JAPAN ADVANCED INSTITUTE OF SCIENCE AND TECHNOLOGY

Doctoral Dissertation

Enhancing the Reliability of Phase Retrieval in Single-Frame Coherent X-ray Diffraction Imaging

---

Author:
Adam Mukharil BACHTIAR

Supervisor:
Hieu-Chi DAM

A thesis submitted in fulfillment of the requirements for the degree of Knowledge science

in the

Dam Laboratory
School of Knowledge Science

September, 2025

Declaration of Authorship

I, Adam Mukharil BACHTIAR, declare that this thesis titled, "Enhancing the Reliability of Phase Retrieval in Single-Frame Coherent X-ray Diffraction Imaging" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work I did jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: Adam Mukharil Bachtiar

Date: July 22nd, 2025

"Knowledge is a hunter, and writing is its restraint. It would be wise to restrain your hunts with reliable ropes. Catching a deer and leaving it among the creatures to be divorced would be unwise."

Imam Syafi'i Rahimahullah

JAPAN ADVANCED INSTITUTE OF SCIENCE AND TECHNOLOGY

Abstract

Co-Creative Intelligence research area
School of Knowledge Science

Knowledge science

Enhancing the Reliability of Phase Retrieval in Single-Frame Coherent X-ray Diffraction Imaging

by Adam Mukharil BACHTIAR

Single-frame Coherent X-ray Diffraction Imaging (CXDI) promises an ultrafast, computational imaging technique of nanomaterials, yet its practical impact is still limited by the low reliability of phase-retrieval reconstructions. In the absence of scan redundancy, the inverse problem becomes severely ill-posed, especially when multiple real-space objects can satisfy exactly the same diffraction intensities. This dissertation addresses the reliability gap through a two-distinct strategy that integrates optical system optimization with uncertainty-aware computational inference. First, we revisit the optical-instrument design by treating the aperture not as a fixed component but as a tunable parameter. This study systematically explores how the corner radius of a triangular aperture modulates the performance of single-frame CXDI. Large-scale simulations and error metrics in both reciprocal and real space reveal an optimal curvature range that balances fabrication feasibility with maximal reconstruction fidelity. The result is a quantitative guideline that elevates hardware design from empirical choice to data-driven optimization, directly enhancing sample reconstruction stability. Second, we introduce Ensemble Phase Retrieval, a framework that boosts reliability beyond image fidelity by explicitly characterizing epistemic uncertainty. Multiple phase-retrieval solutions are generated through stochastic masking using the sliding windows concept, diverse algorithmic initializations, and hybrid solver choices (iterative and deep learning phase retrieval). Multiple–solution aggregation yields a consensus image, while pixel-wise variance localizes regions of ambiguity. The ensemble strategy, therefore, provides two simultaneous deliverables, namely a higher-confidence reconstruction less sensitive to initialization or missing data, and an interpretable ambiguity map that guides scientific judgment. Together, both studies form a cohesive pipeline that enhances the reliability of phase retrieval in single-frame CXDI.

Keywords: Computational Imaging, Nanomaterials Imaging, Single-Frame CXDI, Phase Retrieval, Ensemble Phase Retrieval, Multi-solution Paradigm, Ambiguity Quantification.

Acknowledgements

First and foremost, my deepest gratitude is to Allah SWT, the Almighty, whose continuous guidance, mercy, and infinite wisdom have illuminated my path, granting me the strength and clarity to undertake and complete this dissertation. All praise is due to Allah, who inspires and sustains every step of my journey. My profound respect and reverence also extend to Prophet Muhammad SAW, whose teachings and noble life have brought humanity from darkness into the light of knowledge, wisdom, and compassion.

I express my sincere appreciation and deepest gratitude to my esteemed supervisor, Professor Hieu-Chi Dam, whose invaluable guidance, insightful discussions, and unwavering encouragement have significantly enriched this work. Professor Dam's mentorship and thoughtful direction have allowed me the privilege to immerse myself in rigorous scientific research, ultimately allowing me to experience the profound joy and fulfillment inherent in scholarly discovery.

I am deeply grateful to my colleagues and lab mates, Dr. Minh-Quyet Ha, Dr. Tien-Sinh Vu, Dr. Nguyen Dong Nguyen, Duc-Dao Anh, Dinh-Khiet Le, Vuong Cong, Vu Ding-Hong Phuc, and Abdurrazak Syakir Muharam for their constant support, encouragement, and camaraderie throughout this journey. Their companionship and intellectual insight have been indispensable in overcoming many challenges.

Special thanks to my beloved wife, Elsa Widiati, whose unwavering patience, steadfast support, and loving companionship have helped me through both the brightest and most challenging moments of this endeavor. Her presence has always been my most excellent comfort and motivation.

My heartfelt gratitude extends to my cherished family members—my late father, Amaludin Bachtiar, my mother, Marhaen Dyah Ismoyowati, siblings Anrio Sonri Bachtiar, Shativa Sonrisa, my late father-in-law, Muhammad Nurdjaja, my mother-in-law, Suesty, and relatives Dody Widjaja, Silviana, Herti Widjajanthy, Zaenal Abidin, Ivone Chormedian, and others for their unconditional love, continuous prayers, encouragement, and unwavering belief in my capabilities.

Finally, I extend my gratitude to Professor Eddy Soeryanto Soegoto, Rector of the Universitas Komputer Indonesia (UNIKOM), for providing the financial support necessary to carry out and complete this academic pursuit. His support has been instrumental in ensuring the successful completion of this dissertation.

Contents

List of Figures

List of Tables

List of Abbreviations

CXDI	**C**oherent **X**-ray **D**iffraction **I**maging
FZP	**F**resnel **Z**one **P**late **I**maging
GS	**G**erchberg **S**axton
ER	**E**rror **R**eduction
HIO	**H**ybrid **I**nput **O**utput
ePIE	**e**xtended **P**tychographical **I**terative **E**ngine
CNN	**C**onvolutional **N**eural **N**etwork
MSE	**M**ean **S**quared **E**rror
PSNR	**P**eak **S**ignal-to-**N**oise **R**atio
SSIM	**S**tructural **S**imilarity **I**ndex **M**easure
ASM	**A**ngular **S**pectrum **M**ethod
CB	**C**onvolutional **B**locks
TB	**T**emporal **B**lock
RB	**R**efinement **B**lock
FrFT	**Fr**actional **F**ourier **T**ransform

List of Symbols

$\mathcal{F}$	Fourier Transformation
$\mathcal{F}^{-1}$	inverse Fourier Transformation

Dedicated to all researchers in Computational Imaging and Knowledge Science domain.

Chapter 1

Introduction

1.1 Overview of Background and Motivation

Imaging has become one of the most revolutionary fields in modern technological development, transforming how we understand, record, and interpret the world around us. Since the first photography invention by Joseph Nicéphore Niépce in the 1820s, imaging technology has undergone a remarkable transformation, evolving from simple chemical-based processes to highly sophisticated digital systems (Bann, 2002). In today's digital era, imaging has become an integral part of various aspects of human life, ranging from personal documentation to critical medical applications. The evolution of imaging technology has led to a paradigm shift in the way we process and analyze visual information. The transition from analog film to digital sensors represents a significant technological leap that enables high-resolution image acquisition, efficient storage, and previously impossible processing capabilities. Modern imaging sensors can capture not only the light spectrum visible to the human eye but also various other electromagnetic wavelengths, opening new horizons in applications such as infrared, ultraviolet, and spectral imaging (Ferreira et al., 2017).

From a historical standpoint, imaging was dominated by conventional imaging where lenses or mirrors produced a direct real-space projection, and higher resolution was achieved mainly by enlarging the physical aperture or thoroughly correcting aberrations (Cossairt, M. Gupta, and Nayar, 2012). Because the light field was fully captured and immediately focused onto a detector, the image could be viewed with minimal post-processing. Over the past three decades, however, shrinking target scales and rapid gains in detector and computing power have shifted the bottleneck from hardware to algorithms. In computational imaging, the optics are intentionally designed to encode the scene into a sparse or incomplete measurement, whether a coded-aperture snapshot, a diffraction pattern, or a set of interferometric visibilities. Meanwhile, a back-end algorithm decodes that measurement by solving an ill-posed inverse problem, guided by physics-based or data-driven priors (Bhandari, Kadambi, and Raskar, 2022; Mait, Euliss, and Athale, 2018). Image quality now depends less on perfect glass and more on the synergy between optical encoding and mathematical reconstruction. This paradigm decouples resolution from aperture size, enables lens-less systems such as single-frame CXDI and astronomical interferometry, and elevates hardware and computational design to co-equal, tightly coordinated pillars of modern imaging science.

Computational imaging has been a key tool in science for understanding and exploring things that cannot be seen with the naked eye. From early optical systems using simple lenses to modern electron microscopes and X-ray imaging technologies, the ability to create visual representations of hidden structures has helped scientists make important discoveries in many fields (Falcone et al., 2011; Y. Zhu and Dürr,

2015). These representations are more than just pictures. They are tools for gaining knowledge, allowing us to see the shape, structure, and behavior of objects on microscopic and even atomic scales. The evolution of imaging systems has followed a clear transition from purely optical setups to advanced computational frameworks (C. Li et al., 2021). In conventional imaging, the formation of an image depends primarily on the direct interaction of light with lenses and mirrors. These systems rely on physical optics to focus light and produce a visual representation of the object, as seen in early microscopes and cameras. Although such methods have enabled significant scientific progress, they are fundamentally limited by the properties of light, particularly its wavelength, and by the imperfections of optical components. For instance, the diffraction limit imposes a natural boundary on the resolution that can be achieved using visible light, making it difficult to observe structures smaller than a few hundred nanometers (Miao, Sandberg, and C. Song, 2011).

To overcome these limitations, researchers began integrating computational techniques into the imaging process. This development marked the beginning of computational imaging, where image formation is no longer solely dependent on optical hardware but also involves mathematical models and algorithms (Mait, Euliss, and Athale, 2018). In computational imaging, the system is designed with the understanding that the raw measurements may be incomplete or indirect, and that a significant part of the image reconstruction occurs through computation. This approach allows for the recovery of hidden features, enhancement of image resolution, and even the reconstruction of images' atomic arrangement, lattice defects' presence, as well as diffraction patterns. Figure 1.1 shows the progress of imaging techniques, which started from conventional to computational imaging.

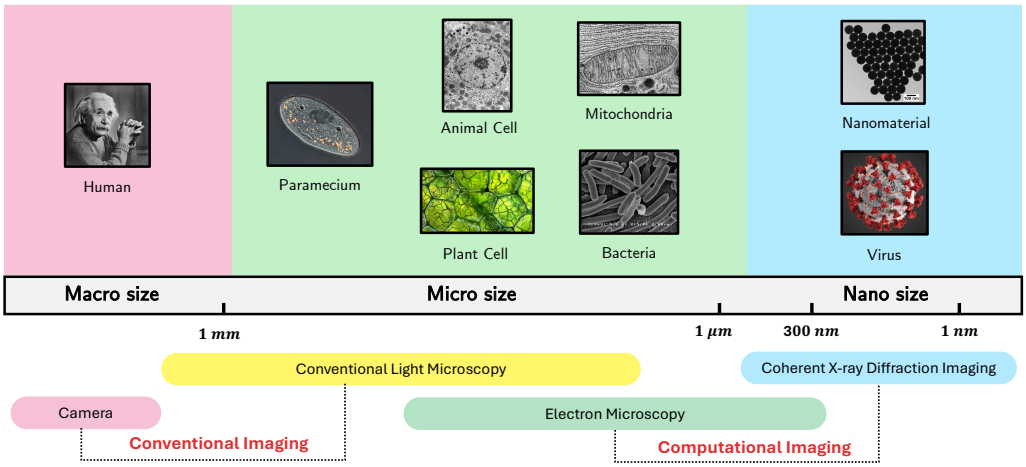


FIGURE 1.1: The progress of imaging techniques from conventional to computational imaging.

The key idea behind computational imaging is the joint design of the physical sensing process and the computational reconstruction. By modeling the physics of image formation and incorporating prior knowledge or constraints, it becomes possible to retrieve meaningful information from limited data. As a result, computational imaging opens new possibilities for studying microscopic and nanoscale structures, where traditional optical systems fall short. The integration of algorithms, optimization, and simulation into the imaging pipeline reflects a broader shift in science, where measurement and computation work together to reveal what cannot be directly seen. The hardware or physical design and computational design became two essential pillars

in the domain of computational imaging. In computational imaging, the reliability encompasses more than numerical fidelity. This word refers to how confidently each reconstructed feature can be trusted. Because the raw measurements are often sparse, noisy, or contain information loss, the imaging process is ill-posed and can produce multiple solutions that satisfy the recorded data in the reconstruction process. A reliable pipeline must therefore integrate two pillars: **reliable hardware design**, which maximizes information content and reduces systematic bias at signal acquisition and **robust computational design**, which regularize the inverse problem and guide the reconstruction toward physically plausible images and exposes residual ambiguities rather than hiding them behind a single point estimate. When hardware design, algorithmic inversion, and confidence diagnostics are co-optimized, computational imaging can deliver outputs that are not only sharp but also scientifically trustworthy and interpretable. Figure 1.2 illustrates the computational to clarify the concept of computational imaging and its pillars.

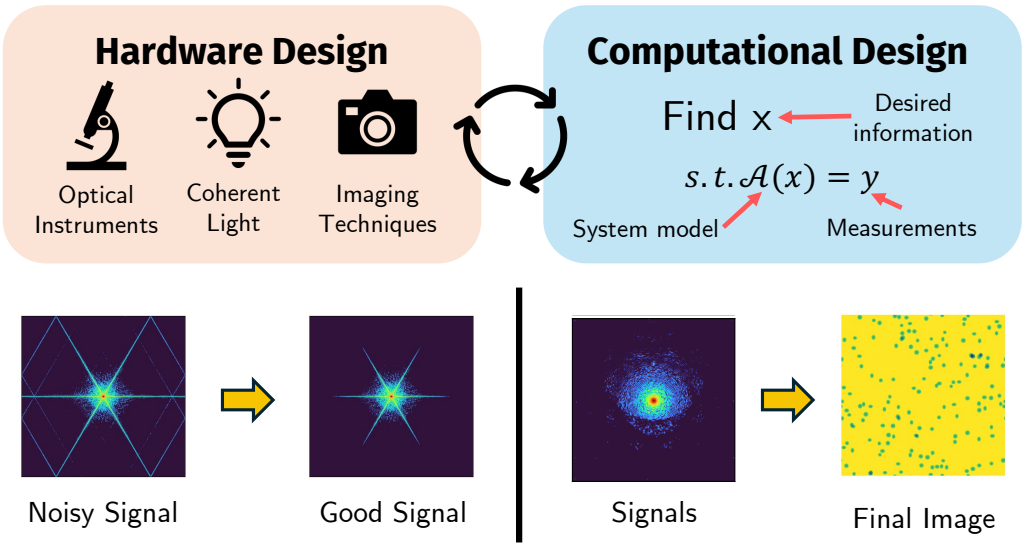


FIGURE 1.2: Illustration of computational imaging system and its essential pillars.

Material science is one of the domains that gets significant advantages from the advancement of computational imaging, particularly for studying structures at micro- and nanoscale. Many of the most critical properties of material, such as strength, conductivity, or reactivity, are determined by features that are invisible to conventional imaging methods (Levoy, 2006). These include the arrangement of atoms, the presence of lattice defects, and dynamic changes in phase or composition. Conventional imaging approaches often lack the reliability, sensitivity, or speed required to observe such phenomena, especially when dealing with non-crystalline or rapidly evolving samples. Computational imaging addresses these challenges by enabling high-resolution and reliable image reconstruction from indirect or incomplete measurements, often without the need for complex optical components.

The emergence of data-driven and physics-informed imaging techniques has significantly transformed the landscape of material characterization and computational imaging, especially in the pillars of computational design (Barbastathis, Ozcan, and Situ, 2019; Z. Wang et al., 2022; Himanen et al., 2019). By leveraging advanced algorithms, physical modeling, and data-driven techniques, computational imaging allows researchers to visualize internal structures and material behavior with greater

clarity and flexibility. This capability is particularly valuable in scenarios where direct imaging is difficult or impossible, such as in high-energy environments, dynamic experiments, or imaging of thick or heterogeneous materials. Furthermore, because computational imaging methods often incorporate domain-specific constraints and priors, they can improve reconstruction accuracy even in the presence of noise or data loss. As a result, computational imaging has opened new possibilities for in-situ and real-time characterization of materials, making it easier to monitor changes during fabrication, assess degradation over time, or study phenomena that occur under extreme conditions (Patterson et al., 2018; Panwar, Singh, and Sehgal, 2020). In essence, the significance of computational imaging in material science lies in its ability to extend what can be observed and measured. It transforms raw experimental data into meaningful structural and functional information, helping scientists to explore new materials, validate theoretical models, and design systems with desired properties. As material science increasingly intersects with fields like nanotechnology, energy storage, and quantum computing, the role of computational imaging will become even more central to discovery and innovation.

The well–defined drive to enhance the reliability of computational imaging for nanomaterials is based on two pillars: hardware design and computational design. On the hardware side, researchers exploit the diffraction principle with highly coherent, short-wavelength X-rays. Single-frame Coherent X-ray Diffraction Imaging (CXDI) incorporates this effort. This technique is a lens-less arrangement that records a far-field diffraction pattern from a single exposure, eliminating lens aberrations and pushing spatial resolution below $10nm$. However, the detector captures only intensity, losing the wavefront phase and rendering image formation an inherently ill-posed inverse problem. Addressing this shortfall constitutes the second pillar, which is phase retrieval. Progress spans physics-constrained iterative solvers such as ER, HIO, and ePIE, through to deep-learning models that embed data-driven priors (Fienup, 1982; Marchesini, 2004). By co-optimizing diffraction-oriented optics and phase-recovery algorithms, the community is steadily elevating CXDI from proof-of-concept to a quantitative tool for nanomaterial characterisation.

Despite these advances, single-frame CXDI still confronts two intertwined challenges: improving diffraction acquisition through proper aperture, as one optical component in the CXDI, and enhancing reconstruction reliability while quantifying phase ambiguity. The present dissertation tackles both fronts. In several further sections, we detail the computational imaging for nanomaterials, dissect the inverse problem of it, and outline how our research systematically strengthens each pillar to deliver a more reliable, interpretable single-frame CXDI pipeline.

1.2 Computational and Nanomaterial Imaging

The term "imaging" generally refers to the process of mapping interactions within a sample by forming a spatial representation of physical properties or categorizing elements of the surrounding environment (Ellenbogen, 2013). Historically, imaging began with basic optical techniques, initially relying solely on visible light and lenses to produce direct visual representations. Early imaging systems were predominantly optical, evolving significantly with the advent of photography, electron microscopy, and radiography throughout the 19th and 20th centuries (Haguenau et al., 2003). These advances steadily increased the ability to observe smaller scales and finer details. With the growth of computational capabilities and digital technologies in

the late 20^{th} century, imaging evolved further into computational imaging. Computational imaging represents a sophisticated methodology in which computational algorithms play an integral role, either in generating images directly from raw sensor data or in enhancing images during post-processing (Bhandari, Kadambi, and Raskar, 2022; Mait, Euliss, and Athale, 2018). Unlike conventional imaging systems, computational imaging integrates advanced mathematical algorithms and signal processing techniques within the image acquisition and reconstruction stages, enabling the extraction of richer, more detailed, and higher-quality images from limited or indirect measurements. This domain is a discipline that involves the collaborative design of front-end optics and post-detection signal processing within an imaging system.

The domain of computational imaging systems involves two primary research focuses: hardware design and computational design. Figure 1.2 shows an illustration of the computational imaging system. Hardware design emphasizes the development and optimization of sensors, optics, and other physical components that capture encoded measurements efficiently and effectively. Computational design involves formulating and solving mathematical and algorithmic models to reconstruct desired information accurately from the system's encoded data, addressing various inverse problems and optimization challenges. Some recent studies employed a data-driven approach, which involves machine learning and artificial intelligence methods, to discover and refine optimal strategies for encoding and decoding information, facilitating adaptive and intelligent imaging solutions that continually improve with additional data and experience. These interconnected research areas synergistically improve computational imaging systems, expanding their capabilities and applications across diverse scientific and technological domains.

As computational power continues to advance rapidly, computational imaging stands to revolutionize numerous fields by providing increasingly sophisticated techniques capable of yielding deeper insights into many scientific domains. The application of computational imaging principles extends to diverse domains such as molecular biology and astronomy. In the case of protein structure prediction, systems like AlphaFold demonstrate how spatial reconstruction can be achieved despite incomplete direct data (Jumper et al., 2021; Varadi et al., 2024). AlphaFold predicts the three-dimensional structure of proteins from their amino acid sequences by solving an inverse problem that resembles phase retrieval. The relationship between input (sequence) and output (structure) is complex and nonlinear. To address inherent ambiguities, AlphaFold employs a strategy that generates multiple structure models and aggregates them into a consensus structure with confidence scores such as pLDDT. This approach parallels the ensemble framework used in CXDI, where multiple reconstructions are combined to produce a consensus image accompanied by a variance map, thereby quantifying uncertainty and enhancing reliability. Meanwhile, in astronomy, similar principles are applied through interferometry, where multiple telescopes are combined to form a large virtual aperture (Saha, 2002; Kamalabadi, 2009). Rather than capturing a complete image, astronomical interferometry samples sparse spatial-frequency data, effectively recording an incomplete Fourier transform of the sky. This creates an ill-posed inverse problem, since many possible brightness distributions can explain the same set of measurements. Post-processing algorithms such as CLEAN or more advanced Bayesian methods act as computational lenses to reconstruct a physically plausible image.

Computational imaging has also emerged as a prominent area of study in recent years, particularly in microscopy and nanomaterial imaging. Nanomaterial imaging has evolved significantly over the past century, driven by the necessity to visualize

and characterize structures at increasingly smaller scales. For example, computational imaging allows scientists to monitor structural changes during battery charging cycles, phase transitions in shape-memory alloys, or stress-induced crack propagation in materials (Boyce et al., 2022; Deng et al., 2021; Shariat et al., 2019). This domain is an essential component of materials science, enabling researchers to visualize, characterize, and understand materials at atomic and subatomic scales. Given the diminutive size and unique properties of the nanomaterial, specialized imaging techniques are essential to resolving features well below the diffraction limit of visible light. Due to the extremely small dimensions of the nanomaterial, visible light, which has wavelengths typically between 400 and 700 nm, is insufficient to resolve nanoscale features (Sliney, 2016). This necessitates the use of invisible radiation with much shorter wavelengths capable of penetrating and interacting with materials at the atomic scale.

According to Abbe's diffraction limit, the smallest resolvable feature size follows this equation:

$$d = \frac{\lambda}{2\mathrm{NA}}, \tag{1.1}$$

where λ is the illumination wavelength and $\mathrm{NA} = n \sin\theta$ denotes the numerical aperture of the collecting optics. Even with a high–NA objective, visible light ($\lambda \approx$ 400–700 nm) is fundamentally restricted to resolutions on the order of $\sim$ 200 nm. To probe nanomaterials whose critical dimensions lie well below this threshold, the wavelength must be reduced by two orders of magnitude. The X-rays, with wavelengths in the 0.05–0.2 nm range, satisfy this requirement. Moreover, their high penetration depth and weak scattering cross-section allow thick or buried structures to be imaged non-destructively. The necessity of using invisible light introduces the domain to a new challenge in creating special techniques that utilize diffraction principles to improve the result of nanomaterial imaging (Bhagyaraj and Oluwafemi, 2018). Figure 1.3 shows the diagram of the required light spectrum to examine nanomaterials.

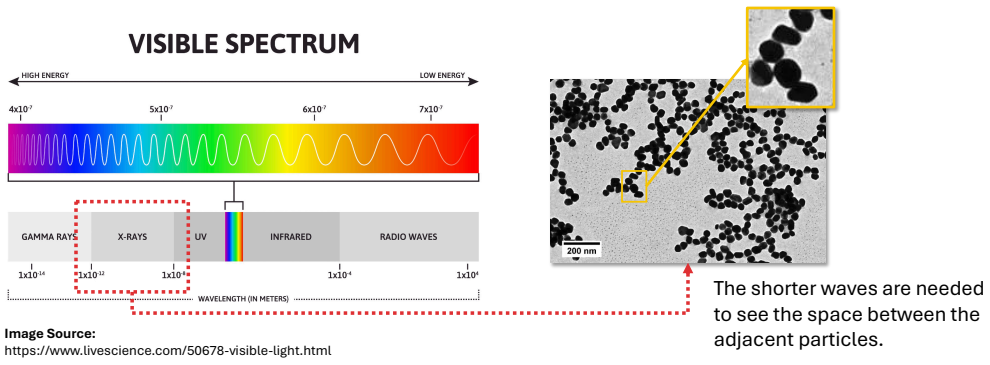


FIGURE 1.3: The light spectrum and the need for shorter wavelength light to examine nanomaterials.

At nanometer length scales, individual structural features are far smaller than the wavelength of visible light, rendering conventional lens-based microscopy inadequate. The diffraction principle offers a way to circumvent this resolution barrier. When coherent X-ray illuminate a nanoscale specimen, the scattered wavefronts interfere and form a far-field diffraction pattern. This pattern represents the squared magnitude of the object's Fourier transform, encoding high-frequency information that corresponds to sub-wavelength details. A practical limitation arises because the

detector records only the intensity of this pattern. The indispensable phase of each Fourier component is lost at acquisition. Nevertheless, if intensities are measured over a sufficiently large angular range, computational algorithms can recover or retrieve the missing phase and invert the Fourier relationship, reconstructing the real-space structure with resolution set chiefly by the probe wavelength rather than by lens quality. Consequently, diffraction-based imaging enables non-destructive, lens-free visualisation of nanomaterials, capturing both crystalline order and amorphous morphology with near-atomic precision.

Unlike conventional laboratory sources, the nanomaterial imaging with diffraction principle should utilize a **highly coherent X-ray beam**. The coherent X-ray are a specific X-rays whose wavefronts share a common wavelength (temporal coherence) and maintain a fixed phase relationship across the beam profile (spatial coherence). Coherence is essential because single-frame CXDI relies on the interference of scattered waves to encode the specimen's structure in the far-field diffraction pattern (Nakasako et al., 2020; Van der Veen and Pfeiffer, 2004). If the illumination were incoherent, the phase relationships would fluctuate, the interference fringes would wash out, and the measured pattern would reduce to a diffuse intensity ring that contains little recoverable structural information. A coherent X-ray source therefore provides the well-defined, phase-locked illumination needed to record interpretable diffraction data and to enable accurate phase-retrieval–based reconstruction. Figure 1.4 illustrates the diffraction principle and its source of light, which is coherent X-ray.

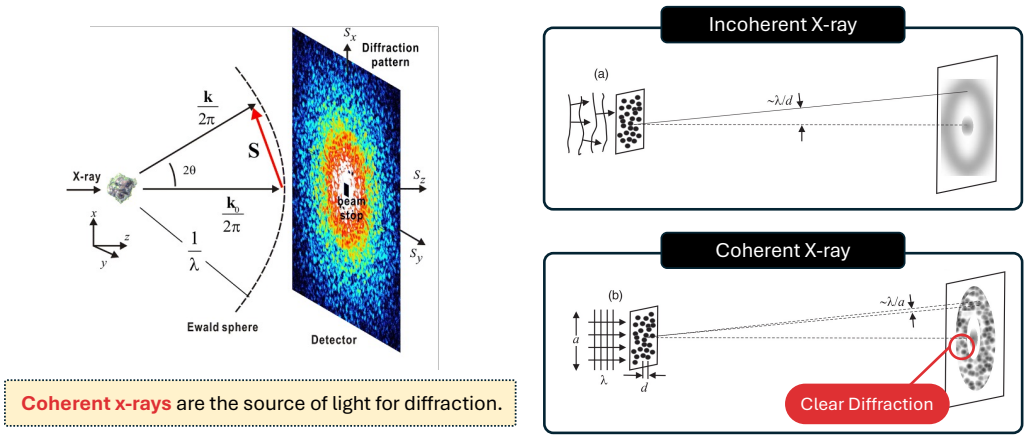


FIGURE 1.4: The principle of diffraction and why we need the coherent X-ray.

In addition, computational imaging underpins non-destructive evaluation (NDE) and quality-control pipelines in materials engineering (R. Gupta et al., 2021). Techniques such as diffraction imaging allow researchers to probe internal features without sectioning or otherwise damaging the specimen, making them invaluable in both laboratory and industrial settings (Withers et al., 2021). Crucially, these advances now extend from the microscale down to the nanoscale. Diffraction-based methods like ptychography push spatial resolution to the sub–10nm regime by replacing conventional lenses with computational phase-retrieval. When combined with proper computational design, these nanomaterial imaging techniques deliver quantitatively accurate reconstructions. Thus, integrating computational imaging across length-scales not only expands the observational reach of material scientists but also enables

new experimental designs and deeper insight into structure–property relationships, fostering advances in both fundamental research and technological innovation.

1.2.1 Techniques in nanomaterial Imaging

The advancement of computational imaging has dramatically impacted the realm of material science. Beginning with crystallography, numerous sophisticated techniques have been developed to advance research in this field (Smyth and Martin, 2000). The evolution of imaging methodologies has been fundamentally driven by scientists' increasing need to examine and understand increasingly minute and intricate material structures. Traditional imaging systems, encompassing optical microscopy, Transmission Electron Microscopy (TEM), have been instrumental in providing crucial understanding of material compositions and have remained fundamental tools in material analysis for many years. These conventional approaches typically employ physical lens systems and well-defined principles of optical science or electron-matter interactions to create magnified visualizations of both surface features and internal material structures (Tang and Z. Yang, 2017; Winey et al., 2014; Inkson, 2016).

TEM revolutionized the utilization of electrons, with a substantially shorter wavelength than visible light, to examine metals, alloys, ceramics, glasses, polymers, semiconductors, and composite materials. An electron beam passes through the nanomaterial specimen to create an image (Fultz and Howe, 2012). Transmission Electron microscopes unveil high spatial resolution and complexity that are inaccessible by conventional microscopes. It allows scientists to explore the specimen's structure, composition, and characteristics. The Transmission Electron Microscope, the primary instrument in TEM, emits an electron beam generated by an electron gun that travels down the column through a series of electromagnetic lenses, passes through the specimen, and is ultimately captured by an image collector at the bottom of the microscope. As the electrons traverse the specimen, certain areas of the material obstruct or deflect the electrons in contrast to other areas. This phenomenon leads to electron scattering, resulting in dark regions where the electrons do not pass through the sample and bright regions where the electrons remain unscattered.

The TEM instrument generally comprises several components, including an electron gun, electron column, electromagnetic lenses, apertures, detectors, specimen chamber, control panel, and image capture system (Javed et al., 2018; Bergen et al., 2013). The electron gun generates the electron beam and is typically situated at the top of the optical instrument. The electrons are emitted through the electron column maintained under vacuum conditions. The vacuum environment is essential to prevent interference with the electron beam from particles other than the specimen. Components within the electron column, such as lenses and apertures, can be adjusted to obstruct or allow the passage of the electron beam. The configuration of these movable parts determines the output, which can be a real image of the specimen or a diffraction pattern. The electromagnetic lenses and apertures focus the electron beam onto the image plane. The spatial arrangement of the lenses not only influences the outcome of the imaging process but also determines whether dark or bright-field images are produced. The focused electron beam then passes through the specimen chamber, while the scattered beam, carrying information about the projected specimen image, is captured by detectors, typically a phosphor viewing screen. The projected image is then displayed on a screen in the viewing chamber and can be captured by a camera. Adjustment of the exposure time allows for control

over the quality of the desired image. In order to ensure the orderly progression of all imaging processes, the electron microscope required $200keV$ of power.

The operation of TEM has the capability to produce both an actual image of the specimen and diffraction, depending on the configuration of lenses and aperture (D. B. Williams et al., 1996). Within TEM, diffraction patterns arise from the elastic interference of the electron beam passing through the specimen. A multi-lens system in TEM employs specific configurations to generate diffraction and undergoes a two-stage imaging process. When the specimen scatters the electron beam, the scattered beam then proceeds through an objective lens to form the actual image. The region known as the Back Focal Plane lies between the initial lens and other additional lenses in the electron microscope. The diffraction pattern can be produced by focusing the supplementary lenses on the back focal plane of lens 1. An illustration depicting the lens configuration for creating diffraction patterns in TEM can be observed in Figure 1.5.

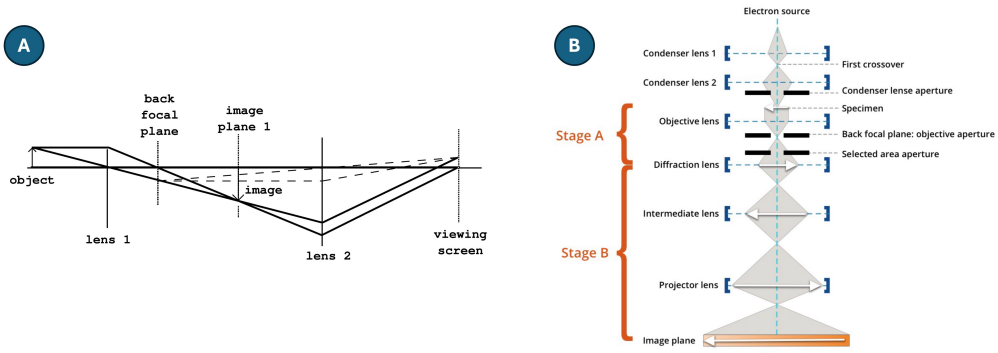


FIGURE 1.5: Configuration of the electron microscope for diffraction imaging (a) Underlying concept of lens configuration in TEM for diffraction pattern imaging, and (b) Two stages of diffraction imaging in the electron microscope.

While the TEM approach has been crucial in developing our fundamental understanding of material systems, it faces increasing challenges as research advances into the realm of dynamic systems, low-dimensional materials, and quantum structures—areas where structural changes occur rapidly and material interactions are particularly subtle (Tripathi et al., 2018). While TEM can achieve sub-nanometer resolution on ultrathin, electron-transparent specimens, its applicability narrows for many practical nanomaterials. The samples must be prepared to tens of nanometers in thickness, remain stable under high vacuum and intense electron irradiation, and fit within a limited field of view. For nanostructures that are beam-sensitive, embedded in thick matrices, or require in-situ environments, coherent X-ray diffraction techniques provide a more suitable imaging route. These limitations have catalyzed the development of new approaches to imaging, including computational methods and lensless techniques namely Ptychography. This technique offers enhanced resolution capabilities, greater operational flexibility, and compatibility with time-based studies (Miao, 2025).

Ptychography is a computational imaging method that focuses on extracting high-resolution details about the structure and properties of nanoscale specimens by reconstructing both the amplitude and phase of the object from diffraction data (J. Rodenburg and A. Maiden, 2019). In a typical experiment, a coherent X-ray probe is raster-scanned across the sample so that successive illumination spots overlap

by 60–80% (Nashed et al., 2014). At each probe position, the detector records a far-field diffraction pattern that contains only intensity information; the corresponding phase is lost. The deliberate overlap between adjacent scan positions provides a shift-invariance constraint, often termed the Ptychographic constraint supplies the redundancy needed to recover the missing phase. A phase-retrieval algorithm then enforces consistency across all overlapping frames, ultimately reconstructing a high-fidelity real-space image of the specimen. By combining coherent illumination, scanning redundancy, and computational inversion, Ptychography delivers quantitative phase and amplitude maps at spatial resolutions well below the probe diameter.

The Ptychography instrument is notably more intricate. The generation of coherent X-ray, the primary light source in Ptychography, was achieved through the Synchrotron Facility. The absence of a coherent X-ray would prevent the observation of interference effects resulting from scattering a fully coherent X-ray beam off an object. In diffraction scenarios, incoherent scattering leads to a continuous diffraction ring due to the inability to recover wave phases from various object segments in the diffracted intensity pattern, as shown in Figure 1.4. Furthermore, the execution of Ptychography presents challenges due to the substantial energy demand, estimated at $5000keV$, and the high costs associated with the development of the synchrotron facility (Boutet and G. J. Williams, 2010). Consequently, this requires researchers to be able to exploit the results of this technique properly and accurately.

In addition to the Synchrotron Facility, specific optical tools played a role in the process. The coherent X-ray was directed toward an aperture and a Fresnel Zone Plate to focus the X-rays and enhance resolution before interacting with the specimen. The detector then captured the resulting light wave emitted by the specimen as a diffraction pattern. Regrettably, as mentioned previously, the detector only captures light intensity and loses phase information during this process. This deficiency gives rise to a challenge known as phase retrieval, as the diffraction data must be transformed into the actual image utilizing both intensity and phase details. An illustration of Ptychography is shown in Figure 1.6.

In recent years, TEM and Ptychography have achieved significant milestones in computational imaging. Utilizing specific light and instrument requirements has enabled the visualization and analysis of materials from the microscale to the nanoscale. The spatial and temporal dimensions of the nanomaterial have been brought to light. The spatial dimension concentrates on the specimen resolution, in terms of diffraction, generated by an optical imaging system, while the temporal dimension sequentially captures dynamic scenes and their time resolution, considering certain limiting factors (Hu et al., 2017).

1.2.2 From Ptychography to Single-Frame Coherent X-ray Diffraction Imaging (CXDI)

With the continuous advancement and emergence of smaller-sized materials, Ptychography has become a prominent lensless imaging technique within the field of nanomaterials imaging. Ptychography utilizes coherent X-ray, which illuminate the sample object at multiple overlapping positions, systematically scanning across the specimen (Nashed et al., 2014; J. Rodenburg and A. Maiden, 2019; Zheng et al., 2021). Due to the absence of traditional lenses, the coherent X-ray interact directly with the sample, and their subsequent diffraction patterns are recorded. Each recorded diffraction pattern results from coherent illumination confined to specific localized

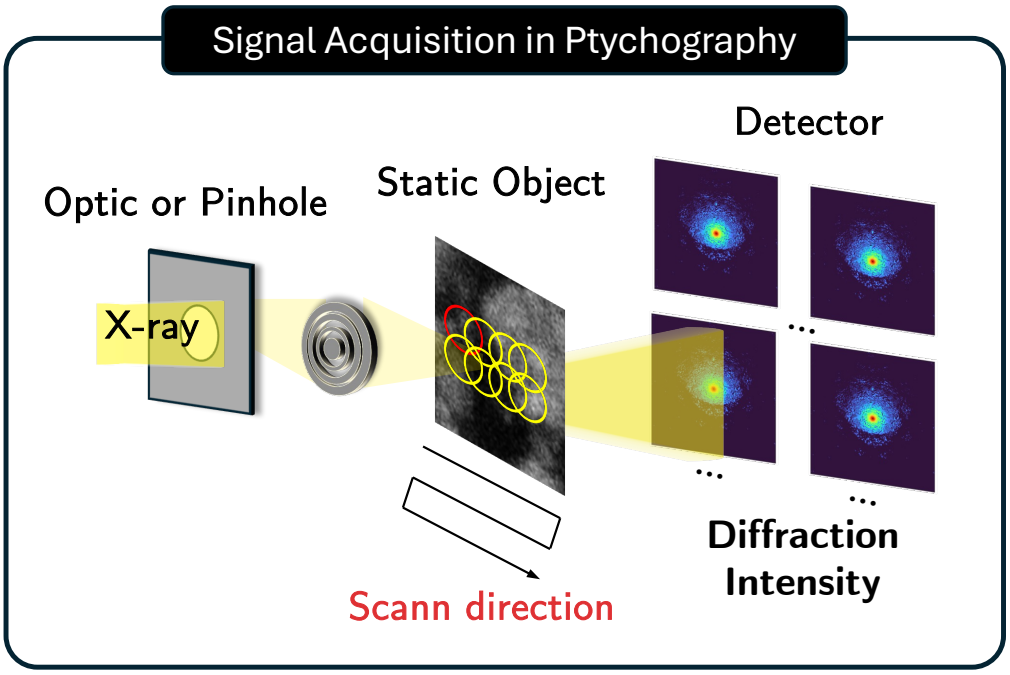


FIGURE 1.6: Illustration of Ptychography.

areas, captured through multiple raster scans to ensure significant overlap between adjacent illumination regions. This overlap is crucial for enforcing the Ptychographic shift-invariance constraint, enabling precise reconstruction of the object's complex wavefront.

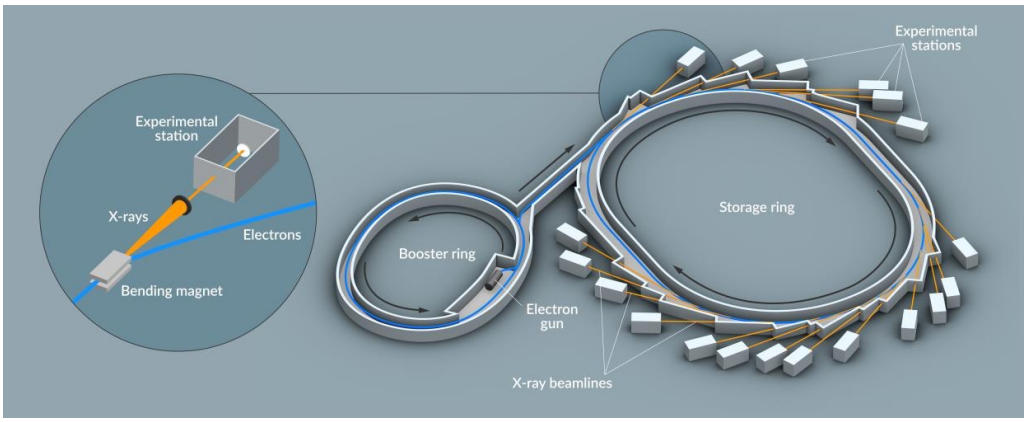


FIGURE 1.7: The illustration of a synchrotron facilitator for producing coherent X-ray.

In a Ptychographic imaging setup, each coherent illumination position (x, y) produces a corresponding two-dimensional diffraction pattern (k_x, k_y), collectively forming a comprehensive data set that encodes information about the structure of the sample, as shown in Figure 1.6. This diffraction data, called the exit wave, inherently contains both amplitude and phase information. However, practical detectors only measure the intensity of these exit waves, fundamentally losing the critical phase information (Henry N Chapman and Nugent, 2010). The overlapping requirement in Ptychography, while essential for accurate image reconstruction, introduces several practical and operational challenges. First, the capture of overlapping diffraction

patterns results in a considerable increase in the time required for data acquisition, since a greater number of raster scans are necessary to adequately sample the sample. This extended scanning time can be prohibitive, especially in high-throughput or time-sensitive applications. Secondly, the necessity for coherent X-ray sources renders Ptychography a costly technique, as coherent radiation generation requires sophisticated equipment and substantial power consumption. A Synchrotron Facilitator is needed to produce coherent X-rays. Figure 1.7 illustrates the synchrotron facilitator for coherent X-rays.

Finally, the efficacy of the technique depends on the stability of the sample during the scanning process. Any movement or instability can result in distortions and inaccuracies in the final reconstructed image. It will also have an impact on the inconsistency in the overlapping constraint, which is really essential for the reconstruction process. Consequently, Ptychography is less suitable for imaging dynamic or delicate samples that cannot be maintained in a fixed position during long scanning periods. All these challenges led to the research of a specialized technique called single-frame Coherent X-ray Diffraction Imaging (CXDI).

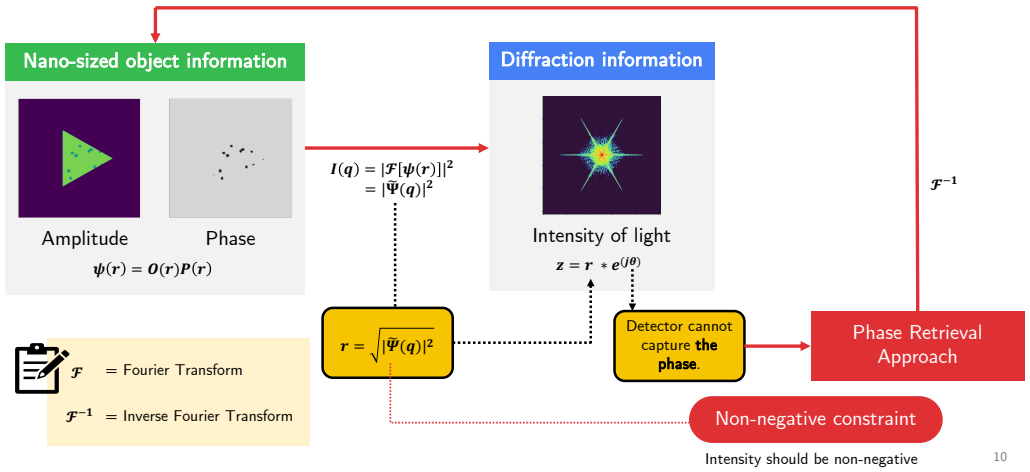


FIGURE 1.8: The fundamental problem of single-frame CXDI. The process of CXDI focuses on producing a diffraction pattern from a sample. Meanwhile, the detector only captures the intensity and loses the phase information.

In materials science, the capture of dynamic phenomena, such as rapid structural transformations, phase transitions, or other transient processes, is crucial to understanding the properties and behavior of materials under various conditions (Takazawa, Dao, et al., 2023). Single-Frame CXDI emerges as a valuable alternative solution specifically tailored for observing such rapid events. This technique involves the capture of a single diffraction pattern from the entire illuminated area in the forward process, eliminating the need for overlapping regions and extensive raster scanning (Takazawa, Kang, et al., 2021; Kang et al., 2021), as shown in Figure 1.8. This approach significantly reduces total imaging time and reduces associated costs by requiring fewer measurements and utilizing fewer resources, as explained in the mechanism of Ptychography. Moreover, single-frame CXDI is inherently suitable for imaging dynamic samples and rapid processes, as it avoids the need for sustained sample stability during data acquisition. Despite these advantages, it still faces critical challenges in the inverse process, especially in solving the complex phase retrieval problem from limited intensity-only data. The loss of overlapping information adds

an additional challenge in the process. Advanced computational techniques and innovative reconstruction algorithms become essential for effective image reconstruction, making single-frame CXDI a powerful yet computationally demanding imaging methodology within nanomaterial characterization.

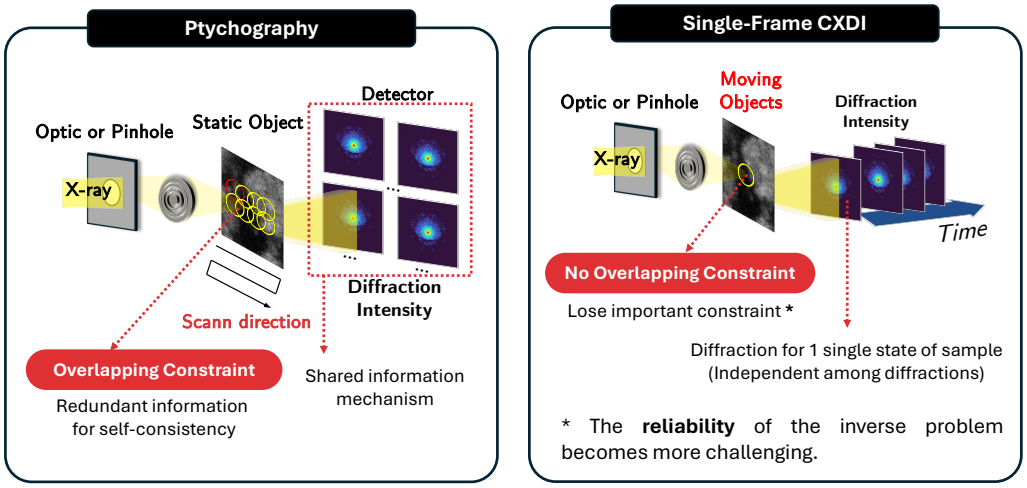


FIGURE 1.9: The fundamental difference between Ptychography and single-frame CXDI in diffraction acquisition.

Ptychography and single–frame Coherent X-ray Diffraction Imaging (CXDI) both leverage lens-less, diffraction-principle acquisition, yet they differ fundamentally in how redundancy is introduced and how quickly data can be collected. To clarify the difference in concept, Figure 1.9 shows the mechanism of each technique in acquiring the diffraction. Ptychography scans a highly coherent probe across the specimen in a dense, overlapping raster. Each probe position yields an individual diffraction pattern, and the enforced overlap supplies strong shift–invariance constraints that make phase retrieval well-conditioned and generally unique. The trade-off is acquisition time. Multiple exposures are required, and mechanical stability must be maintained throughout the scan. In contrast, single-frame CXDI illuminates the entire sample once, capturing a single diffraction snapshot that contains no positional redundancy. This single-shot approach excels in temporal resolution, allowing dynamic or radiation–sensitive samples to be imaged with minimal dose. Consequently, CXDI relies more heavily on optimized optical design (e.g. aperture engineering) and sophisticated phase-retrieval strategies to achieve reliable reconstructions. Advanced computational algorithms and reconstruction methods are required to recover the missing phase data from diffraction images, especially for the CXDI, which has lost the overlapping constraint. Consequently, this limitation poses a significant challenge known as the phase retrieval problem, also known as the inverse problem (Rosenblatt, 1984; Klibanov, Sacks, and Tikhonravov, 1995; Shechtman et al., 2015). In principle, the inverse problem is impossible to conduct without involving some well-defined constraints, and it is more challenging in single-frame CXDI. The comparison between Ptychography and single-frame CXDI can be examined in the Table 1.1.

TABLE 1.1: The comparison of Ptychography and single-frame CXDI.

Technique	Pros	Cons
Ptychography	• High phase-retrieval reliability (overlap acts as a strong constraint). • Robust to noise due to redundant measurements.	• Slow signal acquisition—requires mechanical scanning. • Limited dynamic imaging; scanning delays blur fast events. • Higher accumulated radiation dose.
Single-Frame CXDI	• Ultrafast, single-shot acquisition—ideal for transient dynamics. • Lower radiation dose; only one exposure per frame. • No scanning stages.	• Lower intrinsic phase-retrieval reliability; more ambiguity. • Highly sensitive to noise and missing data. • Requires advanced algorithms or ensembles to stabilize reconstructions.

1.3 Phase Retrieval and Inverse Problem

One of the central problems of Ptychography and single-frame CXDI is the inverse problem. The diffraction image with only intensity information should be reverted to the real-space image. Without phase information, the inverse process cannot be conducted directly. The phase retrieval process handles this problem with some mechanism and involves well-defined constraints in physics to predict the phase so that the inverse process can be conducted. The ultimate goal of phase retrieval is to reconstruct the original image, identical to the original, from the diffraction image produced by the imaging process (Nakasako et al., 2020; Jaganathan, Eldar, and Hassibi, 2016). To examine the phase retrieval problem and its challenge, we should consider the following information.

Figure 1.10 displays the mechanism of the inverse process in single-frame CXDI. In coherent diffraction imaging systems such as CXDI, the object of interest, typically represented as a complex-valued electron density distribution $O(\mathbf{r})$, is illuminated by a coherent probe, and the resulting far-field diffraction pattern is recorded in the

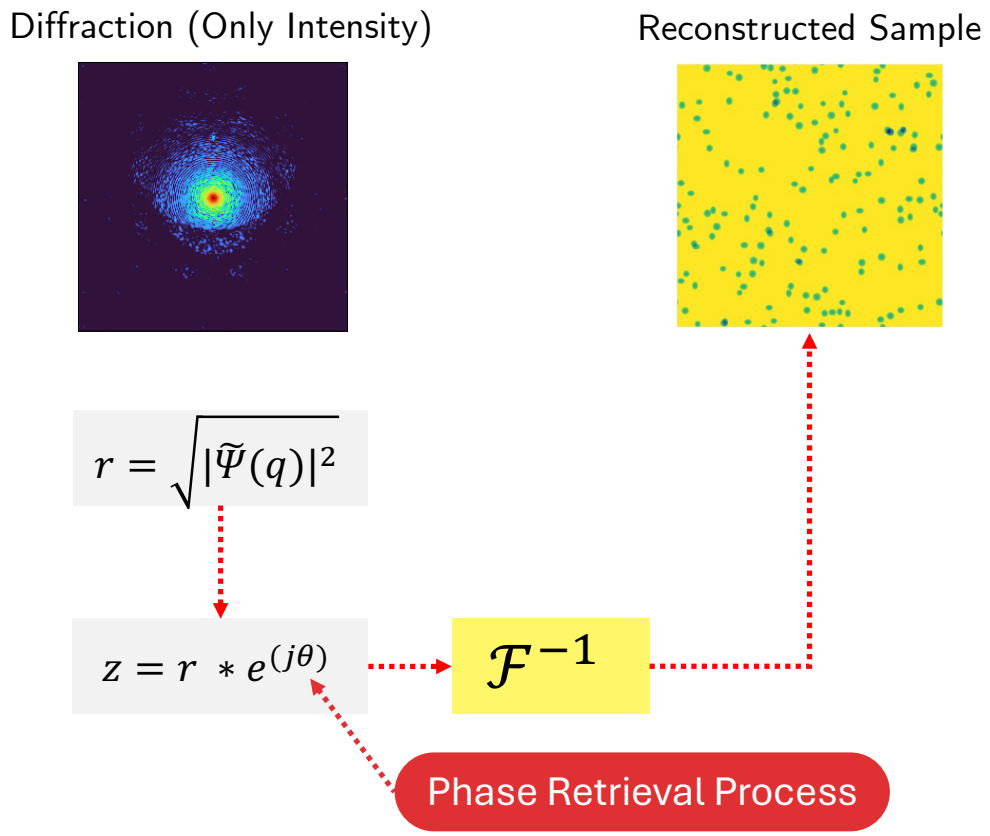


FIGURE 1.10: The inverse process in single-frame CXDI.

detector plane. This measured diffraction intensity corresponds to the squared modulus of the Fourier transform of the object (or its product with a known illumination probe), while the essential phase information is lost. Mathematically, the recorded intensity $I(\mathbf{k})$ is expressed as:

$$I(\mathbf{k}) = |\mathcal{F}\{O(\mathbf{r})\}|^2 = \left|\hat{O}(\mathbf{k})\right|^2, \tag{1.2}$$

where $\hat{O}(\mathbf{k})$ is the Fourier transform of $O(\mathbf{r})$, and $\mathbf{k}$ represents spatial frequency coordinates in reciprocal space. The measured quantity $I(\mathbf{k})$ thus provides only the magnitude:

$$\hat{O}(\mathbf{k}) = \sqrt{I(\mathbf{k})} \cdot e^{i\phi(\mathbf{k})}, \tag{1.3}$$

where $\phi(\mathbf{k})$ is the unknown phase component. Without $\phi(\mathbf{k})$, it is not possible to perform a direct inverse Fourier transform to recover the object $O(\mathbf{r})$. We should recover the phase information previously before conducting the inverse process using the inverse Fourier Transform. Intuitively, the phase recovery process is impossible without involving some well-defined physics constraints coming from the imaging process in CXDI.

1.3.1 The importance of phase information

The phase retrieval problem arises because diffraction patterns only provide the squared magnitude of the Fourier transform of the object, effectively discarding the phase component that is essential for performing an accurate inverse Fourier

transform. Without this phase, the mapping from the frequency domain back to real space is ambiguous, leading to multiple possible reconstructions that satisfy the measured intensity data. This ambiguity is particularly problematic in high-resolution imaging, where fine structural details are heavily dependent on precise phase information. Therefore, phase retrieval is an inherently ill-posed inverse problem, and solving it reliably requires incorporating additional constraints or prior knowledge about the sample.

The importance of phase information cannot be overstated, as it encodes the spatial distribution and depth-related features of the object being imaged. While intensity describes the strength of the signal at each point, phase determines how different parts of the wavefront interfere with each other, influencing the position, shape, and contrast of reconstructed structures. In nanomaterial imaging, where precise localization of atoms or defects is critical, missing or inaccurate phase data can lead to distorted reconstructions or complete misinterpretation of material properties. Thus, recovering the correct phase is central to obtaining a faithful real-space image of the sample. Figure 1.11 illustrates the role of phase information in the inverse image problem (Shechtman et al., 2015).

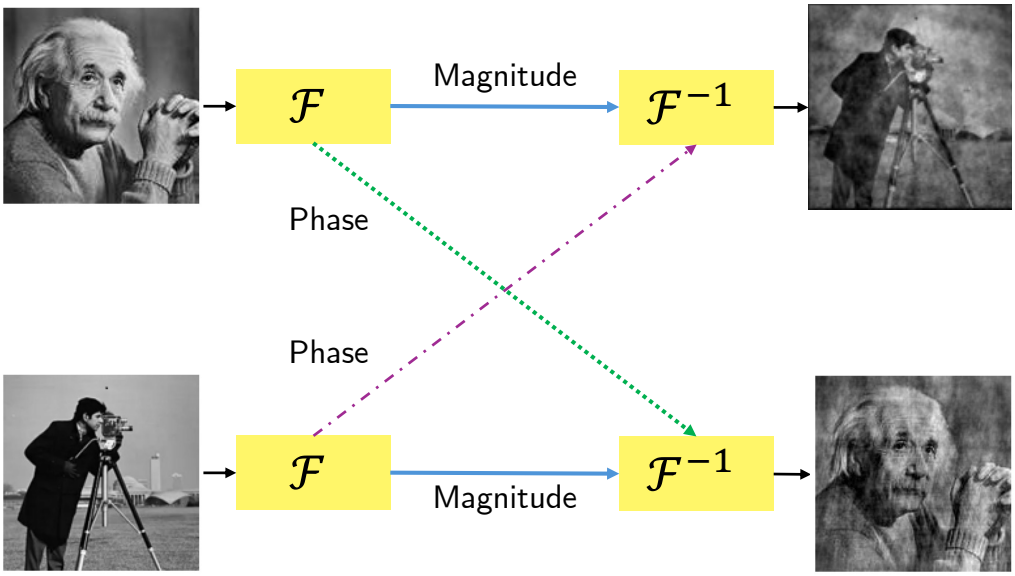


FIGURE 1.11: The significant role of phase information in accurately inverting the image.

1.3.2 The reliability of single-frame CXDI

Reliability in single-frame CXDI, as the computational imaging technique for nanomaterials, begins at the hardware layer, where every component, from X-ray source to detector, governs the informational content of the recorded diffraction pattern. Because the technique captures only a single snapshot with no scanning redundancy, any weakness in the optical train is immediately reflected in the data and cannot be averaged out. A highly coherent X-ray beam is therefore essential. Temporal coherence ensures a well-defined wavelength, while spatial coherence preserves a stable phase relationship across the beam profile. Together, these properties produce sharp interference fringes that faithfully encode high-frequency information needed for nanoscale resolution. Even minor reductions in coherence length or spectral

purity broaden fringe contrast, lowering signal-to-noise ratio (SNR) and decreasing the quality of diffraction.

Other than ensuring the high-quality diffraction, in lens-less diffraction imaging, the hardware design of CXDI should support the effort of breaking the symmetry. The term "symmetry" refers to the existence of mathematical operations, translations, rotations, and inversions that can transform one real-space object into another while leaving the measured diffraction intensity unchanged. When such symmetries are present, the phase retrieval problem becomes non-unique, where multiple structurally distinct objects satisfy the same data constraints, and a solver can converge to any of them. **Breaking the symmetry** means deliberately engineering the experiment so that the forward model admits as few equivalent solutions as possible, thereby steering phase retrieval toward a unique or at least strongly favoured reconstruction. One of the most practical levers for symmetry breaking in hardware design is the aperture. An aperture with non-point symmetry, for example, a triangle, imparts an asymmetric boundary condition on the incident wavefront. Because the illumination profile itself lacks rotational invariance, the resulting far-field diffraction pattern encodes orientation, specific information that cannot be reproduced by simply rotating or conjugating the object. In effect, the asymmetric aperture "tags" the object with a unique angular fingerprint, collapsing whole families of degenerate solutions into a single admissible structure.

Still, from the hardware design perspective, the detector's pixel pitch p, dynamic range (DR), and point-spread function (PSF) also set hard limits on reconstruction fidelity. A coarse pixel grid undersamples high-angle scattering, truncating the Fourier space and capping spatial resolution. Insufficient dynamic range clips bright central peaks or buries weak high-frequency speckles in readout noise. Furthermore, any non-Gaussian PSF smears diffraction spots, reducing contrast. Reliability thus hinges on selecting a detector that pairs small pixels with high bit-depth and a well-characterised, near-Gaussian PSF, alongside flat-field and dark-current corrections to suppress systematic artefacts. In summary, hardware reliability in single-frame CXDI is a multi-factor product: source coherence, aperture geometry, detector performance, and mechanical stability jointly determine whether the single recorded diffraction image contains sufficient, high-quality information for a unique and accurate phase retrieval. Optimizing these elements constitutes the first pillar of computational imaging.

From another point of view, computational design, reconstructing an accurate real-space image from intensity-only diffraction data, presents several fundamental and practical challenges. The core difficulty lies in the nature of the phase retrieval problem itself, which is both nonlinear and undetermined. Without direct access to the phase information in the Fourier domain, the inversion process becomes ambiguous, and multiple candidate solutions can satisfy the same measured intensity pattern, also known as the phase ambiguity problem. Figure 1.12 illustrates the phase ambiguity problem in single-frame CXDI. This non-uniqueness problem is further complicated in single-frame CXDI, where the absence of overlapping measurements, typically used in methods like Ptychography, eliminates the redundancy that would otherwise help stabilize the reconstruction process. Beyond the inherent ambiguity, the success of phase retrieval heavily depends on the quality and completeness of the measured data. In experimental settings, diffraction patterns are often affected by noise, detector limitations, and partial data loss, such as missing central pixels due to beam stops. These imperfections reduce the fidelity of the intensity measurements and propagate uncertainty into the reconstructed image. Even small errors in the

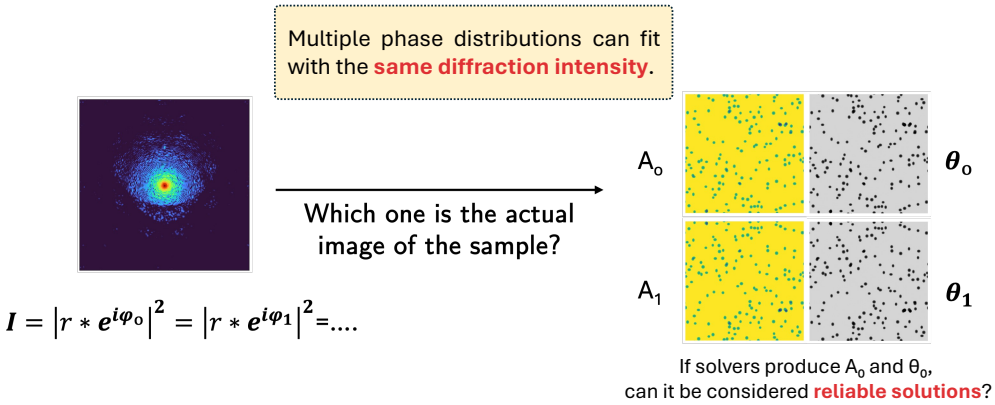


FIGURE 1.12: The concept of phase ambiguity in single-frame CXDI.

measured intensities can lead to significant distortions in phase estimation, thereby degrading the quality of the final reconstruction.

To address the challenges posed by the phase retrieval problem, a wide range of state-of-the-art approaches have been developed, spanning classical iterative algorithms, model-based optimization strategies, and modern machine learning techniques. Traditional iterative methods such as the Gerchberg-Saxton algorithm, Error Reduction (ER), and Hybrid Input-Output (HIO) remain widely used due to their simplicity and physical interpretability (Zalevsky, Mendlovic, and Dorsch, 1996; Jaganathan, Eldar, and Hassibi, 2016; Changliang Guo, S. Liu, and Sheridan, 2015; Fienup, 1982). These algorithms rely on alternating projections between real and reciprocal space, enforcing consistency with measured intensity data and applying real-space constraints such as finite support, non-negativity, or known object priors. While effective under ideal conditions, these methods often suffer from stagnation, sensitivity to initialization, and convergence to local minima when data is noisy or incomplete (Marchesini, 2004; Fienup, 1982). To enhance robustness and solution uniqueness, more recent approaches incorporate adaptive constraint refinement and multi-resolution strategies. The Shrinkwrap algorithm, for instance, updates the support region dynamically during iteration based on the current reconstruction, allowing for better localization of object boundaries (Latychevskaia, 2018). In parallel, optimization-based formulations using convex and non-convex regularizers, such as total variation minimization, sparsity-promoting norms, or maximum entropy priors, have been introduced to reduce ambiguity and suppress artifacts in the reconstructed image. These formulations view phase retrieval as a constrained optimization problem, often solved via gradient-based methods or proximal splitting techniques.

Despite the advantages and the simplicity in employing the physics constraints, reconstruction algorithms must balance between enforcing constraints and avoiding convergence to local minima. Many iterative methods are sensitive to initial conditions and may stagnate or converge to incorrect solutions if the imposed constraints are insufficient or poorly defined. Adaptive strategies like the Shrinkwrap algorithm attempt to refine constraints dynamically, but these approaches still require careful tuning and are not guaranteed to yield physically meaningful results in every case. Another critical challenge involves quantifying the reliability of the reconstructed image. In the absence of ground truth, it is difficult to assess which features of the reconstruction are trustworthy and which may be artifacts of the algorithm or data noise. This issue becomes particularly important when the results are used to infer

structural properties at the nanoscale, where incorrect interpretations can mislead scientific conclusions. Therefore, the development of robust computational frameworks that incorporate uncertainty estimation and solution validation is essential for advancing the practical use of CXDI in material characterization.

Moreover, phase retrieval is therefore an archetypal inverse problem, wherein the task is to reconstruct $O(\mathbf{r})$ from only $|\hat{O}(\mathbf{k})|^2$. The non-linearity of the modulus operator, along with the absence of direct phase information, makes the inversion ill-posed and non-unique. Several mathematically distinct objects may yield the same diffraction intensity, especially in dimensions $D \geq 2$, where the solution space can become uncountably large (J. M. Rodenburg, 2008). Even trivial transformations, such as global phase shifts, spatial translations, or conjugate inversions, leave $I(\mathbf{k})$ unchanged (Jaganathan, Eldar, and Hassibi, 2016; Shechtman et al., 2015). Despite algorithmic advances, the solution obtained may represent only one of many possibilities consistent with the measured data. As such, without sufficient constraints or redundancy, the inverse mapping remains ambiguous. This formulation provides the mathematical foundation upon which deeper challenges, such as solution non-uniqueness and uncertainty quantification, must be addressed in modern CXDI frameworks.

The explanation above underscores a central message, which is that enhancing the reliability of single-frame CXDI cannot be achieved by hardware or computation separately. Well-designed optical devices will not have significant value if the phase recovery algorithm remains vulnerable to its non-uniqueness. Conversely, even the most sophisticated reconstruction code cannot reliably recover information if the diffraction acquired by the imaging process does not fulfill the good criteria of good diffraction. True robustness emerges only when the hardware design deliberately maximizes informational content through coherence control, symmetry-breaking aperture geometry, and the computational design exploits that content via physics-informed constraints, ensemble inference, and uncertainty quantification. By treating these two pillars as co-optimized components of a single imaging system, we move closer to a dependable, interpretable single-frame CXDI pipeline capable of quantitative nanomaterial characterization. Figure 1.13 performs the connection of two pillars of computational imaging to the reliability in single-frame CXDI.

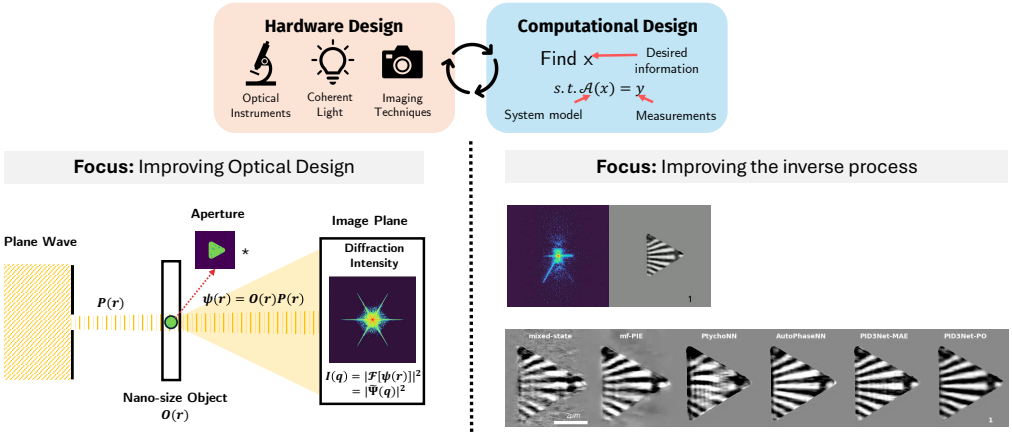


FIGURE 1.13: Two pillars of computational imaging and its connection to the reliability in single-frame CXDI.

1.4 Research Gap and Scientific Challenges

One of the most fundamental challenges in single-frame CXDI is the inherent ambiguity of the phase retrieval problem (Grohs, Koppensteiner, and Rathmair, 2020). Since the recorded diffraction pattern contains only the intensity, corresponding to the squared modulus of the Fourier transform of the object, the critical phase information is entirely lost during acquisition. This results in an inverse problem that is highly ill-posed, where multiple structurally distinct objects may produce identical intensity patterns in reciprocal space. These ambiguities include both trivial transformations (such as global phase shifts, spatial translations, and conjugate inversion) and non-trivial transformations such as zero-flipping and complex-valued reconstructions in higher dimensions. Consequently, without additional information or constraints, image reconstruction becomes fundamentally ambiguous and unreliable.

This non-uniqueness challenge is especially acute in single-frame CXDI, where only one diffraction image is recorded without overlapping scans or multi-frame redundancy. Unlike ptychography, which provides additional spatial context through overlapping probe positions, single-frame CXDI relies solely on a single intensity snapshot, making it more vulnerable to degenerate or unstable solutions. Even when standard constraints such as finite support, non-negativity, or sparsity are applied, they may not be sufficient to guarantee uniqueness or stability, particularly for non-crystalline, noisy, or weakly scattering samples.

The effort to overcome this problem came from the study of Takazawa et.al. Improving reconstruction reliability may require optimizing not only the computational inference model but also the physical system design, such as the aperture geometry that shapes the illumination profile (Takazawa, Kang, et al., 2021; Kang et al., 2021). Prior studies have shown that triangular apertures can enhance phase retrieval performance by breaking inversion symmetry and improving uniqueness conditions in the reconstruction process. The comparison of several geometric shapes of apertures in the effort of breaking the symmetry can be examined in Figure 1.14 (Takayama et al., 2021; Kang et al., 2021). However, these theoretical advantages are met with

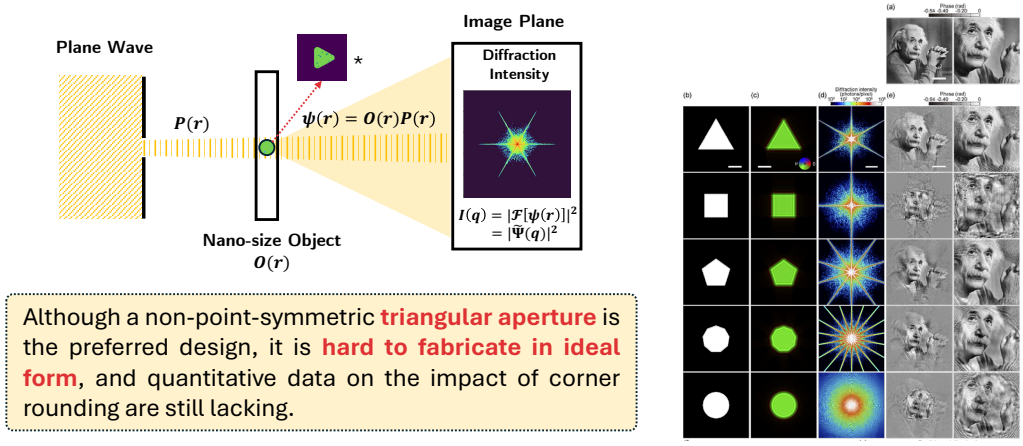


FIGURE 1.14: The existing study of enhancing optical design in single-frame CXDI by tuning some options of aperture design.

significant fabrication challenges, as manufacturing apertures with perfectly sharp corners is technically difficult and often impractical in real experimental setups. This gap between theoretical ideal and experimental feasibility highlights the need for further investigation into relaxed geometric designs, such as triangular apertures

with rounded corners, that can maintain desirable diffraction characteristics while being more robust to fabrication limitations.

Efforts on the computational design front have likewise evolved. Traditional phase-retrieval solvers, iterative schemes such as Error Reduction (ER), Hybrid Input–Output (HIO), and extended Ptychographic Iterative Engine (ePIE), enforce a rigid set of support and Fourier modulus constraints until they converge to one candidate reconstruction image. More recently, data-driven approaches have been introduced to mitigate noise sensitivity and accelerate inference. Supervised networks such as PtychoNN (Cherukara et al., 2020) learn a direct diffraction-to-image mapping from large synthetic libraries, whereas the unsupervised PID3Net Vu et al., 2025 embeds physics-informed priors, dropout-based stochasticity, and domain-specific constraints to recover highly resolved, high-contrast reconstructions. These machine learning solvers markedly improve visibility and robustness, yet they still operate within an under-determined solution space. In practice, each algorithm converges to a single admissible minimum, often the first one encountered, implicitly treating that local optimum as the "true" solution. Consequently, the foundational ambiguity persists: neither iterative nor deep learning methods, by themselves, guarantee that alternative, phase-equivalent solutions have been excluded. From a data science perspective, such behaviour resembles overfitting. The algorithm fits the measured diffraction intensities perfectly, yet offers no evidence that alternative, equally valid solutions have been ruled out.

A natural safeguard against overfitting is cross-validation. In conventional machine learning workflows, the dataset is partitioned into disjoint folds, a model is trained on one subset, and its performance is verified on the held-out portion to detect overfitting and assess generalization. Single-frame CXDI, however, poses a unique obstacle. Its entire diffraction pattern encodes a single unknown object via a non-linear Fourier relationship. Randomly splitting a diffraction would break the physical consistency between real and reciprocal space, violating the forward model and corrupting the phase-retrieval constraints. Therefore, a physics-respecting alternative is required. One that perturbs the data in a controlled manner, preserves the governing wave-propagation equations, and still exposes how sensitive a solver is to missing information.

Thus, a more comprehensive strategy is needed, one that explicitly targets both reconstruction reliability and ambiguity reduction, while enhancing interpretability. Reducing phase ambiguity demands methods that do not merely output a single reconstruction, but instead explore the solution space, identify critical regions in the diffraction data, and quantify the stability of reconstructed features.

Modern scientific inference increasingly moves away from reporting a single deterministic "best" answer and toward modeling a distribution of plausible solutions. Many inverse problems are ill-posed: data are limited, noisy, and compatible with multiple explanations. A distributional view, grounded in Bayesian reasoning, treats each reconstruction as one sample from a posterior rather than as the truth itself. This shift enables researchers to quantify uncertainty, expose where the data underconstrains the solution, and make decisions that openly account for risk and ambiguity.

As shown in the Figure 1.15, in seismology, full waveform inversion exemplifies this paradigm. Instead of publishing one velocity model of the subsurface, ensemble-based approaches generate many models that all fit the recorded waves, then analyze their spread to map spatially varying uncertainty. The outcome is not only a geologically coherent image but also a reliability layer that highlights where additional data

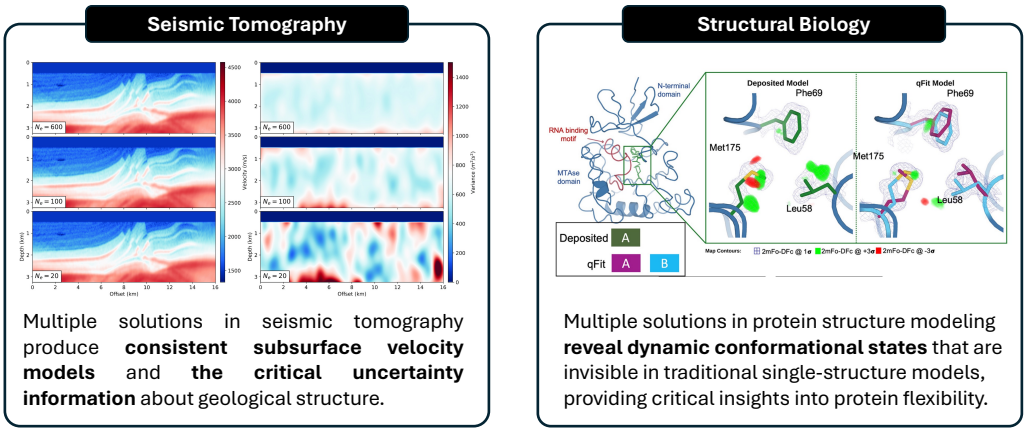


FIGURE 1.15: The existing study of shifting the paradigm from seeking a single deterministic solution to modeling a distribution of plausible solutions.

or constraints are needed (Thurin, Brossier, and Métivier, 2019). Structural biology shows a parallel development. Proteins rarely occupy a single static conformation; they populate families of states. Automated multiconformer and ensemble modeling explain the experimental electron density with multiple conformations, revealing functionally relevant flexibility that a single structure would hide (Wankowicz et al., 2024).

Taken together, these examples capture a broader methodological shift: scientific claims are strengthened when we report both representative reconstructions and the distribution from which they are drawn. Modeling solution ensembles—rather than a lone optimum—supports uncertainty quantification, guides adaptive experiment design, and ill-posed settings.

Bridging this gap calls for a mechanism to create multiple possible solutions from one solver, employing ensemble inference frameworks, spatial attribution techniques, and ambiguity-aware evaluation metrics. Each is grounded in physical modeling and capable of generalizing beyond synthetic simulations. Ultimately, addressing these dual challenges, reliable reconstruction results and ambiguity-aware, will significantly advance CXDI as a trustworthy and explainable nanoscale imaging technique. Such advances are essential not only for improving reconstruction fidelity but also for enabling deeper scientific interpretation and more robust knowledge extraction from complex diffraction data.

1.5 Research Objectives and Methodology

Based on the research gap and scientific challenges previously outlined, the primary objective of this study is to establish a computational framework for single-frame CXDI that enhances both the reliability of image reconstruction and the reduction of uncertainty arising from phase ambiguity. Recognizing that the inverse problem in CXDI is fundamentally ill-posed and non-unique, this research does not treat reconstruction solely as an algorithmic challenge, but as a coupled system, as shown in the pillars of computational imaging, involving both optical design and epistemic computational modeling. This effort is linking the physical imaging process to the extraction of structured knowledge from uncertain measurements.

Based on that aim, the first objective is to **improve reconstruction accuracy by simultaneously optimizing the aperture geometry and phase-retrieval algorithms to improve image fidelity**, specifically by investigating the effect of the corner radius of the aperture on the accuracy and stability of phase retrieval results. Prior studies have demonstrated that triangular apertures produce superior diffraction patterns for phase retrieval due to their asymmetric support, which helps break inversion symmetry and reduce ambiguity. However, in practical applications, fabricating apertures with perfectly sharp corners is technically challenging and often impractical. This motivates the need to examine how varying the corner radius, a geometric relaxation of the ideal triangle shape, affects the reconstruction quality. Building on the findings in (Mukharil Bachtiar et al., 2024). This study evaluates how systematic modifications to the corner radius influence the spatial coherence, diffraction fringe contrast, and downstream phase retrieval performance. The analysis is performed through large-scale simulations, where a range of radius values is applied while maintaining a constant aperture area to ensure comparability. For each configuration, iterative and deep learning phase retrieval are conducted, and performances are assessed using statistical metrics. This optimization is implemented within a data science framework, where reconstruction outcomes are systematically profiled across geometric configurations to identify a sweet spot in the trade-off between physical realizability and computational performance. By treating aperture geometry not as a fixed hardware constraint but as an optimizable design parameter, this objective advances both the experimental feasibility and the robustness of the CXDI pipeline.

The second objective is to **establish a physics-respecting cross-validation framework that mitigates overfitting in phase-retrieval solvers, unifies all plausible reconstructions into a single ensemble estimate, and generates pixel-wise variance maps that explicitly visualize phase ambiguity**. To this objective, we introduce a sliding–window skip region masking protocol that selectively omits real- or reciprocal-space pixels while preserving the forward diffraction model. Each masked data set acts as a pseudo-validation fold. The solver is trained on the unmasked region and implicitly tested on the masked region when the modulus constraint is re-imposed. By cycling the window across the CXDI dataset, we create a population of alternative reconstructions that expose how sensitive each spatial-frequency band is to missing information. These diverse reconstructions, obtained from both iterative algorithms and physics-informed deep learning, are then aggregated within an **ensemble phase-retrieval** framework. The voting on the reconstruction results yields a consensus image that is statistically more reliable than any single solver run, while the pixel-wise variance quantifies residual ambiguity. Low-variance regions mark features that are firmly data-driven, whereas high-variance zones reveal where phase information is under-constrained or extrapolated. In this mechanism, the framework simultaneously holds back solver overfitting via cross-validated masking, boosts reconstruction fidelity through ensemble voting, and delivers an interpretable ambiguity map that links phase ambiguity to concrete spatial structures, thereby advancing both the reliability and explainability of single-frame CXDI.

Together, these objectives aim to unify physical system modeling, statistical inference, and constraint-based reconstruction into a coherent pipeline. This pipeline is not only designed to improve numerical performance, but also to facilitate a comprehensive scientific understanding of single-frame CXDI. In doing so, the work contributes to the broader discourse on explainable and uncertainty-aware imaging at the nanoscale by using a data science approach. Based on both objectives, we performs the research methodology as shown in Figure 1.16

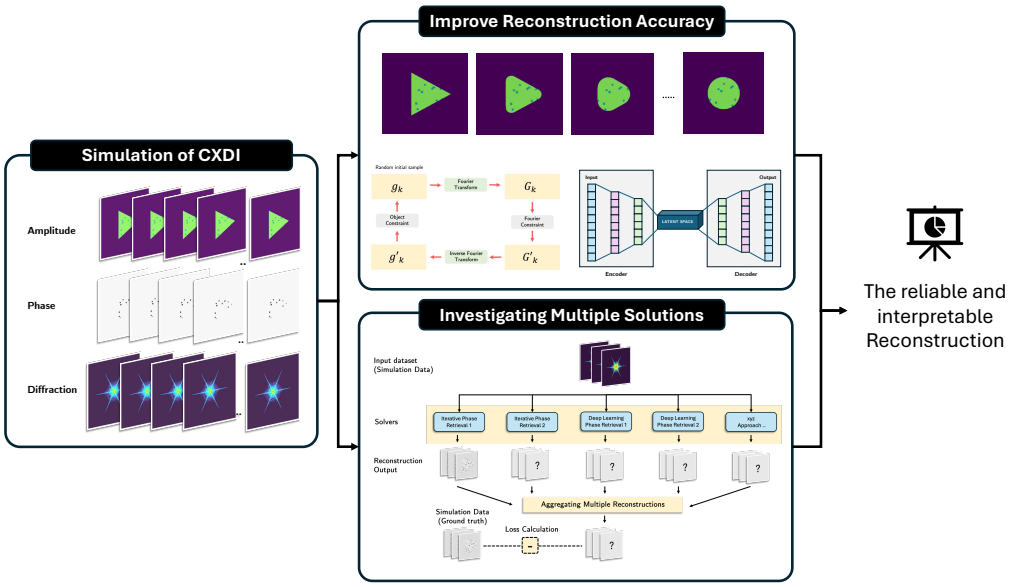


FIGURE 1.16: The research methodology of the study.

1.6 Structure of the Dissertation

This dissertation is organized into six chapters and a set of appendices. Each chapter is structured to progressively guide the reader from theoretical foundations to practical implementation and interpretation, culminating in a novel, uncertainty-aware reconstruction strategy for single-frame Coherent X-ray Diffraction Imaging (CXDI).

- **Chapter 1: Introduction**
  This chapter introduces the scientific context and motivation of the research. It provides a broad overview of imaging in materials science, the significance of computational imaging, and the emergence of CXDI as a lensless nanomaterial imaging technique. It also discusses the challenges of phase retrieval and non-uniqueness, leading to a formulation of the research objectives. The chapter closes with a description of the dissertation structure.

- **Chapter 2: Fundamental Concepts of Single-Frame CXDI and Phase Retrieval Problem**
  This chapter presents the conceptual and mathematical background underlying the CXDI imaging pipeline. It covers topics such as Fourier optics, diffraction modeling, the phase retrieval problem, and commonly used iterative and deep learning-based algorithms. It also reviews physics-informed constraints and introduces ensemble learning concepts from machine learning, highlighting their relevance to inverse problems and uncertainty quantification.

- **Chapter 3: Geometric Optimization of Aperture Design in Enhancing the Single-Frame CXDI Reconstruction**
  This chapter presents the first study, which investigates how variations in aperture geometry, specifically the corner radius of a triangular aperture, impact phase retrieval performance. Both iterative (ePIE) and deep learning-based models are applied to simulated diffraction patterns. The results demonstrate the relationship between optical system design and reconstruction quality, offering insights for optimizing hardware constraints and improving reliability.

- **Chapter 4: Ensemble Phase Retrieval for Cross-Validated Reliability and Phase Ambiguity Mapping of Single-Frame CXDI Reconstruction**
  This study presents a unified ensemble framework that treats skip region masking as a physics-respecting form of cross-validation for single-frame CXDI. By systematically omitting sliding windows of pixels in either real or reciprocal space, the method generates a family of perturbed diffraction data sets that drive a diversity of reconstructions from both iterative and deep-learning solvers. Voting aggregation yields a consensus image, while the pixel-wise variance map exposes where phase information is firmly data-driven and where ambiguity remains. The chapter compares different reconstruction algorithms in terms of stability, robustness, and interpretability, and demonstrates how uncertainty maps can support spatially localized confidence assessment in CXDI reconstructions.

- **Chapter 5: Conclusion and Limitation**
  This chapter synthesizes the findings from both studies and reflects on their contribution to improving the explainability and practical reliability of CXDI. It outlines the main limitations of the approach and proposes future directions, including validation with experimental data, expansion to time-resolved imaging, and deeper integration of physics-informed learning strategies.

- **Chapter 6: Publication List**
  This chapter lists the peer-reviewed publications and conference presentations that have resulted from the research conducted in this dissertation. It provides documentation of scholarly dissemination and recognition within the research community.

- **Appendices**
  The appendices provide detailed implementation data, including simulation parameters, algorithm configurations, and additional visual results. Appendix A covers the full simulation setup and parameterization. Appendix B includes supplementary reconstructions and animations supporting both aperture optimization and ensemble phase retrieval studies.

This structure reflects a deliberate progression from foundational knowledge to applied methods, experimental insights, and scientific contributions. It also demonstrates how computational modeling, physical optics, and ambiguity interpretation can be co-integrated to advance the goals of explainable computational imaging within the nanomaterial imaging domain or even the broader field of Knowledge Science.

Chapter 2

Fundamental Concepts of Single-Frame CXDI and Phase Retrieval Problem

The advancement of computational imaging techniques, particularly in the domain of lensless Coherent X-ray Diffraction Imaging (CXDI), requires a solid understanding of the physical principles, mathematical frameworks, and computational strategies that govern image formation and reconstruction. Given that this dissertation is situated at the intersection of physics, data science, and inverse problem theory, it is essential to establish a coherent conceptual foundation that spans these domains. This foundation must also be accessible to readers from diverse disciplinary backgrounds, particularly those within the field of Knowledge Science, where emphasis is placed not only on the correctness of computational models, but also on their explanatory power and interpretability.

This chapter introduces the fundamental concepts, terminologies, and theoretical models that underlie the development of phase retrieval and uncertainty-aware image reconstruction in CXDI. The discussion begins by framing imaging as a process of knowledge construction, in which the aim is not merely to produce visual representations but to infer and validate structural information from incomplete and indirect measurements. It proceeds to describe the physical and mathematical characteristics of coherent diffraction, the nature of the inverse problem it poses, and the algorithmic strategies used to recover meaningful information from intensity-only data.

Particular attention is given to the challenges of phase retrieval, especially the problem of phase ambiguity, as well as to the epistemic implications of ambiguity and interpretability in CXDI image reconstruction. These challenges motivate the core contribution of this dissertation: the development of an ensemble-based phase retrieval framework that integrates physical constraints, algorithmic diversity, and uncertainty quantification to improve both the reliability and interpretability of phase retrieval outcomes. The conceptual building blocks introduced in this chapter serve as the theoretical scaffolding upon which this novel model is constructed, ensuring that the methodology is technically robust and grounded in a broader understanding of inference, reliability, and scientific explanation.

2.1 Single-Frame CXDI: Definitions and Principles

Single-frame CXDI is an imaging Technique used to obtain structural information from an object by utilizing the diffraction pattern produced by coherent light or x-ray scattered by the object (Nakasako et al., 2020). This technique is a lensless imaging technique that reconstructs the internal structure of a sample by analyzing

the diffraction pattern produced when a coherent X-ray beam illuminates the sample. Unlike traditional imaging methods that use optical lenses to form images, CXDI operates in the far-field regime and records only the intensity of the scattered X-rays on a two-dimensional detector. Because of the necessity of a far-field regime, single-frame CXDI fundamentally relies on the principles of Fraunhofer diffraction, which is the far-field approximation of wave propagation. The Fraunhofer equation is a mathematical equation used to describe the distribution of diffraction patterns at long distances from an aperture or object (Uozumi, Kimura, and Asakura, 1990). In the context of optical and X-ray physics, Fraunhofer diffraction describes the pattern formed when a coherent wavefront interacts with an object and is observed at a sufficiently large distance, or equivalently, when a lens is used to bring the far-field condition into focus at the detector plane. Under this approximation, the recorded diffraction pattern is mathematically equivalent to the squared magnitude of the Fourier transform of the object's complex-valued transmission function. The formula of Fraunhofer diffraction can be examined as follows:

$$E(x_0, y_0) \propto \iint e^{\left(-\frac{jk}{z}(x_0 x_1 + y_0 y_1)\right)} \cdot \text{Aperture}(x_1, y_1)\, dx_1\, dy_1, \tag{2.1}$$

The equation above represents the complex field $E(x_0, y_0)$ observed at a point (x_0, y_0) in the detector plane under the Fraunhofer diffraction approximation. This expression is central to single-frame CXDI, where the sample acts as an aperture that modulates the incoming coherent wavefront. The integral describes the superposition of all point contributions from the aperture, with each term modulated by a phase factor determined by the spatial coordinates and the wavenumber $k = \frac{2\pi}{\lambda}$.

The function $\text{Aperture}(x_1, y_1)$ defines the amplitude transmission of the sample at each point (x_1, y_1) in the object plane, and the exponential term accounts for the phase delay accumulated during propagation to the detector. The far-field or Fraunhofer condition ensures that the resulting diffraction pattern is equivalent to the Fourier transform of the aperture function. Therefore, the measured intensity on the detector is given by $I(x_0, y_0) = |E(x_0, y_0)|^2$, which contains only the amplitude information, while the phase component is lost. The illustration of the Fraunhofer regime of diffraction can be examined in Figure 2.1.

In CXDI, this intensity-only data forms the basis of image reconstruction through computational phase retrieval. Equation 2.1 thus establishes a direct mathematical link between the physical scattering process and the computational inverse problem. It highlights the fact that the observed diffraction pattern is a non-invertible transformation unless we employ the additional assumptions or constraints. Understanding this relationship is key to modeling both the forward and inverse aspects of coherent diffraction imaging. The experimental geometry is specifically configured to satisfy the Fraunhofer condition, ensuring that the intensity pattern measured on the detector accurately represents the spatial frequency content of the object. This means that the measured intensity $I(\mathbf{k})$ corresponds to:

$$I(\mathbf{k}) = |\mathcal{F}\{O(\mathbf{r})\}|^2, \tag{2.2}$$

where $O(\mathbf{r})$ is the object function in real space, and $\mathbf{k}$ denotes the spatial frequency in reciprocal space. The Fourier transform $\mathcal{F}$ arises directly from the Fraunhofer diffraction model, which assumes planar wavefronts and paraxial propagation, conditions typically met in coherent X-ray experiments using synchrotron or XFEL sources.

This relationship is crucial for the feasibility of single-frame CXDI as an imaging

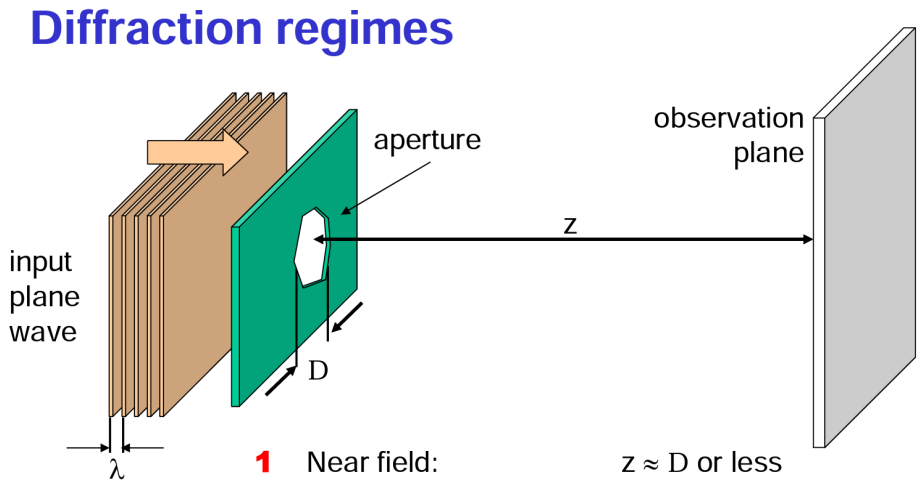


FIGURE 2.1: Illustration of Fraunhofer diffraction geometry in single-frame CXDI. The Fraunhofer approximation holds when the propagation distance z satisfies $z \gg \frac{\pi D^2}{\lambda}$ and $z \gg D$, where D is the characteristic size of the aperture and λ is the X-ray wavelength. Under this condition, the detector plane is treated as the Fourier domain, enabling lensless imaging via computational phase retrieval.

method. Because the detector captures only the intensity (magnitude squared) of the Fourier-transformed wave, the missing phase information must be reconstructed computationally in a process known as phase retrieval. The assumption that the far-field pattern represents a clean Fourier domain measurement is what enables this reconstruction to proceed under a mathematically consistent model. Under the Fraunhofer approximation, also known as the far-field regime, the diffraction pattern observed at a distant detector can be mathematically interpreted as the Fourier transform of the object's transmission function. This approximation becomes valid when the propagation distance z from the object to the detector is much larger than both the aperture size D and the term $\frac{\pi D^2}{\lambda}$, where λ is the wavelength of the illuminating wave.

This relationship is foundational in single-frame Coherent X-ray Diffraction Imaging (CXDI), where no lenses are used and the detector captures only the far-field intensity pattern. Because the intensity corresponds to the squared magnitude of the Fourier transform of the object, reconstructing the object requires recovering the lost phase information and then applying an inverse Fourier transform. Thus, the Fraunhofer approximation provides the critical theoretical link between the physical scattering process and the computational image reconstruction pipeline, enabling lensless imaging to be formulated as a Fourier-domain inverse problem.

2.2 Forward Process in Single-Frame CXDI

As mentioned in the chapter 1, single-frame CXDI performs two groups of processes, namely forward and inverse processes. The forward process in single-frame CXDI describes how a coherent X-ray beam interacts with a sample and produces a measurable diffraction pattern on a far-field detector. In simple terms, the forward process focuses on how to produce a dataset in CXDI, which is a diffraction image. With the lensless imaging setup, no imaging optics are used between the sample and

the detector. The image is formed not through direct magnification, but via wave propagation governed by diffraction principles. The resulting pattern encodes the structural information of the object in the form of spatial frequency content, requiring computational methods to recover the real-space image.

In assumption there are no extra optical instruments between the source of light and the sample or the sample and the detector, let the sample be represented by a complex transmission function $O(\mathbf{r})$, where $\mathbf{r} = (x_1, y_1)$ denotes the coordinates in the object plane. The sample can be considered as the object in the real space. When a coherent X-ray beam illuminates the sample, the wave exiting the sample is given by the product of the object function and the illumination probe $P(\mathbf{r})$:

$$\psi(\mathbf{r}) = O(\mathbf{r}) \cdot P(\mathbf{r}). \tag{2.3}$$

This function $\psi(\mathbf{r})$ represents the exit wavefield at the sample plane, containing both amplitude and phase information determined by the object and the beam profile. The probe can be transformed by utilizing some extra optical instruments, such as a Fresnel Zone Plate. In the far-field (Fraunhofer) regime, the field observed at the detector plane is the scaled Fourier transform of the exit wave. Denoting the spatial frequency coordinates as $\mathbf{q} = (q_x, q_y)$ in reciprocal space, the detector records the following complex field:

$$\Psi(\mathbf{q}) = \mathcal{F}\{\psi(\mathbf{r})\} = \mathcal{F}\{O(\mathbf{r}) \cdot P(\mathbf{r})\}, \tag{2.4}$$

where $\mathcal{F}\{\cdot\}$ denotes the two-dimensional Fourier transform.

However, in practice, only the intensity of this field can be recorded by the detector:

$$I(\mathbf{q}) = |\Psi(\mathbf{q})|^2 = |\mathcal{F}\{O(\mathbf{r}) \cdot P(\mathbf{r})\}|^2. \tag{2.5}$$

This intensity $I(\mathbf{q})$ forms the primary data collected in single-frame CXDI. Importantly, the detector captures no direct information about the phase of $\Psi(\mathbf{q})$, which is essential for reconstructing the object in real space. This missing information makes the inverse problem ill-posed and necessitates the use of phase retrieval algorithms.

In realistic experimental configurations, the probe function $P(\mathbf{r})$ is shaped by the optical elements placed upstream of the sample, typically a combination of an aperture and a Fresnel zone plate (FZP). The aperture determines the spatial coherence and shape of the beam, while the FZP focuses the X-rays into a desired spot size. In the model described in the aperture analysis, the total illumination $P(\mathbf{r})$ is obtained from wave propagation through these optical components, and often modeled via the Angular Spectrum Method (ASM) or Fresnel propagation:

$$P(\mathbf{r}) = \text{Propagate}_{\text{aperture}\to\text{sample}}\{A(\mathbf{r})\}, \tag{2.6}$$

where $A(\mathbf{r})$ is the shape of the aperture, and the propagation function depends on distance, wavelength, and the specific diffraction regime. In simulations, $P(\mathbf{r})$ can be precomputed and fixed for use in synthetic forward models.

Overall, the forward model in single-frame CXDI can be concisely written as:

$$I(\mathbf{q}) = |\mathcal{F}\{O(\mathbf{r}) \cdot P(\mathbf{r})\}|^2, \tag{2.7}$$

which serves as the basis for both numerical simulations and inverse reconstruction. This model assumes ideal coherence and noiseless propagation, but in practice, detector noise, partial coherence, and sampling effects must also be considered.

Understanding this forward process is critical not only for data generation but also for evaluating reconstruction fidelity and for developing uncertainty-aware phase retrieval algorithms.

2.3 Fourier Transform in CXDI

The Fourier transform plays a critical role in diffractive imaging, as we can consider based on the formulation of Fraunhofer diffraction. This calculation relates a sample in the real space and its exit-wave spectrum in reciprocal space, also known as the Fourier or diffraction space. We only discussed the 2D discrete Fourier Transform since we only worked with the image, and because a computer only works with discrete functions. In the context of single-frame CXDI, the image of the object, typically represented as a complex-valued function $O(x, y)$ in real space, is not directly observed. Instead, the imaging process captures the far-field diffraction pattern, which corresponds to the squared magnitude of the object's Fourier transform. Formally, the discrete 2D Fourier transform (DFT) of $O(x, y)$ is given by:

$$\hat{O}(u, v) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} O(x, y) \cdot e^{-2\pi i \left(\frac{ux}{M} + \frac{vy}{N}\right)}, \tag{2.8}$$

where (u, v) denotes spatial frequency indices in the reciprocal space, and $M \times N$ is the image size. The corresponding inverse Fourier transform, used to recover the image from the diffraction spectrum (once the phase is known), is:

$$O(x, y) = \frac{1}{MN} \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} \hat{O}(u, v) \cdot e^{+2\pi i \left(\frac{ux}{M} + \frac{vy}{N}\right)}. \tag{2.9}$$

The Fourier domain representation $\hat{O}(u, v)$ is inherently complex-valued, meaning it contains both real and imaginary components at each spatial frequency point. To work effectively with this complex data, especially in the context of image reconstruction and interpretation. It is essential to understand two fundamental ways of expressing complex numbers: the **rectangular form** and **the polar form**. In the rectangular form, each complex value in the Fourier domain is described in terms of its real and imaginary components:

$$\hat{O}(u, v) = \Re(u, v) + i \cdot \Im(u, v), \tag{2.10}$$

where $\Re(u, v)$ represents the real part and $\Im(u, v)$ represents the imaginary part. This form is especially convenient for implementation in computational algorithms, as it aligns naturally with array-based representations in numerical Fourier transforms and allows for straightforward addition and multiplication of complex-valued matrices.

However, the rectangular form often conceals the physical meaning embedded in the diffraction pattern. For this reason, the *polar form* is more intuitive and widely used in physical interpretation. In polar form, each complex number is expressed as a magnitude (modulus) and a phase angle:

$$\hat{O}(u, v) = |\hat{O}(u, v)| \cdot e^{i\phi(u,v)}, \tag{2.11}$$

where $|\hat{O}(u, v)|$ denotes the amplitude (the strength of a particular spatial frequency component), and $\phi(u, v)$ denotes the phase (which determines how these frequency

components interfere to form a spatial image). In imaging systems such as CXDI, the phase component $\phi(u, v)$ is essential for recovering the positional and structural information of the object in real space.

This polar decomposition becomes particularly relevant when considering the nature of diffraction data. In single-frame CXDI, the detector only records the intensity of the diffracted wave, which corresponds to the squared amplitude of the Fourier transform:

$$I(u, v) = |\hat{O}(u, v)|^2. \tag{2.12}$$

As a result, only the magnitude component $|\hat{O}(u, v)|$ is available from measurement, while the phase $\phi(u, v)$ is completely lost. This loss renders the inverse Fourier transform inapplicable in its direct form, because the full complex information needed to reconstruct $O(x, y)$ is incomplete. The missing phase is what leads to the classical phase retrieval problem in CXDI.

Thus, understanding the distinction between rectangular and polar representations is more than a mathematical formality. It directly affects how data is interpreted, how constraints are applied, and how reconstructions are conducted. The rectangular form is preferred in algorithmic implementations for manipulating wavefields numerically, while the polar form provides the interpretive framework through which physical phenomena like diffraction, interference, and image formation are understood and modeled. In CXDI, the transition from intensity-only measurements to complete image reconstruction hinges on our ability to reason about and reconstruct the missing phase component in the polar domain. This loss of phase means that although we can directly compute $|\hat{O}(u, v)|$ from the measured intensity $I(u, v)$, we cannot reconstruct the object $O(x, y)$ without recovering or approximating the missing phase. This phase retrieval step is essential to apply the inverse Fourier transform and obtain a physically meaningful image. The difference between what is known (the modulus) and what is unknown (the phase) fundamentally defines the ill-posedness of the CXDI inverse problem. In practice, both rectangular and polar forms of the Fourier spectrum are utilized during the computational phase retrieval process. Iterative algorithms typically switch between real-space and Fourier-space representations, using the polar form to enforce measured amplitude constraints and updating the phase until a self-consistent solution is found. Understanding the dual representation of the Fourier transform, both in its discrete computational implementation and in its physical interpretation, is critical to grasping how CXDI systems convert raw diffraction measurements into interpretable structural information.

A foundational concept in coherent diffraction imaging, especially in single-frame CXDI, is the reciprocal relationship between the spatial domain (real space) and the frequency domain (Fourier space). This relationship emerges from the mathematical structure of the Fourier transform and governs the core principle behind how diffraction data is collected and computationally inverted. In real space, the object $O(x, y)$ represents a distribution of physical properties such as electron density or transmission phase. When illuminated by a coherent X-ray beam, this object scatters the incident wavefront in a manner that depends on its internal structure. The scattered wave propagates and reaches a detector placed at a far-field distance, where the recorded intensity corresponds to the squared magnitude of the Fourier transform of the object.

The Fourier domain, or reciprocal space, encodes the spatial frequency content of the object. Low-frequency components, which correspond to large-scale smooth structures in real space, appear near the center of the diffraction pattern, while high-frequency components, arising from fine details such as edges or sharp transitions,

are distributed further from the center. The resolution of the reconstructed image is therefore directly tied to the extent of high-frequency components captured in the diffraction pattern. The inverse Fourier transform allows reconstruction of the object from its complex spectrum as shown in this equation:

$$O(x,y) = \mathcal{F}^{-1}\{\hat{O}(u,v)\}, \tag{2.13}$$

but in practice, only the intensity $I(u,v) = |\hat{O}(u,v)|^2$ is measured, and the phase $\phi(u,v)$ is lost.

Crucially, the real and reciprocal spaces are not just duals. They are reciprocally scaled. Let Δr be the sampling interval in real space and Δq the spacing in reciprocal space (Candes and Wakin, 2008; Van der Veen and Pfeiffer, 2004). These are inversely related:

$$\Delta r = \frac{1}{k}, \qquad \Delta q = \frac{1}{L}, \tag{2.14}$$

where k is the frequency bandwidth and L is the object size or field of view. A smaller Δr (finer sampling in real space) results in a broader frequency range in Fourier space, enabling the capture of higher-resolution diffraction features. Conversely, a larger real-space object yields more densely spaced frequency components (smaller Δq), improving frequency resolution but reducing bandwidth.

This reciprocal grid relationship is tightly linked to the **Shannon Sampling Theorem**, which states that to reconstruct an object accurately, it must be sampled at least twice the highest frequency it contains (Arsenault and Chalasinska-Macukow, 1983). If this condition is not met, aliasing occurs, resulting in distortions and ambiguities in the final image. In CXDI, the real-space grid (object pixelation), detector resolution, and propagation distance collectively determine how reciprocal space is sampled, and thus the quality of the reconstructed structure. Figure 2.2 illustrates the reciprocal space between real space and Fourier space.

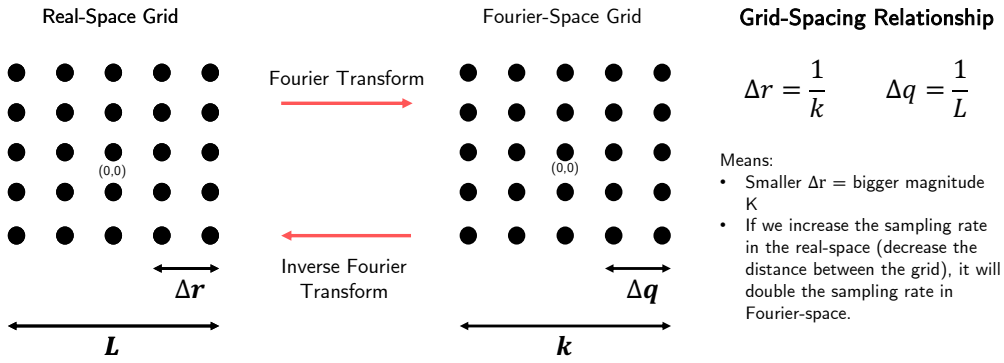


FIGURE 2.2: Illustration of reciprocal relationship between real space and Fourier space in CXDI.

Understanding this concept is essential not only for the physical interpretation of diffraction patterns but also for designing effective reconstruction algorithms. During phase retrieval, algorithms iteratively switch between the real and Fourier domains, applying structural constraints in real space and enforcing amplitude consistency in reciprocal space. This iterative bridging reflects the operational essence of the reciprocal relationship, making Fourier theory not just a mathematical foundation but a practical and physical framework in coherent imaging. The clearer our understanding of this duality, the more effectively we can extract meaningful, high-resolution

structural knowledge from intensity-only diffraction measurements.

2.4 Phase Retrieval Algorithms

After discussing the fundamental concepts in the forward process in CXDI, the further discussion should focus on the inverse process. This inverse problem, commonly referred to as phase retrieval, is central to coherent diffraction imaging because it addresses the key limitation of lensless systems. While the forward process encodes the structural information of a sample into the diffraction pattern via a Fourier transform, the lack of recorded phase in the reciprocal space renders direct inversion infeasible. To overcome this, numerous computational strategies have been developed to estimate the missing phase and reconstruct the object in real space. These phase retrieval methods can be broadly categorized into iterative phase retrieval algorithms, rooted in signal processing and constraint enforcement, and data-driven approaches, based on deep learning and statistical inference. The following section provides a review of both iterative and deep learning phase retrieval techniques, emphasizing their algorithmic principles, operational assumptions, and relevance to single-frame CXDI reconstruction.

2.4.1 Iterative Phase Retrieval

Iterative phase retrieval methods form the foundation of classical approaches to solving the inverse problem in coherent diffraction imaging. These methods are designed to recover the missing phase information by enforcing consistency between the measured intensity data in Fourier space and prior knowledge or constraints in real space. The central idea is to iteratively refine an estimate of the object by alternating between real and reciprocal domains using the Fourier transform and its inverse, while applying physically meaningful constraints, such as support boundaries, non-negativity, or known object features, at each step. In general, Figure 2.3 shows how the flow of the iterative algorithm works, based on the Gerchberg-Saxton algorithm (GS algorithm).

Iterative phase retrieval methods are classical approaches used to solve the inverse problem in coherent diffraction imaging, including single-frame CXDI. At their core, these methods are designed as optimization procedures that aim to reconstruct a complex-valued object from incomplete Fourier measurements, specifically, from intensity-only data where the phase is missing. The problem can be formulated as a non-convex optimization task, which is to find an object $O(x, y)$ whose Fourier transform has a magnitude matching the experimentally measured intensity, and which satisfies a set of known physical or structural constraints in real space. Mathematically, the objective is to minimize a loss function that quantifies the inconsistency between the modulus of the current estimate's Fourier transform and the observed diffraction intensity. This is often expressed as:

$$\min_{O(x,y)} \left| \mathcal{F}\{O(x,y)\}| - \sqrt{I(u,v)} \right|^2, \tag{2.15}$$

subject to constraints such as finite support, non-negativity, or prior knowledge about the object. Because the phase is not observed, the optimization landscape is highly non-convex, and solutions are sensitive to initialization and algorithmic strategy. To solve this problem, iterative algorithms alternate between the real and reciprocal

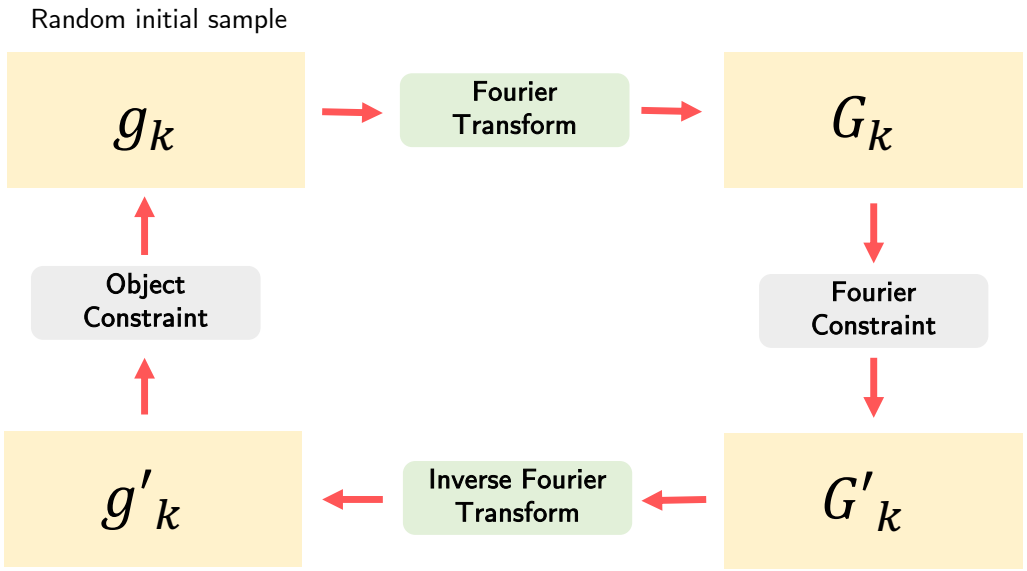


FIGURE 2.3: General workflow of iterative phase retrieval in single-frame CXDI.

domains using the Fourier transform and its inverse. At each step, the algorithm applies the measured amplitude constraint in Fourier space and enforces physical constraints in real space. This iterative strategy incrementally refines the solution in an effort to satisfy the constraints in both domains simultaneously.

Representative methods in this category include Gerchberg-Saxton (GS), Error Reduction (ER), and Hybrid Input-Output (HIO), each of which implements different projection rules and update mechanisms to balance convergence, stability, and avoidance of local minima (Fienup, 1982; Marchesini, 2004). These iterative methods remain widely used due to their algorithmic transparency, ease of implementation, and applicability in experimental settings where labeled data for learning-based approaches is unavailable. The success of these methods relies on the assumption that while the phase is unknown, certain properties of the object are known or can be reasonably approximated. Through repeated projection onto constraint sets, the algorithm seeks convergence toward a solution that simultaneously satisfies the modulus constraint in Fourier space and the structural constraints in real space. Importantly, these methods do not require labeled data or training procedures, making them robust, interpretable, and directly applicable in experimental contexts.

Fundamentally, iterative algorithms share similar steps, as shown in Figure 2.3. For better understanding, this is the calculation inside the GS algorithm. The GS algorithm is one of the earliest and most widely known iterative phase retrieval methods. It operates by iteratively switching between the spatial (real) space and the frequency (reciprocal) space, enforcing known constraints in each. The sequence of mathematical expressions used in the GS algorithm illustrates how the algorithm estimates the missing phase information while preserving measured amplitude data. Let $g_k(x)$ denote the complex-valued estimate of the object at iteration k in the spatial domain. Applying the forward Fourier transform yields:

$$G_k(u) = |\hat{G}_k(u)|e^{i\phi_k(u)} = \mathcal{F}[g_k(x)], \tag{2.16}$$

where $|G_k(u)|$ and $\phi_k(u)$ are the amplitude and phase of the spectrum at iteration

k, respectively. However, since only the measured diffraction amplitude $|F(u)|$ is known, the algorithm replaces the amplitude of $G_k(u)$ with the known magnitude, the one which came from the real experiment, while retaining the phase $\phi_k(u)$:

$$G'_k(u) = |F(u)|e^{i\phi_k(u)}. \tag{2.17}$$

An inverse Fourier transform is then applied to obtain a modified estimate in real space as follows:

$$g'_k(x) = |\tilde{g}'_k(x)|e^{i\theta'_k(x)} = \mathcal{F}^{-1}[G'_k(u)], \tag{2.18}$$

where $\theta'_k(x)$ is the phase in the spatial domain obtained from the current update. Finally, the object estimate is updated by replacing the amplitude with a known or assumed amplitude (e.g., from support or modulus constraints in real space), while retaining the updated phase:

$$g_{k+1}(x) = |f(x)|e^{i\theta_{k+1}(x)} = |f(x)|e^{i\theta'_k(x)}. \tag{2.19}$$

This cycle is repeated iteratively, and all iterations are supposed to minimize the loss between the calculated and experimental diffraction. The algorithm refines the phase information by maintaining consistency with measured amplitude in Fourier space and enforcing known constraints in real space. Although the GS algorithm assumes that both the amplitude in Fourier space and real space are known, which is not always the case in practice, it establishes the basic iterative structure later used and generalized in more flexible algorithms. In single-frame CXDI, this algorithm is often used as an initialization or baseline method due to its conceptual simplicity and foundational role in phase retrieval theory. The workflow of the GS algorithm can be examined in Figure 2.4.

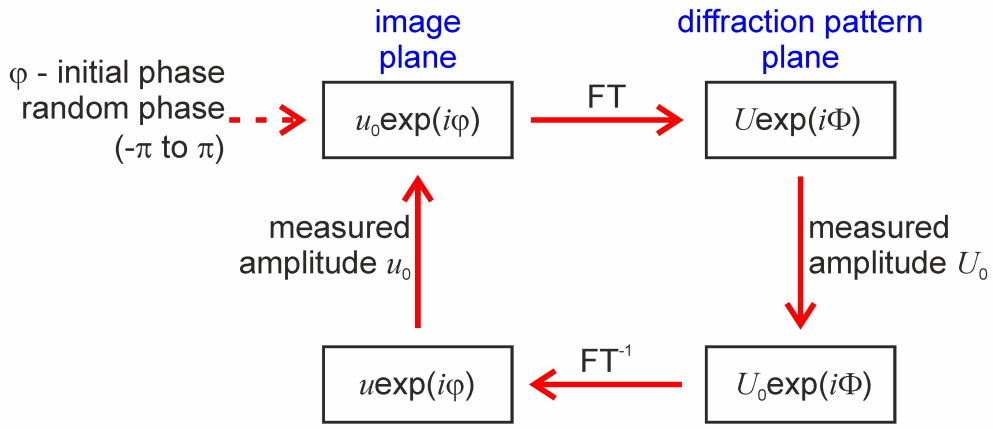


FIGURE 2.4: The workflow of GS algorithm.

The other iterative algorithms, such as ER and HIO, employed the first three steps in the GS algorithm. The difference is only in the constraint in the object space and how they use it to update the predicted reconstructed object for the next iteration. In the deep meaning, the major distinction lies not in how the transition between domains is performed, but rather in how the real-space constraints are applied and how the updated estimate is generated.

In the ER algorithm, the real-space constraint is applied in a straightforward and conservative manner. After each inverse Fourier transform, the object is updated by strictly enforcing the known support and optional amplitude conditions: values

inside the support region are retained, while those outside are set to zero or some background value. This conservative projection often leads to stagnation, especially in non-convex solution spaces, but it provides stable convergence when a good support is known. By contrast, the HIO algorithm introduces a more aggressive update rule in the object space to help the iteration escape local minima. In HIO, if a point lies outside the support region, its new value is not simply zeroed, but updated using a feedback parameter β to incorporate information from the previous estimate. This dynamic projection mechanism encourages exploration of the solution space and typically results in faster convergence and better robustness against noise or imperfect constraints.

Both ER and HIO, therefore, follow the same domain-transition structure inherited from GS but diverge significantly in their philosophy of constraint enforcement in real space. ER focuses on strict adherence, while HIO introduces controlled deviation. In CXDI, the choice between these methods often depends on the complexity of the object, the accuracy of the support constraint, and the desired trade-off between stability and reconstruction fidelity. To examine clearly the difference between those three algorithms, Figure 2.5 illustrates it.

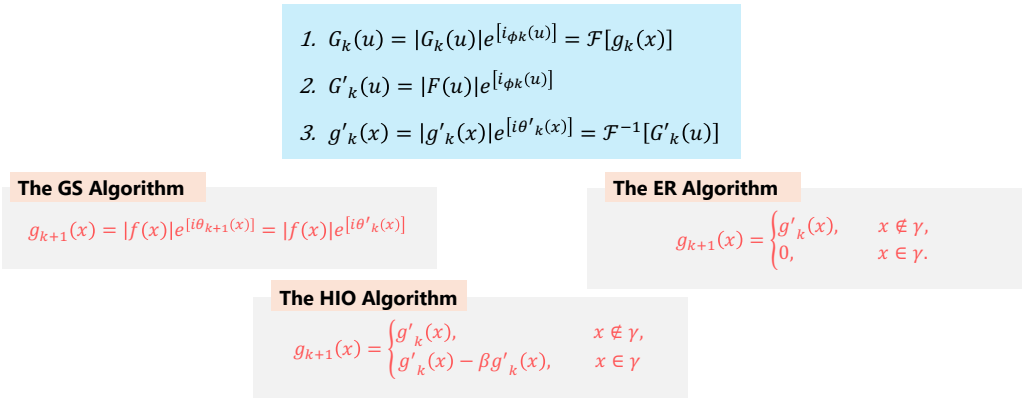


FIGURE 2.5: The difference among the three well-known iterative phase retrieval algorithms.

Other than the three initial algorithms, the extended Ptychographical Iterative Engine (ePIE) is also a part of a well-known iterative algorithm. In the first version, ePIE focused on Ptychography's phase retrieval. The ePIE improves phase retrieval in diffractive imaging by utilizing a series of diffraction patterns recorded as a known illumination function is translated to a set of overlapping positions relative to a target sample. The algorithm is an iterative phase retrieval algorithm designed for Ptychographic imaging, where a sample is scanned with overlapping illumination positions, and diffraction patterns are recorded at each position. While originally developed for multi-frame datasets, the algorithm's flexible structure also makes it useful in coherent diffraction imaging frameworks such as multi-probe CXDI and scanning-based modalities.

At its core, ePIE operates by refining both the object and the illumination probe simultaneously, in contrast to traditional algorithms, which assume a fixed probe. This makes ePIE particularly powerful in experimental settings where the beam profile may fluctuate or is not precisely known. The algorithm iteratively updates both the object estimate $O(\mathbf{r})$ and the probe $P(\mathbf{r})$ using local diffraction measurements, enforcing consistency with measured data and leveraging redundancy across overlapping regions. The basic update rules of ePIE can be expressed as follows. Let

the exit wave at scan position j be:

$$\psi_j(\mathbf{r}) = P(\mathbf{r} - \mathbf{R}_j) \cdot O(\mathbf{r}), \tag{2.20}$$

where $\mathbf{R}_j$ is the j-th scan position. After propagating to the detector and comparing the amplitude to the measured intensity, the exit wave is updated in Fourier space:

$$\tilde{\psi}_j(\mathbf{q}) = \mathcal{F}\{\psi_j(\mathbf{r})\}, \qquad \tilde{\psi}_j^{\text{new}}(\mathbf{q}) = \sqrt{I_j(\mathbf{q})} \cdot \frac{\tilde{\psi}_j(\mathbf{q})}{|\tilde{\psi}_j(\mathbf{q})|}. \tag{2.21}$$

The updated exit wave in real space is obtained by the inverse Fourier transform:

$$\psi_j^{\text{new}}(\mathbf{r}) = \mathcal{F}^{-1}\{\tilde{\psi}_j^{\text{new}}(\mathbf{q})\}. \tag{2.22}$$

The object and probe are then updated based on the difference between the new and old exit waves:

$$O^{\text{new}}(\mathbf{r}) = O(\mathbf{r}) + \alpha \cdot \frac{P^*(\mathbf{r} - \mathbf{R}_j)}{\max |P(\mathbf{r} - \mathbf{R}_j)|^2} \cdot \left[\psi_j^{\text{new}}(\mathbf{r}) - \psi_j(\mathbf{r})\right], \tag{2.23}$$

$$P^{\text{new}}(\mathbf{r}) = P(\mathbf{r}) + \beta \cdot \frac{O^*(\mathbf{r} + \mathbf{R}_j)}{\max |O(\mathbf{r} + \mathbf{R}_j)|^2} \cdot \left[\psi_j^{\text{new}}(\mathbf{r} + \mathbf{R}_j) - \psi_j(\mathbf{r} + \mathbf{R}_j)\right], \tag{2.24}$$

where α and β are step-size parameters controlling the update strength for the object and probe, respectively, and the denominators normalize the updates to prevent divergence. Figure 2.6 shows how the ePIE algorithm works.

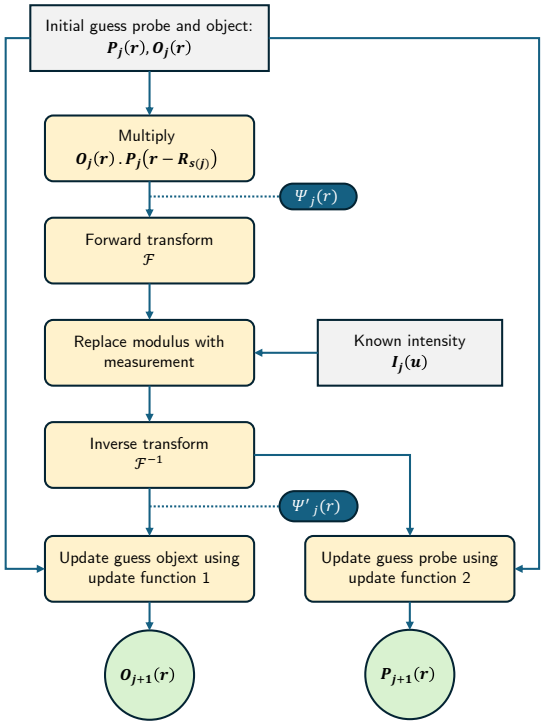


FIGURE 2.6: The workflow of the ePIE algorithm.

Although ePIE is more commonly associated with Ptychography than with single-frame CXDI, its mathematical formulation and underlying iterative principles are shared across both domains. In multi-frame CXDI variants where multiple partially

overlapping exposures are available, or in scenarios with structured probe movement (e.g., scanning FZPs), ePIE provides a powerful framework for joint phase retrieval and system calibration. Later in this study, we will implement the ePIE for the single-frame CXDI mechanism.

Many studies focused on how to optimize the iterative algorithm, as the classical projection-based methods often face significant practical challenges during reconstruction. One of the most common issues is the tendency of these algorithms to become stuck in local minima after several iterations, especially in cases where the object complexity is high or the support constraint is imprecise. Since the phase retrieval problem is inherently non-convex, the solution landscape contains many suboptimal basins of attraction, which can trap the iterative process and lead to reconstructions that satisfy the amplitude constraints but fail to reflect the true object structure.

Additionally, iterative algorithms are highly sensitive to their initialization. A poor initial guess can cause the algorithm to converge slowly, oscillate, or diverge entirely. Even when convergence is achieved, the solution may exhibit artifacts or be misaligned due to ambiguities such as global phase shift, lateral shift, or conjugate inversion. This sensitivity is further amplified in single-frame CXDI, where the absence of data redundancy (as found in Ptychography) reduces the algorithm's ability to self-correct through overlapping constraints. To address these issues, numerous optimization strategies have been proposed, including adaptive constraint modulation, hybrid update rules, feedback parameters, and stochastic perturbations (Changliang Guo, S. Liu, and Sheridan, 2015). Some approaches introduce regularization terms or external priors to steer the solution toward physically meaningful reconstructions, while others focus on dynamically updating the support or combining multiple reconstruction runs through ensemble methods (Mukherjee et al., 2023). Despite these advances, achieving stable, high-fidelity reconstruction across a broad range of samples remains an active area of research. Optimizing iterative phase retrieval thus involves not only refining the update mechanisms but also understanding the broader dynamics of convergence, robustness, and uncertainty in the context of the inverse imaging process. Table 2.1 explains several improvements of iterative phase retrieval algorithms in several studies.

TABLE 2.1: Some notable improvements of the iterative phase retrieval algorithm in some studies.

Researched by	**Summary of Improvement**
A. M. Maiden and J. M. Rodenburg, 2009	ePIE, a development of the PIE algorithm, enables simultaneous recovery of both object and probe wavefronts.
Moravec, Romberg, and Baraniuk, 2007	Offering CPR (Compressive Phase Retrieval) that consists of two main ways, such as using compressibility constraint for measurement and taking the advantages of the random Fourier coefficients to recover structured signals from fewer measurements.

Continued from previous page

Researched by	Summary of Improvement
Guizar-Sicairos and Fienup, 2012	Introducing the method for eliminating the twin-image problem by taking two reconstructions and selecting the one with the lowest support error as a reference.
Eldar and Mendelson, 2014	Guarantee the stability and recovery by studying the conditions under which the input vector can be stably recovered from quadratic measurements.
Changliang Guo, S. Liu, and Sheridan, 2015	Offering the improvement of the iterative phase retrieval algorithm by developing the Spatial Phase Perturbation GS Algorithm and a combined GS/HIO algorithm.
Elser, Lan, and Bendory, 2018	This research combined the HIO + ER algorithm (20 iterations of HIO and 1 iteration of ER) for speeding up the convergence.
Cheng Guo et al., 2019	The shrink-wrap constraint revolutionizes iterative phase retrieval by implementing a dual-optimization strategy that simultaneously enhances reconstruction quality through weighted feedback in the support region while eliminating redundant data outside it, resulting in faster convergence, improved contrast, and reduced computational demands.

In summary, the main idea of improving the iterative phase retrieval algorithm is focusing on two groups of study, either conjugating some iterative algorithms into one workflow or applying the gradient search method to speed up the convergence of the reconstruction process. The application of the gradient search method opened the opportunity for a data-driven approach in the phase retrieval problem.

2.4.2 Data-Driven Approaches for Phase Retrieval

In recent years, data-driven approaches, particularly those based on deep learning, have emerged as powerful alternatives to classical iterative methods for phase retrieval. These models aim to learn the inverse mapping from diffraction intensity patterns to real-space images directly from data, bypassing the need for hand-crafted constraint enforcement and repeated domain switching. The motivation behind using deep learning lies in its ability to approximate highly non-linear functions, making it well-suited for solving ill-posed inverse problems such as phase retrieval, where the solution space is vast and ambiguous.

Unlike iterative algorithms that rely on strict physical priors and deterministic updates, deep learning-based models exploit patterns learned from large datasets of simulated or experimental samples. These models can generalize across different

sample geometries, noise levels, and imaging conditions, offering rapid inference and scalability once trained. Architectures such as Convolutional Neural Networks (CNN) and physics-informed neural networks have been adapted to map intensity-only input data to complex-valued object reconstructions (Ye, L.-W. Wang, and Lun, 2023). However, this flexibility also introduces new challenges, including data dependence, lack of interpretability, and difficulty in quantifying reconstruction confidence.

In general, the deep learning phase retrieval works under a supervised or unsupervised learning mechanism. For supervised learning, the model is trained using pairs of input-output data in which the diffraction pattern (intensity-only measurements) serves as the input, and the corresponding ground-truth image of the object in real space serves as the output label. The learning objective is to directly approximate the inverse mapping from the diffraction domain to the object domain, i.e., from intensity $I(\mathbf{q})$ to reconstructed object $O(\mathbf{r})$. This requires access to large datasets of simulated or experimentally obtained diffraction patterns with known ground truth, which can be expensive or impractical to obtain at scale. Supervised approaches are often favored for their straightforward training process and direct interpretability of performance via reconstruction loss metrics such as mean squared error (MSE).

In contrast, unsupervised learning for phase retrieval does not rely on ground-truth object images during training. Instead, the model is trained to minimize the discrepancy between the predicted diffraction pattern, obtained via forward propagation of the model's current object estimate, and the original input diffraction data. This corresponds to a self-supervised or physics-consistent training paradigm, where the network learns to map from diffraction to object and back to diffraction again. The loss function is defined entirely in the reciprocal space, such as:

$$\mathcal{L} = \|\mathcal{F}\{O_{\text{pred}}(\mathbf{r})\}|^2 - I_{\text{meas}}(\mathbf{q})|^2. \tag{2.25}$$

This formulation allows the model to operate without labeled data, using only the measured intensities and physical constraints, thus making it more applicable to real-world CXDI experiments. In practice, unsupervised learning is often more suitable for CXDI, particularly in single-frame configurations where ground truth reconstructions are fundamentally unavailable. While simulated datasets can be used to generate synthetic ground truth for supervised learning, the realism and completeness of such simulations directly influence the model's generalization to experimental data. Any mismatch in noise characteristics, probe structure, or sample complexity between the simulated and real domains may degrade performance. As such, unsupervised models, by learning directly from experimental diffraction data using physics-consistent objectives, offer a more flexible and scalable approach for phase retrieval, especially when interpretability and adaptability to unknown samples are prioritized.

We conducted some literature reviews of several studies in deep learning phase retrieval to get some intuitive concepts. Table 2.2 explains several improvements of iterative phase retrieval algorithms in several studies.

TABLE 2.2: Some notable deep learning phase retrieval studies to mark the improvement of data-driven implementation in the phase retrieval problem.

Researched by	**Summary of Architecture**
Nishizaki et al., 2020	This paper discusses a supervised learning-based non-iterative phase retrieval method using the ResNet (Residual Network) architecture, which consists of two variants with different depths: ResNet1 (one down-and-up-sampling process) and ResNet2 (two down-and-up-sampling processes), where the uniqueness of the architecture is the use of skip connections to prevent vanishing/exploding gradients during the training process.
Metzler et al., 2018	prDeep, a supervised learning-based approach for phase retrieval that combines the Regularization by Denoising (RED) framework with DnCNN (a convolutional neural network), where the unique architectural feature is its integration of a neural network denoiser within a traditional optimization algorithm framework that allows it to handle various measurement models without requiring retraining. The algorithm proves particularly effective in handling noisy measurements while maintaining flexibility across different measurement models, such as coded diffraction patterns (CDP) and Fourier measurements.
Işıl, Oktem, and Koç, 2019	A supervised learning-based phase retrieval algorithm that combines classical Hybrid Input-Output (HIO) method with two Deep Neural Networks (DNNs), where the unique architectural features include a first DNN trained to remove HIO artifacts used iteratively with HIO, followed by a second DNN for final artifact removal, resulting in a robust solution that outperforms traditional methods while maintaining computational efficiency and showing resilience to different initialization conditions and noise levels.

Continued from previous page

Researched by	Summary of Architecture
Guan et al., 2019	This study introduces PtychoNet, a supervised learning-based non-iterative method for Ptychographic reconstruction, where its unique architectural features include a convolutional encoder-decoder network that processes diffraction patterns individually and combines them to generate the amplitude and phase of the original object, proving particularly effective in low-overlap conditions where traditional iterative methods fail, and when used as initialization for iterative algorithms, enables reconstruction of fine features even with sparse scanning patterns.
Cherukara et al., 2020	PtychoNN, a supervised learning-based approach that employs a deep convolutional neural network to directly map far-field coherent diffraction data to real-space image structure and phase, where its unique architectural features include an encoder arm for learning data representations and two decoder arms for amplitude and phase prediction, enabling 300 times faster reconstruction than traditional iterative methods while maintaining high accuracy even with sparse-sampled data (up to 5x fewer measurement points)
L. Wu et al., 2021	This paper presents a comprehensive approach for 3D coherent X-ray diffraction imaging that innovatively bridges supervised and unsupervised learning paradigms, initially utilizing a supervised convolutional encoder-decoder network architecture for rapid ($9ms$) reconstructions, then transitioning to an unsupervised transfer learning mode that adapts to new experimental conditions without prior training, demonstrating superior performance over traditional iterative methods while being more robust to noisy experimental data.

Continued from previous page

Researched by	Summary of Architecture
Vu et al., 2025	PID3Net, a novel supervised learning-based approach for phase retrieval in CXDI that uniquely transitions between supervised and unsupervised learning paradigms by combining a physics-informed deep learning strategy with temporal convolution blocks for capturing spatiotemporal correlations, where its distinctive architectural features include measurement-informed refined neural network blocks for integrating optical settings and mathematical constraints, enabling superior reconstruction quality and higher spatial resolution compared to traditional iterative methods.

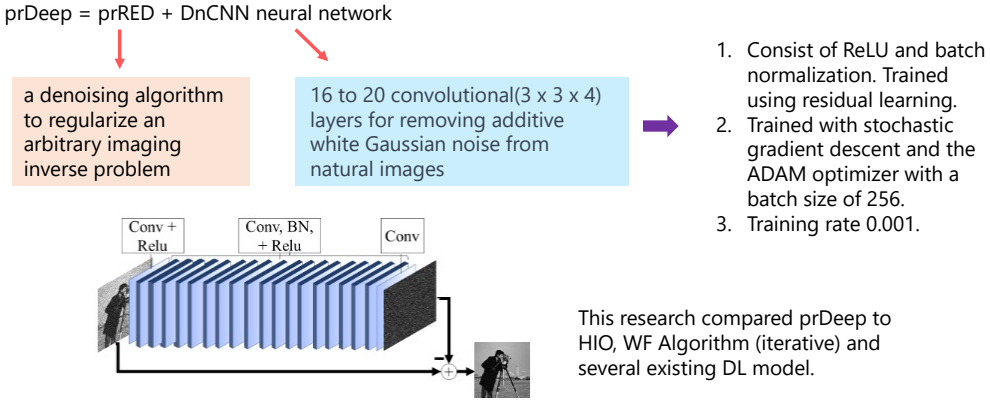


FIGURE 2.7: The architecture of prDeep phase retrieval.

Despite the promising performance of deep learning-based approaches for phase retrieval, several critical challenges remain, particularly in the context of single-frame CXDI, where data is inherently sparse and experimental ground truth is unavailable. Most existing architectures rely heavily on supervised learning paradigms. These models are trained to map diffraction patterns to real-space reconstructions using synthetic datasets that pair input intensity measurements with ground-truth phase or image data. While effective within the domain they are trained on, this reliance on simulated ground truth introduces a significant limitation. The learned models may not generalize well to real experimental conditions, where factors such as noise, beam instability, detector imperfections, or deviations in sample geometry can substantially alter the data distribution. Furthermore, the accuracy of the learned mapping in supervised models is inherently bounded by the quality and representativeness of the training data. If the training dataset fails to capture the diversity and complexity of real-world experimental scenarios, the network's performance may degrade, leading to reconstructions that appear plausible but are physically inconsistent or misleading. This limitation is compounded in CXDI, where the uniqueness of the inverse problem is not guaranteed, and multiple plausible solutions can exist for the same input data.

Architectures like prDeep (Metzler et al., 2018), as shown in Figure 2.7 and the dual-stage HIO + DNN pipeline proposed by (Işıl, Oktem, and Koç, 2019) attempt to mitigate this issue by hybridizing learning with traditional algorithms or denoising steps. However, while these strategies improve robustness and noise resilience, they do not fundamentally address the interpretability or uncertainty of the results. Similarly, models such as PtychoNN (Cherukara et al., 2020) and PtychoNet (Guan et al., 2019) demonstrate high reconstruction speed and accuracy in sparse conditions but provide limited insight into how specific features in the diffraction data influence the reconstructed image as shown in Figure 2.8. PID3Net (Vu et al., 2025) represents a

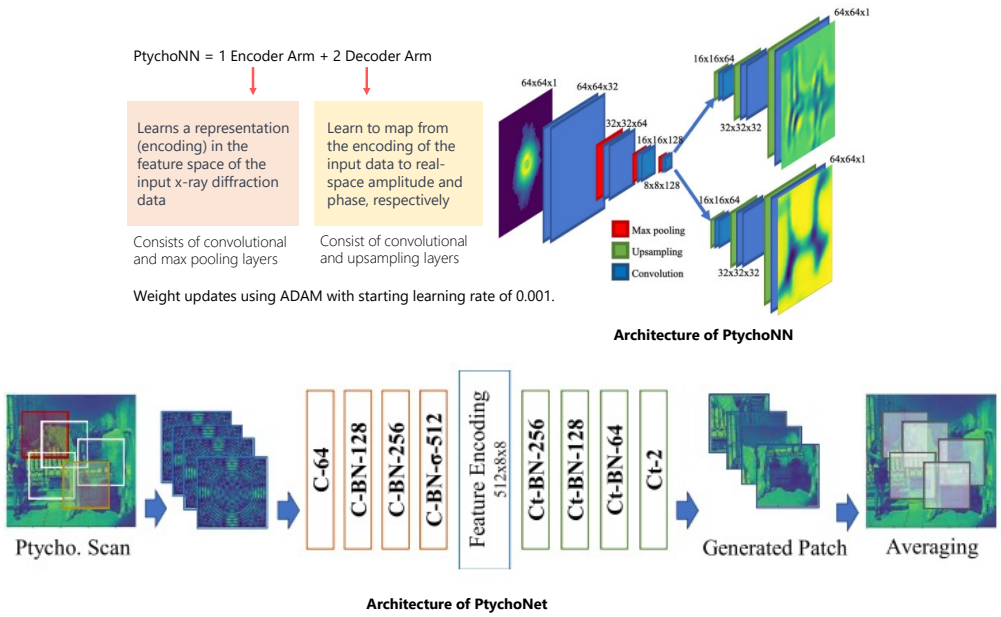


FIGURE 2.8: The architecture of ptychoNN and PtychoNet.

more recent effort to bridge the gap between supervised and unsupervised learning by incorporating physics-informed constraints and ensemble-based regularization. Nevertheless, even this approach, while showing improvement in robustness and uncertainty estimation, still operates within an under-determined solution space and does not fully resolve the epistemic ambiguity inherent in phase retrieval.

Overall, a major research gap remains in developing deep learning frameworks that can operate reliably on real experimental data, minimize reliance on simulated ground truth, and provide meaningful uncertainty estimates that can guide scientific interpretation. These limitations underscore the need for hybrid approaches that combine the speed and flexibility of deep learning with the physical rigor and explainability of model-based inference.

2.4.3 Physics-Informed in Phase Retrieval

Phase retrieval in single-frame CXDI is fundamentally an ill-posed inverse problem. This reconstruction is inherently infeasible through direct inversion, as the phase component of the diffracted wave is entirely lost during measurement. To overcome this limitation, phase retrieval algorithms incorporate well-defined prior information derived from the underlying physical phenomena observed in actual experiments. These physical constraints help guide the algorithm towards realistic solutions by restricting the solution space to a finite and physically meaningful domain. This

approach transforms a problem that would otherwise be difficult to solve into one that is manageable, enhancing both the stability and reliability of the reconstructed image. Several studies refer to this approach as physics-informed or physics-constrained. To implement such guidance effectively, phase retrieval algorithms introduce a range of physics-informed constraints that reflect the physical properties of the imaging system and the sample under investigation. These constraints are typically enforced during the iterative update process, either in the real space or reciprocal space domain, and serve to regularize the solution by eliminating physically implausible candidates from the solution space.

One of the most fundamental constraints is the **Fourier magnitude constraint**, which ensures consistency with the measured diffraction data. Since the detector captures only the intensity of the scattered wave, the reconstructed object must produce a Fourier transform whose magnitude matches the square root of the measured intensity. Mathematically, this is enforced by projecting the current estimate in reciprocal space onto the measured amplitude while keeping the current phase:

$$\hat{O}^{(k+1)}(\mathbf{q}) = \sqrt{I(\mathbf{q})} \cdot \exp\left(i \cdot \phi^{(k)}(\mathbf{q})\right), \tag{2.26}$$

where $I(\mathbf{q})$ is the measured intensity, and $\phi^{(k)}(\mathbf{q})$ is the current phase estimate at iteration k.

In real space, the **finite support constraint** is widely used to restrict the object to lie within a predefined region S. This assumption is physically motivated by the knowledge that the sample occupies only a limited spatial extent, and therefore:

$$O(\mathbf{r}) = 0 \quad \text{for} \quad \mathbf{r} \notin S. \tag{2.27}$$

This constraint is particularly powerful in reducing non-uniqueness, as it effectively truncates the spatial domain and eliminates extraneous structures outside the expected object area. It is most commonly employed in iterative projection algorithms such as ER and HIO. For example, the ER algorithm performs a projection step in real space that enforces the support constraint by setting all values of the object estimate outside the region S to zero. This projection acts as a hard boundary that guides the algorithm toward plausible solutions and helps to suppress non-physical artifacts. The value can be adjusted based on the underlying concept of each algorithm.

Another essential real-space constraint is the **non-negativity constraint**, especially relevant for objects representing electron density or intensity distributions. Since such quantities are inherently non-negative, enforcing $O(\mathbf{r}) \geq 0$ helps to discard non-physical reconstructions and improve convergence. This constraint is especially critical when the sample does not produce complex phase shifts, such as in absorption-dominant imaging setups.

Closely related is the **object constraint**, a general term that encompasses assumptions about the sample's structural and material properties. These can include:

- **Atomicity constraint**, which assumes that the sample is composed of discrete atomic units arranged periodically. This is particularly effective in crystallographic imaging, where the object can be modeled as a sparse arrangement of atomic peaks.

- **Connectivity constraint**, which imposes that the object forms a contiguous region rather than a collection of disjointed fragments. This helps suppress

reconstruction artifacts that break the object into non-physical disconnected regions.

- **Smoothness or total variation constraints**, which penalize abrupt changes in intensity to promote more physically plausible solutions in samples known to be continuous or slowly varying.

Figure 2.9 shows how to implement physical constraints in the general form of iterative phase retrieval.

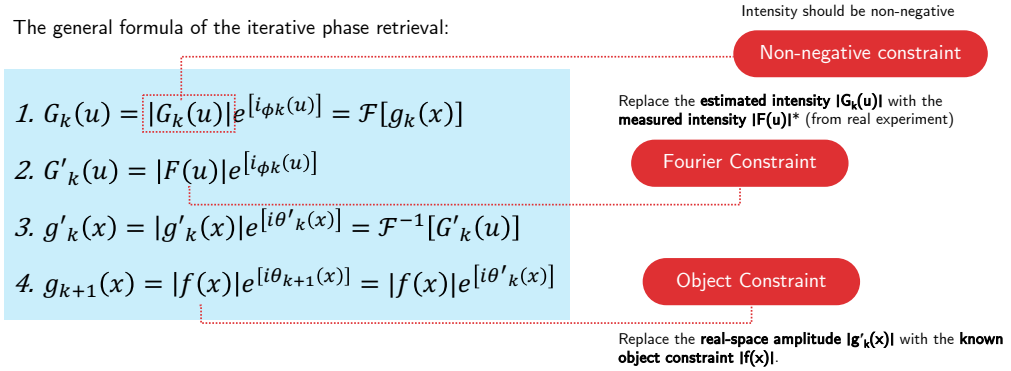


FIGURE 2.9: Explanation of how physical constraints were employed in iterative phase retrieval.

Physical constraints can also be employed in the deep learning phase retrieval. An example of a deep learning architecture that explicitly integrates physics-informed constraints into the phase retrieval pipeline is **PID3Net** (Vu et al., 2025). Unlike purely data-driven models that learn a direct mapping from diffraction patterns to object reconstructions, PID3Net embeds physical consistency directly into the network design and training process. One of its core mechanisms is the enforcement of the **Fourier constraint** through a physics-based loss function that penalizes discrepancies between the measured and predicted diffraction intensities. Specifically, the network is trained to minimize the amplitude error between the square root of the measured intensity $I(\mathbf{q})$ and the Fourier modulus of the predicted object by utilizing Equation 2.25. This encourages the network to generate outputs that are physically consistent with the actual diffraction measurements, thereby maintaining forward-model fidelity throughout training.

In addition to the Fourier constraint, PID3Net incorporates domain-specific priors such as non-negativity and spatial support during pre- and post-processing stages. Non-negativity is enforced by applying a ReLU activation at the final output layer, ensuring that the reconstructed electron density remains physically meaningful. The network also uses spatial masking operations to emulate finite support during data augmentation and inference, further constraining the solution space. Beyond these hard constraints, PID3Net embeds the physics of wave propagation through the use of Fourier layers and dropout-based ensemble modeling, allowing it to both respect the known physics and model the uncertainty arising from data ambiguity. By combining deep neural inference with embedded physical constraints, PID3Net demonstrates a hybrid approach that leverages the strengths of data-driven learning while maintaining interpretability and reliability grounded in the physical imaging model.

Each of those constraints serves a specific role in guiding the algorithm through a vast and ambiguous solution space. While the Fourier constraint ensures consistency

with the observed data, the real-space constraints encode prior knowledge about the object's physical structure and composition. Together, they synergistically improve the stability, reliability, and physical interpretability of the phase retrieval process, particularly in the challenging context of single-frame CXDI, where minimal data redundancy amplifies the ill-posedness of the inverse problem.

2.5 Evaluating Phase Retrieval Performance

After receiving the reconstruction results, the products of each phase retrieval algorithm should be examined to determine the performance of the inverse process. Because phase retrieval operates in two interconnected domains, real space and reciprocal space, evaluation must be conducted in both. Each domain provides unique insights into reconstruction quality, consistency, and physical validity.

In the **reciprocal space**, the primary goal is to assess how well the reconstructed object reproduces the measured diffraction pattern. This comparison reflects the physical fidelity of the solution with respect to the experimental observations. One commonly used metric is the *R-factor (Rf score)*, which quantifies the relative discrepancy between the reconstructed and measured diffraction amplitudes across all spatial frequencies. The Rf score used in this study follows the formulation from (Miao, Chen, et al., 2006):

$$R_f = \frac{\sum_{u,v} \left| \sqrt{I_{\text{rec}}(u,v)} - \sqrt{I_{\text{ref}}(u,v)} \right|}{\sum_{u,v} \sqrt{I_{\text{ref}}(u,v)}}, \tag{2.28}$$

where $I_{\text{rec}}(u,v)$ is the reconstructed diffraction intensity and $I_{\text{ref}}(u,v)$ is the reference or target intensity, which in simulation corresponds to the ground truth diffraction pattern. The R_f score evaluates the agreement between the magnitude of the reconstructed and reference Fourier spectra, independent of phase. It is normalized to the total reference amplitude, allowing consistent comparison across different samples or experimental conditions. Conceptually, the Rf score reflects how faithfully the reconstruction satisfies the Fourier magnitude constraint derived from the experimental data. A lower Rf value indicates better physical consistency with the observed diffraction pattern, suggesting that the inverse model has produced a plausible and stable reconstruction. This metric is particularly valuable in both simulated and experimental settings where validating the physical realism of the reconstruction is essential, even in the absence of a known real-space ground truth.

For the real space, performance is evaluated by comparing the reconstructed object $O_{\text{rec}}(\mathbf{r})$ with a known or simulated ground truth $O_{\text{true}}(\mathbf{r})$. This is typically possible only in a simulation context, where the true structure is available. Several metrics are commonly used:

- **Mean Squared Error (MSE)**:

$$\text{MSE} = \frac{1}{N} \sum_{\mathbf{r}} \left(O_{\text{rec}}(\mathbf{r}) - O_{\text{true}}(\mathbf{r}) \right)^2, \tag{2.29}$$

  where N is the total number of pixels. This metric penalizes large deviations and is sensitive to outliers, as larger differences between the reconstructed image O_{rec} and the ground truth O_{true} are squared, amplifying their contribution to the final score. MSE is not only a common evaluation metric for quantifying

reconstruction accuracy in simulation-based phase retrieval, but also serves as a primary loss function during the training of supervised deep learning models. In many architectures, such as PtychoNN (Cherukara et al., 2020), PtychoNet (Guan et al., 2019), and PID3Net (Vu et al., 2025), the network is trained to minimize the MSE between predicted and ground-truth objects, guiding the optimization process toward reconstructions that are numerically closer to the desired solution. Although MSE offers clear mathematical interpretability and ease of implementation, it does not account for structural or perceptual similarities and is therefore often combined with additional metrics such as SSIM or perceptual losses in practice.

- **Peak Signal-to-Noise Ratio (PSNR)**:

$$\mathrm{PSNR} = 10 \cdot \log_{10} \left(\frac{(\max(O_{\mathrm{true}}))^2}{\mathrm{MSE}} \right), \tag{2.30}$$

where $\max(O_{\mathrm{true}})$ represents the maximum possible pixel value of the ground truth object, and MSE is the mean squared error between the reconstructed and ground truth images (Hore and Ziou, 2010; Setiadi, 2021). PSNR is a logarithmic metric that quantifies the ratio between the maximum signal power and the power of the reconstruction error. A higher PSNR value indicates better fidelity of the reconstruction, as it implies lower overall distortion relative to the dynamic range of the image. PSNR is commonly used in simulation-based studies of phase retrieval, particularly in the training and evaluation of deep learning models. In some studies, PSNR is used alongside MSE or SSIM as a benchmark for reconstruction quality.

- **Structural Similarity Index Measure (SSIM)**:

$$\mathrm{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}, \tag{2.31}$$

where μ_x, μ_y are the local means, σ_x, σ_y are local variances, and σ_{xy} is the covariance between the reconstructed image x and the ground truth image y (Hore and Ziou, 2010). SSIM captures perceptual similarity based on three components: luminance, contrast, and structural consistency. Unlike MSE or PSNR, which treat all pixel differences equally, SSIM emphasizes the preservation of structural information that aligns more closely with human visual perception. SSIM is widely used in image reconstruction tasks, including phase retrieval models that utilize supervised learning and simulation data. SSIM often complements MSE or PSNR to provide a more comprehensive evaluation of model performance. It is typically used as a validation metric after training or to compare visual outcomes between alternative reconstruction strategies. In some cases, SSIM is also integrated into the training loss as a perceptual component to promote more human-aligned reconstructions. Its robustness in highlighting structure-preserving performance makes SSIM particularly useful when the application requires accurate spatial localization of features or preservation of fine structural elements.

However, real-space evaluation is inherently dependent on simulation data. In real experimental scenarios, the true object is typically unknown, making direct

comparison impossible. This poses a significant challenge: the reliability of real-space metrics is only as strong as the realism of the simulated ground truth. If the simulation fails to replicate the complexity, noise characteristics, or experimental conditions of real-world data, then performance evaluations may overestimate algorithmic capability or fail to reveal true limitations. Therefore, a central challenge in evaluating phase retrieval performance lies in constructing simulated datasets that are physically realistic, noise-aware, and experimentally grounded, enabling meaningful validation of inverse algorithms. Bridging the gap between simulated and experimental conditions remains an open challenge, and future research in phase retrieval must continue to explore strategies for calibrating simulated reconstructions against physically observable proxies, ensuring that performance metrics retain their relevance in real-world applications.

2.6 Designing Ensemble Learning Mechanism for Phase Retrieval

Reconstructing real-space images from intensity-only diffraction patterns in CXDI often leads to multiple candidate solutions that are mathematically valid but structurally distinct. This ambiguity is a consequence of the phase loss inherent in the measurement process, as well as the lack of overlapping measurements in single-frame acquisition schemes. Most conventional phase retrieval methods, whether based on iterative solvers or deep learning, aim to converge to a single solution, often without revealing how sensitive that solution is to initialization, algorithmic settings, or data variations. This singularity in output masks the underlying uncertainty of the problem and may result in reconstructions that appear reliable but fail to capture the full range of plausible interpretations.

To address this limitation, ensemble-based phase retrieval, the core of this study, offers a framework that embraces solution diversity rather than suppressing it. By generating multiple reconstructions under varied conditions, such as different algorithmic initializations, phase retrieval algorithms, or masked input configurations, an ensemble can reveal both stable features that consistently emerge and regions of high variability that signal ambiguity. This approach shifts the paradigm of phase retrieval from single-point estimation toward distributional inference, enabling the estimation of reconstruction uncertainty and enhancing the scientific interpretability of the output. It provides a pathway to not only assess the quality of a reconstruction but also to reason about its reliability.

Ensemble learning is a fundamental strategy in machine learning that combines multiple predictive models to improve overall performance, robustness, and generalization. The core idea is that while individual models may make errors due to variance, bias, or overfitting, their aggregated predictions can compensate for each other's weaknesses and lead to more reliable outcomes (Mienye and Y. Sun, 2022; Polikar, 2012; Sagi and Rokach, 2018). Ensemble methods operate under the assumption that no single model can perfectly capture the underlying data distribution, especially in complex or noisy tasks. Instead, a diverse set of models, each trained or configured differently, can jointly produce a solution that is statistically more accurate and stable than any constituent member alone.

The effectiveness of ensemble learning is grounded in three fundamental pillars: **diversity**, **training ensemble members**, and **combining ensemble members** (Polikar, 2012). These elements define the essential design strategies required to construct

ensemble systems that are robust, reliable, and capable of expressing uncertainty, especially in solving ill-posed inverse problems such as phase retrieval.

- **Diversity**: Diversity refers to the need for ensemble members to produce different outputs when exposed to the same or similar inputs. This is critical because if all models in the ensemble behave identically, the ensemble provides no advantage over a single model. Diversity can be introduced through various means, such as initializing models differently, training them on bootstrapped or masked datasets, or using different algorithms or network architectures. In phase retrieval, diversity might arise from random initial phase guesses, algorithmic variation (e.g., ER, HIO, PID3Net), or applying distinct masking strategies to the input diffraction data.

- **Training Ensemble Members**: This pillar focuses on how the ensemble members are constructed and trained. Ensemble members must be designed to exploit the diversity introduced during initialization or data transformation. Models can be trained in parallel (as in bagging) or sequentially (as in boosting), and can either share the same architecture or vary across a set of heterogeneous configurations. In phase retrieval, training can be implemented by running multiple solvers in parallel with different inputs, constraints, or dropout patterns—each capturing a distinct region of the solution space. The objective is to ensure that each member explores a plausible but distinct reconstruction path.

- **Combining Ensemble Members**: Once diverse and independently trained ensemble members are available, their outputs must be combined to form a final prediction. This process, known as aggregation, can be achieved through methods such as majority voting, averaging, or weighted fusion. In phase retrieval, this step involves combining real-space reconstructions through voxel-wise averaging or computing voxel-wise variance to estimate uncertainty. The aggregation step is crucial not only for improving stability but also for providing spatial maps that reflect the confidence level of the ensemble in different regions of the reconstructed object.

These three pillars work synergistically to enable ensemble learning systems to provide better generalization, higher resilience to noise and outliers, and improved epistemic transparency, making them especially valuable for high-stakes scientific imaging applications like CXDI.

The classical ensemble paradigm is widely recognized in the machine learning literature, each offering a distinct mechanism for combining multiple models to enhance prediction quality and robustness. The paradigm is **Bagging**. Bagging (Bootstrap Aggregating) aims to reduce model variance by generating multiple base models, each trained on a different bootstrapped subset of the original dataset. Bootstrapping involves sampling with replacement, resulting in diverse training subsets that introduce variability among the models. Once trained, the models produce independent predictions that are aggregated, typically via averaging for regression or majority voting for classification. The main advantage of bagging is its ability to stabilize high-variance models (e.g., decision trees), producing more consistent outputs across different data realizations. In probabilistic inference contexts, bagging also allows estimation of prediction uncertainty by analyzing the spread or variance among ensemble members. Figure 2.10 illustrates the bagging mechanism in ensemble learning.

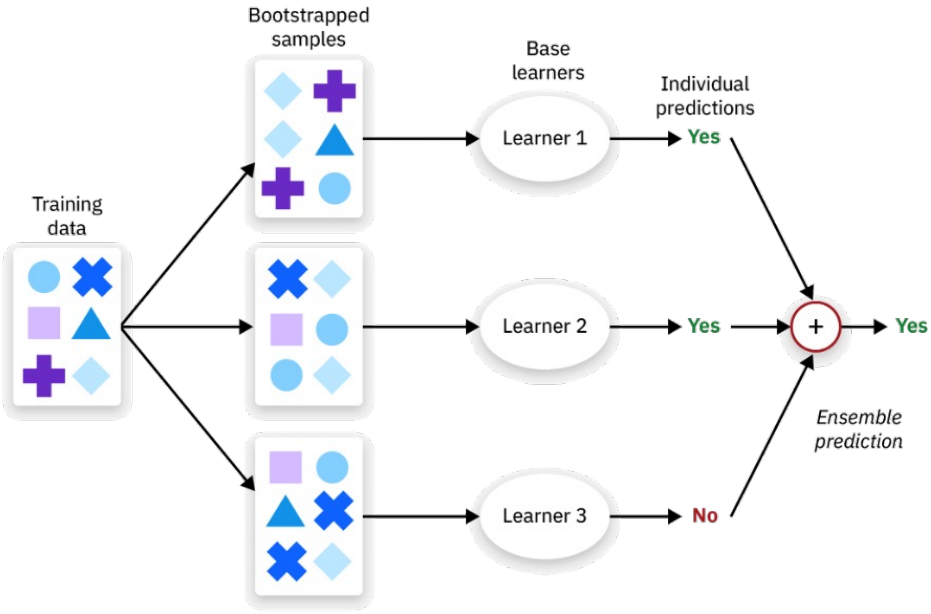


FIGURE 2.10: Illustration of ensemble learning with bagging mechanism (source: https://www.ibm.com/think/topics/ensemble-learning).

In bagging-based ensemble learning, once multiple base models have been trained on bootstrapped subsets of the dataset, their individual predictions must be aggregated to produce a final output. This aggregation is typically performed using either **hard voting** or **soft voting**, depending on whether the models return class labels or probability distributions over classes. Hard voting involves a majority rule, where each model contributes a single categorical prediction, and the final decision corresponds to the most frequently predicted label across the ensemble. While simple and effective in many cases, hard voting ignores the confidence or uncertainty associated with each model's prediction. Soft voting, in contrast, leverages the probabilistic output of each model. Instead of voting on the final label, each model provides a probability distribution over the possible outcomes. These probability vectors are then averaged element-wise across all ensemble members, and the final prediction corresponds to the class with the highest average probability. Mathematically, if $P_i(c)$ denotes the predicted probability for class c from model i, and M is the total number of models, the soft voting score for class c is given by:

$$P_{\text{ensemble}}(c) = \frac{1}{M} \sum_{i=1}^{M} P_i(c). \tag{2.32}$$

Soft voting is especially useful when base models have varying degrees of confidence or produce probabilistic estimates derived from uncertainty-aware mechanisms (e.g., Bayesian neural networks or dropout-based inference). In the context of inverse problems like phase retrieval, where ensemble members may represent reconstructions under different conditions, soft voting allows for the quantification of solution consistency and supports uncertainty estimation by analyzing the dispersion of the averaged predictions. This probabilistic aggregation aligns well with the goals of ensemble phase retrieval, where capturing and interpreting the variability among

plausible reconstructions is more informative than a single hard classification.

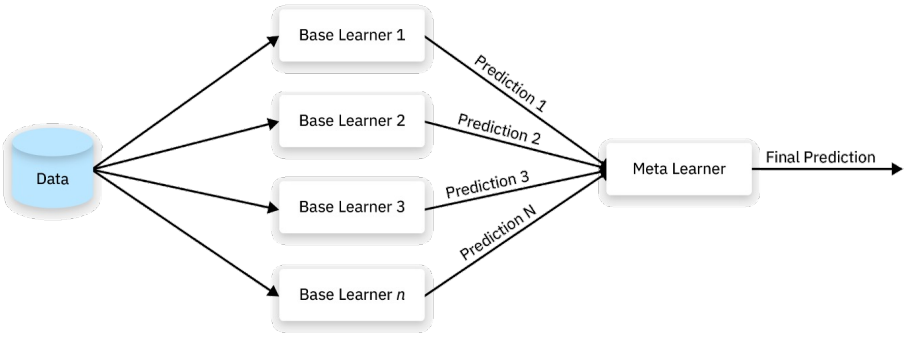


FIGURE 2.11: Illustration of ensemble learning with stacking mechanism (source: https://www.ibm.com/think/topics/ensemble-learning).

Other than the bagging approach, the other two approaches in Ensemble Learning, stacking and boosting, also offer the benefit of their implementation. Stacking, or stacked generalization, is an ensemble learning technique that aims to improve predictive performance by combining the strengths of multiple different learning algorithms. Unlike methods such as bagging or boosting, which typically use homogeneous models, stacking incorporates heterogeneous base learners (e.g., decision trees, support vector machines, and neural networks) and combines their outputs to produce a more accurate final prediction. Stacking is typically structured in two levels. In the **first level**, several base learners are trained in parallel on the same dataset, each generating predictions on a validation or test set. These predictions are then passed to a **second-level model**, known as the **meta-learner** or **blender**, which learns how to best combine the outputs of the base models, as shown in Figure 2.11. The meta-learner effectively learns the strengths and weaknesses of each base model and produces the final prediction based on their combined behavior. One of the main advantages of stacking is its ability to leverage model diversity. It exploits the complementary strengths of different algorithms to improve generalization. However, the effectiveness of stacking depends heavily on choosing sufficiently diverse base learners and a meta-learner capable of capturing their relative contributions.

On the other hand, as the last approach, Boosting is an ensemble learning technique designed to improve prediction accuracy by sequentially combining multiple weak learners. Unlike bagging, which trains models independently in parallel, boosting builds models iteratively, where each subsequent model focuses on the errors made by its predecessor, as shown in Figure 2.12. In essence, boosting directs more attention to instances that are hard to predict, progressively reducing both bias and variance. The process begins by training a base learner on the original dataset. After evaluating its performance, the weights of misclassified instances are increased so that the next learner pays more attention to them. This cycle continues for a predefined number of rounds or until performance stabilizes. The final prediction is typically obtained through a weighted majority vote (for classification) or weighted average (for regression) of all base learners. Prominent examples of boosting algorithms include AdaBoost (Schapire, 2013) and Gradient Boosting (Natekin and Knoll, 2013). The strength of boosting lies in its ability to transform simple models—such as shallow

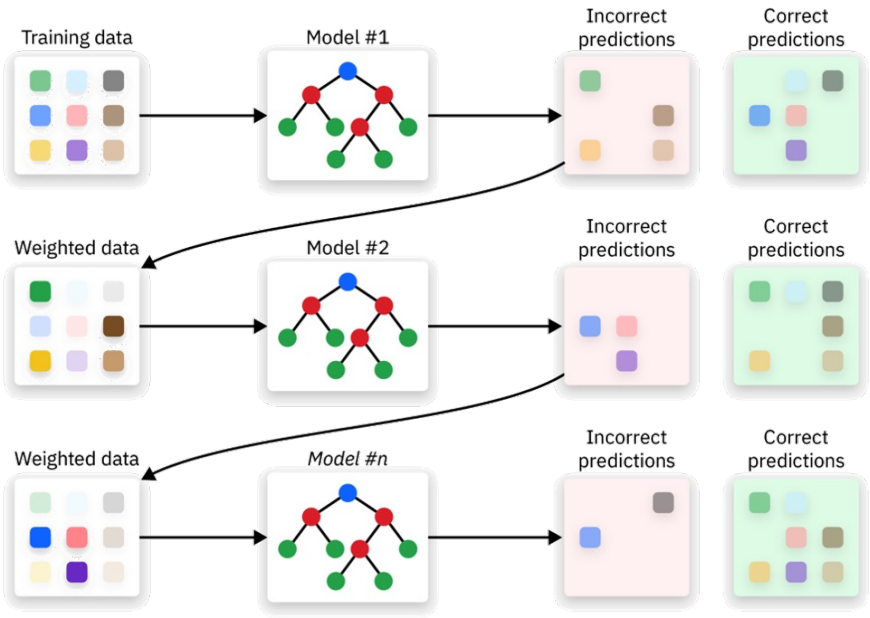


FIGURE 2.12: Illustration of ensemble learning with boosting mechanism (source: https://www.ibm.com/think/topics/ensemble-learning).

decision trees—into a powerful predictive system. However, it also has a tendency to overfit, especially when too many iterations are used or when the data is noisy.

The integration of ensemble methodologies into phase retrieval represents a significant advancement beyond conventional reconstruction approaches, shifting the focus from single-solution determinism to a systematic exploration of multiple viable solutions. This ensemble-based framework addresses two core challenges in single-frame CXDI: reconstruction instability and epistemic uncertainty. Through its principles of solution diversity, methodical generation of ensemble members, and principled aggregation, the approach provides a more nuanced view of the reconstruction space. Importantly, while it resembles ensemble learning, it is not a direct adoption of bagging or boosting: there are no labels or train–test splits, and the ensemble members are physics-constrained reconstructions produced from the same diffraction data under controlled perturbations (e.g., skip-region masking), distinct initializations, and heterogeneous solvers. The objective is to sample the manifold of solutions consistent with the Fourier-magnitude constraint, then fuse them into a consensus estimate and a pixel-wise variance map that quantifies residual ambiguity. Rather than concealing non-uniqueness, the method makes it explicit and measurable, yielding outputs such as uncertainty distributions, coherence analyses, and feature-stability maps that enhance both interpretability and reliability. In this way, the ensemble strategy simultaneously stabilizes the inversion numerics and delivers uncertainty-aware scientific insight, aligning phase retrieval with contemporary probabilistic inference and underpinning the physics-constrained, statistically robust workflows developed in this dissertation.

Chapter 3

Geometric Optimization of Aperture Design in Enhancing the Single-Frame CXDI Reconstruction

3.1 Introduction

Ptychography represents an advanced computational imaging methodology that aims to extract high-resolution structural and property information from specimens at the nanoscale through the reconstruction of both amplitude and phase data. This technique has evolved significantly, leading to the development of single-frame CXDI (J. M. Rodenburg, 2008; J. Rodenburg and A. Maiden, 2019; Konda et al., 2020; Pfeiffer, 2018; Nakasako et al., 2020; Sidorenko and Cohen, 2015; Goldberger et al., 2020). Single-frame CXDI, established as a sophisticated nano-imaging technique, has emerged as a powerful approach for investigating nanomaterial structures using X-ray radiation (Takazawa, Kang, et al., 2021; Kang et al., 2021; Takazawa, Dao, et al., 2023; Levitan et al., 2020). The methodology encompasses two fundamental processes: the Forward Process and the Inverse Process. The Forward Process fundamentally involves the acquisition of diffraction patterns from nanomaterial specimens using specialized instrumentation (Zheng et al., 2021; Chang et al., 2019; Jiang et al., 2021; Weker, Huang, and Toney, 2016). In contrast, the Inverse Process focuses on reconstructing the original image from these diffraction patterns through phase retrieval algorithms, employing either iterative methods or deep learning approaches (Yeh et al., 2015; Latychevskaia, 2018; Fienup, 1982). A comprehensive illustration of these forward and inverse processes in single-frame CXDI is presented in Figure 3.1. The methodology of single-frame CXDI is aligned with the two pillars of computational imaging, where the hardware design focuses on how to provide better signal acquisition, and computational design gives more concern for the inverse problem.

X-rays present a significant advantage in the forward process due to their characteristically short wavelength, which enables effective penetration of nanomaterials (Fromm, 2015; Withers, 2013). A critical requirement for obtaining interpretable diffraction patterns is the coherence of the X-rays utilized in the imaging procedure (Van der Veen and Pfeiffer, 2004; Wilke, 2015; Henry N Chapman and Nugent, 2010; Miao, Sandberg, and C. Song, 2011). However, generating coherent X-ray presents substantial technical challenges, necessitating the employment of specialized Synchrotron facilities (M. Liang et al., 2015). Unlike traditional Ptychography, single-frame CXDI faces unique challenges due to the absence of overlapping conditions, stemming from the singular nature of diffraction pattern generation for individual experimental specimens. Furthermore, the technical limitations of diffraction pattern detection systems underscore two critical requirements: adequate sampling of the

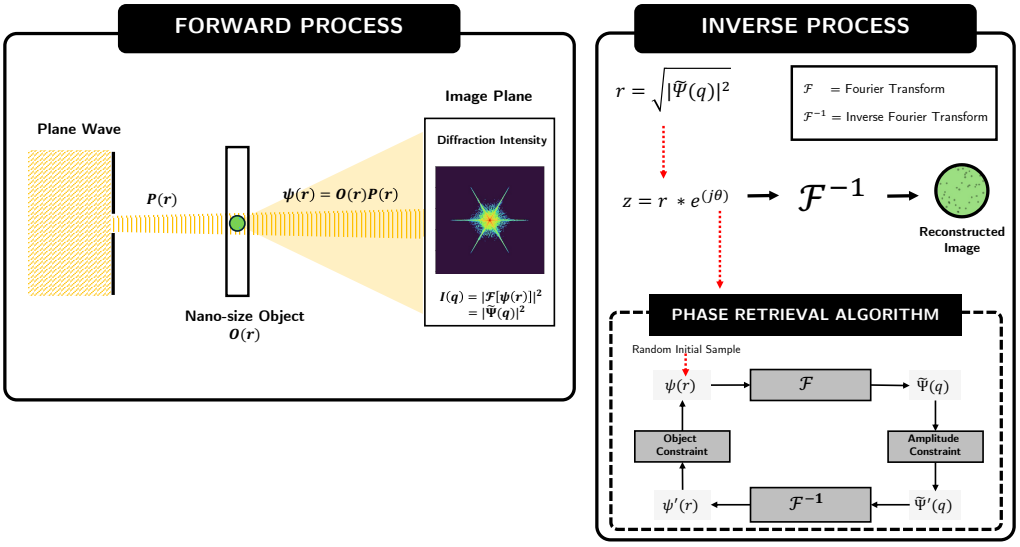


FIGURE 3.1: Illustration of the single-frame CXDI framework. In the forward process, a diffraction pattern is generated from the measured data, inherently discarding phase information. The inverse process attempts to reconstruct the original object from this intensity-only data by applying constraints in either real space or Fourier space. This study focuses on analyzing the effect of aperture geometry, particularly using a triangle-to-circle transition as the central design variable.

object image and the implementation of sophisticated phase recovery algorithms (Edo et al., 2013). Beyond the overlapping constraint issues, the diffraction pattern introduces additional complexities in the inverse process. The non-unique nature of inverse solutions emerges during the transformation of diffraction patterns back to their original image form. These imaging process characteristics and associated interference effects can significantly impact both the accuracy and quality of reconstructed images (Guizar-Sicairos and Fienup, 2012).

Numerous research initiatives have been undertaken to address these fundamental challenges. Contemporary investigations have primarily concentrated on two aspects: optimization of diffraction imaging mechanisms and enhancement of phase retrieval algorithms. In the context of diffraction imaging optimization, the strategic implementation of apertures within optical systems has emerged as a promising solution. A significant advancement in this direction has been the integration of high-precision triangle apertures into the diffraction imaging process (Takazawa, Kang, et al., 2021). Research conducted by Takazawa et al. has demonstrated the efficacy of triangle apertures in generating reconstructed object images through mixed-state reconstruction techniques, achieving remarkable resolutions approximately ten nanometers from single-frame diffracted intensity patterns with exposure times of 10 seconds, and exceeding $50nm$ resolution in certain conditions (Thibault and Menzel, 2013).

A comprehensive comparative analysis was conducted examining five distinct geometric configurations. The triangle geometry, characterized by its low symmetrical and non-symmetrical properties, emerged as the optimal aperture configuration (Kang et al., 2021). These geometric configurations demonstrated comparable contrast characteristics to the original image within illuminated regions, with reconstructed images showing minimal artifact presence. Current research has identified two crucial considerations in aperture fabrication. From a practical perspective, achieving

idealized triangle aperture geometries presents significant manufacturing challenges, with fabrication processes typically resulting in rounded corners or edge smoothing effects. Additionally, empirical evidence suggests that non-point symmetrical and low symmetrical aperture configurations yield superior reconstruction results. The non-point symmetry requirement stems from the fundamental principle that aperture geometries in single-frame CXDI must exhibit rotational symmetry around a central axis. This symmetry condition specifically requires that the aperture configuration maintains visual consistency under specific rotational transformations (e.g., 90°, 180°) relative to its central point. Within these constraints, various symmetrical triangle configurations present viable candidates for single-frame CXDI applications, particularly in satisfying primary geometric requirements.

As a complementary approach to enhancing the inverse process in single-frame CXDI, the implementation of deep learning architectures has emerged as a fundamental strategy for improving phase retrieval algorithms. Significant research efforts have focused on developing self-supervised learning networks, effectively eliminating the requirement for extensive pre-labeled datasets. Initial investigations introduced a fractional Fourier Transform (FrFT) framework incorporating self learning reconstruction methodologies (Y. Yang et al., 2024). Concurrent research proposed an untrained neural network architecture integrated with physics-enhanced mechanisms, capable of operation without preliminary training processes (F. Wang et al., 2020), designated as PhysenNet. These advanced learning networks operate autonomously through self-directed learning mechanisms, independent of diffraction imaging process optimizations such as aperture configurations in optical systems, as previously discussed.

Building upon existing research foundations, this investigation presents a comprehensive analytical framework. We have developed an advanced simulation methodology that enables detailed analysis of triangle aperture configurations and their variants in single-frame CXDI phase retrieval applications. The simulation framework generates triangle, square, and pentagon apertures, incorporating variable corner radius modifications to produce rounded corners, representing potential aperture configurations for single-frame CXDI applications. These aperture configurations are applied to nanomaterial specimens for single Fourier intensity measurements. Subsequently, phase retrieval methodologies, specifically ePIE and deep learning approaches, are employed to evaluate reconstruction outcomes for each aperture configuration (A. M. Maiden and J. M. Rodenburg, 2009; A. Maiden, D. Johnson, and P. Li, 2017; Manekar et al., 2020). This methodological approach addresses research requirements for both optical system optimization in diffraction imaging and phase retrieval algorithm enhancement.

In conclusion, this research part presents two significant contributions to the field. Primarily, our simulation-based methodology enables comprehensive modeling of diffraction imaging processes for nanomaterial specimens utilizing coherent X-ray radiation and single-frame CXDI techniques for both Forward and Inverse Processes. The simulation framework incorporates various aperture configurations with systematically varied corner radius, implementing incremental modifications until circular geometries are achieved. Additionally, this investigation provides comparative performance analyses of aperture configurations in phase retrieval applications, evaluating both real and reciprocal space domains across iterative and deep learning methodologies. These findings establish a foundation for future research in optimal triangle aperture design for single-frame CXDI experimental applications. The simulation parameters are carefully aligned with real experimental conditions, facilitating more efficient and optimized experimental protocols for both complete

and limited diffraction datasets.

3.2 Aperture Implementation in Single-Frame CXDI

Within the field of CXDI, single-frame methodologies have emerged as a crucial advancement, particularly for their ability to capture dynamic material changes with exceptional spatial and temporal precision. However, this technique faces two significant challenges. The first challenge is limited data availability from single measurements, and the second is the complex nature of image reconstruction. This stands in contrast to scanning Ptychography, which gets benefits from multiple overlapping measurements. Instead, single-frame CXDI's success largely depends on the precise configuration of its optical components, with particular emphasis on the illumination aperture design. Researchers have investigated various approaches to improve phase retrieval accuracy and solution uniqueness by optimizing experimental conditions. A particularly successful strategy involves the careful design of aperture geometry, which enhances both the physical collection of phase information and the mathematical efficiency of image reconstruction processes. Notable contributions from Kang et al. (Kang et al., 2021) and Takazawa et al. (Takazawa, Kang, et al., 2021) have shown that utilizing triangular apertures without point symmetry substantially improves image reconstruction quality, even without additional support constraints. This geometric approach fundamentally alters the diffraction pattern formation, enabling it to capture both amplitude and implicit phase information, similar to holographic imaging techniques.

The selection of triangular geometry for illumination apertures in single-frame CXDI represents a carefully researched decision, supported by extensive computational simulations and experimental verification. Research teams conducted comprehensive analyses to understand how various aperture geometries affect both the quality of diffraction patterns and phase retrieval outcomes. This systematic investigation was driven by a crucial observation, namely, traditional aperture designs, particularly those utilizing circular or rectangular geometries. Both apertures generally produce symmetrical diffraction patterns that increase the complexity of phase retrieval challenges, especially when support constraints are not available.

Through detailed computational studies, Kang et al. conducted extensive investigations into how different aperture geometries affect image reconstruction quality, specifically examining the stability of convergence and image resolution capabilities, as shown in Figure 1.14. Their findings revealed that triangular apertures possess unique characteristics for encoding phase information in diffraction patterns, primarily due to their asymmetric structure and distinct geometric features. These properties effectively disrupt the symmetrical patterns that typically cause algorithmic stagnation or produce multiple possible solutions in iterative phase retrieval processes. The triangular configuration essentially functions as an integrated phase reference system, generating structured interference patterns that facilitate phase information recovery from individual measurements. Complementing these theoretical findings, Takazawa et al. implemented practical experiments utilizing precisely manufactured triangular apertures. These apertures were fabricated through focused ion beam milling techniques applied to platinum foil, with subsequent testing conducted using synchrotron-based hard CXDI systems. Their experimental results demonstrated a direct correlation between the aperture's edge quality and the diffraction pattern's frequency characteristics, specifically, sharper edge definitions produced more extensive and well-defined high-frequency interference patterns. Significantly, their

research achieved high-resolution object reconstruction without requiring preliminary sample information, with spatial resolution capabilities exceeding those of the Fresnel Zone Plate (FZP) focusing optics employed in the experimental setup. Comparative analyses between triangular and alternative aperture geometries consistently showed that triangular configurations produced more reliable and higher-resolution reconstructions, maintaining effectiveness even under restricted exposure conditions.

These complementary research efforts provide comprehensive theoretical and experimental validation for implementing triangular apertures in single-frame CXDI applications. Their combined findings emphasize that aperture geometry serves beyond mere beam modification and can actively contribute to the encoding of recoverable phase information, thereby advancing the capabilities of lensless high-resolution imaging techniques.

While the triangular aperture has been shown to provide substantial advantages in single-frame CXDI, particularly in resolving phase ambiguity and enhancing spatial resolution, the practical realization of an ideal triangular shape introduces significant fabrication challenges. In both studies, it is emphasized that producing a perfectly sharp-edged, equilateral triangular aperture is difficult due to inherent limitations in micro-fabrication techniques such as Focused Ion Beam (FIB) processing. One of the most notable imperfections is the introduction of a finite corner radius at each vertex of the triangle, which deviates from the theoretical model that assumes ideal, infinitely sharp corners. Takazawa et al. addressed this issue directly by experimentally fabricating a high-precision triangular aperture on a platinum foil and characterizing its geometry using scanning ion microscopy. Their results revealed that the fabricated corners exhibit a radius of curvature on the order of 700 nm, which softens the aperture profile compared to the sharp-edged model used in simulations. To assess the impact of this imperfection, they compared experimentally measured diffraction intensity profiles with numerically simulated profiles of triangular apertures having varying degrees of edge smoothing 0 nm, 10 nm, 13.5 nm, and 20 nm. The experimental data most closely matched the simulated pattern for an edge smoothing of approximately 13.5 nm, confirming that the non-idealities introduced during fabrication were within tolerable limits for high-resolution reconstruction.

These findings underscore the importance of carefully characterizing and accounting for aperture edge quality in both the optical design and simulation of CXDI systems. Although corner rounding slightly reduces the sharpness of the diffraction pattern, particularly at high spatial frequencies, it does not preclude successful image reconstruction. In fact, Kang et al. noted that such imperfections may even aid in stabilizing the phase retrieval process by diffusing strong aliasing features, though this trade-off remains application-dependent. Nevertheless, continued refinement of micro-fabrication methods remains crucial for pushing the resolution boundaries of single-frame CXDI while maintaining precise phase encoding characteristics through geometrically controlled apertures.

While initial research suggests that moderate corner rounding in triangular apertures does not significantly compromise image reconstruction capabilities, this assumption requires more detailed investigation. Although the studies demonstrated successful phase retrieval using triangular apertures with corner rounding in the nanometer range, their conclusions relied primarily on qualitative image comparison and diffraction pattern matching. A more comprehensive evaluation is needed to quantify how corner geometry influences reconstruction stability, convergence characteristics, and phase retrieval precision across diverse imaging scenarios. Corner radius modifications introduce complex changes to diffraction patterns, particularly

affecting high-frequency components where phase information plays a crucial role. These alterations potentially influence both resolution capabilities and the uniqueness of phase retrieval solutions, especially when working without support constraints. From a practical perspective, the effects of corner smoothing may interact with various system parameters, including probe coherence characteristics, detector sampling specifications, and noise factors, potentially creating compound effects not fully captured in current simulation models or standard test patterns.

This research context necessitates a more thorough analytical approach to evaluate triangular aperture performance under varying experimental conditions. Such analysis should encompass both simulation studies and experimental validation, examining corner radius effects on reconstruction quality across different object types, edge profiles, and aperture curvature specifications. Quantitative assessment metrics, including phase error calculations, Structural Similarity Index (SSIM) measurements, and convergence rate analysis across multiple experimental iterations, can provide a deeper understanding of the relationship between manufacturing tolerances and imaging performance. These comprehensive findings would offer valuable insights for optimizing both aperture design specifications and fabrication in high-resolution CXDI systems.

3.3 Methodology

This research investigation encompasses two fundamental experimental components, which are a comprehensive simulation of single-frame CXDI implementing symmetrical triangle apertures with systematic variations in corner radius, and a detailed evaluation of the phase retrieval performance metrics associated with these aperture configurations. The first component focuses on establishing a robust methodology for diffraction pattern generation, which provides essential baseline data from nanoscale particles. The second investigation component employs both iterative algorithmic approach and advanced deep learning methodology to quantitatively assess the effectiveness of various aperture geometries in the phase retrieval process. A detailed schematic representation of this research methodology is illustrated in Figure 3.2.

3.3.1 Underlying Concepts of Single-Frame CXDI for the Simulation

In single-frame CXDI, diffraction patterns are generated when coherent X-ray interact with nanostructured materials through a specialized probing wavefront (A. M. Maiden and J. M. Rodenburg, 2009). This interaction occurs as coherent X-ray illumination, modified by a carefully designed aperture, encounters a specific sample region, resulting in complex wave scattering that the detector captures in the far-field domain. The detector system can only record light intensity, expressed as $|E(u,v)|^2$, while completely losing phase information $\phi(u,v)$. This phase loss prevents direct image reconstruction through inverse Fourier transformation without additional computational processing.

This study implements a computational framework that simulates the complete diffraction imaging process within single-frame CXDI parameters. This framework incorporates four essential components:

- **Computational design and generation of diverse aperture geometries:** This step involves the numerical modeling of apertures with varied shapes and parameters, such as circular, triangular, and rectangular, taking into account geometric fidelity and fabrication constraints.

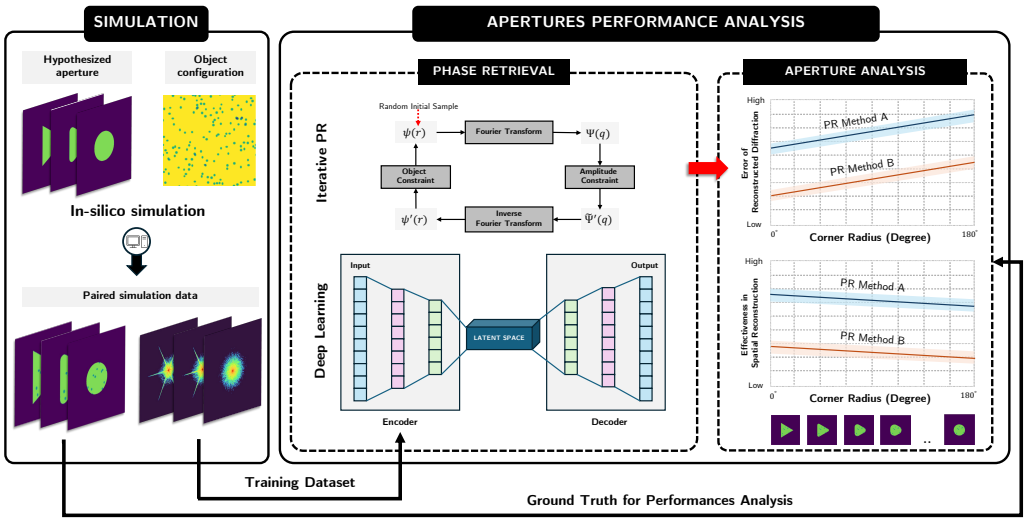


FIGURE 3.2: Our research methodology encompasses two interconnected stages. The first stage focuses on the systematic production of diffraction patterns, specifically utilizing apertures that demonstrate a controlled geometric evolution from triangular to circular forms. Following this, the second stage employs a dual-method analytical approach, incorporating both traditional iterative phase retrieval and deep learning phase retrieval. This comprehensive evaluation framework enables detailed performance assessment of various aperture configurations across both real and reciprocal space.

- **Probing wavefront creation utilizing the Angular Spectrum Method (ASM):** The ASM is used to propagate the wavefront from the aperture to the object plane, enabling the construction of a spatially resolved probe function that reflects physical propagation behavior.
- **Simulation of probe-sample interactions at the object plane:** The complex transmission function of the sample is numerically combined with the incoming probe via point-wise multiplication to form the exit wavefield.
- **Numerical computation of resulting diffraction patterns at the detector level:** The exit wavefield is propagated to the detector using a Fourier transform to compute the intensity-only diffraction pattern, which forms the forward data used in phase retrieval.

Aperture formation involves precise mathematical modeling of geometric characteristics, including shape parameters, boundary conditions, and dimensional specifications based on predetermined optical requirements. The aperture structure is digitally constructed using analytical geometry and grid-based sampling techniques, creating either binary or graduated transmission functions across defined spatial regions. For non-circular or asymmetric configurations, particularly triangular geometries, special consideration is given to edge smoothness and corner radius characteristics to accurately reflect manufacturing constraints that significantly influence diffraction characteristics. The aperture is numerically represented as a transmission matrix, effectively modeling spatial wavefront modulation.

The ASM provides precise probe function generation (Leeman and Healey, 1997). This method decomposes wavefield propagation into plane wave spectra that propagate independently through space. The fundamental ASM equation, expressed in

the frequency domain through the Fourier transformation, enables efficient computational implementation. The probe calculation at distance z is achieved through:

$$P(x,y;z) = \mathcal{F}^{-1}\left[\mathcal{F}\{A(x,y)\} \cdot H(k_x,k_y,z)\right], \tag{3.1}$$

where $A(x,y)$ represents the aperture function and $H(k_x,k_y,z)$ denotes the angular spectrum transfer function.

This probe function interacts with the complex-valued object $x(\mathbf{r})$ through pointwise multiplication, generating the exit wave:

$$\psi(\mathbf{r}) = x(\mathbf{r}) \cdot P(\mathbf{r}), \tag{3.2}$$

where $\mathbf{r}$ indicates real-space coordinates. Fourier transformation of this exit wave produces the reciprocal-space diffraction pattern.

The resulting diffraction pattern, expressed in polar form as $z = r \cdot e^{j\theta}$, comprises both amplitude ($r = |\psi(u,v)|$) and phase (θ) components. However, detector systems only capture intensity values ($|z|^2 = r^2$), losing all phase information. Phase information plays a crucial role in image reconstruction by encoding essential spatial positioning information required for accurate object recovery. Without phase data, the inverse problem becomes mathematically underdetermined, making direct inverse Fourier transformation insufficient for accurate reconstruction (Shechtman et al., 2015).

Consequently, successful object image reconstruction requires preliminary phase recovery or estimation using appropriate computational methods. This phase retrieval challenge represents a fundamental aspect of coherent diffraction imaging and constitutes the core inverse problem in single-frame CXDI applications. All these concepts are used to develop the simulation, so the ground-truth for real space comparison will be available.

3.3.2 Customizing the Iterative Phase Retrieval

While the extended Ptychographical Iterative Engine (ePIE) was initially developed for ptychographic imaging utilizing multiple overlapping diffraction patterns for simultaneous object and probe reconstruction, its application to single-frame CXDI requires significant adaptation. Standard ePIE implementations rely on multiple scan positions and spatial redundancy, elements unavailable in single-frame configurations. Adapting ePIE for single-frame applications necessitates specific algorithmic modifications and constraint redefinitions, maintaining the core alternating update mechanism while accommodating single-diffraction frame limitations and fixed aperture geometries.

Phase recovery in CXDI requires specialized approaches, including iterative algorithms and deep learning methods. Our study implements a modified ePIE algorithm to determine appropriate phase values for each diffraction pattern, enabling exit-wave reconstruction through inverse Fourier transformation. As depicted in Figure 3.2, this iterative process requires both real and Fourier space constraints for phase estimation and refinement. The single-frame context, limited to one object, one probe waveform, and one diffraction pattern per dataset, increases sensitivity to initialization conditions and noise factors. The reciprocal-space constraint utilizes known intensity values, calculated through:

$$I(q) = |\mathcal{F}[O(r)P(r)]|^2, \tag{3.3}$$

where $O(r)$ represents object transmission function and $P(r)$ denotes probe wavefield. The real-space constraint applies through:

$$\psi(r) = O(r)P(r), \tag{3.4}$$

defining object-plane exit wave characteristics.

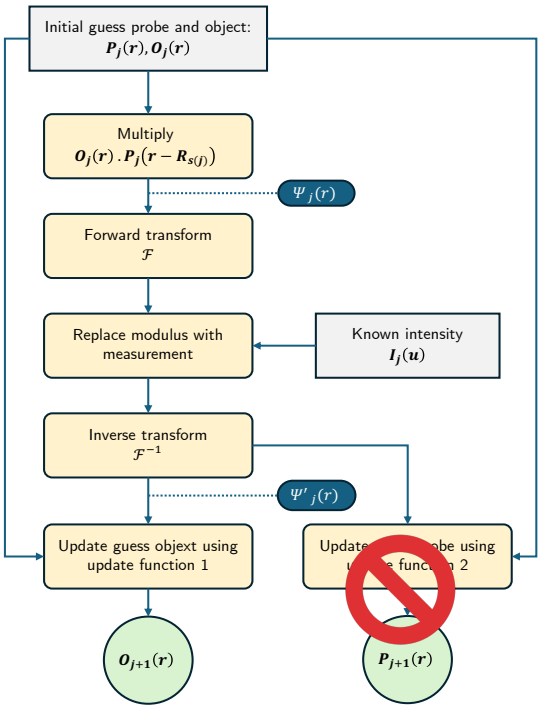


FIGURE 3.3: The customization of the ePIE algorithm of the Ptychography mechanism to single-frame CXDI.

The ePIE implementation requires initial estimates for both object and probe waveforms. Our methodology employs the ASM mechanism for initial probe function simulation, while initializing the object function as uniform transmission. This modification aligns with the intrinsic nature of single-frame CXDI, wherein only one diffraction pattern is captured per measurement without any scanning or probe position modulation. Unlike Ptychographic systems that collect multiple overlapping frames with shifted probes, thereby requiring the simultaneous update of both the object and the probe, the single-frame configuration relies on a fixed probe shape and position throughout the acquisition. Consequently, the probe function can be accurately simulated in advance, often using wave propagation models such as the ASM, and treated as a known, static input in the reconstruction process.

A significant challenge involves managing zero or near-zero exit wave values that can destabilize convergence. This is addressed by updating the exit wave modulus in reciprocal space using measured diffraction intensity:

$$\tilde{\Psi}(q) = \sqrt{I(q)} \frac{\mathcal{F}[\psi(r)]}{|\mathcal{F}[\psi(r)]|}. \tag{3.5}$$

Following the inverse Fourier transformation of this updated pattern, the object function updates according to:

$$O(r) = O(r) + \alpha \frac{P^*(r)}{\max(|P(r)|^2)}(\psi'(r) - \psi(r)), \tag{3.6}$$

where α represents a tunable parameter controlling update magnitude, typically ranging between 0.6 and 0.9 for stable convergence.

The single-frame configuration's lack of probe position diversity presents specific challenges, requiring increased iteration counts and reducing tolerance to noise and model discrepancies. Nevertheless, the adapted ePIE implementation provides an effective single-frame CXDI solution through precise probe modeling and targeted constraint application. The customization of the ePIE algorithm can be examined in Figure 3.3.

3.3.3 Incorporating Deep Learning Phase Retrieval

This research also utilized a customized deep learning-based phase retrieval framework inspired by the PtychoNN model Cherukara et al., 2020, with the objective of complementing and enhancing the reconstruction results obtained through iterative methods. Given the nature of the inverse imaging problem, phase retrieval can be reformulated as an optimization task where the objective is to minimize the discrepancy between the measured and predicted diffraction magnitudes. This is formally expressed as:

$$\arg\min_{O_n} \left| \sqrt{I(u)} - |\mathcal{F}[\psi(r)]| \right|, \tag{3.7}$$

where O_n denotes the reconstructed object at the n-th iteration, $I(u)$ is the measured diffraction intensity, and $\mathcal{F}[\psi(r)]$ is the Fourier transform of the estimated exit wave.

The deep learning model is implemented as a convolutional encoder-decoder network designed to learn the mapping between diffraction patterns and their corresponding amplitude and phase distributions in real space. The network architecture consists of three main components: an encoder, dual decoders (for amplitude and phase), and output activations. In the encoder block, the input diffraction image is processed through a series of four Convolutional Blocks (CB), each consisting of two 3×3 convolutional layers, followed by a 2×2 max-pooling layer to reduce spatial resolution and extract multiscale features. This hierarchical compression enables the model to capture spatially distributed information relevant to the object's global structure.

Following the encoding stage, the network splits into two decoder branches—one for amplitude reconstruction and one for phase reconstruction. Each decoder consists of CB similar to the encoder, but with 2×2 transpose convolution layers to upsample the feature maps and recover spatial detail. Transpose convolutions are used instead of naive interpolation to allow learnable upsampling, improving reconstruction accuracy, particularly near edges and fine structures.

At the final stage, each decoder branch outputs a single-channel prediction: the amplitude branch uses a 3×3 convolution followed by a **sigmoid(.)** activation function to constrain values between 0 and 1, while the phase branch uses the same convolutional size but applies a $\pi \cdot \tanh(.)$ activation to map output values into the range $[-\pi, \pi]$. This separation allows the network to explicitly model the dual

nature of the exit wave, facilitating clearer training dynamics and post-processing interpretability.

The entire network, as shown in Figure 2.8, is trained under an unsupervised learning regime, where the loss function is constructed to compare the predicted diffraction intensity, generated from the predicted amplitude and phase, against the input diffraction pattern. The key idea is to ensure that the forward-generated intensity (via Fourier modulus) matches the measured pattern, without requiring ground-truth object data. This strategy enables the model to learn consistent representations of amplitude and phase over multiple diffraction instances and time steps. Ultimately, this framework provides an end-to-end tool to reconstruct object wavefields and simultaneously evaluate the impact of varying aperture conditions in single-frame CXDI.

3.3.4 Evaluating the Phase Retrieval Performances

Our study implements two complementary evaluation metrics, the R_F factor and the SSIM, to quantitatively assess phase retrieval performance in both real and reciprocal space domains (Hore and Ziou, 2010; Takayama et al., 2021). These metrics enable objective comparison of reconstruction accuracy across different aperture configurations and reconstruction methodologies, encompassing both iterative and deep learning approaches.

For reciprocal space evaluation, we employ the R_F factor to quantify the correspondence between reconstructed and reference diffraction patterns. This metric measures the relative intensity difference between calculated and ground truth diffraction patterns, defined mathematically as (Miao, Chen, et al., 2006):

$$R_F = \frac{\sum_{k_x,k_y} \left| |\mathcal{F}_{sim}(k_x,k_y)| - \alpha |\mathcal{F}_{cal}(k_x,k_y)| \right|}{\sum_{k_x,k_y} |\mathcal{F}_{sim}(k_x,k_y)|}, \tag{3.8}$$

where $|\mathcal{F}_{sim}(k_x,k_y)|$ represents the simulated diffraction intensity from the known object, $|\mathcal{F}_{cal}(k_x,k_y)|$ denotes the calculated diffraction intensity from reconstruction, and α serves as a normalization factor compensating for amplitude scaling variations. Lower R_F values indicate superior matching between predicted and reference diffraction data, suggesting more accurate reciprocal space reconstruction.

For real space assessment, we utilize the SSIM, a metric in image processing that evaluates perceptual similarity through comparison of luminance, contrast, and structural characteristics. In our application, SSIM quantifies the correspondence between reconstructed and ground truth object functions, with particular emphasis on phase components. The SSIM calculation follows:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}, \tag{3.9}$$

where x and y represent the compared images, μ_x and μ_y their respective means, σ_x and σ_y their variances, and σ_{xy} their covariance. Constants C_1 and C_2 ensure computational stability by preventing zero-division scenarios. Our implementation specifically applies SSIM to evaluate phase map reconstruction accuracy against known ground truth values.

The utilization of simulated nanomaterial objects as reference standards enables the generation of precise ground-truth diffraction data and object functions. This approach facilitates accurate quantification of spatial resolution and structural fidelity

in reconstructed images. By applying both R_F and SSIM metrics across iterative and deep learning methodologies, our evaluation framework provides a comprehensive assessment of reconstruction effectiveness under single-frame CXDI conditions.

3.4 Experimental Design of the Simulation

Our simulation framework replicates realistic single-frame CXDI experimental conditions to evaluate phase retrieval methodologies under practical constraints. The framework implements comprehensive modeling of optical systems, aperture geometries, and sample characteristics within a fully parameterized environment, generating datasets suitable for both iterative and deep learning-based phase retrieval analysis.

The simulation architecture incorporates coherent X-ray diffraction principles using Fraunhofer diffraction-based forward imaging models. The implementation encompasses four primary components:

- **Comprehensive parameter definition:** Specification of optical, aperture, and particle characteristics based on realistic physical constraints, including energy, geometry, and sampling considerations.
- **Systematic generation of probe and object wavefunctions:** Construction of spatially resolved probe and object fields on a discrete grid using physical models and propagation techniques.
- **Exit wave propagation modeling:** Simulation of wave propagation from the object plane to the detector plane, governed by the Fraunhofer diffraction condition.
- **Precise numerical sampling of diffraction patterns:** Calculation and digitization of intensity distributions captured at the detector, incorporating effects such as resolution, dynamic range, and noise models.

The illustration of the simulation process in this study can be examined in Figure 3.4.

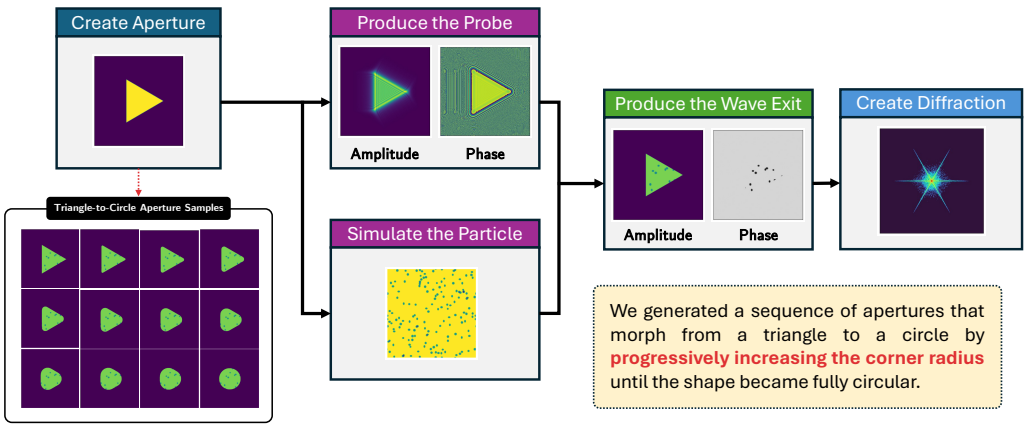
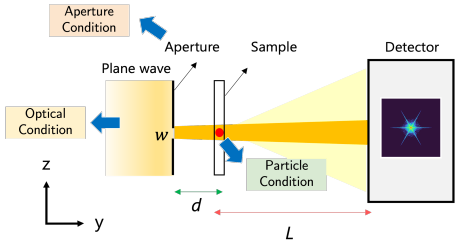


FIGURE 3.4: The illustration of the simulation process for aperture study.

The simulation process in this study involved three components, namely optical, particle, and aperture conditions, as shown in Figure 3.5. Table 3.1 presents the complete simulation parameter specifications. This study involved the morphing of a



Optical Conditions*

Atribute	Definition	Value	Unit
L	Distance between Sample and Detector	3	m
Npixel	Pixel size of sample	512 or 1024	pixel
Energy	Energy for coherent X-ray	5000	eV

Aperture Conditions*

Atribute	Definition	Value	Unit
sideNumber	Number of aperture side	Triangle	-
sideLength	Length of the aperture side	4e-6	m
Lpropagation	The aperture distance to the sample	5e-4	m

Particle Conditions*

Atribute	Definition	Value	Unit
diameter	diameter of the particle	300e-9	m
ratio	aperture : area	calculated	-

* These settings are only for the essential attribute of each condition.

FIGURE 3.5: Three conditions in the simulation settings for aperture study.

symmetrical triangular aperture into a circular one by incrementally increasing the corner radius, while maintaining a consistent aperture area.

The system employs 5.0keV X-ray energy, selected for optimal nanometer-scale feature resolution. Implementation parameters include a 3.14m sample-to-detector distance and 75 μm detector resolution, incorporating a 7-level dynamic range to reflect realistic detector capabilities. The specimen model utilizes gold nanoparticles ranging from 150 to 350nm in diameter, with 50nm incremental steps. Particle distribution occupies 10% of the imaging window, ensuring adequate diffraction signal while maintaining characteristic nanostructure sparsity.

TABLE 3.1: The settings for diffraction imaging simulation using Single-Frame CXDI concept.

Parameter Category	**Setting Description**
OPTICAL CONDITION	
X-ray Energy	$5keV$
Sample-to-detector distance	$3.14m$
Detector resolution	75μm
Pixel resolution at the sample plane	$0.2nm/pixel$
Dynamic range of detector	7
APERTURE CONDITION	
Number of aperture sides	3 (e.g., triangle)
Side length of aperture	5 μm
Aperture-to-sample distance	$5 \times 10^{-4}m$

Continued from previous page

Parameter Category	Setting Description
Corner radius	0° to 180° (increased by 1 per iteration)
PARTICLE CONDITION	
Diameter of nanoparticles	$150 - -350nm$ (increased by $50nm$)
Ratio of particle area to window area	0.1

The computational framework implements 512×512 pixel arrays for object functions, probe representations, and resulting diffraction patterns. Exit wave propagation follows the Fraunhofer diffraction principle. Aperture geometry investigation represents a central component of our simulation strategy. The baseline configuration implements a symmetrical triangle with 5 μm side length. Corner radius variations from 0° to 180° generate 181 distinct aperture configurations, each maintaining consistent area in accordance with Nyquist–Shannon sampling requirements to prevent aliasing effects (Arsenault and Chalasinska-Macukow, 1983). The implementation of aperture includes a 3×3 Gaussian filter application to the aperture function, simulating probe softening and minimizing high-frequency artifacts from sharp edges (Hummel, Kimia, and Zucker, 1987). Figure 3.6 demonstrates Gaussian smoothing effects on diffraction patterns, highlighting reduced speckle patterns and enhanced intensity gradient stability.

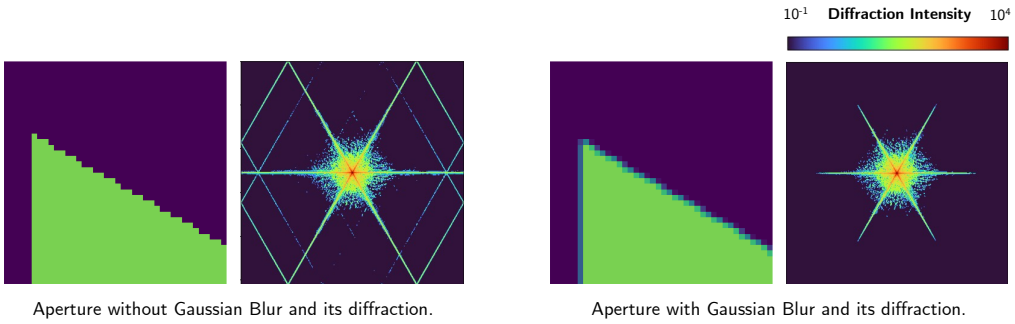


FIGURE 3.6: A comparative analysis of Gaussian function effects on diffraction pattern characteristics, where both diffraction patterns were generated under identical particle and aperture parameters. The diffraction pattern generated without Gaussian blur application exhibits significant noise artifacts compared to its Gaussian-filtered counterpart. For enhanced visualization of the Gaussian blur effects, the aperture region is presented at increased magnification to highlight the distinctions in edge treatment and noise reduction characteristics.

Probe-object interaction modeling employs point-wise multiplication to generate exit waves, subsequently propagated to the detector plane via Fourier transformation for far-field diffraction pattern generation. The detection system incorporates Poisson distribution modeling for intensity values, accurately representing photon count variations. System resolution implements 7-level intensity discrimination to reflect

discrete detector response characteristics. Each phase retrieval methodology underwent 30 experimental iterations, varying three primary parameters: aperture corner radius, nanoparticle dimensions, and spatial positioning. Results aggregation across multiple trials ensures statistical validity. Additional square and pentagon aperture configurations support previous research suggesting enhanced phase retrieval performance with non-point-symmetric apertures in single-frame CXDI applications (Kang et al., 2021).

The experimental framework generates consistent paired datasets for both customized ePIE implementation and deep learning used in this study, enabling direct performance comparison under identical physical and computational conditions. This approach ensures equitable evaluation across real and reciprocal space metrics. Figure 3.7 illustrates the paired datasets that, either in real or reciprocal space.

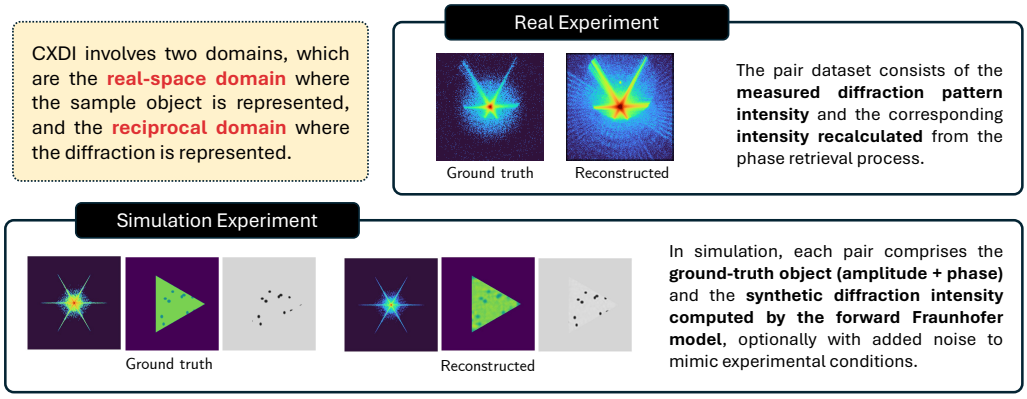


FIGURE 3.7: The paired datasets of single-frame CXDI.

3.5 Case Study 1: Aperture Analysis with Iterative Phase Retrieval

This initial case study examines how aperture geometry transitions, specifically, the progressive transformation from triangular to circular configurations through systematic corner radius modifications, influence iterative phase retrieval performance in single-frame CXDI applications. We implement the ePIE for reconstruction analysis using simulated diffraction patterns generated from diverse aperture configurations. Unlike Ptychography, which relies on overlapping scanning positions to constrain the inverse problem, single-frame CXDI operates without redundancy. Thus, the quality of reconstruction heavily depends on the physical configuration, including the aperture shape and the algorithmic design.

Figure 3.8 shows the reconstruction results for both amplitude and phase. To ensure fair comparison, all output images were clipped to a uniform dynamic range. Qualitatively, it can be observed that reconstructions from apertures with larger corner radius exhibit lower fidelity and stronger artifact presence, particularly near the aperture edges. The increasing presence of noise and spatial distortion is apparent, especially in the phase images, which are more sensitive to inconsistencies in the forward model.

This observation is supported by reciprocal-space evaluation using the R_F score, shown in Figure 3.10 (a). Across the range of corner radius values, R_F scores fall

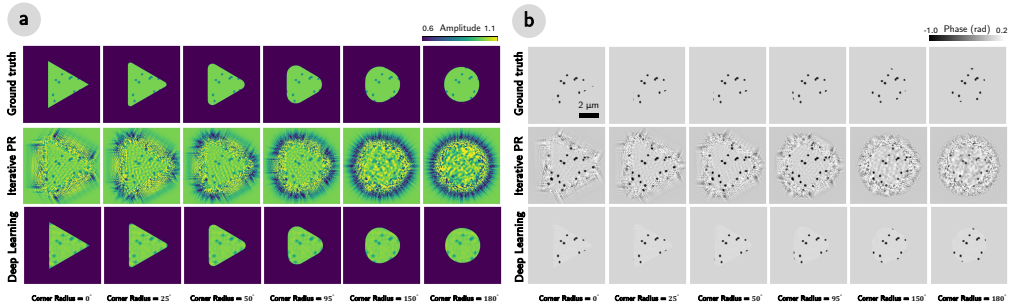


FIGURE 3.8: A comparative visualization of reconstruction outcomes demonstrating both ground truth references and reconstructed results utilizing two distinct methodological approaches. The analysis implements systematic aperture geometry variations, progressing from triangular to circular configurations through incremental corner radius modifications. The results are presented in two complementary components: (a) amplitude reconstruction outcomes and (b) phase reconstruction results.

between 0.83 and 0.96, with an average around 0.90, indicating a high level of inconsistency between the reconstructed and reference diffraction patterns. These results suggest that the enlargement of the corner radius weakens the optical encoding power of the aperture, resulting in less reliable reconstruction and higher error in the reciprocal domain. This finding aligns with practical fabrication challenges, where achieving sharp triangular apertures is often unfeasible (Takazawa, Kang, et al., 2021; Kang et al., 2021), and thus motivates the need for investigating geometric relaxations like rounded corners.

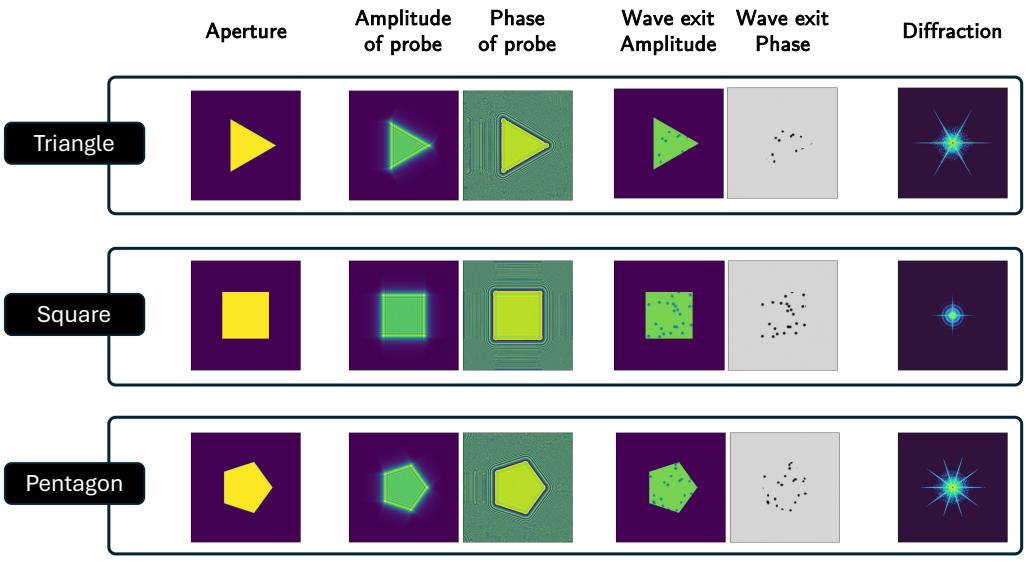


FIGURE 3.9: The supplementary aperture in the comparison of aperture performances.

The real-space evaluation, depicted in Figure 3.10 (b), uses the SSIM to assess the reconstructed images against the ground truth. SSIM values ranged from 0.51 to 0.54, with an average of 0.53, indicating moderate similarity with low variance across all apertures. The triangle aperture with minimal corner radius consistently outperformed more circular variants, suggesting that sharper geometries provide

stronger spatial encoding. However, the narrow SSIM variance implies that real-space reconstruction is less sensitive to corner radius adjustments compared to the reciprocal space.

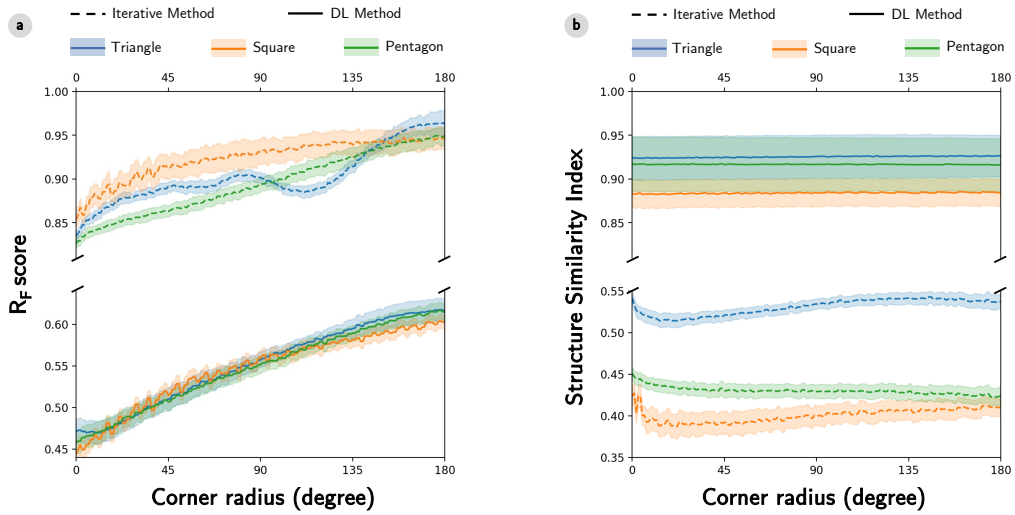


FIGURE 3.10: Comprehensive analysis of aperture performance metrics evaluating phase retrieval effectiveness across multiple domains. The evaluation employs two complementary metrics: the R_F factor, quantifying reconstruction accuracy in reciprocal space, and the SSIM, assessing performance in real space. Extended analysis incorporates comparative evaluation of triangle, square, and pentagon aperture configurations. The results present: (a) comparative R_F scores across both iterative and deep learning methodologies, and (b) corresponding SSIM measurements for both computational approaches, providing a comprehensive performance assessment across reconstruction frameworks.

Interestingly, fluctuations in both R_F and SSIM scores across the aperture configurations can be attributed to geometric factors such as varying aperture coverage of the sample area. Although all apertures maintain the same surface area, the corner radius affects how much of the particle is captured within the aperture projection. As illustrated in Figure 3.1, the circular aperture, having the largest radius, captures a smaller portion of the sample compared to the ideal triangle, which explains some of the performance inconsistencies.

Despite these nuances, the results clearly indicate that iterative phase retrieval in single-frame CXDI performs best with minimal corner radius. This reinforces the notion that while relaxing the aperture shape may ease fabrication, it degrades reconstruction fidelity. Therefore, if iterative algorithms like ePIE are to be employed in practical scenarios, preserving the sharpness of the triangular aperture while keeping its fabrication feasible becomes a critical trade-off.

The supplementary analysis also includes comparisons with other polygonal shapes, such as squares and pentagons, as shown in Figure 3.9. Interestingly, pentagon apertures exhibited lower average diffraction error (R_F) compared to triangles. However, triangle apertures still delivered better real-space image reconstruction. Notably, within a moderate range of corner radii (between 90° and 145°), triangle apertures occasionally outperformed pentagons in diffraction fidelity. These findings

highlight the complex interplay between optical geometry and algorithmic reconstruction behavior and emphasize the importance of aperture optimization in enhancing the reliability of single-frame CXDI systems.

3.6 Case Study 2: Aperture Analysis with Deep Learning Phase Retrieval

For the second case study, we investigate the use of deep learning to recover phase information in single-frame CXDI, with a focus on evaluating how variations in aperture geometry affect reconstruction performance. Unlike iterative phase retrieval methods that rely on repeated forward–inverse operations to gradually refine the phase, the deep learning approach adopts a direct data-driven strategy by learning a functional mapping from the input diffraction data to the reconstructed object. The core objective is to minimize the loss between the predicted diffraction intensity $|\psi_n|^2$, derived from the estimated object function O_n, and the simulated ground truth intensity.

Using the same simulation dataset as in the previous case study, the deep learning model was trained to process diffraction images, along with the corresponding probe and aperture configurations, as its input. The probe function encodes the illumination profile, while the aperture is applied as a multiplicative mask to constrain the reconstruction within the region of interest. This architecture ensures that the predicted object adheres to physical boundary conditions defined by the aperture shape, enabling the network to learn meaningful structural features from the diffraction patterns. Unlike the iterative solver, which updates the object gradually, the trained neural network infers the object in a single forward pass, leveraging the learned relationship between data and structure.

The reconstruction results, shown in Figure 3.8, exhibit higher visual clarity and structural integrity than those obtained with the iterative approach. The network was trained with a batch size of 2 over 20 epochs using a dataset of 181 inputs. Despite the limited dataset, the model achieved a low diffraction intensity loss of 1.73, demonstrating effective learning of the forward imaging physics. Evaluation using the R_F factor and SSIM confirms the superior performance of deep learning. The R_F factor, used to assess reconstruction accuracy in reciprocal space, ranges from 0.47 to 0.62 with an average of 0.55, significantly lower than those achieved via iterative methods. These results suggest that the deep learning model is more resilient to small imperfections in aperture geometry and can maintain reconstruction fidelity even as the corner radius increases.

In contrast to the fluctuations observed in the iterative case, the R_F scores in the deep learning approach exhibit a more stable and gradual trend. This stability implies that deep learning is less sensitive to variations in corner geometry, an important property for real-world applications where precisely fabricated apertures are difficult to produce. The model's robustness is attributed to the unsupervised learning paradigm embedded in the architecture, which allows the network to extract invariant and salient features from noisy and variably shaped diffraction data. This capability enables the deep learning method to achieve high-fidelity reconstructions without relying on stringent physical conditions.

The real-space evaluation, as shown in Figure 3.10 (b), yields SSIM values ranging between 0.926 and 0.929, with an average of 0.928, indicating very high structural

similarity between the reconstructed and ground truth images. Notably, the reconstructions display fewer artifacts within the aperture region, even at higher corner radius values, suggesting that the model is capable of suppressing noise and stabilizing phase recovery across geometric variations. Although the SSIM variance is similar to that observed in the iterative approach, the overall structural quality of the reconstructions is consistently higher.

Additional findings from supplementary experiments further validate the geometric robustness of deep learning. While the symmetrical triangle aperture offers high object reconstruction accuracy, its performance is only marginally lower, by less than 1%, compared to the circular aperture, and the difference is statistically negligible across multiple trials. Furthermore, the pentagon aperture emerges as a viable alternative to the triangle due to its similarly asymmetric structure, outperforming the square aperture in real-space reconstruction. These findings suggest that, although corner radius variations do not drastically impact reciprocal-space performance, geometric asymmetry plays a critical role in real-space image quality. Compared to iterative phase retrieval, the deep learning approach exhibits superior tolerance to shape perturbations, reinforcing its potential for practical deployment in experimental CXDI systems.

3.7 Contributions and Limitations

This research provides a multi-perspective investigation into the effect of aperture geometry, particularly the use of triangle apertures with varying corner radii, on the performance of phase retrieval in single-frame CXDI. The study combines two reconstruction paradigms, iterative phase retrieval using ePIE and deep learning using an encoder-decoder architecture. Through this dual approach, the research delivers new insights into the trade-offs between optical system design and computational reconstruction performance.

The primary contribution lies in the development of an in-silico simulation framework that enables quantitative analysis of reconstruction quality across both real and reciprocal domains. This framework supports both iterative and deep learning-based phase retrieval, allowing direct comparison of accuracy, stability, and sensitivity to aperture shape variations. Notably, the simulation incorporates complex object functions and physically realistic diffraction patterns, establishing a reliable benchmark for evaluating reconstruction fidelity. The results show that while iterative methods are highly sensitive to corner sharpness and geometry-induced artifacts, deep learning models exhibit greater robustness and superior performance across a range of corner radius values.

From a physical instrumentation standpoint, the study introduces an optimized aperture design strategy. The triangle aperture, already known for its favorable diffraction properties, is re-evaluated with a focus on its manufacturability. The analysis reveals that slight rounding of the triangle corners, often unavoidable in practice, has a negligible impact on reconstruction quality when using deep learning models. This finding relaxes the strict requirement for perfectly sharp corners and provides a more realistic design threshold, bridging the gap between simulation and physical fabrication constraints. The simulation further highlights that while circular apertures introduce increased reconstruction artifacts, triangular shapes remain consistently effective across varying conditions.

In terms of computational methodology, the deep learning approach demonstrates clear advantages in suppressing reconstruction noise and stabilizing performance

across different aperture configurations. The network learns to extract latent features that are invariant to minor geometric perturbations, allowing it to maintain high accuracy in both phase and amplitude reconstruction. Although the model requires training with a moderate amount of data, the use of synthetic diffraction data generated from simulation effectively addresses this challenge, making the approach practical for real-world implementation where experimental data is limited.

However, several limitations should be acknowledged. First, the performance of the deep learning model remains dependent on the representativeness and realism of the simulated training data. If the simulation does not adequately reflect experimental noise characteristics or beam coherence conditions, model generalization to real data may be limited. Second, while this study focuses on single-frame setups, the methods have not yet been extended to multi-frame or time-resolved imaging scenarios, which may involve different physical constraints and dynamic considerations. Third, although the use of corner radius as a geometric parameter offers a practical degree of control, other shapes, such as edge curvature or asymmetry, remain unexplored and may further impact reconstruction quality.

Despite these limitations, the contributions of this work are substantial in both the computational and physical dimensions of CXDI. The study establishes a replicable pipeline for evaluating aperture influence on phase retrieval, offers practical guidance for optical design optimization, and introduces a deep learning-based method that improves reconstruction reliability. By integrating physics-informed modeling with data-driven inference, this research lays the groundwork for more robust and interpretable phase retrieval strategies in nanoscale imaging.

Chapter 4

Ensemble Phase Retrieval for Cross-Validated Reliability and Phase Ambiguity Mapping of Single-Frame CXDI Reconstruction

4.1 Introduction

Single-frame CXDI has revolutionized computational imaging, especially in nanoscale imaging, by enabling non-invasive, high-resolution visualization of internal structures within nanomaterials (Y. Wang et al., 2008; Withers et al., 2021; A. S. Johnson et al., 2023). This technique plays a crucial role in material science research, particularly in exploring structural dynamics, defect formation, and interface behaviors in real or near-real time. By combining coherent X-ray illumination with advanced computational reconstruction techniques, CXDI surpasses the limitations of traditional lens-based systems, offering access to detailed internal features that were previously inaccessible (Keller, 1957; Henry N Chapman and Nugent, 2010; I. Robinson and Harder, 2009). The fundamental imaging principle involves illuminating a sample with a coherent X-ray beam and recording the resulting diffraction pattern. This pattern captures the intensity distribution of scattered waves but inherently omits the corresponding phase information, making reconstruction of the original object non-trivial and computationally intensive.

The image formation process in CXDI is governed by the interference of scattered wavefronts in reciprocal space. Although the measured diffraction pattern encodes valuable information, the absence of phase prevents a direct inverse Fourier transform. Consequently, the challenge of phase retrieval, reconstructing the lost phase from intensity-only measurements, becomes central to the technique's success (C. G. Schroer et al., 2008; I. K. Robinson et al., 2001; Prosekov, Nosik, and Blagov, 2021; Van der Veen and Pfeiffer, 2004). Compared to scanning-based techniques such as X-ray ptychography, which utilizes overlapping probe positions to introduce redundancy (J. Rodenburg and A. Maiden, 2019; Pfeiffer, 2018; Zheng et al., 2021; Konda et al., 2020; Sidorenko and Cohen, 2015), single-frame CXDI offers the advantage of rapid acquisition and reduced radiation exposure. However, this benefit comes at the cost of increased ambiguity in the inverse solution, as the lack of overlapping regions limits available constraints for robust phase recovery (Marchesini et al., 2003; Takazawa, Dao, et al., 2023; Takazawa, Sasaki, et al., 2025). The experimental setup used in this study is illustrated in Figure 4.1.

Recent advances in CXDI have emerged from parallel progress in both experimental instrumentation and computational algorithms. On the hardware side, innovations

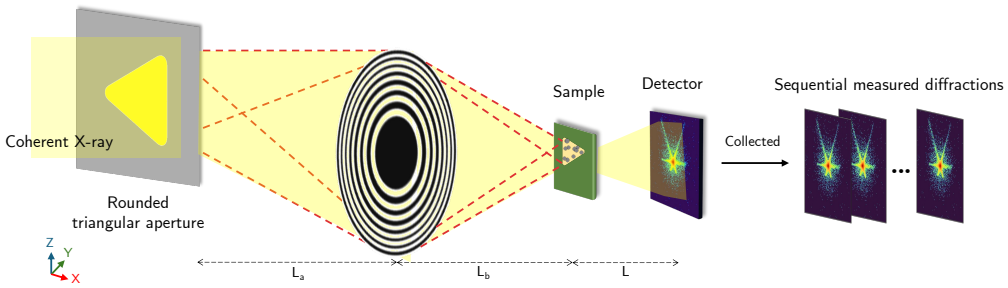


FIGURE 4.1: A schematic of a single-frame CXDI experiment. The forward process of single-frame CXDI (also known as single-shot CXDI). The simulation in this study followed the schematic to produce the sequence of diffraction images from simulated sample images, which served as the ground truth.

such as the use of FZP and custom-designed aperture geometries have improved illumination coherence and wavefront control (Caro et al., 2013; C. Kim, Y. Kim, C. Song, et al., 2016; Shi, 2018; Takazawa, Kang, et al., 2021; Kang et al., 2021; Mukharil Bachtiar et al., 2024). On the algorithmic front, iterative phase retrieval methods, including ER, HIO, and difference-map techniques, have been extensively developed to enhance reconstruction robustness and convergence reliability (Fienup, 1982; Bauschke, Combettes, and Luke, 2002; Jaganathan, Eldar, and Hassibi, 2016; Elser, 2003; Y. Zhang, Tian, and R. Liang, 2019; Luke, 2017; Latychevskaia, 2018). Complementing these efforts, machine learning-based approaches such as PtychoNN and PID3Net have demonstrated significant promise in learning complex, nonlinear mappings from diffraction intensity to phase information. These models incorporate convolutional neural networks and physics-informed constraints, respectively, to improve generalization and resistance to experimental noise and missing data (Cherukara et al., 2020; Vu et al., 2025.

At the core of the phase retrieval problem lies its ill-posed nature. Diffraction patterns are invariant to certain transformations of the object, such as translations, global phase shifts, or inversions. Multiple phase distributions can fit with the same diffraction intensity. More problematically, complex ambiguities such as zero-flipping in the Fourier domain can generate drastically different reconstructions from the same data (Muoi et al., 2018; Barnett et al., 2020). Figure 4.2 illustrates the ambiguity problem in single-frame CXDI. To mitigate these ambiguities, most algorithms apply additional constraints informed by physical or experimental knowledge (Kumar, 2021; Xiang et al., 2024; X. Yang et al., 2024). These typically include finite support constraints that limit the object to a known spatial region, non-negativity assumptions that are relevant to quantities like electron density, and oversampling conditions to increase redundancy and reduce degrees of freedom (Barnett et al., 2020; Osherovich, 2012; Beinert, 2017; J. Dong et al., 2023). While these strategies reduce solution ambiguity, they are not always sufficient, particularly in data-limited or noise-corrupted experimental contexts.

The source of the ambiguity problem in phase retrieval lies in the behavior of a single solver, which typically enforces the reconstruction toward one solution that is highly sensitive to initial conditions, algorithmic hyperparameters, and implicit biases introduced by the imposed constraints. This tendency to converge to a single, seemingly plausible result, without adequately exploring the space of other mathematically valid reconstructions, as shown in Figure 4.3. This also limits the phase

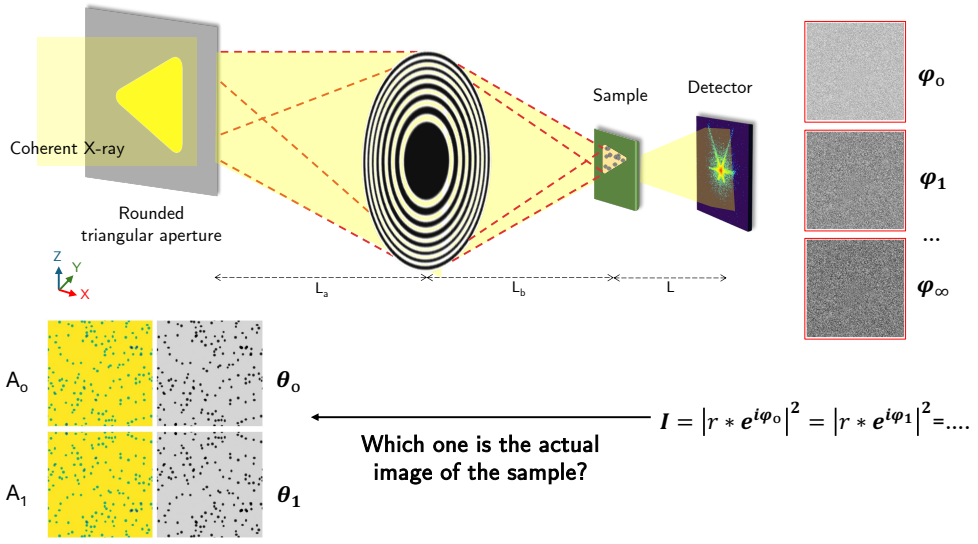


FIGURE 4.2: The illustration of the phase ambiguity problem in single-frame CXDI.

retrieval solver's ability to capture the full ambiguity inherent in the inverse problem. In the specific context of single-frame CXDI, this phenomenon can be interpreted as a form of overfitting. Much like in machine learning, where a model becomes overly adapted to a narrow representation of training data and fails to generalize, a phase retrieval solver may become overly committed to a particular reconstruction path, often settling into a local optimum that satisfies the numerical criteria but lacks epistemic robustness.

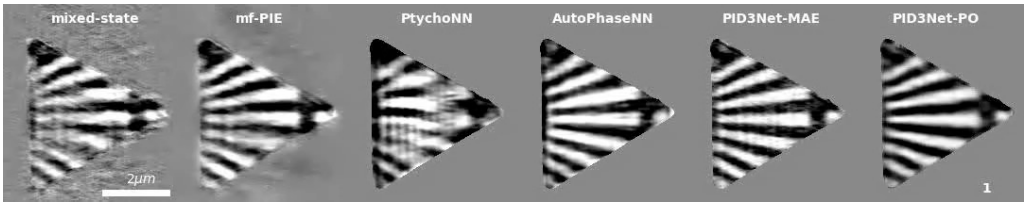


FIGURE 4.3: Overcommitment of the phase retrieval solver to one exact solution as a source of ambiguity.

This overcommitment is particularly problematic in CXDI, where the missing phase information renders the inverse problem ill-posed. Therefore, relying on a single deterministic reconstruction obscures the fact that alternative, equally valid solutions may exist. Without mechanisms to expose this multiplicity, the resulting image can appear as the precise solution while in reality failing to reflect the underlying ambiguity of the measurement. Moreover, standard error metrics can be misleading because they may rate a reconstruction as highly accurate simply because it fits the measured intensity data well, even if it is significantly incorrect in regions where no data were observed or where the solution had to be extrapolated. Addressing this issue requires a shift from single-solution determinism to ensemble-based inference frameworks that explicitly acknowledge and quantify uncertainty across the solution space.

To mitigate the risk of overfitting in phase retrieval, where a single solver may overly commit to one reconstruction path based on its initialization and constraints,

this study introduces a cross-validation-inspired mechanism using spatial data skipping. The key idea is to probe reconstruction stability and sensitivity by temporarily removing subsets of the input during the reconstruction process. However, unlike standard learning tasks where the dataset can be cleanly partitioned into independent training and validation sets, coherent X-ray diffraction imaging (CXDI) presents a unique constraint where the diffraction data and the real-space object are coupled through a deterministic, one-to-one Fourier relationship. This mutual dependence makes arbitrary data splitting physically invalid, as it disrupts the underlying coherence required for successful phase retrieval. Therefore, the proposed data-skipping mechanism must be applied with spatial locality in mind. Specifically, a sliding-window skipped region, as depicted in Figure 4.4, is applied to mask different localized regions of either the diffraction pattern or the object domain. Each skipped region acts as a blind spot, compelling the solver to infer missing information from the remaining visible data, and thus revealing how robust the reconstruction is to partial data omission. This mechanism mirrors the logic of cross-validation in conventional machine learning, where models are trained on partial data and evaluated on held-out segments to assess generalization performance. By using structured windowed masking rather than disjoint partitioning, the framework preserves the physical consistency of the imaging process while still enabling a principled evaluation of reconstruction reliability.

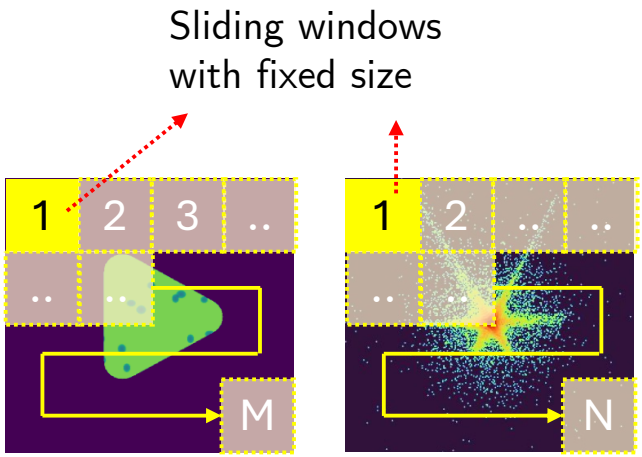


FIGURE 4.4: A schematic representation demonstrating the implementation of sliding window methodology for dataset processing. The illustration depicts how the sliding window concept systematically masks and segments the dataset, providing a visual explanation of the windowing operation and its application to the data analysis framework.

That mechanism will lead the process to produce multiple reconstructions from each solver. These multiple reconstructions, each derived from a distinct masking configuration, are then aggregated within an Ensemble Phase Retrieval. Unlike traditional ensemble learning, which typically combines predictions from heterogeneous models to reduce generalization error, the ensemble phase retrieval approach emphasizes epistemic robustness. It uses architectural or loss-function symmetry to explore the solution space and, in parallel, quantify the variance of aggregated reconstructions as the reconstruction ambiguity. The final output is obtained through soft voting, where the pixel-wise mean provides the consensus image and the variance

map reflects localized ambiguity. This dual output enhances both interpretability and reliability in single-frame CXDI, aligning phase retrieval with the principles of explainable and uncertainty-aware inference.

This ensemble phase retrieval strategy allows for two complementary objectives. The first is to evaluate the stability and reproducibility of reconstruction across varying information availability. The second, to spatially localize areas of high ambiguity, offering interpretability into the phase retrieval process. The proposed method can be applied uniformly across multiple phase retrieval algorithms, iterative and learning-based, enabling algorithm-independent analysis of ambiguity and solution variability. By leveraging skipped region masking and ensemble inference, this approach transforms phase retrieval from a deterministic solution process into a probabilistic framework, enabling richer scientific interpretation and greater robustness under real-world constraints.

4.2 Applying Cross Validation and Ensemble Learning Mechanism into Phase Retrieval Problem

As mentioned in the previous section, there are two mechanisms to overcome the problem. The first one is implementing a cross-validation mechanism to prevent overfitting in each phase retrieval solver, and the second is aggregating the multiple solutions using Ensemble Phase Retrieval. In conventional machine learning, cross-validation is achieved by splitting the dataset into disjoint subsets, allowing the model to be trained on one portion and validated on another. However, this approach cannot be directly applied to single-frame CXDI, because the diffraction pattern and the reconstructed real-space object are linked through a deterministic Fourier transform. Any arbitrary partitioning of the data would break this relationship and compromise the physical consistency required for valid reconstruction. To address this, the cross-validation concept must be reinterpreted in a spatially localized manner. Instead of dividing the dataset, small regions of the diffraction or object domain are systematically masked, temporarily not involved, during each reconstruction. These masked regions are referred to as **skipped regions**.

Each skipped region introduces a controlled blind spot, forcing the solver to reconstruct the object without access to a specific subset of data. This setup effectively simulates a validation scenario, where the solver's ability to infer missing information from surrounding signals is tested. By systematically shifting the skipped region across different spatial locations using a sliding-window approach, it is possible to highlight the robustness and generalizability of the solver across the entire domain. If the reconstruction quality significantly degrades when a particular region is skipped, it indicates high sensitivity and potentially poor generalization in that area. This strategy mimics the principles of cross-validation while maintaining the physical coherence of the CXDI inverse problem, making it a suitable adaptation for evaluating and stabilizing single-frame reconstructions. The outcome of this process is not only a set of reconstructed images generated from different masked versions of the input but also a deeper understanding of which parts of the reconstruction are stable and which are vulnerable to perturbations. These insights form the basis for the second mechanism, which is the Ensemble Phase Retrieval. This approach was inspired by the Ensemble learning approach in the Machine Learning domain, even though the mechanism would be different.

Ensemble Learning has become a foundational approach in machine learning to enhance model reliability, generalization, and ambiguity quantification by aggregating multiple predictors. When applied to ill-posed inverse problems such as phase retrieval, ensemble learning offers an opportunity to move beyond single-solution outputs and explore the broader solution landscape. In the context of single-frame CXDI, where the lack of phase information and data redundancy amplifies the non-uniqueness of the inverse mapping, ensemble learning can be repurposed to systematically evaluate the stability, robustness, and ambiguity of reconstructed images.

Mapping ensemble learning to the phase retrieval problem involves reinterpreting core ensemble components, such as model diversity, independent training pathways, and output aggregation, within the constraints of physics-based imaging. In conventional machine learning, ensemble diversity is typically achieved by training models on different subsets of data or using different random seeds. In phase retrieval, the recorded diffraction data represents a single time-invariant state of the object. It captures its structural configuration at one specific moment in the temporal dimension. Therefore, to introduce diversity, the ensemble must be generated by varying the inference conditions under which the reconstruction is performed. This includes:

- Using different phase retrieval algorithms (e.g., ER, HIO, PID3Net) to capture variability in solution behavior across different reconstruction strategies.
- Applying data masking strategies, specifically skipped region masking, serves as a targeted diagnostic tool to assess reconstruction sensitivity and local ambiguity across the object domain. By systematically omitting localized portions of the diffraction pattern using a sliding window, this approach introduces controlled perturbations that reveal how much the reconstruction depends on specific spatial features. The size of the skipped region plays a critical role in this diagnostic process. Smaller masks highlight fine-grained sensitivity without significantly disrupting the global reconstruction, while larger masks introduce more substantial information loss, challenging the solver to extrapolate broader missing content. By sweeping the skipped region across the diffraction domain, the sliding window mechanism enables comprehensive spatial coverage. This allows for a dense sampling of the reconstruction's response to localized data absence. This technique not only highlights regions of ambiguity, where the reconstruction is less stable or more ambiguous, but also enables the generation of pixel-wise variance maps that quantify the local reliability of the inferred image.

Each reconstruction obtained from varying the phase retrieval process, whether through algorithm choice, initialization, data masking, or hyperparameter tuning, can be viewed as a **weak model** in the ensemble learning framework. Individually, these reconstructions may only partially capture the true object structure, each introducing its own artifacts or biases due to the underdetermined nature of the phase retrieval problem. Similar to the bagging paradigm in machine learning, aggregating such diverse weak models can enhance robustness and suppress idiosyncratic errors, leading to a more stable and interpretable solution. However, while the analogy to ensemble learning is useful, it cannot be directly implemented in the CXDI domain. In classical ensemble methods like bagging or boosting, there is no privileged or ground-truth model available during inference. Each weak learner contributes equally toward the final decision. In contrast, in the context of CXDI, we retain access to a reconstruction generated without any data masking, the one that uses the full

diffraction pattern, which serves as a reference or anchor solution. This anchor is not guaranteed to be correct, but it provides a useful comparative baseline for interpreting ensemble outputs. It allows us to assess how the ensemble result aligns with or improves upon the unperturbed reconstruction, offering a unique hybrid between conventional ensemble logic and domain-specific physical reasoning. This difference underscores the epistemic flexibility of the ensemble phase retrieval approach, where multiple weak reconstructions are not just fused blindly but interpreted relative to a full-data benchmark within a physically grounded framework.

By treating each distinct phase retrieval configuration as a weak learner, the ensemble framework operates analogously to bagging, where multiple noisy or uncertain predictions are combined to reduce variance and suppress overfitting to a particular reconstruction pathway. This is particularly relevant in phase retrieval, where multiple solutions may satisfy the same diffraction intensity constraint but differ in phase configuration or spatial features. The diversity introduced by algorithmic and input variation thus becomes a source of strength, allowing the ensemble to form a consensus solution that is more reliable than any individual reconstruction.

After producing the distinct reconstructions using the skipped region mechanism, the final step in ensemble learning is aggregation. In regression tasks, this often involves averaging predicted outputs in classification. Soft voting over probability distributions is used. In phase retrieval, we adopt a similar principle by aggregating real-space reconstructions pixel-wise across ensemble members. For each pixel, the ensemble average represents the consensus estimate, while the variance provides an estimate of reconstruction uncertainty using the formula as follows:

$$\bar{O}(\mathbf{r}) = \frac{1}{M}\sum_{i=1}^{M} O_i(\mathbf{r}), \tag{4.1}$$

$$\text{Uncertainty}(\mathbf{r}) = \sqrt{\frac{1}{M}\sum_{i=1}^{M}\left(O_i(\mathbf{r}) - \bar{O}(\mathbf{r})\right)^2}, \tag{4.2}$$

where $O_i(\mathbf{r})$ is the reconstruction from the i-th ensemble member, and M is the total number of reconstructions.

This ensemble aggregation not only improves the robustness of the final output by averaging over noise and artifacts but also provides interpretability through spatial uncertainty maps. Regions with low variance indicate structural features that are consistently recovered across multiple reconstructions. Thus, likely to be reliable, whereas high-variance regions reveal areas where ambiguity exists. Such spatial interpretation is especially valuable in CXDI, where the physical structure of the sample must be inferred without direct imaging and where experimental artifacts can obscure the reconstruction.

In summary, adapting ensemble learning to the context of phase retrieval involves redefining ensemble members not as independently trained models but as reconstructions generated under systematically varied physical and algorithmic conditions, such as skipped region masking and solver options. Unlike classical ensemble learning, where each model operates in isolation and no reference solution exists, the ensemble framework in CXDI benefits from the presence of a full-data reconstruction that can serve as an anchor for comparison. This re-imagined ensemble paradigm enables not only aggregation for robustness, but also structured probing of reconstruction sensitivity and ambiguity. It allows researchers to assess where reconstructions are consistent, where they diverge, and how these differences relate to data support and

algorithmic behavior. In doing so, it transforms phase retrieval from a deterministic point estimation task into a distribution-aware inference process. This is better aligned with the ill-posed and probabilistic nature of coherent imaging problems.

4.3 Methodology

The central challenge in CXDI extends beyond the technical task of phase retrieval. It includes the more fundamental and unresolved issue of **non-uniqueness**, wherein multiple distinct object configurations in real space can yield identical diffraction patterns in reciprocal space. This ambiguity persists even when the diffraction intensity is perfectly measured, as the lost phase component introduces degrees of freedom that are not constrained by the data alone. In such a context, relying on a single reconstruction may give a false sense of accuracy while masking deeper uncertainties inherent to the inverse process.

This study has a focus on designing around two interrelated objectives. First, we simulate the forward imaging process of single-frame CXDI using an FZP as a focusing optic, allowing for physically realistic generation of diffraction data from known object structures. This simulation produces a ground-truth dataset against which reconstruction outcomes can be objectively compared. The use of FZP ensures high spatial resolution and coherence, mimicking conditions typical in modern synchrotron experiments. Second, we evaluate the robustness of the phase retrieval process by introducing controlled data loss through a skipped region mechanism. By systematically masking portions of the diffraction image using a sliding window, we simulate scenarios where parts of the measurement are missing or degraded, whether due to detector limitations, noise, or beam instabilities. Multiple reconstructions are then performed across these varied masking configurations, effectively forming an ensemble of plausible solutions.

By analyzing the consistency across this ensemble, the proposed framework enables the identification of structurally stable regions, those features that persist across reconstructions despite partial data loss. Furthermore, spatial variability in the ensemble is leveraged to construct ambiguity maps, offering a visual and quantitative assessment of which regions of the reconstructed image are reliable and which are ambiguous. This dual analysis, reconstruction reliability, and uncertainty interpretation, provides a more comprehensive understanding of phase retrieval quality and opens a new direction in addressing the non-uniqueness problem in single-frame CXDI.

4.3.1 CXDI Simulation using FZP

This study performs a CXDI simulation, which is similar to the real experiment. This simulation builds upon the modeling strategy used in the previous aperture design study, but with a key modification: the incorporation of FZP as a focusing optic in the beamline. In contrast to the earlier study, where the probe wavefield was approximated more generically, this approach simulates a physically realistic illumination system. The FZP is modeled as a binary amplitude lens composed of alternating transparent and opaque concentric rings, designed to satisfy the constructive interference condition:

$$r_n^2 = n\lambda f, \tag{4.3}$$

where r_n is the radius of the n-th zone, λ is the X-ray wavelength, and f is the focal length of the FZP. This configuration ensures coherent focusing of the incident beam and is commonly used in synchrotron and XFEL-based nanofocusing experiments.

To incorporate realistic experimental asymmetries, the FZP is placed with a slight lateral offset from the optical axis. This introduces a degree of spatial variation in the probe profile, mimicking imperfections commonly encountered in practice. The beam is then propagated from the FZP to the triangular aperture and subsequently to the sample plane using the ASM (Leeman and Healey, 1997), a numerically efficient approach for modeling near-field and far-field wave propagation. This results in a complex, spatially structured probe wavefield that captures phase curvature, intensity gradients, and beam divergence.

Following propagation, the probe is cropped around its high-intensity region and normalized to match realistic photon flux conditions. This normalized probe field illuminates the sample, generating the complex-valued exit wave:

$$\psi_{\text{exit}}(x, y) = O(x, y) \cdot P(x, y), \tag{4.4}$$

where $O(x, y)$ is the complex transmission function of the sample and $P(x, y)$ is the probe wavefield. This exit wave then propagates to the far-field detector, where its Fourier transform gives the measured diffraction intensity:

$$I(q_x, q_y) = \left| \mathcal{F} \left\{ \psi_{\text{exit}}(x, y) \right\} \right|^2. \tag{4.5}$$

The resulting diffraction intensity $I(q_x, q_y)$ serves as the input for downstream phase retrieval tasks. Since the phase component of the wave is irretrievably lost in the measurement, computational methods are required to reconstruct the object from this intensity-only data. The simulated dataset produced by this forward model is particularly well-suited for studying the effect of data perturbation (e.g., skipped regions) and for quantifying the stability and uncertainty of phase retrieval algorithms under the real space domain.

4.3.2 Two Concepts of Iterative Phase Retrieval

Phase retrieval in Coherent X-ray Diffraction Imaging (CXDI) is commonly approached through iterative algorithms that alternate between real-space and reciprocal-space constraints. In this study, we employ two specific algorithms from the Ptychographic Iterative Engine (PIE) family: the **Mixed-state PIE** and the **Multiframe PIE**. These algorithms are chosen due to their robustness in handling non-ideal experimental conditions, such as partial coherence and temporal object variation, which are prevalent in realistic single-frame CXDI data.

The Mixed-state PIE algorithm (A. M. Maiden and J. M. Rodenburg, 2009) is an extension of the standard PIE framework and is designed to accommodate scenarios involving partially coherent illumination, multiple object modes, or fluctuations in the sample state. In contrast to the extended PIE (ePIE), which assumes a single coherent state and updates both the object and probe at each step, the mixed-state variant holds the probe $P(x, y)$ constant and attributes the reconstruction variability to multiple coherent object components.

The forward model in this approach begins with computing the exit wave as:

$$\psi_{\text{exit}}(x, y) = P(x, y) \cdot O(x, y), \tag{4.6}$$

where $O(x, y)$ is the current estimate of the object's complex transmission function. The exit wave is Fourier-transformed to reciprocal space, and the modulus is replaced with the square root of the measured intensity, preserving the current phase:

$$\widetilde{\psi}_{\text{exit}}(q_x, q_y) = \sqrt{I(q_x, q_y)} \cdot \frac{\mathcal{F}[\psi_{\text{exit}}]}{|\mathcal{F}[\psi_{\text{exit}}]| + \varepsilon}. \tag{4.7}$$

An inverse Fourier transform brings the corrected wave back to real space, and the object function is updated using the residual between the corrected and original exit waves:

$$O(r) = O(r) + \alpha \frac{P^*(r)}{\max |P(r)|^2} \left(\psi'(r) - \psi(r)\right), \tag{4.8}$$

where α is a feedback parameter and $P^*(r)$ denotes the complex conjugate of the probe. This algorithm is particularly effective when the recorded diffraction pattern results from a statistical mixture of object states, which may arise due to sample drift, radiation damage, or beamline instability. Figure 4.5 illustrates the customization of mixed-state PIE for this study.

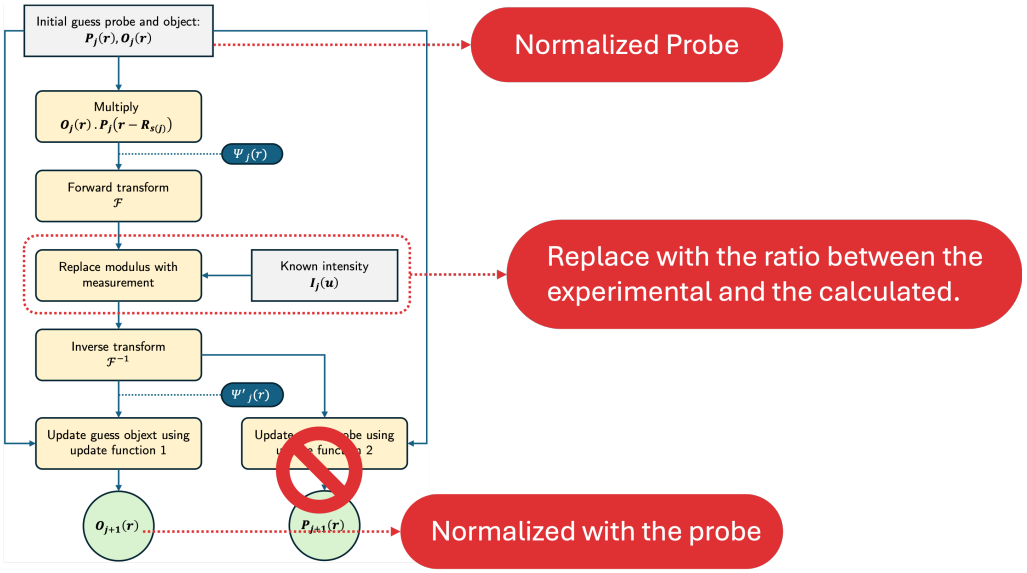


FIGURE 4.5: The customization of mixed-state PIE for this study.

To address time-dependent changes in the sample and add the variant of weak model, we also implemented the Multiframe PIE algorithm (Takayama et al., 2021), which extends the PIE framework by explicitly modeling object evolution across multiple temporal frames, as shown in Figure 4.6. Unlike the mixed-state approach, which treats variability as a statistical mixture, Multiframe PIE introduces a temporally coupled object $O_t(x, y)$ that evolves across exposure intervals. The recorded intensity for frame n is modeled as an integral over time:

$$I_n = \int_{t_{n-1}}^{t_n} |\mathcal{F}[P \cdot O(t)]|^2 \, dt, \tag{4.9}$$

which is approximated in practice as an incoherent sum over overlapping subframes. Each subframe is associated with a temporally local object slice O_{m+l} that contributes to the integrated intensity.

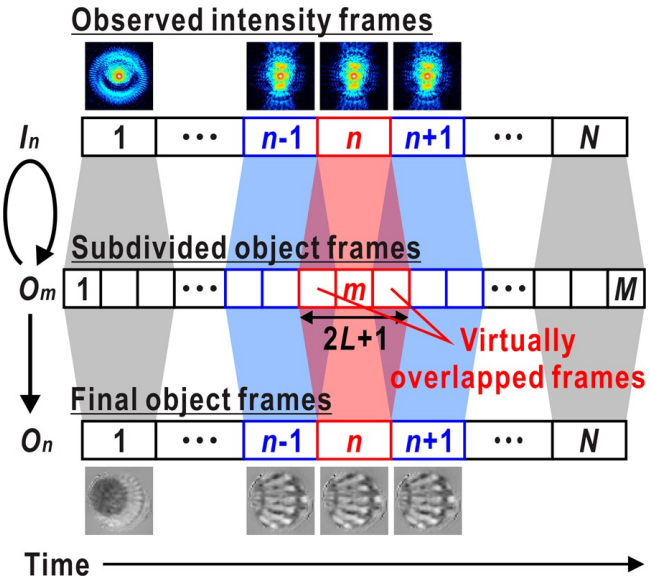


FIGURE 4.6: The illustration of multiframe PIE.

The object update rule generalizes the standard PIE formulation by spreading the feedback from each intensity frame to its corresponding subframes:

$$O'_{m+l} = O_{m+l} + \alpha \frac{P^*(r)}{\max |P(r)|^2} \left(\psi'_{m+l}(r) - P \cdot O_{m+l}\right). \tag{4.10}$$

This formulation ensures smooth temporal evolution by allowing the object to adapt gradually across frames. It is especially beneficial for weakly dynamic systems or low-contrast features that may otherwise be lost in conventional static reconstructions.

Both Mixed-state PIE and Multiframe PIE are conceptually and operationally distinct from the more commonly used extended PIE (ePIE) algorithm. In ePIE, both the object and the probe are updated iteratively and synchronously, making it ideal for systems with unknown or fluctuating probe profiles. However, ePIE assumes full coherence and single-state object behavior, which limits its effectiveness in more complex or noisy imaging scenarios. In contrast, Mixed-state PIE holds the probe fixed and models multiple object modes, improving robustness under partial coherence. Multiframe PIE further generalizes this by introducing explicit temporal dynamics, making it better suited for dynamic imaging tasks or frame-integrated measurements.

In our implementation, both Mixed-state PIE and Multiframe PIE were run with 100 iterations and identical feedback parameters, enabling a fair comparison across conditions. While Mixed-state PIE was effective for static reconstructions under data masking, Multiframe PIE demonstrated higher stability and improved detail retention in sequences with simulated temporal variation and incomplete data due to skipped regions.

4.3.3 Employing the Deep Phase Retrieval in the Ensemble Model

To represent the deep learning branch within the ensemble phase retrieval framework, we utilize **PID3Net**, a physics-informed deep neural network specifically developed for phase retrieval in single-shot (also known as single-frame) CXDI of

dynamic and complex samples (Vu et al., 2025). PID3Net stands out from conventional neural network architectures due to its explicit integration of physical models into its computational structure. Rather than functioning as a purely data-driven black box, PID3Net embeds domain-specific physics constraints to ensure consistency between predicted reconstructions and the fundamental wave propagation behavior of coherent scattering systems.

The network architecture of PID3Net is divided into two key modules, namely the **Temporal Block (TB)** and the **Refinement Block (RB)**. The TB module utilizes 3D convolutional layers to extract spatial and temporal features from sequences of diffraction images. Although our study does not focus on dynamic samples, the temporal structure in the TB provides a powerful mechanism for learning contextual correlations across multiple data representations, such as masked or perturbed inputs. This design allows PID3Net to operate on a stack of varied input conditions, including those generated by the sliding window masking strategy, making it highly adaptable to the ensemble learning scenario.

The RB module serves as the core physics-informed engine of the network. It imposes a forward-model constraint by ensuring that the predicted object $\hat{O}(x, y)$, when combined with the probe function $P(x, y)$ and transformed via Fourier propagation, yields a diffraction intensity

$$\hat{I}(q_x, q_y) = \left| \mathcal{F} \left\{ P(x, y) \cdot \hat{O}(x, y) \right\} \right|^2 \tag{4.11}$$

that closely approximates the measured diffraction pattern $I(q_x, q_y)$. The difference between $\hat{I}$ and I is quantified using a Poisson-based loss function that models the stochastic nature of photon-counting detectors, making it well-aligned with the statistical properties of simulated and experimental CXDI data.

Furthermore, PID3Net enforces physical consistency through multiple levels of constraint:

- **Forward model alignment:** The predicted object must satisfy the diffraction relation with the measured intensity.
- **Backpropagation of physical error:** The RB module calculates residuals in reciprocal space and refines the object in real space to minimize model mismatch.
- **Total Variation (TV) Regularization:** Applied in both spatial and temporal dimensions to promote smooth, noise-suppressed reconstructions.
- **Self-supervised learning:** The model is trained entirely without access to ground truth object images, relying solely on physics-based constraints and the observed intensity patterns.

In this ensemble-based study, PID3Net is executed multiple times with variations introduced via sliding window masking. Each masked input represents a slightly different observation of the diffraction data, allowing PID3Net to generate a diverse set of reconstructions. These reconstructions are aggregated along with other iterative phase retrieval outcomes as part of the ensemble. Within this context, each PID3Net output serves as a **weak model** in the ensemble learning sense, individually imperfect but collectively informative.

PID3Net is chosen as the representative deep learning component in this ensemble not only because of its state-of-the-art reconstruction performance but also due to its unique ability to balance data fidelity with physical interpretability. Its integrated

forward model and Poisson-based optimization make it inherently suitable for the noisy and incomplete data scenarios simulated in this study. Additionally, the network's ability to operate in a self-supervised manner complements the ensemble design, which does not rely on access to ground truth. These properties collectively make PID3Net an ideal candidate for deep phase retrieval under ensemble-based uncertainty analysis.

4.3.4 Aggregating Multiple Reconstructions in Ensemble Phase Retrieval

To systematically explore reconstruction reliability and spatial uncertainty in CXDI, this study introduces a novel Ensemble Phase Retrieval framework, grounded in the principle of aggregating multiple weak models generated under controlled data omission conditions. The core mechanism of this framework is the **skipped region strategy**, which serves both as a cross-validation mechanism and as a sensitivity probe for evaluating the dependence of phase retrieval algorithms on localized data regions.

Unlike conventional ensemble learning, where diversity is typically introduced by resampling the input data or re-initializing model parameters, the ensemble in this context is created by executing multiple reconstructions, each with a different spatial region selectively excluded from algorithmic computation. Importantly, the skipped region strategy does not alter the raw diffraction data or inject artificial perturbations. Instead, it bypasses critical operations, such as loss computation, gradient update (in deep learning phase retrieval), or constraint enforcement (in iterative phase retrieval), within the specified region. This ensures that the physical consistency and mathematical integrity of each algorithm remain intact, while inducing sufficient variation across reconstructions.

The skipped region is defined using a sliding-window masking approach, where a fixed-size rectangular region systematically moves across the spatial domain. For each position of the window, a new reconstruction is produced, with the algorithm restricted from using the masked region during optimization. This process is repeated across the entire sample domain, resulting in a set of reconstructions that collectively form the ensemble. Depending on the type of phase retrieval algorithm, the implementation of skipping is adapted:

- In **iterative phase retrieval**, constraint updates (e.g., Fourier modulus projection or support enforcement) are omitted within the masked region.
- In **deep learning phase retrieval**, gradient backpropagation and loss computation are conditionally disabled over the skipped region, preserving the network's architectural flow but reducing its learning signal in that area.

The motivation behind this strategy is twofold. First, it enables **cross-validated reconstruction**, allowing the model to infer missing regions based on surrounding contextual information, mirroring the inductive nature of robust inverse process. Second, it provides a principled way to **quantify ambiguity** by analyzing variability in the reconstructions generated under different skipped regions.

Once the ensemble is constructed, the individual reconstructions are aggregated through a mechanism that resembles a **soft voting approach**, in which pixel-wise averaging is applied across the ensemble to produce a consensus image. However, unlike classical soft voting in ensemble learning, the reconstructions in CXDI share a common physical origin and differ only through controlled perturbations, such as skipped region masking. As a result, the averaging process not only combines diverse

estimates but also reflects the underlying spatial ambiguity in the reconstruction. This condition allows the ensemble to serve a dual purpose. The first is generating a more robust consensus image by following this formula:

$$\bar{O}(x,y) = \frac{1}{M}\sum_{i=1}^{M} O_i(x,y), \tag{4.12}$$

with $O_i(x,y)$ representing the i-th reconstruction and M the total number of ensemble members. In parallel, a pixel-wise variance is computed to construct an **ambiguity map**:

$$\sigma(x,y) = \sqrt{\frac{1}{M}\sum_{i=1}^{M}\left(O_i(x,y) - \bar{O}(x,y)\right)^2}. \tag{4.13}$$

The ambiguity map reveals which regions of the reconstruction are consistently inferred (low variance) and which regions are sensitive to missing data (high variance), offering spatial insight into the reliability of the reconstruction process. Figure 4.7 illustrates the full ensemble phase retrieval pipeline, including the skipped region masking, reconstruction generation, and aggregation steps.

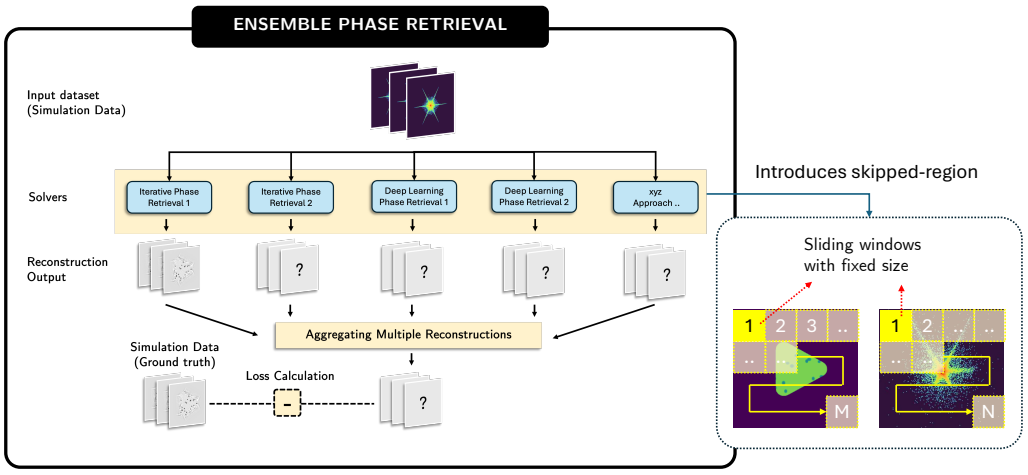


FIGURE 4.7: This illustration presents a comprehensive overview of the Ensemble Phase Retrieval framework, incorporating skipped region methodology. The system integrates multiple reconstruction models, encompassing both iterative and deep learning approaches, to generate bootstrapped reconstruction outputs. The framework implements soft voting mechanisms for result aggregation, enabling detailed uncertainty quantification and sensitivity analysis capabilities. The methodology selectively bypasses specific operational components, including loss computation and gradient update processes, within each algorithm's processing pipeline, with bypassing patterns determined by individual phase retrieval method specifications.

In essence, this ensemble framework transforms phase retrieval from a single-output inverse estimation into a **distributional inference process**. It enables deeper analysis of algorithmic behavior under partial data conditions, enhances interpretability through spatial ambiguity visualization, and improves the overall robustness of the reconstructed output. This methodology is applicable across various types of phase retrieval algorithms and is especially relevant for experimental conditions characterized by noise, missing measurements, or physical constraints.

4.4 Experiment Setting

This study employs two strategies for the skipped region mechanism in the Ensemble Phase Retrieval. Each strategy targets a different domain of the reconstruction process, real space and reciprocal space. This enables complementary insights into region-specific sensitivity and the reliability of phase retrieval outcomes. Strategy 1 omits amplitude correction or loss computation in specified regions of reciprocal space, while strategy 2 is implemented by suppressing object updates in selected regions of the real-space domain during iterative phase retrieval. Both strategies are systematically applied using a sliding window approach to generate a diverse ensemble of reconstructions. In this section, we detail the simulation parameters, model configurations, and evaluation metrics used to conduct and assess the ensemble-based uncertainty analysis.

4.4.1 Strategy 1: Skipped Region in Reciprocal Space

This section introduces Skipped Region Strategy 1, which applies selective omission in the reciprocal space domain. This strategy targets the Fourier constraint step in phase retrieval algorithms, where it omits amplitude correction within designated regions of the measured diffraction pattern. Unlike the other strategy, which modifies the object update in real space, Strategy 1 prevents the enforcement of modulus constraints over a localized area in reciprocal space, effectively simulating a missing or untrusted measurement region.

In traditional iterative algorithms such as PIE and its variants, the modulus constraint ensures that the amplitude of the predicted exit wave matches the square root of the measured diffraction intensity:

$$\widetilde{\Psi}_{\text{exit}}(q_x, q_y) = A(q_x, q_y) \cdot \frac{\Psi_{\text{exit}}(q_x, q_y)}{|\Psi_{\text{exit}}(q_x, q_y)| + \varepsilon}, \tag{4.14}$$

where $A(q_x, q_y) = \sqrt{I(q_x, q_y)}$ is the target amplitude and ε is a small constant to prevent division by zero. By implementing Skipped Region Strategy 2, this step is modified to skip amplitude correction in a specified region defined by a binary mask $M(q_x, q_y)$. The updated constraint becomes:

$$\widetilde{\Psi}_{\text{exit}}(q_x, q_y) = M(q_x, q_y) \cdot \left[A(q_x, q_y) \cdot \frac{\Psi_{\text{exit}}(q_x, q_y)}{|\Psi_{\text{exit}}(q_x, q_y)| + \varepsilon} \right] + \big(1 - M(q_x, q_y)\big) \cdot \Psi_{\text{exit}}(q_x, q_y). \tag{4.15}$$

This modification ensures that no correction is applied within the skipped region. It effectively preserves the amplitude predicted by the algorithm in those locations. The updated wavefield is then propagated back to real space for standard object updates. This technique is also compatible with multiframe PIE, where the skipped region is consistently applied across all temporal frames. Figure 4.8 illustrates this process. This strategy enables the algorithm to attempt reconstruction while explicitly withholding constraint feedback from part of the measured data. When repeated across various masking positions using a sliding window, this approach generates an ensemble of reconstructions that reflect the system's dependence on different regions of reciprocal space. Each ensemble member thus provides insight into how local diffraction information affects global image formation.

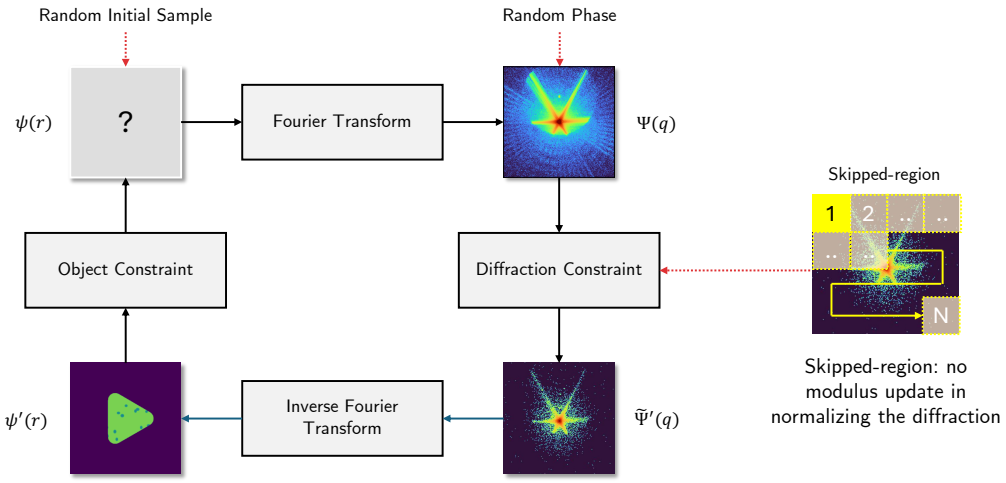


FIGURE 4.8: This figure demonstrates the implementation of skipped region Strategy 1 within iterative phase retrieval frameworks. The methodology employs a binary mask to define specific regions within the measured diffraction pattern $\Psi(q)$ where amplitude corrections are suspended.

For the deep learning-based approach, we adopt a compatible implementation within the PID3Net framework. Instead of modifying the input diffraction pattern, a spatial mask is applied during the computation of the diffraction-domain loss function. Specifically, the full diffraction pattern is partitioned into non-overlapping square tiles, and each ensemble model is trained with one tile designated as the skipped region. The diffraction loss in PID3Net is based on the negative log-likelihood of the Poisson distribution, which models the statistical noise of photon-counting detectors:

$$\mathcal{L}_{\text{diff}} = \sum_{q} M_{m,n}(q) \cdot \left[\hat{I}(q) - I(q) \log \hat{I}(q)\right], \tag{4.16}$$

where $I(q)$ is the measured intensity at reciprocal coordinate q, $\hat{I}(q) = |\mathcal{F}\{\hat{O}(r)\}|^2$ is the predicted intensity from the reconstructed object $\hat{O}(r)$, and $M_{m,n}(q)$ is the binary mask that excludes the selected tile from the loss calculation.

This masking ensures that during training, the skipped region does not contribute to gradient updates. Importantly, the full diffraction pattern is still fed into the network's temporal feature extractor and object decoder, meaning the network attempts to infer the missing region based on contextual relationships learned across the entire diffraction domain. An illustration of this mechanism is shown in Figure 4.9. The final loss function used in training combines the masked diffraction loss with a regularization term based on 3D Total Variation (TV), which promotes spatial and temporal smoothness:

$$\mathcal{L} = \mathcal{L}_{\text{diff}} + \beta \cdot \mathcal{L}_{\text{TV}}(\hat{O}). \tag{4.17}$$

We intentionally introduce a skipped region in each train model to behave as a different model. The masked pixels do not contribute to the loss, so each PID3Net instance "trains on what it can see" and must infer the missing Fourier amplitudes through learned priors and the image-domain losses. Different skip locations yield slightly different optimization landscapes, producing an ensemble of diverse but physically consistent reconstructions. Systematically applying the skipped region masks removes a distinct set of diffraction pixels for every run, so the network is

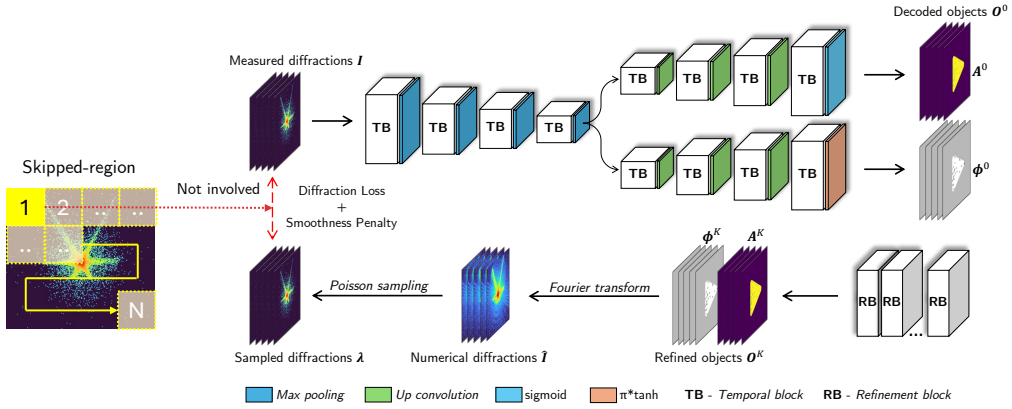


FIGURE 4.9: This figure performs the implementation of skipped region methodology within a deep learning-based phase retrieval framework utilizing PID3Net architecture. The approach designates specific regions within the measured diffraction pattern (highlighted by yellow boxes) for systematic exclusion from loss computation processes. While the complete diffraction pattern dataset serves as input for TB processing, only non-skipped regions participate in diffraction loss calculations and smoothness penalty assessments.

trained (or fine-tuned) on a slightly different input-target relationship and optimizes a correspondingly different loss calculation. Because each mask hides unique frequency information, the gradient path and final weights diverge, yielding multiple, independently biased versions of PID3Net.

4.4.2 Strategy 2: Skipped Region in Real Space

To explore spatial sensitivity and assess region-specific influence in iterative phase retrieval, we introduce a localized masking approach referred to as **Skipped Region Strategy 2**. This strategy operates in the object (real-space) domain and is exclusively applied to iterative phase retrieval algorithms, particularly the Mixed-State PIE and Multiframe PIE methods. By systematically preventing updates in selected object regions during reconstruction, this method enables targeted analysis of how information omission impacts global phase recovery.

In this strategy, a binary spatial mask $M(x,y)$ is defined over the object domain. During each iteration of the phase retrieval process, updates to the object $O(x,y)$ are suppressed within a predefined rectangular region, while the remainder of the object continues to evolve according to the standard algorithm. The update rule is modified such that the masked (skipped) region remains unchanged between iterations, as illustrated in Figure 4.10. The binary mask $M(x,y)$ is defined as:

$$M(x,y) = \begin{cases} 1, & \text{if } (x,y) \in \text{non-skipped region}, \\ 0, & \text{if } (x,y) \in \text{skipped region}. \end{cases} \tag{4.18}$$

The object update at iteration $n+1$ is computed as:

$$O^{(n+1)}(x,y) = M(x,y) \cdot O_{\text{full}}^{(n+1)}(x,y) + (1 - M(x,y)) \cdot O^{(n)}(x,y), \tag{4.19}$$

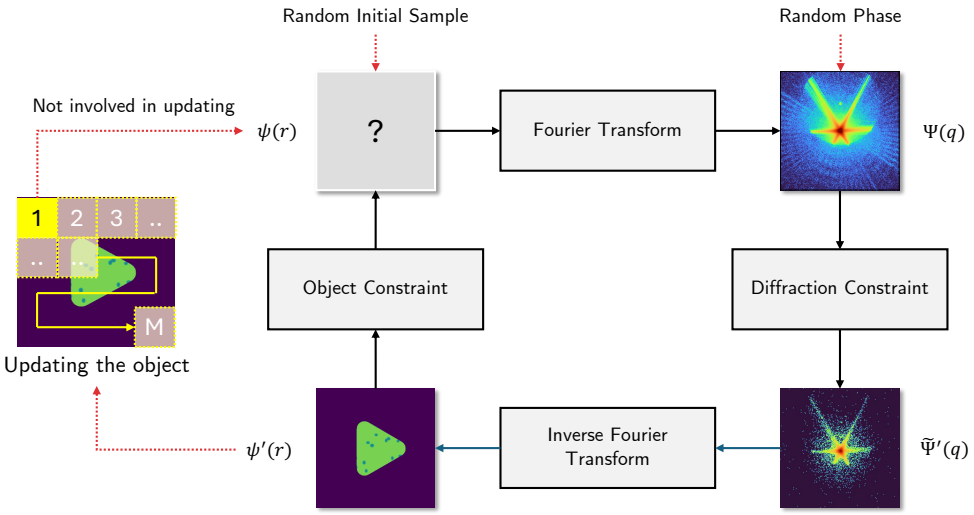


FIGURE 4.10: This illustration demonstrates the implementation of Skipped Region Strategy 2 within an iterative phase retrieval framework. The methodology employs selective regional exclusion during each iteration cycle, where a specifically designated area within the object domain (visually highlighted in yellow within the upper-left quadrant) is systematically omitted from update procedures through the application of a binary mask function $M(x, y)$. This designated region maintains its previous iteration values, effectively bypassing the phase retrieval feedback loop while preserving its original characteristics.

where $O_{\text{full}}^{(n+1)}$ is the object result from the standard update step without masking. In the case of **Multiframe PIE**, this operation is extended across all temporal frames:

$$O_t^{(n+1)}(x,y) = M(x,y) \cdot O_{t,\text{full}}^{(n+1)}(x,y) + (1 - M(x,y)) \cdot O_t^{(n)}(x,y), \quad \forall t \in \{1,\ldots,T\}. \tag{4.20}$$

This masking mechanism is applied at every iteration and at every frame, allowing the algorithm to reconstruct the object based only on partial spatial information. The skipped region effectively becomes a blind spot for feedback, meaning no phase error correction or constraint enforcement is performed in that region. Over multiple runs, the mask is shifted across the object domain in a sliding-window fashion, generating an ensemble of reconstructions which each corresponding to a different region being skipped. This allows for localized sensitivity analysis and ensemble-based uncertainty estimation.

This approach is specifically designed for iterative phase retrieval because it directly manipulates the algorithmic update step in real space, which is a core and accessible component of iterative phase retrieval, like PIE. In these algorithms, the object update is explicitly formulated and applied in each iteration, enabling selective suppression in the update loop with minimal disruption to the algorithm's physical or mathematical assumptions. In contrast, deep learning-based phase retrieval methods, such as PID3Net, lack this modular. The object function is inferred end-to-end via backpropagation through a fixed architecture, and manipulating only part of the output or suppressing gradients in a spatially selective manner during inference risks breaking the architectural flow or introducing unintended learning artifacts. This strategy allows us to highlight the spatial robustness of reconstruction by observing how omissions in localized regions influence the final object estimate. Areas of the

object that lead to significant degradation in reconstruction fidelity when skipped are identified as informationally critical, whereas regions with minimal impact suggest redundancy or stability in the reconstruction process. Aggregating these results across the object domain provides both a consensus reconstruction and a pixel-wise uncertainty map that informs users about which regions of the reconstruction can be trusted with higher confidence.

4.4.3 Simulation Settings for CXDI Simulation

To closely replicate real experimental conditions, we constructed a comprehensive simulation environment for single-frame CXDI. The system integrates both optical modeling and physical parameters that reflect realistic nanomaterial imaging scenarios. This simulation was designed to serve as a synthetic testbed for evaluating the robustness and reliability of the proposed ensemble phase retrieval strategies. The simulation parameters are grouped into five categories: user-defined computational conditions, optical properties of the X-ray system, aperture specifications, FZP configurations, and gold nanoparticle (AuNP) characteristics. Table 4.1 provides a full overview of all simulation parameters used in this study. This setup extends our prior simulation work (Mukharil Bachtiar et al., 2024) by incorporating an FZP to improve focusing realism and spatial resolution.

The simulation employs a coherent X-ray source with a photon energy of $5.0keV$. A triangular aperture with rounded corners, defined by an effective side length of 0.08 mum, is placed downstream of the focusing optic to shape the incident beam profile. To avoid sharp-edge artifacts, a Gaussian filter with a kernel size of 3×3 pixels is applied to the probe wavefield. The simulated sample consists of $400nm$ gold nanoparticles distributed such that their total volume occupies approximately 10% of the simulation region.

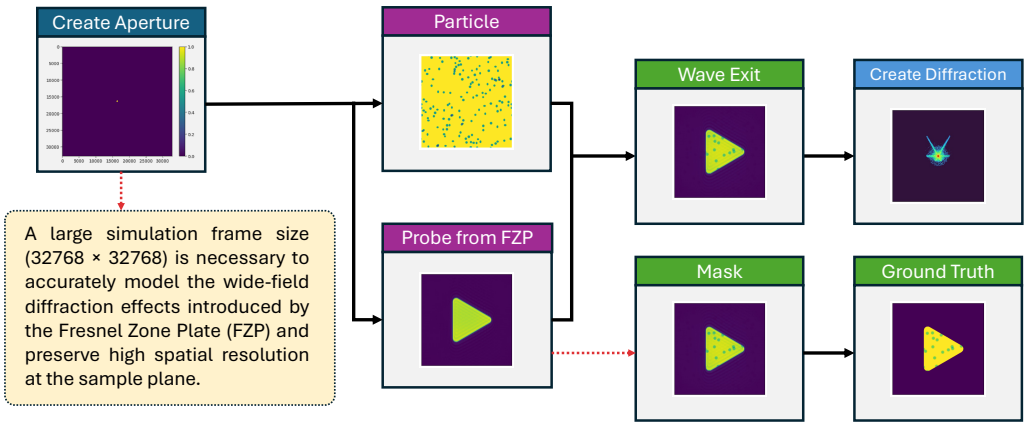


FIGURE 4.11: The illustration of the CXDI simulation in this study.

We adopted a wave-optical propagation model incorporating an off-axis FZP configuration. This design enables precise modeling of complex beam-shaping effects. A diffraction image sequence was generated using an exposure time of $100ms$ per frame, with photon-counting noise modeled by a Poisson distribution. The dynamic behavior of the sample was simulated by setting the nanoparticle velocity to $200nm/s$, a value derived from X-ray Photon Correlation Spectroscopy (XPCS) in our previous experimental work. This consistency between simulation and experiment ensures that the observed image variations accurately reflect both optical physics and particle dynamics.

TABLE 4.1: The numerical setting for CXDI Simulation using Fresnel Zone Plate.

Parameter Category	**Setting Description**
User-Defined Condition	
Window size of probe, object, diffraction pattern	≥ 520 pixels (recommended)
Window size of wavefield	$\geq 2^{15}$ pixels (recommended)
Optical Condition	
X-ray Energy	$5.0keV$
Sample-to-detector distance	$3.14m$
Detector pixel size	75μm
Pixel resolution at sample plane	$0.2nm/pixel$
Dynamic range of detector	$7bits$
Aperture Condition	
Aperture geometry	Triangular with rounded corners
Effective side length	$10^{-5}m$ (10μm)
Curvature radius of corners	0.08μm
Blur applied	Gaussian kernel 3×3 pixels
Fresnel Zone Plate (FZP) Condition	
Number of zones	1875
Thickness	250×10^{-9} m (250 nm)
Particle (AuNP) Condition	
Particle diameter	$400nm$
Estimated velocity	$200nm/s$ (from XPCS data)
Volume fraction in sample region	$\approx 10\%$
Exposure time per frame	$100ms$
Max intensity per pixel	$\sim 104.1photons/pixel$
Noise model	Poisson-distributed photon-counting

All values are adjustable based on specific experimental goals. Default settings correspond to actual physical parameters used in synchrotron-based dynamic CXDI experiments. The flow process of simulation can be examined in Figure 4.11.

To further evaluate the robustness and reliability of the proposed ensemble phase retrieval framework, we conducted a series of controlled simulations that incorporate realistic shot noise into the diffraction data. This simulation mimics the photon-count fluctuations naturally present in single-frame CXDI experiments, especially in low-dose or high-speed imaging regimes. In actual experiments, a typical diffraction frame contains approximately 3×10^6 photons distributed across a 512×512 detector, resulting in an average of about 10 photons per pixel. To replicate this condition, we first scaled our clean simulated diffraction patterns so that the average intensity per pixel equals approximately 10. Then, for each pixel, we drew a random sample from a Poisson distribution with this mean and replaced the clean intensity value with the sampled count. This process yields a noise-contaminated diffraction pattern that captures the statistical properties of real photon detection events. An illustration of this simulation is shown in Figure 4.12.

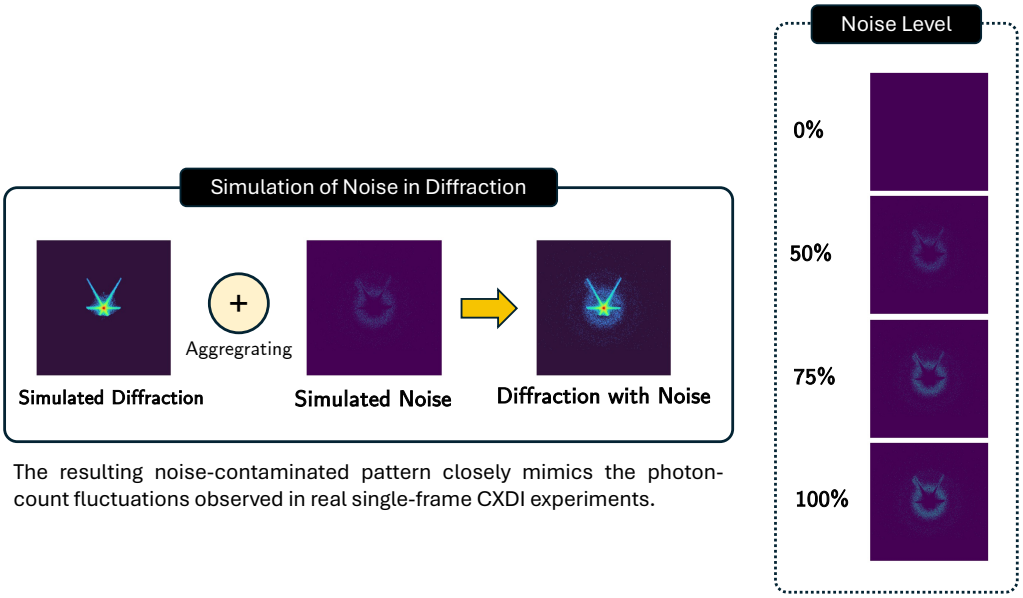


FIGURE 4.12: The illustration of noise simulation settings.

By systematically adding noise at varying intensities, we assess whether the ensemble-based ambiguity maps remain stable, meaningful, and predictive under increasingly degraded input conditions. A key hypothesis is that the ensemble framework should not only reflect intrinsic ambiguity due to missing phase information, but also adapt to extrinsic variability introduced by measurement noise. In this context, we examine whether high-uncertainty regions in the ensemble reconstructions align with those most affected by noise, thereby validating the ensemble variance map as a reliable proxy for confidence under realistic data scenarios. Taken together, this noise simulation protocol plays a dual role. It ensures physical realism in the evaluation pipeline and offers a framework for benchmarking ensemble sensitivity to noise. This is especially important for real-world deployment, where photon noise is inherent and interpretive robustness is essential for reliable downstream analysis in CXDI imaging workflows.

4.5 Examining the Results of Skipped Region Mechanism

We conducted a series of controlled ensemble phase retrieval experiments, incorporating both strategies, to systematically investigate the spatial sensitivity and reconstruction reliability of single-frame CXDI. The primary goal of these experiments was to quantify how the exclusion of localized information, either in the object domain or in the measured diffraction pattern, affects the consistency and fidelity of the phase retrieval outcome. This investigation is critical in understanding the non-uniqueness of CXDI, particularly in scenarios involving missing or incomplete data.

For the initial examination, we applied a sliding-window skipped region across the spatial domain to explore the impact of local data omission. Square skip masks of sizes 32, 64, and 128 pixels were tested, with sliding step sizes of 32 and 64 pixels. The smaller skip sizes offer higher resolution for sensitivity mapping, while the 128-pixel window is particularly important for probing the central region of the diffraction pattern, which encodes low-frequency information critical for recovering global object structure. For ensemble analysis, we employed three phase retrieval methods that were already mentioned previously. Each method was run across 1000 simulated frames with known ground-truth reconstructions. For every skipped region configuration, the reconstructed images were evaluated using the SSIM. To ensure consistency across evaluations, a preprocessing pipeline was applied to both predicted and ground-truth phase images. The reconstructions were cropped to a fixed 270×270 pixel window centered on the illuminated region, eliminating boundary effects and unsupported regions.

A subset of the experiments focused on Strategy 1 (reciprocal-space skipping) with a fixed skip size of 128 pixels and step size of 64 pixels. Figure 4.13 visualizes the degradation in SSIM across all skip positions. Each matrix element represents averaged SSIM values obtained when excluding a 128×128 pixel region of the diffraction pattern centered at the corresponding position. The central diffraction pattern region is specifically highlighted for emphasis. The figure's lower section provides qualitative reconstruction comparisons between cases. Notably, all three phase retrieval methods showed a pronounced SSIM drop when the skipped region was centered near the origin of the diffraction pattern, corresponding to low spatial frequencies. This confirms that the central region of the Fourier domain encodes global object features and contrast, and is thus essential for maintaining reconstruction fidelity. Conversely, skipping peripheral regions had minimal effect on SSIM, suggesting that high-frequency components are less critical to the coarse structural representation of the object.

From Figure 4.13, it is evident that the spatial region located at coordinates $[128:256, 128:256]$ plays a critical role in the overall success of the phase retrieval process. The analysis highlights several key insights into the spatial sensitivity of different reconstruction methods when exposed to localized data omission in the diffraction domain. Specifically, all three phase retrieval approaches exhibited consistent patterns: the lowest SSIM values were observed when the skipped region was positioned near the center of the diffraction image. This central region corresponds to low spatial frequency components, which are responsible for encoding the global shape, contrast, and envelope of the real-space object. Omitting these components severely disrupts the algorithm's ability to reconstruct coherent structures, resulting in substantial loss of fidelity and, in some cases, complete reconstruction failure. This trend reflects the well-known principle in Fourier imaging that low-frequency

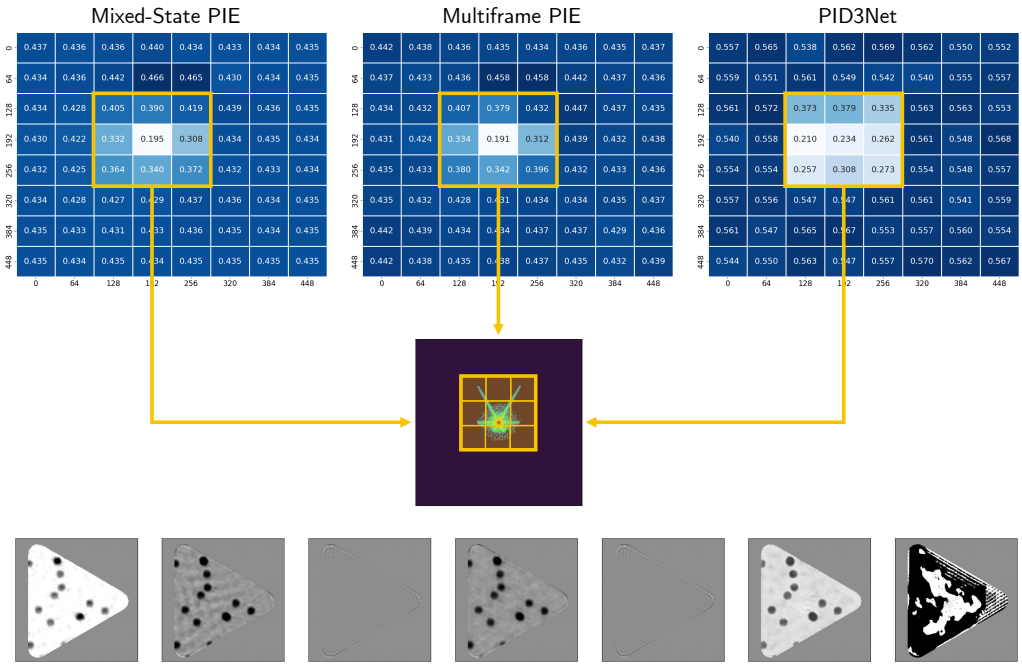


FIGURE 4.13: Comparative visualization of SSIM degradation patterns through heatmap representations, analyzing three distinct phase retrieval methodologies under Strategy 1.

information governs the macro-scale structural coherence of the object.

Conversely, when the skipped region was located toward the periphery of the diffraction pattern, corresponding to high spatial frequencies, the SSIM values remained relatively stable, and the visual quality of the reconstructed object was largely preserved. While high-frequency components contribute to edge sharpness and fine-grained texture, they appear to be less critical for retaining the overall form and recognizable features of the object. The reconstructions displayed below the heatmap in Figure 4.13 provide qualitative confirmation of this trend. For example, skipping the region around $[0, 0]$, which lies at the edge of the Fourier domain, results in negligible distortion. In contrast, omitting the region around $[192, 192]$, which is located near the diffraction center, leads to severe structural degradation or complete collapse of the reconstruction.

These results emphasize the centrality of low-frequency regions in the diffraction pattern and their disproportionate influence on reconstruction reliability. They further underscore the value of skipped region ensemble strategies in not only quantifying uncertainty but also localizing critical zones in Fourier space that contribute most significantly to successful phase recovery.

TABLE 4.2: Comparison between skipped region strategies with small and large step sizes.

Aspect	Small Step Size	Large Step Size
Definition	The skipped window is shifted by small intervals between positions.	The skipped window is shifted by larger intervals between positions.
Sampling Density	High-density scanning with fine granularity.	Sparse sampling with coarse granularity.
Sensitivity Map Resolution	High: can capture subtle variations and fine local sensitivity.	Low: only reveals coarse global sensitivity.
Number of Reconstructions	Large (many skip configurations).	Small (fewer skip configurations).
Computational Cost	High (increased time and resources).	Low (more efficient).
Strengths	Detects small critical regions and local artifacts.	Quickly identifies major sensitive zones.
Limitations	Time-consuming and resource-intensive.	May miss localized sensitivity patterns.

To examine how scan resolution affects the quality of ensemble-based uncertainty quantification, we conducted a comparative analysis of reconstruction results using two different step sizes—32 pixels and 64 pixels, while keeping the size of the skipped region constant. These step sizes determine how frequently the masked window is shifted across the diffraction image to generate ensemble members. The results, as summarized in Table 4.2, provide compelling evidence that finer scan resolution (i.e., smaller step size) contributes to more stable and informative ambiguity maps. Specifically, reconstructions generated with a step size of 32 pixels consistently demonstrated lower uncertainty and higher structural consistency across ensemble members for both Mixed-State PIE and Multiframe PIE algorithms. This outcome can be attributed to the higher sampling density afforded by the smaller step size, which allows the ensemble to more thoroughly explore localized sensitivities and subtle variations in the diffraction space. The increased coverage of overlapping skipped regions ensures that critical zones, particularly around low-frequency components, are adequately probed in multiple configurations, thereby enhancing the granularity of the uncertainty map.

In contrast, reconstructions performed using a larger step size of 64 pixels resulted in sparser sampling across the diffraction domain. This coarser resolution reduces the number of ensemble members and leads to a more limited view of the solution variability, ultimately producing ambiguity maps with lower spatial resolution and diminished interpretive power. Although this approach is computationally more efficient, it trades off sensitivity and detail in the uncertainty quantification process.

These findings suggest that while larger step sizes offer computational speed-ups, smaller step sizes are preferable when the goal is to achieve fine-grained uncertainty localization and higher confidence in structural interpretations. Thus, scan resolution emerges as a critical design parameter in ensemble-based phase retrieval workflows, influencing both the accuracy of the consensus reconstruction and the quality of ambiguity estimation.

Other than that comparison, we also compared the results between the two strategies. The results can be examined in Table 4.3.

TABLE 4.3: Comparison between two strategies of skipped region.

Aspect	**Strategy 1**	**Strategy 2**
Reconstruction Impact	Significant distortion, especially when an important part of the diffraction was skipped.	Final image almost unchanged due to the core intention of the algorithm, which is more concerned with reciprocal space.
Suitability for Ensemble	Reconstructions have a large variance, so it is highly informative for ambiguity interpretation.	Reconstruction is homogenous (small variance) but less useful to Ensemble, especially in interpreting the variance in the phase.
Sensitivity Map Resolution	High: can capture subtle variations and fine local sensitivity.	Low: only reveals coarse global sensitivity.
Compatibility	Can be implemented in both phase retrieval approaches.	Only works in the iterative phase retrieval.
Main Summary	By selecting a proper skip size, the nature of the skip process in reciprocal gives more highlight to the phenomenon that happened in the real experiment.	Does not expose weakness in frequency (diffraction) information, so it is not very natural.

Building on the comparative analyses of step sizes and skip-region strategies, this study adopts a refined configuration by selecting smaller skipped region sizes, specifically, 8×8 and 16×16 pixels. This design decision is grounded in the observed benefits of fine-grained sampling, as demonstrated by the superior performance of smaller step sizes in generating high-resolution sensitivity maps and more stable consensus reconstructions. By using smaller skip windows, the ensemble can highlight local variations in the diffraction pattern with greater precision, enabling the

detection of subtle sensitivity patterns that are otherwise obscured in coarser configurations. Moreover, smaller regions increase the overlap across different skip locations, enriching the statistical diversity of the ensemble and improving the granularity of uncertainty quantification, especially around low-frequency zones that are known to dominate phase retrieval fidelity.

In addition to refining the skip size for further examination, this study exclusively adopts **Strategy 1**, which applies the skip region directly in the reciprocal space (i.e., the diffraction pattern), rather than in the object domain. This choice is motivated by several empirical and conceptual considerations. As shown in Table 4.3, skipping in reciprocal space results in more significant variation among reconstructions. Thereby, producing informative variance maps that better reflect the system's sensitivity to missing frequency components. In contrast, object-domain skipping (Strategy 2) yields reconstructions with minimal variance, limiting its utility for ambiguity interpretation and ensemble integration. Furthermore, Strategy 1 maintains physical interpretability by directly probing the influence of specific diffraction components. The strategy is aligning more closely with the nature of experimental CXDI, where data incompleteness primarily manifests in the reciprocal domain. Thus, for both theoretical coherence and practical effectiveness, the reciprocal-space skipping strategy with small window sizes is adopted as the primary design in this ensemble-based phase retrieval framework.

4.6 Aggregating Phase Retrieval Results

Once all reconstructed objects have been generated from various skipped region configurations, the next critical step involves aggregating these results into a unified representation. This aggregation process serves a dual purpose. It not only produces a consensus reconstruction via a voting mechanism but also quantifies the variance to be an ambiguity map associated with each pixel in the reconstructed image. Through this ensemble mechanism, we gain both a stable solution and a spatially resolved confidence map, offering valuable insights into which regions of the reconstruction are consistently determined versus those that are prone to ambiguity or instability. The illustration of aggregating the reconstruction results of this study can be examined in Figure 4.14.

The ensemble-based phase retrieval framework allows for a statistical interpretation of reconstruction reliability. Given a set of N reconstructed phase maps $\{\theta_i(r)\}_{i=1}^{N}$, where r denotes the pixel location, the consensus reconstruction is computed as the pixel-wise mean:

$$\mu(r) = \frac{1}{N}\sum_{i=1}^{N}\theta_i(r), \tag{4.21}$$

which acts as the final prediction derived from soft voting over the ensemble. The corresponding pixel-wise uncertainty is captured by the standard deviation:

$$\sigma(r) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\theta_i(r) - \mu(r)\right)^2}, \tag{4.22}$$

which reflects the dispersion of predictions at each location and serves as a direct measure of epistemic uncertainty in the reconstruction. To ensure that the aggregated reconstruction reflects both diversity and quality, we implement a selection

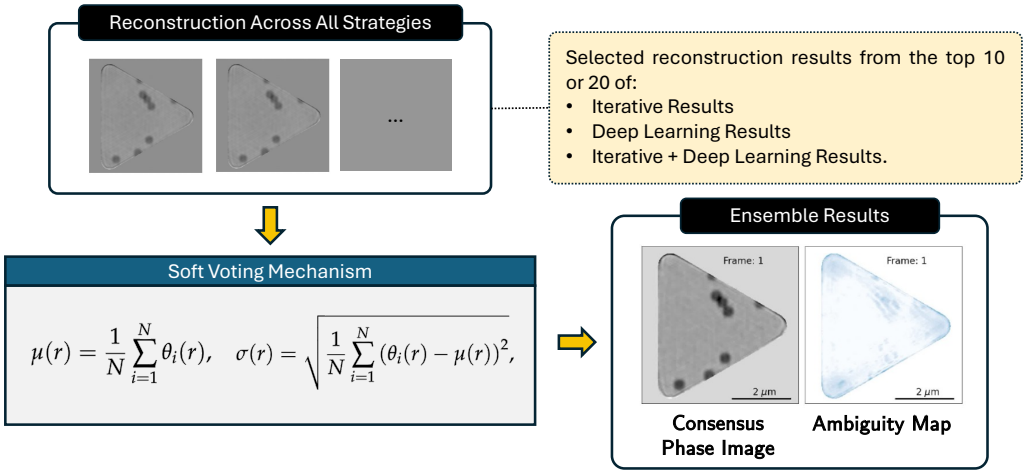


FIGURE 4.14: The mechanism of aggregating the reconstruction result from each solver and the correspondence results of it in the form of mean map and variance map.

mechanism prior to the ensemble aggregation step based on the R_F score. Instead of aggregating all ensemble members indiscriminately, we rank the individual reconstructions according to their R_f scores and retain only the top-N best-performing models for final integration. This strategy, henceforth referred to as the *N-best model selection*, allows us to balance two competing goals, which are preserving epistemic ambiguity across reconstructions while filtering out clearly poor solutions that may degrade the consensus result.

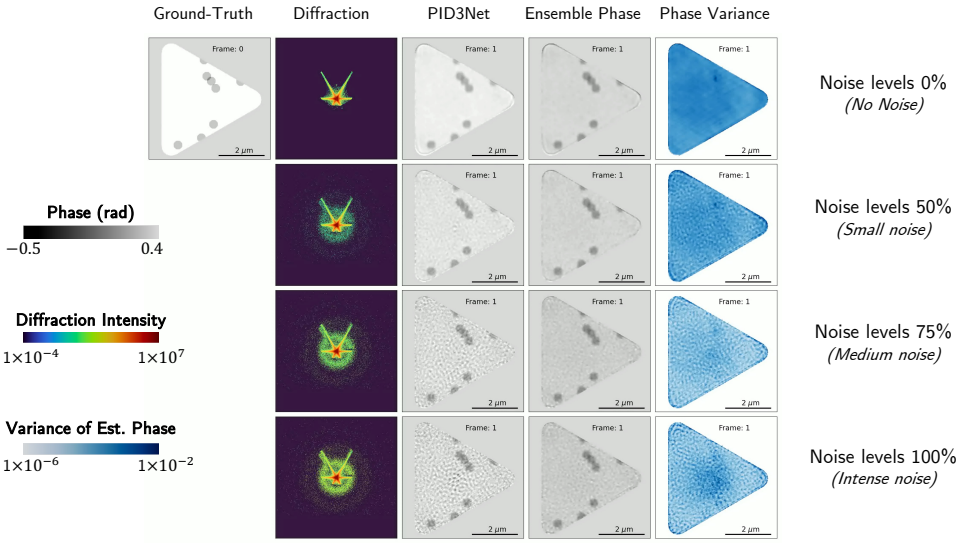


FIGURE 4.15: Ensemble phase retrieval result using the top 10 reconstructions from the combination of iterative and deep learning solvers. The phase variance is highly localized at the corners of the triangular aperture.

The rational reason behind using the R_f score is that it directly quantifies how well a given reconstruction respects the physical diffraction data, independent of real-space interpretability. Since ensemble phase retrieval approaches are inherently

sensitive to the quality of individual members, including highly inconsistent reconstructions may introduce artificial variance or blur structurally important features in the aggregated image. To probe how R_f–guided model selection influences ensemble quality, we first stratified the full pool of phase-retrieval solutions into three source categories:

1. **iterative only** (Mixed-State PIE and Multiframe PIE),
2. **deep-learning only** (PID3Net),
3. a **combined iterative + DL** set,

within each category. We retained either the top-20 or the top-10 (also known as the N-best selection) reconstructions ranked by their diffraction-fidelity score R_f, which directly gauges how closely a candidate respects the measured intensities independent of real-space appearance. This filtering step serves two purposes: it removes physically implausible outliers that would otherwise inflate ensemble variance, and it limits computational burden by discarding solutions unlikely to improve the aggregate. Empirically, the top-20 ensembles preserve greater solver diversity but also admit members with marginal R_f scores. In contrast, the top-10 ensembles concentrate on the highest-quality reconstructions, sacrificing some diversity for sharper physical consistency.

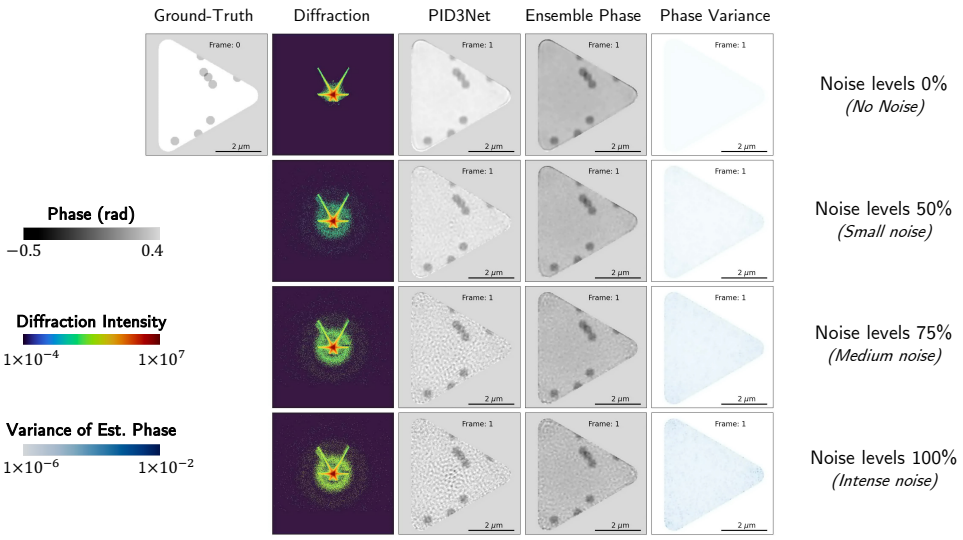


FIGURE 4.16: Ensemble phase retrieval result using the top 10 reconstructions using only iterative solvers.

The comparative results are revealing. Across all three categories, aggregating the **top-20** members produces consensus images accompanied by **broader variance maps**, where high uncertainty spreads into both the central core and the triangular-aperture corners—an indication that lower-quality reconstructions inject inconsistent phase information and blur fine structure. When the selection is tightened to the **top-10**, uncertainty contracts dramatically. These observations confirm that retaining roughly 10–20 high-R_f members strikes an effective compromise between ensemble diversity and reliability while reducing post-processing time. Thus, R_f-based N-best selection emerges as a simple yet powerful mechanism for curating ensemble inputs, delivering computationally efficient and epistemically transparent reconstructions. Preliminary tests show that an ensemble trimmed to the $N_{best} = 10$ reconstructions consistently

yields a markedly lower phase variance than its top-20 counterpart, indicating tighter agreement among the highest–R_f members and reduced artefactual blurring. In view of this finding, all subsequent analyses concentrate on the **top-10** setting while retaining the three aggregation schemes introduced earlier. Focusing on this variance-minimising configuration ensures that comparisons across solver categories reflect intrinsic algorithmic behaviour rather than noise injected by lower-quality reconstructions, thereby providing a clearer assessment of ensemble performance and uncertainty localization. The first experiment focused on the $N_{best} = 10$ combining iterative and deep learning solvers. Figure 4.15 illustrates the performance of the *top-10* ensemble built from the combined pool of iterative and deep-learning reconstructions across four noise levels (0 %, 50 %, 75 %, and 100 %). Even in the no-noise case, the consensus phase image closely reproduces the ground-truth triangular object, and the accompanying variance map shows uncertainty confined mainly to the aperture corners, which are the locations that correspond to the weakest low-frequency diffraction intensities.

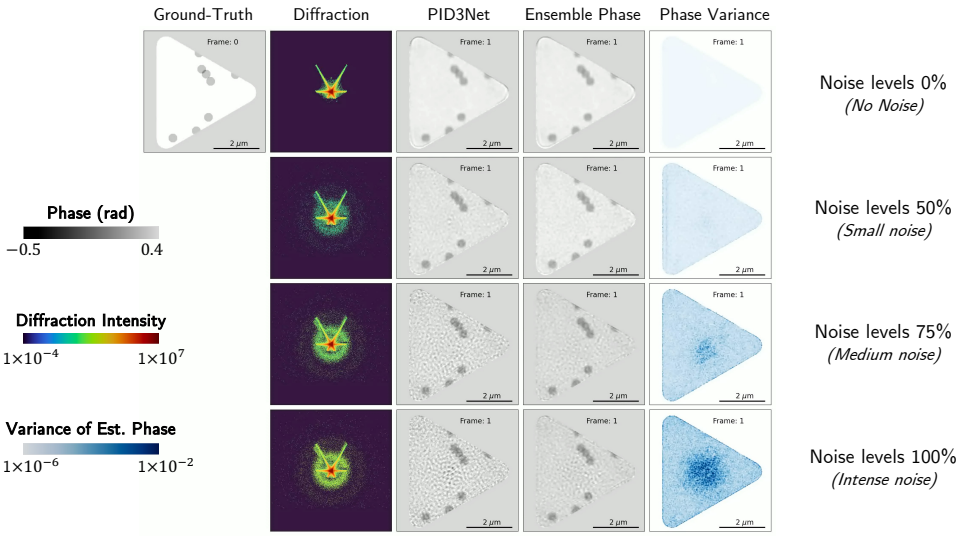


FIGURE 4.17: Ensemble phase retrieval result using the top 10 reconstructions using only deep learning solvers.

Figure 4.16 shows the *top-10* ensemble built exclusively from iterative solvers (Mixed-State and Multiframe PIE). Across all four noise levels, the consensus phase image preserves the triangular morphology and internal particle distribution with minimal artefacts. The accompanying variance maps remain tightly confined to the aperture corners and a few low-signal voids, even under intense photon noise, indicating that the iterative solvers alone supply highly self-consistent phase information. When compared with the *top–10 combined* ensemble (iterative + DL in Figure 4.15), the iterative-only consensus appears marginally sharper at high noise. Still, its variance map is slightly less sensitive to subtle central ambiguities than the combined set detects. This trade-off highlights the complementary strengths of the two approaches: iterative solvers enforce strict modulus and support constraints, yielding uniformly low variance, whereas deep-learning members inject prior-driven flexibility that can expose residual uncertainty in regions where iterative methods over-regularize. Figure 4.17 added an extra illustration about the phenomenon.

We further quantified the impact of experimental noise on ensemble performance. As shown in Figure 4.18, the left-hand plot traces the total phase variance versus

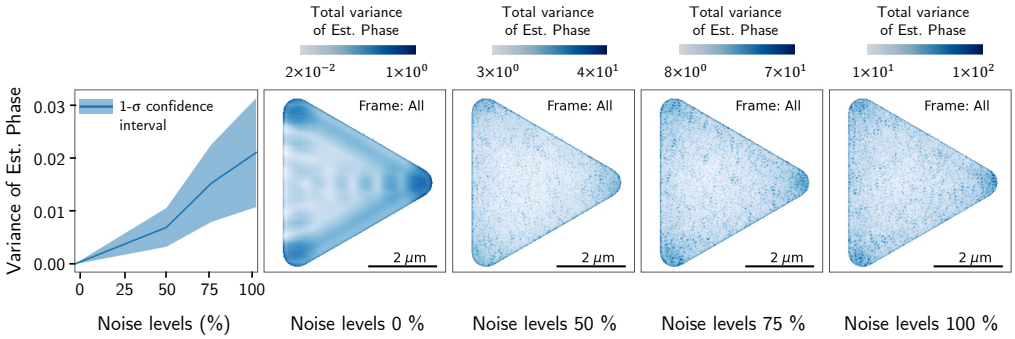


FIGURE 4.18: Correlation between noise level and phase ambiguity in ensemble phase retrieval using only the top 10 iterative reconstruction results. The noise level increases from 0% to 100%.

noise level, with the shaded band denoting the 1-σ confidence interval obtained from repeated trials. A clear monotonic trend emerges: as photon-count noise rises from 0% to 50%, 75%, and 100%, the ensemble variance increases accordingly, signalling diminishing certainty in the reconstructed phase. The spatial maps on the right clarify how this uncertainty is distributed. At 0% noise, variance is minimal and confined mainly to the triangular-aperture corners—regions already weakly constrained by diffraction intensity. At 50% noise, fine-grained ambiguity appears across the object, indicating that even moderate shot noise perturbs global phase stability. The 75% noise condition amplifies this effect: variance becomes denser yet still heterogeneous, revealing that certain interior zones begin to rival the aperture corners in instability. Under 100% noise, uncertainty is both widespread and nearly uniform, demonstrating that severe photon fluctuations degrade confidence throughout the field. Together, these results confirm that ensemble-derived variance maps are sensitive not only to the intrinsic ill-posedness of phase retrieval but also to data quality, reliably distinguishing high-confidence from low-confidence regions across a realistic noise spectrum, making the method a practical diagnostic for real-world single-frame CXDI.

The ambiguity maps produced through ensemble phase retrieval, as shown in Figure 4.19, reveal systematic patterns of uncertainty that closely correlate with the physical aperture geometry, specifically, the triangular aperture design introduced by Takazawa et al. (Takazawa, Kang, et al., 2021; Kang et al., 2021). Even under noise-free conditions, the spatial variance maps exhibit prominent ambiguity localized at the three corners of the reconstructed object. This distribution reflects intrinsic limitations of the diffraction information itself, where the corners of the triangular aperture yield relatively low-intensity signals in the reciprocal domain. Such observations validate the hypothesis that ambiguity in CXDI is not solely due to algorithmic instability or external noise but is also a structural consequence of the experimental setup. The insights align with Takazawa's experimental findings, which demonstrated that while triangular apertures enhance spatial resolution and break symmetry better than circular or square apertures, they inherently contain regions of lower modulation transfer, most notably at the aperture vertices.

Figure 4.20 summarizes how deep learning phase retrieval responds to increasing shot noise and highlights a fundamental difference from the iterative ensemble. The left-hand plot demonstrates a monotonic rise in total phase variance as noise is raised from 0 % to 100 %, while the shaded band denotes a 1-σ confidence interval across trials. The accompanying variance maps reveal that, unlike the iterative

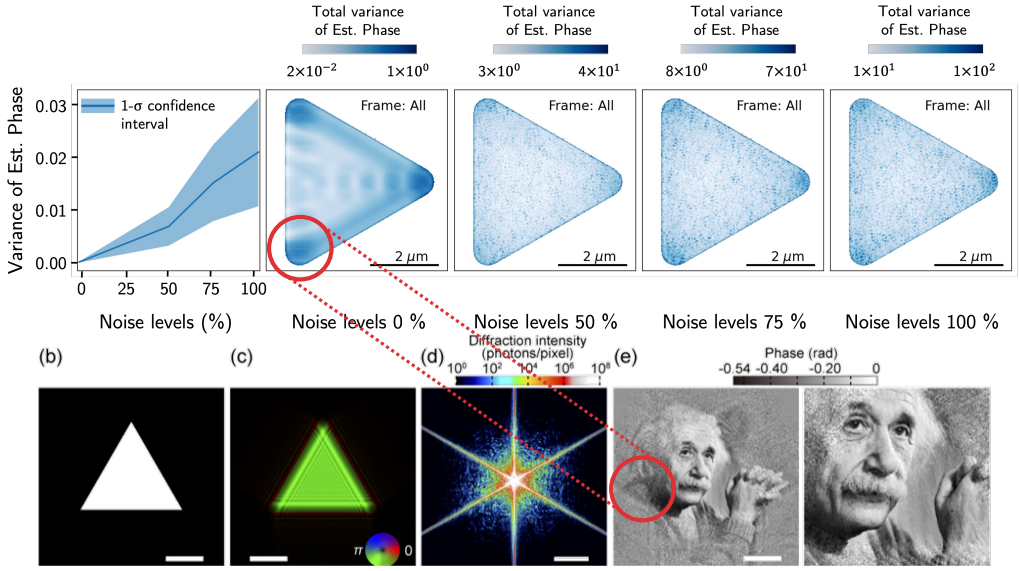


FIGURE 4.19: Ensemble phase retrieval reveals localized ambiguity at the corners of the triangular aperture, even in the absence of noise. This spatial uncertainty pattern aligns with the low-intensity regions in the diffraction domain and supports findings by Takazawa et al. regarding the structural sensitivity induced by triangular aperture geometry.

ensemble, whose uncertainty manifests mainly at the triangular-aperture corners and thus reflects genuine ambiguity about overlapping particle presence, the DL ensemble exhibits a pronounced, noise-amplified "hot spot" in the image center. This centrally concentrated variance persists even at 0 % noise and expands with 50 %, 75 %, and 100 % noise, indicating that the dominant source of ambiguity is not structural but algorithmic: a regularization bias in the network that under-constrains low-entropy central pixels. Consequently, the DL variance map encodes two layers of information: peripheral uncertainty comparable to the particle-presence ambiguity seen in iterative results, and a superimposed central uncertainty driven by model bias and training error. Recognizing this dual origin is essential for interpreting DL-based CXDI reconstructions, as mitigation strategies differ—physical constraints for particle ambiguity versus architectural or training refinements for solver-induced variance.

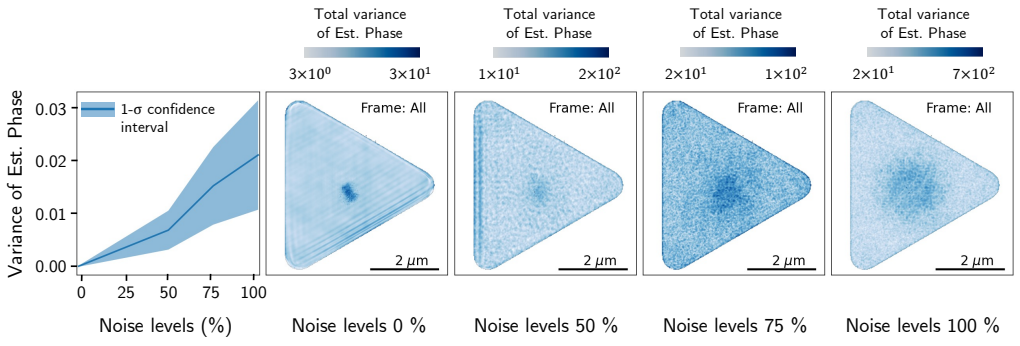


FIGURE 4.20: Correlation between noise level and phase ambiguity in ensemble phase retrieval using the top 10 reconstruction results of deep learning phase retrieval.

This connection highlights the value of ensemble-based ambiguity quantification not only as a post-reconstruction diagnostic but also as a tool for evaluating the effectiveness of optical system design. By tracing how local variances in the diffraction pattern propagate into phase instability in the reconstructed image, the ensemble framework enables a new layer of interpretability grounded in physical optics. The linkage shown in the figure, from the corner of the variance map to the affected region in the final phase image, illustrates how local weaknesses in the diffraction domain can translate into globally visible reconstruction artifacts. In doing so, the method provides an epistemically motivated lens through which both reconstruction performance and optical design decisions can be jointly evaluated.

4.7 Contributions and Limitations

This study introduces a principled and interpretable framework to address the challenges of non-uniqueness and ambiguity in single-frame CXDI. The central idea begins with a cross-validation-inspired mechanism that uses spatial data-skipping to probe the stability and sensitivity of the reconstruction process. By temporarily masking different regions of the diffraction pattern or object domain, the method forces solvers to infer missing information and tests their ability to generalize from incomplete data. This strategy helps mitigate overfitting, a common issue in phase retrieval where a single solver may lock onto a solution that is overly influenced by initialization or specific constraints. Each of these masked reconstructions is treated as a weak learner that captures a different perspective of the solution space. The resulting ensemble of reconstructions is then aggregated using a voting scheme to generate a consensus image. In parallel, the variability across ensemble members is quantified to produce a pixel-wise ambiguity map, offering localized insight into reconstruction reliability. By systematically varying skipped regions in the diffraction or object domain, the method generates an ensemble of reconstructions that capture different aspects of the solution space. This approach enables compatibility across different solvers, such as Mixed-State PIE, Multiframe PIE, and PID3Net, without requiring manual hyperparameter tuning or architecture-specific design.

Further contribution is by separating the spatial-variance landscape into two diagnostic layers, our ensemble analysis delivers a nuanced contribution to the long-standing phase-ambiguity problem in CXDI. The first layer, tight clusters of high variance surrounding overlapping particles, pinpoints regions where the diffraction data themselves are under-constrained, revealing genuine epistemic limits on particle presence and positioning. The second layer, an extended band of variance anchored at the image center, traces back to architectural priors and regularization biases in the deep-learning solver, exposing algorithm-induced uncertainty that is absent in fully iterative reconstructions. Recognizing this dual origin is pivotal. It enables researchers to attribute ambiguity either to the physics of signal formation or to the heuristics of model training. Practically, the finding translates into a two-track remediation strategy: introduce additional physical constraints or complementary measurements to resolve particle-presence ambiguity, while addressing central variance through hyperparameter tuning, alternative network architectures, or domain-aware regularizers. In doing so, the study elevates variance mapping from a qualitative visualization to a quantitative decision-making tool, guiding both experimental design and algorithmic development toward a more reliable and interpretable phase-retrieval pipeline.

Each ensemble member is generated under controlled data perturbations, implemented through structured masking either in reciprocal or real space. These

perturbations expose varying sensitivities within the reconstruction process and provide an interpretable mechanism to simulate missing information. Importantly, before aggregation, reconstructions are filtered using a quality-based selection mechanism based on the R_F score. Only the top-N solutions with the highest fidelity to the measured diffraction pattern are retained. This N-best model selection improves the stability of the ensemble output by eliminating poor or noisy reconstructions that may introduce artificial variance. The results show that selecting the top-10 reconstructions yields a more focused and accurate output than the top-20 case, with ambiguity maps that are largely confined to physically interpretable regions, such as the low-intensity corners of a triangular aperture. In contrast, larger ensembles that include lower-quality members exhibit broader variance distributions, highlighting the importance of careful aggregation in ensemble design.

Another significant contribution lies in the generation of spatial ambiguity maps (σ) that serve as localized indicators of reconstruction confidence. These maps shift evaluation from global scores like SSIM or R_f toward spatially explicit interpretability. Much like attention heatmaps in explainable AI (XAI), these uncertainty maps reveal which regions are stably reconstructed and which are susceptible to ambiguity, enabling researchers to assess the trustworthiness of individual object features. In addition, the study systematically investigates the impact of scan granularity and skipping strategy. Reciprocal-space skipping proves more informative for probing Fourier-domain sensitivity, especially in identifying the dominant role of low-frequency bands in maintaining object structure. Fine-grained skip steps yield high-resolution sensitivity maps but at increased computational cost. The trade-off between resolution and efficiency defines a tunable design space for future applications.

To validate robustness under realistic conditions, Poisson-distributed noise is injected into the diffraction data, simulating photon-count fluctuations in real experiments. The ensemble framework shows a consistent correlation between noise levels and the extent of phase ambiguity. At higher noise levels, the uncertainty maps reflect broader variance, yet still localize ambiguity in a physically meaningful way. This demonstrates the method's sensitivity to data quality and its applicability to experimental scenarios.

All the explanations lead us to a key conceptual innovation of this work, which is the shift from seeking a single "best" reconstruction to modeling a distribution of plausible solutions. By treating phase retrieval as an ensemble problem, the framework not only improves accuracy through aggregation but also quantifies uncertainty, allowing researchers to judge how reliable and therefore how trustworthy each reconstructed feature truly is.. Instead of relying on a single solver with fixed constraints, ensemble phase retrieval generates multiple reconstructions under systematically varied data conditions and aggregates them to produce both a consensus image and a pixel-wise ambiguity map. This approach enhances robustness against initialization sensitivity and noise, while providing interpretability through spatially resolved confidence estimates. By doing so, it bridges the gap between computational reconstruction and epistemic insight, enabling users to assess not only how accurate a reconstruction is, but also where it is trustworthy or potentially ambiguous. This dual focus on fidelity and interpretability positions ensemble phase retrieval as a significant methodological advancement within computational imaging, especially in applications where data incompleteness and non-uniqueness are intrinsic to the imaging process.

Despite the contributions, the framework has several limitations:

- **Simulated Data Dependence:** The framework is developed and validated

using simulated datasets. Its generalizability to real-world experimental data remains to be tested, particularly under conditions involving partial coherence, beamstop occlusion, detector nonlinearity, and saturation.

- **Computational Overhead:** The approach requires the generation and evaluation of multiple reconstructions across varied masking conditions, which incurs high computational cost, especially when smaller (fine-grained) skip sizes are used.
- **Mask Design Sensitivity:** The framework's effectiveness is influenced by the masking configuration. While systematic sliding windows offer full coverage, the method may overlook critical zones if patch design is not sufficiently informed by data or experimental priors.
- **Non-Uniqueness Mitigation, Not Resolution:** Although the method narrows the plausible solution space, it does not fully resolve the non-uniqueness inherent in phase retrieval. Additional constraints or external priors are needed to isolate a unique solution.
- **Single-Frame Specificity:** The current design is tailored to single-frame CXDI. Extending the framework to time-resolved imaging will require additional mechanisms to account for temporal coherence and dynamic consistency.

Nonetheless, the proposed framework represents a significant advancement by incorporating ensemble reasoning and spatial data-skipping strategies into the phase retrieval pipeline. By introducing an explicit interpretability layer, the method redefines reconstruction analysis from a solely output-focused task to one centered on stability, confidence, and data-informed trustworthiness. These contributions have meaningful implications for both algorithm development and experimental design. For example, informing decisions about which spatial frequencies to prioritize during data acquisition or where to enforce additional constraints for improved reconstruction quality. Ultimately, this work provides a practical and extensible foundation for ambiguity-aware inference in nanoscale imaging, bridging the gap between numerical reconstruction fidelity and epistemic interpretability.

Chapter 5

Conclusion and Limitation

5.1 Conclusion

This dissertation presented two integrated studies that collectively advance the state of phase retrieval and interpretability in single-frame Coherent X-ray Diffraction Imaging (CXDI). Both studies are grounded in a shared motivation: to improve the reliability and explainability of nanoscale reconstructions under the constraints of limited and phase-deficient diffraction data. While Chapter 3 addressed reconstruction robustness via physical system design, Chapter 4 introduced a new strategy for uncertainty-aware inference by leveraging ensemble learning and spatial data perturbation. Together, these contributions present a unified pipeline that spans from optical design optimization to confidence quantification in reconstructed structures.

The first part of this work focused on the impact of aperture geometry on reconstruction quality in single-frame CXDI. Through a systematic simulation study, we investigated how the corner radius of a triangular aperture influences the stability and accuracy of iterative and deep learning-based phase retrieval methods. The results revealed that a small but non-zero corner radius, closer to realistically fabricable geometries, maintains high reconstruction fidelity, particularly when interpreted through metrics in both real and reciprocal space. This study offered a practical pathway for relaxing fabrication constraints without sacrificing imaging performance, and it positioned aperture geometry as an active variable within the CXDI system design, rather than a fixed component.

By converting aperture shape from a fixed hardware choice into a tunable, data-driven variable, our study makes three concrete contributions to the computational-imaging community as follows:

- The systematic sweep of corner-radius and symmetry generates quantitative reliability curves, empirical plots that link each geometric parameter to reconstruction error in both real and reciprocal space. These curves now function as a ready-made lookup table. The optical engineer can read off the maximum tolerable rounding before fabrication, trimming costly trial-and-error runs at the beamline or in the clean room.

- The same curves translate directly into manufacturing tolerances. A mask designer can specify, for example, "corner radius $\leq 0.08\mu$ m for $\geq 95\%$ SSIM," turning our simulation output into actionable engineering limits.

- Most importantly, from a conceptual standpoint, the work forges a bridge between hardware and algorithm design. By showing how optical choices reshape the ill-posed inverse problem, the study demonstrates that reliability in computational imaging must be co-optimized across the entire pipeline, from

photons to reconstruction metrics, rather than treated as an isolated software or hardware tweak.

The second part of this dissertation introduced a novel ensemble phase retrieval framework that incorporates skipped-region masking as a strategy to probe sensitivity and quantify uncertainty. By generating multiple reconstructions under controlled data perturbations and aggregating them through soft voting, we were able to produce both a high-fidelity consensus image and a pixel-wise uncertainty map. These maps revealed structural zones that were consistently stable, as well as regions of high variability, particularly around object boundaries and overlapping particle areas, providing an interpretable layer of confidence that goes beyond standard evaluation metrics like SSIM or R_F. This mechanism parallels explainable AI approaches, allowing scientists to reason not only about what is reconstructed, but also how much confidence to place in each region of the result.

Both studies emphasized the importance of integrating domain-specific physical insight, whether through aperture geometry or diffraction space analysis, with algorithmic flexibility. In doing so, we demonstrated how optical design and computational strategies can work in tandem to enhance the robustness and interpretability of reconstructions in ill-posed inverse imaging problems. Furthermore, this dissertation proposed that ensemble phase retrieval is not just a tool for boosting accuracy, but a framework for navigating non-uniqueness and conveying reconstruction trustworthiness in a quantifiable manner.

Taken together, the two research streams of this thesis offer a comprehensive methodology for improving the practical and scientific value of single-frame CXDI. They provide not only higher-quality reconstructions but also clarity about where those reconstructions can be trusted, laying a foundation for more interpretable, adaptive, and uncertainty-aware imaging systems in the future. Beyond its technical and methodological contributions to phase retrieval and nanoscale imaging, this dissertation also advances the broader agenda of Knowledge Science, particularly within the domain of **Co-Creative Intelligence**. At its core, Co-Creative Intelligence explores how humans and intelligent systems can collaboratively generate, refine, and evaluate knowledge representations. The ensemble-based framework introduced in this work embodies this paradigm by transforming the reconstruction process from a deterministic, algorithm-driven output into an interpretable and uncertainty-aware dialogue between model behavior and scientific inference. The pixel-wise uncertainty maps, in particular, serve as epistemic indicators that invite human interpretation, providing feedback on which regions of the reconstruction are trustworthy, ambiguous, or potentially misleading. This aligns with the principles of co-creative systems, which aim to empower human experts through transparent algorithmic reasoning rather than replace their judgment. By enabling spatially explicit confidence visualization, the framework bridges computational reconstruction with expert decision-making, thereby fostering a hybrid intelligence loop where human intuition and machine inference co-evolve. As such, this work contributes not only to imaging science but also to the development of collaborative, explainable, and adaptive tools for knowledge creation in complex scientific domains.

5.2 Limitation

While this dissertation provides several novel contributions to phase retrieval in single-frame CXDI, there remain limitations that should guide future research directions.

- **Simulation-Only Validation:** All findings were obtained using simulated datasets under controlled conditions. While this allows for systematic and repeatable probing of aperture design and uncertainty quantification strategies, and offers the significant advantage of providing access to ground-truth object structures for evaluating phase retrieval performance in real space, it does not fully capture the complexities of real experimental conditions. Key phenomena such as detector imperfections, beamstop, induced occlusions, partial spatial or temporal coherence, and stochastic noise characteristics, particularly those arising from photon counting or environmental instabilities, are often difficult to model accurately or comprehensively. As such, the generalizability of the proposed methods to actual experimental datasets remains to be validated, and future work should include empirical studies to test whether the insights and uncertainty patterns observed in simulation persist under real-world conditions.

- **Computational Cost:** The ensemble-based framework—especially when applied with small skip-step sizes—incurs significant computational load due to the large number of reconstructions required. This may limit its scalability for high-resolution imaging or real-time applications unless acceleration strategies are employed.

- **Sensitivity to Masking Design:** The effectiveness of skipped-region strategies depends on how masking is applied. Although a systematic sliding window was used here, more adaptive masking informed by physics or data statistics might yield better diagnostic value. Random masking may miss critical zones or lead to under-informative ensembles.

- **No Direct Resolution of Non-Uniqueness:** Although the ensemble framework pinpoints which image regions are stable and which remain ambiguous, it does not guarantee a single, uniquely correct reconstruction. Multiple phase-equivalent solutions can still satisfy the measured diffraction data. Breaking this residual symmetry and fully resolving non-uniqueness will require additional physical constraints or independent experimental information. The revised version is stronger because it explicitly links the limitation to the ensemble's actual role, mapping stable versus ambiguous regions rather than enforcing uniqueness, and specifies what is still needed, additional physical constraints or independent measurements. By tying the limitation directly to the CXDI non-uniqueness problem and indicating a concrete path forward, the revised wording is easier to understand and more closely aligned with the study's scope.

- **Convergence Homogenization Risk:** Algorithms may converge to similar solutions across ensemble runs, particularly when strong priors are used. This may lead to underestimated uncertainty and falsely high confidence in consensus reconstructions.

- **Single-Frame Focus:** The proposed methods were developed and tested exclusively for single-frame CXDI. Their extension to multiframe, ptychographic, or time-resolved CXDI—where dynamics and temporal coherence matter—remains an open challenge for future studies.

In conclusion, this dissertation establishes a new framework for addressing both reconstruction accuracy and interpretability in CXDI. It highlights that improving phase retrieval performance is not solely about recovering sharper images, but about building mechanisms for trust, explanation, and informed analysis. Future work should focus on experimental validation, computational optimization, and extending the proposed strategies to dynamic and high-throughput imaging regimes. These directions will move the field closer to deploying CXDI not just as a high-resolution imaging tool, but as an explainable and reliable inference engine for complex material systems.

Chapter 6

Publication list

Publications

Here are all the publications that represent major contributions to the field of Knowledge Science within the scope of this thesis. These works collectively form the foundation and framework of the dissertation. All publications are centered on optimizing the reliability and accuracy of Single-Frame Coherent X-ray Diffraction Imaging (CXDI) through a data science-driven approach.

Published Paper

1. Mukharil Bachtiar, A., Vu, T.S., Ha, M.Q., Takazawa, S., Takahashi, Y. and Dam, H.C., 2024. *Impact of corner radius of the aperture on the accuracy of phase retrieval analysis in single-frame CDI for nanomaterials*. **Molecular Physics**, p.e2388303.

2. Vu, T.S., Ha, M.Q., Bachtiar, A.M., Dao, D.A., Tran, T., Kino, H., Takazawa, S., Ishiguro, N., Sasaki, Y., Abe, M. and Uematsu, H., 2025. *PID3Net: a deep learning approach for single-shot coherent X-ray diffraction imaging of dynamic phenomena*. **npj Computational Materials**, 11(1), p.66.

Paper in Progress (under review)

- Mukharil Bachtiar, A., Vu, T.S., Ha, M.Q., and Dam, H.C. *The Non-Uniqueness Problem in Single-Frame Coherent X-ray Diffraction Imaging (CXDI): A Detailed Exploration and Its Solution*. **Molecular Physics**.

International Conference Presentations (Poster)

1. Adam Mukharil Bachtiar, Tien-Sinh Vu, and Hieu-Chi Dam. *Optimizing Geometric Shape as Apertures in Single-Frame CDI*, Sanibel Symposium 2024, 27th February 2024, Florida, USA.

2. Adam Mukharil Bachtiar, Tien-Sinh Vu, and Hieu-Chi Dam. *Mitigating the Uncertainty in Single-Frame CXDI: an Ensemble Phase Retrieval Approach*, Sanibel Symposium 2025, 27th February 2025, Florida, USA.

3. Tien-Sinh Vu, Adam Mukharil Bachtiar, Minh-Quyet Ha, and Hieu-Chi Dam. *Attention Deep Learning Framework for Exploring Structure-Property Relationship in Molecular Dynamics*, Sanibel Symposium 2025, 24th February 2025, Florida, USA.

Appendix A

Detail Explanation of Single-Frame CXDI Simulation

A.1 CXDI Simulation in the First Study

This section describes the physical and computational parameters used to simulate the forward imaging process in single-frame CXDI as presented in Chapter 3. The simulation is designed to replicate a realistic synchrotron-based CXDI experimental setup, with emphasis on the propagation of coherent X-rays through a sample and the subsequent recording of far-field diffraction patterns on a detector. The simulation framework encodes essential optical conditions in a configuration class named **Optical Condition**, aperture conditions in an **Aperture Condition** class, and particle conditions in a **Particle Condition** class.

The **optical condition** class encodes the physical constants and core parameters that define the coherent X-ray imaging system in the simulation. These include the geometric layout (such as propagation distances), detector specifications, and photon-related quantities essential for modeling diffraction in the far-field (Fraunhofer) regime. Together, they represent a realistic model of experimental CXDI setups.

- **L**: Distance from the object plane to the detector in meters. This parameter sets the physical propagation scale of the X-ray wavefront and is critical in determining whether the system satisfies the far-field approximation.
- **p**: Detector pixel size in meters. This controls the spatial sampling density of the diffraction pattern and impacts reciprocal-space resolution.
- **Npixel**: Number of pixels in each dimension for the simulation window. This value defines the spatial extent of the simulated probe, object, and resulting diffraction patterns.
- **energy**: Incident X-ray energy in electron volts (eV). This governs the wavelength of the illumination beam and ultimately the imaging resolution.
- **h**: Planck constant, used to convert photon energy into wavelength.
- **c**: Speed of light in a vacuum, part of the fundamental constants required to calculate wavelength.
- **lamb**: X-ray wavelength computed from the energy using $\lambda = \frac{hc}{E}$. This value is central to the system's resolution capability and diffraction behavior.
- **k**: Wavenumber defined as $k = \frac{2\pi}{\lambda}$, which quantifies the phase oscillation per unit length.

- **dx**: Sample-plane pixel resolution derived from the detector geometry and wavelength:

$$dx = \frac{L \cdot \lambda}{N_{\text{pixel}} \cdot p}$$

  This parameter defines how each pixel in the object corresponds to a physical length scale in real space.

- **DR**: Detector dynamic range, expressed as the logarithm of the ratio between the maximum and minimum measurable photon counts. This affects contrast sensitivity in high dynamic range diffraction patterns.

- **Wnoise**: Boolean flag that determines whether photon shot noise is applied to the simulated diffraction data. When enabled, it injects Poisson noise to approximate real-world experimental variability.

The **aperture condition** class defines the geometric structure and placement of the aperture in the CXDI setup. In lensless coherent imaging, the aperture modulates the incoming coherent wavefront and determines the shape of the illumination profile reaching the sample. Aperture geometry directly affects the encoded diffraction pattern and plays a key role in phase retrieval effectiveness.

- **sideNumber**: Number of sides of the aperture shape. This determines the symmetry class of the aperture (e.g., 3 for triangle), which affects spatial frequency distribution in the diffraction pattern.

- **sideLength**: Physical length of each side of the polygonal aperture in meters. Larger sizes allow more light transmission and influence resolution and contrast in the diffraction domain.

- **Lpropagation**: Distance from the aperture plane to the sample in meters. This simulates the Fresnel propagation between the aperture and the object, as in real synchrotron beamlines.

- **needPropagation**: Boolean flag indicating whether forward propagation from the aperture to the sample is numerically required. It is activated when Lpropagation is non-zero.

The **particle condition** class defines the properties of the sample, composed of gold nanoparticles (AuNPs), which are used to simulate realistic object scattering in the CXDI imaging process. These particles affect the incident wavefield by introducing both amplitude attenuation and phase shifts, governed by the material's complex refractive index.

- **beta_Au**: Imaginary component of the refractive index for gold at the specified energy. It determines the absorption coefficient of the sample.

- **delta_Au**: Real component of the refractive index decrement, which quantifies the phase delay introduced by the sample.

- **diameter**: Diameter of each nanoparticle in meters. This sets the spatial resolution scale of the sample features.

- **ratio**: Ratio between the area occupied by particles and the total simulation window area. It controls the sample density.

- **number**: Total number of gold nanoparticles placed in the window. This is derived from the object area, particle size, and area ratio.
- **velocity**: Translational velocity of the particles per frame, used to simulate dynamic scenes over multiple diffraction frames (e.g., for time-resolved reconstruction).

After defining the optical, aperture, and particle configurations, the simulation proceeds to construct the actual illumination mask—a two-dimensional numerical representation of the aperture function used in the forward CXDI simulation. This step models how a physical aperture modulates the incoming wavefront before it reaches the object. The generation of this aperture is handled by a dedicated function that constructs a regular polygonal shape, such as a triangle, square, or pentagon, based on user-specified parameters such as the number of sides and the side length. These parameters allow flexible modeling of different aperture geometries, enabling controlled experiments to examine how shape influences reconstruction performance. The polygon is then rendered within a simulation grid and acts as a binary mask, marking transmissive regions where X-rays are allowed to pass and opaque regions where they are blocked.

To simulate realistic experimental conditions, the binary aperture mask is further processed using a Gaussian smoothing operation. This step accounts for non-idealities that occur in fabricated apertures, such as edge rounding due to lithographic limitations or diffraction-induced blurring. The smooth transitions at the edges of the mask better approximate how an actual X-ray beam is modulated before interacting with the sample. Collectively, this process yields an illumination function that governs the shape, symmetry, and spatial coherence of the wavefront in the object domain. In the context of Chapter 3, this model is central to investigating how variations in aperture design, particularly changes in corner sharpness, impact the spatial frequency content of the resulting diffraction pattern and the quality of phase retrieval. The aperture, therefore, acts not merely as a passive component but as a tunable design variable that directly influences reconstruction fidelity in single-frame CXDI.

Following the generation of the aperture function, the simulation constructs the complex-valued illumination wavefield that interacts with the object. This wavefield, often referred to as the incident wave or input wavefront, represents the spatial profile of the coherent beam shaped by the aperture and propagated to the sample plane. If the simulated geometry includes a nonzero distance between the aperture and the sample, free-space propagation must be computed to account for diffraction effects occurring between these two planes. In this simulation, the angular spectrum method is used to perform this propagation. The angular spectrum approach models the wavefront in terms of its spatial frequency components and simulates how each component evolves during propagation over a defined distance. It is particularly well-suited for near-field and Fresnel regime simulations, where wave curvature and angular divergence must be accurately captured. The aperture function is first transformed into its angular frequency domain using a forward Fourier transform. Each frequency component is then modulated by a complex exponential factor that encodes phase advancement over the propagation distance. Finally, an inverse Fourier transform returns the modulated wavefront to the spatial domain, now representing the field as it would appear at the sample plane after propagating from the aperture.

If no propagation is required (i.e., the aperture is placed directly adjacent to the object), the incident wave is modeled directly as the aperture function itself. In both cases, the wavefield is normalized to ensure that its total intensity matches a defined

illumination energy, typically expressed in terms of the detector's dynamic range. This normalization ensures consistency in simulated intensity levels across different experimental conditions.

After constructing the illumination wavefield, the simulation proceeds to model the wave interaction between the incident beam and the object, as well as the formation of the resulting diffraction patterns at the detector plane. This step represents the core of the forward imaging process in single-frame Coherent X-ray Diffraction Imaging (CXDI). The object in this simulation consists of gold nanoparticles distributed across a sample window. These particles modulate the incoming wavefield in both amplitude and phase, as determined by their material-specific complex refractive index. The interaction is modeled through element-wise multiplication between the object transmission function and the incident wavefield, extended across multiple frames to simulate temporal dynamics or sequential exposures. The diffracted wavefronts are then propagated to the detector plane using a two-dimensional Fourier transform, in accordance with the Fraunhofer diffraction approximation. This approximation is valid given the far-field condition assumed in the simulation geometry. The result is a complex-valued field at the detector plane, whose squared modulus yields the simulated diffraction intensity distribution.

To match experimental conditions, the intensity patterns are normalized such that their dynamic range spans from 1 photon to 10^{DR} photons, where `DR` is a user-defined parameter representing the logarithmic dynamic range of the detector. This normalization step ensures consistency in illumination energy across simulation runs and facilitates comparison with real-world measurements.

Further, the simulation accounts for the discrete nature of photon detection by suppressing intensity values below 1 photon, effectively removing low-count pixels that would fall below the detector's sensitivity threshold. If enabled, a Poisson-distributed noise model is applied to the normalized diffraction intensities. This step introduces photon shot noise, an intrinsic property of photon-counting detectors, and increases the realism of the synthetic data. The noise model is especially important for evaluating the robustness of phase retrieval algorithms under experimental variability and low-signal conditions. The final output of this process is a sequence of diffraction images that reflect both the physical interaction of the wavefield with the sample and the stochastic properties of photon detection. These images serve as the foundational data for downstream phase retrieval and reconstruction analysis. All the results from each stage can be examined in the Figure A.1.

A.2 CXDI Simulation in the Second Study

The second study in the Chapter 4, a comprehensive simulation framework was established to model the full optical pathway of a single-frame CXDI system, including illumination from a triangular aperture, focusing by a Fresnel Zone Plate (FZP), and interaction with a nanoparticle-based sample. This configuration allows the generation of synthetic diffraction datasets used in ensemble-based phase retrieval and uncertainty quantification.

The simulation begins by defining user-dependent numerical resolutions. Two key parameters govern spatial sampling: **Npixel**, which sets the resolution of the object and detector grids, and **Nwindow**, which defines the simulation size for wave propagation. Their appropriate ratio ensures that fine structures—such as those in FZP zones—can be adequately sampled while avoiding truncation during

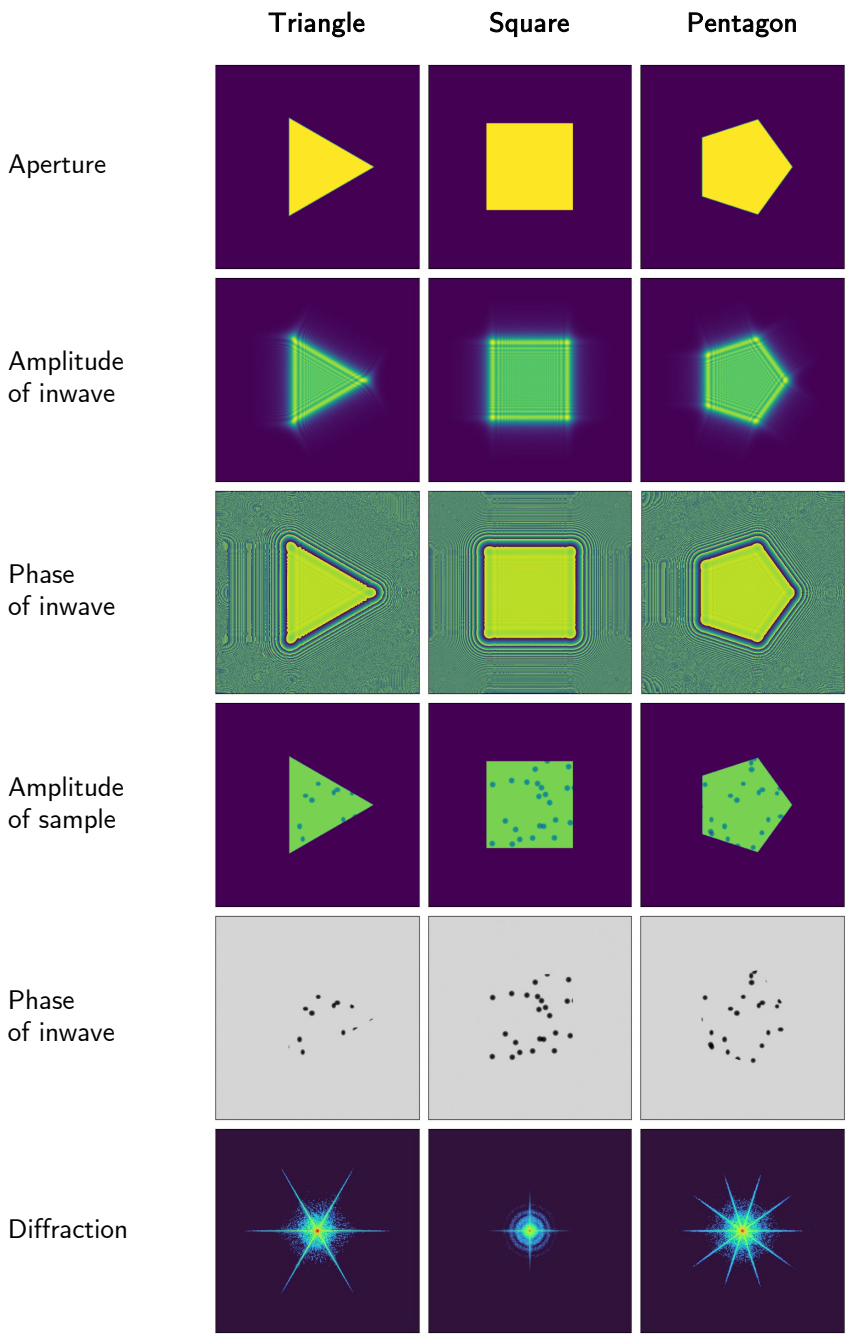


FIGURE A.1: Illustration samples of the CXDI forward simulation pipeline for three different aperture geometries: triangle, square, and pentagon. From top to bottom, each row shows: (1) the binary aperture shape; (2) the amplitude of the propagated wavefield at the sample plane; (3) the corresponding wavefront phase; (4) the amplitude modulation introduced by the object (gold nanoparticles); (5) the resulting phase of the wavefield after interaction with the sample; and (6) the intensity of the diffraction pattern recorded at the detector plane. This figure captures the evolution of the wavefield from aperture modulation to diffraction, highlighting the influence of aperture design on wavefront structure and diffraction outcome.

propagation. A pixel resolution finer than 20nm is required to resolve 40nm-wide FZP zones, making values of Npixel greater than 520 essential.

The **optical configuration** includes physical constants and parameters characterizing the system geometry and photon source. This comprises the X-ray energy, wavelength, sample-to-detector distance, and pixel size. In addition, simulation-specific elements such as beam intensity, off-axis shift, and FZP focal length are also defined. These values reflect experimental settings typically observed in synchrotron-based CXDI setups. Derived quantities such as pixel size at the sample (`dx`) and wavenumber (`k`) are calculated from these inputs, ensuring physical consistency across all simulation stages.

- **L**: Distance between the sample plane and the detector, determining the physical propagation length and establishing the validity of the far-field approximation.
- **p**: Pixel size of the detector in meters. Defines the spatial sampling interval in the diffraction plane.
- **Npixel**: Number of pixels used to discretize the object, probe, and diffraction windows. Influences the spatial resolution and computational grid size.
- **energy**: Incident X-ray energy in electron volts (eV), used to determine the wavelength of illumination.
- **h**: Planck constant, used in the calculation of photon energy to wavelength conversion.
- **c**: Speed of light in a vacuum, used alongside Planck's constant to compute the X-ray wavelength.
- **lamb**: Wavelength of the X-ray beam, derived from fundamental constants and energy. Central to diffraction and imaging scale.
- **k**: Wavenumber calculated from the wavelength, representing the spatial frequency of the wavefield.
- **dx**: Pixel resolution at the sample plane, derived from physical system dimensions. Influences the ability to resolve small-scale structures.
- **photon**: Total number of incident photons per second. Controls intensity scaling and photon count statistics.
- **Nwindow**: Size of the wavefield propagation window. Ensures correct sampling during free-space propagation.
- **mag**: Magnification factor of the Fresnel Zone Plate (FZP), influencing image scale and resolution.
- **Lf**: Focal length of the FZP, dictating the optical distance between the FZP and the image plane.
- **La**: Distance from the aperture plane to the FZP plane. Part of the three-part propagation path.
- **Lb**: Distance from the FZP to the sample. Completes the optical path for FZP-based imaging.

- **Loffax**: Off-axis shift applied to the FZP position, used to model beam steering or asymmetric setups.

The **aperture configuration** models a triangular aperture with rounded corners. The effective side length and curvature ratio are specified to simulate fabrication limitations observed in experimental apertures. This aperture acts as the initial spatial filter that defines the beam's symmetry and coherence properties before it enters the focusing optics.

- **Leff**: Effective side length of the triangular aperture in meters. Sets the physical scale of the aperture pattern.
- **curvature**: Dimensionless ratio representing the corner curvature relative to side length. Models edge rounding typical in experimental apertures.

Next, the **FZP configuration** defines the focusing element responsible for beam shaping. The number of concentric zones, their physical thickness, and the complex refractive index of the material (tantalum) are specified. These parameters govern the diffractive focusing behavior of the FZP, which is essential for imaging in the absence of conventional lenses. The zone count and thickness reflect typical designs used in hard X-ray nano-imaging.

- **Nzone**: Total number of concentric zones in the Fresnel Zone Plate. Determines the focusing precision and resolution.
- **beta_Ta**: Imaginary component of the refractive index of tantalum, representing absorption.
- **delta_Ta**: Real component of the refractive index decrement of tantalum, representing phase shift.
- **thickness**: Physical thickness of the FZP material. Impacts diffraction efficiency and phase modulation.

The sample itself is modeled in the **particle configuration** as a dynamic distribution of gold nanoparticles. The parameters include the complex refractive index (real and imaginary components), particle diameter, spatial density, number of frames, and translational velocity per frame. These values allow the simulation of a realistic dynamic object in CXDI, with frame-to-frame movement that mimics nanometer-scale dynamics. The total number of particles is computed to ensure that a predefined area ratio is maintained across the imaging window.

- **beta_Au**: Imaginary component of the refractive index for gold, responsible for absorption effects.
- **delta_Au**: Real component of the refractive index decrement for gold, contributing to phase modulation.
- **diameter**: Diameter of each gold nanoparticle in meters. Affects spatial scattering resolution.
- **ratio**: Fraction of the total simulation window occupied by nanoparticles. Defines sample density.
- **number**: Total number of gold particles, computed based on window area, particle size, and ratio.

- **Nframe**: Number of temporal frames simulated. Supports dynamic imaging scenarios.
- **velocity**: Displacement of particles per frame in meters. Enables simulation of translational motion for dynamic phase retrieval.

Together, these components establish a complete and physically grounded simulation environment that enables the generation of high-resolution, dynamic, and statistically meaningful diffraction data. This simulated data is then used as the input for the ensemble phase retrieval framework discussed in Chapter 4, enabling analysis of reconstruction stability, sensitivity to missing data, and uncertainty quantification.

After the setting of all parameters have been initialized, the triangular aperture is generated as the initial modulator of the coherent X-ray wavefield. The aperture defines the spatial support and angular spectrum of the illumination before it reaches the focusing optics, and plays a central role in shaping the information content of the resulting diffraction pattern. The triangular aperture is constructed using a user-defined function that takes into account the system's spatial resolution and the specified geometry. In this case, the aperture is modeled as a regular triangle with rounded corners, a geometry chosen to reflect realistic fabrication constraints encountered in experimental setups. The corner curvature is controlled parametrically, allowing systematic exploration of its influence on diffraction structure and phase retrieval stability.

The output of this process is a two dimensional array representing the aperture's amplitude transmission profile on a discretized grid. The grid size is defined by the object-plane resolution (**Npixel**), ensuring alignment with downstream simulation components such as the object, Fresnel Zone Plate, and propagated wavefields. The simulated aperture function is symmetric and centered within the window, although its coordinates may be flipped or transformed for display or alignment purposes. This triangular aperture serves as the starting point of the forward CXDI simulation, encoding spatial constraints that propagate through the entire optical system and influence the interpretability and reliability of the reconstructed object. Its systematic modulation—together with downstream ensemble strategies—enables a robust framework for uncertainty-aware phase retrieval in the second study.

Following the generation of the triangular aperture that defines the initial wavefront geometry, the next step in the simulation involves constructing the downstream optical elements and the object to be imaged. In this stage, both the Fresnel Zone Plate (FZP) and the gold nanoparticle (AuNP) object model are synthesized to complete the forward model of the CXDI system. The Fresnel Zone Plate is simulated as a diffractive focusing optic, consisting of a series of concentric zones designed to produce constructive interference at a focal point. The FZP is implemented as a complex-valued phase and amplitude mask on a high-resolution grid defined by `Nwindow`. Its optical properties are governed by material-specific parameters such as refractive index decrement and absorption coefficient, as well as geometric attributes including the number of zones and total thickness. The resulting FZP pattern introduces precise phase modulation across the wavefront, enabling high-resolution focusing required for nanoscale imaging.

To simulate practical conditions where perfect alignment is not always achievable, an off-axis shift can be introduced to the FZP by laterally displacing its center relative to the optical axis. This feature allows the exploration of misalignment effects and their impact on reconstruction robustness, which is particularly relevant for ensemble-based interpretability analysis. Simultaneously, the object is modeled as a distribution

of gold nanoparticles with properties that match experimental parameters. Each nanoparticle alters the wavefront by imposing both amplitude attenuation and phase shift, based on the complex refractive index of gold at the given energy. The spatial distribution of particles is randomized while controlling for coverage density and particle size. This ensures statistical diversity across the field of view and simulates realistic sample conditions.

The object is configured as a three-dimensional array, where each slice along the temporal dimension corresponds to a different frame. This setup enables the simulation of dynamic imaging sequences, in which nanoparticles translate incrementally between frames. Such temporal variation is essential for evaluating phase retrieval algorithms under conditions of structural change, and for validating the utility of ensemble strategies in tracking uncertainty across time.

After generating the triangular aperture, Fresnel Zone Plate (FZP), and gold nanoparticle object, the simulation continues with the propagation of the wavefield through the multi-element optical system. This stage models how the spatially coherent X-ray beam, shaped by the aperture, is further transformed as it passes through each optical component en route to the sample. The first propagation step simulates the free-space travel of the modulated wavefront from the aperture plane to the FZP plane. This is achieved using the angular spectrum method, which decomposes the wavefront into its constituent spatial frequencies and accurately models their evolution over a defined propagation distance. The method captures both near-field diffraction and wavefront curvature, making it suitable for modeling propagation within sub-meter-scale imaging systems.

At the FZP plane, the propagated wavefield is modulated by the complex-valued FZP transmission function. This interaction simulates the diffraction-based focusing behavior of the FZP, where each concentric zone introduces phase shifts that cause constructive interference at the focal plane. When the FZP is slightly displaced off-axis, as simulated here, the modulation introduces asymmetry in the transmitted wavefront, which can affect the focal spot profile and downstream diffraction characteristics. Following this interaction, the wavefield is again propagated, this time from the FZP plane to the sample plane. This second angular spectrum propagation simulates the focused illumination that interacts directly with the object. The resulting complex-valued wavefront—referred to as the incident wave at the sample—contains the combined effects of aperture geometry, FZP focusing behavior, and propagation physics.

At this stage, the spatial structure, intensity distribution, and phase front of the wavefield fully reflect the integrated behavior of the optical system. This wavefield serves as the input for object interaction and diffraction simulation, representing a high-fidelity model of how real X-ray beams behave in multi-optic lensless imaging systems. The use of physically accurate propagation methods ensures that the simulation captures essential effects such as beam shaping, off-axis distortions, and coherence-dependent interference, all of which contribute to the interpretability and accuracy of phase retrieval in downstream analysis.

After the wavefield is propagated to the sample plane, the simulation performs a spatial cropping operation to extract the region of interest (ROI) that corresponds to the illuminated area of the sample. This step is critical to isolate the physically meaningful portion of the wavefield that will interact with the object and contribute to the resulting diffraction pattern.

The region to be cropped is determined based on system geometry, including the pixel resolution, magnification factor, and any off-axis displacement introduced by

the FZP. By centering the crop region around the peak intensity of the wavefront and aligning it to the physical center of the simulation window, the extracted subregion accurately represents the focused probe that illuminates the sample. This ensures that subsequent simulation steps operate only on the wavefield portion relevant to the imaging process, while discarding peripheral artifacts that may arise from edge effects or numerical padding in the propagation window. Once extracted, the cropped wavefield is visualized in amplitude to inspect its structure and symmetry. This visualization allows qualitative evaluation of the focusing performance and any aberrations introduced by the aperture or FZP.

To further analyze the spatial frequency content of the illumination, a two-dimensional Fourier transform is applied to the cropped wavefield. The resulting intensity spectrum reveals how spatial frequencies are distributed in the probe, which is directly related to the resolving power of the imaging system. By examining the log-scaled power spectrum, one can assess the extent to which high-frequency components are preserved or attenuated—information that is crucial for interpreting diffraction image quality and reconstruction limits. This cropped probe image serves as the final illumination input to the object interaction phase. Its amplitude and phase characteristics govern the coherence, contrast, and dynamic range of the diffraction patterns, making this stage essential for ensuring physical fidelity and alignment with realistic beamline conditions.

Following the extraction of the focused probe at the sample plane, the simulation introduces a filtering operation in the Fourier domain to suppress unwanted spatial frequency components. This step is performed to emulate experimental conditions where aperture stops or beamline geometries inherently restrict the transmitted angular spectrum. The filter is defined as a circular binary mask, calculated based on system geometry, including distances between optical elements and the effective propagation magnification, and applied to the Fourier transform of the probe. The filter is centered relative to the dominant spatial frequency component in the illumination, ensuring that the most significant energy contribution is preserved while peripheral components are attenuated. The application of this mask suppresses noise and edge effects while enhancing the coherence and contrast of the illumination. After masking, the modified wavefront is transformed back into the spatial domain, producing a filtered and shifted wavefield that retains physically relevant features of the initial illumination.

Subsequently, the wavefield is normalized to ensure that the total energy corresponds to the specified photon flux. This normalization accounts for the total intensity of the probe and rescales it so that the integrated energy across the spatial field matches the desired number of incident photons per second. This step is essential to simulate photon-limited imaging conditions and to provide physically meaningful diffraction data for statistical and ensemble analysis. Once the wavefield is finalized, it is propagated into the object domain by multiplying it with the complex transmission function of the gold nanoparticle sample. This interaction produces a three-dimensional exit wavefield across all simulated frames, each containing different spatial configurations of nanoparticles due to their prescribed velocity.

The final step involves simulating the diffraction process by computing the two-dimensional Fourier transform of the exit wave for each frame. The modulus squared of this transformed wavefield yields the simulated diffraction intensities recorded at the detector plane. A photon threshold is applied, setting intensities below 1 photon to zero to replicate the sensitivity limit of real detectors. To further approximate experimental conditions, Poisson-distributed shot noise is added to the diffraction

images, modeling the stochastic nature of photon detection. This diffraction dataset serves as the core input for downstream ensemble phase retrieval and uncertainty quantification. It reflects both the physical propagation and interaction mechanisms of the CXDI system as well as practical noise characteristics inherent to experimental acquisition. Some process results of CXDI simulation in this section can be examined in Figure A.2

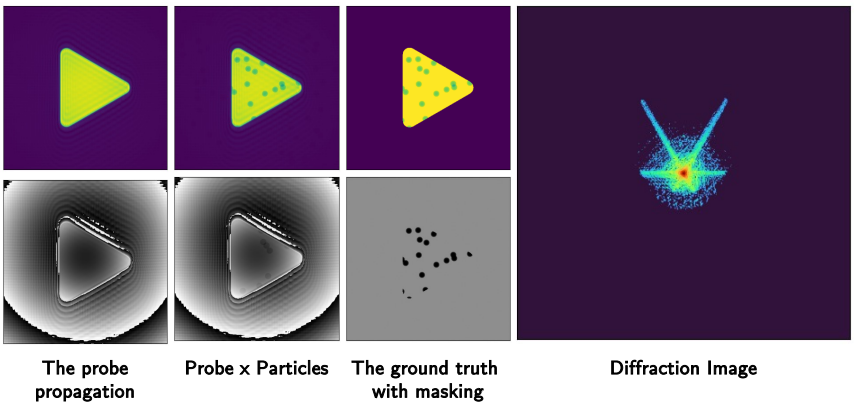


FIGURE A.2: Visualization of the CXDI simulation pipeline using a Fresnel Zone Plate (FZP) focusing system as applied in Chapter 4. From left to right: (1) the complex amplitude and phase of the focused probe after multi-stage wave propagation through a triangular aperture and FZP; (2) the interaction between the probe and a gold nanoparticle sample, producing the exit wavefield; (3) the masked ground-truth object used for ensemble-based reconstruction and interpretability analysis; and (4) the resulting diffraction pattern recorded at the detector, generated under photon-limited conditions with Poisson-distributed noise. This image illustrates the optical and computational flow from probe generation to data acquisition, forming the foundation for uncertainty-aware phase retrieval.

Appendix B

Phase Retrieval Results

This appendix provides additional visualizations of phase retrieval results to complement the main findings of this dissertation. It is divided into two parts:

- **Appendix B.1** focuses on supplementary results from Chapter 3, which examined the phase retrieval results in the study of the impact of aperture geometry on phase retrieval performance using iterative and deep learning-based methods.
- **Appendix B.2** presents complementary outputs from Chapter 4, which investigated the results of the ensemble-based phase retrieval strategy and uncertainty quantification using skipped-region analysis.

The supplementary figures or videos listed below offer a more detailed and time-resolved perspective on structural consistency, reconstruction stability, and spatial uncertainty across different algorithmic and physical conditions.

All supplementary materials referenced here are accessible via the following link: https://bit.ly/dissertationsup2220401

B.1 Phase Retrieval Results in the First Experiment

This section presents additional visualizations related to the first experimental study described in Chapter 3, which investigated the effect of aperture geometry, particularly the corner radius of a triangular aperture, on the quality of phase retrieval in single-frame CXDI. The supplementary materials compiled here aim to enhance the interpretation of the main findings by illustrating how both iterative and deep learning-based reconstruction methods respond to systematic variations in aperture shape.

The included videos and plots offer a dynamic perspective on the evolution of reconstruction performance, visual clarity, and error metrics (such as R_F and SSIM) across different aperture configurations. These materials support the conclusion that small modifications in aperture curvature can significantly affect the stability and fidelity of phase retrieval outcomes, especially in regions where the object contains fine spatial features or sharp transitions. Collectively, these visualizations provide deeper insight into the coupling between optical design and computational reconstruction fidelity. These are some explanations of the repository as follows:

- **Folder name: Experiment 1 Ground Truth**
  This folder contains the reference data used in the first experiment, serving as the simulated ground truth for both amplitude and phase of the object. It is structured into two subfolders, one for the amplitude distribution and one for

the phase profile, each representing the true object function used as a benchmark for evaluating phase retrieval accuracy. These ground truth datasets are essential for validating reconstruction results under different aperture configurations and allow quantitative comparison using error metrics such as SSIM and R_F across all tested methods.

Each file within these subfolders follows the naming convention **Angle_{corner radius}**, which indicates the specific aperture configuration associated with that data. For instance, a file named `Angle_10` corresponds to the ground truth used for reconstruction involving an aperture with a corner radius of 10°. This naming structure ensures clear traceability between ground truth and the corresponding reconstructed outputs across different reconstruction strategies.

- **Folder name: Experiment 1 Iterative PR**
  This folder contains the phase retrieval results obtained using iterative algorithms for the first experiment, which focused on evaluating the effect of aperture geometry on reconstruction performance. The contents are organized by corner radius variation, allowing direct visual and numerical comparison across different aperture configurations. Each subfolder includes both amplitude and phase reconstructions for a specific aperture shape, produced using the ePIE algorithm. This folder has a similar structure to the ground truth folder.

- **Folder name: Experiment 1 Deep Learning Result**
  This folder contains the phase retrieval results produced by the deep learning-based reconstruction framework applied in the first experiment. The deep learning model used in this study was trained to infer both amplitude and phase of the object from simulated diffraction patterns, under varying aperture configurations—specifically, different corner radius values of triangular apertures.

B.2 Phase Retrieval Results in the Second Experiment

This section presents supplementary materials related to the second experimental study described in Chapter 4, which introduced an ensemble-based phase retrieval framework combined with skipped region masking to quantify spatial uncertainty and improve interpretability. The results provided here include dynamic visualizations and reconstructions that illustrate the behavior of different phase retrieval algorithms, Mixed-State PIE, Multiframe PIE, and PID3Net, under systematic region-skipping strategies in both real and reciprocal space.

The supplementary materials are provided as a set of videos, each illustrating the dynamic behavior of phase retrieval reconstructions over 1000 consecutive frames. These videos visualize the evolution of the estimated phase, the ensemble consensus result, and the corresponding uncertainty maps under various reconstruction strategies and noise conditions. Each video corresponds to a distinct ensemble configuration and includes both qualitative and uncertainty-aware interpretations of the reconstructed object.

The video will have a format like the one which is displayed in Figure 4.15. It presents representative frames used in the videos. From left to right, each column shows:

- The ground-truth object.

- The simulated diffraction pattern with increasing noise levels (0%, 50%, 75%, and 100%).
- The phase reconstruction obtained from a single PID3Net model corresponds to the noise level.
- The consensus phase image obtained from ensemble aggregation corresponds to the noise level.
- The associated variance map visualizing pixel-wise ambiguity corresponds to the noise level.

This structure allows viewers to compare the impact of noise and reconstruction strategies on both the image fidelity and confidence distribution.

To enable systematic analysis and traceability, all video files follow a standardized naming convention with the following format:

```
ensemble_cxdi_<reconstruction_option>_<skip_size>.mp4
```

where:

- `reconstruction_option`: Indicates the aggregation strategy used in the ensemble reconstruction.
- `skip_size`: Indicates the size (in pixels) of the skipped region used during masking, which can be one of `8`, `16`, or `32`.

Available Reconstruction Options

- `all`: Aggregates all reconstructions across all solvers (Mixed-state PIE, Multiframe PIE, PID3Net).
- `onlyiterative`: Aggregates only the iterative-based reconstructions (Mixed-state PIE, Multiframe PIE).
- `onlydl`: Aggregates only PID3Net-based reconstructions.
- `top10_each`: Top-10 reconstructions per solver (Mixed-state PIE, Multiframe PIE, PID3Net), aggregated independently.
- `top20_each`: Top-20 reconstructions per solver (Mixed-state PIE, Multiframe PIE, PID3Net), aggregated independently.
- `top10_each_category`: Top-10 reconstructions from each algorithmic category (iterative group: Mixed-state PIE, Multiframe PIE; DL group: PID3Net).
- `top20_each_category`: Top-20 reconstructions from each algorithmic category (iterative group: Mixed-state PIE, Multiframe PIE; DL group: PID3Net).
- `top10_onlydl`: Top-10 reconstructions from PID3Net.
- `top20_onlydl`: Top-20 reconstructions from PID3Net.

This naming structure provides clarity, consistency, and navigability across ensemble strategies and experimental settings. Combined with the accompanying videos, it facilitates dynamic analysis of reconstruction quality and spatial ambiguity under varied noise levels and skip conditions.

Bibliography

Arsenault, HA and K Chalasinska-Macukow (1983). "The solution to the phase retrieval problem using the sampling theorem". In: *Optics communications* 47.6, pp. 380–386.

Bann, Stephen (2002). "Photography, printmaking, and the visual economy in nineteenth-century France". In: *History of photography* 26.1, pp. 16–25.

Barbastathis, George, Aydogan Ozcan, and Guohai Situ (2019). "On the use of deep learning for computational imaging". In: *Optica* 6.8, pp. 921–943.

Barnett, Alexander H et al. (2020). "Geometry of the phase retrieval problem". In: *Inverse Problems* 36.9, p. 094003.

Bauschke, Heinz H, Patrick L Combettes, and D Russell Luke (2002). "Phase retrieval, error reduction algorithm, and Fienup variants: a view from convex optimization". In: *Journal of the Optical Society of America A* 19.7, pp. 1334–1345.

Beinert, Robert (2017). "Non-negativity constraints in the one-dimensional discrete-time phase retrieval problem". In: *Information and Inference: A Journal of the IMA* 6.2, pp. 213–224.

Bergen, Michael et al. (2013). "Centralized instrument control for a TEM laboratory". In: *Microscopy and Microanalysis* 19.S2, pp. 1394–1395.

Bhagyaraj, Sneha Mohan and Oluwatobi Samuel Oluwafemi (2018). "Nanotechnology: the science of the invisible". In: *Synthesis of inorganic nanomaterials*. Elsevier, pp. 1–18.

Bhandari, Ayush, Achuta Kadambi, and Ramesh Raskar (2022). *Computational Imaging*. MIT Press.

Boutet, Sébastien and Garth J Williams (2010). "The coherent X-ray imaging (CXI) instrument at the Linac Coherent Light Source (LCLS)". In: *New Journal of Physics* 12.3, p. 035024.

Boyce, Adam M et al. (2022). "Cracking predictions of lithium-ion battery electrodes by X-ray computed tomography and modelling". In: *Journal of Power Sources* 526, p. 231119.

Candes, Emmanuel J and Michael B Wakin (2008). "An Introduction To Compressive Sampling. A sensing/sampling paradigm that goes against the common knowledge in data aquisition". In: *IEEE Signal Processing Magazine* 25.21.

Caro, Liberato De et al. (2013). "Coherent diffraction imaging with hard x-ray waveguides". In: *The X-Ray Standing Wave Technique: Principles and Applications*. World Scientific, pp. 427–440.

Chang, Huibin et al. (2019). "Advanced denoising for X-ray ptychography". In: *Optics express* 27.8, pp. 10395–10418.

Chapman, Henry N and Keith A Nugent (2010). "Coherent lensless X-ray imaging". In: *Nature photonics* 4.12, pp. 833–839.

Cherukara, Mathew J et al. (2020). "AI-enabled high-resolution scanning coherent diffraction imaging". In: *Applied Physics Letters* 117.4.

Cossairt, Oliver, Mohit Gupta, and Shree K Nayar (2012). "When does computational imaging improve performance?" In: *IEEE transactions on image processing* 22.2, pp. 447–458.

Deng, Zhe et al. (2021). "Recent progress on advanced imaging techniques for lithium-ion batteries". In: *Advanced energy materials* 11.2, p. 2000806.

Dong, Jonathan et al. (2023). "Phase retrieval: From computational imaging to machine learning: A tutorial". In: *IEEE Signal Processing Magazine* 40.1, pp. 45–57.

Edo, TB et al. (2013). "Sampling in x-ray ptychography". In: *Physical Review A—Atomic, Molecular, and Optical Physics* 87.5, p. 053850.

Eldar, Yonina C and Shahar Mendelson (2014). "Phase retrieval: Stability and recovery guarantees". In: *Applied and Computational Harmonic Analysis* 36.3, pp. 473–494.

Ellenbogen, Paul H (2013). "Imaging 3.0: what is it?" In: *Journal of the American college of Radiology* 10.4, p. 229.

Elser, Veit (2003). "Phase retrieval by iterated projections". In: *Journal of the Optical Society of America A* 20.1, pp. 40–55.

Elser, Veit, Ti-Yen Lan, and Tamir Bendory (2018). "Benchmark problems for phase retrieval". In: *SIAM Journal on Imaging Sciences* 11.4, pp. 2429–2455.

Falcone, Roger et al. (2011). "New directions in X-ray microscopy". In: *Contemporary Physics* 52.4, pp. 293–318.

Ferreira, Mário FS et al. (2017). "Roadmap on optical sensors". In: *Journal of Optics* 19.8, p. 083001.

Fienup, James R (1982). "Phase retrieval algorithms: a comparison". In: *Applied optics* 21.15, pp. 2758–2769.

Fromm, Cayla (2015). *Ptychography: Pushing the limits of X-ray microscopy*. Tech. rep. SLAC National Accelerator Lab., Menlo Park, CA (United States).

Fultz, Brent and James M Howe (2012). *Transmission electron microscopy and diffractometry of materials*. Springer Science & Business Media.

Goldberger, David et al. (2020). "Three-dimensional single-shot ptychography". In: *Optics Express* 28.13, pp. 18887–18898.

Grohs, Philipp, Sarah Koppensteiner, and Martin Rathmair (2020). "Phase retrieval: uniqueness and stability". In: *SIAM Review* 62.2, pp. 301–350.

Guan, Ziqiao et al. (2019). *Ptychonet: Fast and high quality phase retrieval for ptychography*. Tech. rep. Brookhaven National Lab.(BNL), Upton, NY (United States).

Guizar-Sicairos, Manuel and James R Fienup (2012). "Understanding the twin-image problem in phase retrieval". In: *Journal of the Optical Society of America A* 29.11, pp. 2367–2375.

Guo, Changliang, Shi Liu, and John T Sheridan (2015). "Iterative phase retrieval algorithms. I: optimization". In: *Applied optics* 54.15, pp. 4698–4708.

Guo, Cheng et al. (2019). "Multi-distance phase retrieval with a weighted shrink-wrap constraint". In: *Optics and Lasers in Engineering* 113, pp. 1–5.

Gupta, Ranjeetkumar et al. (2021). "A review of sensing technologies for non-destructive evaluation of structural composite materials". In: *Journal of Composites Science* 5.12, p. 319.

Haguenau, F et al. (2003). "Key events in the history of electron microscopy". In: *Microscopy and Microanalysis* 9.2, pp. 96–138.

Himanen, Lauri et al. (2019). "Data-driven materials science: status, challenges, and perspectives". In: *Advanced Science* 6.21, p. 1900808.

Hore, Alain and Djemel Ziou (2010). "Image quality metrics: PSNR vs. SSIM". In: *2010 20th international conference on pattern recognition*. IEEE, pp. 2366–2369.

Hu, Xue-mei et al. (2017). "Emerging theories and technologies on computational imaging". In: *Frontiers of Information Technology & Electronic Engineering* 18.9, pp. 1207–1221.

Hummel, Robert A, B Kimia, and Steven W Zucker (1987). "Deblurring gaussian blur". In: *Computer Vision, Graphics, and Image Processing* 38.1, pp. 66–80.

Inkson, Beverley J (2016). "Scanning electron microscopy (SEM) and transmission electron microscopy (TEM) for materials characterization". In: *Materials characterization using nondestructive evaluation (NDE) methods*. Elsevier, pp. 17–43.

Işıl, Çağatay, Figen S Oktem, and Aykut Koç (2019). "Deep iterative reconstruction for phase retrieval". In: *Applied optics* 58.20, pp. 5422–5431.

Jaganathan, Kishore, Yonina C Eldar, and Babak Hassibi (2016). "Phase retrieval: An overview of recent developments". In: *Optical Compressive Imaging*, pp. 279–312.

Javed, Yasir et al. (2018). "TEM for atomic-scale study: Fundamental, instrumentation, and applications in nanotechnology". In: *Handbook of materials characterization*, pp. 147–216.

Jiang, Shaowei et al. (2021). "Resolution-enhanced parallel coded ptychography for high-throughput optical imaging". In: *ACS Photonics* 8.11, pp. 3261–3271.

Johnson, Allan S et al. (2023). "Ultrafast X-ray imaging of the light-induced phase transition in VO2". In: *Nature Physics* 19.2, pp. 215–220.

Jumper, John et al. (2021). "Highly accurate protein structure prediction with AlphaFold". In: *nature* 596.7873, pp. 583–589.

Kamalabadi, Farzad (2009). "Multidimensional image reconstruction in astronomy". In: *IEEE Signal Processing Magazine* 27.1, pp. 86–96.

Kang, Jungmin et al. (2021). "Single-frame coherent diffraction imaging of extended objects using triangular aperture". In: *Optics Express* 29.2, pp. 1441–1453.

Keller, Joseph B (1957). "Diffraction by an aperture". In: *Journal of Applied Physics* 28.4, pp. 426–444.

Kim, Chan, Yoonhee Kim, Changyong Song, et al. (2016). "Enhancing resolution in coherent x-ray diffraction imaging". In: *Journal of Physics: Condensed Matter* 28.49, p. 493001.

Klibanov, Michael V, Paul E Sacks, and Alexander V Tikhonravov (1995). "The phase retrieval problem". In: *Inverse problems* 11.1, p. 1.

Konda, Pavan Chandra et al. (2020). "Fourier ptychography: current applications and future promises". In: *Optics express* 28.7, pp. 9603–9630.

Kumar, Sanjeev (2021). "Phase retrieval with physics informed zero-shot network". In: *Optics Letters* 46.23, pp. 5942–5945.

Latychevskaia, Tatiana (2018). "Iterative phase retrieval in coherent diffractive imaging: practical issues". In: *Applied optics* 57.25, pp. 7187–7197.

Leeman, Sidney and andrew J Healey (1997). "Field propagation via the angular spectrum method". In: *Acoustical Imaging*, pp. 363–368.

Levitan, Abraham L et al. (2020). "Single-frame far-field diffractive imaging with randomized illumination". In: *Optics Express* 28.25, pp. 37103–37117.

Levoy, Marc (2006). "Light fields and computational imaging". In: *Computer* 39.8, pp. 46–55.

Li, Chong et al. (2021). "The challenges of modern computing and new opportunities for optics". In: *PhotoniX* 2, pp. 1–31.

Liang, Mengning et al. (2015). "The coherent X-ray imaging instrument at the Linac Coherent Light Source". In: *Synchrotron Radiation* 22.3, pp. 514–519.

Luke, D Russell (2017). "Phase retrieval, what's new". In: *SIAG/OPT Views and News* 25.1, pp. 1–5.

Maiden, Andrew, Daniel Johnson, and Peng Li (2017). "Further improvements to the ptychographical iterative engine". In: *Optica* 4.7, pp. 736–745.

Maiden, Andrew M and John M Rodenburg (2009). "An improved ptychographical phase retrieval algorithm for diffractive imaging". In: *Ultramicroscopy* 109.10, pp. 1256–1262.

Mait, Joseph N, Gary W Euliss, and Ravindra A Athale (2018). "Computational imaging". In: *Advances in Optics and Photonics* 10.2, pp. 409–483.

Manekar, Raunak et al. (2020). "Deep learning initialized phase retrieval". In: *NeurIPS 2020 Workshop on Deep Learning and Inverse Problems*.

Marchesini, S (2004). "Benchmarking iterative projection algorithms for phase retrieval". In: *arXiv preprint physics/0404091*.

Marchesini, S et al. (2003). "Coherent X-ray diffractive imaging: applications and limitations". In: *Optics Express* 11.19, pp. 2344–2353.

Metzler, Christopher et al. (2018). "prDeep: Robust phase retrieval with a flexible deep network". In: *International Conference on Machine Learning*. PMLR, pp. 3501–3510.

Miao, Jianwei (2025). "Computational microscopy with coherent diffractive imaging and ptychography". In: *Nature* 637.8045, pp. 281–295.

Miao, Jianwei, Chien-Chun Chen, et al. (2006). "Three-dimensional GaN-Ga 2 O 3 core shell structure revealed by x-ray diffraction microscopy". In: *Physical review letters* 97.21, p. 215503.

Miao, Jianwei, Richard L Sandberg, and Changyong Song (2011). "Coherent x-ray diffraction imaging". In: *IEEE Journal of selected topics in quantum electronics* 18.1, pp. 399–410.

Mienye, Ibomoiye Domor and Yanxia Sun (2022). "A survey of ensemble learning: Concepts, algorithms, applications, and prospects". In: *Ieee Access* 10, pp. 99129–99149.

Moravec, Matthew L, Justin K Romberg, and Richard G Baraniuk (2007). "Compressive phase retrieval". In: *Wavelets XII*. Vol. 6701. SPIE, pp. 712–722.

Mukharil Bachtiar, Adam et al. (2024). "Impact of corner radius of the aperture on the accuracy of phase retrieval analysis in single-frame CDI for nanomaterials". In: *Molecular Physics*, e2388303.

Mukherjee, Subhadip et al. (2023). "Learned reconstruction methods with convergence guarantees: A survey of concepts and applications". In: *IEEE Signal Processing Magazine* 40.1, pp. 164–182.

Muoi, Pham Quy et al. (2018). "Inverse problems with nonnegative and sparse solutions: algorithms and application to the phase retrieval problem". In: *Inverse Problems* 34.5, p. 055007.

Nakasako, Masayoshi et al. (2020). "Methods and application of coherent X-ray diffraction imaging of noncrystalline particles". In: *Biophysical reviews* 12, pp. 541–567.

Nashed, Youssef SG et al. (2014). "Parallel ptychographic reconstruction". In: *Optics express* 22.26, pp. 32082–32097.

Natekin, Alexey and Alois Knoll (2013). "Gradient boosting machines, a tutorial". In: *Frontiers in neurorobotics* 7, p. 21.

Nishizaki, Yohei et al. (2020). "Analysis of non-iterative phase retrieval based on machine learning". In: *Optical Review* 27, pp. 136–141.

Osherovich, Eliyahu (2012). "Numerical methods for phase retrieval". In: *arXiv preprint arXiv:1203.4756*.

Panwar, Aditya Singh, Armender Singh, and Shankar Sehgal (2020). "Material characterization techniques in engineering applications: A review". In: *Materials Today: Proceedings* 28, pp. 1932–1937.

Patterson, Brian M et al. (2018). "Data challenges of in situ X-ray tomography for materials discovery and characterization". In: *Materials Discovery and Design: By Means of Data Science and Optimal Learning*, pp. 129–165.

Pfeiffer, Franz (2018). "X-ray ptychography". In: *Nature Photonics* 12.1, pp. 9–17.

Polikar, Robi (2012). "Ensemble learning". In: *Ensemble machine learning: Methods and applications*, pp. 1–34.

Prosekov, PA, VL Nosik, and AE Blagov (2021). "Methods of coherent X-ray diffraction imaging". In: *Crystallography reports* 66, pp. 867–882.

Robinson, Ian and Ross Harder (2009). "Coherent X-ray diffraction imaging of strain at the nanoscale". In: *Nature materials* 8.4, pp. 291–298.

Robinson, Ian K et al. (2001). "Reconstruction of the shapes of gold nanocrystals using coherent x-ray diffraction". In: *Physical review letters* 87.19, p. 195505.

Rodenburg, John and Andrew Maiden (2019). "Ptychography". In: *Springer Handbook of Microscopy*, pp. 819–904.

Rodenburg, John M (2008). "Ptychography and related diffractive imaging methods". In: *Advances in imaging and electron physics* 150, pp. 87–184.

Rosenblatt, Joseph (1984). "Phase retrieval". In: *Communications in mathematical physics* 95.3, pp. 317–343.

Sagi, Omer and Lior Rokach (2018). "Ensemble learning: A survey". In: *Wiley interdisciplinary reviews: data mining and knowledge discovery* 8.4, e1249.

Saha, Swapan K (2002). "Modern optical astronomy: technology and impact of interferometry". In: *Reviews of Modern Physics* 74.2, p. 551.

Schapire, Robert E (2013). "Explaining adaboost". In: *Empirical inference: festschrift in honor of vladimir N. Vapnik*. Springer, pp. 37–52.

Schroer, Christian G et al. (2008). "Coherent x-ray diffraction imaging with nanofocused illumination". In: *Physical review letters* 101.9, p. 090801.

Setiadi, De Rosal Igantius Moses (2021). "PSNR vs SSIM: imperceptibility quality assessment for image steganography". In: *Multimedia Tools and Applications* 80.6, pp. 8423–8444.

Shariat, Bashir S et al. (2019). "Computational and experimental analyses of martensitic transformation propagation in shape memory alloys". In: *Journal of Alloys and Compounds* 806, pp. 1522–1528.

Shechtman, Yoav et al. (2015). "Phase retrieval with application to optical imaging: a contemporary overview". In: *IEEE signal processing magazine* 32.3, pp. 87–109.

Shi, Xiaowen (2018). "General Overview of Coherent X-Ray Diffraction Imaging and Ptychography and Their Developments and Applications". In: *Advances in Optics Reviews 1*, p. 251.

Sidorenko, Pavel and Oren Cohen (2015). "Single-shot ptychography". In: *Optica* 3.1, pp. 9–14.

Sliney, DH (2016). "What is light? The visible spectrum and beyond". In: *Eye* 30.2, pp. 222–229.

Smyth, MS and JHJ Martin (2000). "x Ray crystallography". In: *Molecular Pathology* 53.1, p. 8.

Takayama, Yuki et al. (2021). "Dynamic nanoimaging of extended objects via hard X-ray multiple-shot coherent diffraction with projection illumination optics". In: *Communications Physics* 4.1, p. 48.

Takazawa, Shuntaro, Duc-Anh Dao, et al. (2023). "Coupling x-ray photon correlation spectroscopy and dynamic coherent x-ray diffraction imaging: Particle motion analysis from nano-to-micrometer scale". In: *Physical Review Research* 5.4, p. L042019.

Takazawa, Shuntaro, Jungmin Kang, et al. (2021). "Demonstration of single-frame coherent X-ray diffraction imaging using triangular aperture: Towards dynamic nanoimaging of extended objects". In: *Optics Express* 29.10, pp. 14394–14402.

Takazawa, Shuntaro, Yuhei Sasaki, et al. (2025). "In situ heating coherent X-ray diffraction imaging for visualizing nanometer-scale structural changes in metallic materials". In: *Materialia* 39, p. 102311.

Tang, CY and Z Yang (2017). "Transmission electron microscopy (TEM)". In: *Membrane characterization*. Elsevier, pp. 145–159.

Thibault, Pierre and Andreas Menzel (2013). "Reconstructing state mixtures from diffraction measurements". In: *Nature* 494.7435, pp. 68–71.

Thurin, Julien, Romain Brossier, and Ludovic Métivier (2019). "Ensemble-based uncertainty estimation in full waveform inversion". In: *Geophysical Journal International* 219.3, pp. 1613–1635.

Tripathi, Shalini et al. (2018). "Phase-change materials; the challenges for TEM". In: *Microscopy and Microanalysis* 24.S1, pp. 1904–1905.

Uozumi, Jun, Hiroyuki Kimura, and Toshimitsu Asakura (1990). "Fraunhofer diffraction by Koch fractals". In: *Journal of modern optics* 37.6, pp. 1011–1031.

Van der Veen, Friso and Franz Pfeiffer (2004). "Coherent x-ray scattering". In: *Journal of Physics: Condensed Matter* 16.28, p. 5003.

Varadi, Mihaly et al. (2024). "AlphaFold Protein Structure Database in 2024: providing structure coverage for over 214 million protein sequences". In: *Nucleic acids research* 52.D1, pp. D368–D375.

Vu, Tien-Sinh et al. (2025). "PID3Net: a deep learning approach for single-shot coherent X-ray diffraction imaging of dynamic phenomena". In: *npj Computational Materials* 11.1, p. 66.

Wang, Fei et al. (2020). "Phase imaging with an untrained neural network". In: *Light: Science & Applications* 9.1, p. 77.

Wang, Yujie et al. (2008). "Ultrafast X-ray study of dense-liquid-jet flow dynamics using structure-tracking velocimetry". In: *Nature Physics* 4.4, pp. 305–309.

Wang, Zhuo et al. (2022). "Data-driven materials innovation and applications". In: *Advanced Materials* 34.36, p. 2104113.

Wankowicz, Stephanie A et al. (2024). "Automated multiconformer model building for X-ray crystallography and cryo-EM". In: *Elife* 12, RP90606.

Weker, Johanna Nelson, Xiaojing Huang, and Michael F Toney (2016). "In situ X-ray-based imaging of nano materials". In: *Current opinion in chemical engineering* 12, pp. 14–21.

Wilke, Robin Niklas (2015). *Coherent X-ray diffractive imaging on the single-cell-level of microbial samples*. Göttingen University Press.

Williams, David B et al. (1996). "Imaging in the TEM". In: *Transmission Electron Microscopy: A Textbook for Materials Science*, pp. 349–366.

Winey, Mark et al. (2014). "Conventional transmission electron microscopy". In: *Molecular biology of the cell* 25.3, pp. 319–323.

Withers, Philip J (2013). "Synchrotron X-ray Diffraction". In: *Practical residual stress measurement methods*, pp. 163–194.

Withers, Philip J et al. (2021). "X-ray computed tomography". In: *Nature Reviews Methods Primers* 1.1, p. 18.

Wu, Longlong et al. (2021). "Three-dimensional coherent x-ray diffraction imaging via deep convolutional neural networks". In: *npj Computational Materials* 7.1, p. 175.

Xiang, Mingjun et al. (2024). "Amplitude/phase retrieval for terahertz holography with supervised and unsupervised physics-informed deep learning". In: *IEEE Transactions on Terahertz Science and Technology* 14.2, pp. 208–215.

Yang, Xiaogang et al. (2024). "Physics-informed generative network for phase retrieval from a single intensity measurement". In.

Yang, Yixiao et al. (2024). "Single-shot phase retrieval from a fractional Fourier transform perspective". In: *IEEE Transactions on Signal Processing*.

Ye, Qiuliang, Li-Wen Wang, and Daniel PK Lun (2023). "Towards practical single-shot phase retrieval with physics-driven deep neural network". In: *Optics express* 31.22, pp. 35982–35999.

Yeh, Li-Hao et al. (2015). "Experimental robustness of Fourier ptychography phase retrieval algorithms". In: *Optics express* 23.26, pp. 33214–33240.

Zalevsky, Zeev, David Mendlovic, and Rainer G Dorsch (1996). "Gerchberg–Saxton algorithm applied in the fractional Fourier or the Fresnel domain". In: *Optics Letters* 21.12, pp. 842–844.

Zhang, Yu, Xiaobo Tian, and Rongguang Liang (2019). "Timesaving phase retrieval approach based on difference map normalization and fast iterative algorithm". In: *Optics and Lasers in Engineering* 121, pp. 18–28.

Zheng, Guoan et al. (2021). "Concept, implementations and applications of Fourier ptychography". In: *Nature Reviews Physics* 3.3, pp. 207–223.

Zhu, Yimei and Hermann Dürr (2015). "The future of electron microscopy". In: *Physics Today* 68.4, pp. 32–38.