LUND UNIVERSITY

Advancing X-ray imaging with deep learning

Physics-inspired reconstruction approaches

Zhang, Yuhe

2024

Document Version:
Publisher's PDF, also known as Version of record

Link to publication

Citation for published version (APA):
Zhang, Y. (2024). *Advancing X-ray imaging with deep learning: Physics-inspired reconstruction approaches*. Department of Process and Life Science Engineering, Lund University.

Total number of authors:
1

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Advancing X-ray imaging with deep learning

Physics-inspired reconstruction approaches

YUHE ZHANG

DEPARTMENT OF PHYSICS | FACULTY OF SCIENCE | LUND UNIVERSITY

Advancing X-ray imaging with deep learning

Advancing X-ray imaging with deep learning

Physics-inspired reconstruction approaches

Yuhe Zhang

DOCTORAL DISSERTATION
by due permission of the Faculty of Science, Lund University, Sweden.
To be defended in the Rydberg Lecture Hall at the Department of Physics.
Date: 2024-06-14 and time: 9:15

Faculty opponent
Prof. Chris Jacobsen, Northwestern University

Organization
LUND UNIVERSITY

Department of Physics
Box 118
SE–221 00 LUND
Sweden

Document name
DOCTORAL DISSERTATION

Date of disputation
2024-06-14

Sponsoring organization

Author(s)
Yuhe Zhang

Title and subtitle
Advancing X-ray imaging with deep learning : Physics-inspired reconstruction approaches

Abstract

The development of high-brilliance X-ray sources, such as the fourth-generation diffraction-limited storage rings and X-ray free-electron lasers, have opened up new possibilities for X-ray imaging and pushed the temporal resolutions of imaging techniques to unprecedented levels. Capturing fast dynamics in two-dimensional (2D), three-dimensional (3D), and even four-dimensional (4D, 3D + time) beyond microsecond temporal resolution has become possible. To fully exploit the unique capabilities of these facilities, challenges such as the data problem must be addressed. Automated tools are needed to handle the large amount of data acquired from each experiment. As a data-driven approach, deep learning has undergone rapid development over the past decade and offers a promising solution to this problem. However, state-of-the-art deep learning methods applied to X-ray imaging ignore the physics of X-ray propagation and interaction with matter and require paired training datasets. In this thesis, we show that combining the physical principles of X-ray imaging with deep learning greatly improves the performance and robustness of the approaches, and it is possible to construct reliable unsupervised approaches, where no paired datasets are needed.

Firstly, we present a theoretical background on X-ray imaging and various imaging methods. Secondly, we provide an overview of deep learning, including training strategies and common frameworks for addressing imaging tasks. Lastly, we introduce novel algorithms developed during this thesis:

1. FFCGAN, a supervised approach for shot-to-shot flat-field correction at X-ray free-electron lasers.
2. PhaseGAN, a phase-retrieval approach for unpaired datasets.
3. ONIX, a self-supervised approach for 3D reconstruction from sparse views.
4. 4D-ONIX, a self-supervised approach for reconstructing 3D movies from sparse projections.

These approaches offer high-quality image reconstructions for X-ray imaging techniques, enabling further exploration and understanding of the structure and dynamic properties of various samples.

Key words
X-ray imaging, deep learning, image reconstruction, phase-contrast, tomography

Classification system and/or index terms (if any)

Supplementary bibliographical information

Language
English

ISSN and key title

ISBN
978-91-8039-982-1 (print)
978-91-8039-983-8 (pdf)

Recipient's notes

Number of pages
216

Price

Security classification

I, the undersigned, being the copyright owner of the abstract of the above-mentioned dissertation, hereby grant to all reference sources the permission to publish and disseminate the abstract of the above-mentioned dissertation.

Signature

Date 2024-5-6

DOKUMENTDATABLAD enl SIS 61 41 21

Advancing X-ray imaging with deep learning

Physics-inspired reconstruction approaches

by Yuhe Zhang

LUND
UNIVERSITY

A doctoral thesis at a university in Sweden takes either the form of a single, cohesive research study (monograph) or a summary of research papers (compilation thesis), which the doctoral student has written alone or together with one or several other author(s).

In the latter case the thesis consists of two parts. An introductory text puts the research work into context and summarizes the main points of the papers. Then, the research publications themselves are reproduced, together with a description of the individual contributions of the authors. The research papers may either have been already published or are manuscripts at various stages (in press, submitted, or in draft).

Cover image by Yuhe Zhang.

Faculty of Science, Department of Physics

ISBN: 978-91-8039-982-1 (print)
ISBN: 978-91-8039-983-8 (pdf)

Printed in Sweden by Media-Tryck, Lund University, Lund 2024

致我的家人，朋友，师长，
还有所有一路上帮助过我的人

To my family, friends, teachers,
and all those who have helped me along the journey

Contents

List of publications

This thesis is based on the following publications, referred to by their Roman numerals:

I **Shot-to-shot flat-field correction at X-ray free-electron lasers**

K. Buakor[†], **Y. Zhang**[†], Š. Birnšteinová, V. Bellucci, T. Sato, H. Kirkwood, A. P. Mancuso, P. Vagovic, P. Villanueva-Perez
[†] These authors contributed equally
Optics Express 30, 10633-10644 (2022)

II **PhaseGAN: a deep-learning phase-retrieval approach for unpaired datasets**

Y. Zhang, M. Andreas Noack, P. Vagovic, K. Fezzaa, F. Garcia-Moreno, T. Ritschel, P. Villanueva-Perez
Optics Express 29, 19593-19604 (2021)

III **Reusability Report: Unpaired deep-learning approaches for holographic image reconstruction**
Y. Zhang, T. Ritschel, P. Villanueva-Perez
Nature Machine Intelligence, 1-7 (2024)

IV **ONIX: an X-ray deep-learning tool for 3D reconstructions from sparse views**

Y. Zhang, Z. Yao, T. Ritschel, P. Villanueva-Perez
Applied Research 2, 4 (2023)

V **4D-ONIX: A deep learning approach for reconstructing 3D movies from X-ray multi-projection imaging**
Y. Zhang, Z. Yao, R. Klöfkorn, T. Ritschel, P. Villanueva-Perez
arXiv preprint arXiv:2401.09508 (2024)

VI **Megahertz X-ray Multi-projection imaging**

P. Villanueva-Perez, V. Bellucci, **Y. Zhang**, Š. Birnšteinová, R. Graceffa, L. Adriano, E. M. Asimakopoulou, I. Petrov, Z. Yao, M. Romagnoni, A. Mazzolari, R. Letrun, C. Kim, J. Koliyadu, C. Deiter, R. Bean, G. Giovanetti, L. Gelisio, T. Ritschel, A. Mancuso, H. N. Chapman, A. Meents, T. Sato, P. Vagovic
arXiv preprint arXiv:2305.11920 (2023)

All papers are reproduced with permission of their respective publishers.

Publications not included in this thesis, to which I contributed:

vii **PyPhase–a Python package for X-ray phase imaging**

M. Langer, **Y. Zhang**, D. Figueirinhas, J.-B. Forien, K. Mom, C. Mouton, R. Mokso, P. Villanueva-Perez
Journal of Synchrotron Radiation 28, 1261 (2021)

viii **X-ray in-line holography and holotomography at the NanoMAX beamline**

S. Kalbfleisch, **Y. Zhang**, M. Kahnt, K. Buakor, M. Langer, T. Dreier, H. Dierks, P. Stjärneblad, E. Larsson, K. Gordeyeva, L. Chayanun, D. Söderberg, J. Wallentin, M. Bech, P. Villanueva-Perez
Journal of Synchrotron Radiation, 29, 224 (2022)

ix **Dose-efficient multimodal microscopy of human tissue at a hard X-ray nanoprobe beamline**
S. Sala, **Y. Zhang**, N. De La Rosa, T. Dreier, M. Kahnt, M. Langer, L. B. Dahlin, M. Bech, P. Villanueva-Perez, S. Kalbfleisch
Journal of Synchrotron Radiation, 29, 807 (2022)

x **Deep-learning image enhancement and fibre segmentation from time-resolved computed tomography of fibre-reinforced composites**
R. Guo[†], J. Stubbe[†], **Y. Zhang**, C. M. Schlepütz, C. R. Gomez, M. Mehdikhani, C. Breite, Y. Swolfs, P. Villanueva-Perez
[†] These authors contributed equally
Composites Science and Technology, 244, 110278 (2023)

xi **Online dynamic flat-field correction for MHz Microscopy data at European XFEL**
S. Birnsteinova, D. E. Ferreira de Lima, E. Sobolev, H. J. Kirkwood, V. Bellucci, R. J. Bean, C. Kim, J. Koliyadu, T. Sato, F. Dall' Antonia, E. M. Asimakopoulou, Z. Yao, K. Buakor, **Y. Zhang**, A. Meents, H. N. Chapman, A. P. Mancuso, P. Villanueva-Perez, P. Vagovic
Journal of Synchrotron Radiation 30.6 (2023)

xii **Ultrasound cavitation and exfoliation dynamics of 2D materials re-vealed in operando by X-ray free electron laser megahertz imaging**
K. Xiang, S. Huang, H. Song, V. Bazhenov, V. Bellucci, S. Birnsteinova, R. de Wijn, J. Koliyadu, F. H. M. Koua, A. Round, E. Round, A. Sarma, T. Sato, M. Sikorski, **Y. Zhang**, E. M. Asimakopoulou, P.Villanueva-Perez, K. Porfyrakis, I. Tzanakis, D. G. Eskin, N. Grobert, A. Mancuso, R. Bean, P. Vagovic, J. Mi
arXiv preprint arXiv:2305.08538 (2023)

xiii **Development towards high-resolution kHz-speed rotation-free volumetric imaging**
E. M. Asimakopoulou, V. Bellucci, S. Birnsteinova, Z. Yao, **Y. Zhang**, I. Petrov, C. Deiter, A. Mazzolari, M. Romagnoni, D. Korytar, Z. Zaprazny, Z. Kuglerova, L. Juha, B. Lukic, A. Rack, L. Samoylova, F. G. Moreno, S. A. Hall, T. Neu, X. Liang, P. Vagovic,and P. Villanueva-Perez
Optics Express 32(3), 4413-4426 (2024)

xiv **ForMAX–a beamline for multiscale and multimodal structural characterization of hierarchical materials**
K. Nygård, S. A. McDonald, J. B. González, V. Haghighat, C. Appel, E. Larsson, R. Ghanbari, M. Viljanen, J. Silva, S. Malki, Y. Li, V. Silva, C. Weninger, F. Engelmann, T. Jeppsson, G. Felcsuti, T. Rosén, K. Gordeyeva, L. D. Söderberg, H. Dierks, **Y. Zhang**, Z. Yao, R. Yang, E. M. Asimakopoulou, J. K. Rogalinski, J. Wallentin, P. Villanueva-Perez, R. Krüger, T. Dreier, M. Bech, M. Liebi, M. Bek, R. Kádár, A. E. Terry, H. Tarawneh, P. Ilinski, J. Malmqvist, Y. Cerenius
arXiv preprint arXiv:2312.07904 (2023)

xv **Revealing the origins of vortex cavitation in a Venturi tube by high speed X-ray imaging**
H. Soyama, X. Liang, W. Yashiro, K. Kajiwara, E. M. Asimakopoulou, V. Bellucci, Š. Birnšteinová, G. Giovanetti, C. Kim, H. Kirkwood, J. Koliyadu, R. Letrun, **Y. Zhang**, J. Uličný, R. Bean, A. P. Mancuso, P. Villanueva-Pérez, T. Sato, P. Vagovič, D. E. Eakins, A. M. Korsunsky
Ultrasonics Sonochemistry 101, 106715 (2023)

xvi **Development of crystal optics for Multi-Projection X-ray Imaging for synchrotron and XFEL sources**
V. Bellucci, Š. Birnšteinová, T. Sato, R. Letrun, J. Koliyadu, C. Kim, G. Giovanetti, C. Deiter, L. Samoylova, I. Petrov, L. L. Morillo, R. Graceffa, L. Adriano, H. Huelsen, H. Kollmann, T. N. T. Calliste, D. Korytar, Z. Zaprazny, A. Mazzolari, M. Romagnoni, E. M. Asimakopoulou, Z. Yao, **Y. Zhang**, J. Ulicny, A. Meents, H. N. Chapman, R. Bean, A. Mancuso, P. Villanueva-Perez, P. Vagovic
arXiv preprint arXiv:2402.13262 (2024)

Popular summary

If you have ever had an X-ray or CT scan in a hospital, you are most likely familiar with X-ray imaging. An X-ray source, a sample, and a detector are all that are needed for the simplest X-ray imaging experiments. By measuring the different levels of absorption of X-rays as they pass through different materials, an image can be created that reveals the internal structures and details of the sample.

Conventionally, X-ray experiments were mainly performed in laboratories using small-scale X-ray sources, similar to the ones used in hospitals. Nowadays, X-ray imaging experiments are carried out not only in laboratories but also at large-scale facilities. These facilities, including synchrotron radiation sources and X-ray free-electron lasers, can produce intense X-ray beams that allow scientists to study the structure and properties of materials with unprecedented detail. Researchers from various fields, from biology and chemistry to materials science and physics, use these facilities to conduct complicated experiments that are impossible with conventional lab-based X-ray sources. These large-scale X-ray sources are generally shared resources, with scientists from around the world submitting research proposals to apply for beamtime, i.e., the time allotted for experiments. Approved proposals are granted a specific amount of beamtime, typically between 3 and 7 days, to conduct experiments at these facilities.

During the beamtime, experiments can easily produce hundreds of gigabytes or even terabytes of data every day. Given the large amount of data, manually analyzing it is impractical due to the time and effort required. As a result, automated data analysis tools are required to process and interpret the data efficiently. Since the 2010s, deep learning has been rapidly developed as an essential branch of artificial intelligence (AI). Deep learning involves having machines perform tasks based on available data without human interference. For instance, internet companies use deep learning algorithms to categorize photos in your smartphone's album based on visual content, allowing you to search for images by keywords like "cats" or "selfies." Deep learning algorithms are also employed to recommend personalized advertisements, using data about your online behavior to tailor content that aligns with your interests. Additionally, deep learning is at the heart of virtual assistants, enabling them to understand voice commands, respond to questions, and carry on conversations. Beyond these applications, deep learning also plays a critical role in X-ray imaging. It helps doctors to analyze medical images, create 3D reconstructions, and predict patient outcomes based on historical data.

Deep learning methods are driven by data, making them particularly well-suited to handle the large amount of data generated by these experiments. Once trained, a deep learning algorithm can process data rapidly, nearly in real-time, dramatically cutting down the time needed for analysis, and can help researchers make decisions during experiments.

However, deep learning methods also come with challenges. The first is a lack of explainability in most current deep-learning approaches. AI is often described as a "black box" because we can observe what goes into the system and what comes out, but the underlying processes remain obscure. This lack of transparency makes it difficult to understand how a deep learning model reaches its conclusions, leading researchers to be cautious about relying on AI for critical decisions.

Another challenge is that many commonly used deep learning methods are supervised, requiring paired training data, i.e., sets of inputs and corresponding outputs. Acquiring these large paired datasets can be difficult and impossible in some experimental settings. This limitation can restrict the application of deep learning when sufficient data is unavailable or too costly to obtain.

Finally, the robustness and reliability of deep learning are barriers to more general applications in scientific research. Robustness refers to a model's ability to perform consistently despite variations in data or unexpected environmental changes. Reliability involves the consistency and stability of results over time. Both robustness and reliability are critical for scientific applications.

In this thesis, we investigate deep learning X-ray image reconstruction algorithms for experiments at large-scale X-ray facilities. Specially, we combine deep learning with the physics of X-ray imaging. By incorporating physics, we show that AI approaches can achieve improved explainability, robustness, and reliability. It also enables us to design robust deep learning models without requiring large paired training datasets, which is particularly beneficial for experiments when such data is limited or difficult to get.

Chapter 1

Introduction

Since the discovery of X-rays by Wilhelm Conrad Röntgen in 1895, X-ray imaging has been widely used in many aspects of our daily life, from medical diagnosis to industrial non-destructive inspection. The short wavelength and high penetration power of X-rays make them an excellent scientific probe for studying microstructures from micrometer down to nanometer scales. Early X-ray imaging experiments were carried out in laboratories. Nowadays, with the advancement of science and technology, many large-scale X-ray sources have emerged, and X-ray imaging experiments are carried out not only in laboratories but also at large-scale facilities.

Two of the most important large-scale X-ray sources are synchrotron radiation sources and X-ray free-electron lasers. In 1947, scientists at the General Electric Research Laboratory in Schenectady, New York, first observed the radiation that is now known as synchrotron radiation. Synchrotron radiation is the electromagnetic radiation emitted when charged particles are being accelerated. The main difference between synchrotron radiation sources and laboratory X-ray sources is the brilliance. Simply speaking, brilliance describes the quantity of desired photons that illuminate a unit area of the sample per unit time. A synchrotron light source is up to 10^{15} more brilliant than a laboratory X-ray source. The higher the brilliance, the more precise the information that can be obtained from the X-ray. Since the 1970s, X-ray free-electron laser have been developed, pushing the brilliance of X-rays to unprecedented levels and producing laser-like X-ray beams.

The development of these large-scale X-ray sources has offered opportunities for new X-ray imaging approaches that were not possible before. The enhanced X-ray brilliance facilitates imaging of fast dynamics using short exposures, even with single X-ray pulses. Moreover, phase-contrast imaging has been introduced, which not only captures the absorption but also the phase shift of X-rays as they pass through an object, offering opportunities like a laser.

New opportunities come with new challenges. Reconstructing high-quality representation from these imaging techniques requires computational efforts. Conventional image reconstruction methods are not always applicable to these newly developed techniques. Since the 2010s, deep learning approaches have undergone rapid development and revolutionized the approach to addressing imaging problems. Deep learning methods employ deep neural networks to automatically learn image representations from the given data, allowing for the handling of large datasets without the need for human intervention. Moreover, neural networks are differential operators, facilitating the integration of physical prior knowledge into the learning process.

Deep learning approaches have been widely applied in X-ray imaging. However, current deep-learning methods for X-ray imaging face two major challenges. First, they usually require paired training datasets. One can easily think of scenarios where it is difficult or impossible to obtain a large number of paired datasets. Second, the robustness and reliance of such methods continue to be a barrier to more general applications.

In the present work, we explore how deep learning approaches can potentially address the challenges associated with image reconstructions for X-ray imaging experiments. Specifically, we study image reconstructions of phase-contrast imaging and single-pulse imaging experiments for both static and dynamic scenes.

Chapter 2 introduces the fundamentals of X-ray imaging. This chapter describes the fundamental principles of X-ray propagation and interaction with matter and presents X-ray imaging methods relevant to the thesis.

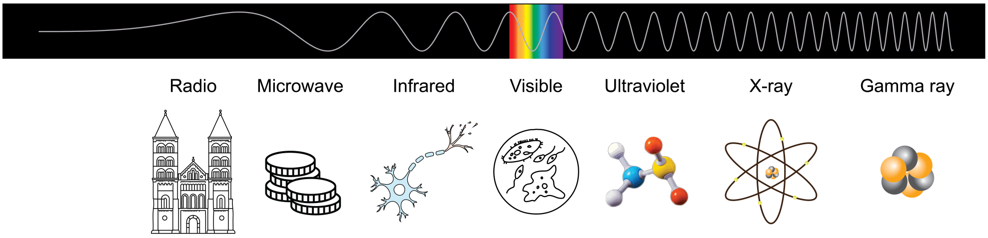
Chapter 3 provides an overview of deep learning and introduces state-of-the-art deep learning models for image reconstruction.

Chapter 4 presents the development of four image reconstruction algorithms for the thesis. We describe the scientific questions and the design of the deep learning algorithms, each addressing specific challenges in X-ray imaging.

Chapter 2

X-ray imaging

Light allows us to perceive and understand the world around us. From a physical perspective, light is an electromagnetic wave. The electromagnetic spectrum contains multiple different components, including visible light and also light that we cannot see, such as radio waves, microwaves, infrared, ultraviolet, X-rays, and gamma rays. The properties of these electromagnetic waves are depicted in Figure 2.1, including their wavelength, frequency, and energy.



	Buildings	Coins	Neurons	Cells	Molecules	Atoms	Nuclei
Wavelength (m)	10^{3}	10^{-2}	10^{-5}	0.5×10^{-6}	10^{-8}	10^{-10}	10^{-12}
Frequency (Hz)	10^{4}	10^{8}	10^{12}	10^{15}	10^{16}	10^{18}	10^{20}
Energy (eV)	10^{-9}	10^{-4}	10^{-1}	10^{1}	10^{2}	10^{4}	10^{6}

Figure 2.1: Illustration of the electromagnetic spectrum. From left to right, the electromagnetic waves decrease in their wavelength and increase in their frequency and energy. The wavelength, frequency, and energy are connected by Equation 2.1. Icons show the length scale of the wavelengths of the electromagnetic waves.

Quantum mechanics revealed that light exhibits wave-particle duality. From a particle view, the energy of the electromagnetic wave is carried by elementary particles called photons. The energy of a photon is proportional to the frequency and inversely proportional to the wavelength of the electromagnetic wave, as in

$$E = h\nu = \frac{hc}{\lambda} = \frac{1.24\,\mathrm{keV}\cdot\mathrm{nm}}{\lambda}, \tag{2.1}$$

where E is the energy of the photon, $h = 6.626 \times 10^{-34}$J·s is the Planck constant, $c = 3 \times 10^8$m/s is the speed of light in vacuum, ν is the frequency, and λ is the wavelength.

Figure 2.1 also depicts example objects that are of the same length scale as the wavelength of each electromagnetic wave. When we illuminate an object with light, it becomes challenging to resolve microscopic structures within the object that are smaller than the wavelength of the light due to the diffraction nature of light. Considering an electromagnetic wave with wavelength λ, traveling in a medium with refractive index n and converging to a spot with half-angle θ, the minimum structure that an optical system can resolve is determined by the Abbe diffraction limit [1]:

$$d = \frac{\lambda}{2n\sin(\theta)}, \tag{2.2}$$

The smaller the d, the better the resolving power of the system. Consequently, the study of structures at smaller scales necessitates electromagnetic waves with shorter wavelengths.

X-rays are electromagnetic waves with a wavelength between 10 nm to 0.01 nm. The short wavelength property of X-rays allows the investigation of microscopic structures down to the atomic level. Specifically, X-rays with wavelength below 0.1 nm (or 1 Å, with 1 Å=10^{-10} m) are categorized as *hard* X-rays, while those with longer wavelength are called *soft* X-rays.

As seen from Equation 2.1, the energy of hard X-rays is higher than that of soft X-rays. As we will see in Section 2.1.1, higher X-ray energy indicates potentially higher penetration ability. For example, the attenuation length of 2 nm soft X-rays in water is less than 1 µm, while it is 3.8 mm for 0.1 nm hard X-rays [2]. The high penetration power of hard X-ray makes it an effective probe for studying the internal features of objects in a non-destructive manner.

This chapter introduces the fundamentals of X-ray imaging and concepts relevant to the thesis work. Firstly, we discuss the interaction between X-rays and matter, both from the atomic level, how X-ray photons interact with the electrons within atoms, and from a computational perspective, how to quantify the perturbation of X-ray wave function when passing through an object. Secondly, we explain the propagation of X-rays and different regimes of diffraction imaging. Following this, the phase-contrast imaging methods are introduced. Expanding the understanding developed in two-dimensional (2D) imaging, we then transition to three-dimensional (3D) and four-dimensional (4D) imaging methods and discuss the possibility of capturing fast dynamics with time-resolved imaging techniques. Concluding the chapter, a brief background on the evolution of advanced large-scale X-ray sources is provided in the end.

2.1 X-ray interaction with matter

Imaging involves mapping interactions in both space and time. Therefore, understanding how X-rays interact with matter forms the basics of X-ray imaging, which is discussed in this section. First, from a particle perspective, we discuss the interaction between individual X-ray photons and individual atoms. Then, from a wave perspective, we introduce the concept of complex refractive index and explain how the X-ray wavefield changes after traveling through a uniform medium.

2.1.1 Atomic interactions

When X-rays travel through a medium, they are absorbed, transmitted, or scattered. From an atomic view, the interaction between X-rays and matter can be treated as the interaction between X-ray photons and the electrons within atoms. X-ray photons interact with electrons in two ways: photoelectric absorption and scattering, as illustrated in Figure 2.2.

As the energy of hard X-ray photons is higher than the binding energy of many core electrons, an incoming X-ray photon can remove a core electron from the atom and ionize it. This process is called *photoelectric absorption* (Figure 2.2 **a**), and the ejected electron is called a *photoelectron*. If the X-ray photon is absorbed, it transfers all its energy to the photoelectron. Following the ejection of the photoelectron, a vacancy is created at the position previously occupied by the photoelectron. Because atoms tend to stay at their lowest energy state with all orbitals filled up in energy order, the vacancy formed by the photoelectron will be filled up by an outer shell electron. The transition of the outer shell election will lead to secondary processes, such as the emission of an X-ray photon (X-ray fluorescence) or the ejection of an Auger electron.

As shown in Figure 2.2 **b**, X-ray photons can also be scattered by electrons. This process can occur in two main ways: coherent scattering or incoherent scattering. Coherent scattering is also called *Rayleigh scattering*. It is an elastic scattering process where the X-ray photon undergoes scattering without losing any energy. The process of incoherent scattering, also known as *Compton scattering*, is inelastic, meaning that the photon loses part of the energy as it is scattered. The energy is transferred to the kinetic energy of the scattering electron, causing it to be ejected from the atom. The ejected electron is called a *Compton recoil electron*. According to the law of conservation of energy and momentum, the energy of the Compton recoil electron is determined by the scattering angle of the X-ray photon.

The probability of these processes is characterized by the cross section, which depends on the element type, electron density, and the physical properties of the X-rays. Two examples of the interaction cross sections for C and Fe atoms are shown in Figure 2.3. As can be seen, from 1 keV to 1 MeV, the total attenuation cross sections are initially dominated

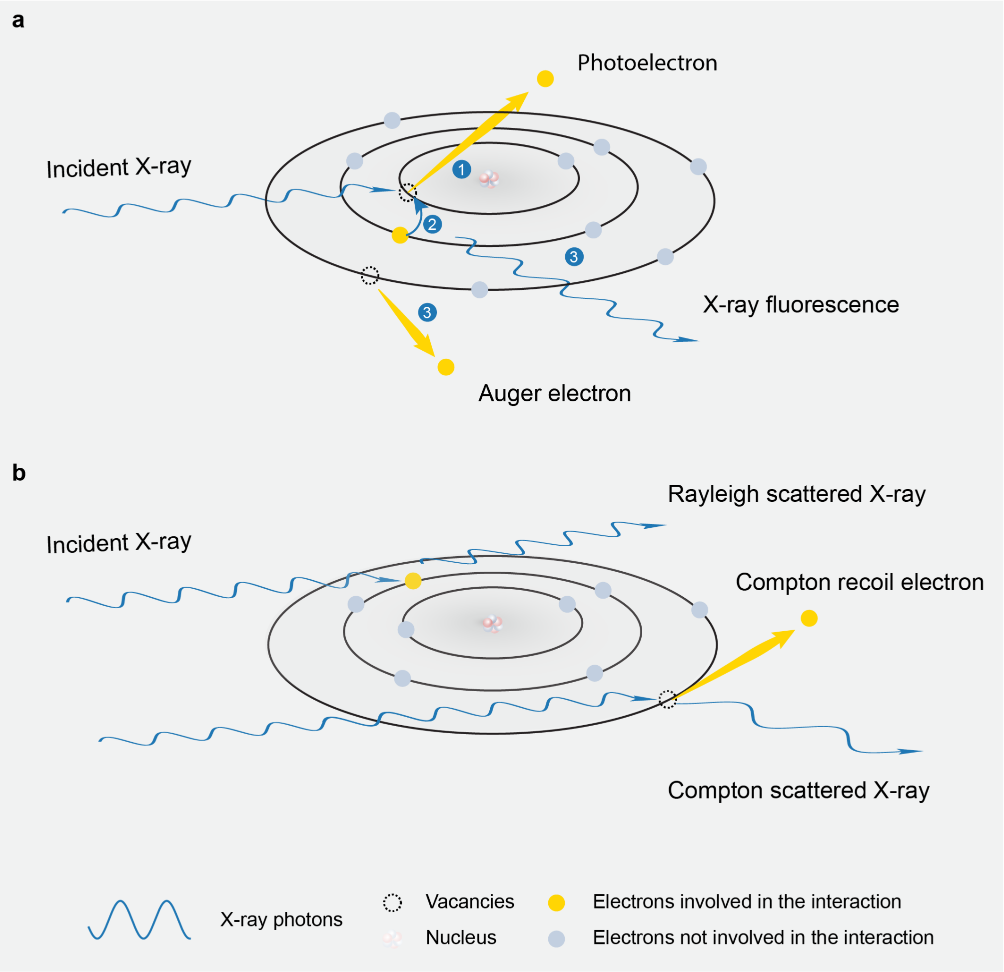


Figure 2.2: Demonstration of X-ray interaction with matter. **a** Photoelectric absorption: The incident X-ray photon is absorbed, leading to the ejection of a core electron. The ejected electron, known as a photoelectron, creates a vacancy in the core, subsequently filled by an outer-shell electron. This process may result in the creation of a photon causing fluorescence or the ejection of an Auger electron. **b** Scattering: The incident X-ray photons are scattered by electrons. The scattering may be elastic (Rayleigh scattering) or inelastic (Compton scattering). In the latter case, the wavelength of the scattered X-rays increases due to energy loss, and the scattering electron is ejected (Compton recoil electron) as a result of the energy transfer.

by photoelectric absorption, transitioning gradually to incoherent scattering. The cross section of the coherent scattering is at least one order of magnitude smaller than the total attenuation cross section. For C, photoelectric absorption dominates the total attenuation when the photon energy is below $\sim$ 20 keV, while incoherent scattering dominates when the energy is above that. For Fe, photoelectric absorption dominates the total attenuation until the photon energy reaches $\sim$ 100 keV, and incoherent scattering takes precedence over photoelectric absorption when the energy is beyond that. At 7.1 keV, the incident photon energy matches the binding energy of the innermost electrons, leading to a sharp jump in

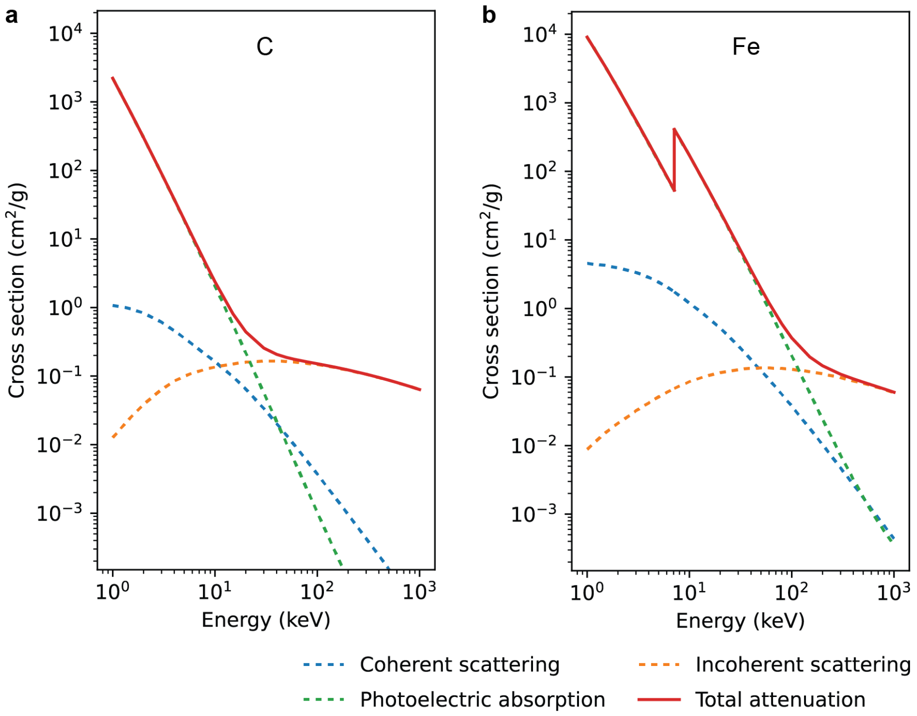


Figure 2.3: Photon cross sections in (**a**) C and (**b**) Fe as a function of energy. The three dashed lines show the cross sections for coherent scattering (blue), incoherent scattering (orange), and photoelectric absorption (green), respectively. The red solid line shows the total attenuation cross section formed by the absorption and scattering processes. Data extracted from the NIST XCOM database [3].

the photoelectric absorption cross section. This is usually referred to as an *absorption edge*. Depending on the energy level of the electrons, absorption edges can be categorized, e.g., a K-shell absorption edge indicates that the X-ray energy is equal to the binding energy of the innermost electron shell, while an L-shell absorption edge indicates that the X-ray energy matches the binding energy of the second innermost electron shell. The absorption edge shown in Figure 2.3 **b** represents the K-shell absorption edge of Fe.

2.1.2 Complex refractive index

So far, we have treated X-rays as particles and discussed the interaction between individual X-ray photons and electrons within atoms. In imaging applications, it is more common to treat X-rays as waves and consider what will happen when X-rays pass through an object instead of an atom.

When X-rays pass through a uniform medium, as depicted in Figure 2.4, it undergoes amplitude attenuation and phase shift. The amplitude attenuation is mainly due to the process of photoelectric absorption and incoherent scattering, while the phase shift is the result of coherent scattering.

The attenuation and phase shift of X-rays is determined by the refractive index of the medium. In X-ray physics, the refractive index n is complex and extremely close to unity. Because of this, it can be expressed by

$$n = 1 - \delta + i \cdot \beta \, , \tag{2.3}$$

where δ, β are real and positive numbers that are much less than unity, as

$$\delta, \beta \in \mathbb{R}^+ \text{ and } |\delta| \, , |\beta| \ll 1 \, . \tag{2.4}$$

In general, both δ and β decrease with the increase of X-ray energy (with exemptions such as absorption edges), while β decreases faster than δ, as will be discussed later (see Figure 2.7). In the hard X-ray regime, δ is of order 10^{-5} in solids and around 10^{-8} in air, and β is orders of magnitude smaller than δ.

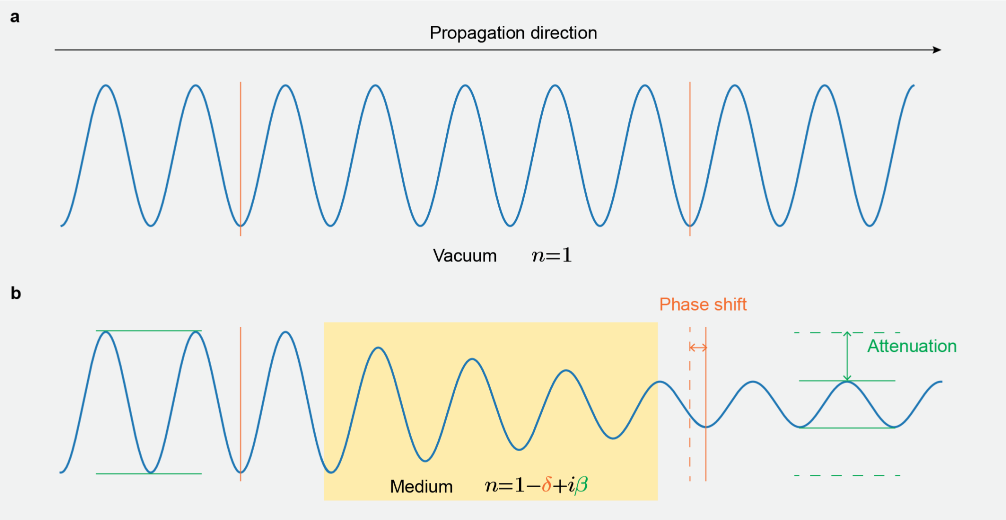


Figure 2.4: Comparison of X-ray propagation in free space (**a**) and through a medium of constant index of refraction n and thickness T (**b**).

The refractive index describes how an incident wavefield is modulated by the object. Let us consider a time-independent incoming plane wave traveling in the z direction, as expressed by

$$\psi_{z_0} = A \exp\left(ikz\right) . \tag{2.5}$$

Here, A describes the amplitude of the incoming X-rays, and $k = 2\pi/\lambda$ is the wavenumber, which is inversely proportional to the X-ray wavelength λ.

After passing through an object with a thickness T and a refractive index n described by Equation 2.3, the transmitted wavefield becomes

$$\begin{aligned}\psi_{z_{\text{exit}}} &= \psi_{z_0} \exp\left(inkT\right) \\ &= \psi_{z_0} \exp\left\{ikT\left[1 - \delta(x, y; z) + i\beta(x, y; z)\right]\right\} \\ &= \psi_{z_0} \exp\left(ikT\right) \exp\left[-ik\delta(x, y; z)T\right] \exp\left[-k\beta(x, y; z)T\right].\end{aligned} \tag{2.6}$$

As can be seen, the result of Equation 2.6 contains a phase shift $\exp\left(ikT\right)$ due to the propagation of X-rays in vacuum. On top of that, the transmitted wavefield undergoes an amplitude attenuation of $\exp\left[-k\beta(x, y; z)T\right]$ and a phase shift of $\exp\left[-ik\delta(x, y; z)T\right]$ with respect to vacuum.

Equation 2.6 is commonly referred to as the *projection approximation*. Consider varying refractive index or object thickness in the (x,y) plane, Equation 2.6 tells us that the exit wave also varies in the (x,y) plane and is a projection of the object. This approximation is usually valid for thin objects where multiple scattering can be neglected. For thicker objects, multi-slice methods may be needed [4, 5]. As we can see in Figure 2.3, when photoelectric absorption dominates the total attenuation, the scattering cross sections are at least an order of magnitude smaller than absorption. In this scenario, the scattered photons are more likely to be absorbed rather than scattered. Therefore, the probability of multiple scattering events is much lower compared to absorption events.

The intensity of the transmitted wavefield can be derived by taking the squared modulus of Equation 2.6, as expressed by

$$I_{\text{exit}} = A^2 \exp(-2k\beta T). \tag{2.7}$$

If we write the intensity of the incoming wavefield as I_0 and the linear absorption coefficient as $\mu = 2k\beta$, we end up at the well-known Beer-Lambert law

$$I_{\text{exit}} = I_0 \exp(-\mu T). \tag{2.8}$$

From Equation 2.8, we see that when the thickness T is equal to the inverse of the linear absorption coefficient $1/\mu$, the intensity of the exit wave is attenuated to about $1/e \approx 37\%$ of I_0. This distance is usually referred to as the *attenuation length*.

As an empirical rule, the linear absorption coefficient of an element is approximately proportional to the fourth power of the atomic number ($\mu \propto Z^4$) and inversely proportional to the third power of the X-ray energy ($\mu \propto E^{-3}$) [6]. Therefore, absorption is more significant in low-energy X-ray regions with high-Z materials. From Equation 2.8, it is evident that absorption is more visible in thick and dense areas of an object compared to thin and light areas. When imaging an object with X-rays, these differences in absorption will create contrasts in the obtained image, which contains information about material properties and thickness of the internal structure of the studied object. Absorption-contrast imaging techniques are commonly used in medical diagnosis and industrial inspection, such as X-ray radiography and computed tomography.

From Equation 2.6, we notice that it is also possible to examine an object by measuring the phase shift of X-rays. Obtaining phase contrast in X-rays is more challenging than absorption due to the difficulty in directly measuring the phase of X-rays caused by their short wavelength and rapid oscillation. Therefore, we are unable to directly measure the phase of the X-rays at the exit plane. The way to measure phase is to use interference. One way to interfere is through propagation-based methods, where we allow X-rays to propagate over a certain distance to let interference patterns build up, and then measure the diffraction patterns at a specific plane located a distance away from the exit plane. In order to produce interference, coherence is required. Details of coherence and the phase-contrast imaging will be discussed in Section 2.3.

2.2 X-ray propagation and imaging regimes

Before introducing phase-contrast imaging methods, we first discuss the propagation of X-ray wavefields in free space and the generation of interference effects. This is presented in the first half of this section. In the second half of this section, we introduce imaging regimes. X-ray imaging experiments can be categorized into different regimes depending on the visibility of the interference patterns. Knowledge of the imaging regimes and wavefield propagation is fundamental for phase imaging experiments.

2.2.1 Free-space propagation

Consider a simple scenario where a plane wave is passing through a single slit, and the diffraction patterns are observed from a distance behind it, as shown in Figure 2.5. If the plane of observation is very close to the slit, almost in contact, the shape of the slit is easily discernible, and interference patterns are minimal. Interference patterns build up as the plane is moved away from the slit, particularly at the edges. As the propagation distance increases, the interference fringes become more structured, while the shape of the slit remains recognizable. This region is commonly referred to as the *near field* or the *Fresnel regime*. With further increase of the propagation distance, diffraction patterns become consistent and nearly independent of the shape of the slit. In this region, as the observation plane is moved further away, the shape of the diffraction pattern remains almost constant, with only the size varying. This area is commonly referred to as the *far field* or the *Fraunhofer regime*.

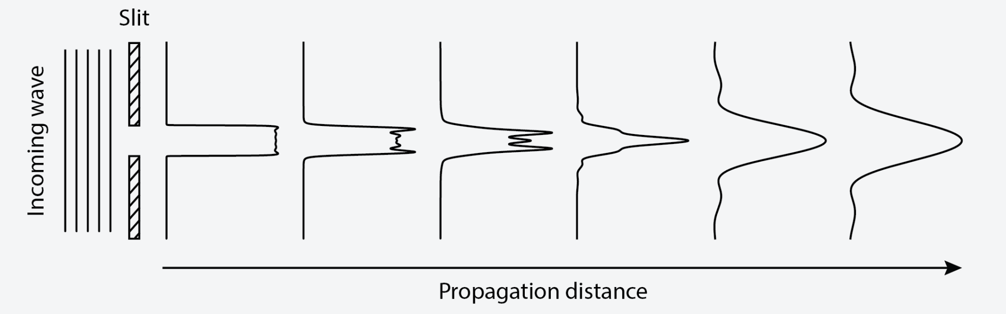


Figure 2.5: Single slit diffraction patterns at different propagation distances. Note that the distances shown between the diffraction patterns are not proportional to the real propagation distances.

In general, the diffraction intensity I of wavefield ψ propagated through a distance L can be calculated by [7]

$$\begin{aligned} I(x, y, z = L) &= |\psi(x, y; z = L)|^2 \\ &= |\mathcal{H}[\psi(x_0, y_0; z = 0)]|^2 , \end{aligned} \tag{2.9}$$

where (x, y) denotes coordinates on the observation plane and (x_0, y_0) denotes coordinates on the aperture plane $(z = 0)$. $\mathcal{H}$ represents the free-space propagation operator, or the *propagator*, that can be expressed by

$$\begin{aligned} \mathcal{H}(\,\cdot\,) &= \mathcal{F}^{-1}\left\{H\mathcal{F}(\,\cdot\,)\right\} \\ &= \mathcal{F}^{-1}\left\{\exp\left[ikL\sqrt{1-(\lambda f_x)^2-(\lambda f_y)^2}\right]\mathcal{F}(\,\cdot\,)\right\} . \end{aligned} \tag{2.10}$$

Here, $\mathcal{F}$ represents the Fourier transform, and H is the Helmholtz *free-space transfer function* representing the phase delay of each spatial frequency component over the propagation distance, and f_x and f_y are the spatial frequencies in the x and y direction.

The method of studying the propagation of light in free space using Equation 2.9 and Equation 2.10 is generally known as the angular spectrum method or momentum decomposition, which formulates a complex wavefield as a linear combination of a series of plane waves of the same frequency and different directions.

A thorough derivation of Equation 2.10 involves solving the time-independent Helmholtz equation, which originates from the Maxwell wave equation. Helmholtz equation governs the propagation of electromagnetic waves in free space, as expressed by

$$\left(\nabla^2 + k^2\right) \psi\left(x, y, z\right) = 0, \tag{2.11}$$

where $\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$ is the Laplace operator. Note that wavefields are treated as scalars in this equation, which neglects the vectorial nature (polarization) of electromagnetic fields. This assumption is valid when the process of light propagation does not involve optical components close to the size of the wavelength of light, and the study of the diffraction problem is not adjacent to the plane of diffraction. Beyond these conditions, the scalar theory can still be accurate to some extent [8]. We refer the readers to [7] for detailed discussions about the theory of wavefield propagation.

Fresnel approaximation

Now we further consider the case where $|\lambda f_x| \ll 1$ and $\left|\lambda f_y\right| \ll 1$ in Equation 2.10. Physically, this means the waves are propagated in small angles with respect to the propagation direction. In this case, we can simplify Equation 2.10 by applying Maclaurin expansion $\sqrt{1-(\lambda f_x)^2-(\lambda f_y)^2} \approx 1 - \frac{\lambda^2}{2}(f_x^2 + f_y^2)$. Under this approximation, the propagator can be written as

$$\mathcal{H}(\,\cdot\,) = \mathcal{F}^{-1}\left\{\exp(ikL)\exp\left[-i\pi\lambda L(f_x^2 + f_y^2)\right]\mathcal{F}\left(\,\cdot\,\right)\right\}. \tag{2.12}$$

This is the propagator for the near field or the Fresnel regime. In other words, Fresnel propagation is equivalent to paraxial propagation, i.e., the wavevectors of the plane wave components of the complex wave field have small angles with respect to the propagation direction.

Fresnel diffraction equation is also commonly written in its integral form, given by

$$\psi(x, y; L) = \frac{\exp(ikL)}{i\lambda L}\iint_{-\infty}^{\infty}\psi(x_0, y_0; 0)\exp\left\{i\frac{k}{2L}\left[(x-x_0)^2 + (y-y_0)^2\right]\right\}dx_0dy_0. \tag{2.13}$$

This equation, known as the *Fresnel diffraction integral*, can also be represented as a convolution

$$\psi(x, y; L) = \iint_{-\infty}^{\infty}\psi(x_0, y_0; 0)\, h(x-x_0, y-y_0; L)dx_0dy_0, \tag{2.14}$$

where

$$h(x, y; L) = \frac{\exp(ikL)}{i\lambda L}\exp\left[\frac{ik}{2L}\left(x^2 + y^2\right)\right]. \tag{2.15}$$

The convolution kernel $h(x, y; L)$ is also called the *real-space transfer function* or the *impulse response function*, which describes the response of the system to an impulse function. As can be seen, the Fourier transform of the impulse response function $h(x, y; L)$ is equal to the free-space transfer function derived in Equation 2.12, as

$$\mathcal{F}[h(x, y; L)] = H(f_x, f_y; L) = \exp(ikL) \exp\left[-i\pi\lambda L(f_x^2 + f_y^2)\right]. \tag{2.16}$$

Therefore, according to the convolution theorem, Equation 2.13 and Equation 2.12 are equivalent.

By moving the $\exp\left[\frac{ik}{2L}\left(x^2 + y^2\right)\right]$ term outside the integral, the Fresnel diffraction integral (2.13) can also be viewed as a Fourier transform, as expressed by

$$\psi(x, y; z = L) = h(x, y; L)\, \mathcal{F}\left\{\psi(x_0, y_0; z = 0)\, \exp\left[\frac{ik}{2L}\left({x_0}^2 + {y_0}^2\right)\right]\right\}. \tag{2.17}$$

This result indicates that the propagated wavefield can be calculated by multiplying the real-space transfer function at the observation plane by the Fourier transform of the product of the wavefield before propagation and a quadratic phase exponential.

Fraunhofer approaximation

If the propagation distance is further increased, a more stringent assumption can be considered, expressed as

$$L \gg \frac{k}{2}({x_0}^2 + {y_0}^2)_{\max}. \tag{2.18}$$

Under this assumption, the quadratic phase exponential in Equation 2.17 can be approximated as 1. Consequently, Equation 2.17 can be greatly simplified, yielding

$$\begin{aligned} \psi(x, y; z = L) &= h(x, y; L)\, \mathcal{F}\left[\psi(x_0, y_0; z = 0)\right] \\ &= \frac{\exp(ikL)}{i\lambda L} \exp\left[\frac{ik}{2L}\left(x^2 + y^2\right)\right] \mathcal{F}\left[\psi(x_0, y_0; z = 0)\right]. \end{aligned} \tag{2.19}$$

The factor $\frac{\exp(ikL)}{i\lambda L}$ is independent of x and y. Therefore, it can be neglected when calculating the relative distribution of the diffraction intensity. In this context, the propagated wavefield can be simply expressed as the product of a quadratic phase exponential and the Fourier transform of the unpropagated wavefield, as

$$\mathcal{H}(\,\cdot\,) = \exp\left[\frac{ik}{2L}\left(x^2 + y^2\right)\right] \mathcal{F}(\,\cdot\,). \tag{2.20}$$

Equation 2.20 represents the propagator for the far field or the Fraunhofer regime.

2.2.2 Imaging regimes

As we can see from Figure 2.5, the visibility of the interference fringes increases with the increase of the propagation distance. In fact, the diffraction effect is not only dependent on the propagation distance (L) but also on the wavelength of the incoming wave (λ) and the size of the imaging feature (a). The visibility of diffraction effects can be determined by the *Fresnel number*, which is defined as

$$N_F = \frac{a^2}{\lambda L}. \tag{2.21}$$

Imaging experiments can be classified into three regimes depending on the Fresnel number:

- $N_F \gg 1$: Contact regime, where the interference pattern cannot be resolved.
- $N_F \sim 1$: Fresnel regime or the near field, where the interferences fringes become visible, but the shape of the aperture is still recognizable. Diffraction largely reflects local effects in this regime.
- $N_F \ll 1$: Fraunhofer regime or the far field, where the shape of the aperture is invisible and the diffraction pattern represents the Fourier transform of the original aperture (non-locality).

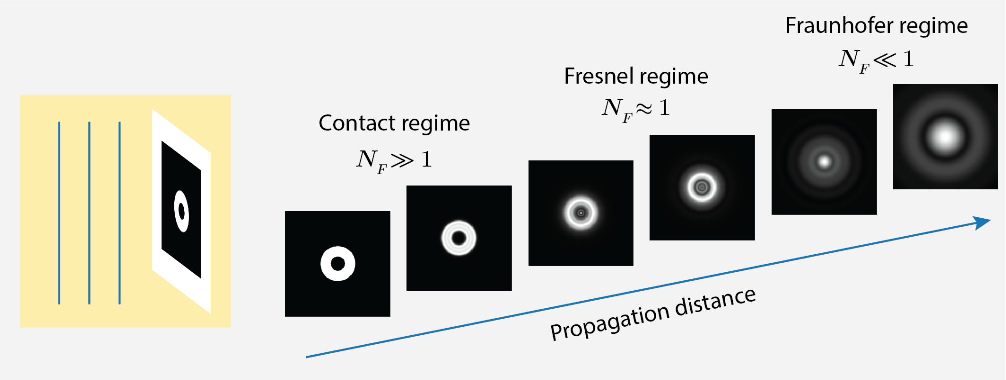


Figure 2.6: Illustration of different imaging regimes. The diffraction patterns develop with the increase of the propagation distance, and the imaging regime shifts from the contact regime to the Fresnel regime, and finally to the Fraunhofer regime.

As can be seen in Figure 2.6 and Equation 2.21, for a fixed light wavelength and feature size, the visibility of diffraction effects builds up with the increase of the propagation distance. In the contact regime, the detector intensity contains no contrast that is due to the phase shifts, i.e., the diffraction effect is negligible. As the propagation distance increases, the interference patterns start to show up, and the intensity image contains not only the amplitude information but also the phase.

Imaging samples with hard X-rays ($\lambda < 0.1$nm) in all three imaging regimes is not always feasible. If the size of the feature of interest is below 1 nm, it can be seen from Equation 2.21 that $N_F \sim \frac{10^{-8}(\mathrm{m})}{L}$, indicating that obtaining information from the contact or Fresnel regime is not realistic. Therefore, achieving nanometer imaging resolution is constrained to the Fraunhofer regime, for instance, through methods such as coherent diffraction imaging (CDI) [9] and ptychography [10]. For imaging at micrometer resolution, $N_F \sim \frac{10^{-2}(\mathrm{m})}{L}$, allowing experiments to be performed at different regimes depending on the position of the detector. If the feature of interest is larger than 1 mm, $N_F \sim \frac{10^{4}(\mathrm{m})}{L}$, suggesting that diffraction patterns may not be visible unless a very long propagation distance is used.

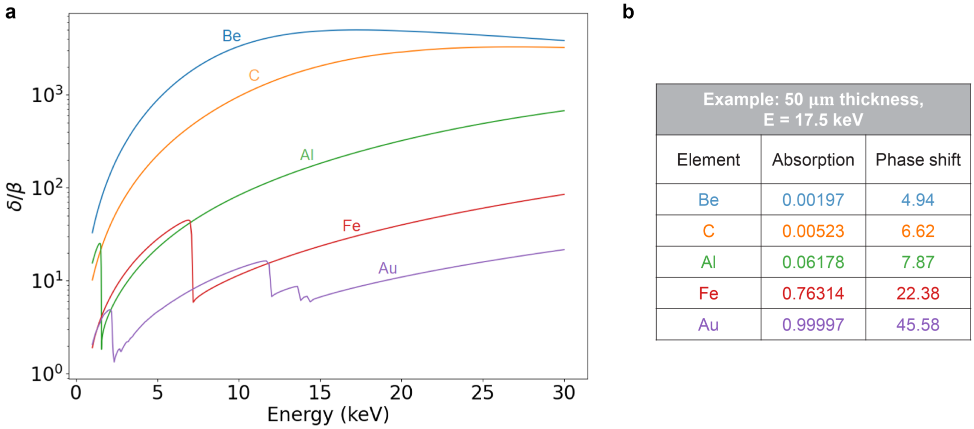
In the present work, we mainly focus on the contact regime and the Fresnel regime. **Paper II** and **Paper III** address holographic phase reconstruction problems, which fall within the Fresnel regime. The experiments performed in the remaining papers are focused on the contact regime. From a computational standpoint, it is important to note that the propagators used in these imaging regimes, despite being described in different mathematical forms, are all unitary differentiable operators. The differentiable property is of great importance for deep learning applications. It allows for automatic differentiation, which is the basis of deep learning. In addition, the deep learning methodologies designed for one imaging regime can be easily adapted to other regimes as well. For example, the idea of including a physical propagator into neural networks, as presented in **Paper II**, can be applied to the Fraunhofer regime by substituting the Fresnel propagator (Equation 2.12) with the Fraunhofer propagator (Equation 2.20), as evidenced in [11].

2.3 Phase-contrast imaging

In this section, we introduce phase-contrast imaging, i.e., imaging the phase shift of the X-rays when passing through an object.

Although X-ray absorption-contrast imaging has proven to be a versatile and straightforward technique, its effectiveness may be compromised when imaging elements with similar absorption contrast or light materials like soft tissues. Phase-contrast imaging emerges as a promising alternative, enabling the imaging of light elements that pose challenges for absorption methods.

When imaging light materials like soft tissues using absorption methods, variations within the materials are mainly visible at low photon energies [12]. As the photoelectric absorption increases greatly with the decrease of energy (Figure 2.3), imaging at lower energies



Example: 50 µm thickness, E = 17.5 keV		
Element	Absorption	Phase shift
Be	0.00197	4.94
C	0.00523	6.62
Al	0.06178	7.87
Fe	0.76314	22.38
Au	0.99997	45.58

Figure 2.7: The absorption and phase shift determined by the complex refractive index as a function of X-ray energy for different materials. **a** The ratio δ/β of five common materials at different X-ray energies. **b** The comparison between absorption and phase shift for the five elements at an example condition: $50\,\mu\mathrm{m}$ thickness and a photon energy of 17.5 keV.

may lead to higher radiation damage on the sample being measured, although the contrast is enhanced. An exception to this is the "water window" spectral region, which lies between the carbon K-shell absorption edges at 290 eV and the oxygen K-shell absorption edges at 540 eV. In this region, organic materials such as biological cells show particularly good absorption contrast because water is relatively transparent while organic materials are not [5, 13]. In many cases, phase-contrast imaging provides enhanced image quality with reduced radiation exposure compared to absorption-contrast techniques. As discussed in Section 2.1.2, the refractive index of a medium contains a real part δ and an imaginary part β, where the phase shift is proportional to the real part δ and the amplitude attenuation is proportional to the imaginary part β. Figure 2.7 **a** shows the ratio δ/β for five common materials at different X-ray energies. For light materials like beryllium and carbon, δ is generally two to three orders of magnitude greater than β in the hard X-ray region, indicating the potential for low-dose imaging. An example is given in Figure 2.7 **b**, where, at a photon energy of 17.5 keV and a thickness of 50 µm, the difference in absorption for light materials is minimal, while the phase shift is apparent. Therefore, phase imaging offers a low-dose possibility for imaging materials with low or similar absorption contrasts.

This section is divided into three subsections. We first explain the concept of coherence, which is required for phase-contrast imaging experiments. Then, we introduce in-line holography, a commonly used phase-contrast imaging technique. Finally, we discuss the phase problem, i.e., reconstructing phase information from the intensity measurements.

2.3.1 Coherence

In our discussion of wave field propagation, we have confined ourselves to monochromatic plane waves traveling in the same direction. In reality, the incoming X-ray waves may have different wavelengths and may not propagate precisely in the same direction, which can affect the visibility of the interference pattern. As an example, if the waves are completely out of phase, no interference pattern will be visible. Certain criteria must met for the interference pattern to be visible.

Coherence describes the capability of waves to interfere with each other. Waves are considered fully coherent if they maintain a constant phase difference during propagation. However, it is challenging to achieve full coherence in illumination. In this context, it is relevant to discuss under what conditions light waves can be considered coherent.

Coherence involves both temporal coherence and spatial coherence. Temporal coherence describes the correlation between the phases of a light wave at different points along the direction of propagation, which indicates the monochromaticity of the light wave. The degree of temporal coherence is expressed by longitudinal coherence length. Let us consider two waves from the same source, one with a wavelength of $\lambda_1 = \lambda$ and the other with a wavelength of $\lambda_2 = \lambda_1 + \Delta\lambda$. Longitudinal coherence length quantifies the distance over which the two waves become fully out of phase, as can be calculated by

$$L_L = \frac{\lambda^2}{2\Delta\lambda}, \tag{2.22}$$

Spatial coherence describes the correlation between the phases of a light wave at different points transverse to the direction of propagation. It indicates how much the waves deviate from a well-defined propagation direction due to the finite size of the light source. The degree of spatial coherence is quantified by the transverse coherence length, expressed as

$$L_T = \frac{\lambda R}{2D}. \tag{2.23}$$

Here, D is the size of the incoherent light source, and R is the distance between the source and the imaging plane. The size of the X-ray beam is usually different in the horizontal and vertical directions. Therefore, it is more helpful to split the transverse coherence length into an x and a y component, as $L_{Tx} = \frac{\lambda R}{2D_x}$, and $L_{Ty} = \frac{\lambda R}{2D_y}$. As can be seen, the transverse coherence length increases as light propagates, suggesting that waves from a finite source will naturally become coherent over distance.

An illustration of the longitudinal coherence length and transverse coherence length is illustrated in Figure 2.8. Typically, the volume formed by the product of the longitudinal coherence length and the transverse coherence length in the two directions is referred to as

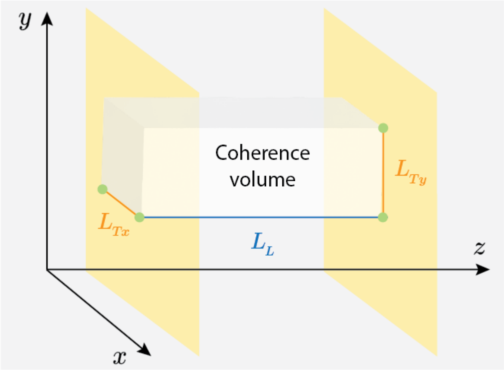


Figure 2.8: Conceptual illustration of longitudinal coherence length (L_L), transverse coherence length along the two directions (L_{Tx} and L_{Ty}), and coherence volume. Light propagates along the z-axis.

the coherence volume, as can be expressed by

$$V = L_L \cdot L_{Tx} \cdot L_{Ty} = \lambda^3 \frac{\lambda}{2\Delta\lambda} \frac{R^2}{4D_x D_y}. \tag{2.24}$$

It determines the maximum scale at which light can be considered coherent, or partially coherent, and will give rise to the interference effect. For phase-contrast imaging experiments, it is crucial that the imaging feature falls within the coherence volume of the incoming waves to capture interference patterns effectively.

2.3.2 In-line holography

As mentioned in Section 2.1.2, phase contrast is achieved by analyzing the diffraction patterns of X-rays following a certain distance of propagation. The diffraction patterns carry the phase information of the measuring object. One of the simplest phase-contrast methods is *in-line holography*, which differs from conventional absorption imaging only in that a propagation distance is introduced between the object and the detector to provide sufficient distance for the refracted rays to be separable from the unperturbed ones.

In-line holography was proposed by Dennis Gabor, who was awarded the Nobel Prize in Physics “for his invention and development of the holographic method” in 1971. As shown in Figure 2.9, in-line holography is a lensless imaging approach, where the light illuminates directly on the sample without any optical components. Holography experiments involve two wavefields: a reference wavefield, which is a coherent light beam, and an object wavefield formed by light scattered from the object. A *hologram* is created by recording the interference pattern between the reference wave and the object wave. The name “hologram” came from the Greek word “holos” — the whole, because it contains the whole information, i.e., amplitude and phase [14].

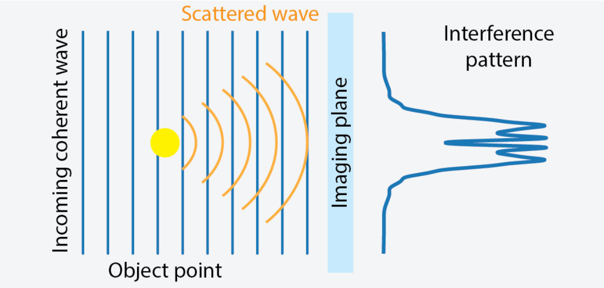


Figure 2.9: Schematic setup of in-line holography.

Originally, holograms were recorded on photographic films, which capture the intensity distribution generated by the interference of the two wavefields. To read the hologram and reconstruct the object wavefront, a coherent light source similar to the reference one is then used to illuminate the hologram [14]. Nowadays, holograms are usually recorded by digital cameras, and the subsequent reconstruction process is carried out on computers.

Reconstructing the phase of the object from the detector intensity involves phase-retrieval methods, as will be discussed in the next subsection. Before describing the solutions, we first provide a brief explanation of the twin-image problem, a longstanding challenge in in-line holography.

If we write the reference wavefield as U and the object wavefield as O, then the captured intensity of the hologram can be expressed as

$$I = |U + O|^2 = |U|^2 + |O|^2 + UO^* + U^*O. \tag{2.25}$$

The recorded hologram includes the object field O and its conjugation O^*, representing a real and virtual image. As the reference wave and the object wave share the same optical path for the in-line holography, these two images are not separable. Therefore, this problem is known as the twin image problem. One way to overcome the twin image problem is to use off-axis holography, which employs a small angle between the reference beam and the object beam to separate the real image from the virtual one [15]. As off-axis holography usually requires additional experimental effort than in-line holography, it is beyond the scope of this thesis. The twin image problem can also be solved via computational approaches, e.g., iterative algorithms and deep learning methods, as will be discussed in the next section.

2.3.3 Phase retrieval methods

In phase-contrast imaging experiments like in-line holography, the phase information of the wave after passing through an object is recorded in the interference measurement. However, as the detectors are only sensitive to the intensity of the signal, the phase information is lost. Reconstructing phase from an intensity-only measurement is referred to as *the phase problem*.

The phase problem is an inverse problem [16]. Generally speaking, phase retrieval methods address the inverse problem of reconstructing original sample images from detector intensity images measured by a specific optical system. It can be described by

$$\widehat{O} = \arg\min_{O} \|I - F(O)\|\,, \tag{2.26}$$

where $\widehat{O}$ is the reconstructed object images, I is the measured intensity, O is the ground truth. F is the forward imaging model that related the object O to the measured intensity I. The goal of solving this equation is to accurately reconstruct the object image to match the ground truth.

Conventional phase retrieval methods can be categorized into two main types: analytical methods and iterative methods.

Analytical methods reconstruct the phase by solving the forward imaging model based on certain assumptions, The transport-of-intensity equation (TIE) is one of the analytical methods that describes the relationship between intensity and the second derivative of the phase of an object. Under the short-propagation assumption, phase can be determined by recording intensity at different propagation distances [17]. Another commonly used analytical approach is the contrast transfer function (CTF) method, which relies on the assumption of weak absorption and slowly varying phase shifts. The CTF method is not limited by the short-propagation requirement, but it typically requires more intensity measurements at different distances [18]. As it is not always feasible to capture images at different distances, single-distance phase retrieval approaches are often favored. While single-distance phase reconstruction approaches have also been proposed, they often rely on stricter assumptions that may not always be valid [19, 20, 21].

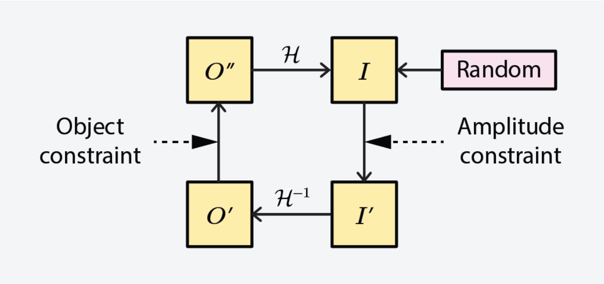


Figure 2.10: Diagram of iterative phase retrieval method.

On the other hand, iterative approaches can be applied to reconstruct phases from single-intensity images. As illustrated in Figure 2.10, iterative phase retrieval methods project between the object space and the detector space using physical propagators and impose constraints on both spaces [22, 23]. These methods start by randomly initializing a phase to the intensity measurement, and then iteratively minimize the difference between the projected intensity and the measured intensity until convergence is achieved. However, these methods may suffer from long computation times, and convergence is not always guaranteed [20].

Recently, there has been a growing interest in deep learning phase retrieval approaches, with demonstrations of real-time image reconstructions [24, 25]. In **Paper II**, we introduce a deep-learning-based phase retrieval approach, which incorporates the physical propagator into the learning process of deep neural networks and eliminates the need for paired datasets during network training. This approach provides a solution for scenarios where conventional phase retrieval approaches fail or are not applicable. Once trained, it enables rapid and high-quality phase reconstruction using single detector images.

2.4 Extend to 3D and 4D imaging

Up to this point, we have discussed the interaction of X-rays with matter and contrast mechanisms of X-ray imaging, including absorption and phase contrast imaging. Since we live in a 3D world, it is natural to think about imaging an object directly in 3D. Moreover, understanding the evolution of objects over time is fundamental to the study of many scientific phenomena. Therefore, 4D imaging, which integrates 3D spatial information with temporal evolution, is indispensable for dynamic studies.

In this section, we present state-of-the-art 3D and 4D X-ray imaging methods.

The 3D information of an object can be reconstructed from its 2D projections [26]. These projections are typically acquired by scanning the sample from different directions over a period of 180° or 360°, as performed in computed tomography (CT). In recent years, rotation-free 3D imaging techniques have been developed, which illuminate the sample simultaneously from multiple angles without scanning [27, 28, 29]. We first explain the theory and reconstruction methods of CT. Following that, we briefly describe tomoscopy, a 4D imaging technique based on tomography. Finally, we introduce X-ray multi-projection imaging, a rotation-free 4D imaging technique for fast dynamics. In particular, we emphasize the difficulties of tomoscopy for dynamic imaging and the need to apply X-ray multi-projection imaging to capture fast dynamics.

Please note that the methods presented in this section are applicable to both absorption and phase contrast imaging, due to the fact that both the absorption and phase shift can be considered as projections through, e.g., line integrals, as described in Section 2.1.2.

2.4.1 Computed tomography

Computed tomography (CT) was first developed by Allan M. Cormack and Godfrey N. Hounsfield in 1972 [26], who were awarded the Nobel Prize in Physiology or Medicine "for the development of computer assisted tomography" in 1979. The concept behind CT involves capturing multiple projections of the sample from various angles and subsequently combining them through computational methods to reconstruct a 3D *tomogram*. Since its development, CT has been widely used in medical diagnosis and non-destructive industrial inspections. In medical CT, the sample (patient) remains fixed, while the X-ray source is rotated, with the patient lying in the center. For imaging at large-scale facilities like synchrotron radiation sources, the light usually travels a long distance before reaching the experimental stage. Therefore, it is impractical to rotate the light sources. Luckily, as the samples being measured are usually much smaller than a patient, it is possible to rotate the sample instead of the light source. An illustration of the CT setup is shown in In this context, the setup of a CT experiment can be simplified as a beam of X-rays, a sample rotation stage, and an X-ray detector, as shown in Figure 2.11.

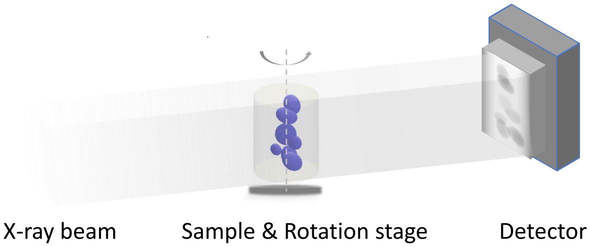


Figure 2.11: Basic components of a CT experiment.

Fourier slice theorem

The collected 2D projections from different angles can be combined into 3D using back projections according to the *Fourier slice theorem*.

An illustration of the Fourier slice theorem is shown in Figure 2.12. Here, we use the 2D Shepp–Logan phantom image as an example, which was created by Larry Shepp and Benjamin F. Logan in the 1970s. This image simulates the structure of a human head and serves as a standard test image for the evaluation of image reconstruction algorithms [30]. The Fourier slice theorem states that when we project an image $f(x, y)$ along a specific angle θ and then perform a 1D Fourier transform on the projection, the result is equivalent to taking a 2D Fourier transform of the image $F(u, v)$ and extracting the slice corresponding to the same angle θ [6, 5]. Here, $F(u, v) = \mathcal{F}(f(x, y))$.

Extending this to 3D, the Fourier slice theorem states that the slice of the 3D Fourier trans-

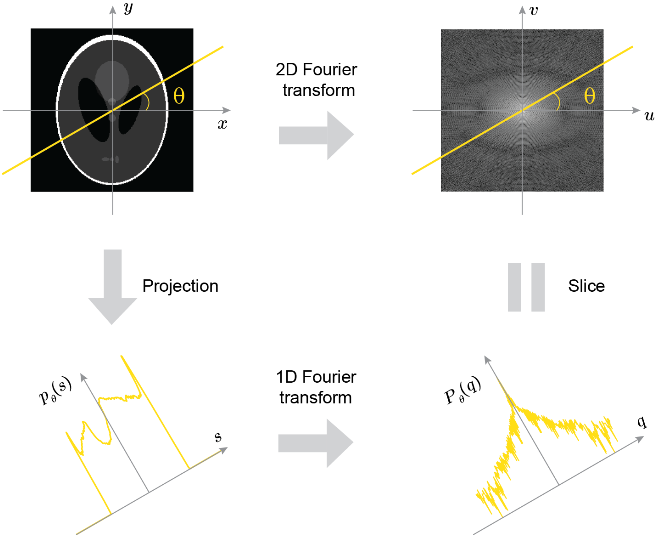


Figure 2.12: Illustration of the Fourier slice theorem.

form of an object along a plane perpendicular to the projection direction is identical to the 2D Fourier transform of the corresponding projection image. Therefore, the 3D image can be reconstructed by combining the 2D Fourier transform of projections from various angles to fill the 3D Fourier domain and then performing the inverse 2D Fourier transform. In the experiment, we usually treat 3D volume as a stack of 2D slices and reconstruct the 3D by reconstructing each individual 2D slice.

Crowther criterion

A certain number of projections is necessary to adequately fill the Fourier space for CT reconstruction. This number, denoted by N_θ, is determined by the *Crowther criterion*

$$N_\theta = \frac{\pi D}{d}, \tag{2.27}$$

where D denotes the maximum size of the imaging object and d is the spatial resolution of the 3D reconstruction. Depending on the dimension and resolution of the experiment, several hundred to thousands of projections are typically required for a proper reconstruction.

Reconstruction algorithms

Although the Fourier slice theorem provides a theoretical foundation for reconstructing sample details from projections, directly applying it for the reconstruction may cause problems.

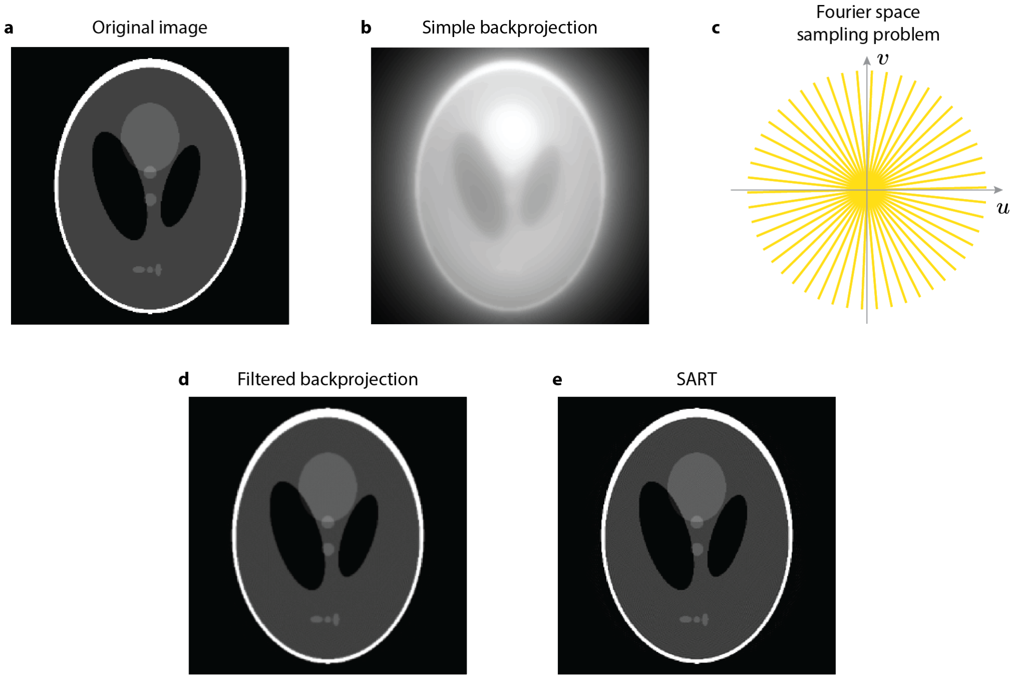


Figure 2.13: Comparison between CT reconstructions from simple backprojection and filtered backprojection. **a** The Shepp–Logan test phantom. $N_\theta = 400$ projections of this phantom were simulated from different angles, equally spaced from $0°$ to $180°$, which satisfy the Crowther criterion for the 256×256 image. **b** The reconstruction results of simple backprojection. **c** Illustration of sampling problem in Fourier space. The reconstruction of simple backprojection is oversampled in the low-frequency area and undersampled in the high-frequency area. **d** The reconstruction results of filtered backprojection. The ramp filter was applied for the reconstruction. **e** The reconstruction results of simultaneous algebraic reconstruction technique (SART). The result is obtained after ten iterations.

To illustrate the problem, we simulate $N_\theta = 400$ projections from the Shepp–Logan phantom image (Figure 2.13 **a**) and then show the results reconstructed from the projections. The number of projections sufficiently satisfies the Crowther criterion for the 256×256 image. Figure 2.13 **b** shows the results of simple backprojection, where the projections are simply smeared back in the direction from which they were measured. As can be seen, this back-smearing process causes blurring in the reconstructed images. When we fill the Fourier space with slices, the sampling density is different for low and high frequencies. As illustrated in Figure 2.13 **c**, the sampling density decreases with the increase of the frequency, causing oversampling for the low-frequency area and undersampling for the high-frequency area. This blurring can be mathematically corrected by applying filters in the backprojection process, which is known as filtered backprojection. As shown in Figure 2.13 **d**, filtered backprojection fixes the blurring problem in simple backprojection and properly reconstructs the shape of the original image.

Although filtered backprojection can effectively reconstruct the sample from its projections, complete projections are required to sample sufficiently in the Fourier space, as described by the Crowther criterion. In many scenarios, it is challenging to acquire a complete set of projections due to practical concerns. For example, medical imaging aims to limit the number of projections needed to minimize the radiation doses exposed to the sample. In addition, the projection angles may be limited in a specific range because of the constraints of the scanning environment. Filtered backprojection may run into difficulties when handling incomplete projections. In such a scenario, iterative approaches may be applicable. Iterative approaches solve the reconstruction problem by minimizing the difference between the measured projections and the predicted projections. Figure 2.13 **e** shows the reconstruction result of simultaneous algebraic reconstruction technique (SART), an effective iterative method.

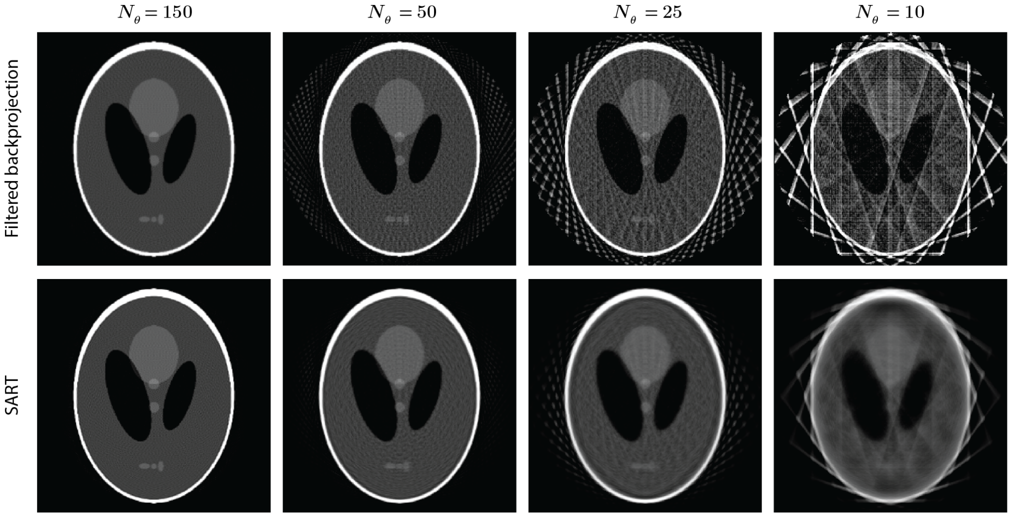


Figure 2.14: Reconstruction results of filtered backprojection and SART with different numbers of projections (N_θ).

Figure 2.14 shows the reconstruction results with a limited number of input projections. The number of projections decreases from 150 to 50, 25, and down to 10. The quality of the reconstruction results appears acceptable when using 150 projections, but drops significantly when using fewer than 50 projections. When comparing the two reconstruction methods, SART outperforms filtered backprojection; nonetheless, neither of them is able to reconstruct the details of the phantom using fewer than 25 projections.

Deep learning approaches that incorporate prior knowledge can also be applied to CT reconstructions. These approaches have demonstrated the capability to achieve high-quality reconstructions from extremely sparse projections [31, 32]. In **Paper IV**, we present a deep-learning-based sparse-view reconstruction method, which integrates state-of-the-art deep-learning concepts with the physics of X-ray imaging. This approach can effectively reconstruct 3D from less than ten projections.

2.4.2 Tomoscopy

Tomography can be further extended to 4D if temporal information is accessible. Time-resolved tomography experiment is also termed *tomoscopy*. Tomoscopy exploits fast rotation stages and high-brilliance X-ray sources to capture multiple tomograms per second, enabling the recording of 3D movies (4D) showcasing dynamic processes.

State-of-the-art tomoscopy techniques can record up to 1000 tomograms per second, facilitating the capture of dynamics at kHz rates. This has been demonstrated in [33], where the foaming process of metal bubbles was recorded with a temporal resolution of 1 ms and a spatial resolution of 8.2 µm.

However, the rotation process inherent in tomoscopy has led to some problems, limiting its applications and the achievable temporal resolution. These difficulties arise from both the equipment used for rotation and the nature of the sample itself. Firstly, the rotation itself is constrained by mechanical limitations, preventing it from exceeding a certain speed. As a result, the projections obtained during rotation are not simultaneous, meaning that dynamics occurring faster than the rotation speed will not be captured or reconstructed. Secondly, the sample being imaged needs to withstand rapid rotations, which may not always be possible. The centrifugal force F acting on an object of mass m can be calculated by

$$F = ma = m\omega^2 r, \tag{2.28}$$

where a represents the radial acceleration, ω is the angular velocity of the rotation, and r is the distance between the center of mass of the object and the axis of rotation. Achieving a temporal resolution of 1 ms with tomoscopy necessitates a minimum of 500 rotations per second, resulting in a radial acceleration that is 500 times the gravitational acceleration for an object with a diameter of 1 mm. This could potentially alter the structure or dynamics of the sample. Besides, preparing a sample environment for fast rotations can be challenging.

One way to improve temporal resolution while avoiding sample rotation is to use multi-projection approaches [27, 28, 29, 34, 35], which use multiple beams to simultaneously illuminate the sample from different angles. These approaches offer improved temporal resolution without sample rotation at the cost of a reduced number of projections, as wii be discussed in the next subsection.

2.4.3 X-ray multi-projection imaging

X-ray multi-projection imaging (XMPI) is a rotation-free imaging approach that has been developed to address the aforementioned problems in tomoscopy.

In contrast to tomoscopy experiments, which capture projections by rotating the sample, XMPI relies on advanced X-ray sources and a series of beam splitters to split the beam and simultaneously record projections from various angles.

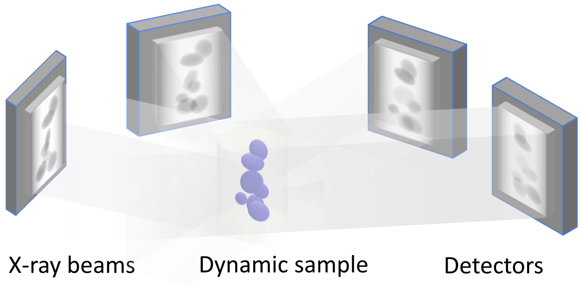


Figure 2.15: XMPI conceptual illustration.

The concept of XMPI is illustrated in Figure 2.15. As can be seen, the sample remains fixed at the center, while the incoming X-ray beam is divided into multiple beamlets by a group of beam splitters, often crystals. The split beamlets simultaneously illuminate the sample from different angles, and the projections are recorded using ultrafast cameras from various angles. The recorded projections can then be used for 3D and 4D reconstruction.

As a rotation-free method, XMPI has several advantages as well as disadvantages over tomoscopy. Firstly, it eliminates the constraint on temporal resolution imposed by the rotation process. The resolution is ultimately determined by the speed of the detector and the flux of the X-ray source. Secondly, XMPI allows for in situ and operando studies of rotation-sensitive samples and dynamical processes. There is no need to prepare a reaction environment capable of tolerating fast rotations, and dynamic processes can be measured in real situations without the disturbance of rotations. Thirdly, it is a single-shot 3D approach. When combined with the unique capabilities of high-brilliance X-ray sources, such as X-ray free-electron lasers and diffraction-limited storage rings, XMPI enables the recording of 3D information of dynamical processes beyond ms and $\mu\mathrm{s}$ temporal resolutions using single X-ray pulses.

However, XMPI's rotation-free advantage comes at the cost of reduced data: the number of acquired projections is much less compared with tomoscopy. The Crowther criterion (Equation 2.27) indicates that a sufficient number of projections are needed for each timestamp to achieve a reliable 4D reconstruction. For example, more than one hundred projections are needed for reconstructing 128×128 images with a spatial resolution of three pixels. However, due to practical constraints, current XMPI configurations record no more than eight projections in a range of less than 90°, constrained by practical limitations [27, 28, 35]. This results in a highly ill-defined sparse-view reconstruction problem, making it extremely challenging to reconstruct 4D from such sparse projections using classic algorithms.

In **Paper V**, we introduce a deep-learning-based sparse-view reconstruction approach for reconstructing 4D from the extremely sparse projections recorded by XMPI.

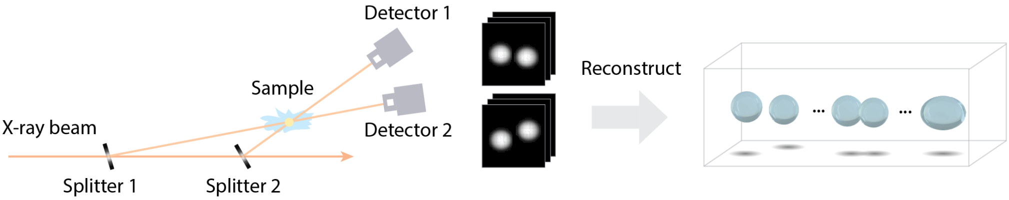


Figure 2.16: Illustration of the MHz XMPI setup performed at the European X-Ray Free-Electron Laser Facility (European XFEL). The collision process of water droplets was recorded from two directions and reconstructed using deep learning methods.

In **Paper VI**, we present the first megahertz (MHz) XMPI experiment performed at the European XFEL. The experimental setup of this experiment is shown in Figure 2.16. The collision process of two water droplets was imaged from two angles simultaneously at a 1.128 MHz frame rate. The incoming X-ray beam was split by two crystal beam splitters, with the diffracted beams illuminating the sample from two directions. Diamond crystals were chosen for this experiment to mitigate potential crystal damage from the intense XFEL pulses. At the end of each beamlet, two MHz detectors were positioned to record projections of the sample. The deep learning approach reported in **Paper V** was employed to reconstruct the collision process from the recorded projections.

2.5 X-ray sources

We have discussed various imaging techniques in this chapter. Many of these techniques would not have been possible without the advancement of high-brilliance X-ray sources. Brilliance quantifies the coherent photon flux of an X-ray light source, as defined by

$$\text{Brilliance} = \frac{\text{photons / second}}{\text{mrad}^2 \cdot \text{mm}^2 \cdot 0.1\% \text{ bandwidth}}. \tag{2.29}$$

As can be seen from Equation 2.29, it calculates the number of photons per unit time per unit photon beam size per unit photon beam divergence, within a specific bandwidth interval. The higher the brilliance, the brighter and more coherent the light source is.

The development of synchrotron radiation sources significantly improves the brilliance of the X-ray sources. More recently, advancements in X-ray free-electron lasers have propelled the brilliance of X-ray sources to unprecedented levels, enabling the generation of "laser-like" X-rays. In this section, we provide a brief background of the development of high-brilliance sources, from synchrotron radiation sources to X-ray free-electron lasers.

2.5.1 Synchrotron radiation sources

Synchrotron radiation is the electromagnetic radiation generated when relativistic charged particles are accelerated in a magnetic field. It was first observed from an electron synchrotron at the General Electric Research Laboratory in New York in 1947 [36]. The early uses of synchrotron radiation were parasitic in electron synchrotrons designed for high-energy physics. High-energy physicists consider synchrotron radiation as a drawback because it consumes energy from the electron synchrotron and thus reduces the energy of the electrons. Over time, scientists gradually realized this radiation can be used as standalone scientific probes without interference with high-energy physics, and dedicated storage rings for synchrotron radiation have been developed since then.

Nowadays, synchrotron radiation facilities are widely used all over the world. They are powerful tools that generate high-intensity X-ray beams, enabling scientists to conduct cutting-edge research in diverse fields. The development of synchrotron radiation sources has gone through four generations. *The first generation* of synchrotron radiation sources can be dated back to the 1960s and early 1970s, when the use of synchrotron radiation was parasitic to storage rings used for conducting electron-positron colliding experiments. In the 1980s, dedicated storage rings designed for synchrotron radiation were built, which were *the second generation* of synchrotron radiation sources [37]. For example, Synchrotron Radiation Source (SRS) in the UK and National Synchrotron Light Sources (NSLS) in the USA. Bending magnets were employed to generate X-rays at the second-generation synchrotron radiation sources, and insertion devices comprising periodic magnetic arrays like wigglers and undulators were developed during this period [38]. The synchrotron radiation facilities designed with the primary use of insertion devices were *the third generation*. Starting from the 1990s, the widespread use of insertion devices has significantly increased the brightness and coherence of synchrotron radiation sources. Examples of the third generation synchrotron radiation sources include European Synchrotron Radiation Facility (ESRF) [39] in France, Advanced Photon Source (APS) [40] in the USA, Swiss Light Source (SLS) [41] in Switzerland, and Super Photon Ring –8 GeV (SPring-8) [42] in Japan. The development of *fourth-generation* synchrotron radiation sources aimed to achieve even higher brightness and coherence of light, primarily through hardware improvements, particularly in multi-bend achromat lattices [43, 44]. MAX IV Laboratory [45] in Lund, Sweden is the first fourth-generation storage ring. ESRF and APS have upgraded to fourth-generation sources. The upgraded facilities are named ESRF–EBS (Extremely Brilliant Source) and APS-U (Upgrade), respectively. SLS is currently undergoing an upgrade toward the fourth generation. Other major fourth-generation synchrotron radiation sources include the Sirius synchrotron light source in Brazil and the High Energy Photon Source (HEPS) in China.

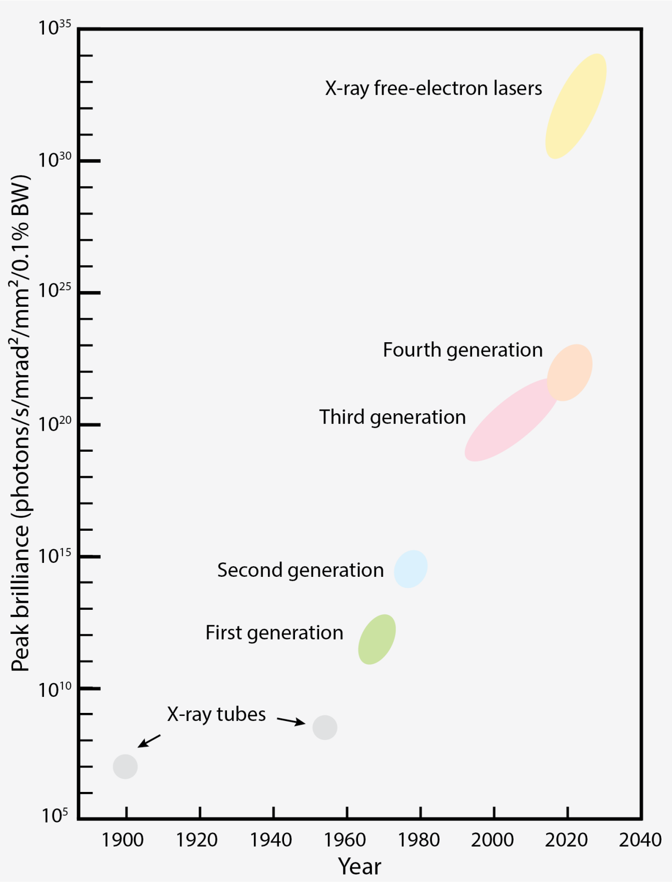


Figure 2.17: The evolution of the peak brilliance of X-ray sources.

2.5.2 X-ray free-electron lasers

The brightness and coherence of synchrotron radiation sources are, to some extent, intrinsically limited by their photon generation mechanism [46]. To further improve the brilliance of X-rays, X-ray free-electron lasers (XFELs) has been developed. The advent of XFELs enables the generation of "laser-like" X-rays with extremely brilliant X-rays that were not possible before [47].

Unlike synchrotron radiation sources, which accelerate electrons in circular accelerators, an XFEL accelerates electrons by a long (typically several kilometers) linear accelerator and then applies periodic magnetic structures (undulators) to generate X-rays from the electron beam. The most common type of XFELs is based on a mechanism called self-amplified spontaneous emission (SASE), where electrons are grouped into micro-bunches, amplifying the generated X-rays as they propagate through the undulator. This process yields short-duration, high-intensity, and coherent X-ray pulses. The peak brilliance of XFELs can reach 10^{30} photons/second/mrad^2/mm^2/0.1% bandwidth. The unique brilliance of XFELs has opened the opportunity for imaging with unprecedented spatial and temporal resolutions [48, 49], and made it possible to perform imaging experiments using single X-ray pulses [50]. European XFEL is the first operational XFEL, which can generate femtosecond duration short pulses at megahertz repetition rate [51], as used in this thesis.

Figure 2.17 shows the evolution of the peak brilliance of X-ray sources over the last century. With the development of the fourth-generation synchrotron radiation sources and XFELs, the brilliance of X-ray light sources has increased by more than 20 orders of magnitude over the past 50 years, providing new scientific opportunities for a variety of fields. The increased photon flux enables imaging with the same quality but much shorter exposure times. This reduction in exposure time enables the study of rapid dynamics in both 2D and 3D, opening up new possibilities for dynamic imaging applications, as presented in **Paper I**, **Paper IV**, **Paper V**, and **Paper VI**. The enhanced coherence greatly progresses the development of phase-contrast imaging methods, as presented in **Paper II** and **Paper III**.

Chapter 3

Fundamentals of deep learning

This chapter introduces the basic concepts and fundamentals of deep learning. We start by giving a general overview of deep learning. Next, we introduce supervised learning and unsupervised learning, two different learning types depending on the requirement of the input data. Then, we introduce three neural network architectures, which are used in the thesis work. In the end, we describe neural implicit representation and deep learning frameworks based on it.

3.1 Overview of deep learning

Deep learning is a subset of machine learning and artificial intelligence, which involves having machines perform tasks based on available data without interference from humans. Classic machine-learning approaches focus on learning the features of the data. However, extracting these features typically involves human input. Machine learning algorithms learn to make predictions based on the selected features, but the selection of features strongly influences the model's performance, leading to feature engineering challenges, i.e., how to extract features from raw data to enhance the performance of models.

Deep learning is a form of machine learning that relies on deep neural networks to learn data representations and features automatically. The neural networks learn by themselves the features of the data and the weights of each feature, allowing them to handle more complex tasks.

Mathematically, deep learning can be expressed as the process of finding the optimal parameters ($\boldsymbol{\theta}$) for a neural network through the optimization of a loss function $\mathcal{L}$ using provided data. For the simplest case of supervised learning, the loss function can be written as

$$\mathcal{L}(\boldsymbol{\theta}) = \mathbb{E}_{(\boldsymbol{x},\boldsymbol{y})\sim p_{\text{data}}} L(f(\boldsymbol{x}, \boldsymbol{\theta}), \boldsymbol{y}) \tag{3.1}$$

where p_{data} denotes the data distribution, $f(\boldsymbol{x}, \boldsymbol{\theta})$ denotes the network predictions with the input $\boldsymbol{x}$, and $\boldsymbol{y}$ is the ground truth or the target output of the model. L specify the loss function of each sample and $\mathbb{E}$ denotes expectation. The choice of loss functions is crucial and usually depends on the specific task.

Once determined, machines can minimize the loss function using optimization methods such as backpropagation and gradient descent [52, 53]. The network optimization process is commonly referred to as *training*. Usually, this involves a large number of operations and is performed on graphics processing units (GPUs) for its parallelism.

The mathematical foundations of deep learning can be traced back to the 1950s [54, 55]. However, its significant influence and widespread use did not occur until the 21st century. The key factors driving this surge in popularity were the availability of powerful hardware and the emergence of massive datasets. Deep learning algorithms are data-driven, requiring large amounts of high-quality training data to achieve optimal performance. In the mid-20th century, computational resources were limited, and data collection was a laborious and expensive process. This constrained the development and application of deep learning models. The beginning of the 21st century was a changing point. The rise of the “big data era” enabled researchers to train deep learning models on massive datasets, which has greatly expedited the progress of deep learning. Projects like ImageNet [56], established in 2009, collected millions of carefully labeled images, allowing researchers to train deep learning models with unprecedented accuracy and generalizability.

3.2 Supervised learning and unsupervised learning

Deep learning approaches can be categorized into supervised learning and unsupervised learning, depending on the difference in input data [52]. In certain scenarios, the training data comes with labels or ground truth, which specifies the target output for each data; whereas in other scenarios, only the data itself is accessible. The former is referred to as supervised learning, while the latter is referred to as unsupervised learning.

The difference between supervised learning and unsupervised learning is visualized in Figure 3.1. As shown in Figure 3.1 **a**, in supervised learning, each data is paired with a corresponding label or a ground truth. There is a one-to-one relationship between the data and the labels. Therefore, supervised learning is also called a paired approach [57]. In unsupervised learning, the training data comes without labels, as shown in Figure 3.1 **b**. The algorithm learns by itself the underlying relationship of the training data, enabling it to accomplish various tasks.

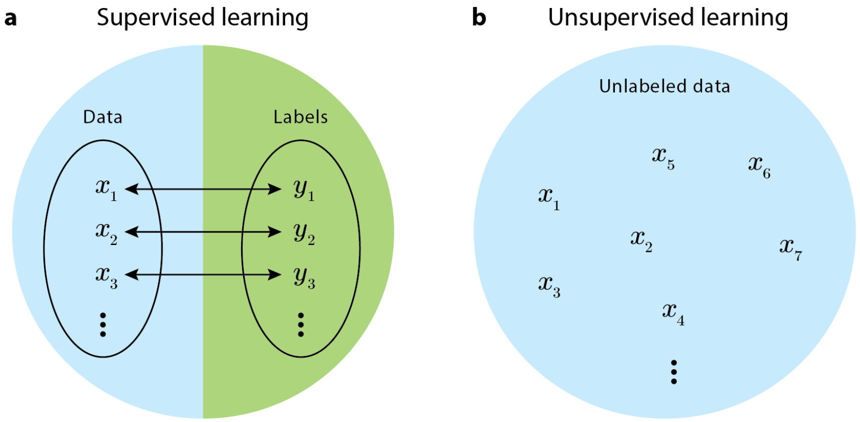


Figure 3.1: Comparison between supervised learning (**a**) and unsupervised learning (**b**).

The data is typically divided into three datasets for supervised training: the training dataset, the validation dataset, and the testing dataset. The model is trained on the training dataset, selected based on the validation data, and in the end, applied to the testing dataset, which contains new data that is never shown to the model. Unsupervised learning approaches do not require data division. The entire dataset can be used for training, and the performance of the model can be evaluated on the same data.

While supervised learning generally achieves higher accuracy, it is constrained by the need for labeled training data. Unsupervised learning is particularly effective in scenarios where labeled training data is unavailable or difficult to acquire, making it a valuable tool for many practical applications.

Depending on the difference in training strategy, unsupervised learning can be further categorized. Numerous unsupervised learning approaches have been developed. Here, we introduce the unpaired approach and the self-supervised approach, which are two common unsupervised learning strategies in image reconstruction tasks. An illustration of the two approaches is shown in Figure 3.2.

Unpaired approaches rely on unpaired training datasets [57]. If we write the distribution of input images as X, and the distribution of target images as Y, unpaired approaches require training data from both of the two distributions. Instead of learning the $\{\boldsymbol{x}_i \leftrightarrow \boldsymbol{y}_i\}$ relationship as in supervised approaches, an unpaired approach learns the translation between the two domains $X \leftrightarrow Y$, without the correspondence of each individual sample. Please note the difference between a paired approach (Figure 3.1 **a**) and an unpaired approach (Figure 3.2 **a**). While the target domain is expressed using the same notation, in supervised learning $\boldsymbol{y}_i$ is the label of the training data, whereas in an unpaired approach, it is part of the training data.

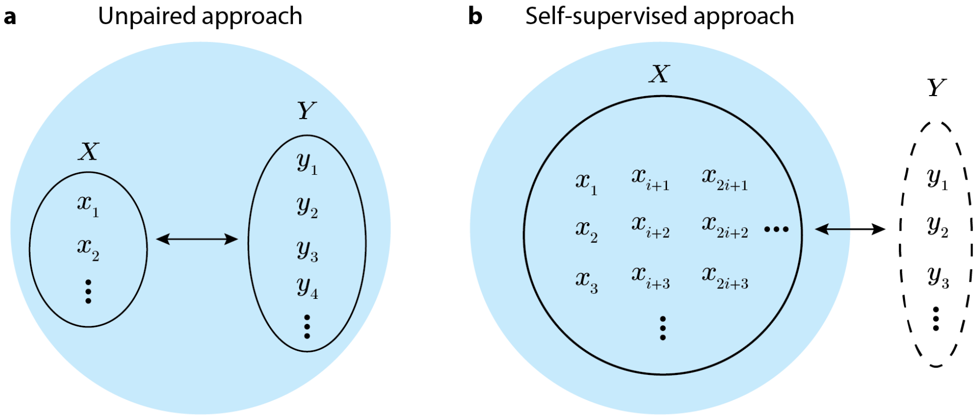


Figure 3.2: Two unsupervised learning approaches: unpaired approach (**a**) and self-supervised approach (**b**). The blue circles mark the range of the training data, as in Figure 3.1.

Self-supervised learning is another type of unsupervised learning. As shown in Figure 3.2 **b**, a self-supervised approach differs from an unpaired approach in that no data from the target distribution is available, and the model learns only from the distribution of the input images. The target can be inferred from the input data. Typically, self-supervised approaches are trained on a part of the input and supervised by another part of the input. For example, if we equally divide the training data into i ($i \in \mathbb{Z}$ and $i \geqslant 2$) batches, during each training iteration, the model can be trained on one batch of data ($x_1, x_{i+1}, x_{2i+1}, ...$) while being supervised by the other $i-1$ batches. With self-supervised learning, networks can be trained to reconstruct missing details in masked images [58] or to reconstruct 3D representations from parts of the 2D views while being supervised by the other 2D projections [59].

In the papers presented, **Paper I** introduces a supervised method that corrects shot noise in detector images by learning the mapping from noisy-clean image pairs. **Paper II** and **Paper III** present unsupervised phase-retrieval approaches using unpaired datasets. **Paper IV** applies self-supervised learning to learn the 3D self-consistency between the projection images measured by X-ray multi-projection imaging. **Paper V** forces self-consistency on all 3D reconstructions in order to learn a 4D representation of dynamical processes.

3.3 Deep neural networks

Deep learning approaches use deep neural networks to extract features and make predictions. In this section, we introduce network architectures that are commonly used for imaging tasks.

3.3.1 Fully connected neural networks

A neural network is constituted of fundamental units called neurons, which are also commonly known as perceptrons. As illustrated in Figure 3.3 **a**, a neuron consists of several inputs and an output. The neuron's input accepts a vector $\boldsymbol{x}$, multiplies it by a weight vector $\boldsymbol{w}$, and then adds a bias term b to it. After that, the result is transferred to the output, which generates a prediction y by applying a non-linear activation function f to the earlier results. This process can be expressed as

$$y = f(\boldsymbol{w} \cdot \boldsymbol{x} + b). \tag{3.2}$$

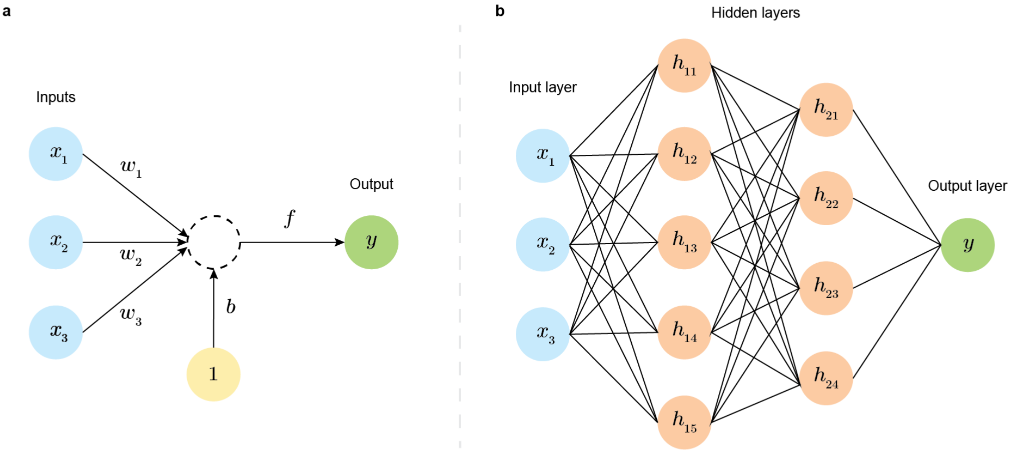


Figure 3.3: Illustration of neural networks. **a** A neuron with three inputs (x_1, x_2, x_3) and a single output (y). The middle dashed circle shows an intermediate state, which multiplies each input by a corresponding weight (w_1, w_2, w_3) and adds a bias term (b) on top of it. The intermediate result is sent into a non-linear activation function f, which produces the final output y. **b** An example of a fully connected neural network with two hidden layers. Similar to the neuron shown in **a**, a fully connected neural network also contains input and output layers, but with extra hidden layers (h) in between. The first hidden layer contains five neurons (h_{11} - h_{15}), while the second hidden layer contains four neurons (h_{21} - h_{24}). For simplicity, the biases and activation functions are omitted.

Figure 3.3 **a** shows a fully connected neural network, also referred to as a multi-layer perceptron [60]. A fully connected neural network differs from a neuron in that it includes hidden layers between the input and the output, where each hidden layer contains a series of neurons. The output takes the results from the last hidden layer and generates a prediction based on the contributions from all neurons.

It is important to note that the activation functions need to be non-linear. As a linear combination of a group of linear functions is still linear, neural networks with linear activation functions are equivalent to a single perceptron. Therefore, having non-linear activation functions is crucial for the network to approximate non-linear functions. As stated in [60], multi-layer perceptrons can be considered as universal approximators, i.e., in principle, they can approximate any continuous functions. Commonly used activation functions are the sigmoid function $f(a) = 1/(1 + e^{-a})$ and the ReLU function $f(a) = \max(0, a)$, as illustrated in Figure 3.4.

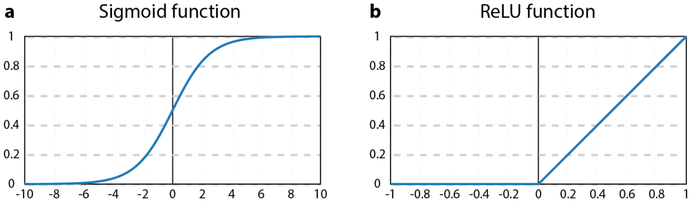


Figure 3.4: Two common activation functions: the sigmoid function (**a**) and the ReLU function (**b**).

In a fully connected neural network, each neuron is connected to all neurons in its adjacent layers, which comes with drawbacks. The number of parameters grows quadratically as the number of neurons increases, potentially resulting in memory issues and training difficulties. A common issue in training is overfitting, i.e., the neural network performs well on the training data but not on the test data. Therefore, fully connected neural networks are not ideal for imaging tasks. As an example, consider a simple five-neuron network for imaging tasks, where the input and output are both images with dimensions of 1000×1000. The images are treated as vectors in a fully connected neural network, so the number of parameters needed to train this network is around $10^6 \times 5 + 5 \times 10^6 = 10^7$ (with the bias neglected). Typical fully connected neural networks contain hundreds to thousands of neurons distributed in multiple hidden layers, which will further magnify the number of trainable parameters needed.

It is intuitive to think that, for imaging tasks, many of these parameters may be redundant. In general, the value of a pixel in an image is more dependent on its neighboring pixels and has less correlation with pixels that are farther away. Fully connected neural networks neglect this local correlation of pixels, potentially leading to increased model complexity without effectively capturing essential features of an image.

Paper IV and **Paper V** employ fully connected neural networks to learn the refractive index of an object from its latent features and encoded coordinates.

3.3.2 Convolutional neural networks

For the above reasons, it is usually more common to use convolutional neural networks [61] for imaging tasks.

As shown in Figure 3.5 **a**, convolutional neural networks consist of a stack of convolutional layers and pooling layers. Convolutional layers include a sequence of weight-sharing convolutional kernels to extract features from the images (Figure 3.5 **b**). Commonly used convolutional kernels are 3×3, 5×5, and 7×7 kernels. The use of these small kernels

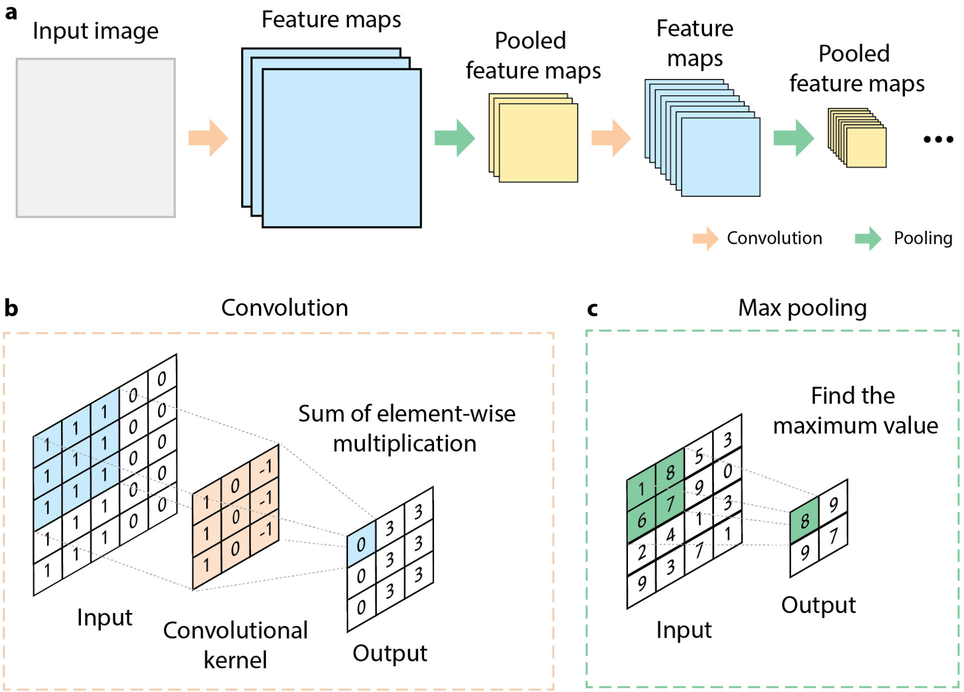


Figure 3.5: Illustration of a convolutional neural network. **a** Layout of a convolutional neural network. It is composed of a series of convolutional layers and pooling layers. The convolutional layers extract feature maps from the input, while the pooling layers reduce the dimension of the extracted feature maps. **b** In the convolution operation, convolutional kernels slide through the input matrix and apply the sum of element-wise multiplication. For simplicity, we only show a single-channel convolutional kernel with the size of 3×3, and the input is not padded. In most scenarios, multi-channel convolutional kernels are used, and the input is padded with zeros in a way that the output has the same dimension as the input. **c** Pooling operation reduces the dimension of the input. Here, we show an example of a max pooling with a 2×2 kernel and a stride of 2.

effectively addresses the memory problems of multi-layer perceptrons, as the size of the convolutional kernels is independent of the dimension of the input. In addition, convolutional kernels allow the extracting of local features of the images [62]. The pooling layers reduce the dimension of the images by pixel binning, usually by taking the maximum or the average (Figure 3.5 **c**). The combination of convolutional layers and pooling layers allows convolutional neural networks to learn multi-scale features from local to global. More importantly, it indicates the translation invariance of convolutional neural networks, i.e., they are capable of recognizing features regardless of the positions. Translation invariance is particularly important for tasks such as feature extraction and image denoising, where features of interest may appear in different parts of an image or undergo shifts due to practical considerations.

Paper I uses a convolutional neural network to learn and correct shot-to-shot noise on the detector images. **Paper II** and **Paper III** both use two convolutional neural networks, one learning to reconstruct the phase and attenuation of the object from the holographic images, while the other learning the experimental effects of the holographic intensity measurements, such as the experimental noise and the point spread function. **Paper IV** and **Paper V** use convolutional neural networks as encoders to extract latent features from 2D projections measured from different directions. These features are then integrated in fully connected neural networks to reconstruct 3D representations of the object. In addition, **Paper I**, **Paper II**,**Paper III**, and **Paper V** also use convolutional neural networks as discriminators, as will be discussed in the next subsection.

3.3.3 Generative adversarial networks

Though multi-layer perceptrons and convolutional neural networks are good at learning the data distribution and extracting features, the creativity for generative tasks of such frameworks is limited [63]. This limitation has been overcome by generative adversarial networks [52, 63].

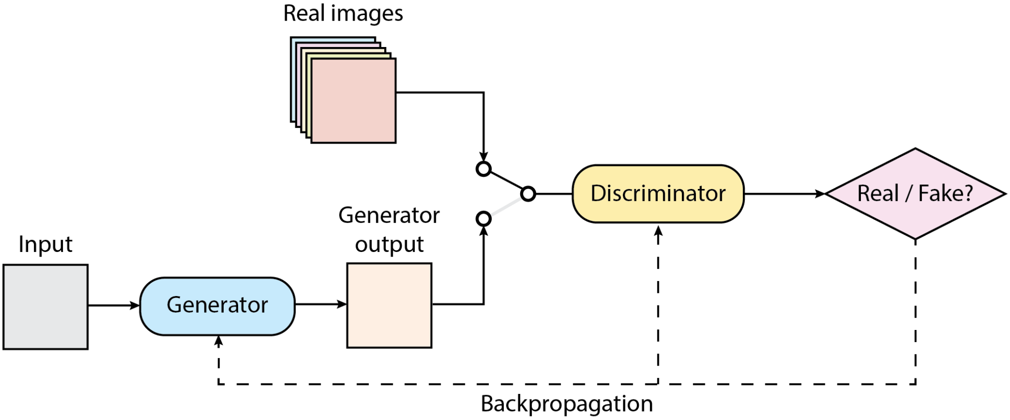


Figure 3.6: Illustration of a generative adversarial network.

Generative adversarial networks enhance the capabilities of single-network approaches by incorporating a second network. As shown in Figure 3.6, a generative adversarial network consists of two neural networks: a generator and a discriminator. These networks can be implemented as multi-layer perceptrons or convolutional neural networks. Like an artist, the generator learns to create images that resemble real images. On the other hand, the discriminator works like a critic. It accesses both real images and the generator outputs and learns to distinguish the generated ones from the real images. Both networks are trained simultaneously, with the generator aiming to generate images that can fool the discriminator, while the discriminator improves its ability to better distinguish between the generated and the real data. In the end, the generator can be trained to produce high-quality images that are nearly indistinguishable to the discriminator.

With the inclusion of the adversarial component, generative adversarial networks allow for learning the probability distribution of target images without the need for ground truth (labels), i.e., training in an unsupervised way. Therefore, it opens up possibilities for not only supervised but also unsupervised learning.

Paper I includes adversarial learning in a supervised way. A discriminator is used to distinguish between noisy detector images and clean denoised images so that the generator can produce better noise-free images. **Paper II** and **Paper III** employ generative adversarial networks in unsupervised phase retrieval tasks to learn the image transformation between detector images and object images. **Paper V** exploits adversarial training to confine the 3D reconstruction of X-ray multi-projection imaging experiment. The objective is to ensure that the generated 2D projections are indistinguishable from the experimental projections.

3.4 Neural implicit representation learning

Typically, an image is treated as a matrix in deep learning, with the matrix elements filled up by the pixel values, as shown in Figure 3.5. When it comes to higher dimensions, it is natural to think of representing 3D (4D) information of an object as a high-dimension matrix, and train high-dimension neural networks to directly reconstruct the matrices, same as in 2D. However, this comes with several problems. The most straightforward is the memory problem. As the dimensionality of the data and the network increases, the size of the matrices and network parameters grows geometrically, causing memory problems for the network training. As a result, training 3D grids with 3D convolutional neural networks is usually performed on low-resolution objects with a size below 128^3. Beyond this limit, excessive GPU memory is required [64]. Additionally, another drawback of the grid-based representation is its limited resolution. Grids cannot represent continuous variations. Therefore, grid-based representation is not always suitable for capturing the fine details required in high-quality representations.

These drawbacks are addressed by neural implicit representation [65, 66, 59]. Just like the shape of an object can be expressed by an implicit function, and the function can be approximated with a neural network. Neural implicit representation implies training neural networks to represent relevant information about an object. For example, the refractive index of a 3D object can be expressed as a continuous function $f : (x, y, z) \rightarrow n$, with (x, y, z) representing the 3D spatial coordinates and n the refractive index of the object at each spatial point. As it is flexible to sample the function at different rates, the memory constraints imposed by grid-based representations can be addressed.

Neural radiance field (NeRF) [59], introduced by Mildenhall et al. in 2021, is one of the popular applications of neural implicit representation. It is designed to reconstruct 3D images from multiple 2D views. In NeRF, a multi-layer perceptron is trained to learn the mapping from the spatial coordinates and viewing directions to the density of the object and view-dependent RGB color of a static scene. After the training, the network learns a continuous representation of the 3D object and is able to produce novel views from any given viewpoint.

The effectiveness of neural implicit representation and NeRF has been widely applied in many fields, including X-ray imaging. The first application of implicit neural representation on X-ray imaging was performed by Sun et al. [67], where a neural network was trained to reconstruct 2D sinograms for sparse-view CT measurement. However, the authors neglected the physics model of X-ray imaging. As NeRF was originally designed for the optical field, where the object is opaque to the light and the color of the object is dependent on the viewpoint, it is not directly applicable to X-rays. As we introduced in Chapter 2, X-rays process much higher penetration power than visible light, and most of the objects are transparent to X-rays. Additionally, the 3D object formed by X-ray imaging is not view-dependent, and the 3D consistency should be maintained regardless of the view angles.

Paper IV applied neural implicit representation for addressing the problem of reconstructing 3D from X-ray multi-projection imaging experiments, where a neural network was trained to represent the mapping from the spatial coordinates to the refractive index of an object $(x, y, z) \rightarrow n$. The physics of X-ray propagation and interaction with matter was also considered. Similarly, **Paper V** exploits neural implicit representation for reconstructing 4D (3D+time) dynamic processes from X-ray multi-projection imaging experiments, where the collision process of water droplets was studied by a neural network with the mapping $(x, y, z; t) \rightarrow n$.

Chapter 4

Advancing X-ray imaging with deep learning

In the previous chapters, we discussed the challenges of X-ray imaging and introduced common deep-learning frameworks addressing imaging tasks. In this chapter, we demonstrate several applications of deep learning as enabling techniques for X-ray imaging, addressing experimental challenges encountered in X-ray image reconstruction.

We start this chapter with 2D image reconstructions and move on to 3D and 4D volumetric dynamical reconstructions. In Section 4.1, we introduce the deep-learning approach presented in **Paper I**. This is a supervised learning approach developed for shot-to-shot flat-field correction at XFELs when conventional approaches do not work, i.e., to mitigate the imaging artifacts caused by non-uniform illumination and detector noise. Section 4.2 introduces unpaired deep-learning approaches for phase retrieval and holographic image reconstruction. We present PhaseGAN, our physics-based unpaired approach, and its comparison to other state-of-the-art unpaired approaches. The results are presented in **Paper II** and **Paper III**. Section 4.3 describes the application of deep learning in 3D and 4D X-ray imaging. We introduce ONIX and 4D-ONIX, two self-supervised learning approaches for reconstructing 3D and 4D information from sparse X-ray projections. The 3D reconstruction results of ONIX are reported in **Paper IV**. The dynamical 3D reconstruction results of 4D-ONIX are presented in **Paper V** and **Paper VI**.

It is worth mentioning that, the deep learning (DL) approaches presented in this chapter do not apply neural networks simply to the experimental data of X-ray imaging. Instead, we include the physics of X-ray imaging into the neural network architecture to improve the performance and explainability of the model. Beyond these benefits, incorporating physics also accelerates the training process of deep learning models. State-of-the-art large models often require millions of training data and heavy neural networks with up to 10^{12} parameters, trained on high-performance GPUs [68]. Such resource-demanding processes are not generally accessible and can cause severe environmental effects [69]. The inclusion of physics in the neural network design enhances the performance and robustness of existing deep learning models, making the training process more sustainable and accessible [70], as evidenced in **Paper II** and **Paper III**.

4.1 FFCGAN: A supervised approach for shot-to-shot flat-field correction at X-ray free-electron lasers

In this section, we describe our deep learning approach for solving the flat-field correction problem of X-ray free-electron lasers. This approach is based on a generative adversarial network (GAN). For brevity, we refer to this approach as FFCGAN.

First, we explain the requirement of flat-field correction for imaging experiments and introduce the conventional method for tackling this problem. Then, we introduce dynamic flat-field correction approaches, which are effective when the conventional approach fails. In the end, we present FFCGAN, our deep-learning approach that can perform flat-field correction at the level of state-of-the-art dynamic approaches and is suitable for real-time image analysis.

The results of this approach are presented in detail in **Paper I**. Please note that **Paper I** introduces two methods, a dynamic flat-field correction approach adapted by K. Buakor (see Section 4.1.2, this approach is originally reported in [71]) and a deep learning approach developed by myself. In this section, the focus will be on my contributions to the deep learning part.

4.1.1 Conventional flat-field correction

Typically, an imaging process involves illuminating an object with a probe light and then recording the signal responses on the detector. In an ideal scenario, uniform illumination and a noise-free detector are desirable, with the signal response of the detector being linear with the intensity of the incoming light. However, this is not always valid. Firstly, for the X-ray imaging experiments at high-brilliant sources, the incoming X-ray illumination (flat-field) is often non-uniform and may contain specific patterns [72]. Secondly, commonly used indirect X-ray detectors convert X-rays into visible light using scintillators for detection, and the sensitivity of the scintillator can be non-uniform. Thirdly, the detector usually contains thermal or electronic noises, leading to non-zero intensities even in the absence of any illumination.

The process of obtaining a noise-free and artifact-free image of the probed object is referred to as flat-field correction. If we write the measured sample images as I_{sample}, flat-field correction normally requires measuring two additional images: the flat-field image without any sample (I_{flat}) and the dark-field image without any illumination (I_{dark}). The flat-field corrected image (I_{FFC}) can then be calculated by

$$I_{\text{FFC}} = \frac{I_{\text{sample}} - I_{\text{dark}}}{I_{\text{flat}} - I_{\text{dark}}}. \tag{4.1}$$

This method, described in Equation 4.1, is the conventional and the most common approach for flat-field correction [73]. It is important to note that as it can be challenging to measure I_{sample}, I_{flat}, and I_{dark} simultaneously, Equation 4.1 is only valid for imaging systems with slow temporal variations, e.g., imaging systems with stable illumination and detector responses.

The conventional flat-field correction approach is not suitable for single-pulse imaging experiments performed at XFELs. As described in Chapter 2, XFELs commonly use self-amplified spontaneous emission (SASE) to produce high-intensity coherent X-ray pulses. This process, however, is intrinsically stochastic due to the random distribution of the electron beam's initial current density. As a result, the generated X-rays exhibit relatively poor temporal coherence and contain spatial and frequency fluctuations [74]. The illumination carries rapid-varying shot-to-shot noise that can not be corrected by the conventional flat-field correction method described in Equation 4.1 as it is not slowly varying.

4.1.2 Dynamic flat-field correction

Dynamic flat-field correction approaches, which estimate the flat-field component of each image from frame to frame, are suitable for correcting the shot noise for rapid-varying illumination [71]. These approaches use principal component analysis (PCA) [75] to determine the eigenvectors associated with shot noise from a set of flat-field images. After selecting the eigenvectors, the contribution of each eigenvector is evaluated and subsequently removed from the image to be corrected.

K. Buakor has implemented a PCA-based dynamic flat-field correction approach [71], which selects the flat-field eigenvectors based on the total variance explained by each element. This approach has been successfully applied to the data collected at MAX IV and European XFEL, and the results are presented in [76] as well as **Paper I**. One drawback of this approach, however, is that it is computationally slow, and thus not suitable for real-time image corrections. Additionally, this approach necessitates a significant number of flat-field images to adequately represent all potential artifact distributions and variations. It is unable to rectify feature shifts, such as drifts in illuminations, due to the locality of the method. From a practical standpoint, acquiring flat-field images before and after each sample measurement is often necessary, thereby reducing the overall acquisition time for each sample.

4.1.3 Deep learning flat-field correction

To address the aforementioned problems, we developed FFCGAN, a deep-learning flat-field correction approach. FFCGAN applies an end-to-end fully-convolutional neural network to learn the image mapping from the noisy detector images to the noise-free flat-field corrected images.

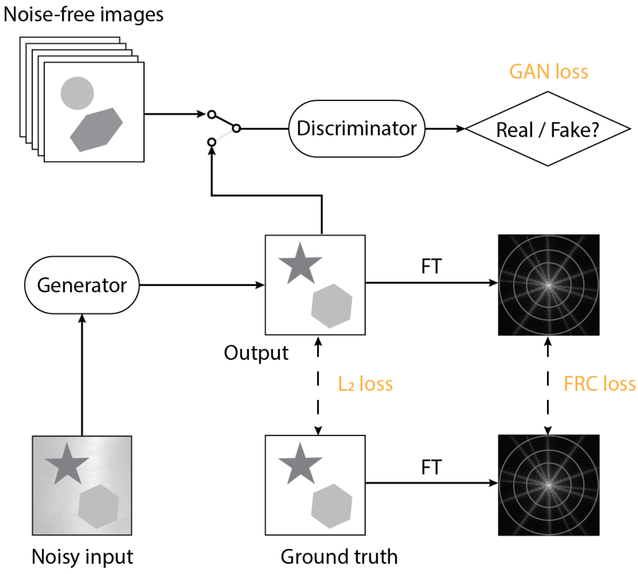


Figure 4.1: Schema of FFCGAN

The architecture of FFCGAN is presented in Figure 4.1. As can be seen, FFCGAN is based on a generative adversarial network, which consists of a generator and a discriminator. Both the generator and the discriminator are convolutional neural networks. The generator of FFCGAN takes the detector image as input and learns to eliminate the shot noise, whereas the discriminator sees both the noise-free images and the generator output to minimize the difference between the two distributions.

The training of FFCGAN is constrained by three losses. The adversarial process between the generator and the discriminator is governed by the GAN loss. Additionally, FFCGAN is supervised by the L_2 loss, minimizing the L_2 distance between the ground truth and the generator output. Aside from the GAN loss and the L_2 loss, we also include a novel loss function in FFCGAN based on Fourier ring correlation (FRC) [77]. FRC loss maximizes the normalized cross-correlation of the two images $\mathbf{x}_1$ and $\mathbf{x}_2$ in the frequency space over rings, as expressed by

$$L_{FRC} = \left\| 1 - \frac{\sum_{r\in R} \mathcal{F}[\mathbf{x}_1](r) \cdot \mathcal{F}[\mathbf{x}_2](r)^*}{\sqrt{\sum_{r\in R} |\mathcal{F}[\mathbf{x}_1](r)|^2 \cdot \sum_{r\in R} |\mathcal{F}[\mathbf{x}_2](r)|^2}} \right\|_2 , \tag{4.2}$$

where R stands for the radius of the ring in Fourier space, and r denotes individual rings. The introduction of constraints in the frequency space has been demonstrated to be effective in mitigating artifacts and improving results [25, 78, 79].

We assess the performance of FFCGAN using both simulated data and experimental data of single-pulse X-ray imaging experiments collected at the European XFEL. The results are reported in **Paper I**. After the training, FFCGAN can perform flat-field correction at a speed of 10^{-4} s at the level of dynamic flat-field correction approaches. In comparison to dynamic flat-field correction approaches, which require seconds of execution time for each image, FFCGAN operates four orders of magnitude faster. This significantly accelerates the process, opening up the possibility for real-time image correction and on-line image analysis during the experiment. In addition, as we mentioned in Section 3.3.2, convolutional neural networks exhibit translation invariance. This property enables FFCGAN to correct artifacts regardless of their locations, which can be crucial for practical applications, e.g., when the beam oscillates from frame to frame as in the SASE process of XFELs.

Although effective, it's important to note that FFCGAN is a supervised approach, necessitating ground truth data for each input during training. For the rest of this chapter, we will focus on unsupervised deep-learning approaches, where ground truth is not needed.

4.2 PhaseGAN: a phase-retrieval approach for unpaired datasets

In this section, we demonstrate PhaseGAN, an unpaired deep-learning phase-retrieval approach. PhaseGAN includes the physics of image formation in the learning process, which further improves the capabilities of unpaired approaches.

First, we recapitulate the phase problem and traditional methods for solving it. Next, we introduce state-of-the-art paired deep-learning approaches addressing the phase problem. Then, we introduce unpaired phase-retrieval approaches and discuss the design of PhaseGAN. A comparison of state-of-the-art unpaired approaches is given in the end.

The results of this section are presented in **Paper II** and **Paper III**.

4.2.1 Classic approach for the phase problem

As discussed in Chapter 2, the phase problem, i.e., retrieving samples' original phase from the intensity measurement, is a fundamental problem in coherent imaging experiments.

Conventional methods addressing the phase problem rely on either analytical or iterative solutions. Analytical approaches have limited application scenarios, because they are based on mathematical derivations calculated under specific assumptions about the characteristics of the light source, sample, and propagation distance. Iterative solutions are not limited by application scenarios. They map between the object and detector spaces, applying constraints in both spaces (see Figure 4.2a). However, they come with the drawback of extended computational time, and the convergence is not guaranteed.

4.2.2 Paired deep-learning phase-retrieval approaches

In recent years, deep learning methods have also been applied to solve the phase problem. In 2017, A. Sinha et al. [80] first reported the use of a deep neural network to solve the phase problem in computational imaging. They trained a convolutional neural network to reconstruct phase from the diffraction patterns. The phase object was generated by a spatial light modulator, using the images from ImageNet [56]. Ten thousand image-diffraction pattern pairs were used for the training, and the trained network was applied to an unseen test dataset for evaluation. The first deep learning for phase retrieval on real experimental samples was reported by Y. Rivenson et al. [24] in 2018, where sample-specific deep neural networks were trained to perform phase recovery on biological tissue samples measured by holography experiments. For each sample, 100 pairs of image-diffraction patterns were used in the training process. Although these approaches demonstrate the effectiveness of deep learning methods on the phase problem, they are all paired approaches, i.e., they learn from paired training datasets.

As discussed, paired approaches rely on a large amount of paired training data, i.e., for each diffraction intensity pattern, a corresponding phase image is needed. The phase image can be obtained in one of the following two ways. First, conventional phase retrieval algorithms are suitable when the quality of the diffraction intensity is good enough (e.g., low noise, high spatial resolution), and the experimental conditions meet the required assumptions. However, the effectiveness of the deep learning model in this case is constrained by the quality of the phase retrieval methods, which may be prone to artifacts. Second, if the hologram image cannot be accurately phase-reconstructed using conventional phase retrieval algorithms, but the image quality can be enhanced through experimental condition improvements (e.g., increased exposure time, selection of a higher resolution detector, etc.), then additional experiments can be conducted on a small dataset to obtain enhanced images suitable for phase retrieval. These images can then be phase-retrieved and paired with the holographic images to form a training dataset. After the training, the model can be applied to other holographic images for phase reconstruction.

However, acquiring many high-quality holographic images can be challenging. For example, images collected at XFELs often suffer from shot-to-shot noise that can hinder the performance of phase reconstructions. Additionally, when imaging sub-microsecond dynamical processes, the short exposure time may lead to low spatial resolution and high noise that pose difficulties in the reconstruction process. If the amount of training data is inadequate, or the phase reconstructions provided are inaccurate, the deep learning model will not be able to get the correct reconstructions.

4.2.3 PhaseGAN: a physics-based unpaired phase-retrieval approach

In cases where the paired approach is not applicable, an unpaired approach may be considered as an alternative. As introduced in Section 3.2, unpaired approaches use unpaired training datasets. In unpaired datasets, the one-to-one match between the phase and intensity images is not required. The diffraction intensity image and the phase image can be obtained through different imaging modalities using similar but not exactly identical samples. Therefore, the use of unpaired datasets allows for phase reconstruction in the case where prior phase reconstructions are not available.

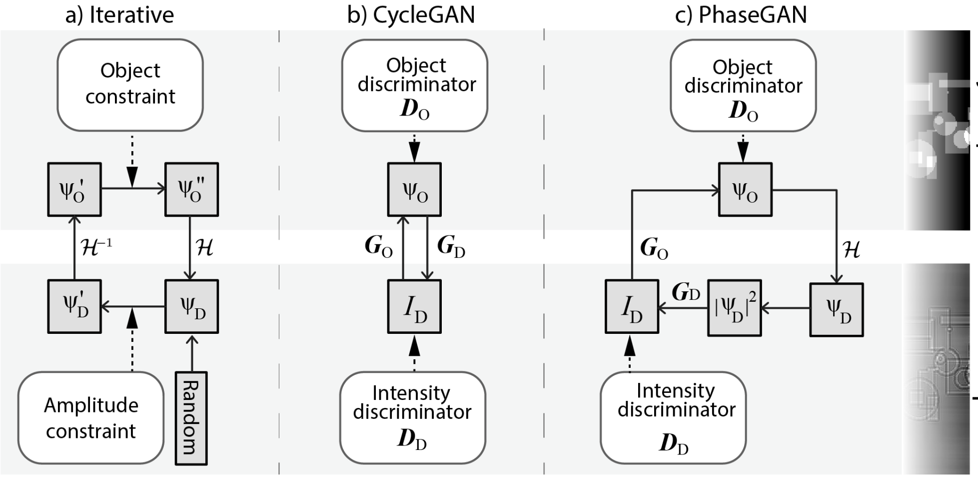


Figure 4.2: Comparison of phase retrieval methods: (a) iterative approach, (b) CycleGAN, and (c) PhaseGAN.

PhaseGAN is an unpaired deep learning phase reconstruction approach. It is based on CycleGAN [57]. As shown in Figure 4.2 **b**, CycleGAN is a standard unpaired deep learning approach that allows learning the image transformation between two domains by using two generator neural networks: the object generator G_O and the detector generator G_D. Two discriminator neural networks (D_O,D_D) are employed to apply constraints in both spaces.

In PhaseGAN, we include a physics propagator H, as shown in Figure 4.2 **c**, similar to the one used in the iterative approach. By including the propagator, the detector generator G_D does not need to learn the physical process of image formation, and it needs to learn experimental effects such as the response function of the optical system and noises. Aside from including the physics propagator, we also include the Fourier ring correlation loss in PhaseGAN to better constrain the model in the frequency space, as explained in Section 4.1.3.

We validated PhaseGAN using simulated holographic images based on the CelabA dataset [81] and compared the results with state-of-the-art paired approaches [82] and CycleGAN. The reconstruction results are presented in **Paper II**. PhaseGAN outperformed CycleGAN and reconstructed results at the level of state-of-the-art paired approaches. By comparing the results of PhaseGAN trained with and without the Fourier ring correlation loss, we also proved the effectiveness of adding constraints in the frequency space.

We also applied PhaseGAN to time-resolved imaging experiments using single pulses of the Advanced Photon Source (APS), capturing the cell wall rupture of metallic foams. As shown in **Paper II**, the measured holographic images are very noisy and thus can not be phase reconstructed using conventional approaches. We formed an unpaired dataset by combining the single-pulse noisy images with phase images reconstructed from a different imaging modality, which were captured using a high-resolution camera integrated over 31 pulses. PhaseGAN successfully learned the image reconstruction from the unpaired training dataset, and was able to perform phase reconstructions for the time-resolved imaging experiments where paired approaches were not applicable.

4.2.4 A comparison: Unpaired deep-learning approaches for holographic image reconstruction

In recent years, several unpaired phase-retrieval approaches have been developed for holographic image reconstruction, with a wide range of applications [83, 84]. Biomedical imaging, in particular, poses challenges where collecting paired datasets can be difficult or impractical due to various considerations. A recent study by C. Lee et al. [70] has introduced a parameterized physical model for adaptive holographic imaging. This model, referred to as FMGAN, replaces the forward generator with a physical model that is parameterized by the propagation distance of the probing light. Unlike PhaseGAN, which requires knowing the accurate physical model, FMGAN can learn to determine the physical model from the training data. FMGAN can predict not only the phase and attenuation images but also the propagation distance of the probing light.

In **Paper III**, we evaluated the reusability and generalizability of FMGAN. Furthermore, we compared the reconstruction results of three unpaired approaches—CycleGAN, PhaseGAN, and FMGAN [70]—using polystyrene microsphere samples measured by optical holography [85], and investigated their performance under non-perfect optical systems. The results show that both PhaseGAN and FMGAN are capable of reconstructing the phase and absorption of the polystyrene microsphere samples from their holographic images, while CycleGAN fails. FMGAN outperforms PhaseGAN in both accuracy and generalizability. However, FMGAN encounters difficulties when dealing with blurring non-perfect optical systems, while PhaseGAN succeeds because the detector generator allows it to learn the response of the optical system.

4.3 ONIX and 4D-ONIX: Self-supervised approaches for reconstructing (dynamical) 3D from sparse X-ray projections

In this section, we shift our focus from 2D to 3D and 4D reconstructions. We introduce ONIX and 4D-ONIX, two deep-learning approaches for reconstructing 3D and 4D from sparse X-ray projections.

ONIX and 4D-ONIX are developed mainly for solving the 3D reconstruction problems in X-ray multi-projection imaging (XMPI). As discussed in Section 2.4.3, XMPI captures fast dynamics of the sample simultaneously from different angles, but reconstructing dynamical 3D information from the sparsely recorded projections is extremely difficult. This section first discusses the reconstruction challenges in XMPI. Then, we introduce ONIX, a self-supervised sparse-view reconstruction approach capable of reconstructing 3D from fewer than ten projections. Last, we introduce 4D-ONIX, a self-supervised 4D reconstruction approach based on ONIX, and demonstrate its reconstruction results on experimental XMPI data. By integrating adversarial learning, 4D-ONIX further enhances the capability of ONIX, enabling reconstruction with as few as two projections separated by approximately 24°.

The results of ONIX are presented in **Paper IV**. The results of 4D-ONIX are presented in **Paper V** and **Paper VI**.

4.3.1 Reconstruction challenges for X-ray multi-projection imaging

As discussed in Section 2.4.3, X-ray multi-projection imaging is a rotation-free single-pulse imaging technique that captures multiple projections of a sample simultaneously from different angles. When combined with high-brilliance X-ray sources such as the fourth-generation synchrotron radiation sources and XFELs, it can record volumetric information of a sample or dynamical processes at kHz up to MHz rate. However, the current XMPI design allows for recording just a small and limited set of sparse projections [28, 35]. Therefore, there is a need for a reconstruction method that can reconstruct a 3D representation from these sparsely recorded XMPI projections.

Deep learning approaches have shown convincing results in 2D and 3D image reconstructions [86]. Specifically, as discussed in Section 3.4, deep learning approaches based on neural implicit representation, such as NeRF, have shown their capability to learn a continuous implicit 3D representation of objects from 2D views [59, 87, 88]. Unlike voxel-based 3D reconstruction approaches, where the quality of the reconstruction is constrained by the memory limitations of available hardware, NeRF-based approaches can go beyond the memory limit and reconstruct high-resolution 3D representations without the need to train on slices [89, 90] or cropped patches [91, 92]. Nevertheless, these approaches are designed for visible light imaging, where most of the objects are opaque to the light. By applying the principles of X-ray propagation and interaction with matter as outlined in Chapter 2, it is possible to extend these methods to reconstruct objects imaged using X-rays. ONIX is specifically designed for this purpose.

4.3.2 Optimized neural implicit X-ray imaging (ONIX)

ONIX stands for Optimized Neural Implicit X-ray imaging. It employs neural implicit representation to learn the mapping from the 3D coordinate to the complex refractive index of an object: $(\mathbf{x}) \mapsto n(\delta, \beta)$.

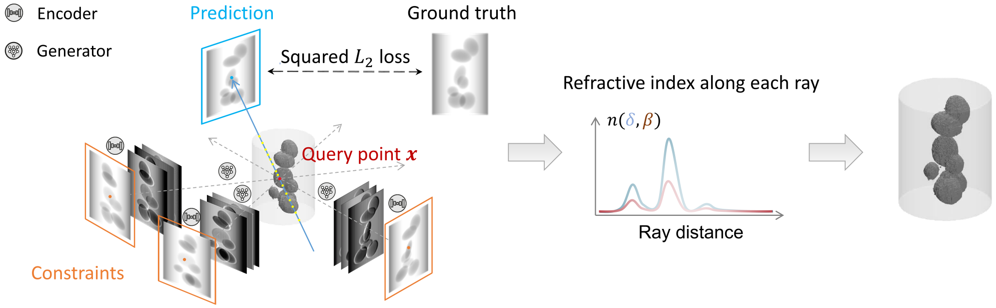


Figure 4.3: Overview of the ONIX learning process. By forcing the self-consistency of the imaging object over recorded projections, ONIX learns the complex refractive index at each spatial coordinate and reconstructs a comprehensive 3D representation of the object.

ONIX is composed of convolutional neural networks and fully connected neural networks, as illustrated in Figure 2 of **Paper IV** and Figure 4.3. The convolutional neural networks serve as encoders, extracting features from the recorded projections and learning common features among the measured samples. The fully-connected neural networks take the input from the encoders and generate the refractive index at each point.

To understand the learning process of ONIX, let us consider a hypothetical scenario where four projections are available for each XMPI measurement, as illustrated in Figure 4.3. We randomly select three of them as constraints. For each query point in the 3D space, we project the corresponding pixel in the camera space for each of the three constraints and collect the pixel-aligned latent features extracted by the encoder. These extracted vectors are then fed into the fully connected neural network, together with the positional-encoded spatial coordinates of the query point. Positional encoding is a technique commonly used in NeRF-based 3D reconstruction approaches. It maps the spatial coordinates to a high-dimensional space, enhancing the model's performance on learning high-frequency details [59].

With a sufficient number of query points sampled within the object, we integrate along rays to form pixels in the projection plane, resulting in a prediction image. The integration is based on the projection approximation. Nonetheless, it can be extended beyond this by using multi-slice methods. To train ONIX, we use the fourth projection. By minimizing the squared L_2 loss between the recorded projection and the prediction, ONIX learns the self-consistency of the projections in 3D space without using any 3D information.

Although only four projections are shown in Figure 4.3 for simplicity, in reality, we use a minimum of eight projections and four of them as constraints. It is important to note that ONIX learns from all of the measurements, not just the projections of a single sample. It learns common features of the samples and generalizes across similar samples and experiments.

We validated ONIX using a simulated dataset of ellipsoids and an experimental dataset of metallic foams. For each object in both datasets, eight projections in a range of less than 140° were prepared for the 3D reconstruction. The reconstruction results were compared with three other approaches: i) simultaneous algebraic reconstruction technique (SART), a classic sparse-view reconstruction approach, ii) 3D supervised method, a supervised deep-learning approach trained on stacks of individual 2D slices, and iii) Noise2Inverse, a self-supervised deep-learning approach [93]. The results are presented in detail in **Paper IV**. As can be seen, ONIX successfully reconstructed the 3D details of the objects in both datasets, using only eight projections for each object. The reconstruction results of ONIX outperform SART and Noise2Inverse and are comparable to the 3D supervised approach.

4.3.3 4D-ONIX: Reconstructing 4D for X-ray multi-projection imaging

Although ONIX has demonstrated its ability to reconstruct 3D from sparse X-ray projections, acquiring eight projections simultaneously from XMPI, as required by ONIX, is extremely cumbersome. The metallic foam data presented in **Paper IV** was extracted from a tomographic dataset rather than an XMPI one, i.e., we selected eight projections from a total of 96 projections for evaluating the performance of ONIX.

During the development of ONIX, the experimental setup of XMPI has been successfully deployed at the European XFEL, achieving acquisition rates beyond MHz. While not yet capable of capturing eight projections simultaneously, the XMPI setup allows for the simultaneous capture of two or more projections using a single X-ray pulse.

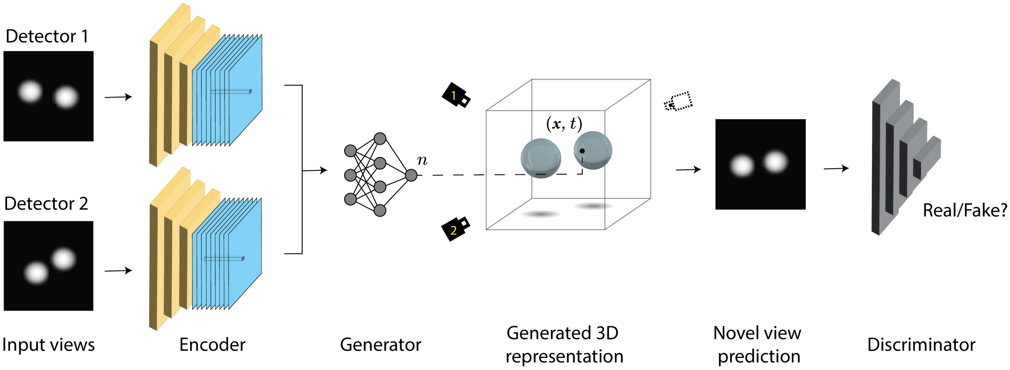


Figure 4.4: Overview of the 4D-ONIX learning process.

In this subsection, we present 4D-ONIX, a self-supervised deep-learning approach for reconstructing 3D movies from the ultra-sparse projections measured with XMPI. As shown in Figure 4.4, 4D-ONIX is based on ONIX, with two main differences. First, the time dimension is included as a fourth dimension of the input of the model. With time included, the goal is to learn the mapping from 4D spatial-temporal coordinates to the refractive index: $(\mathbf{x}, t) \mapsto n$. Second, we include adversarial learning in 4D-ONIX. As reported in **Paper VI**, the first MHz XMPI experiment at European XFEL allows for the simultaneous measurement of only two projections. Due to the limited number of projections from different angles, assessing the self-consistency of the 3D sample using only squared L_2 loss, as in ONIX, is unrealistic. Instead, 4D-ONIX introduces a discriminator neural network and trains with both the squared L_2 loss and an additional adversarial loss. The discriminator learns to distinguish between the predicted projections and the measurements, thereby compelling the reconstructions to adhere to the distribution of the measured projections of all the experiments from any given angle on the experimental plane.

We validated 4D-ONIX using simulated data of water droplet collision, with detailed results reported in **Paper V**. Our evaluation of 4D-ONIX includes two scenarios: reproducible processes and quasi-reproducible processes. For reproducible processes, the dynamical process can be measured multiple times with different sample orientations, but only two projections per experiment. However, achieving reproducibility in many cases poses experimental challenges. Therefore, it is more practical to measure multiple similar dynamical processes within the experimental tolerance, and each process being measured only once. We refer to this as the scenario of quasi-reproducible processes. We simulated XMPI experiments for both scenarios, mimicking the experimental setup at European XFEL, and assessed the reconstruction results of 4D-ONIX. As detailed in **Paper V**, our results demonstrate that 4D-ONIX effectively reconstructs the dynamics of the droplet collision process for both scenarios, with slightly better results observed for reproducible processes compared to quasi-reproducible ones. The reconstruction results from 16 experiments provide a 3D spatial resolution of 4 ± 1 voxels for reproducible processes and 6 ± 1 voxels for quasi-reproducible processes. The voxel dimension corresponds to the pixel dimension of the input projections. These validation results indicate the potential of 4D-ONIX to achieve dynamical 3D reconstructions from as few as two projections. With additional projections or experiments, the results can be further improved. When combined with XMPI, this opens up possibilities to study ultrafast dynamical processes at a rate that was not possible before.

In **Paper VI**, we applied 4D-ONIX to experimental data of water droplet collision collected at the European XFEL. Here, we present an additional analysis of the experimental reconstruction results.

4.3.4 4D-ONIX results on experimental data

As described in **Paper VI**, we applied 4D-ONIX on experimental data collected at European XFEL, where collisions of water droplets were recorded at 10 keV in 1.128 MHz frame rate using XMPI with two split beamlets. The relative angle between the two beamlets was 23.8°. Unfortunately, only two sequences of the droplet collision were captured, with a minimal shift in sample orientation. Each sequence contains 127 timestamps, corresponding to $\sim 113\,\mu\text{s}$ recording time. We trained 4D-ONIX using the two sequences of data collected, and present the reconstruction results for one of the droplet sequences in Figure 4.5.

Figure 4.5(a) shows the preprocessed detector frames and reconstruction results at different timestamps, demonstrating different stages of the water droplet collision. Figure 4.5(b) shows the distribution of the center of mass along the direction of collision (x-axis) for each droplet and for the whole system over time. Before the collision, the two droplets moved towards each other. The motion of the two individual droplets is estimated with a linear fit, assuming that each droplet moved with a constant velocity. The quality of the fit is evaluated by the coefficient of determination R^2, ranging from 0 (no fit) to 1 (perfect fit). The fitting result for the movement of Droplet_1 is $x_1 = (1.51 \pm 0.04) \cdot t + (100.8 \pm 0.6)$, where x_1 is the position of the center of mass of Droplet_1 along the collision axis and t is time. The coefficient of determination for the fitting is $R^2 = 0.98$. The fitting result for the movement of the center of mass of Droplet_2 is $x_2 = (-1.10 \pm 0.05) \cdot t + (297.6 \pm 0.8)$, with $R^2 = 0.94$. These fitting results indicate the velocity and position of the two droplets. The two fitted lines collide at the time of $75\,\mu\text{s}$. The collision point of the two fitted lines was marked by a red star, indicating the completion of the droplet collision. Figure 4.5(c) shows the center of mass of the droplets as a function of time for the y-axis and z-axis,

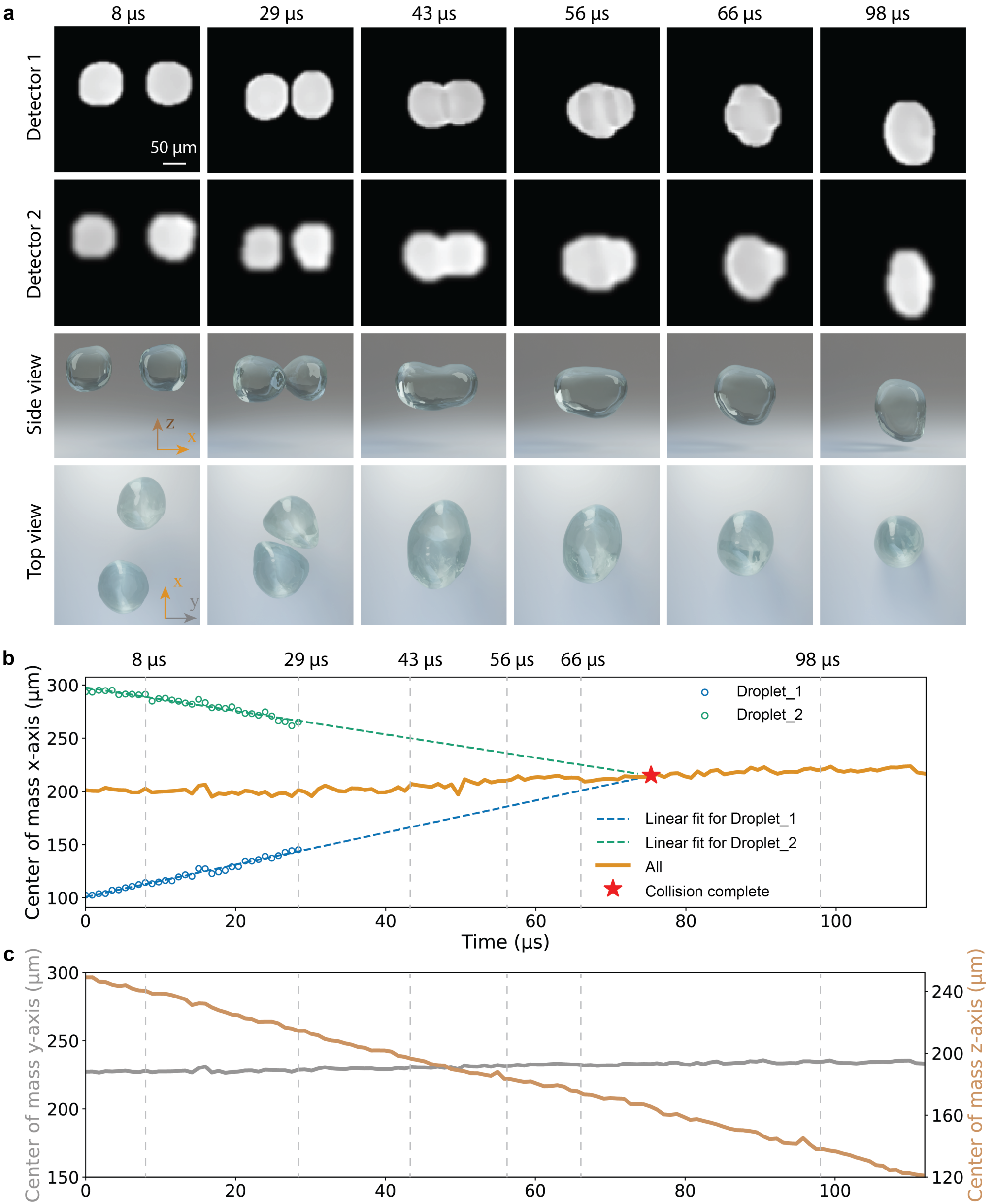


Figure 4.5: Collision of the two water droplets as measured by XMPI. **a** The preprocessed projections obtained from the two detectors and the 3D rendering of the reconstructions at different stages of the collision. **b,c** Movement of the center of mass over time for a period of $\sim 120\,\mu\mathrm{s}$, shown for the x-axis (**b**), y-axis and z-axis (**c**), respectively. The movement of the center of mass of each individual droplet and the collision point are also shown in **b**.

respectively, where the z-axis is the axis of gravity and the y-axis is perpendicular to the collision axis and gravity.

As can be seen from Figure 4.5, prior to the collision, the two droplets could be observed moving toward each other at a consistent speed. They started at a distance of 197 μm apart from each other. The velocity of the two droplets were derived to be 1.5 m/s and −1.1 m/s, respectively, and the relative velocity between the droplets were 2.6 m/s. After a period of 30 μs, the droplets coalesced into a single droplet and became inseparable. The convergence point of each droplet's center of mass, denoted by the red star in Figure 4.5(b), lies on the trajectory of the combined center of mass, marking the completion of the collision between the two droplets. The collision process lasted for around 45 μs. As shown in Figure 4.5(c), the overall center of mass for both droplets remains relatively steady along the y-axis, while the two droplets descend along the z-axis due to gravity.

The reconstruction results include some level of error. First, as depicted in Figure 4.5(a), the recorded projections contain some degree of noise and artifacts, which may affect the accuracy and consistency of the reconstructions. The imperfections are particularly noticeable for the second detector. The signal-to-noise ratio for detector 2 is significantly inferior to that of detector 1, largely attributed to the second beamlet being generated from a higher order of diffraction, which leads to decreased diffraction efficiency. Second, only two experiments of the collision process were collected, which may not be sufficient for the reconstruction. As evidenced in **Paper V**, the performance of the model may be compromised when trained with only two experiments. Increasing the number of experiments for identical or similar samples is crucial to improve both convergence and the accuracy of the reconstructions.

Chapter 5

Conclusion and outlook

5.1 Conclusion

In this thesis, we have presented different X-ray imaging methods and discussed the challenges in the image reconstruction process. In addition, we have provided an overview of deep-learning models relevant to imaging tasks. Specially, we presented how deep learning can be used as an enabling tool to address the challenges in X-ray imaging. We introduced four deep-learning-based image reconstruction algorithms, and demonstrated their performance on both simulated and experimental X-ray imaging data.

Firstly, we introduced FFCGAN, a flat-field correction approach designed to mitigate imaging artifacts for X-ray free-electron lasers. It provides comparable performance to dynamic flat-field correction methods, but with significantly faster reconstruction speeds, enabling real-time image correction. In addition, the translation-invariant property of FFCGAN allows it to correct for non-local artifacts, such as systematic drifts in the illumination, where conventional approaches fail.

Secondly, we present PhaseGAN, a phase-retrieval approach for unpaired datasets. PhaseGAN enables high-quality holographic image reconstructions without requiring paired datasets, making it a valuable algorithm for scenarios where conventional phase reconstruction is difficult or unavailable. Examples of these scenarios include single-pulse imaging, where conventional methods encounter difficulties due to low image quality, and biological imaging, where acquiring paired datasets poses challenges. Additionally, we compared the performance of state-of-the-art unpaired image reconstruction approaches and evaluated their efficacy for imperfect optical systems.

Lastly, we presented ONIX and 4D-ONIX, two self-supervised deep-learning approaches for sparse-view reconstructions. ONIX enables the reconstruction of 3D information from fewer than ten projections, while 4D-ONIX extends this capability to reconstruct 4D from as few as two projections. We envision these approaches as enabling computational tools for dynamical 3D analysis, offering new spatiotemporal resolutions to study ultrafast dynamical processes at a rate that was not possible before. Specifically, these approaches offer the potential to record 3D movies beyond kHz rate when combined with X-ray multi-projection imaging and exploit the unique capabilities of high brilliance X-ray sources such as X-ray free-electron lasers and synchrotron radiation facilities.

5.2 Outlook

Looking to the future, there are still lots of possibilities to improve the efficiency and accuracy of the models. As we mentioned in Chapter 3, deep learning approaches rely on three fundamental components: training data, neural networks, and loss functions. From a physical standpoint, physical principles and known physical prior can serve as a fourth pillar. As presented in Chapter 4, including known physical principles in the learning process improves the performance, robustness, and interpretability of the deep learning approaches. Future advancements in deep learning can be achieved by enhancing these aspects.

Firstly, the quality and quantity of training data greatly influence the performance of deep-learning models. Acquiring better quality and a larger quantity of experimental data can enhance the performance of the model. Besides, building extensive databases for X-ray imaging would facilitate the development of models, enabling techniques like transfer learning, which involves pretraining on large datasets and fine-tuning on specific data. This approach has already shown promise in medical X-ray imaging for patient diagnosis [94] and allows for 3D reconstruction from as few as single projection images [95, 96]. Applying it to general microscopy faces challenges due to the diversity and complexity of imaging samples. Nevertheless, it can serve as valuable prior knowledge, especially for 4D imaging of rapid dynamics, where data availability is limited.

Secondly, from the architecture perspective, incorporating advanced deep architectures would also enhance the model's performance. DL is a rapidly evolving field, with new network architectures rapidly replacing old ones. Staying updated with these developments and incorporating state-of-the-art network architectures can enhance the effectiveness and adaptability of deep learning models. For example, attention-based transformer models and diffusion models have demonstrated improved performance on imaging tasks, including noise reduction and image reconstruction [97, 98, 99].

Thirdly, investigating a variety of loss functions and their combinations has the potential to further enhance the performance of the model. For instance, integrating Fourier domain constraints can help the model to better capture image details in the frequency space. Including perceptual loss functions can potentially improve the visual quality of the reconstructions [100, 101].

Finally, physics can serve as a crucial enhancement to the model. If the imaging process adheres to well-known physical principles, integrating this knowledge directly into the network architecture can significantly improve the model's performance and robustness, as demonstrated in Section 4.2. On the other hand, physics can also be integrated into the loss function to better condition the deep learning model. For example, if the sample dynamics are governed by a partial differential equation, incorporating it into the loss function can enhance training efficiency and overall results [102, 103].

In the present work, our primary objective is to address real experimental challenges encountered in X-ray imaging with deep learning and apply physical principles to improve the performance of the models. While enhancing the first three aspects will certainly improve model effectiveness, from a physics perspective, the incorporation of physical principles into learning algorithms, as a fourth pillar, is indispensable to making deep learning methods more robust, interpretable, and sustainable [69, 104].

Acknowledgements

The work presented in this thesis would not have been possible without the support and guidance of many people.

First and foremost, my deepest gratitude goes to my supervisor, Pablo Villanueva-Pérez. Thank you for offering me the opportunity to work on the fascinating topic of X-ray imaging and providing me with the best guidance and patience that I could have hoped for. Your dedication to research has influenced me and helped me become the researcher I am today. I would also like to express my sincere gratitude to Tobias Ritschel for guiding me along the path of learning computer vision and deep learning since the beginning of my PhD journey. Discussions with you have always been inspiring and enlightening. Thanks also to my co-supervisor, Jesper Wallentin, for teaching me the basics of X-ray imaging and for always providing helpful suggestions and being patient with my questions.

The papers presented in this thesis are a product of collaboration among many different teams. I want to thank the team of Robert Klöfkorn for providing droplet collision simulations, the team of Patrik Vagovic for collaborating on multiple experiments, and the team of Francisco Garcia-Moreno for providing samples and sharing datasets of metallic foams. Thank you, Yentl Swolfs, Christian Breite, and Rui Guo, for explaining the dynamics of different fiber materials. Thank you, Jiawei Mi, Kang Xiang, and Shi Huang, for the tour of the chemistry lab and interesting discussions regarding ultrasound cavitation. Thanks also to Emanuel Larsson, Till Dreier, Nathaly De La Rosa, Martin Bech, Max Langer, Kannara Mom, and Simone Sala for shared beamtimes and collaboration in different projects. I would also like to thank the researchers and beamline scientists at MAX IV, PSI, and European XFEL for supporting the experiments I participated in during my PhD. Special thanks go to Zdenek Matej and Luca Gelisio for their constant support and for providing access to the GPU clusters at MAX IV and European XFEL.

It has been a great pleasure to work at Sljus. For this, I would like to thank my current and previous colleagues. First, I would like to thank everyone from the X-ray imaging group, Myrto, Zisheng, Julia, Zhe, and Runqing, for the outstanding experimental work on the development of the XMPI setup, which laid the groundwork for the algorithms presented

in this thesis, and for the valuable discussions we had together with cakes and fika. The same gratitude goes to our previous group members, Khachiwan, Johannes, and Maria. Thank you for being the first to test my code and improve it. Thank you to my office mates, Smita and Nelia, for making my PhD journey so lovely and joyful. Thank you, Yi, Huaiyu, Zhaojun, Lingjia, Ziyun, Fatemeh, Zhihua, Foqia, Annika, Hanna, Rohit, Vidar, and Klara, for the great time we have shared and for always being there to help when I needed it. Thank you, Edvin, for your guidance during the ISP meeting. Thank you, Patrik, for answering all my naive non-academic questions.

Thank you, Di Lu, Roger Hutton, and Tomas Brage, for introducing me to Lund and offering the opportunity to come as an exchange student when I was just an undergraduate. Thank you, Zhongshan Li and Wubin Weng, for supervising my master's project and teaching me the basics of spectroscopy. Thank you, Märta Lewander Xu, for being my mentor and guiding me in achieving my goals.

On a more personal level, I would like to thank the people who have brought joy to my life outside of Ph.D. work. Thank you, Jingwei and Weihua, for housing me at the beginning of my PhD and always treating me like family. Thank you, Xin, Yupan, Qingshuang, Yingzhe, and Ruiyu, for helping me so much when you did not have to. Thank you, Andrea, Hannah, Shivani, Lisa, Anki, and Swati, for the beers and jokes. Thank you to everyone in the Friday badminton group for having junk food together after working out. Thanks to "Lund Five Overlords" for the trips and board games. Thank you, Qi, for opening the door to an amazing world and guiding me every step of the way. Thank you, Eleven (the cat), for being a constant source of inspiration and for teaching me valuable life lessons. Thanks also to my childhood friends Ahuang, Helun, and Fanfan, whose support has never been absent, even though we have always been in different places around the world.

Last, and most importantly, I want to express my sincere gratitude to my parents. Mom and dad, thank you for all your love and support. You will always be my best friends and teachers.

References

[1] Eugene Hecht. Optics, global edition. *Harlow: Pearson Education Limited*, 2017.

[2] Burton L Henke, Eric M Gullikson, and John C Davis. X-ray interactions: photoabsorption, scattering, transmission, and reflection at e= 50-30,000 ev, z= 1-92. *Atomic data and nuclear data tables*, 54(2):181–342, 1993.

[3] Martin Berger, J Hubbell, Stephen Seltzer, J Coursey, and D Zucker. Xcom: Photon cross section database (version 1.2), 1999-01-01 1999.

[4] David Paganin. *Coherent X-ray optics*. Number 6. Oxford University Press, USA, 2006.

[5] Chris Jacobsen. *X-ray Microscopy*. Cambridge University Press, 2019.

[6] Jens Als-Nielsen and Des McMorrow. *Elements of modern X-ray physics*. John Wiley & Sons, 2011.

[7] Joseph W Goodman. *Introduction to Fourier optics*. Roberts and Company publishers, 2005.

[8] Stephen D Mellin and Gregory P Nordin. Limits of scalar diffraction theory and an iterative angular spectrum algorithm for finite aperture diffractive optical element design. *Optics Express*, 8(13):705–722, 2001.

[9] Jianwei Miao, Pambos Charalambous, Janos Kirz, and David Sayre. Extending the methodology of x-ray crystallography to allow imaging of micrometre-sized non-crystalline specimens. *Nature*, 400(6742):342–344, 1999.

[10] Pierre Thibault, Martin Dierolf, Andreas Menzel, Oliver Bunk, Christian David, and Franz Pfeiffer. High-resolution scanning x-ray diffraction microscopy. *Science*, 321(5887):379–382, 2008.

[11] Yudong Yao, Henry Chan, Subramanian Sankaranarayanan, Prasanna Balaprakash, Ross J Harder, and Mathew J Cherukara. Autophasenn: unsupervised physics-aware deep learning of 3d nanoscale bragg coherent diffraction imaging. *npj Computational Materials*, 8(1):124, 2022.

[12] Robert A Lewis. Medical phase contrast x-ray imaging: current status and future prospects. *Physics in medicine & biology*, 49(16):3573, 2004.

[13] Ping-chin Cheng and Gwo-jen Jan. *X-ray microscopy: instrumentation and biological applications*. Springer Science & Business Media, 2012.

[14] Dennis Gabor. Holography, 1948-1971. *Science*, 177(4046):299–313, 1972.

[15] George W Stroke. Lensless fourier-transform method for optical holography. *Applied Physics Letters*, 6(10):201–203, 1965.

[16] Charles W Groetsch and CW Groetsch. *Inverse problems in the mathematical sciences*, volume 52. Springer, 1993.

[17] KA Nugent, TE Gureyev, DF Cookson, D Paganin, and ZNMN Barnea. Quantitative phase imaging using hard x rays. *Physical review letters*, 77(14):2961, 1996.

[18] Peter Cloetens, W Ludwig, J Baruchel, D Van Dyck, J Van Landuyt, JP Guigay, and Ma Schlenker. Holotomography: Quantitative phase tomography with micrometer resolution using hard synchrotron radiation x rays. *Applied physics letters*, 75(19):2912–2914, 1999.

[19] David Paganin, Sheridan C Mayo, Tim E Gureyev, Peter R Miller, and Steve W Wilkins. Simultaneous phase and amplitude extraction from a single defocused image of a homogeneous object. *Journal of microscopy*, 206(1):33–40, 2002.

[20] Max Langer, Peter Cloetens, Jean-Pierre Guigay, and Françoise Peyrin. Quantitative comparison of direct phase retrieval algorithms in in-line phase tomography. *Medical physics*, 35(10):4556–4566, 2008.

[21] Max Langer, Yuhe Zhang, Diogo Figueirinhas, J-B Forien, Kannara Mom, Claire Mouton, Rajmund Mokso, and Pablo Villanueva-Perez. Pyphase–a python package for x-ray phase imaging. *Journal of Synchrotron Radiation*, 28(4):1261–1266, 2021.

[22] Ralph W Gerchberg. A practical algorithm for the determination of plane from image and diffraction pictures. *Optik*, 35(2):237–246, 1972.

[23] Reconstruction of an object from the modulus of its fourier transform. *Optics letters*, 3(1):27–29, 1978.

[24] Yair Rivenson, Yibo Zhang, Harun Günaydın, Da Teng, and Aydogan Ozcan. Phase recovery and holographic image reconstruction using deep learning in neural networks. *Light: Science & Applications*, 7(2):17141–17141, 2018.

[25] Yuhe Zhang, Mike Andreas Noack, Patrik Vagovic, Kamel Fezzaa, Francisco Garcia-Moreno, Tobias Ritschel, and Pablo Villanueva-Perez. Phasegan: a deep-learning phase-retrieval approach for unpaired datasets. *Optics express*, 29(13):19593–19604, 2021.

[26] Godfrey N Hounsfield. Computerized transverse axial scanning (tomography): Part 1. description of system. *The British journal of radiology*, 46(552):1016–1022, 1973.

[27] Masato Hoshino, Kentaro Uesugi, James Pearson, Takashi Sonobe, Mikiyasu Shirai, and Naoto Yagi. Development of an x-ray real-time stereo imaging technique using synchrotron radiation. *Journal of synchrotron radiation*, 18(4):569–574, 2011.

[28] P Villanueva-Perez, B Pedrini, R Mokso, P Vagovic, VA Guzenko, SJ Leake, PR Willmott, P Oberta, C David, HN Chapman, et al. Hard x-ray multi-projection imaging for single-shot approaches. *Optica*, 5(12):1521–1524, 2018.

[29] Wolfgang Voegeli, Kentaro Kajiwara, Hiroyuki Kudo, Tetsuroh Shirasawa, Xiaoyu Liang, and Wataru Yashiro. Multibeam x-ray optical system for high-speed tomography. *Optica*, 7(5):514–517, 2020.

[30] Lawrence A Shepp and Benjamin F Logan. The fourier reconstruction of a head section. *IEEE Transactions on nuclear science*, 21(3):21–43, 1974.

[31] Guangming Zang, Ramzi Idoughi, Rui Li, Peter Wonka, and Wolfgang Heidrich. Intratomo: self-supervised learning-based tomography via sinogram synthesis and prediction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1960–1970, 2021.

[32] Liyue Shen, John Pauly, and Lei Xing. Nerp: implicit neural representation learning with prior embedding for sparsely sampled image reconstruction. *IEEE Transactions on Neural Networks and Learning Systems*, 2022.

[33] Francisco García-Moreno, Paul Hans Kamm, Tillmann Robert Neu, Felix Bülk, Mike Andreas Noack, Mareike Wegener, Nadine von der Eltz, Christian Matthias Schlepütz, Marco Stampanoni, and John Banhart. Tomoscopy: Time-resolved tomography for dynamic processes in materials. *Advanced Materials*, 33(45):2104659, 2021.

[34] Pablo Villanueva-Perez, Valerio Bellucci, Yuhe Zhang, Sarlota Birnsteinova, Rita Graceffa, Luigi Adriano, Eleni Myrto Asimakopoulou, Ilia Petrov, Zisheng Yao,

Marco Romagnoni, et al. Megahertz x-ray multi-projection imaging. *arXiv preprint arXiv:2305.11920*, 2023.

[35] Eleni Myrto Asimakopoulou, Valerio Bellucci, Sarlota Birnsteinova, Zisheng Yao, Yuhe Zhang, Ilia Petrov, Carsten Deiter, Andrea Mazzolari, Marco Romagnoni, Dusan Korytar, et al. Development towards high-resolution khz-speed rotation-free volumetric imaging. *Optics Express*, 32(3):4413–4426, 2024.

[36] John P Blewett. Synchrotron radiation—1873 to 1947. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 266(1-3):1–9, 1988.

[37] Arthur L Robinson. History of synchrotron radiation. *Synchrotron Radiation News*, 28(4):4–9, 2015.

[38] Pascal Elleaume. Insertion devices for the new generation of synchrotron sources: A review. *Review of scientific instruments*, 63(1):321–326, 1992.

[39] B Buras and G Materlik. The european synchrotron radiation facility: An overview. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 246(1-3):21–31, 1986.

[40] John N Galayda. The advanced photon source. In *Proceedings Particle Accelerator Conference*, volume 1, pages 4–8. IEEE, 1995.

[41] M Böge. First operation of the swiss light source. In *Proc. EPAC*, volume 2, 2002.

[42] H Kamitsubo. Spring-8 program. *Journal of Synchrotron Radiation*, 5(3):162–167, 1998.

[43] Nils Martensson and Mikael Eriksson. The saga of max iv, the first multi-bend achromat synchrotron light source. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 907:97–104, 2018.

[44] Seunghwan Shin. New era of synchrotron radiation: fourth-generation storage ring. *AAPPS Bulletin*, 31(1):21, 2021.

[45] Aymeric Robert, Yngve Cerenius, Pedro Fernandes Tavares, Anna Hultin Stigenberg, Olof Karis, Ann-Christine Lloyd Whelan, Caroline Runéus, and Marjolein Thunnissen. Max iv laboratory. *The European Physical Journal Plus*, 138(6):495, 2023.

[46] Gianluca Geloni, Evgeni Saldin, Evgeni Schneidmiller, and Mikhail Yurkov. Transverse coherence properties of x-ray beams in third-generation synchrotron radiation sources. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 588(3):463–493, 2008.

[47] Brian WJ McNeil and Neil R Thompson. X-ray free-electron lasers. *Nature photonics*, 4(12):814–821, 2010.

[48] Henry N Chapman, Carl Caleman, and Nicusor Timneanu. Diffraction before destruction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1647):20130313, 2014.

[49] Philip H Bucksbaum and Nora Berrah. Brighter and faster: The promise and challenge of the x-ray free-electron laser. *Physics Today*, 68(7):26–32, 2015.

[50] SN Luo, BJ Jensen, DE Hooks, K Fezzaa, KJ Ramos, JD Yeager, K Kwiatkowski, and T Shimada. Gas gun shock experiments with single-pulse x-ray phase contrast imaging and diffraction at the advanced photon source. *Review of Scientific Instruments*, 83(7), 2012.

[51] Massimo Altarelli. The european x-ray free-electron laser: Toward an ultra-bright, high repetition-rate x-ray source. *High Power Laser Science and Engineering*, 3:e18, 2015.

[52] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.

[53] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015.

[54] Warren S McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5:115–133, 1943.

[55] Herbert Robbins and Sutton Monro. A stochastic approximation method. *The annals of mathematical statistics*, pages 400–407, 1951.

[56] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[57] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017.

[58] Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked autoencoders are scalable vision learners. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16000–16009, 2022.

[59] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021.

[60] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural networks*, 2(5):359–366, 1989.

[61] Yann LeCun, Yoshua Bengio, et al. Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, 3361(10):1995, 1995.

[62] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[63] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.

[64] Zhijian Liu, Haotian Tang, Yujun Lin, and Song Han. Point-voxel cnn for efficient 3d deep learning. *Advances in Neural Information Processing Systems*, 32, 2019.

[65] Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. Deepsdf: Learning continuous signed distance functions for shape representation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 165–174, 2019.

[66] Vincent Sitzmann, Julien Martel, Alexander Bergman, David Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions. *Advances in neural information processing systems*, 33:7462–7473, 2020.

[67] Yu Sun, Jiaming Liu, Mingyang Xie, Brendt Wohlberg, and Ulugbek S Kamilov. Coil: Coordinate-based internal learning for imaging inverse problems. *arXiv preprint arXiv:2102.05181*, 2021.

[68] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023.

[69] Aimee Van Wynsberghe. Sustainable ai: Ai for sustainability and the sustainability of ai. *AI and Ethics*, 1(3):213–218, 2021.

[70] Chanseok Lee, Gookho Song, Hyeonggeon Kim, Jong Chul Ye, and Mooseok Jang. Deep learning based on parameterized physical forward model for adaptive holographic imaging with unpaired data. *Nature Machine Intelligence*, pages 1–11, 2023.

[71] Vincent Van Nieuwenhove, Jan De Beenhouwer, Francesco De Carlo, Lucia Mancini, Federica Marone, and Jan Sijbers. Dynamic intensity normalization using eigen flat fields in x-ray imaging. *Optics express*, 23(21):27975–27989, 2015.

[72] SV Milton, E Gluskin, ND Arnold, C Benson, W Berg, SG Biedron, M Borland, Y-C Chae, RJ Dejus, PK Den Hartog, et al. Exponential gain and saturation of a self-amplified spontaneous emission free-electron laser. *Science*, 292(5524):2037–2041, 2001.

[73] James Anthony Seibert, John M Boone, and Karen K Lindfors. Flat-field correction technique for digital detectors. In *Medical Imaging 1998: Physics of Medical Imaging*, volume 3336, pages 348–354. SPIE, 1998.

[74] Peter Schmüser, Martin Dohlus, and Jörg Rossbach. *Ultraviolet and soft X-ray free-electron lasers: introduction to physical principles, experimental results, technological challenges*, volume 229. Springer Science & Business Media, 2008.

[75] Michael Greenacre, Patrick JF Groenen, Trevor Hastie, Alfonso Iodice d' Enza, Angelos Markos, and Elena Tuzhilina. Principal component analysis. *Nature Reviews Methods Primers*, 2(1):100, 2022.

[76] Khachiwan Buakor. X-ray imaging at high brilliance sources. 2020.

[77] P Guigay. Fourier-transform analysis of fresnel diffraction patterns and in-line holograms. *Optik*, 49:121–125, 1977.

[78] Andreas Berberich, Andreas Kurz, Sebastian Reinhard, Torsten Johann Paul, Paul Ray Burd, Markus Sauer, and Philip Kollmannsberger. Fourier ring correlation and anisotropic kernel density estimation improve deep learning based smlm reconstruction of microtubules. *Frontiers in Bioinformatics*, 1:752788, 2021.

[79] Yuhe Zhang, Tobias Ritschel, and Pablo Villanueva-Perez. Reusability report: Unpaired deep-learning approaches for holographic image reconstruction. *Nature Machine Intelligence*, pages 1–7, 2024.

[80] Ayan Sinha, Justin Lee, Shuai Li, and George Barbastathis. Lensless computational imaging through deep learning. *Optica*, 4(9):1117–1125, 2017.

[81] Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Deep learning face attributes in the wild. In *Proceedings of the IEEE international conference on computer vision*, pages 3730–3738, 2015.

[82] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017.

[83] Tianjiao Zeng, Yanmin Zhu, and Edmund Y Lam. Deep learning for digital holography: a review. *Optics Express*, 29(24):40572–40593, 2021.

[84] Juyeon Park, Bijie Bai, DongHun Ryu, Tairan Liu, Chungha Lee, Yi Luo, Mahn Jae Lee, Luzhe Huang, Jeongwon Shin, Yijie Zhang, et al. Artificial intelligence-enabled quantitative phase imaging methods for life sciences. *Nature Methods*, 20(11):1645–1660, 2023.

[85] Chanseok Lee, Gookho Song, Hyeonggeon Kim, Jong Chul Ye, and Mooseok Jang. 3um polystyrene bead, red blood cell and histological slide datasets. `https://figshare.com/articles/dataset/3um_polystyrene_bead_red_blood_cell_datasets/21378744`, 2022.

[86] Ge Wang, Jong Chul Ye, and Bruno De Man. Deep learning for tomographic image reconstruction. *Nature Machine Intelligence*, 2(12):737–748, 2020.

[87] Alex Yu, Vickie Ye, Matthew Tancik, and Angjoo Kanazawa. pixelnerf: Neural radiance fields from one or few images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4578–4587, 2021.

[88] Albert Pumarola, Enric Corona, Gerard Pons-Moll, and Francesc Moreno-Noguer. D-nerf: Neural radiance fields for dynamic scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10318–10327, 2021.

[89] Daniel Otero Baguer, Johannes Leuschner, and Maximilian Schmidt. Computed tomography reconstruction using deep image prior and learned reconstruction methods. *Inverse Problems*, 36(9):094004, 2020.

[90] Silja Flenner, Stefan Bruns, Elena Longo, Andrew J Parnell, Kilian E Stockhausen, Martin Müller, and Imke Greving. Machine learning denoising of high-resolution x-ray nanotomography data. *Journal of synchrotron radiation*, 29(1):230–238, 2022.

[91] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017.

[92] Rui Guo, Johannes Stubbe, Yuhe Zhang, Christian Matthias Schlepütz, Camilo Rojas Gomez, Mahoor Mehdikhani, Christian Breite, Yentl Swolfs, and Pablo Villanueva-Perez. Deep-learning image enhancement and fibre segmentation from time-resolved computed tomography of fibre-reinforced composites. *Composites Science and Technology*, 244:110278, 2023.

[93] Allard Adriaan Hendriksen, Daniël Maria Pelt, and K Joost Batenburg. Noise2inverse: Self-supervised deep convolutional denoising for tomography. *IEEE Transactions on Computational Imaging*, 6:1320–1335, 2020.

[94] Laith Alzubaidi, Mohammed A Fadhel, Omran Al-Shamma, Jinglan Zhang, Jorge Santamaría, Ye Duan, and Sameer R. Oleiwi. Towards a better understanding of

transfer learning for medical imaging: a case study. *Applied Sciences*, 10(13):4523, 2020.

[95] Liyue Shen, Wei Zhao, and Lei Xing. Patient-specific reconstruction of volumetric computed tomography images from a single projection view via deep learning. *Nature biomedical engineering*, 3(11):880–888, 2019.

[96] Philipp Henzler, Niloy J Mitra, and Tobias Ritschel. Escaping plato's cave: 3d shape from adversarial rendering. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9984–9993, 2019.

[97] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

[98] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

[99] Yunping Zhang, Xihui Liu, and Edmund Y Lam. Single-shot inline holography using a physics-aware diffusion model. *Optics Express*, 32(6):10444–10460, 2024.

[100] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14*, pages 694–711. Springer, 2016.

[101] Gustav Grund Pihlgren, Fredrik Sandin, and Marcus Liwicki. Improving image autoencoder embeddings with perceptual loss. In *2020 International Joint Conference on Neural Networks (IJCNN)*, pages 1–7. IEEE, 2020.

[102] Maziar Raissi, Paris Perdikaris, and George E Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational physics*, 378:686–707, 2019.

[103] Majid Rasht-Behesht, Christian Huber, Khemraj Shukla, and George Em Karniadakis. Physics-informed neural networks (pinns) for wave propagation and full waveform inversions. *Journal of Geophysical Research: Solid Earth*, 127(5):e2021JB023120, 2022.

[104] Ayana Ghosh. Towards physics-informed explainable machine learning and causal models for materials research. *Computational Materials Science*, 233:112740, 2024.

Scientific publications

Author contributions

Paper i: Shot-to-shot flat-field correction at X-ray free-electron lasers

K. Buakor[†], **Y. Zhang**[†], Š. Birnšteinová, V. Bellucci, T. Sato, H. Kirkwood, A. P. Mancuso, P. Vagovic, P. Villanueva-Perez
[†] These authors contributed equally
Optics Express 30, 10633-10644 (2022)

I developed the deep learning algorithm and contributed to the data analysis and manuscript writing.

Paper ii: PhaseGAN: a deep-learning phase-retrieval approach for unpaired datasets

Y. Zhang, M. Andreas Noack, P. Vagovic, K. Fezzaa, F. Garcia-Moreno, T. Ritschel, P. Villanueva-Perez
Optics Express 29, 19593-19604 (2021)

I developed the algorithm, generated the simulated data, analyzed the data, and contributed to the manuscript writing.

Paper iii: Reusability Report: Unpaired deep-learning approaches for holographic image reconstruction

Y. Zhang, T. Ritschel, P. Villanueva-Perez
Nature Machine Intelligence, 1-7 (2024)

I contributed to the experimental work, data analysis, and writing of the manuscript.

Paper iv: ONIX: an X-ray deep-learning tool for 3D reconstructions from sparse views

Y. Zhang, Z. Yao, T. Ritschel, P. Villanueva-Perez
Applied Research 2, 4 (2023)

I developed the algorithm, generated the simulated data, and contributed to the data analysis and manuscript writing.

Paper v: 4D-ONIX: A deep learning approach for reconstructing 3D movies from X-ray multi-projection imaging

Y. Zhang, Z. Yao, R. Klöfkorn, T. Ritschel, P. Villanueva-Perez
arXiv preprint arXiv:2401.09508 (2024)

I developed the algorithm and contributed to the data analysis and manuscript writing.

Paper vi: Megahertz X-ray Multi-projection imaging

P. Villanueva-Perez, V. Bellucci, **Y. Zhang**, Š. Birnšteinová, R. Graceffa, L. Adriano, E. M. Asimakopoulou, I. Petrov, Z. Yao, M. Romagnoni, A. Mazzolari, R. Letrun, C. Kim, J. Koliyadu, C. Deiter, R. Bean, G. Giovanetti, L. Gelisio, T. Ritschel, A. Mancuso, H. N. Chapman, A. Meents, T. Sato, P. Vagovic
arXiv preprint arXiv:2305.11920 (2023)

I contributed to the experimental work, algorithm development, data analysis, and writing of the manuscript.

Printed by Media-Tryck, Lund 2024 NORDIC SWAN ECOLABEL 3041 0903

LUND UNIVERSITY

Faculty of Science
Department of Physics

ISBN 978-91-8039-982-1