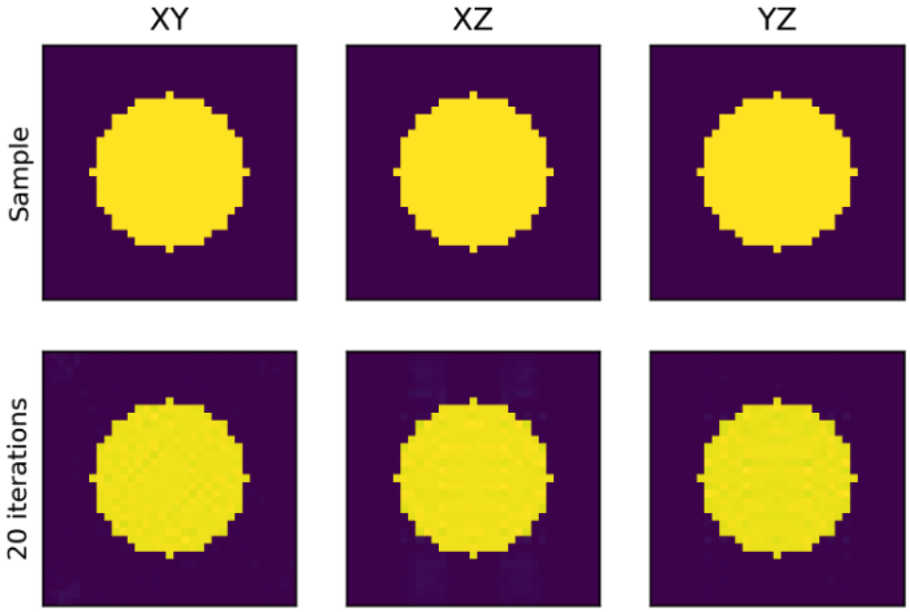


Figure 57 – Central slices in each of the three directions for the original sample (top) and reconstruction using the M-matrix for 20 iterations of the ART algorithm (bottom).

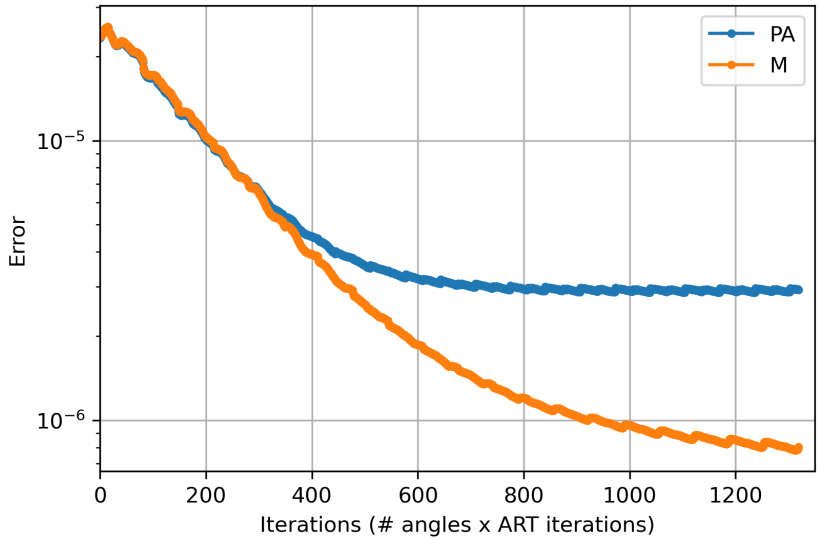


Figure 58 – Reconstruction error of the sphere along iterations for both PA-ART and M-ART. The red-dashed line indicates the thickness limit given by equation 3.106.

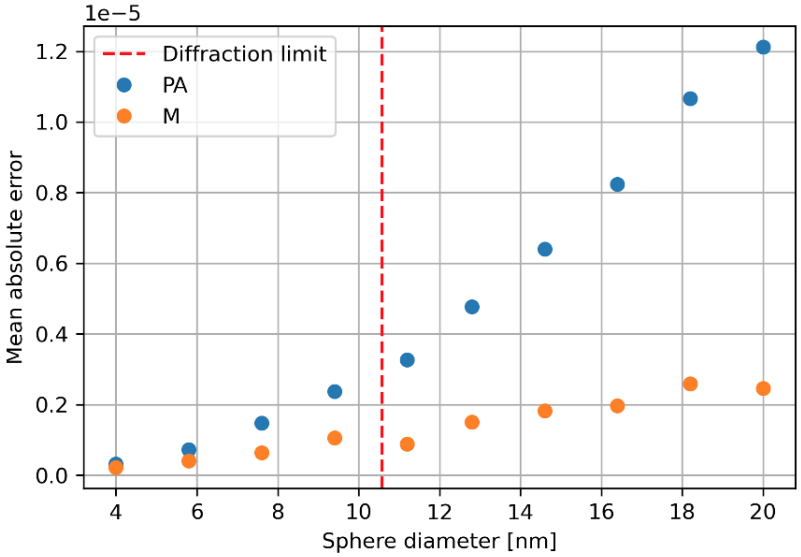


Figure 59 – Mean absolute error of the sphere reconstruction using both the PA and M matrices as a function of the diameter of the sphere.

increase the size of the sphere, especially above the thickness limit. On the other hand, the reconstruction error for M-ART barely increases for all sphere diameters.

This result was the strongest indication of the superiority of M-ART over PA-ART that we obtained so far. Nonetheless, rigorously speaking it is still not a conclusive result because of the already mentioned "inverse-crime" problem. From the above results, we could not yet affirm that the superiority of M-ART over PA-ART came from the quality of the model itself, or from the fact that the sinogram was first synthesized using the M-matrix instead of the PA-matrix.

Table 3 – Time estimate of ten complete M-ART reconstructions as a function of matrix size N, each with ten iterations of the outermost loop.

N	**17**	**33**	**65**	**129**	**257**	**513**
Time [hours]	0.17	0.58	17.24	1131	164706	11882468
Time [days]	0.01	0.02	0.72	47	6862	495102

4.4.1 Full simulation

To overcome the inverse-crime bias that might be present in the above results, we moved forward to perform a "full simulation": first doing the Phase-Retrieval from the analytical sphere form-factor (which we already presented the result in section 4.3, Figure 53) followed by the tomographic reconstruction.

Nonetheless, performing the full simulation turned out to be quite problematic, primarily because of the biggest issue we had been facing: the enormous computational cost of the M-ART method. To further elucidate on this problem, we first note that all the tomography results presented so far were performed for very small N values. Because we needed to use an "oversized" M-matrix, as mentioned in the beginning of this section, we usually had an M-matrix of $N = 65^3$ and a sample matrix of $N = 33^3$ points.

We estimated in Table 3 the time it would take to repeat the analysis presented in Figure 59, for oversized M-matrices with different N-values. The estimate was calculated by performing a single operation of the innermost loop of the algorithm and then multiplying the time this operation took by $10 \times 2 \times 10 \times N^3$ (since $M = 10$ and we wanted to perform a total of 10 reconstructions, each for a different sphere diameter). One can see that the $\mathcal{O}(N^3)$ complexity indeed makes the problem unfeasible to solve for $N > 65$, since a single analysis would take 47 days to complete.[3] Therefore, we had to limit our analysis in a first moment to such small matrices with $N \leqslant 65$. This shows how the computational cost of the method is indeed challenging for the practical implementation at a beamline such as Cateretê, where $N \leqslant 3072$.

Reconstruction from the phase-retrieved projections

The initial M-ART reconstructions were performed by obtaining the projection from the HIO+ER approach shown in Figure 53. Because in principle one expects complete symmetry of the sphere over rotations, the retrieved amplitude from the analytical diffraction pattern should, in principle, be the same. Therefore, we performed a single phase-retrieval and repeated the obtained projection for all sinogram angles. Subsequently, this sinogram was used as the input to the M-ART algorithm. The results showed no difference between M-ART and PA-ART. We hypothesised that this could be a consequence of

[3] All simulations results presented were performed in a personal computer CPU (Intel Core i5) with 8 GBs of RAM.

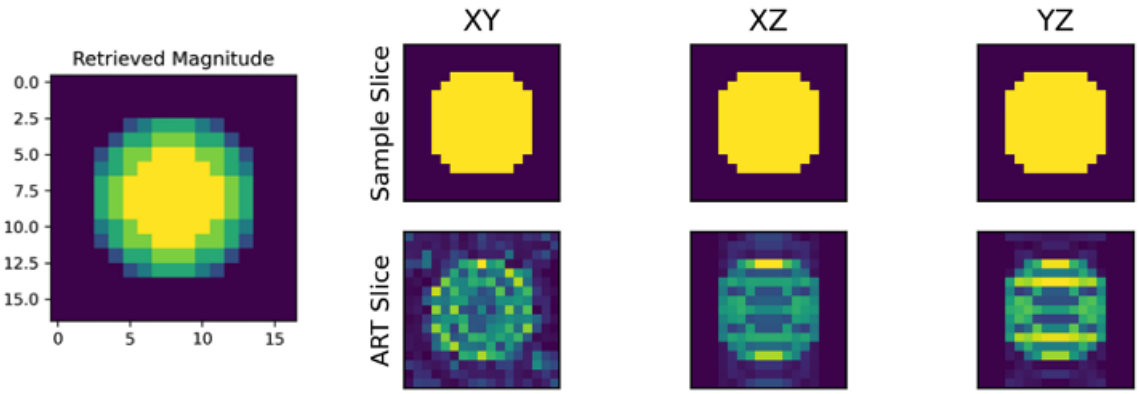


Figure 60 – Reconstruction using M-ART from a sinogram containing the "perfect retrieved magnitude" shown in the left. The lack of spherical symmetry due to the low number of pixels made it impossible to properly reconstruct the sphere, since a rotation of the sphere would not provide the same projection of it for each angle.

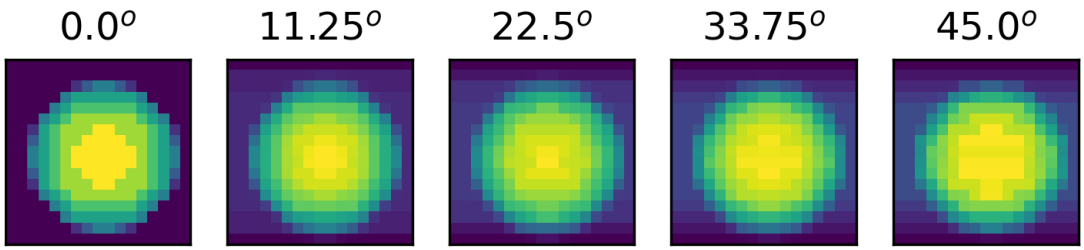


Figure 61 – Difference in the projection of a sphere of $N = 17$ when rotated by different angle values. This lack of symmetry impedes the faithful reconstruction using the M-ART method.

the big errors from the phase-retrieval step. Therefore, we further simplified the analysis by supposing that the phase-retrieval step was "perfect", that is, that the retrieved amplitude was the perfect projection shown in Figure 60 (left). This amplitude was calculated simply by integrating the spherical sample over one axis, and then repeating it for all angles of the sinogram. In this manner, if the reconstruction was not successful, it could only be because of the tomographic part. Indeed, that was the case. As we shown in Figure 60, the reconstruction of the central slices is far from perfect. It differs a lot from the model and from the reconstructed slices we had obtained before (Figure 57).

After thorough investigation, we concluded that this was a consequence of the lack of symmetry of the sphere due to the low number of points of the object. That is, ideally, when we calculated the projection of a rotated sphere, the resulting projection should remain the same. Nonetheless, we can see from Figure 61 that it is not the case when we are dealing with such low values of N: the pixelated "sphere" lacks spherical symmetry. Consequently, we concluded that we could not escape the fact that the computational cost of the M-ART method had to be reduced if we wanted to reliably confirm its superiority.

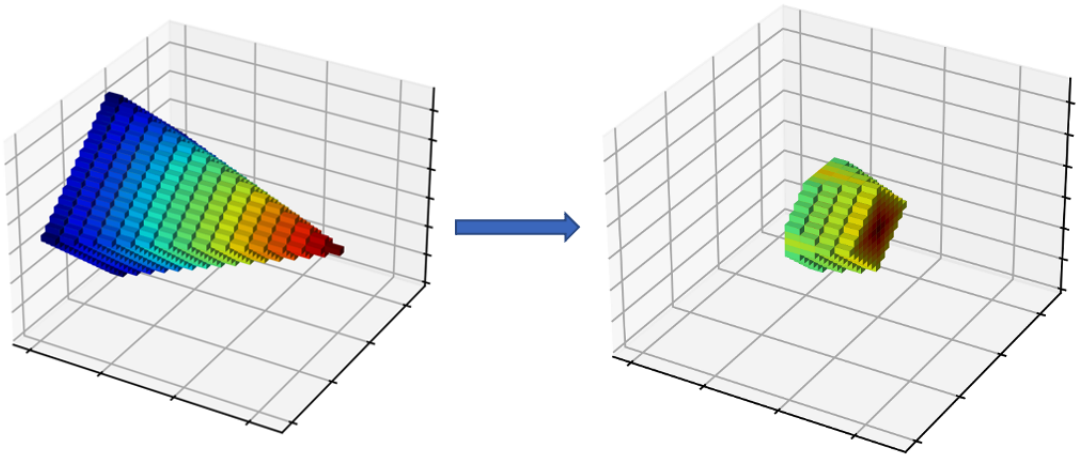


Figure 62 – The M-matrix maps each voxel of the sample independently and may allow us to perform a local reconstruction and resolution of smaller regions of the sample that have been pre-reconstructed using more traditional methods.

4.5 Perspectives

In this final section we lay out some future perspectives for this project in face of the challenges we encountered. The natural target for the M-ART method would be, first, to further validate it with more robust simulations. After doing so, we would eventually use it to reconstruct images using experimental data and verify the cost-benefit of the method in comparison to other techniques. A first robust simulation would be the full CDI simulation mentioned in section 4.4.1. To achieve that nonetheless, we must improve the computational speed of the method, as the results have evidenced. We carefully timed the different sections of the code and confirmed that the main issue lies not on the computational cost of the operations themselves, but at the sheer number of operations that need to be performed. Among the operations, the most expensive one is the rotation of the M-matrix.

A few strategies have already been tried to improve speed. First, we tried to perform the rotations in GPU using the CuPy library [34]. This strategy turns to be advantageous only when the mathematical operations involves big matrices. Otherwise, the time (also known as overhead) taken for the matrices to be transferred from CPU to GPU and, after operations, back from GPU to CPU, may be even longer the the computing time itself. For the matrix sizes we tested (up to $N = 512$), CuPy did not seem to be advantageous. This is likely because the vectorized form of the rotation function we implemented was already considerably optimized. We also tried parallel computing strategies to perform the rotation. Since the rotation is around the z-axis, each slice of the 3D matrix could be rotated in parallel. However, some overhead is also present to perform such operations, and it turned out to be roughly equivalent to the vectorized form we had implemented for matrices with sizes up to $N = 512$.

We then tried to encapsulate the entire nested loops of the algorithm within the GPU calculations (i.e. ART iteration, translations and rotations), so that the writing/reading operations between CPU and GPU had to occur only before starting the loop and after finishing it. By doing this, we seem to have gained a factor of approximately 2 times in speed for small matrices ($N \leqslant 32$). Notwithstanding the fact that this is expected to increase when tested for bigger values of N, it is far from the orders of magnitude improvement we would need to actually simulate it with bigger matrices and, hopefully, with experimental data.

Another proposition is to perform the update of the ART operations in batches, so we can make use of parallel computing. Since the rotation operation is much slower than the ART update itself, we believe we should be able to perform some rotations of different matrices in parallel, and then apply the Kaczmarz method for the batch of rotated matrices consecutively. However, how fast this could get would also depend on resources, because of the huge RAM required to story many matrices, which would require powerful machines with dozens gigabytes of RAM available.

The most promising idea would be implement the code, or at least the entire M-ART nested loops in C language. C is a compiled language which may become even thousands of times faster than Python, an interpreted language. However, the downside comes from the more complicated low-level syntax of the C language. We estimate it would require at least a few months to complete such code translation, since we currently lack the expertise in such a language. This strategy, by itself, may be sufficient to gain sufficient speed up to allow the full simulation to be performed in feasible time using matrices of $N \geqslant 128$.

Last but not least, we illustrate in Figure 62 the idea of M-ART as a refinement method. Since the M-matrix maps a specific region of the sample, one may be able to use a smaller portion of it to reconstruct a local region of the sample. Consequently, even if the big computational cost cannot be circumvented up to the required level, smaller matrices can be used to refine the resolution of reconstructions in a local region of interest – the price to pay being a small deterioration of reconstruction quality, since the model would not be strictly followed. In other words, one would first obtain images using traditional methods based on the Projection Approximation, select a particular region of scientific interest, and then later refine the resolution in such smaller location using M-ART.

5 Conclusion

In this Masters project, we proposed a discrete model of the inhomogenous Helmholtz equation that allows us to perform three-dimensional image reconstruction with Plane-Wave Coherent Diffractive Imaging without using the Projection Approximation (PA). Traditional tomographic reconstruction relies on this approximation and therefore presents an intrinsic limit to the resolution that can be achieved in the final image for a sample of certain thickness. We intended on evaluating if this limitation could be circumvented if the Projection Approximation was avoided.

We demonstrate that this discrete model can be written in a matrix form, which linearly relates the system-matrix and object-matrix. Consequently, Algebraic Reconstruction Techniques may be applied to perform the tomographic reconstruction. We name the system-matrix as the "M-Matrix" and the ART iterative algorithm adapted to it as the "M-ART" algorithm. Furthermore, we demonstrate a simple recursive relation to generate a system M-Matrix of arbitrary size.

The downside of the method turns out to be its computational cost. The iterative character of the ART algorithm, together with the fact that the we need a full system-matrix to model a single point of the wavefield at the output plane of the sample, results in a problem of $\mathcal{O}(N^3)$ complexity. Therefore, we had to perform simulations with very small samples of size $N^3 \leqslant 33^3$. Even for such small matrices, the obtained results indicate that the M-ART algorithm is indeed superior to the PA based reconstruction method, in the sense that it improves the reconstruction error. Nonetheless, the low number N of pixels prevented us from performing a more reliable analysis in the absence of an "inverse-crime" when solving the inverse problem. Notwithstanding the challenges we faced, we believe the computational-cost can still be reduced with the application of high-performance computing strategies. This may at least allow the method to work as a local refinement method for the current Coherent Diffracting Imaging reconstruction techniques.

Bibliography

1. Jacobsen, C. *X-ray Microscopy* (Cambridge University Press, 2019).
2. Jiang, Y. *et al.* Electron ptychography of 2D materials to deep sub-ångström resolution. *Nature* **559,** 343–349 (July 2018).
3. Hubbell, J. H. *Radiation Physics* 2003. http://dx.doi.org/10.1016/B0-12-227410-5/00634-7.
4. Seltzer, S. *Tables of X-Ray Mass Attenuation Coefficients and Mass Energy-Absorption Coefficients, NIST Standard Reference Database 126* 1995. http://www.nist.gov/pml/data/xraycoef/index.cfm.
5. Willmott, P. *An Introduction to Synchrotron Radiation* Mar. 2019. http://dx.doi.org/10.1002/9781119280453.
6. *Caterete beamline, Brazilian Synchrotron Light Laboratory* Accesed on May 10th, 2022. https://wiki-sirius.lnls.br/mediawiki/index.php/Main_Page. 2022.
7. *Sirius Project Brazilian Synchrotron Light Laboratory* Accesed on May 10th, 2022. https://www.lnls.cnpem.br/facilities/caterete/. 2022.
8. Pfeiffer, F. X-ray ptychography. *Nature Photonics* **12,** 9–17. https://doi.org/10.1038/s41566-017-0072-5 (2018).
9. Chapman, H. N. *et al.* High-resolution ab initio three-dimensional x-ray diffraction microscopy. *J. Opt. Soc. Am. A* **23,** 1179–1200. http://opg.optica.org/josaa/abstract.cfm?URI=josaa-23-5-1179 (May 2006).
10. Paganin, D. *Coherent X-Ray Optics* Jan. 2006. http://dx.doi.org/10.1093/acprof:oso/9780198567288.001.0001.
11. Feeman, T. G. *The Mathematics of Medical Imaging* http://dx.doi.org/10.1007/978-0-387-92712-1 (Springer New York, 2010).
12. Rinkel, J., Polli, J. M. & Miqueles, E. X. X-ray coherent diffraction imaging: Sequential inverse problems simulation. *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **912,** 43–47. ISSN: 01689002. https://doi.org/10.1016/j.nima.2017.10.032 (2018).
13. James, R. W. *The Optical Principles of the Diffraction of X-Rays* (G. Bell and Sons Ltd, 1962).
14. Saleh, B. E. A. & Teich, M. C. *Fundamentals of Photonics* Aug. 1991. http://dx.doi.org/10.1002/0471213748.

15. Liboff, R. L. *Introductory Quantum Mechanics* 3rd ed. ISBN: ISBN: 0201878798 (Addison-Wesley, 1998).

16. Attwood, D. *Soft X-Rays and Extreme Ultraviolet Radiation: Principles and Applications* (Cambridge University Press, 1999).

17. Als-Nielsen, J. & McMorrow, D. *Elements of Modern X-ray Physics* http://dx.doi.org/10.1002/9781119998365 (Wiley, Mar. 2011).

18. Griffiths, D. J. *Introduction to Electrodynamics* 4th ed. (Cambridge University Press, 2017).

19. Zwiebach, B. *Lecture notes from MIT 8.06 Quantum Physics III, Spring 2018. Chapter 7 - Scattering. Massachusetts Institute of Technology: MIT OpenCouseWare* Accesed on May 4th, 2022. https://ocw.mit.edu/courses/8-06-quantum-physics-iii-spring-2018/pages/lecture-notes/. 2018.

20. Anitas, E. M. *Small-Angle Scattering (Neutrons, X-Rays, Light) from Complex Systems* http://dx.doi.org/10.1007/978-3-030-26612-7 (Springer International Publishing, 2019).

21. Jackson, J. D. *Classical Electrodynamics* 3rd ed. (New York, NY, 1999).

22. Ida, N. & Meyendorf, N. *Handbook of Advanced Nondestructive Evaluation* http://dx.doi.org/10.1007/978-3-319-26553-7 (Springer International Publishing, 2019).

23. Miao, J., Sayre, D. & Chapman, H. N. Phase retrieval from the magnitude of the Fourier transforms of nonperiodic objects. *J. Opt. Soc. Am. A* **15,** 1662–1669. http://josaa.osa.org/abstract.cfm?URI=josaa-15-6-1662 (June 1998).

24. Henke, B., Gullikson, E. & Davis, J. X-Ray Interactions: Photoabsorption, Scattering, Transmission, and Reflection at E = 50-30,000 eV, Z = 1-92. *Atomic Data and Nuclear Data Tables* **54,** 181–342 (1993).

25. Born, M. & Wolf, E. *Principle of Optics, 60th Anniversary Edition* 7th edition (Cambridge University Press, 2020).

26. Van der Veen, F. & Pfeiffer, F. Coherent x-ray scattering. *Journal of Physics: Condensed Matter* **16,** 5003–5030. https://doi.org/10.1088/0953-8984/16/28/020 (July 2004).

27. Sayre, D. Some implications of a theorem due to Shannon. *Acta Crystallographica* **5,** 843. https://doi.org/10.1107/S0365110X52002276 (Nov. 1952).

28. Gerchberg, R. W. & Saxton, W. O. A practical algorithm for the determination of phase from image and diffraction plane pictures. *Optik* **35,** 237. https://ci.nii.ac.jp/naid/10010556614/en/ (1972).

29. Fienup, J. R. Reconstruction of an object from the modulus of its Fourier transform. *Opt. Lett.* **3,** 27–29 (July 1978).

30. Miao, J. & Sayre, D. On possible extensions of X-ray crystallography through diffraction-pattern oversampling. *Acta Crystallographica Section A* **56,** 596–605. https://doi.org/10.1107/S010876730001031X (Nov. 2000).

31. Miao, J., Ishikawa, T., Anderson, E. H. & Hodgson, K. O. Phase retrieval of diffraction patterns from noncrystalline samples using the oversampling method. *Physical Review B - Condensed Matter and Materials Physics* **67,** 1–6. ISSN: 1550235X (2003).

32. Marchesini, S. Invited Article: A unified evaluation of iterative projection algorithms for phase retrieval. *Review of Scientific Instruments* **78,** 011301. eprint: https://doi.org/10.1063/1.2403783. https://doi.org/10.1063/1.2403783 (2007).

33. Shepp, L. A. & Logan, B. F. *The Fourier reconstruction of a head section* June 1974. http://dx.doi.org/10.1109/TNS.1974.6499235.

34. *Cupy* Accesed on May 16th, 2022. https://cupy.dev/. 2022.