The Development of Non-iterative Electron Ptychography Methods for Fast and Low-fluence Imaging



Zhiyuan Ding

Oriel College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas 2024

Abstract

Electron ptychography has been a widely used imaging method in TEM due to its high resolution and adaptability to various imaging conditions, including low-fluence (low-dose) condition, focused and defocused probes. This thesis explored the development of non-iterative electron ptychography for fast imaging under low-fluence conditions, offering insights into its potential and limitations.

In this thesis, the impact of inelastic electron scattering on ptychographic reconstruction is discussed. Experiment and calculation results indicate that inelastically scattered electrons have a minimal effect on typical atomic resolution experiments. The optimized defocus condition for various accelerating voltages and convergence angles are calculated. Additionally, TDS (Thermal Diffuse Scattered) electrons do not significantly affect ptychographic reconstructions, although they may degrade resolution if dark field signals are considered. Therefore, inelastically scattered electrons can be included into the dataset for ptychography reconstruction to improve electron fluence efficiency.

Furthermore, a phase reconstruction method named as Single-Side Band masked Centre-of-mass (SSBm-COM), which combines SSB ptychography and Centre-of-mass (COM) imaging is proposed and discussed. This method applies the weak phase object approximation (WPOA) and can compensate for aberrations in 4DSTEM datasets. This method reduces the total electron fluence by minimizing the need for fine-tuning defocus before data collection.

In addition, Arbitrary Detector Ptychography (AD-ptycho), which can reconstruct data recorded on arbitrary multi-channel detectors under the weak phase object approximation

(WPOA) is introduced. This method can compensate defocus and can be applied to various detector layouts, including segmented detectors which can reduce data collection time. However, the reliance on WPOA remains a theoretical constraint.

Based on AD-ptycho, Linear System Solution ptychography method (LSS) for analysing 4DSTEM datasets without relying on the WPOA is then developed. Similar to AD-ptycho, LSS can be configured to operate with non-pixelated multichannel detectors. LSS has potential for imaging non-WPOA samples with a fast segmented detector.

Dedication

Dedicated to my wife, Minyi

Acknowledgements

Firstly, I would like to thank my supervisors, Professors Peter Nellist and Angus Kirkland, for their consistent mentoring and advice during my DPhil studies.

I would like to thank Dr Chen Huang for his support on the 4DSTEM experiment of graphene in ePSIC, thank Dr Adrián Pedrazo Tardajos and Dr Amirafshar Moshtaghpour from the Rosalind Franklin Institute (RFI) for providing monolayer graphene sample and discussing about optimization algorithms. I would like to thank Dr Judy Kim, Dr Marcus Gallagher-Jones, Dr Jonathan Barnard and all RFI staff for their support during the writing of my thesis.

I am also thankful to students and post-docs in Nellist group and Kirkland group. Thanks Dr Weixin Song, Dr Ali Mostaed, Dr Zezhong Zhang, Dr Xiaonan Luo, Dr Botao Hao, Dr Ping-Luen Ho, Dr Emma Hedley, Dr Ruomu Zhang, Mr Zhiyang Zhong, Mr Felix Bennemann and Mr Alessandro Zanre.

I would like to thank all the staff in DCCEM and ePISC for their support to my experiment sessions.

Finally, I would like to give special thanks to my family and my wife, Minyi Zhang for their constant encouragement and support, without whom this thesis would not exist.

Disclaimer

I hereby declare that this thesis is my own original work and was solely undertaken during my DPhil studies at the University of Oxford. I have acknowledged all main sources of help. Where the work in the thesis is based on a collaboration with others, I have explicitly stated what work was done by me, and what was done by others. For all the images used in this thesis which are not my own, I have received permission from the copyright holders and accredited them in the figure captions. I have clearly cited those research works that I have consulted throughout my work.

Abbreviations

AD-ptycho	Arbitrary Detector ptychography
CBED	Convergent beam electron diffraction
COM	Centre-of-mass
Cryo-EM	Cryogenic transmission electron microscopy/microscope
CTF	Contrast transfer function
dCOM	Differential Centre-of-mass
DO	Double overlap region
DPC	Differential phase contrast
EDX/EDS	Energy-dispersive X-ray spectroscopy
EELS	Electron energy loss spectroscopy
ePIE	Extended Ptychographical iteration engine
FCC	Face-centred cubic
iCOM	Integrated Centre-of-mass
iDPC	Integrated differential phase contrast
LSM	Least square method
LSS	Linear system solution
pCTF	Phase contrast transfer function
PIE	Ptychographical iteration engine
POA	Phase object approximation
QEP	Quantum Excitation of Phonons model
RMSE	Root mean square error

SSB	Single-side band
SSBm-COM	Single-side band masked centre-of-mass
SSBm-iCOM	Single-side band masked integrated centre-of-mass
SSBm-dCOM	Single-side band masked differential centre-of-mass
STEM	Scanning transmission electron microscope
SVD	Singular value decomposition
TDS	Thermal diffused scattering
TEM	Transmission electron microscope
TF	Transfer function
TO	Triple overlap region
WDD	Wigner distribution deconvolution
WPOA	Weak phase object approximation
4DSTEM	Four-dimensional STEM

Notation

A	Aperture function
A_t	Gate function, amplitude of aperture function, see Equation (2-5)
$\boldsymbol{COM^f}$	Filtered Centre-of-mass, see Section 5.2
C_c	Chromatic aberration
C_n	Signal from a detector that integrals all electron intensity within the detector area
C_1	Defocus
C_{12}	Astigmatism
C_{21}	Axial coma
C_{23}	3-fold astigmatism
C_3	Spherical aberration
C_{32}	2-fold astigmatism of spherical aberration
C_{34}	4-fold astigmatism
c	Speed of light in vacuum
c_n	The Fourier transform of C_n
c_v	A vector form by c_n, see Section 6.2
$\boldsymbol{com^f}$	Fourier transform of filtered Centre-of-mass, see Section 5.2
D_n	Detector function
d	d-spacing
df_{MAX}	The limiting defocus, see Section 4.2.3.3

E	Envelope function
$\boldsymbol{E_p}$	Projected electrostatic field
E_0	Energy of incident electron beam
e	Charge of an electron
EW	Exit wave
$EW_{E,n}$	Expected exit wave for the nth iteration in ePIE algorithm
$EW_{R,n}$	Exit wave after amplitude replacement in the nth iteration in ePIE algorithm
$\mathcal{F}$	Fourier transform
$\mathcal{F}^{-1}$	Inverse Fourier transform
G	4D array, calculated from performing Fourier transform on 4DSTEM dataset with respect to probe positions
$\boldsymbol{g}$	An arbitrary spatial frequency corresponding to a scattering angle
H	4D array, calculated from performing Fourier transform on 4DSTEM dataset with respect to probe positions and coordinate on diffraction plane.
H_t	HRTEM transfer function
h	Plank's constant
h_t	Point spread function
I	Intensity of a wave function
I_t	Integrated intensity of an EELS spectrum
I_{ZLP}	Integrated intensity of zero-loss peak
$iCOM^f$	Filtered integrated Centre-of-mass, see Section 5.2
I_{obj}	Current of objective lens

$\boldsymbol{K}$	Reciprocal space coordinate, vector
$\boldsymbol{K_f}$	Reciprocal space coordinate, diffraction plane coordinate, vector
$\boldsymbol{K_f'}$	Dummy reciprocal space coordinate, diffraction plane coordinate, vector, see Section 7.2
K_{fx}	Reciprocal space coordinate, diffraction plane coordinate on x axis
K_{fy}	Reciprocal space coordinate, diffraction plane coordinate on y axis
M	Wave function at diffraction plane
$M_{E,n}$	Expected save function at diffraction plane for n^{th} iteration in ePIE algorithm
m	Mask function
m_0	Rest mass of an electron
P	Probe function
P_n	Probe function for n^{th} iteration in ePIE algorithm
P_s	The probability that an inelastically scattered electron with a specific energy loss still lands on the same pixel
P_0	Initial probe function for ePIE algorithm
$\boldsymbol{Q_p}$	Frequency corresponding to probe scanning
$\boldsymbol{R}$	Real space coordinate, sample plane coordinate, vector
$\boldsymbol{R_p}$	Real space coordinate, sample plane coordinate for probe positions, vector
R_{px}	Real space coordinate, sample plane coordinate for probe positions on x axis
R_{py}	Real space coordinate, sample plane coordinate for probe positions on y axis

R_x	Real space coordinate, sample plane coordinate on x axis
R_y	Real space coordinate, sample plane coordinate on y axis
TF	Transfer function
T_m	Matrix of a linear system, see Section 6.2
T_n	An integral term based on detector function and aperture function, see Section 7.2
T_{Neg}	A function calculated from detector function and aperture function, see Section 6.2
T_{Pos}	A function calculated from detector function and aperture function, see Section 6.2
t	Sample thickness
V_p	Projected electric potential
V_0	Acceleration voltage in TEM
W	Wave function at image plane in TEM
α	Convergence angle for STEM
α_P	Update factor for probe function in ePIE algorithm
α_s	Angular spread of the electron source
α_ψ	Update factor for transmission function in ePIE algorithm
Γ	A function constructed by aperture function, see Section 5.3
Δf_{Sch}	Scherzer defocus
ΔE	Energy loss
ΔE_s	Energy spread
ϵ	Super parameter in WDD to prevent the denominator from being zero
δ	Dirac function

δ_c	A factor related to chromatic aberration, see Equation (2-6)
δ_{df}	Defocus spread
θ	Scattering angle
θ_d	Scattering angle corresponding to d-spacing
θ_E	Characteristic scattering angle
λ	Wavelength
λ_{iMFP}	Inelastic mean free path
σ	Interaction parameter of projected electric potential and phase change
Φ	An intermediate result related to ψ, a 4D array, see Equation (3-45)
Φ_v	A vector of Φ under different $\boldsymbol{K}_{\boldsymbol{f}}'$, see Section 7.2
Φ_{WDD}	The Φ calculated by WDD method
χ	Phase distortion function
χ_q	Wigner distribution function with respect to q
Ψ	Fourier transform of transmission function
$\Psi_{AD-ptycho}$	An intermediate result in AD-ptycho algorithm, the Fourier transform of $\psi_{AD-ptycho}$
Ψ_s	Fourier transform of phase of sample transmission function under (weak) phase object approximation
Ψ_{SSB}	An intermediate result in SSB algorithm, the Fourier transform of ψ_{SSB}
Ψ_{WDD}	An intermediate result in WDD algorithm, also used in LSS method.
ψ	Sample transmission function or object function
$\psi_{AD-ptycho}$	Reconstruction result of AD-ptycho algorithm

ψ_n	Transmission function for n^{th} iteration in ePIE algorithm
ψ_{POA}	Sample transmission function or object function under phase object approximation
ψ_s	Phase of sample transmission function or object function under (weak) phase object approximation
ψ_{SSB}	Reconstruction result of SSB algorithm
ψ_{WPOA}	Sample transmission function or object function under weak phase object approximation
ψ_0	Initial transmission function for ePIE algorithm
Ψ_{sv}	A vector in a linear system, see Section 6.2

Contents

Chapter 1: Overview 1
Reference 3

Chapter 2: Literature Review 5
2.1 Introduction 5
2.1.1 A Brief introduction to TEM 5
2.1.2 TEM and STEM Imaging 6
2.2 The Phase Object and Weak Phase Object Approximations 8
2.3 Theoretical Background of Phase Contrast in TEM 9
2.3.1 The Transfer Function for HRTEM 10
2.3.2 The Scherzer Defocus 13
2.3.3 The Transfer Function with Aberration Corrector 14
2.4 Phase Contrast Imaging methods 16
2.4.1 HRTEM 17
2.4.1.1 The Zernike Phase Plate 19
2.4.2 Off-axis Holography 22
2.4.3 Ptychography 25
2.4.3.1 Non-iterative Methods 29
2.4.3.2 Iterative methods 31
2.4.4 Focal Series Reconstruction 36
2.4.5 Differential Phase Contrast and Centre-of-Mass 39
2.5 Conclusion 45
Reference 45

Chapter 3: Methods 58

3.1 Introduction 58
3.2 Source of Information for Ptychography and Centre-Of-Mass 59
3.2.1 Object Transmission Function and the Sample. 59
3.2.1.1 Elastic and Inelastic, Single and Multiple Scattering 59
3.2.1.2 Transmission Function 60
3.2.1.3 Phase Object Approximation 61
3.2.1.4 Weak Phase Object Approximation 62
3.2.2 Probe Function and Aberrations 63
3.2.3 Phase Information in Diffraction Patterns 65
3.3 4DSTEM 71
3.3.1 Mathematical description of a 4DSTEM dataset 72
3.3.2 Phase Information in a 4DSTEM dataset 74
3.4 Algorithm of Wigner Distribution Deconvolution 75
3.5 Algorithm of Single-side Band 77
3.6 Algorithm of Centre-of-mass 80
3.7 Algorithm of Extended Ptychographic Iterative Engine 83
3.8 Conclusion 84
Reference 85

Chapter 4: Influence of Inelastically Scattered Electrons on Electron Ptychography 88
4.1 Introduction 88
4.2 Influence of Plasmon-excited Electrons 89
4.2.1 Energy-filtered 4DSTEM 90
4.2.2 Energy-filtered Ptychography 93
4.2.3 Factors Affecting Energy-filtered Ptychography 95
4.2.3.1 Characteristic Inelastic Scattering Angle 96
4.2.3.2 Defocus and Interference Fringes 96
4.2.3.3 The Limiting Defocus 101

4.2.3.4 Detector Sampling 103
4.3 High-loss CBED patterns 105
4.4 Influence of Thermal Diffuse Scattered Electrons 107
4.4.1 Simulation of 4DSTEM with TDS 107
4.4.2 Simulations of Ptychography with TDS 109
4.5 Conclusion 113
Reference 114

Chapter 5: Single-side Band Masked Centre-of-Mass Imaging 118
5.1 Introduction 118
5.2 SSBm-COM Algorithm 119
5.3 Transfer Function 124
5.4 Experiments on Defocused samples 127
5.5 Simulations 130
5.5.1 SSBm-iCOM under Different Fluence Conditions 130
5.5.2 Low-frequency transfer 131
5.6 Conclusion 138
Reference 139

Chapter 6: Ptychography in the Weak Phase Object Approximation using an Arbitrary Detector 142
6.1 Introduction 142
6.2 The AD-ptycho Method 143
6.3 Solvable Conditions and the Shape of $\boldsymbol{Dn}$ 147
6.4 Weighting of $\boldsymbol{\Psi s}$ and Information Transfer 153
6.5 Aberration Correction 155
6.6 Experimental data 157
6.7 Connections to Other 4DSTEM methods 165
6.7.1 Connection to Single-side Band 165

6.7.2 Connection to COM 167
6.8 Conclusion 168
Reference 169

Chapter 7: Ptychographic Linear System Solution Method without WPOA .. 171
7.1 Introduction 171
7.2 Linear system of 4DSTEM dataset without WPOA 172
7.3 Solving the Transmission function 176
7.4 WPOA on LSS and Connection to AD-ptycho 177
7.5 Visualization of Matrix Elements ($\boldsymbol{Tn}$) 180
7.6 Resolution 182
7.7 Simulations 183
7.7.1 $\boldsymbol{Kf'} = \boldsymbol{0}$ and Simulation of a WPOA Sample 184
7.7.2 Simulation of a non-WPOA Sample 185
7.7.3 $\boldsymbol{Kf'}$ Threshold 189
7.7.4 Calculation Time 192
7.7.5 $\boldsymbol{Kf'}$ Selection of Bragg Spots for Crystalline Samples 194
7.7.6 Aberration correction 196
7.7.7 LSS using Segmented detectors 198
7.8 Conclusion 199
Reference 201

Chapter 8: Conclusions and Future Work 202
8.1 Conclusions 202
8.2 Future Work 204
Reference 205

Appendix 1: WDD reconstructions 206

Chapter 1: Overview

Transmission electron microscope (TEM) is an important technique in materials science and biology and plays an important role in the fields of microstructure, morphology characterisation, crystal structure and chemical component analysis. The wavelength of the accelerated electrons used in TEM is shorter than visible light, which leads to a higher diffraction limit for high-resolution imaging and Modern TEM instruments can achieve sub-Ångstrom resolution.

Over nearly a century of development, TEM technology has expanded from E. Ruska and M. Knoll's first electron microscope [1-3] to various complex optical configurations and imaging modes. In addition to conventional parallel beam TEM imaging and scanning transmission electron microscope (STEM) imaging, electron ptychography [4], as a phase retrieval method, has recently been widely used. Compared to traditional TEM or STEM, electron ptychography can achieve high resolution [5, 6], compensate aberrations in lenses [7, 8] and, as a phase contrast method, is sensitive to light atoms [9-12].

Electron ptychography is a promising alternative method that demonstrates high fluence efficiency, making it an ideal candidate for low-fluence experiments [12-15]. In this thesis, I investigate and develop electron ptychography and related imaging methods to address two key challenges:

Firstly, how to improve the utilisation of electrons in imaging to reduce the electron fluence requirement. For this question, In Chapter 4, using inelastically scattered electrons in

ptychography is considered. In Chapter 5, a new method for centre-of-mass imaging is proposed to reduce the time required for precise focusing during data collection, aiming to make the most of every electron passing through the sample.

Second, how to increase the speed of data acquisition. To enhance data acquisition speed, in Chapters 6 and 7, improved non-iterative ptychography methods (AD-ptycho and LSS) are discussed. Experiments and simulations are conducted using segmented detectors, which are often faster than pixelated detectors for ptychography.

In the following chapters of this thesis, I will outline the context of electron ptychography and related imaging methods, discuss their limitations, and highlight these new developments. The structure of this thesis (from Chapter 2) is as follows:

- Chapter 2: Literature review. This chapter provides a general introduction to phase contrast in transmission electron microscope and related phase contrast imaging methods.
- Chapter 3: This chapter describes algorithms used for non-iterative ptychography and other related methods. This chapter is focused on the background theory of electron ptychography, including the source of information, as well conventional single-side band (SSB), Wigner-distribution deconvolution (WDD) and extended ptychographic iterative engine (ePIE) algorithms. Other imaging algorithms related to this thesis, for example Centre-of-Mass (COM), are also introduced.
- Chapter 4: The influence of inelastically scattered electrons on electron ptychography is discussed in this chapter. The effect of inelastic scattered electron on ptychography reconstructions is experimentally investigated together with theoretical calculations. The aim of this study is to utilise as many electrons, including inelastically scattered electrons, as possible in ptychography to increase fluence efficiency.
- Chapter 5: A combined method of electron ptychography and centre-of-mass (COM) imaging is developed. The new method, SSBm-COM, can perform aberration correction based on the traditional COM method. Furthermore, the relationship in terms of information transfer between ptychography and COM is discussed.

- Chapter 6: A framework for ptychography using an arbitrary detector (AD-ptycho) is described. This method can be used on segmented detectors but relies on the weak phase object approximation (WPOA). Experiments and simulations are performed to test the viability of aberration correction on segmented detectors.
- Chapter 7: A new method based on linear system solution (LSS), is derived without using WPOA as a basis for AD-ptycho. The algorithm parameter ($\boldsymbol{K}_f'$ condition) is discussed and the cost of computation time is compared.
- Chapter 8: Conclusions of this thesis and potential further developments in the future.

Reference

[1] M. Knoll and E. Ruska, "Das elektronenmikroskop," Zeitschrift für physik, vol. 78, pp. 318-339, 1932.

[2] M. Knoll and E. Ruska, "Beitrag zur geometrischen Elektronenoptik. I," Annalen der Physik, vol. 404, no. 5, pp. 607-640, 1932.

[3] M. Knoll and E. Ruska, "Beitrag zur geometrischen Elektronenoptik. II," Annalen der Physik, vol. 404, no. 6, pp. 641-661, 1932.

[4] W. Hoppe, "Beugung im inhomogenen primärstrahlwellenfeld. i. prinzip einer phasenmessung von elektronenbeugungsinterferenzen," Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography, vol. 25, no. 4, pp. 495-501, 1969.

[5] Y. Jiang et al., "Electron ptychography of 2D materials to deep sub-ångström resolution," Nature, vol. 559, no. 7714, pp. 343-349, 2018.

[6] Z. Chen et al., "Electron ptychography achieves atomic-resolution limits set by lattice vibrations," Science, vol. 372, no. 6544, pp. 826-831, 2021.

[7] H. Yang et al., "Simultaneous atomic-resolution electron ptychography and Z-contrast imaging of light and heavy elements in complex nanostructures," Nature Communications, vol. 7, no. 1, p. 12532, 2016.

[8] J. G. Lozano, G. T. Martinez, L. Jin, P. D. Nellist, and P. G. Bruce, "Low-dose aberration-free imaging of Li-rich cathode materials at various states of charge using electron ptychography," Nano letters, vol. 18, no. 11, pp. 6850-6855, 2018.

[9] C. T. Putkunz et al., "Atom-scale ptychographic electron diffractive imaging of boron nitride cones," Physical Review Letters, vol. 108, no. 7, p. 073901, 2012.

[10] P. Wang, F. Zhang, S. Gao, M. Zhang, and A. I. Kirkland, "Electron ptychographic diffractive imaging of boron atoms in LaB6 crystals," Scientific Reports, vol. 7, no. 1, p. 2857, 2017.

[11] Z. Ding et al., "Three-dimensional electron ptychography of organic–inorganic hybrid nanostructures," Nature Communications, vol. 13, no. 1, p. 4787, 2022.

[12] L. Zhou et al., "Low-dose phase retrieval of biological specimens using cryo-electron ptychography," Nature communications, vol. 11, no. 1, p. 2773, 2020.

[13] B. Küçükoğlu et al., "Low-dose cryo-electron ptychography of proteins at sub-nanometer resolution," Nature Communications, vol. 15, no. 1, p. 8062, 2024.

[14] Z. Chen et al., "Mixed-state electron ptychography enables sub-angstrom resolution imaging with picometer precision at low dose," Nature communications, vol. 11, no. 1, p. 2994, 2020.

[15] J. Song et al., "Atomic resolution defocused electron ptychography at low dose with a fast, direct electron detector," Scientific reports, vol. 9, no. 1, p. 3919, 2019.

Chapter 2: Literature Review

2.1 Introduction

This chapter is a review chapter, providing a brief introduction to the development of TEM, a mathematical description of scattering in TEM samples and high-resolution imaging modes, and an overview of common phase-contrast imaging methods. It provides the background necessary for the subsequent discussions on specific algorithms and new developments of phase-contrast imaging methods in the following chapters.

2.1.1 A Brief introduction to TEM

The transmission electron microscope (TEM) is commonly used as a characterisation instrument for organic and inorganic materials. In contrast to optical microscopes, the TEM uses accelerated electrons instead of visible light as the probe to illuminate a sample and form an image. Due to wave-particle duality, an accelerated electron beam shows similar behaviour to light, such as interference and diffraction and the wavelength of accelerated electrons can be calculated from the De Broglie equation. In 1920s, it was proved that a rotationally symmetric magnetic field can be used as a lens to focus electrons [1]. Following that, a magnetic lens for controlling accelerated electron beams was built and became the foundation of the invention of TEM in 1932 [2-4]. For nearly a century after that, until today, electromagnetic lenses remain one of the key components in modern TEMs.

In a modern TEM, electrons are usually accelerated to 60~300 kV. The accelerated

electrons have wavelength of several picometers (pm), which is far shorter than visible light and X-rays. This short wavelength naturally leads to a higher resolution. For thin samples, phase contrast [5, 6] was used in TEM. However, the resolution of phase contrast TEM imaging was limited by aberration of the lenses. In the late 1940s, electron holography [7, 8] was proposed as a solution to reconstruct high resolution sample information and compensate the influence of aberrations. In the 1970s, about 40 years after the scanning transmission electron microscope (STEM) was invented in 1930s [9], it was improved [10] with a field emission electron gun and better lenses. In the late 1990s and 2000s, the development of aberration correctors significantly improved the imaging resolution of TEM and STEM, as a result aberration-corrected TEM achieve sub-Angstrom resolution [11-15].

Sample preparation techniques are crucial for transmission electron microscopy imaging. Obtaining high-quality electron-transparent samples is an important aspect of all experiments. For this reason, many new sample preparation techniques have been developed from the mid-20th century to the present [16]. For bulk samples, besides traditional hand-polished sample, another common sample preparation technique is the use of Focused Ion Beam (FIB) technology [17, 18]. For soft samples, the development of resin embedding techniques [19] and rapid freezing as used in Cryogenic Transmission Electron Microscopy (Cryo-EM) [20, 21] allows the three-dimensional structure of biological samples to be preserved.

After nearly a century of development, the TEM has demonstrated significant applications in the field of microscopic characterisation. However, in specific areas, such as imaging radiation sensitive samples, it is foreseeable that TEM technology will continue to advance towards higher spatial resolution, lower electron fluence, and increased data acquisition speed.

2.1.2 TEM and STEM Imaging

The interaction of an electron beam and the sample in TEM can be separated into 5 steps:

1. An incident electron beam is generated, accelerated, and propagated to the sample.

2. The electron beam passes through the sample and is modulated by the sample, by which information about the sample is carried by the electron beam.
3. Subsequently, the electron beam propagates a distance through lenses.
4. Then the intensity of the electron beam is recorded by detector(s).
5. By analysing the recorded intensity distribution of the electron beam, information about the sample can be retrieved.

These are 5 basic steps involved in TEM image formation and interpretation.

The 1st step, it highlights one fundamental difference between TEM and STEM modes. In TEM mode, the electron beam is typically parallel or nearly parallel as it interacts with the sample. In contrast, in STEM mode, the electron beam is focused into a small convergent probe before interacting with the sample. Instead of using a static probe as in TEM mode, STEM mode uses scanning coils to control the probe, allowing it to scan across the sample.

For the 2nd step, a brief mathematical description of the samples and their interactions with the electron beam will be provided in Section 2.2, with more discussions in Section 3.2 of Chapter 3.

For the 3rd step, the influence of lenses, including aberrations on high-resolution TEM imaging will be discussed in Section 2.3. For STEM mode, the relationship between sample information and diffraction patterns will be discussed in Sections 3.2 and 3.3 of Chapter 3.

For the 4th step, in TEM mode, the electron detector is positioned at the imaging plane to record the intensity distribution of the electron beam. In contrast, in STEM mode, the detector is located at the back focal plane (diffraction plane). Traditionally, one or more annular detectors are used to measure the total intensity within their detection areas. The STEM image is constructed by mapping these intensity values at each scanning position of the probe. In this thesis, the influence of inelastically scattered electrons on diffraction patterns will be discussed in Chapter 4. In chapter 6 and 7, developments of imaging methods that utilise segmented detectors will be discussed.

The 5th step covers all the algorithms used to reconstruct sample information from raw

image or diffraction data. In Section 2.4, several phase contrast imaging methods will be introduced and compared. Chapters 3 (Sections 3.4 to 3.7) provide more detailed mathematical descriptions of some of these methods. Chapters 5 to 7 discuss new developments based on these imaging methods.

2.2 The Phase Object and Weak Phase Object Approximations

Before discussing specific TEM imaging methods, this section will briefly introduce the Weak Phase Object Approximation (WPOA). This approximation is one that describes samples and their interaction with the electron beam. A more detailed description is given in Chapter 3.

In TEM or STEM, a thin sample can be modelled as a 2-dimensional complex function, $\psi(\boldsymbol{R})$, with respect to two spatial coordinates, known as the transmission function or object function. The 2-dimensional model implies that the sample is thin enough to neglect the propagation of the electron beam within it. The amplitude is related to the absorption of electrons, while the phase of the transmission function corresponds to the phase shift of the electron beam after its interaction with the sample.

The phase shift, according to the Aharonov–Bohm effect [22], is related to the electric or magnetic potential of the specimen. This implies that the distribution of the electric and magnetic fields can be quantified using phase-contrast imaging methods [23, 24]. For non-magnetic samples, the projected electric potential of the sample $V_p(\boldsymbol{R})$ is the main factor affecting the phase. A more detailed discussion about the projected electric potential and the interaction parameter (σ) will be provided in Chapter 3. In this chapter, the conclusion that the phase of the transmission function, $\angle\psi(\boldsymbol{R})$, is proportional to the projected electric potential $V_p(\boldsymbol{R})$ is used, as shown in Equation (2-1).

$$\angle\psi(\boldsymbol{R}) = \sigma V_p(\boldsymbol{R}) \tag{2-1}$$

If the amplitude of the transmission function is constant, indicating no electron absorption,

the sample can be treated as a phase object. This approximation is known as the Phase Object Approximation (POA) [25], where $\psi(\boldsymbol{R})$ is expressed as an exponential function of an imaginary function $\psi_s(\boldsymbol{R})$, as shown in Equation (2-2). Furthermore, under the Weak Phase Object Approximation (WPOA) [25], $exp[\psi_s(\boldsymbol{R})]$ is expanded using the Taylor series, assuming only small phase shifts and hance considering only the first two terms, resulting in $1+\psi_s(\boldsymbol{R})$, as shown in Equation (2-2).

$$\psi(\boldsymbol{R}) \xrightarrow{POA} exp[\psi_s(\boldsymbol{R})] = exp\left[-i\sigma V_p(\boldsymbol{R})\right] \xrightarrow{WPOA} 1+\psi_s(\boldsymbol{R}) = 1-i\sigma V_p(\boldsymbol{R})$$
$$Where\ \psi_s(\boldsymbol{R}) = -i\sigma V_p(\boldsymbol{R}), is\ an\ imaginary\ function \tag{2-2}$$

When an incident beam passes through the sample, it can be represented as the wave function of the incident beam, $P(\boldsymbol{R})$, multiplied by the sample transmission function, as shown in Equation (2-3).

$$EW(\boldsymbol{R}) = P(\boldsymbol{R}) \cdot \psi(\boldsymbol{R}) \tag{2-3}$$

In Equation (2-3), $EW(\boldsymbol{R})$ represents the exit wave, describing the electron beam after its interaction with the transmission function but before any propagation. The transmission function $\psi(\boldsymbol{R})$ can be replaced by either the POA or WPOA transmission function from Equation (2-2), depending on the sample. The phase information of the sample is then carried by the exit wave. In the next section, the WPOA transmission function from Equation (2-3) is used to further discuss phase contrast in TEM.

2.3 Theoretical Background of Phase Contrast in TEM

Phase contrast in TEM arises from the changes in the phase of exit wave after it interacts with the sample. However, detectors cannot directly capture phase information; they only record the intensity of the electron beam, which is the square of its amplitude. This loss of phase information is known as the "phase problem" [26]. In High-Resolution TEM (HRTEM),

however, the imaging process translates some of this phase information into intensity variations, allowing detectors to indirectly capture it. In this section, the mechanism of this conversion in HRTEM under the WPOA will be explained. In addition, the effects of aberrations on HRTEM imaging with be briefly discussed.

2.3.1 The Transfer Function for HRTEM

The transfer function describes how effectively different spatial frequencies in the exit wave are transmitted to the image. In HRTEM, the transfer function $H_t(\boldsymbol{K})$, is expressed as the product of an aperture function $A(\boldsymbol{K})$ and an envelope function $E(\boldsymbol{K})$, as in Equation (2-4).

$$H_t(\boldsymbol{K}) = A(\boldsymbol{K}) \cdot E(\boldsymbol{K}) \tag{2-4}$$

In Equation (2-4), the aperture function $A(\boldsymbol{K})$ describes the aberrations in the imaging system through by a function $\chi(\boldsymbol{K})$, as shown in Equation (2-5). In this section, only defocus (C_1) and spherical aberration (C_3) are considered. Aberrations in this thesis follow Krivanek's notation [13], though equivalent notations are also available [27, 28].

$$\begin{aligned} \chi(\boldsymbol{K}) &= \pi C_1 \lambda \boldsymbol{K}^2 + \frac{1}{2}\pi C_3 \lambda^3 \boldsymbol{K}^4 \\ A(\boldsymbol{K}) &= A_t(\boldsymbol{K}) \cdot exp[i\chi(\boldsymbol{K})] \end{aligned} \tag{2-5}$$

In Equation (2-5), λ is the wavelength of the incident electron beam. $A_t(\boldsymbol{K})$ is a real-valued function, with a cut-off determined by the size of the objective aperture. The function $\chi(\boldsymbol{K})$can be expanded to include higher-order aberrations [28]. In Chapter 3, this function in a simplified form with be revisited.

The envelope function $E(\boldsymbol{K})$ in Equation (2-4) is associated with partial spatial and temporal coherence [29]. As a real-valued damping function, it decreases to zero, thereby establishing an information limit in the frequency domain. Frequencies beyond this limit are filtered out by the envelope function. A simplified expression of $E(\boldsymbol{K})$ is given by [29]:

$$E(\boldsymbol{K}) = \exp\left[-\frac{1}{2}(\pi\lambda\delta_c)^2\boldsymbol{K}^4 - \left(\frac{\pi\alpha_s}{\lambda}\right)^2(C_3\lambda^3\boldsymbol{K}^3 + \lambda C_1\boldsymbol{K})^2\right]$$
$$where: \delta_c = C_c\left[4\left(\frac{\Delta I_{obj}}{I_{obj}}\right)^2 + \left(\frac{\Delta E_s}{V_0}\right)^2 + \left(\frac{\Delta V_0}{V_0}\right)^2\right]^{\frac{1}{2}} \quad (2\text{-}6)$$

In Equation (2-6), δ_c is a term related to chromatic aberration. C_c represents the chromatic aberration coefficient, $\frac{\Delta I_{obj}}{I_{obj}}$ and $\frac{\Delta V_0}{V_0}$ represents instabilities in the objective lens current and acceleration voltage, respectively. $\frac{\Delta E_s}{V_0}$ denotes the energy spread of the electron gun. α_s is the angular spread of the electron beam, typically around 0.1 mrad for field emission guns [29].

In HRTEM, $P(\boldsymbol{R})$ represents a plane wave, so in Equation (2-3), the exit wave is equivalent to the transmission function. This exit wave then propagates (via a Fourier transform) to the back focal plane (or diffraction plane), where it is multiplied by $H_t(\boldsymbol{K})$, incorporating the effects of the aperture, aberrations, and the envelope function. The wave function then further propagates (via an inverse Fourier transform) to the image plane, where it is recorded as an intensity by a detector. The effect of the imaging system is characterized by Equation (2-7).

$$W(\boldsymbol{R}) = \mathcal{F}^{-1}\{\mathcal{F}\{\psi(\boldsymbol{R})\} \cdot H_t(\boldsymbol{K})\} = \psi(\boldsymbol{R}) \otimes h_t(\boldsymbol{R}) \quad (2\text{-}7)$$

In Equation (2-7), $\mathcal{F}$ and $\mathcal{F}^{-1}$ represent a Fourier transform and inverse Fourier transform, respectively. $\otimes$ represents convolution. $h_t(\boldsymbol{R})$, the point spread function, is the inverse Fourier transform of $H_t(\boldsymbol{K})$. When the WPOA condition from Equation (2-2) is assumed, Equation (2-7) can be written as [25]:

$$W(\boldsymbol{R}) = 1 - i\sigma V_p(\boldsymbol{R}) \otimes \mathcal{F}^{-1}\{H_t(\boldsymbol{K})\}(\boldsymbol{R}) \quad (2\text{-}8)$$

The intensity distribution on the detector, $I(\boldsymbol{R})$, can then be calculated as [25]:

$$I(\boldsymbol{R}) = |W(\boldsymbol{R})|^2 = 1 + 2\sigma V_p(\boldsymbol{R}) \otimes \mathcal{F}^{-1}\{imag[H_t(\boldsymbol{K})]\}(\boldsymbol{R})$$

$$imag[H_t(\boldsymbol{K})] = A_t(\boldsymbol{K})\, sin[\chi(\boldsymbol{K})]\, E(\boldsymbol{K}) \tag{2-9}$$
$$= E(\boldsymbol{K}) A_t(\boldsymbol{K})\, sin\left[\pi C_1 \lambda \boldsymbol{K}^2 + \frac{1}{2}\pi C_3 \lambda^3 \boldsymbol{K}^4\right]$$

In Equation (2-9), the $imag[H_t(\boldsymbol{K})]$ denotes the imaginary part of $H_t(\boldsymbol{K})$, representing the phase contrast transfer function (pCTF). Unlike $H_t(\boldsymbol{K})$, the pCTF, $imag[H_t(\boldsymbol{K})]$, specifically acts on phase information, $2\sigma V_p(\boldsymbol{R})$, and thus reflects only the transfer of phase information. As shown in Equation (2-9), the modified phase information $V_p(\boldsymbol{R})$ is now represented within the intensity distribution $I(\boldsymbol{R})$.

To visualize the pCTF, it is assumed that the cut-off frequency of $A_t(\boldsymbol{K})$ is sufficiently high. In this case, $sin[\chi(\boldsymbol{K})]\, E(\boldsymbol{K})$ governs the frequency transfer for the phase information from the projected potential $V_p(\boldsymbol{R})$.

The sine function in pCTF introduces oscillations as $\boldsymbol{K}$ varies. To illustrate this, an example of pCTF curve is calculated and shown in Figure 2-1. This curve is based on a 300 kV accelerating voltage, zero defocus ($C_1 = 0$), and an uncorrected spherical aberration (C_3) of 1 mm. The envelope function incorporates chromatic aberration ($C_c = 1.4\, mm$), an energy spread ΔE of 0.15 eV, $\frac{\Delta I_{obj}}{I_{obj}} = 2 \times 10^{-7}$, $\frac{\Delta V_0}{V_0} = 8 \times 10^{-8}$, and α_s of 0.1 mrad.

In Figure 2-1, information transfer in HRTEM under the WPOA condition is governed by two characteristic frequencies. The first, as discussed in the previous section, is the information limit, defined as the frequency at which the envelope function $E(\boldsymbol{K})$ approaches zero. Beyond this point, any information in V_p or ψ_s is lost. The second key frequency is the first zero of $sin[\chi(\boldsymbol{K})]$, which defines the point or interpretable resolution limit, as beyond this frequency, the contrast of transferred information exhibits reversals. As shown in Figure 2-1, the resolution is not the only zero of $sin[\chi(\boldsymbol{K})]$, this implies that contrast can invert multiple times at higher frequencies, making information beyond the resolution increasingly difficult to interpret.

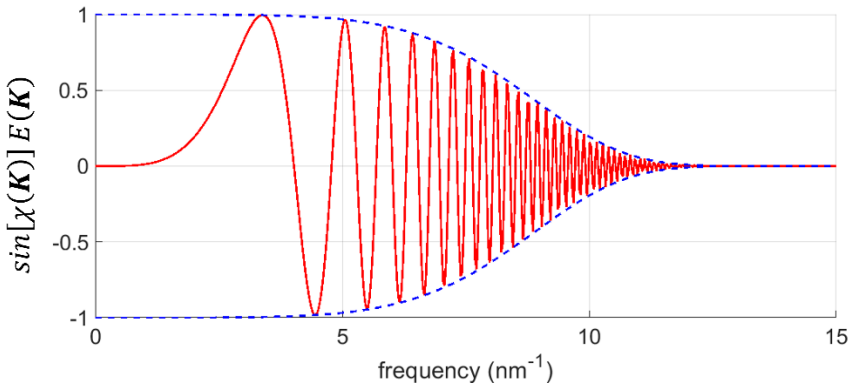

Figure 2-1 | pCTF curve $sin[\chi(\boldsymbol{K})]\,E(\boldsymbol{K})$ (red solid line) and envelop function $E(\boldsymbol{K})$ (blue dashed line), zero defocus and a spherical aberration (C_3) of 1 mm. Calculated based on Equation (2-5), (2-6) and (2-9). Acceleration voltage: 300 kV, chromatic aberration (C_c): 1.4 mm, energy spread (ΔE): 0.15 eV, $\frac{\Delta I_{obj}}{I_{obj}} = 2 \times 10^{-7}$, $\frac{\Delta V_0}{V_0} = 8 \times 10^{-8}$, $\alpha_s = 0.1\,mrad$.

2.3.2 The Scherzer Defocus

If only defocus and spherical aberration are considered, the defocus value that maximizes the frequency range between zero and the point resolution limit (the first zero of $sin[\chi(\boldsymbol{K})]$) is known as the “Scherzer defocus” [30], denoted as Δf_{Sch} in Equation (2-10). At the Scherzer defocus, HRTEM theoretically achieves its highest directly interpretable resolution for a given acceleration voltage and spherical aberration.

$$\Delta f_{Sch} = -(\frac{4}{3}C_3\lambda)^{\frac{1}{2}} \tag{2-10}$$

For the conditions used in Figure 2-1, the "Scherzer defocus" is approximately -51 nm. With this Scherzer defocus applied to Equation (2-9), an updated pCTF is calculated, as shown in Figure 2-2. Comparing Figures 2-1 and 2-2, the first zero of $sin[\chi(\boldsymbol{K})]\,E(\boldsymbol{K})$ shifts from 4.02 nm^{-1} to 5.14 nm^{-1}, indicating a resolution improvement from approximately 0.25 nm (2.5 Å) to 0.19 nm (1.9 Å).

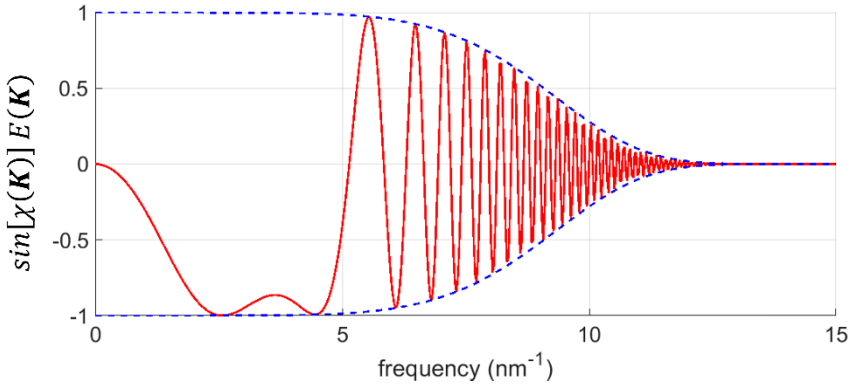

Figure 2-2 | pCTF curve $sin[\chi(\boldsymbol{K})]\,E(\boldsymbol{K})$ (red solid line) and envelop function $E(\boldsymbol{K})$ (blue dashed line), with "Scherzer defocus" (-51 nm) and spherical aberration (C_3) of 1 mm. Calculated based on Equation (2-5), (2-6) and (2-9). Acceleration voltage: 300 kV, chromatic aberration (C_c): 1.4 mm, energy spread (ΔE): 0.15 eV, $\frac{\Delta I_{obj}}{I_{obj}} = 2 \times 10^{-7}$, $\frac{\Delta V_0}{V_0} = 8 \times 10^{-8}$, $\alpha_s = 0.1\,mrad$.

However, even with the application of the "Scherzer defocus," the resolution remains well below the limit set by the envelope function due to the oscillations. In the next section, a scenario with the spherical aberration corrected will be discussed.

2.3.3 The Transfer Function with Aberration Corrector

The use of aberration correctors enables the correction of spherical aberration to a value of zero [31]. However, in aberration corrected TEM, spherical aberration usually is not corrected to zero. A spherical aberration in several micrometre allows for the optimization with defocus of phase contrast transfer across a broader range of spatial frequencies [32]. As shown in Figure 2-3, the pCTF curves for a spherical aberration of 1 µm are plotted. In Figure 2-3 (a), under an in-focus condition, there are no contrast reversals in the pCTF curve. The resolution is now limited by the information limit. However, at low frequencies, the transfer is still very low. Information is only transferred by $sin[\chi(\boldsymbol{K})]\,E(\boldsymbol{K})$ between approximately 5 nm^{-1} and 22 nm^{-1}, corresponding to feature sizes from 0.2 nm to 0.045 nm. This indicates that low-frequency (below 5 nm^{-1}) information is filtered by $sin[\chi(\boldsymbol{K})]$.

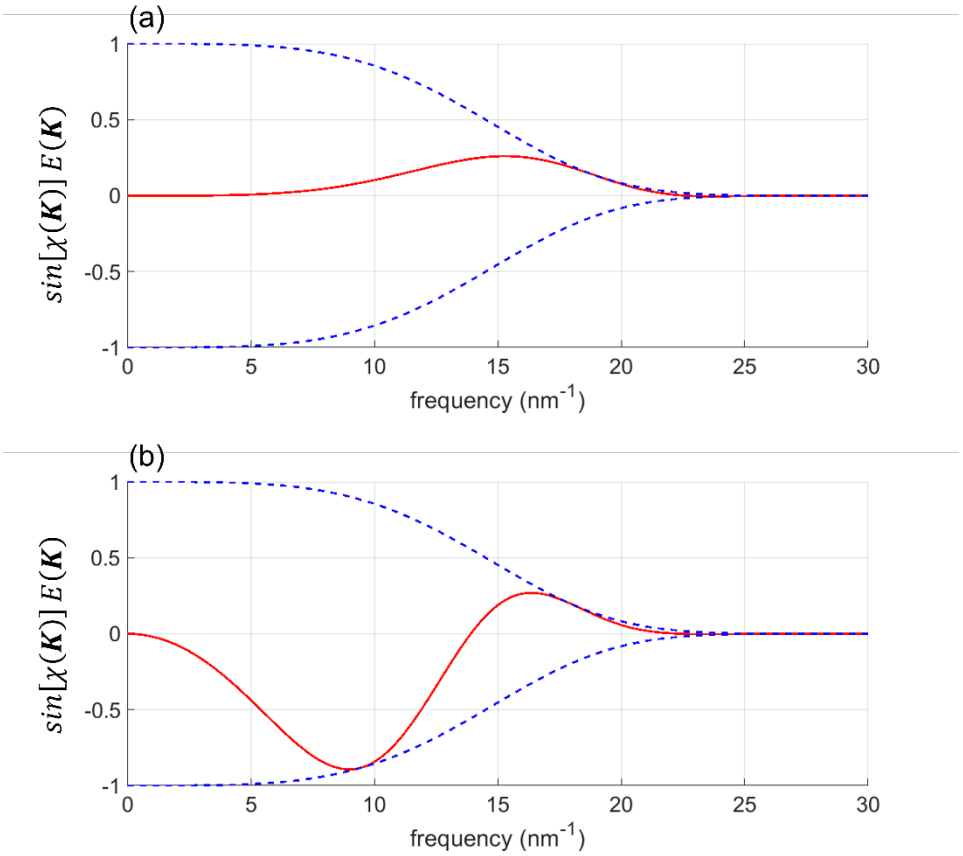

Figure 2-3 | pCTF curve $sin[\chi(\boldsymbol{K})]\,E(\boldsymbol{K})$ (red solid line) and envelop function $E(\boldsymbol{K})$ (blue dashed line), with spherical aberration (C_3) of 1 μm, (a) in-defocus, (b) defocus of -3 nm. Acceleration voltage: 300 kV, chromatic aberration (C_c): 1.4 mm, energy spread (ΔE): 0.15 eV, $\frac{\Delta I_{obj}}{I_{obj}} = 2 \times 10^{-7}$, $\frac{\Delta V_0}{V_0} = 8 \times 10^{-8}$, $\alpha_s = 0.1\ mrad$.

Experimentally, an ideal transfer function should uniformly transmit information across the widest possible frequency range. To enhance the transmission of low-frequency information under such conditions, defocus can be introduced into aberration-corrected HRTEM, as shown in Figure 2-3 (b). However, this comes at the cost of resolution, preventing it from reaching the information limit. For biological samples, defocus values of 2 to 4 micrometres are commonly used to improve low-frequency transfer (less than 5 nm^{-1} in Figure 2-3) , with a consequent trade-off in resolution [33].

To achieve uniform phase information transfer, various phase contrast imaging methods have been proposed and applied in TEM with different beam settings. In the next section,

the most common phase contrast imaging techniques in TEM are introduced and compared across several aspects, including information transfer, low-fluence tolerance, and imaging speed.

2.4 Phase Contrast Imaging methods

In the previous section, the “phase problem” and generation of phase contrast in HRTEM was discussed. However, conventional HRTEM is limited by its pCTF, which makes it challenging to achieve a uniform transfer of phase information from low frequencies to the information limit. Since the concept of phase contrast was introduced in the 1940s, numerous alternative imaging methods have been proposed to address this issue. In this section, various phase contrast imaging methods, focusing on the following key factors, are introduced, and compared.

The first factor is the phase contrast transfer function, which describes the ability of an imaging system to transfer information across different frequencies of a sample. As discussed in the previous section, an ideal imaging system would transfer information uniformly across the broadest possible frequency range. The phase contrast transfer function thus represents the effective pass-band limit of the system. Specifically, the highest interpretable frequency in the imaging process determines the theoretical resolution, making it a crucial factor in imaging. However, the lowest transferred frequency can also be crucial in certain contexts. For instance, in large-scale biological samples, low-frequency information captures the overall shape and outline of structures such as cells or viruses [34].

The second factor to consider is low-fluence tolerance, or noise tolerance. Many samples are beam-sensitive, making low-fluence imaging conditions essential for imaging these samples. Consequently, various low-fluence imaging techniques have been developed for TEM [35]. In phase contrast imaging, low-fluence conditions typically result in noisier raw data, so the low-fluence tolerance is directly linked to a methods ability to treat noise in these datasets.

The third factor is imaging speed, which encompasses both data collection and processing times. Slow data collection requires greater mechanical stability in the specimen holder, as drift can negatively impact high-resolution data. Fast data processing, ideally in real-time, enhances the user experience by allowing easy parameter adjustments during the experiment. Fast data collection enables *in-situ* imaging of dynamic processes and makes experiments more time-efficient for researchers. Therefore, fast imaging speed in both data collection and processing is crucial for optimal usability.

The final factor is the potential for integration with other imaging and analysis techniques, such as HAADF, EDX, EELS and EFTEM. Combining phase contrast imaging with these methods (under TEM or STEM mode) allows for the simultaneous collection of diverse signals, providing an important advantage for sample analysis.

There are several other important factors to consider for an efficient phase imaging method, such as field of view (FOV), depth of field (DOF), the ability to obtain three-dimensional information, experimental complexity, reliability in quantification, and limitations imposed by instrument settings. In this section, while these additional factors are considered when necessary and phase imaging methods are primarily evaluated based on the factors discussed above.

2.4.1 HRTEM

HRTEM is the most widely used method for acquiring phase-contrast images in TEM. In Section 2.3, it was introduced from a theoretical perspective. From an experimental perspective, the key features of HRTEM (outlined in Section 2.1.2) are summarised in Table 2-1. HRTEM in modern TEMs does not require additional hardware, such as biprism required for off-axis holography. It can relatively easily reveal high resolution detail provided the specimen is properly prepared [29].

Table 2-1 | Key features of HRTEM

Beam convergence	Parallel beam
Raw data recorded on	Imaging plane
Data collection	Single exposure, no scanning
Area of illumination	nm^2 ~ μm^2, depend on sample
Detector	Pixelated detector
Resolution limit by	Aberration, information limit
Typical low-fluence limit	Can be less than 100 $e^-/Å^2$ [36, 37] (for imaging, not for Single Particle Analysis)

The brief theoretical background of HRTEM has been discussed is Section 2.3. In this section, the application and practical limitations of HRTEM will be discussed.

One limitation is low-frequency transfer. Since the phase-distortion function is based on a sine function, as shown in Equation (2-9), the transfer at zero frequency always tends toward zero. As a result, the transfer of low-frequency signals is inherently weak. This becomes a concern when imaging large-scale specimens with fields of view in hundreds of nanometre to micrometre [38]. To enhance low-frequency transfer, up to micrometre level defocus can be set to modify the phase contrast transfer function [39]. While the contrast transfer function can be corrected after data collection [40, 41], but duo to the oscillation in pCTF curves, as shown in Figure 2-1 and 2-2, information is still lost at frequencies where the pCTF equals zero. To address this limitation, a Zernike phase plate (ZPP) can be an effective solution. The advantages and disadvantages of ZPP will be discussed in the next section.

HRTEM is a traditional imaging method widely used for low-fluence imaging in both

conventional TEM [42, 43] and cryo-EM [44]. The fluence for HRTEM can vary significantly depending on the sample, ranging from less than 10 $e^-/Å^2$ [36, 37] to higher fluences for more robust samples. It is important to note that the imaging fluence in HRTEM differs from the average fluence used in Cryo-EM single particle analysis (SPA) [45]. In SPA, thousands of images are recorded and averaged for particle reconstruction, with the average fluence being used for reconstruction rather than for a single image.

The speed of HRTEM is constrained by exposure time and data storage. Since only 1 frame is recorded for imaging, HRTEM images can be captured very quickly.

HRTEM can be used in conjunction with an energy filter [46]. Energy-filtered TEM (EFTEM) can reach a resolution of 1 nm with information about chemical composition [47]. However, for thicker samples, inelastically scattered electrons can overwhelm the signals from elastic electrons, potentially obscuring the phase contrast. In such cases, selectively removing inelastically scattered electrons using an energy filter is important for high-resolution imaging [48]. An energy filter can be used to selectively remove inelastically scattered electrons, helping to restore signal. In addition, HRTEM can also be combined with tomography for 3D reconstruction [49], at room temperature [50] and under cryogenic conditions [51].

2.4.1.1 The Zernike Phase Plate

To achieve improved low frequency transfer in HRTEM, phase plates can be introduced to modify the contrast transfer function. The Zernike phase plate (ZPP) [52] is the most commonly used type for enhancing low-frequency transfer in TEM. In TEM, the ZPP is a semi-transparent aperture located at the back focal plane (diffraction plane) of the objective lens, featuring a central hole, as illustrated in Figure 2-4. The unscattered central beam passes through the hole, while the diffracted beams pass through the semi-transparent aperture, experiencing an additional $\frac{\pi}{2}$ phase shift. Considering only defocus and spherical aberration, the phase contrast transfer function (pCTF) can now be expressed as shown in Equation (2-11):

$$E(\boldsymbol{K})A_t(\boldsymbol{K})\,sin\left[\pi C_1\lambda\boldsymbol{K}^2+\frac{1}{2}\pi C_3\lambda^3\boldsymbol{K}^4+\frac{\pi}{2}\right] \quad (2\text{-}11)$$

$$when\ |\boldsymbol{K}|>0$$

The $\frac{\pi}{2}$ phase shift causes the phase contrast transfer function (pCTF) to approach 1 instead of 0 when the frequency approach to 0 (but information at 0 frequency is still not transferred), thereby boosting low-frequency information, as shown in Figure 2-5 [52]. The $\frac{\pi}{2}$ phase shift effectively converts the sine function from the imaginary part of $H(K)$ into a cosine form, and switches the transfer of the sample's amplitude and phase information [53]. Furthermore, when combined with aberration correctors, the Zernike phase plate (ZPP) can achieve a broad frequency transfer range, extending from low frequencies up to the information limit [54].

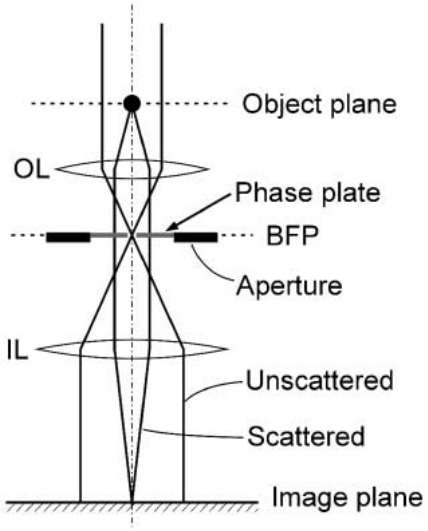

Figure 2-4 | Schematic setup for a Zernike phase plate (ZPP) in TEM [52]. The phase plate is placed at the back focal plane after objective lens. The rest of the beam path is the same as HRTEM. Reprinted with permission from [52]. Copyright 2001 by Elsevier Science B.V.

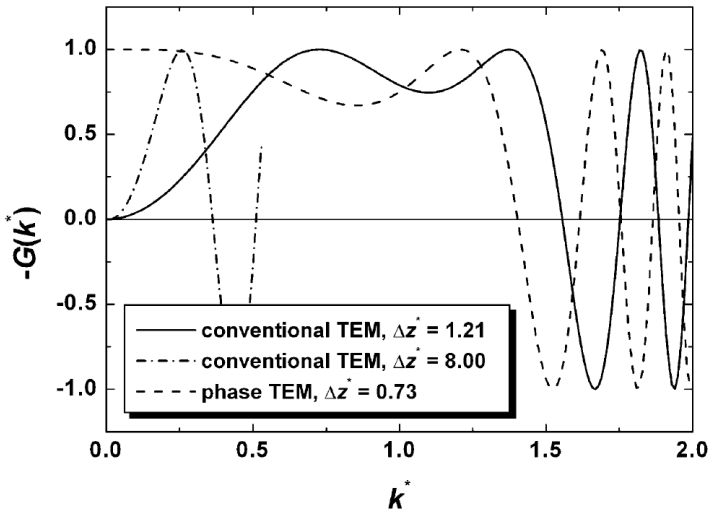

Figure 2-5 | pCTF ($-G(k^*)$, the vertical axis) for HRTEM (solid line and dash-dotted line) and ZPP TEM (dashed line) at 300 kV [52]. The horizontal axis $k^* = K(C_3\lambda^3)^{\frac{1}{4}}$, $\Delta z^* = \frac{C_1}{\sqrt{C_3\lambda}}$. Envelope function is not considered. The pCTF of ZPP TEM (dashed line) is at 1 rather than 0 as for conventional HRTEM (solid line and dash-dotted line). For HRTEM, achieving better low-frequency transfer (under $k^* = 0.5$) requires a sacrifice in resolution, as indicated by the dashed line in the figure. However, ZPP TEM can achieve a flat low-frequency transfer curve with only a minimal loss of resolution. Reprinted with permission from [52]. Copyright 2001 by Elsevier Science B.V.

The Zernike phase plate (ZPP) has been applied to biological samples [55-57], significantly enhancing the contrast transfer in low frequencies of weak-phase samples. However, in practice, the ZPP aperture at the back focal plane can be affected by several issues, including charging, contamination, and misalignment [52]. These factors limit the effectiveness of the ZPP in some experimental settings.

In addition to the ZPP, other phase plates have been used to modify the information transfer in HRTEM, including electrostatic (Boersch) phase plate [58], Zach phase plate [59], and hole-free Volta phase plates [60, 61]. These phase plates share the same basic principle of modifying the information transfer in a conventional HRTEM beam setup to improve contrast transfer. An alternative approach using TEM setting to solving the exit wave,

including phase information, is focal series reconstruction, which will be discussed in Section 2.4.4. In addition, there are phase contrast imaging methods that do not rely on the conventional TEM setup. The following section will introduce off-axis holography using a biprism in a modified TEM configuration.

2.4.2 Off-axis Holography

Holography was first proposed by Gabor in 1949 [8] with the original goal of resolving the phase of the exit wave without contrast oscillations. Over decades of development, it has expanded into a comprehensive field. Broadly speaking, holography now encompasses all methods aimed at reconstructing the complex exit-wave function [62].

Many schemes have been developed to achieve holography in both conventional TEM and STEM modes [62]. Holographic techniques in TEM are primarily categorized into two groups: in-line holography and off-axis holography. In-line holography involves methods that do not require interference with an external reference wave; examples include ptychography and focal series reconstruction, which will be discussed in subsequent sections. Off-axis holography, in contrast, uses a known reference wave to interfere with the exit wave, producing a hologram from which the exit wave can be reconstructed.

Among the various off-axis holography methods reported [62], the most widely used is imaging-plane off-axis holography with a Möllenstedt-Düker biprism [63], which is the method discussed in this section. In this approach, the biprism deflects both the exit wave and a reference wave to create interference fringes on the image plane [64], as illustrated in Figure 2-6. At the image plane, a pixelated camera records the intensity of the interference pattern formed by the exit wave and the reference wave. The amplitude and phase of the exit wave can then be reconstructed using specialised algorithms [65-67]. The phase reconstruction is linear to electrostatic potential if magnetic induction can be ignored [68].

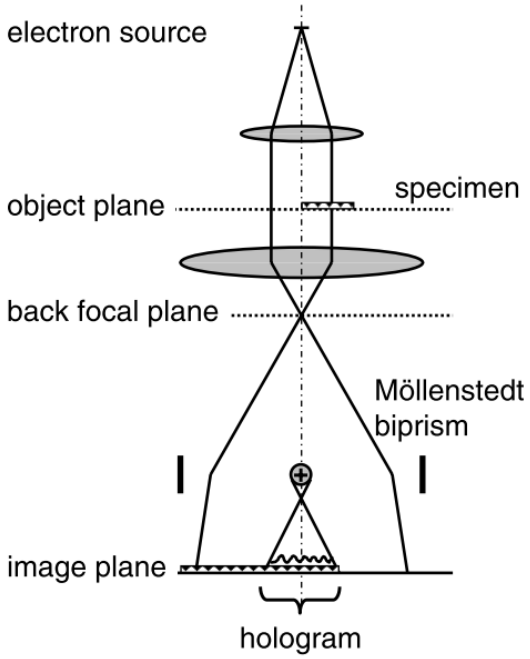

Figure 2-6 | Schematic for image-plane off-axis holography using a Möllenstedt-Düker biprism [53]. Half of the incident beam passes through the specimen, forming the exit wave, while the other half bypasses it, serving as a reference wave. In the back focal plane, the Möllenstedt-Düker biprism deflects both the reference wave and the exit wave, bringing them together to interfere and form holograms on the image plane. Reprinted with permission from [53]. Copyright 2007 by IOP Publishing.

The experimental key features used in off-axis holography with a Möllenstedt-Düker biprism are listed in Table 2-2.

Table 2-2| key features of Off-axis Biprism Holography

Beam convergence	Parallel beam
raw data recorded on	Imaging plane
Data collection	Single exposure, no scanning
Area of illumination	nm^2 ~ μm^2, depend on sample

Detector	Pixelated detector
Resolution limit by	Information limit, coherence of source, signal to noise ratio [69-71]
Additional condition	a Möllenstedt-Düker biprism to form a reference wave

Off-axis holography can also correct coherent lens aberrations [72, 73] , implying that, in theory, resolution is limited only by the information limit. Additionally, off-axis holography provides improved low-frequency transfer, extending to micrometre scales [53]. However, when capturing off-axis holograms, it is crucial to consider the spacing of the interference fringes, which is adjustable through the biprism's intensity, and ensure that these fringes are adequately sampled by the detector [53]. The interference fringes are formed by elastically scattered electrons and the unscattered reference beam, which intrinsically filtered the inelastically scattered electrons from the imaging process [74]. This will be discussed later in this section.

For a given signal-to-noise ratio, or the number of electrons if a perfect detector is considered, the detectable phase difference (phase detection limit) of off-axis holography can be calculated using Equation (2-12) [53].

$$\delta_{\psi lim} = \sqrt{2} \cdot \frac{SNR}{C \cdot \sqrt{N_{pixel}}} \tag{2-12}$$

Where SNR is an intended signal to noise ratio, C is the contrast of interference fringes and N_{pixel} is the average number of electrons counts per reconstructed pixel.

With the development of direct electron detectors [75], particularly counting-mode cameras, low-fluence electron off-axis holography has been reported [76, 77]. Further, applying sparse coding methods to the hologram [78] can help partially eliminate shot noise, thereby enhancing the signal-to-noise ratio.

In terms of imaging speed, off-axis holography requires additional time for biprism alignment compared to conventional TEM imaging. However, once the beam alignment is complete, data collection can proceed at speeds similar to conventional TEM. The data processing using software is generally fast and automated. in most cases, the operator simply selects points on the power spectrum for reconstruction, and the algorithm calculates results within seconds. Furthermore, if a time series of holograms is recorded, off-axis holography can also be applied to *in-situ* studies [79, 80].

In addition, an elliptical illumination in the direction perpendicular to the biprism (achieved by adjusting astigmatism) can result in an improvement of interpretable lateral resolution [81, 82].

Due to its unique instrumental setup, off-axis holography is challenging to combine with other imaging methods. However, exit wave function reconstructed by off-axis holography can effectively be considered an "energy-filtered" image of zero-loss electrons, owing to the introduction of the unscattered reference wave [53]. The reference wave can be assumed to have a single wavelength, allowing only zero-loss electrons in the exit wave to interfere with it [74]. Inelastically scattered electrons contribute only to the background of the hologram and are filtered out when non-zero order spots on the power spectrum are selected for reconstruction. This offers an advantage over HRTEM, particularly for imaging thicker samples.

2.4.3 Ptychography

Ptychography is technique based on coherent diffractive imaging (CDI) [83]. In CDI, the intensity distribution of diffraction patterns is recorded, and information about the specimen and probe is reconstructed by processing these patterns using analytical or iterative methods. Various algorithms have been developed for CDI to reconstruct phase information, including the Gerchberg-Saxton (G-S) algorithm [84, 85], the Yang-Gu algorithm [86, 87] and other algorithms that utilize gradient descent methods [88]. However, a significant limitation of CDI is that the reconstruction area is limited by the size of the

probe and the object support constrain [89].

Ptychography was first proposed by W. Hoppe [90] around the same time as CDI. The modern theories and methods of ptychography are largely based on Rodenburg's work [91-94]. In ptychography, a series of diffraction patterns are collected at various known illumination positions of the specimen. The phase information is embedded in the diffraction patterns through the interference between direct and diffraction discs, which will be discussed in detail in Chapter 3. The beam path of a convergent probe for electron ptychography, which is similar to STEM, is illustrated in Figure 2-7. The illuminated regions of these diffraction patterns partially overlap with adjacent illuminated regions. In a STEM, the probe is typically controlled by a scanning system, and diffraction patterns are recorded during the scanning process. Consequently, compared to CDI, the imaging area is no longer limited by the size of a single probe or the estimated size of object (support constrain) but by the scanning area, data storage capacity, and data collection speed.

In addition to ptychography using a scanning probe, an alternative approach involves tilting the illumination beam at different angles, a technique known as Fourier ptychography [95] or tilt-series reconstruction [96]. In Fourier ptychography, the exit wave is reconstructed from a series of tilted images, demonstrating the ability to achieve resolution beyond the conventional resolution limit of HRTEM. A resolution of 78 pm was reported under 200 kV [96].

However, this section will focus on literature related to the ptychography with a scanning probe. This is because the developments in ptychography methods discussed in the following chapters are based on scanning probe ptychography rather than Fourier ptychography.

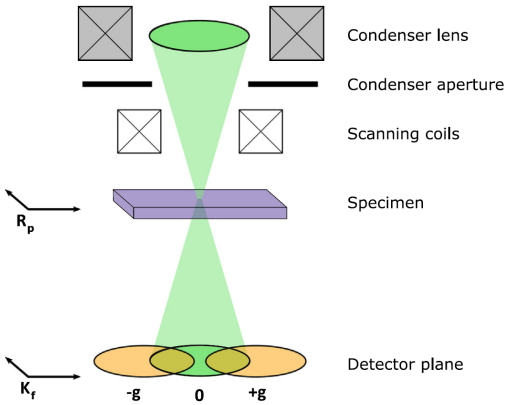

Figure 2-7 | Schematic beam path for convergent-beam ptychography (4DSTEM) with a scanning in-focus probe [97]. A convergent electron probe scans across the specimen (controlled by scanning coils), while a pixelated detector captures diffraction patterns corresponding to each probe position at Detector plane. The interference between direct beam and diffracted beams (at $-g$ and $+g$ frequency) contains phase information of the sample, which will be discussed in Chapter 3. Reprinted with permission from [97]. Copyright 2021 by Elsevier B.V.

The dataset used in ptychography reconstruction is typically referred to as a 4DSTEM dataset [98]. 4DSTEM combines STEM with a pixelated electron detector to record scanning convergent beam electron diffraction (CBED) patterns. The term "4D" comes from the fact that the scanning on the sample plane occurs in two dimensions, while the diffraction pattern adds another two dimensions. As a result, the dataset is considered four-dimensional. The detailed mathematical formulation of a 4DSTEM dataset will be discussed in Chapter 3.

4DSTEM datasets can be recorded under both in-focus and defocus conditions. For a focussed probe, it can be easily combined with traditional STEM imaging techniques. There are two methods for simultaneously obtaining STEM images while collecting ptychographic 4DSTEM datasets. The first method involves physically capturing HAADF or ADF images using a ring-shaped detector. However, this approach can create a shadow on the camera used to record the diffraction patterns, limiting the maximum diffraction angle that can be recorded. Another method is to use a virtual mask on 4DSTEM

datasets to integrate the electron signal within a ring-shaped area, generating a virtual ADF image from the diffraction patterns. This virtual ADF approach is commonly used in experimental datasets because it preserves the raw diffraction pattern data without any loss [99, 100]. However, it is important to note that if a defocused 4DSTEM dataset is collected, the simultaneous HAADF or ADF images will also be defocused.

In addition, Energy Dispersive X-ray Spectroscopy (EDS) can be integrated with 4DSTEM data collection, as EDS detectors do not interfere with the positioning of 4DSTEM detectors. However, for current atomic resolution ptychographic datasets, the pixel time for each diffraction pattern is typically very short (on the order of milliseconds), which may limit the ability of EDS detectors to capture sufficient X-ray signals.

For Electron Energy Loss Spectroscopy (EELS) data collection, the EELS spectrum is typically acquired from the bright-field disc, which is blocked by the camera used for 4DSTEM datasets. To resolve this issue, the concept of hollow electron ptychography has been proposed [101]. This approach shows that ptychography can reconstruct high-resolution phase images from incomplete bright-field discs, potentially enabling the simultaneous detection of both EELS and ptychography data.

After collecting the 4DSTEM data, the dataset needs to be processed using algorithms to reconstruct phase information. Various algorithms have been developed for ptychographic reconstruction [102], which can be categorized into iterative and non-iterative methods. Iterative methods involve a numerical optimization process, where the goal is to find sample and in some cases probe functions that satisfy the constraints of the collected diffraction patterns (in reciprocal space) and the consistency of the overlapping illuminated areas (in real space). In contrast, non-iterative methods rely on analytical approaches that do not use numerical optimisation. Instead, they use the mathematical characteristics and approximations of the scanning diffraction datasets to directly solve for the exit wave.

Detailed algorithms of non-iterative and iterative ptychography will be introduced in Chapter 3. However, in the next two sections of this chapter, these methods with be

introduced and compared with other phase contrast imaging methods based on the key factors outlined in Section 2.4.

2.4.3.1 Non-iterative Methods

In the 1990s, non-iterative ptychography methods including single-side band (SSB) [92] and Wigner-distribution deconvolution (WDD) [91] were introduced. The WDD and SSB methods analytically reconstruct phase information from 4DSTEM datasets, using detailed algorithms to be introduced in Sections 3.4 and 3.5, Chapter 3. Non-iterative ptychography methods have been applied a variety of specimens. For example, Pennycook *et al*. demonstrated non-iterative ptychography on bilayer graphene, comparing them with conventional STEM imaging methods [100]. Hao *et al*. used WDD to image structures of both light and heavy elements and explored information along the Z-axis [99]. Some of the experimental features of non-iterative ptychography are listed in Table 2-3.

Table 2-3 | Key Features of Non-iterative Ptychography

Beam convergence	Convergent beam
Raw data recorded in	Diffraction plane
Data collection	4DSTEM, beam scanning
Area of illumination	nm^2 ~ 100 nm^2, limited by size of dataset
Detector	Pixelated detector
Resolution limit by	Convergence angle (2α limit) for SSB
Typical low-fluence limit	around 200 $e^-/Å^2$ [97, 103]
Additional features	Post-acquisition aberration correction Defocus probe ptychography

The transfer function for SSB method can be analytically calculated [100], with the convergence angle and aberrations being the primary factors influencing the CTF [104]. The theoretical resolution for both SSB and WDD is given by $\frac{\lambda}{2\alpha}$, where λ is the wavelength and α is the convergence semi-angle. O'Leary *et al*. [97] also discuss the influence of convergence angle on the function for SSB and demonstrate the transfer function and reconstruction for different convergence angles, as shown in Figure 2-4. Furthermore, they explore the influence of weighted CTFs on reconstructions. As for WDD, it is theoretically possible for all the information collected in the diffraction plane to contribute to the reconstruction, allowing the resolution to exceed the 2α limit [91, 105].

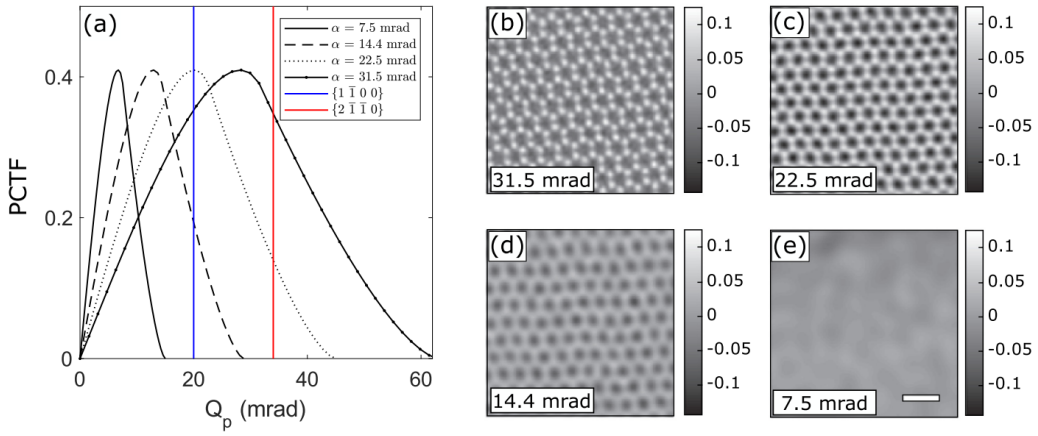

Figure 2-8 | Transfer function (TF) and reconstructions of a graphene sample using SSB at 80 kV under in-focus probe and different convergence angle α [97]. (a) Transfer function for different convergence angle (b-e) corresponding phase reconstructions. Scale bar on (e) is 5 Å. The TF of SSB begins at the origin and extends up to twice the convergence angle, indicating that the resolution of SSB is limited by 2α. The reconstructions shown in (b-e) demonstrate this resolution degradation as the convergence angle decreases from 31.5 mrad to 7.5 mrad. Reprinted with permission from [97]. Copyright 2021 by Elsevier B.V.

Non-iterative ptychography methods can retrieve phase information at low fluence. Simulations for both inorganic [106, 107] and organic [104, 107] specimens under low-

fluence conditions have been reported. Experimentally, non-iterative ptychography has been applied to zeolite, a typical sensitive specimen, with a fluence of 200 $e^-/Å^2$ at atomic resolution [97, 103]. Additionally, it has also been used to image sensitive Li-rich cathode materials [108].

The imaging speed of ptychographic data collection has traditionally been limited by the capabilities of the instruments. For a long time, the speed of 4D-STEM dataset collection was constrained by the readout speed and detection efficiency of pixelated detectors. However, from 2010s, fast direct detection electron cameras [109-111] have significantly accelerated data collection for 4D-STEM datasets. Consequently, a state-of-the-art TEM equipped with a fast electron camera can detect and record a diffraction pattern in milliseconds, enabling large-scale ptychographic data collection within an acceptable time. In terms of data processing speed, non-iterative ptychographic methods are generally faster than iterative methods due to their lower computational demands. With further improvements in algorithms, non-iterative methods can now be processed in real-time [112].

2.4.3.2 Iterative methods

Iterative ptychography methods shares many features with non-iterative methods, as shown in Table 2-4.

Rodenburg *et al*. [94] proposed the Ptychographic Iteration Engine (PIE) in 2004, and further developed the extended Ptychographic Iteration Engine (ePIE) in 2009 [113, 114], which is the algorithm discussed in this section. The flowchart for ePIE is shown in Figure 2-9 and the detailed algorithm will be discussed in Section 3.7 in Chapter 3. Based on PIE and ePIE, many other updates are also used based on PIE algorithm, for example, rPIE(mPIE) [115] and sasPIE [116]. Additionally, other iterative reconstruction algorithms are reported, for example, the conjugate gradient (CG) method [117],difference map (DM) reconstruction [118] and non-convex Bayesian optimization method [119].

Table 2-4 | Key Features of Iterative Ptychography

Beam convergence	Convergent beam
Raw data recorded in	Diffraction plane
Data collection	4DSTEM, beam scanning
Area of illumination	nm^2 ~ μm^2, limited by size of dataset
Detector	Pixelated detector
Resolution limit by	Collection angle
Typical low-fluence limit	Around 5 $e^-/Å^2$ in Cryo-EM to resolve features with size of about 5 nm [120] Information limit is related to fluence [121]
Additional features	Post-acquisition aberration correction Defocus probe ptychography

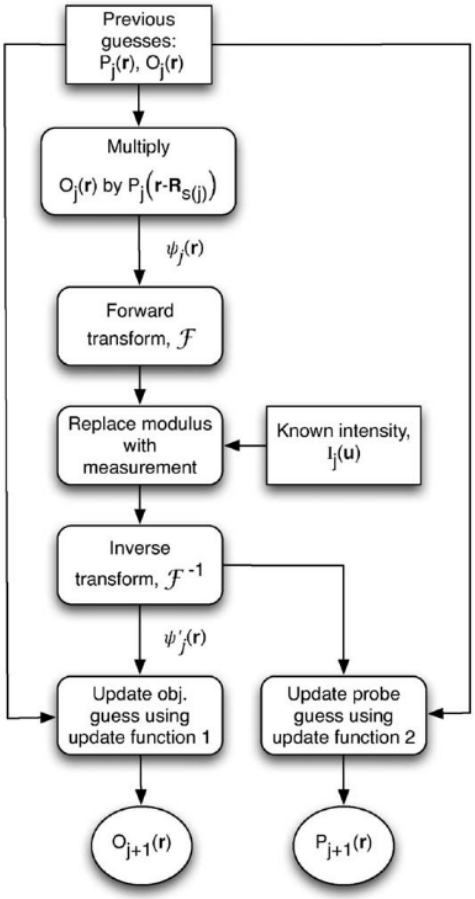

Figure 2-9 | Flowchart of the ePIE algorithm [113]. The detailed algorithm of ePIE will be introduced in Chapter 3. Reprinted with permission from [113]. Copyright 2009 by Elsevier B.V.

For ePIE, the transfer function can also be numerally calculated [97], as shown in Figure 2-10. For high-resolution reconstruction, ePIE has been experimentally proven to achieve super-resolution beyond the 2α limit. A spatial resolution of 0.39 Å under 80 kV has been achieved on 2D materials using ePIE [106]. Furthermore, Chen *et al*. reaches the atomic-resolution limits set by lattice vibrations using multi-slice electron ptychography [122]. ePIE also demonstrates good transfer for low-frequency signals, as shown for large-scale biological samples [120]. Through multiple scans of the same sample, ePIE has demonstrated the ability to achieve a broad range of information transfer from low to high frequencies by adjusting the convergence angle [123].

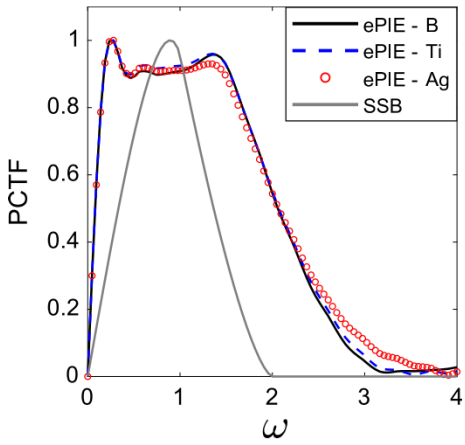

Figure 2-10 | Calculated transfer function for ePIE compared with SSB [97]. Curves were calculated using isolated atom (B, Ti, and Ag) simulation. $\omega = \frac{Q_p}{\alpha}$, Q_p is scattered angle in mrad and α is convergence angle in mrad. This calculation is at 200 kV with a convergence angle of 22.48 mrad. Reprinted with permission from [97]. Copyright 2021 by Elsevier B.V.

Under low-fluence conditions, ePIE has been applied for reconstructing high-resolution phase data across various sample types. For example, Song *et al*. obtained phase-contrast image of monolayer Molybdenum disulfide (MoS_2) using ePIE at a fluence of 403 $e^-/Å^2$, achieving a resolution of 1.58 Å [124]. For organic samples, Zhou *et al*. demonstrated that ePIE could operate at fluences as low as 5.7 $e^-/Å^2$ for imaging virus particles in cryo-EM, successfully resolving features approximately 5 nm in size [120]. Figure 2-11 [121] summarizes the information limit of ptychographic reconstruction under various fluence (dose) conditions, revealing a fluence-dependent information limit when the fluence is below 1000 $e^-/Å^2$. In addition, low-fluence ePIE reconstructions can be used for further SPA reconstructions [123, 125].

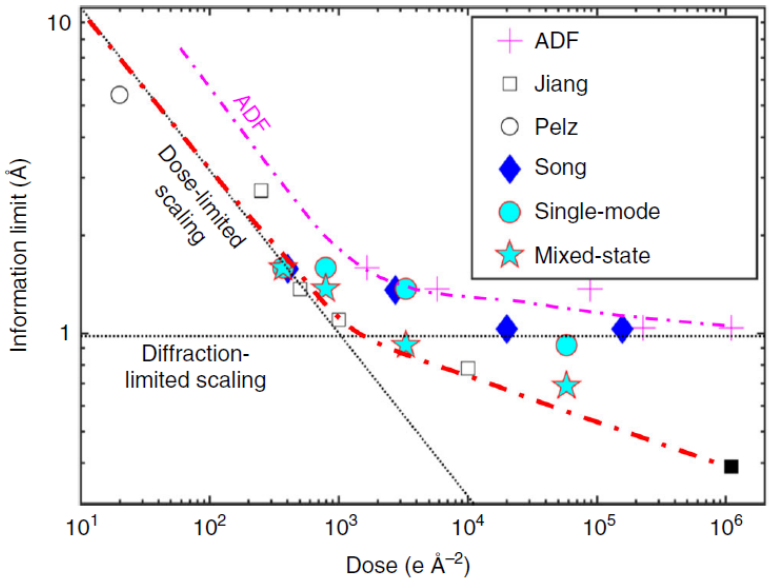

Figure 2-11 | Summary of dose dependence of information limit for ptychography [121]. The opened markers are electron ptychographic results using simulated data from references [106] by Jiang *et al*. and [119] by Pelz *et al*. The filled markers are experimental results by electron ptychography from [124] by Song *et al.* and [121] by Chen et al. The single-mode and mixed-state results are from reference [121]. Resolution by dose-limit, inversely proportional to square root of dose, and by aperture limit are guided by dashed lines. Information limit for ADF (plus makers) is measured from low-angle ADF images of monolayer WS_2 synthesized from in-focused 4D-STEM.

The data collection speed for iterative methods is similar to that for non-iterative methods since both rely on the same raw datasets. However, iterative methods typically require more processing time. For instance, ePIE generally needs 50 to 200 iterations for convergence of the reconstructed object function, with each iteration taking from a few seconds up to one or two hours, depending largely on the number of pixels and scanning points. Using a GPU instead of a CPU can improve the reconstruction speed for different map algorithms [126, 127]. For ePIE/PIE, using a GPU to execute fast Fourier transforms offers substantial speed advantages over using a CPU. Nevertheless, real-time processing for iterative methods continues to be challenging due to the high computational demands.

Beyond the four aspects discussed earlier, 3D imaging with ePIE is another important area.

The first approach to achieve 3D ptychography reconstruction is to combine ptychography and tomography, which has shown improved contrast in 3D reconstruction of organic-inorganic hybrid nanostructures [128]. The second approach is inverse multislice ptychography. Multislice [129, 130] is a widely used method in the simulation of TEM/STEM imaging. It models a thick specimen as a stack of single layers, where each layer is related to a section along the z axis of the sample. The interaction between a probe and the thick specimen is considered as a sequential process where the probe interacts with each layer in order. In electron ptychography, inverse multi-slice treats the transmission function of the specimen as multiple complex transmission functions along the beam direction. Inverse multi-slice ptychography can reconstruct phase-contrast images at different positions along the beam axis, providing information from the third dimension. This approach allows three-dimensional ptychography to analyse specimens at different heights along the beam axis [131] and reducing the effect of multiple scattering on the reconstruction [122]. Multiple scattering occurs when the incident beam interacts multiple times with the sample, affecting the accuracy of single slice ptychography. The depth resolution by applying the inverse multislice method can reach 3.9 nm by reducing the negative affect of the multiple scattering on resolution [122].

2.4.4 Focal Series Reconstruction

Focal series reconstruction reconstructs the exit wave as a complex function through focus variation of images recorded with plane wave illumination. Unlike HRTEM, this technique involves collecting a series of TEM images at different defocus. Experiments using focal series reconstruction based on maximum likelihood (MAL) algorithms were reported in the 1990s [132-134]. Later, iterative wave function reconstruction (IWFR) was proposed and compared with the MAL method [135]. Focal series reconstruction can achieve a resolution of 0.5 Å with a monochromator on a 300kV TEM [136] and is also capable of measuring aberration coefficients (defocus and two-fold astigmatism) [137]. In Table 2-5, the key features from an experimental perspective are listed.

Table 2-5 | Key Features of Focal Series Reconstruction

Beam convergence	Parallel beam
Raw data recorded in	Imaging plane
Data collection	Focal series images
Area of illumination	nm^2 ~ μm^2, limited by size of detector
Detector	Pixelated detector
Resolution limit by	HRTEM information limit
Typical low-fluence limit	26 $e^-/Å^2$ [138]

The shape of the transfer function is determined by the TEM settings (including aberration, energy spread, etc.) and the choice of defocus values for the input focal series [25]. The resolution at high frequencies is limited by the information limit, at the zero point of the envelope function. Since all information in the reconstructed exit wave is based on recorded TEM images, it is impossible to reconstruct information not transferred by HRTEM.

To extend the low-frequency range, images at high defocus values can be included in the focal series, and a non-uniform focal increment can further improve low-frequency transfer when needed [139], as shown in Figure 2-12. In conclusion, focal series reconstruction can effectively improve low-frequency transfer without sacrificing high-frequency resolution. However, high-frequency transfer is still limited by the information limit of HRTEM.

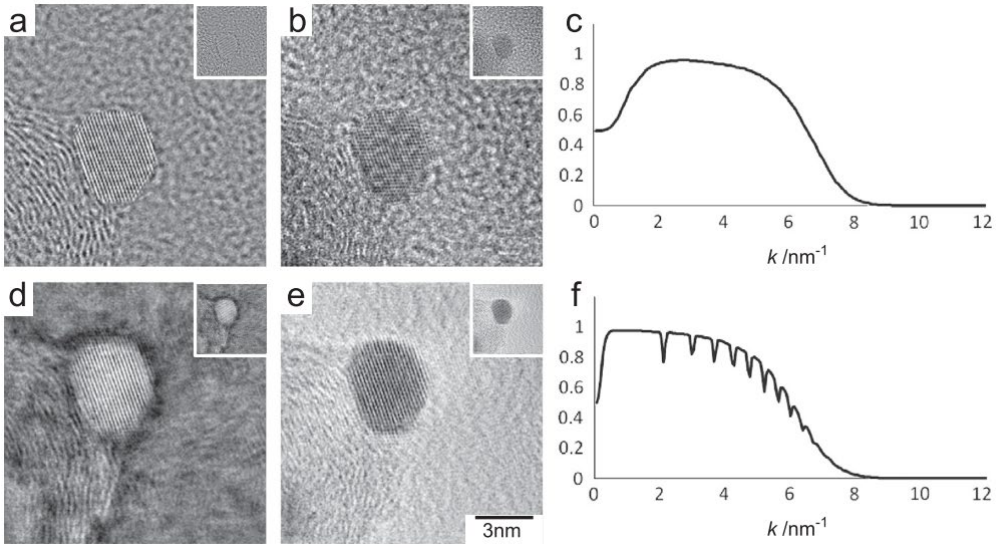

Figure 2-12 | Comparison of the information transfer for uniform and non-uniform focal increments [139]. (a) Phase, (b) modulus and (c) effective contrast transfer function of focal series of images recorded at 30.5, -26, -22, -20, -18, -12, -8, -3, 1, 5 and 10 nm with C_3 = 20 µm and C_{12} = -3.5, 4.7 nm. (d) Phase, (e) modulus and (f) effective contrast transfer function of focal series of images recorded at 528, -432, 336, -244, -148, -52, 43, 139, 235, 335 and 431 nm with C_3 = 0.5 mm and C_{12} = -68.6, -28.2 nm. Reprinted with permission from [139]. Copyright 2013 by Elsevier B.V.

For low-fluence focal series imaging, Huang *et al*. applied focal series reconstruction to zeolite (ZSM-5), achieving near-atomic resolution with a low-fluence reconstruction of 26 $e^-/Å^2$ [138]. In addition, Barton *et al*. compared focal series reconstructions across accelerating voltages from 20 kV to 300 kV [140]. Using a silicon crystal sample in the [110] direction, they achieved a resolution of 1.36 Å for Si dumbbells at 50, 80, and 300 kV, but not at 20 kV. This work demonstrates that focal series reconstruction can reconstruct phase information under low flux (or dose rate), down to 12 $e^-/Å^2 \cdot s$.

Data collection for focal series reconstruction can be automated using commercial software, such as the FTSR plugin for Gatan Digital Micrograph. Additionally, the reconstruction process can be significantly accelerated with GPU computing [141, 142]. it remains slower than conventional HRTEM, which requires only a single image and no post-processing.

Focal series reconstruction is generally not combined with other spectral analysis techniques, such as EDS or EELS, as focal series reconstruction relies on conventional TEM imaging settings and uses a plane wave. However, it can be applied in energy-filtered TEM (EFTEM) [143, 144], enabling the collection of EFTEM image stacks with consistent instrument settings. As an in-line holography technique, focal series reconstruction benefits from the introduction of an energy filter to reduce the influence of chromatic aberrations and inelastic scattering.

As an exit-wave retrieval method, focal series reconstruction is often compared with ptychography and other similar techniques. As an example Liberti *et al*. have compared electron ptychography and focal series reconstruction using a monolayer graphene sample [145]. There are two main differences between these exit-wave retrieval methods. First, the resolution of focal series reconstruction is constrained by the information limit of TEM, whereas ptychography can achieve resolution beyond this limit [93]. Second, the field of view in focal series reconstruction is limited by the camera size and pixel sampling, while in ptychography, it is determined by the scanning area.

2.4.5 Differential Phase Contrast and Centre-of-Mass

Differential phase contrast (DPC) in the scanning transmission electron microscope was proposed in 1974 [146, 147]. It is a phase contrast imaging method operated in STEM mode. Intuitively, the source of differential phase contrast is the shift of bright field disk at the diffraction plane, caused by an electric or magnetic field in the specimen plane [148], recorded by segmented ring-shape detector [149], as shown in Figure 2-13.

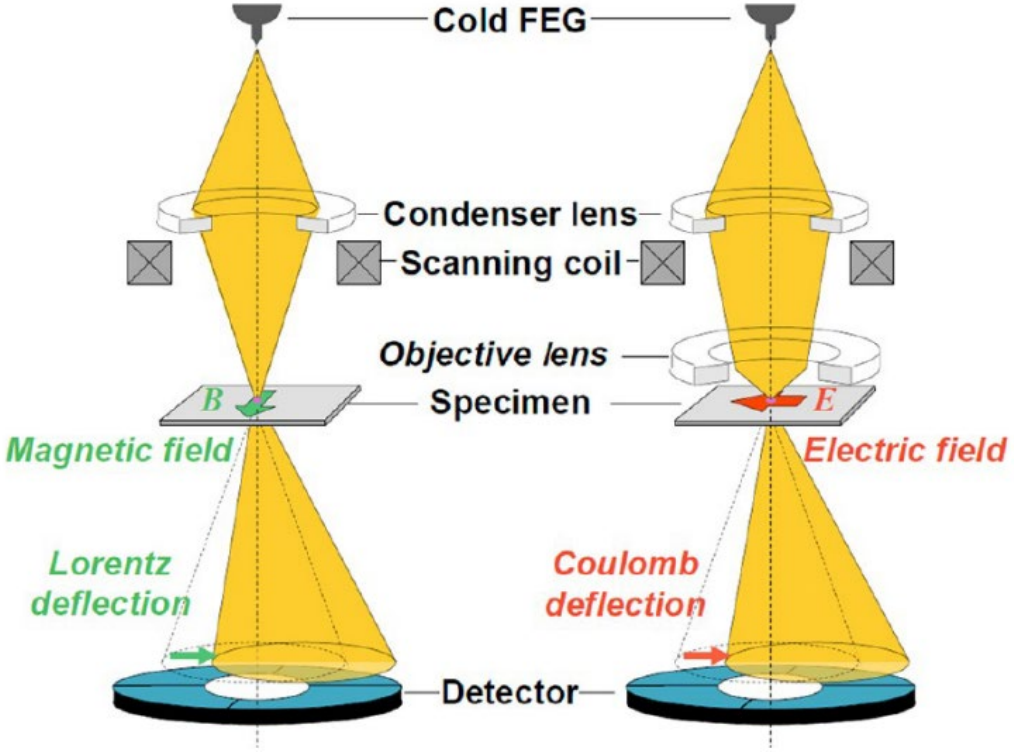

Figure 2-13 | Schematic illustration of STEM-DPC for specimens containing a magnetic field (left) and electric field (right) [148]. The electron beam is deflected by the magnetic field and electric field at the sample plane, which causes a shift of the bright disc in the diffraction plane. Reprinted with permission from [148]. Copyright 2017 by American Chemical Society.

Centre-of-Mass (COM) is an alternative way to obtain the phase contrast in a STEM setting. However, for COM images, the raw data is 4DSTEM dataset and recorded by pixelated detectors. The raw DPC data from segmented detectors can be regarded as an integral on a region of the pixelated detector, and the DPC method with segmented detectors can be seen as an approximation of COM method [150, 151]. The detailed algorithm of the COM method will be discussed in Chapter 3. However, it is noted here that DPC and COM methods share many key features in experiments, as shown in Table 2-6.

Table 2-6 | Key Features of DPC/COM method

Beam convergence	Convergent beam
Raw data recorded on	Diffraction plane
Data collection	Multi-channel STEM images (DPC) 4DSTEM dataset (COM)
Area of illumination	nm^2 ~ μm^2, limited by scanning area
Detector	Segmented detectors (DPC) Pixelated detector (COM)
Resolution limit by	Convergence angle (2α limit)
Typical low-fluence limit	26 $e^-/Å^2$ [138]
Additional condition	In-Focus

If the magnetic field is not considered, the contrast in the COM image is proportional to the distribution of the projected electrostatic field of a sample satisfying the POA, convolved with the probe intensity [152], as shown in Equation (2-12).

$$\boldsymbol{COM}(\boldsymbol{R}) = \boldsymbol{E_p}(\boldsymbol{R}) \otimes |P(\boldsymbol{R})|^2 \tag{2-12}$$

In Equation (2-12), $\boldsymbol{COM}(\boldsymbol{R})$ represents the COM result as a 2D vector, while $\boldsymbol{E_p}$ denotes the projected electrostatic field, $\otimes$ indicates convolution and $P(\boldsymbol{R})$ represents the probe function. Since the COM method is implemented in a STEM configuration, $P(\boldsymbol{R})$ here corresponds to an in-focus convergent beam rather than a plane wave. A detailed description of $P(\boldsymbol{R})$ will be provided in Chapter 3.

To obtain phase contrast imaging based on the COM results, further calculation of an

integrated Centre-of-mass (iCOM) is required. The relationship between COM and iCOM is described by Equation (2-13) [150], where $\boldsymbol{\nabla}$ represents the nabla operator and $\boldsymbol{COM}$ is the gradient of $iCOM$ with respect to x and y direction. Hence, $iCOM$ is proportional to the phase information of the sample (after processing for the contrast transfer function), which corresponds to the projected electrostatic potential, as shown in Equation 2-1. Additional computational details will be discussed in Section 3.6.

$$\boldsymbol{COM}(\boldsymbol{R}) = \left(\frac{\partial iCOM(\boldsymbol{R})}{\partial R_x}, \frac{\partial iCOM(\boldsymbol{R})}{\partial R_y}\right) = \boldsymbol{\nabla} iCOM(\boldsymbol{R}) \tag{2-13}$$

The divergence of a COM image is proportional to the projected charge distribution, referred to as differential Centre-of-mass (dCOM) [150], as given by Equation (2-14).

$$dCOM(\boldsymbol{R_p}) = \frac{\partial COM_x(\boldsymbol{R_p})}{\partial R_{px}} + \frac{\partial COM_y(\boldsymbol{R_p})}{\partial R_{py}} = \boldsymbol{\nabla} \cdot \boldsymbol{COM}(\boldsymbol{R_p}) \tag{2-14}$$

As previously noted, DPC can be considered as an approximation of COM under the condition of a segmented detector. Using the same mathematical framework, integrated DPC (iDPC) and differential DPC (dDPC) can similarly be derived from DPC, following the same methods used to calculate iCOM and dCOM from COM.

In terms of information transfer, the contrast transfer function for DPC (COM), iDPC (iCOM) and dDPC (dCOM) are similar. However, the results differ due to their different sources of contrast: the projected electric field, projected potential, and projected charge distribution, respectively. This section focuses on the CTF of iDPC and iCOM, which reflects the phase information in the sample.

The CTF of iDPC (iCOM) is illustrated in Figure 2-14 [150]. Considering the CTF, the resolution of iDPC (iCOM) is limited to $\frac{\lambda}{2\alpha}$ and iDPC (iCOM) exhibits a descending information transfer from the low-frequency range (lower than frequency corresponding to α) to the high-frequency range (higher than frequency corresponding to α), as shown in Figure 2-14. The information transfer approaches 1 when frequency approaches 0, but 0 frequency is not transferred. Furthermore, it is important to note that when a segmented

detector is used instead of a pixelated detector, the detector's asymmetry influences the phase contrast transfer function. This asymmetry causes a disruption in the symmetry of the phase contrast transfer function, as shown in Figures 2-14(e) and 2-14(f).

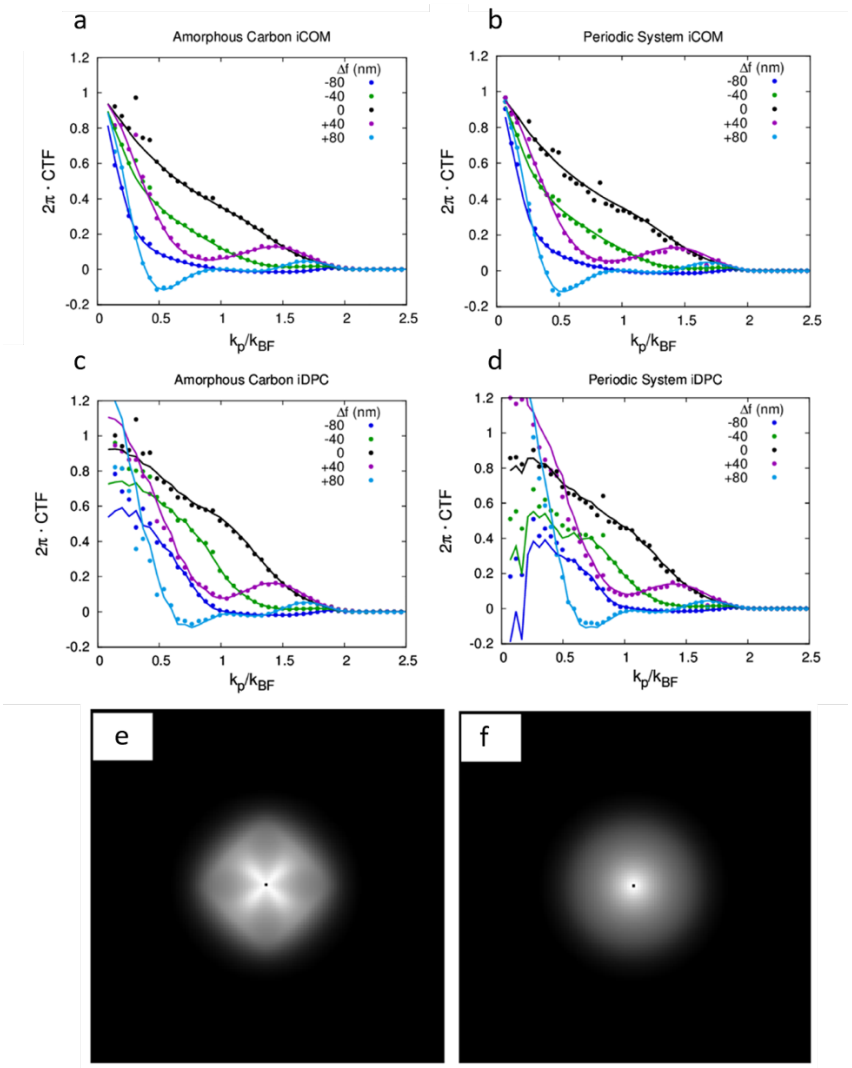

Figure 2-14 | Contrast transfer function of iCOM and iDPC on 4-sector detector [150]. Voltage: 300 kV, C_3: 2.7mm, convergence angle: 7 mrad. The dot lines on (a)-(d) are from simulated images and solid lines are from theoretical calculations. (a) iCOM contrast transfer function on amorphous carbon. (b) iCOM contrast transfer function on a periodic system of atoms arranged in the shape of the periodic table. (c) iDPC contrast transfer function on amorphous carbon. (d) iDPC contrast transfer function on periodic system. (e) 2D contrast transfer function of iDPC. (f) 2D contrast transfer function of iCOM. Reprinted with permission from [150]. Copyright 2015 by Elsevier B.V.

For low-fluence imaging, iDPC (iCOM) has been applied to radiation sensitive samples such as metal-organic frameworks [153], Li-based material [154] and zeolites [154-156]. Unlike traditional STEM-HAADF or STEM-ADF imaging methods, which rely on dark-field signals, iDPC (iCOM) utilises the bright-field disc. As the majority of transmitted electrons contribute to bright-field signals, this approach can achieve a better signal-to-noise ratio (SNR) under conditions where the total number of electrons is limited.

However, traditional DPC and COM methods require a focused probe. Therefore, during data collection, probe need to be accurately tuned to focus on the sample. This focusing process may results in additional illumination time comparing to defocus ptychography, which typically requires only a rough defocus estimate during data collection and allows for defocus refinement during reconstruction. This results in some fluence not contributing to the imaging process. To address this issue, Chapters 5 and 6 will discuss post-aberration correction for the COM method and DPC datasets on segmented detectors.

Another advantage of iDPC (iCOM) is its speed in both imaging and processing. Data collection can be fully automated and performed rapidly on both segmented and pixelated detectors using commercial hardware and software. Furthermore, the computational requirements for iDPC (iCOM) data processing are significantly lower than those for ptychography.

As the optical configuration for iDPC (iCOM) is based on STEM, this allows for easy integration with traditional STEM imaging methods (such as HAADF, ADF, etc.) [157-159]. Furthermore, DPC and COM methods have the potential to sacrifice some signal in the bright disc and incorporate a hole in the detector, enabling combination with EELS, in a geometry similar to proposed for ptychography.

Due to its real-time phase contrast imaging and compatibility with other imaging and spectrum methods, iDPC (iCOM) is widely used for specimens requiring contrast for light atoms in STEM. For example, iDPC (iCOM) has been used to image hydrogen atoms at a metal-metal hybrid interface [160]. However, as for traditional STEM imaging, iDPC (iCOM)

requires in-focus scanning, which is a disadvantage in terms of fast data collection (without fine tuning of defocus) comparing to defocus ptychography.

2.5 Conclusion

In this chapter, a brief introduction to TEM the sample and phase contrast in high-resolution TEM (HRTEM) is provided. Some common phase imaging methods in TEM/STEM are reviewed, including HRTEM, Zernike phase plates (ZPP), off-axis holography, ptychography, focal series reconstruction, and differential phase contrast/centre-of-mass (DPC/COM) methods.

In the next chapter, mathematical descriptions of the sample and its interaction with the electron beam, along with detailed algorithms for ptychography and COM methods, will be further discussed. Subsequently, Chapters 4 to 7 will introduce and discuss new developments built upon the algorithms presented in Chapter 3.

Reference

[1] H. Busch, "Über die Wirkungsweise der Konzentrierungsspule bei der Braunschen Röhre," *Archiv für Elektrotechnik,* vol. 18, no. 6, pp. 583-594, 1927/11/01 1927.

[2] M. Knoll and E. Ruska, "Das elektronenmikroskop," *Zeitschrift für physik,* vol. 78, pp. 318-339, 1932.

[3] M. Knoll and E. Ruska, "Beitrag zur geometrischen Elektronenoptik. I," *Annalen der Physik,* vol. 404, no. 5, pp. 607-640, 1932.

[4] M. Knoll and E. Ruska, "Beitrag zur geometrischen Elektronenoptik. II," *Annalen der Physik,* vol. 404, no. 6, pp. 641-661, 1932.

[5] F. Zernike, "Phase contrast, a new method for the microscopic observation of transparent objects," *Physica,* vol. 9, no. 7, pp. 686-698, 1942.

[6] F. Zernike, "Phase contrast, a new method for the microscopic observation of transparent objects part II," *Physica,* vol. 9, no. 10, pp. 974-986, 1942.

[7] D. Gabor, "A New Microscopic Principle," *Nature,* vol. 161, no. 4098, pp. 777-778, 1948.

[8] D. Gabor, "Microscopy by reconstructed wave-fronts," *Proceedings of the Royal*

Society of London. Series A. Mathematical and Physical Sciences, vol. 197, no. 1051, pp. 454-487, 1949.

[9] M. von Ardenne, "Das Elektronen-Rastermikroskop," *Zeitschrift für Physik,* vol. 109, no. 9, pp. 553-572, 1938/09/01 1938.

[10] A. V. Crewe, M. Isaacson, and D. Johnson, "A Simple Scanning Electron Microscope," *Review of Scientific Instruments,* vol. 40, no. 2, pp. 241-246, 1969.

[11] M. Haider, H. Rose, S. Uhlemann, E. Schwan, B. Kabius, and K. Urban, "A spherical-aberration-corrected 200 kV transmission electron microscope," *Ultramicroscopy,* vol. 75, no. 1, pp. 53-60, 1998.

[12] M. Haider, S. Uhlemann, E. Schwan, H. Rose, B. Kabius, and K. Urban, "Electron microscopy image enhanced," *Nature,* vol. 392, no. 6678, pp. 768-769, 1998.

[13] O. L. Krivanek, N. Dellby, and A. R. Lupini, "Towards sub-Å electron beams," *Ultramicroscopy,* vol. 78, no. 1-4, pp. 1-11, 1999.

[14] P. E. Batson, N. Dellby, and O. L. Krivanek, "Sub-ångstrom resolution using aberration corrected electron optics," *Nature,* vol. 418, no. 6898, pp. 617-620, 2002/08/01 2002.

[15] G. W. Bailey *et al.*, "Seeing is Not Always Believing: Reduction of Artefacts by an Improved Point Resolution With a Spherical Aberration Corrected 200 Kv Transmission Electron Microscope," *Microscopy and Microanalysis,* vol. 3, no. S2, pp. 1179-1180, 1997.

[16] D. B. Williams and C. B. Carter, "Specimen Preparation," in *Transmission Electron Microscopy: A Textbook for Materials Science. Imaging. I*, vol. 1: Springer Science & Business Media, 2008, ch. 10, pp. 173-191.

[17] T. Ishitani, H. Tsuboi, T. Yaguchi, and H. Koike, "Transmission electron microscope sample preparation using a focused ion beam," *Microscopy,* vol. 43, no. 5, pp. 322-326, 1994.

[18] L. A. Giannuzzi and F. A. Stevie, "A review of focused ion beam milling techniques for TEM specimen preparation," *Micron,* vol. 30, no. 3, pp. 197-204, 1999.

[19] S. B. Newman, E. Borysko, and M. Swerdlow, "New sectioning techniques for light and electron microscopy," *Science,* vol. 110, no. 2846, pp. 66-68, 1949.

[20] K. A. Taylor and R. M. Glaeser, "Electron diffraction of frozen, hydrated protein crystals," *Science,* vol. 186, no. 4168, pp. 1036-1037, 1974.

[21] M. Adrian, J. Dubochet, J. Lepault, and A. W. McDowall, "Cryo-electron microscopy of viruses," *Nature,* vol. 308, no. 5954, pp. 32-36, 1984.

[22] Y. Aharonov and D. Bohm, "Significance of electromagnetic potentials in the quantum theory," *Physical Review,* vol. 115, no. 3, p. 485, 1959.

[23] W. Gao *et al.*, "Real-space charge-density imaging with sub-ångström resolution

by four-dimensional electron microscopy," *Nature,* vol. 575, no. 7783, pp. 480-484, Nov 2019.

[24] K. Muller-Caspary *et al.*, "Measurement of atomic electric fields and charge densities from average momentum transfers using scanning transmission electron microscopy," *Ultramicroscopy,* vol. 178, pp. 62-80, Jul 2017.

[25] A. I. Kirkland, S. L.-Y. Chang, and J. L. Hutchison, "Atomic Resolution Transmission Electron Microscopy," in *Springer Handbook of Microscopy*, P. W. Hawkes and J. C. H. Spence Eds. Cham: Springer International Publishing, 2019, ch. 1, pp. 3-47.

[26] D. L. Misell, "The phase problem in electron microscopy," in *Advances in Imaging and Electron Physics*, vol. 225, M. Hÿtch and P. W. Hawkes Eds.: Elsevier, 2023, ch. 4, pp. 163-262.

[27] M. Haider, S. Uhlemann, and J. Zach, "Upper limits for the residual aberrations of a high-resolution aberration-corrected STEM," *Ultramicroscopy,* vol. 81, no. 3-4, pp. 163-175, 2000.

[28] H. Sawada *et al.*, "Measurement method of aberration from Ronchigram by autocorrelation function," *Ultramicroscopy,* vol. 108, no. 11, pp. 1467-1475, 2008.

[29] D. B. Williams and C. B. Carter, "High-resolution TEM," in *Transmission Electron Microscopy: A Textbook for Materials Science. Imaging. III*, vol. 3: Springer Science & Business Media, 2008, ch. 28, pp. 483-509.

[30] O. Scherzer, "The Theoretical Resolution Limit of the Electron Microscope," *Journal of Applied Physics,* vol. 20, no. 1, pp. 20-29, 1949.

[31] F. Hosokawa, H. Sawada, Y. Kondo, K. Takayanagi, and K. Suenaga, "Development of Cs and Cc correctors for transmission electron microscopy," *Microscopy,* vol. 62, no. 1, pp. 23-41, 2013.

[32] N. Dellby, O. L. Krivanek, P. D. Nellist, E. Batson, and A. R. Lupini, "Progress in aberration-corrected scanning transmission electron microscopy," *Microscopy,* vol. 50, no. 3, pp. 177-185, 2001.

[33] H. Jeong, H. Park, J. Kim, and J. Hyun, "Critical importance of the correction of contrast transfer function for transmission electron microscopy-mediated structural biology," *Journal of Analytical Science and Technology,* vol. 4, pp. 1-7, 2013.

[34] K. Nagayama and R. Danev, "Phase contrast electron microscopy: development of thin-film phase plates and biological applications," *Philos Trans R Soc Lond B Biol Sci,* vol. 363, no. 1500, pp. 2153-62, Jun 27 2008.

[35] D. B. Carlson, J. E. Evans, and K. Maaz, "Low-dose imaging techniques for transmission electron microscopy," *The transmission electron microscope,* vol. 95, no. 3, pp. 85-98, 2012.

[36] Z. Zhan, Y. Liu, W. Wang, G. Du, S. Cai, and P. Wang, "Atomic-level imaging of beam-sensitive COFs and MOFs by low-dose electron microscopy," *Nanoscale*

Horizons, 2024.

[37] Y. Lin, T. Zhou, A. Lei, and J. Wen, "Low-dose HRTEM Study of Stacking Fault Structures in Cubic Ice," *Microscopy and Microanalysis,* vol. 28, no. S1, pp. 2258-2260, 2022.

[38] P. Roingeard, P. I. Raynal, S. Eymieux, and E. Blanchard, "Virus detection by transmission electron microscopy: Still useful for diagnosis and a plus for biosafety," *Reviews in medical virology,* vol. 29, no. 1, p. e2019, 2019.

[39] K. H. Downing and R. M. Glaeser, "Restoration of weak phase-contrast images recorded with a high degree of defocus: the “twin image” problem associated with CTF correction," *Ultramicroscopy,* vol. 108, no. 9, pp. 921-928, 2008.

[40] G. Zanetti, J. D. Riches, S. D. Fuller, and J. A. Briggs, "Contrast transfer function correction applied to cryo-electron tomography and sub-tomogram averaging," *J Struct Biol,* vol. 168, no. 2, pp. 305-12, Nov 2009.

[41] A. Rohou and N. Grigorieff, "CTFFIND4: Fast and accurate defocus estimation from electron micrographs," *J Struct Biol,* vol. 192, no. 2, pp. 216-21, Nov 2015.

[42] X. Li *et al.*, "Direct Imaging of Tunable Crystal Surface Structures of MOF MIL-101 Using High-Resolution Electron Microscopy," *J Am Chem Soc,* vol. 141, no. 30, pp. 12021-12028, Jul 31 2019.

[43] D. Zhang *et al.*, "Atomic-resolution transmission electron microscopy of electron beam–sensitive crystalline materials," *Science,* vol. 359, no. 6376, pp. 675-679, 2018.

[44] J. E. Evans, C. Hetherington, A. I. Kirkland, L. Y. Chang, H. Stahlberg, and N. Browning, "Low-dose aberration corrected cryo-electron microscopy of organic specimens," *Ultramicroscopy,* vol. 108, no. 12, pp. 1636-44, Nov 2008.

[45] S. J. Ludtke, P. R. Baldwin, and W. Chiu, "EMAN: semiautomated software for high-resolution single-particle reconstructions," *Journal of structural biology,* vol. 128, no. 1, pp. 82-97, 1999.

[46] L. Chang, F. Chen, C. Chen, and T. Lin, "HRTEM and energy filtering TEM study of interfacial reaction between diamond film and silicon," *Diamond and related materials,* vol. 5, no. 11, pp. 1282-1287, 1996.

[47] A. Berger and H. Kohl, "Elemental mapping using an imaging energy filter: image formation and resolution limits," *Microscopy Microanalysis Microstructures,* vol. 3, no. 2-3, pp. 159-174, 1992.

[48] W. M. Stobbs and W. O. Saxton, "Quantitative high resolution transmission electron microscopy: the need for energy filtering and the advantages of energy-loss imaging," *Journal of Microscopy,* vol. 151, no. 2, pp. 171-184, 1988.

[49] R. K. Leary and P. A. Midgley, "Electron Tomography in Materials Science," in *Springer Handbook of Microscopy*, P. W. Hawkes and J. C. H. Spence Eds. Cham:

Springer International Publishing, 2019, pp. 1279-1329.

[50] M. Bar Sadan *et al.*, "Toward atomic-scale bright-field electron tomography for the study of fullerene-like nanostructures," *Nano letters,* vol. 8, no. 3, pp. 891-896, 2008.

[51] M. Turk and W. Baumeister, "The promise and the challenges of cryo-electron tomography," *FEBS letters,* vol. 594, no. 20, pp. 3243-3261, 2020.

[52] R. Danev and K. Nagayama, "Transmission electron microscopy with Zernike phase plate," *Ultramicroscopy,* vol. 88, no. 4, pp. 243-252, 2001.

[53] H. Lichte and M. Lehmann, "Electron holography—basics and applications," *Reports on Progress in Physics,* vol. 71, no. 1, 2008.

[54] S. Hettler and R. Arenal, "Aberration-corrected transmission electron microscopy with Zernike phase plates," *Ultramicroscopy,* vol. 239, p. 113564, 2022.

[55] M. Malac, M. Beleggia, R. Egerton, and Y. Zhu, "Imaging of radiation-sensitive samples in transmission electron microscopes equipped with Zernike phase plates," *Ultramicroscopy,* vol. 108, no. 2, pp. 126-140, 2008.

[56] Y. Fukuda, Y. Fukazawa, R. Danev, R. Shigemoto, and K. Nagayama, "Tuning of the Zernike phase-plate for visualization of detailed ultrastructure in complex biological specimens," *Journal of structural biology,* vol. 168, no. 3, pp. 476-484, 2009.

[57] W. Chang *et al.*, "Zernike phase plate cryoelectron microscopy facilitates single particle analysis of unstained asymmetric protein complexes," *Structure,* vol. 18, no. 1, pp. 17-27, 2010.

[58] E. Majorovits, B. Barton, K. Schultheiss, F. Perez-Willard, D. Gerthsen, and R. R. Schröder, "Optimizing phase contrast in transmission electron microscopy with an electrostatic (Boersch) phase plate," *Ultramicroscopy,* vol. 107, no. 2-3, pp. 213-226, 2007.

[59] K. Schultheiss *et al.*, "New electrostatic phase plate for phase-contrast transmission electron microscopy and its application for wave-function reconstruction," *Microscopy and Microanalysis,* vol. 16, no. 6, pp. 785-794, 2010.

[60] M. Malac, M. Beleggia, M. Kawasaki, P. Li, and R. F. Egerton, "Convenient contrast enhancement by a hole-free phase plate," *Ultramicroscopy,* vol. 118, pp. 77-89, 2012.

[61] R. Danev, B. Buijsse, M. Khoshouei, J. M. Plitzko, and W. Baumeister, "Volta potential phase plate for in-focus phase contrast transmission electron microscopy," *Proceedings of the National Academy of Sciences,* vol. 111, no. 44, pp. 15635-15640, 2014.

[62] J. M. Cowley, "Twenty forms of electron holography," *Ultramicroscopy,* vol. 41, no. 4, pp. 335-348, 1992.

[63] G. Möllenstedt and H. Düker, "Beobachtungen und messungen an biprisma-interferenzen mit elektronenwellen," *Zeitschrift für Physik,* vol. 145, no. 3, pp. 377-397, 1956.

[64] P. A. Midgley, "An introduction to off-axis electron holography," *Micron,* vol. 32, no. 2, pp. 167-184, 2001.

[65] E. Völkl, L. F. Allard, and B. Frost, "A software package for the processing and reconstruction of electron holograms," *Journal of microscopy,* vol. 180, no. 1, pp. 39-50, 1995.

[66] K. Morawiec, W. Zajkowska, P. Dłużewski, M. Shiojiri, and J. Kusiński, "PyHoLo software, a new tool for electron hologram analysis and magnetic investigation," *Computer Physics Communications,* vol. 256, 2020.

[67] J. Bonevich, "HolograFree Software," *https://www.nist.gov/services-resources/software/holografree*.

[68] R. E. Dunin-Borkowski, A. Kovács, T. Kasama, M. R. McCartney, and D. J. Smith, "Electron Holography," in *Springer Handbook of Microscopy*, P. W. Hawkes and J. C. H. Spence Eds. Cham: Springer International Publishing, 2019, ch. 16, pp. 767-818.

[69] W. D. Rau and H. Lichte, "High Resolution Off-Axis Electron Holography," in *Introduction to Electron Holography*, E. Völkl, L. F. Allard, and D. C. Joy Eds. Boston, MA: Springer US, 1999, pp. 201-229.

[70] M. Linck, H. Lichte, and M. Lehmann, "Off-axis electron holography: Materials analysis at atomic resolution," *International Journal of Materials Research,* vol. 97, no. 7, pp. 890-898, 2006.

[71] M. Lehmann and H. Lichte, "Tutorial on Off-Axis Electron Holography," *Microscopy and Microanalysis,* vol. 8, no. 6, pp. 447-466, 2004.

[72] H. Lichte, M. Linck, D. Geiger, and M. Lehmann, "Aberration correction and electron holography," *Microscopy and Microanalysis,* vol. 16, no. 4, p. 434, 2010.

[73] A. Tonomura, T. Matsuda, and J. Endo, "Spherical-aberration correction of an electron lens by holography," *Japanese journal of applied physics,* vol. 18, no. 7, p. 1373, 1979.

[74] D. Van Dyck, H. Lichte, and J. C. H. Spence, "Inelastic scattering and holography," *Ultramicroscopy,* vol. 81, no. 3-4, pp. 187-194, 2000.

[75] R. Clough and A. I. Kirkland, "Direct digital electron detectors," in *Advances in Imaging and Electron Physics*, vol. 198: Elsevier, 2016, pp. 1-42.

[76] E. Voelkl, R. Herring, B. Bammes, and D. Hoyle, "Low Dose Electron Holography: First Steps," *Microscopy and Microanalysis,* vol. 21, no. S3, pp. 1951-1952, 2015.

[77] S. L. Y. Chang, C. Dwyer, J. Barthel, C. B. Boothroyd, and R. E. Dunin-Borkowski,

"Performance of a direct detection camera for off-axis electron holography," *Ultramicroscopy,* vol. 161, pp. 90-97, Feb 2016.

[78] S. Anada, Y. Nomura, T. Hirayama, and K. Yamamoto, "Simulation-Trained Sparse Coding for High-Precision Phase Imaging in Low-Dose Electron Holography," *Microscopy and Microanalysis,* vol. 26, no. 3, pp. 429-438, Jun 2020.

[79] A. H. Tavabi, Z. Yasenjiang, and T. Tanji, "In situ off-axis electron holography of metal-oxide hetero-interfaces in oxygen atmosphere," *J Electron Microsc (Tokyo),* vol. 60, no. 5, pp. 307-14, 2011.

[80] M. den Hertog *et al.*, "In situ biasing and off-axis electron holography of a ZnO nanowire," *Nanotechnology,* vol. 29, no. 2, p. 025710, Jan 12 2018.

[81] D. J. Smith and M. R. McCartney, "Practical Electron Holography," in *Introduction to Electron Holography*, E. Völkl, L. F. Allard, and D. C. Joy Eds. Boston, MA: Springer US, 1999, pp. 87-106.

[82] M. Lehmann, "Influence of the elliptical illumination on acquisition and correction of coherent aberrations in high-resolution electron holography," *Ultramicroscopy,* vol. 100, no. 1-2, pp. 9-23, 2004.

[83] W. Hoppe, "Principles of electron structure research at atomic resolution using conventional electron microscopes for the measurement of amplitudes and phases," *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography,* vol. 26, no. 4, pp. 414-426, 1970.

[84] R. W. Gerchberg and W. O. Saxton, "A practical algorithm for the determination of phase from image and diffraction plane pictures," *Optik,* vol. 35, pp. 237-246, 1972.

[85] R. W. Gerchberg, "Super-resolution through error energy reduction," *Optica Acta: International Journal of Optics,* vol. 21, no. 9, pp. 709-720, 1974.

[86] B. Gu and G. Yang, "On the phase retrieval problem in optical and electronic microscopy," *Acta Opt. Sin.,* vol. 1, no. 6, pp. 517-522, 1981.

[87] G. Yang, L. Wang, B. Dong, and B. Gu, "On the amplitude-phase retrieval problem in an optical system involved non-unitary transformation," *Optik (Stuttgart),* vol. 75, no. 2, pp. 68-74, 1987.

[88] J. R. Fienup, "Phase retrieval algorithms: a comparison," *Applied optics,* vol. 21, no. 15, pp. 2758-2769, 1982.

[89] J. N. Clark *et al.*, "Use of a complex constraint in coherent diffractive imaging," *Optics Express,* vol. 18, no. 3, pp. 1981-1993, 2010/02/01 2010.

[90] W. Hoppe, "Beugung im inhomogenen primärstrahlwellenfeld. i. prinzip einer phasenmessung von elektronenbeugungsinterferenzen," *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography,* vol. 25, no. 4, pp. 495-501, 1969.

[91] J. M. Rodenburg and R. H. T. Bates, "The theory of super-resolution electron microscopy via Wigner-distribution deconvolution," *Philosophical Transactions of the Royal Society of London. Series A: Physical and Engineering Sciences,* vol. 339, no. 1655, pp. 521-553, 1992.

[92] J. M. Rodenburg, B. C. McCallum, and P. D. Nellist, "Experimental tests on double-resolution coherent imaging via STEM," *Ultramicroscopy,* vol. 48, no. 3, pp. 304-314, 1993.

[93] P. D. Nellist, B. C. McCallum, and J. M. Rodenburg, "Resolution beyond the'information limit'in transmission electron microscopy," *Nature,* vol. 374, no. 6523, pp. 630-632, 1995.

[94] J. M. Rodenburg and H. M. L. Faulkner, "A phase retrieval algorithm for shifting illumination," *Applied Physics Letters,* vol. 85, no. 20, pp. 4795-4797, 2004.

[95] W. Hoppe, "Zur" Abbildung" komplexer Bildfunktionen in der Elektronenmikroskopie," *Zeitschrift für Naturforschung A,* vol. 26, no. 7, pp. 1155-1168, 1971.

[96] S. J. Haigh, H. Sawada, and A. I. Kirkland, "Atomic Structure Imaging Beyond Conventional Resolution Limits in the Transmission Electron Microscope," *Physical Review Letters,* vol. 103, no. 12, p. 126101, 09/17/ 2009.

[97] C. M. O'Leary, G. T. Martinez, E. Liberti, M. J. Humphry, A. I. Kirkland, and P. D. Nellist, "Contrast transfer and noise considerations in focused-probe electron ptychography," *Ultramicroscopy,* vol. 221, p. 113189, Feb 2021.

[98] C. Ophus, "Four-dimensional scanning transmission electron microscopy (4D-STEM): From scanning nanodiffraction to ptychography and beyond," *Microscopy and Microanalysis,* vol. 25, no. 3, pp. 563-582, 2019.

[99] H. Yang *et al.*, "Simultaneous atomic-resolution electron ptychography and Z-contrast imaging of light and heavy elements in complex nanostructures," *Nature communications,* vol. 7, p. 12532, Aug 26 2016.

[100] T. J. Pennycook, A. R. Lupini, H. Yang, M. F. Murfitt, L. Jones, and P. D. Nellist, "Efficient phase contrast imaging in STEM using a pixelated detector. Part 1: experimental demonstration at atomic resolution," *Ultramicroscopy,* vol. 151, pp. 160-167, Apr 2015.

[101] B. Song *et al.*, "Hollow Electron Ptychographic Diffractive Imaging," *Physical Review Letters,* vol. 121, no. 14, p. 146101, Oct 5 2018.

[102] J. M. Rodenburg and A. M. Maiden, "Ptychography," in *Springer Handbook of Microscopy*, P. W. Hawkes and J. C. H. Spence Eds. Cham: Springer International Publishing, 2019, ch. 17, pp. 819-905.

[103] C. M. O'Leary *et al.*, "Phase reconstruction using fast binary 4D STEM data," *Applied Physics Letters,* vol. 116, no. 12, 2020.

[104] H. Yang, P. Ercius, P. D. Nellist, and C. Ophus, "Enhanced phase contrast transfer using ptychography combined with a pre-specimen phase plate in a scanning transmission electron microscope," *Ultramicroscopy,* vol. 171, pp. 117-125, Dec 2016.

[105] P. Li, T. B. Edo, and J. M. Rodenburg, "Ptychographic inversion via Wigner distribution deconvolution: noise suppression and probe design," *Ultramicroscopy,* vol. 147, pp. 106-113, 2014.

[106] Y. Jiang *et al.*, "Electron ptychography of 2D materials to deep sub-ångström resolution," *Nature,* vol. 559, no. 7714, pp. 343-349, Jul 2018.

[107] T. J. Pennycook, G. T. Martinez, P. D. Nellist, and J. C. Meyer, "High dose efficiency atomic resolution imaging via electron ptychography," *Ultramicroscopy,* vol. 196, pp. 131-135, Jan 2019.

[108] J. G. Lozano, G. T. Martinez, L. Jin, P. D. Nellist, and P. G. Bruce, "Low-Dose Aberration-Free Imaging of Li-Rich Cathode Materials at Various States of Charge Using Electron Ptychography," *Nano Letters,* vol. 18, no. 11, pp. 6850-6855, Nov 14 2018.

[109] G. McMullan, A. R. Faruqi, D. Clare, and R. Henderson, "Comparison of optimal performance at 300keV of three direct electron detectors for use in low dose electron microscopy," *Ultramicroscopy,* vol. 147, pp. 156-63, Dec 2014.

[110] J. A. Mir *et al.*, "Characterisation of the Medipix3 detector for 60 and 80keV electrons," *Ultramicroscopy,* vol. 182, pp. 44-53, Nov 2017.

[111] H. Ryll *et al.*, "A pnCCD-based, fast direct single electron imaging camera for TEM and STEM," *Journal of Instrumentation,* vol. 11, no. 04, pp. P04006-P04006, 2016.

[112] A. Bangun, P. F. Baumeister, A. Clausen, D. Weber, and R. E. Dunin-Borkowski, "Wigner Distribution Deconvolution Adaptation for Live Ptychography Reconstruction," *Microscopy and Microanalysis,* vol. 29, no. 3, pp. 994-1008, 2023.

[113] A. M. Maiden and J. M. Rodenburg, "An improved ptychographical phase retrieval algorithm for diffractive imaging," *Ultramicroscopy,* vol. 109, no. 10, pp. 1256-62, Sep 2009.

[114] M. J. Humphry, B. Kraus, A. C. Hurst, A. M. Maiden, and J. M. Rodenburg, "Ptychographic electron microscopy using high-angle dark-field scattering for sub-nanometre resolution imaging," *Nature communications,* vol. 3, p. 730, Mar 6 2012.

[115] A. M. Maiden, D. Johnson, and P. Li, "Further improvements to the ptychographical iterative engine," *Optica,* vol. 4, no. 7, pp. 736-745, 2017.

[116] A. Sun *et al.*, "Sub-aperture switching based ptychographic iterative engine (sasPIE) method for quantitative imaging," *Optics Communications,* vol. 410, pp. 514-519, 2018.

[117] M. Guizar-Sicairos and J. R. Fienup, "Phase retrieval with transverse translation

diversity: a nonlinear optimization approach," *Optics express,* vol. 16, no. 10, pp. 7264-7278, 2008.

[118] P. Thibault, M. Dierolf, A. Menzel, O. Bunk, C. David, and F. Pfeiffer, "High-resolution scanning x-ray diffraction microscopy," *Science,* vol. 321, no. 5887, pp. 379-382, 2008.

[119] P. M. Pelz, W. X. Qiu, R. Bucker, G. Kassier, and R. J. D. Miller, "Low-dose cryo electron ptychography via non-convex Bayesian optimization," *Scientific reports,* vol. 7, no. 1, p. 9883, Aug 29 2017.

[120] L. Zhou *et al.*, "Low-dose phase retrieval of biological specimens using cryo-electron ptychography," *Nature communications,* vol. 11, no. 1, p. 2773, Jun 2 2020.

[121] Z. Chen *et al.*, "Mixed-state electron ptychography enables sub-angstrom resolution imaging with picometer precision at low dose," *Nature communications,* vol. 11, no. 1, p. 2994, 2020.

[122] Z. Chen *et al.*, "Electron ptychography achieves atomic-resolution limits set by lattice vibrations," *Science,* vol. 372, no. 6544, pp. 826-831, 2021.

[123] W. Mao *et al.*, "Multi-Convergence-Angle Ptychography with Simultaneous Strong Contrast and High Resolution," *arXiv preprint arXiv:2403.16902,* 2024.

[124] J. Song *et al.*, "Atomic resolution defocused electron ptychography at low dose with a fast, direct electron detector," *Scientific reports,* vol. 9, no. 1, p. 3919, 2019.

[125] X. Pei *et al.*, "Cryogenic electron ptychographic single particle analysis with wide bandwidth information transfer," *Nature Communications,* vol. 14, no. 1, p. 3027, 2023.

[126] Z. Dong *et al.*, "High-performance multi-mode ptychography reconstruction on distributed GPUs," in *2018 New York Scientific Data Summit (NYSDS)*, 2018: IEEE, pp. 1-5.

[127] S. Marchesini *et al.*, "SHARP: a distributed GPU-based ptychographic solver," *Journal of applied crystallography,* vol. 49, no. 4, pp. 1245-1252, 2016.

[128] Z. Ding *et al.*, "Three-dimensional electron ptychography of organic–inorganic hybrid nanostructures," *Nature Communications,* vol. 13, no. 1, p. 4787, 2022.

[129] J. M. Cowley and A. F. Moodie, "The scattering of electrons by atoms and crystals. I. A new theoretical approach," *Acta Crystallographica,* vol. 10, no. 10, pp. 609-619, 1957.

[130] P. Goodman and A. F. Moodie, "Numerical evaluations of N-beam wave functions in electron scattering by the multi-slice method," *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography,* vol. 30, no. 2, pp. 280-290, 1974.

[131] S. Gao *et al.*, "Electron ptychographic microscopy for three-dimensional imaging,"

Nature Communications, vol. 8, no. 1, p. 163, 2017.

[132] A. H. Buist, A. Van den Bos, and M. A. O. Miedema, "Optimal experimental design for exit wave reconstruction from focal series in TEM," *Ultramicroscopy,* vol. 64, no. 1-4, pp. 137-152, 1996.

[133] W. Coene, G. Janssen, M. Op de Beeck, and D. Van Dyck, "Phase retrieval through focus variation for ultra-resolution in field-emission transmission electron microscopy," *Physical Review Letters,* vol. 69, no. 26, pp. 3743-3746, Dec 28 1992.

[134] W. M. J. Coene, A. Thust, M. Op de Beeck, and D. Van Dyck, "Maximum-likelihood method for focus-variation image reconstruction in high resolution transmission electron microscopy," *Ultramicroscopy,* vol. 64, no. 1-4, pp. 109-135, 1996.

[135] L. J. Allen, W. McBride, N. L. O'Leary, and M. P. Oxley, "Exit wave reconstruction at atomic resolution," *Ultramicroscopy,* vol. 100, no. 1-2, pp. 91-104, Jul 2004.

[136] P. C. Tiemeijer, M. Bischoff, B. Freitag, and C. Kisielowski, "Using a monochromator to improve the resolution in TEM to below 0.5 Å. Part II: application to focal series reconstruction," *Ultramicroscopy,* vol. 118, pp. 35-43, Jul 2012.

[137] R. R. Meyer, A. I. Kirkland, and W. O. Saxton, "A new method for the determination of the wave aberration function for high resolution TEM: 1. Measurement of the symmetric aberrations," *Ultramicroscopy,* vol. 92, no. 2, pp. 89-109, 2002.

[138] C. Huang, K. B. Borisenko, and A. I. Kirkland, "Exit wave reconstruction of radiation-sensitive materials from low-dose data," in *Journal of Physics: Conference Series*, 2014, vol. 522, no. 1: IOP Publishing, p. 012052.

[139] S. J. Haigh, B. Jiang, D. Alloyeau, C. Kisielowski, and A. I. Kirkland, "Recording low and high spatial frequencies in exit wave reconstructions," *Ultramicroscopy,* vol. 133, pp. 26-34, Oct 2013.

[140] B. Barton, B. Jiang, C. Song, P. Specht, H. Calderon, and C. Kisielowski, "Atomic resolution phase contrast imaging and in-line electron holography using variable voltage and dose rate," *Microscopy and Microanalysis,* vol. 18, p. 12, 2012.

[141] M. A. Dyson and J. Sloan, "GPU accelerated image simulation, image processing and exit wave reconstruction," presented at the EMC2012.

[142] K. Morawiec, P. Dłużewski, S. Kret, A. Szczepańska, T. Li, and J. Sloan, "Application of Graphics Processing Unit for In-Line Electron Holography," *Acta Physica Polonica A,* vol. 131, no. 5, pp. 1353-1357, 2017.

[143] C. T. Koch, "A flux-preserving non-linear inline holography reconstruction algorithm for partially coherent electrons," *Ultramicroscopy,* vol. 108, no. 2, pp. 141-50, Jan 2008.

[144] J. L. Hutchison *et al.*, "A versatile double aberration-corrected, energy filtered HREM/STEM for materials science," *Ultramicroscopy,* vol. 103, no. 1, pp. 7-15, Apr 2005.

[145] E. Liberti, H. Yang, G. T. Martinez, P. D. Nellist, and A. I. Kirkland, "Quantitative Comparison of Phase Contrast Imaging in Conventional TEM Focal Series and STEM Ptychography," *Microscopy and Microanalysis,* vol. 22, no. S3, pp. 922-923, 2016.

[146] N. H. Dekkers and H. de Lang, "Differential phase contrast in a STEM," *Optik,* vol. 41, pp. 452-456, 1974.

[147] H. Rose, "Phase contrast in scanning transmission electron microscopy," *Optik,* vol. 39, pp. 416-436, 1974.

[148] N. Shibata *et al.*, "Direct Visualization of Local Electromagnetic Field Structures by Scanning Transmission Electron Microscopy," *Accounts of Chemical Research,* vol. 50, no. 7, pp. 1502-1512, 2017.

[149] N. Shibata, S. D. Findlay, Y. Kohno, H. Sawada, Y. Kondo, and Y. Ikuhara, "Differential phase-contrast microscopy at atomic resolution," *Nature Physics,* vol. 8, no. 8, pp. 611-615, 2012.

[150] I. Lazić, E. G. T. Bosch, and S. Lazar, "Phase contrast STEM for thin samples: Integrated differential phase contrast," *Ultramicroscopy,* vol. 160, pp. 265-280, Jan 2016.

[151] I. Lazić and E. G. T. Bosch, "Analytical Review of Direct Stem Imaging Techniques for Thin Samples," in *Advances in Imaging and Electron Physics*, 2017, ch. 3, pp. 75-184.

[152] K. Müller *et al.*, "Atomic electric fields revealed by a quantum mechanical approach to electron picodiffraction," *Nature communications,* vol. 5, no. 1, p. 5653, 2014.

[153] B. Shen, X. Chen, K. Shen, H. Xiong, and F. Wei, "Imaging the node-linker coordination in the bulk and local structures of metal-organic frameworks," *Nature Communications,* vol. 11, no. 1, p. 2692, Jun 1 2020.

[154] A. Carlsson, I. Alexandrou, E. Yücelen, E. G. T. Bosch, and I. Lazić, "Low Dose Imaging Using Simultaneous iDPC- and ADF-STEM for Beam Sensitive Crystalline Structures," *Microscopy and Microanalysis,* vol. 24, no. S1, pp. 122-123, 2018.

[155] B. Shen *et al.*, "Atomic Spatial and Temporal Imaging of Local Structures and Light Elements inside Zeolite Frameworks," *Advanced Materials,* vol. 32, no. 4, p. e1906103, Jan 2020.

[156] L. Liu *et al.*, "Direct Imaging of Atomically Dispersed Molybdenum that Enables Location of Aluminum in the Framework of Zeolite ZSM-5," *Angew Chem Int Ed Engl,* vol. 59, no. 2, pp. 819-825, 2020.

[157] R. Close, Z. Chen, N. Shibata, and S. D. Findlay, "Towards quantitative, atomic-resolution reconstruction of the electrostatic potential via differential phase contrast using electrons," *Ultramicroscopy,* vol. 159 Pt 1, pp. 124-37, Dec 2015.

[158] E. G. T. Bosch, I. Lazić, E. Yucelen, and P. Trompenaars, "Thick (3D) Sample

STEM Imaging at Nano Scale: iDPC and ADF Simultaneously," *Microscopy and Microanalysis,* vol. 24, no. S1, pp. 226-227, 2018.

[159] I. Lazić, E. Yucelen, and E. G. T. Bosch, "Simultaneous iDPC and ADF STEM Imaging at the Limit of Contrast and Resolution," *Microscopy and Microanalysis,* vol. 24, no. S1, pp. 214-215, 2018.

[160] S. de Graaf, J. Momand, C. Mitterbauer, S. Lazar, and B. J. Kooi, "Resolving hydrogen atoms at metal-metal hydride interfaces," *Science advances,* vol. 6, no. 5, p. eaay4312, 2020.

Chapter 3: Methods

3.1 Introduction

In the previous chapter, a brief overview of common phase contrast imaging methods is provided, including electron ptychography (Section 2.4.3) and centre-of-mass (Section 2.4.5) techniques. These methods have been reported to be effective for imaging beam sensitive materials [1-4]. The subsequent chapters will focus on the advancements towards low-fluence conditions based on these methods. In this chapter, the theoretical background for the subsequent chapters and the details of the related imaging algorithms will be provided.

In Section 3.2, the sources of information for ptychography and centre-of-mass methods will be discussed at first. Then the mathematical representations of probe functions and aberrations used in subsequent chapters will be introduced. The sample is described as an object transmission function, along with the related assumptions and constraints, and how its information is embedded within diffraction patterns is explained. In Section 3.3, mathematical description of a 4DSTEM dataset is provided. Additionally, in Section 3.4 to 3.7, the algorithms of two non-iterative ptychography methods, Single-Side Band (SSB) and Wigner Distribution Deconvolution (WDD), and an iterative ptychography method, extended Ptychographical Iteration Engine (ePIE) will be discussed, as well as traditional centre-of-mass (COM) methods, illustrating how they extract information from diffraction patterns.

3.2 Source of Information for Ptychography and Centre-Of-Mass

In this section, the sample, also known as the object transmission function, is defined and how its information is represented in a diffraction pattern is explained.

3.2.1 Object Transmission Function and the Sample.

When an electron beam passes through a sample, the interactions are complex, involving both elastic and inelastic scattering. Additionally, multiple scattering events may occur beyond the initial single scattering. Because of these complex interactions, simplifying the sample for calculations is a primary concern in simulations. In this section, the methods used in this thesis to model the sample are introduced.

3.2.1.1 Elastic and Inelastic, Single and Multiple Scattering

Firstly, when an electron beam is interacting with a sample, the scattering can be categorized as either elastic or inelastic, depending on whether the incident electrons lose observable energy during the process. Electrons undergoing elastic scattering lose almost no energy, retaining the same wavelength as the incident beam, with only their propagation direction changing. In contrast, electrons undergoing inelastic scattering lose some energy, which allows them to be detected by an electron energy loss spectrometer. Both types of scattering enable the electron beam to carry information about the sample. In this chapter and in Chapters 5, 6, and 7, only elastic scattering will be considered in calculations. The impact of inelastic scattering on ptychography will be discussed in Chapter 4.

Regardless of whether scattering is elastic or inelastic, electrons may undergo single or multiple scattering events within the sample. The impact of multiple scattering on imaging in a thick sample is referred to as dynamical effects, while single scattering is termed kinematical. To address dynamic effects in imaging simulations, methods such as the Bloch

Wave method [5] and the multislice method [6] have been developed. In this thesis, in principle, calculations consider only thin samples, thereby ignoring dynamic effects, in Chapter 5 and 6 as the Phase Object Approximation (POA) and, subsequently, the Weak Phase Object Approximation (WPOA) are applied. In Chapter 7, the calculations are based on an object function that includes both phase and amplitude variations.

3.2.1.2 Transmission Function

A thin sample, where the propagation effects within the sample can be ignored, can be described using a complex function ψ.

$$EW(\boldsymbol{R}) = P(\boldsymbol{R}) \cdot \psi(\boldsymbol{R}) \tag{3-1}$$

In Equation (3-1), P represents the probe function, which describes the incident beam. ψ is the complex object transmission function, and EW is the exit wave. The transmission function ψ can be expressed in terms of amplitude and phase:

$$\psi(\boldsymbol{R}) = |\psi(\boldsymbol{R})| \cdot \exp\{i \cdot \angle\psi(\boldsymbol{R})\} \tag{3-2}$$

In equation (3-2), $|\psi(\boldsymbol{R})|$ and $\angle\psi(\boldsymbol{R})$ represent the amplitude and phase of ψ, respectively. When the probe function P multiplies the transmission function ψ, it multiplies the amplitude of the probe function by $|\psi(\boldsymbol{R})|$ and shifts the phase of the probe function by $\angle\psi(\boldsymbol{R})$ at each $\boldsymbol{R}$. For a vacuum, $\psi(\boldsymbol{R})$ satisfies $|\psi(\boldsymbol{R})| = 1$ and $\angle\psi(\boldsymbol{R}) = 0$. For a sample, $|\psi(\boldsymbol{R})| \in [0,1]$. The closer $|\psi(\boldsymbol{R})|$ is to 0, the more the sample absorbs the electron beam. For most experimental samples, the amplitude changes are small compared to the phase changes and can be ignored. The simplified transmission function, ignoring the amplitude changes, will be discussed in the following sections.

According to the Aharonov–Bohm effect [7], from a quantum mechanics perspective, the electromagnetic potentials in space alter the electron wave function as the electron beam passes through the sample. In electron microscopy, this is typically expressed as the relationship between the projected potential of the sample V_p and the phase of the

transmission function $\angle\psi(\boldsymbol{R})$, as shown in equation (3-3), assuming that the potential V_p is much smaller than the energy of the accelerated electrons.

$$\angle\psi(\boldsymbol{R}) = \sigma V_p(\boldsymbol{R}) \tag{3-3}$$

σ in Equation (3-3) is an interaction parameter, determined by the wavelength of the incident electron beam:

$$\sigma = \frac{2\pi me\lambda}{h^2} \tag{3-4}$$

In Equation (3-4), the $m = \gamma m_0$ represents the relativistic mass of an electron, where e is the electron charge, λ is wavelength of the electron, h is the Plank's constant.

The projected potential $V_p(\boldsymbol{R})$ is the integral of the 3D potential of the sample, $V_{3d}(R_x, R_y, R_z)$, along the beam axis, typically denoted as the z-axis. It can be expressed as:

$$V_p(\boldsymbol{R}) = V_p(R_x, R_y) = \int V_{3d}(R_x, R_y, R_z) dR_z \tag{3-5}$$

Equations (3-2) to (3-5) explain how the phase of a transmission function is related to the projected potential. As mentioned in Chapter 2, phase of a transmission function directly conveys information from the sample's projected potential. However, since detectors can only detect the intensity distribution of the electron beam I, which is the square of the amplitude, the phase information is lost. This is known as "the phase problem".

3.2.1.3 Phase Object Approximation

The Phase Object Approximation (POA) is used for samples that do not exhibit strong absorption when interacting with an incident electron beam. It is applicable to samples that are thin enough such that the propagation effects within the sample can be ignored. In this approximation, $|\psi(\boldsymbol{R})|$ is assumed to be 1, with no constraints on $\angle\psi(\boldsymbol{R})$. Therefore, a sample under the POA can be described by:

$$\psi_{POA}(\boldsymbol{R}) = \exp[i \cdot \angle\psi(\boldsymbol{R})] = exp[\psi_s(\boldsymbol{R})] \tag{3-6}$$

Where $\psi_s(\boldsymbol{R})$ is a complex function that has only imaginary part.

Simulation of the interaction between an object function and an electron beam is described by the multiplicative object approximation, which assume the exit wave as the product of the incident wave and a complex object function, including both amplitude and phase effects. For POA sample, only the phase of object function influences the exit wave.

3.2.1.4 Weak Phase Object Approximation

The Weak Phase Object Approximation (WPOA) is a more stringent approximation based on the POA. As mentioned in Chapter 2, WPOA transmission function is the first two terms of the Taylors expansion of POA transmission function. So, WPOA transmission function $\psi_{WPOA}(\boldsymbol{R})$ assumes that the phase of the transmission function is sufficiently weak to satisfy equation (3-7).

$$\begin{gathered}\psi_{WPOA}(\boldsymbol{R}) \approx 1 + \psi_s(\boldsymbol{R}) \\ \psi_s(\boldsymbol{R})\ is\ an\ imaginary\ function\ with\ a\ real\ part\ of\ 0\end{gathered} \tag{3-7}$$

$\psi_s(\boldsymbol{R})$ is a small value compared to i. The change of $\psi_s(\boldsymbol{R})$ has a negligible effect on the amplitude of $\psi_{WPOA}(\boldsymbol{R})$ and only affects the angle of $\psi_{WPOA}(\boldsymbol{R})$ within a limited range. Since $\psi_s(\boldsymbol{R})$ is pure imaginary, the Fourier transform of $\psi_s(\boldsymbol{R})$ is conjugate asymmetric.

$$\begin{gathered}\Psi_s(\boldsymbol{K}) = \mathcal{F}\{\psi_s(\boldsymbol{R})\} \\ \Psi_s(\boldsymbol{K}) = -\Psi_s^*(-\boldsymbol{K})\end{gathered} \tag{3-8}$$

Where $\mathcal{F}$ indicates Fourier transform.

The WPOA is an important simplification for calculations and simulations. In Chapters 5 and 6, all calculations and simulations are conducted under the WPOA conditions.

3.2.2 Probe Function and Aberrations

In Equation (3-1), the probe function P is another important factor influencing the exit wave. It is dependent on various microscope settings including acceleration voltage and aberrations. In this section, the influence of these factors on probe function are discussed.

The first parameter is monochromacy. The monochromacy is controlled by the electron gun and the accelerating electrode system. Generally, electrons in the electron beam have an energy distribution arising from factors such as thermal spread at the emission source and voltage instabilities in the accelerating system. As a result, electrons do not all have the same wavelength. This energy distribution reduces the temporal coherence of the electron beam. Additionally, due to the chromatic aberration, monochromacy affects the convergence accuracy of the electron beam. With the help of a cold field-emission gun (cold FEG) or a monochromator, modern transmission electron microscopes can typically achieve an energy spread within 0.3 eV. This intrinsic energy spread leads a defocus spread [8]:

$$\delta_{df} = C_c \cdot \frac{\Delta E}{V_{acc}} \tag{3-9}$$

Where C_c is chromatic aberration, ΔE is energy spread and V_{acc} is accelerate voltage. Equation (3-9) is equivalent to a simplified δ_c of Equation (2-6), which ignores instabilities in the objective lens current and acceleration voltage. As discussed in Chapter 2, δ_c and δ_{df} associated with the envelope function and define the information limit in conventional TEM imaging. This effect becomes more important under low acceleration voltages (below 100 kV). However, in this thesis, to simplify the simulation process and enhance the clarity of the results, δ_{df} is neglected in the simulations presented in the Chapter 5 to Chapter 7.

The second parameter is the source size, or the equivalent electron source size in the TEM. If the source is not a point, it reduces the spatial coherence of the electron beam. In this thesis, the equivalent electron source is assumed to be a point for all calculations and simulations.

The third parameter is the convergence angle, which determines whether the electron beam is parallel (convergence angle is 0) or convergent. The convergence angle is the main difference between conventional TEM mode and STEM mode. In TEM mode, a convergence angle close to 0 (a parallel beam) is used. In STEM mode, the convergence angle varies depending on the experimental setting. Typically, for high-resolution STEM (resolution better than 1 Å), the convergence angle is generally above 20 mrad. The convergence angle is usually limited by aberration, which will be discussed in the next paragraph.

The fourth factor are aberrations. As mentioned in Chapter 2, aberrations play a significant role in electron microscopy. Aberrations prevent the magnetic lens from ideally converging the electron beam. In this chapter, circularly symmetric aberrations are used as examples in the equations for simplicity, although non-circularly symmetric aberrations can be treated using a similar approach. The effect of circularly symmetric aberrations can be described using a phase distortion function on the aperture plane, which is the same plane as the inverse Fourier transform of the probe function $P(R)$. A phase distortion function $\chi(\theta)$ is defined as:

$$\chi(\theta) = \frac{2\pi}{\lambda}\left(\frac{1}{2}C_1\theta^2 + \frac{1}{4}C_3\theta^4 + \frac{1}{6}C_5\theta^6 + \cdots\right) \tag{3-10}$$

Various notations and bases can be used to represent the phase distortion function [9, 10]. While Equation (3-10) can be extended to include higher-order terms and asymmetrical aberrations, this thesis focuses only on the fundamental terms: defocus C_1 and spherical aberration C_3.

As previously mentioned, the phase distortion function χ is located at the aperture plane. Consequently, the aperture function for a beam with aberration can be expressed as:

$$A(\boldsymbol{K}) = A_t(\boldsymbol{K}) \cdot \exp[i\chi(\boldsymbol{K})] \tag{3-11}$$

Where $\boldsymbol{K}$ is the frequency related to the angle $\boldsymbol{\theta}$:

$$\boldsymbol{\theta} = \lambda \boldsymbol{K} \tag{3-12}$$

The most common aperture is a circular aperture that is transparent in the centre and opaque at the edges, which can be written as:

$$A_t(\boldsymbol{K}) = \begin{cases} 0, where\ |\boldsymbol{K}| > \alpha \\ 1, where\ \ |\boldsymbol{K}| \leq \alpha \end{cases} \tag{3-13}$$

α represents the maximum angle of the transparent area and simultaneously serves as the convergence angle of the beam.

The probe function $P(\boldsymbol{R})$ can be determined by performing the inverse Fourier transform of the aperture function $A(\boldsymbol{K})$:

$$P(\boldsymbol{R}) = \mathcal{F}\{A(\boldsymbol{K})\} = \int A_t(\boldsymbol{K}) \cdot exp\big(i \cdot \chi(\boldsymbol{K})\big) \cdot exp(2\pi i \boldsymbol{K} \cdot \boldsymbol{R})\, d\boldsymbol{K} \tag{3-14}$$

In the next section, the phase information contained in the exit wave and at the detector plane will be discussed.

3.2.3 Phase Information in Diffraction Patterns

The exit wave encapsulates all the information about the sample conveyed by the electron beam. As discussed in the previous section, under the conditions of a single wavelength and single scattering, the exit wave can be represented by Equation (3-1). In this section, how this information is represented in the diffraction pattern will be explored.

As discussed in the previous section, the probe function P is influenced by aberrations. For simplicity, only defocus (C_1) is considered at this stage. However, higher-order aberrations can, in principle, be incorporated using the same method.

A phase distortion function considering only defocus can be defined using Equations (3-10) and (3-12) as follows:

$$\chi(\boldsymbol{K}) = \pi \lambda C_1 \boldsymbol{K}^2 \tag{3-15}$$

Next, as described in the previous section, the probe function P can be represented by the aperture function. Using the transmission function ψ, the exit wave can be expressed as:

$$EW(\boldsymbol{R}) = \psi(\boldsymbol{R})P(\boldsymbol{R}) = \psi(\boldsymbol{R}) \int A(\boldsymbol{K})\, exp(2\pi i \boldsymbol{K} \cdot \boldsymbol{R})\, d\boldsymbol{K} \tag{3-16}$$

In Equation (3-16), the movement of the probe function is not considered. In this section, only a single diffraction pattern generated by a stationary probe is considered. Probe scanning will be considered in Section 3.3.

After Equation (3-16), the wave function on the diffraction plane (or detector plane) is then calculated by propagating the exit wave to the diffraction plane using a Fourier transform.

$$\begin{aligned} M(\boldsymbol{K_f}) &= \mathcal{F}\{EW(\boldsymbol{R})\} = \int EW(\boldsymbol{R}) \cdot exp(-2\pi i \boldsymbol{R} \cdot \boldsymbol{K_f}) d\boldsymbol{R} \\ &= \int \left[\psi(\boldsymbol{R}) \int A(\boldsymbol{K})\, exp(2\pi i \boldsymbol{K} \cdot \boldsymbol{R})\, d\boldsymbol{K}\right] \cdot exp(-2\pi i \boldsymbol{R} \cdot \boldsymbol{K_f}) d\boldsymbol{R} \end{aligned} \tag{3-17}$$

Equation (3-16) can be simplified by reordering the integrals:

$$M(\boldsymbol{K_f}) = \int A(\boldsymbol{K}) \int \psi(\boldsymbol{R})\, exp[-2\pi i \boldsymbol{R} \cdot (\boldsymbol{K_f} - \boldsymbol{K})]\, d\boldsymbol{R}\, d\boldsymbol{K} \tag{3-18}$$

The second integral in Equation (3-17) has the same form as the Fourier transform of $\psi(\boldsymbol{R})$. Therefore, it is defined as:

$$\Psi(\boldsymbol{K}) = \int \psi(\boldsymbol{R})\, exp[-2\pi i \boldsymbol{R} \cdot \boldsymbol{K}]\, d\boldsymbol{R} \tag{3-19}$$

By combining Equations (3-18) and (3-19), $M(\boldsymbol{K_f})$ can be represented as a product of convolutions.

$$M(\boldsymbol{K_f}) = \int A(\boldsymbol{K}) \Psi(\boldsymbol{K_f} - \boldsymbol{K}) d\boldsymbol{K} = A(\boldsymbol{K_f}) \otimes \Psi(\boldsymbol{K_f}) \tag{3-20}$$

"$\otimes$" represents convolution operator.

In Equation (3-20), the information at all frequencies in $\Psi(\boldsymbol{K_f})$ is convolved with $A(\boldsymbol{K_f})$. By considering only a single spatial frequency ($\boldsymbol{g}$) and its symmetric spatial frequency ($-\boldsymbol{g}$), $\Psi(\boldsymbol{K_f})$ can be simplified as follows:

$$\Psi_g(\boldsymbol{K_f}) = \Psi(\boldsymbol{K_f}) \cdot \left[\delta(\boldsymbol{K_f}) + \delta(\boldsymbol{K_f} + \boldsymbol{g}) + \delta(\boldsymbol{K_f} - \boldsymbol{g})\right] \tag{3-21}$$

In Equation (3-21), δ represents the Dirac function, and $\boldsymbol{g}$ is a distance in the frequency domain, measured in m^{-1}. This equation can be interpreted as a specific three-beam condition by selecting two symmetric diffraction discs (at $\boldsymbol{g}$ and $-\boldsymbol{g}$) and the central bright disc, while ignoring the effects of other diffraction discs and the intensity difference between these three discs.

For now, when Equation (3-21) is applied, the integral in Equation (3-20) can be simplified to a summation of three terms.

$$\begin{aligned} M(\boldsymbol{K_f}) &= \int A(\boldsymbol{K})\Psi(\boldsymbol{K_f} - \boldsymbol{K})d\boldsymbol{K} \\ &= A(\boldsymbol{K_f})\Psi(\boldsymbol{0}) + A(\boldsymbol{K_f} - \boldsymbol{g})\Psi(\boldsymbol{g}) + A(\boldsymbol{Kf} + \boldsymbol{g})\Psi(-\boldsymbol{g}) \end{aligned} \tag{3-22}$$

The intensity distribution at the diffraction plane can be calculated by the modulus square of $M(\boldsymbol{K_f})$, which is also the signal recorded by the detector.

$$\begin{aligned} \left|M(\boldsymbol{K_f})\right|^2 &= M(\boldsymbol{K_f}) \cdot M^*(\boldsymbol{K_f}) \\ &= \left[A(\boldsymbol{K_f})\Psi(\boldsymbol{0}) + A(\boldsymbol{K_f} - \boldsymbol{g})\Psi(\boldsymbol{g}) + A(\boldsymbol{Kf} + \boldsymbol{g})\Psi(-\boldsymbol{g})\right] \\ &\cdot \left[A^*(\boldsymbol{K_f})\Psi^*(\boldsymbol{0}) + A^*(\boldsymbol{K_f} - \boldsymbol{g})\Psi^*(\boldsymbol{g}) + A^*(\boldsymbol{Kf} + \boldsymbol{g})\Psi^*(-\boldsymbol{g})\right] \end{aligned} \tag{3-23}$$

There are nine terms in the expansion of Equation (3-23). Three terms are constant items:

$$\left|A(\boldsymbol{K_f})\right|^2|\Psi(\boldsymbol{0})|^2 + \left|A(\boldsymbol{K_f} - \boldsymbol{g})\right|^2|\Psi(\boldsymbol{g})|^2 + \left|A(\boldsymbol{K_f} + \boldsymbol{g})\right|^2|\Psi(-\boldsymbol{g})|^2 \tag{3-24}$$

The constant terms in Equation (3-24) represent the background intensity of the bright field disc and the diffraction discs. But they do not contribute to the interference fringes. Additionally, the other six terms in the expansion of Equation (3-23) can be divided into

three groups. Two of them, as shown in Equation (3-25) and (3-26) correspond to the interference fringes with defocus in the overlapping areas of the bright disc and the diffraction discs:

$$A(\boldsymbol{K_f})A^*(\boldsymbol{K_f}-\boldsymbol{g})\Psi(\boldsymbol{0})\Psi^*(\boldsymbol{g}) + A(\boldsymbol{K_f}-\boldsymbol{g})A^*(\boldsymbol{K_f})\Psi(\boldsymbol{g})\Psi^*(\boldsymbol{0}) \tag{3-25}$$

$$A(\boldsymbol{K_f})A^*(\boldsymbol{K_f}+\boldsymbol{g})\Psi(\boldsymbol{0})\Psi^*(-\boldsymbol{g}) + A(\boldsymbol{K_f}+\boldsymbol{g})A^*(\boldsymbol{K_f})\Psi(-\boldsymbol{g})\Psi^*(\boldsymbol{0}) \tag{3-26}$$

The last group also includes two terms, which correspond to the interference fringes between the diffraction discs:

$$A(\boldsymbol{K_f}-\boldsymbol{g})A^*(\boldsymbol{K_f}+\boldsymbol{g})\Psi(\boldsymbol{g})\Psi^*(-\boldsymbol{g}) + A(\boldsymbol{K_f}+\boldsymbol{g})A^*(\boldsymbol{K_f}-\boldsymbol{g})\Psi(-\boldsymbol{g})\Psi^*(\boldsymbol{g}) \tag{3-27}$$

Equation (3-25) to (3-27) can be further simplified. For example, consider the first term in (3-25):

$$\begin{aligned} &A(\boldsymbol{K_f})A^*(\boldsymbol{K_f}-\boldsymbol{g})\Psi(\boldsymbol{0})\Psi^*(\boldsymbol{g}) \\ &\quad = |\Psi(\boldsymbol{0})||\Psi(\boldsymbol{g})|A_t(\boldsymbol{K_f})A_t(\boldsymbol{K_f}-\boldsymbol{g}) \\ &\quad \cdot exp[i\angle\Psi(\boldsymbol{0})]\, exp[-i\angle\Psi(\boldsymbol{g})]\, exp\big[i\chi(\boldsymbol{K_f})\big]\, exp\big[-i\chi(\boldsymbol{K_f}-\boldsymbol{g})\big] \\ &\quad = |\Psi(\boldsymbol{0})||\Psi(\boldsymbol{g})|A_t(\boldsymbol{K_f})A_t(\boldsymbol{K_f}-\boldsymbol{g}) \\ &\quad \cdot \begin{Bmatrix} \cos\Big(\angle\Psi(\boldsymbol{0}) - \angle\Psi(\boldsymbol{g}) + \chi(\boldsymbol{K_f}) - \chi(\boldsymbol{K_f}-\boldsymbol{g})\Big) + \\ i\cdot\sin\Big(\angle\Psi(\boldsymbol{0}) - \angle\Psi(\boldsymbol{g}) + \chi(\boldsymbol{K_f}) - \chi(\boldsymbol{K_f}-\boldsymbol{g})\Big) \end{Bmatrix} \end{aligned} \tag{3-28}$$

Similar presentation for the second term (3-25):

$$\begin{aligned} &A(\boldsymbol{K_f}-\boldsymbol{g})A^*(\boldsymbol{K_f})\Psi(\boldsymbol{g})\Psi^*(\boldsymbol{0}) \\ &\quad = |\Psi(\boldsymbol{g})||\Psi(\boldsymbol{0})|A_t(\boldsymbol{K_f}-\boldsymbol{g})A_t(\boldsymbol{K_f}) \\ &\quad \cdot \begin{Bmatrix} \cos\Big(\angle\Psi(\boldsymbol{g}) - \angle\Psi(\boldsymbol{0}) + \chi(\boldsymbol{K_f}-\boldsymbol{g}) - \chi(\boldsymbol{K_f})\Big) + \\ i\cdot\sin\Big(\angle\Psi(\boldsymbol{g}) - \angle\Psi(\boldsymbol{0}) + \chi(\boldsymbol{K_f}-\boldsymbol{g}) - \chi(\boldsymbol{K_f})\Big) \end{Bmatrix} \end{aligned} \tag{3-29}$$

Equation (3-25) can be written as a summation of Equation (3-28) and (3-29):

$$\begin{aligned} &A(\boldsymbol{K_f})A^*(\boldsymbol{K_f}-\boldsymbol{g})\Psi(\boldsymbol{0})\Psi^*(\boldsymbol{g}) + A(\boldsymbol{K_f}-\boldsymbol{g})A^*(\boldsymbol{K_f})\Psi(\boldsymbol{g})\Psi^*(\boldsymbol{0}) \\ &\quad = 2|\Psi(\boldsymbol{0})||\Psi(\boldsymbol{g})|A_t(\boldsymbol{K_f})A_t(\boldsymbol{K_f}-\boldsymbol{g}) \\ &\quad \cdot cos\left(\angle\Psi(\boldsymbol{0}) - \angle\Psi(\boldsymbol{g}) + \chi(\boldsymbol{K_f}) - \chi(\boldsymbol{K_f}-\boldsymbol{g})\right) \end{aligned} \tag{3-30}$$

Equation (3-30) describes the fringes in the overlapping area of diffraction patterns. By combining Equation (3-30) with Equation (3-15), the cosine term can be expanded to calculate the period of the interference fringes. In the cosine term, the terms related to $\chi(\boldsymbol{K_f})$ can be writen as:

$$\chi(\boldsymbol{K_f}) - \chi(\boldsymbol{K_f}-\boldsymbol{g}) = 2\pi\lambda C_1|\boldsymbol{K_f}||\boldsymbol{g}|\,cos[\angle(\boldsymbol{K_f},\boldsymbol{g})] - \pi\lambda C_1|\boldsymbol{g}|^2 \tag{3-31}$$

Where $\angle(\boldsymbol{K_f},\boldsymbol{g})$ represents the angle between vector $\boldsymbol{K_f}$ and vector $\boldsymbol{g}$. The term $|\boldsymbol{K_f}||\boldsymbol{g}|\,cos[\angle(\boldsymbol{K_f},\boldsymbol{g})]$ in Equation (3-31) equals the dot product of vector $\boldsymbol{K_f}$ and vector $\boldsymbol{g}$, *i.e.* $\boldsymbol{K_f}\cdot\boldsymbol{g}$.

Combing with Equation (3-31), Equation (3-30) can be expressed as a real cosine function with respect to $\boldsymbol{K_f}$, as shown in Equation (3-32):

$$\begin{aligned} &2|\Psi(\boldsymbol{0})||\Psi(\boldsymbol{g})|A_t(\boldsymbol{K_f})A_t(\boldsymbol{K_f}-\boldsymbol{g}) \\ &\quad \cdot cos\left(2\pi\lambda C_1|\boldsymbol{g}|\,cos[\angle(\boldsymbol{K_f},\boldsymbol{g})]\,|\boldsymbol{K_f}| - \pi\lambda C_1|\boldsymbol{g}|^2 + \angle\Psi(\boldsymbol{0}) - \angle\Psi(\boldsymbol{g})\right) \end{aligned} \tag{3-32}$$

Therefore, the period of the interference fringes can be calculated as follows:

$$\frac{1}{\lambda C_1|\boldsymbol{g}|\,cos[\angle(\boldsymbol{K_f},\boldsymbol{g})]} \tag{3-33}$$

Equation (3-32) can be illuminated on the $\boldsymbol{K_f}$ plane, as shown in Figure 3-1. In this figure, the orange circles represent the regions where $A_t(\boldsymbol{K_f})$ is non-zero. Similarly, the green and purple regions represent the areas where $A_t(\boldsymbol{K_f}-\boldsymbol{g})$ and $A_t(\boldsymbol{K_f}+\boldsymbol{g})$ are non-zero, respectively. In Equation (3-32), the cosine function can only be non-zero when both $A_t(\boldsymbol{K_f})$ and $A_t(\boldsymbol{K_f}-\boldsymbol{g})$ are non-zero, exhibiting the period expressed in Equation (3-33). For simplicity, $\boldsymbol{K_f}$ and $\boldsymbol{g}$ are chosen to be in the same direction in this figure.

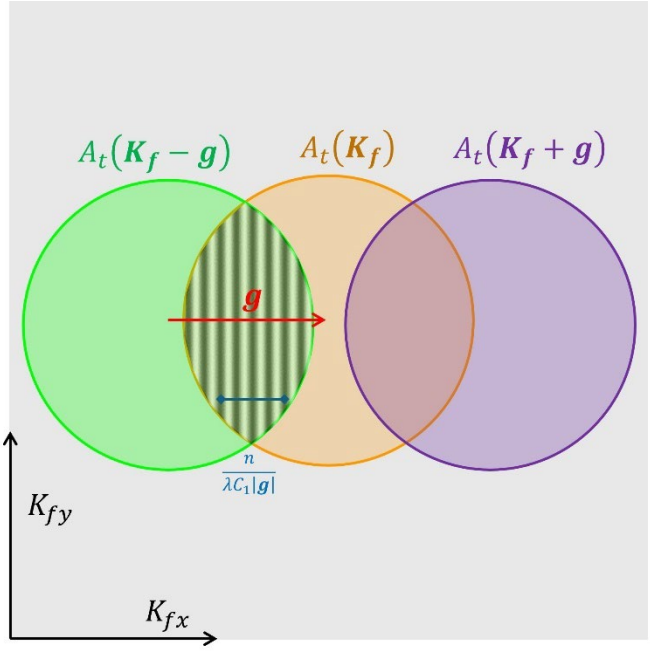

Figure 3-1 | Schematic figure of interference fringes on the overlapping area of diffraction discs. $\boldsymbol{K_f}$ and $\boldsymbol{g}$ are chosen to be in the same direction in this figure ($cos[\angle(\boldsymbol{K_f},\boldsymbol{g})]=1$). Therefore, the n times period of interference fringes is marked as $\frac{n}{\lambda C_1|\boldsymbol{g}|}$.

Using a similar derivation process, equations similar to (3-32) from (3-26) and (3-27) can be derived as:

$$\begin{aligned} &A(\boldsymbol{K_f})A^*(\boldsymbol{K_f}+\boldsymbol{g})\Psi(\boldsymbol{0})\Psi^*(-\boldsymbol{g})+A(\boldsymbol{K_f}+\boldsymbol{g})A^*(\boldsymbol{K_f})\Psi(-\boldsymbol{g})\Psi^*(\boldsymbol{0}) \\ &\quad=2|\Psi(\boldsymbol{0})||\Psi(-\boldsymbol{g})|A_t(\boldsymbol{K_f})A_t(\boldsymbol{K_f}+\boldsymbol{g}) \\ &\quad\cdot cos\left(2\pi\lambda C_1|\boldsymbol{g}|\,cos[\angle(\boldsymbol{K_f},\boldsymbol{g})]\,|\boldsymbol{K_f}|+\pi\lambda C_1|\boldsymbol{g}|^2-\angle\Psi(\boldsymbol{0})+\angle\Psi(-\boldsymbol{g})\right) \end{aligned} \tag{3-34}$$

$$\begin{aligned} &A(\boldsymbol{K_f}-\boldsymbol{g})A^*(\boldsymbol{K_f}+\boldsymbol{g})\Psi(\boldsymbol{g})\Psi^*(-\boldsymbol{g})+A(\boldsymbol{K_f}+\boldsymbol{g})A^*(\boldsymbol{K_f}-\boldsymbol{g})\Psi(-\boldsymbol{g})\Psi^*(\boldsymbol{g}) \\ &\quad=2|\Psi(\boldsymbol{g})||\Psi(-\boldsymbol{g})|A_t(\boldsymbol{K_f}-\boldsymbol{g})A_t(\boldsymbol{K_f}+\boldsymbol{g}) \\ &\quad\cdot cos\left(4\pi\lambda C_1|\boldsymbol{g}|\,cos[\angle(\boldsymbol{K_f},\boldsymbol{g})]\,|\boldsymbol{K_f}|-\angle\Psi(\boldsymbol{g})+\angle\Psi(-\boldsymbol{g})\right) \end{aligned} \tag{3-35}$$

For Equations (3-32), (3-34), and (3-35), the common condition for producing interference fringes is the overlap of the corresponding diffraction discs. Once the overlap occurs, the resulting interference fringes contain the amplitude and phase information of Ψ. Therefore, by processing the diffraction patterns, it is possible to recover the information in Ψ and

subsequently calculate the transmission function ψ.

Furthermore, the interference fringes are related to aberrations. Although, for simplicity, in this section only defocus (C_1) is considered. In the next Chapter, the relation between defocus and interference fringes will be further discussed.

In the Section 3.4, 3.5, 3.6 and 3.7, detailed methods to retrieve information in transmission ψ from a dataset of scanning diffraction patterns will be introduced. Before that, in the next section, 4DSTEM as a data acquisition method of scanning diffraction patterns will be discussed in the next section.

3.3 4DSTEM

4DSTEM is a data acquisition method. It refers to the process in transmission electron microscopy where, while scanning the sample with an electron beam, a pixelated detector records the diffraction pattern at each scanning point. The resulting data includes two coordinate positions in the sample plane (scanning plane) and two coordinate positions in the diffraction plane (detector plane), as shown in Figure 3-2. Therefore, the data can be considered to have four dimensions, hence the name 4DSTEM. 4DSTEM combines scanning of the probe and corresponding diffraction, has been widely used on different materials systems. A range of information can be extracted from 4DSTEM dataset by different analysis methods [11]. This chapter focuses on phase contrast imaging methods based on 4DSTEM dataset.

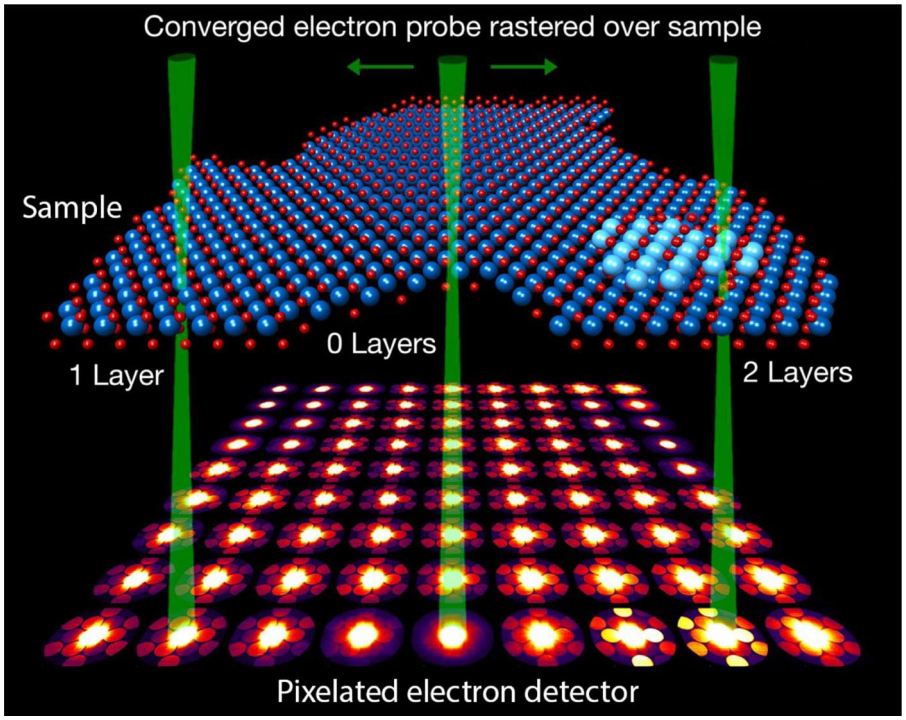

Figure 3-2 | Schematic diagram of 4DSTEM [11]. The probe scans sample in a 2D array of probe positions, while 2D diffraction patterns are recorded on diffraction plane. As a result, the dataset has 4 dimensions. Reprinted with permission from [11]. Copyright 2019 by Microscopy Society of America.

3.3.1 Mathematical description of a 4DSTEM dataset

To describe a 4DSTEM dataset, in addition to considering the intensity distribution in the diffraction plane, it is also necessary to consider the scanning of the probe in the sample plane. Therefore, when considering the interaction between the sample and the probe, an additional variable of probe position $\boldsymbol{R_p}$ can be included. So, Equation (3-1) can be rewritten with $\boldsymbol{R_p}$ as:

$$EW(\boldsymbol{R}, \boldsymbol{R_p}) = P(\boldsymbol{R} - \boldsymbol{R_p}) \cdot \psi(\boldsymbol{R}) \tag{3-36}$$

In Equation (3-36), $\boldsymbol{R}$ represents the coordinates on the sample plane, while $\boldsymbol{R_p}$ denotes the shift in the probe position. Discrete sampling of $\boldsymbol{R_p}$ describes the scanning process of the probe.

The exit wave is propagated to the diffraction plane by a Fourier transform with respect to $\boldsymbol{R}$. The wave function on diffraction plane is defined as $M(\boldsymbol{K_f}, \boldsymbol{R_p})$. Furthermore, $M(\boldsymbol{K_f}, \boldsymbol{R_p})$ can be represented as a product of a convolution, as shown in Equation (3-37).

$$M(\boldsymbol{K_f}, \boldsymbol{R_p}) = \mathcal{F}\{EW(\boldsymbol{R}, \boldsymbol{R_p})\} = \int P(\boldsymbol{R} - \boldsymbol{R_p}) \cdot \psi(\boldsymbol{R}) \cdot \exp(-2\pi i \boldsymbol{R} \cdot \boldsymbol{K_f})\, d\boldsymbol{R}$$
$$= \left[A(\boldsymbol{K_f})\, exp(-2\pi i \boldsymbol{K_f} \cdot \boldsymbol{R_p})\right] \otimes_{\boldsymbol{K_f}} \Psi(\boldsymbol{K_f}) \tag{3-37}$$

Where $\otimes_{\boldsymbol{K_f}}$ represents convolution with respect to $\boldsymbol{K_f}$.

The intensity distribution on the detector plane can be calculated as [12]:

$$\left|M(\boldsymbol{K_f}, \boldsymbol{R_p})\right|^2 =$$
$$\iint P(\boldsymbol{t_1} - \boldsymbol{R_p}) P^*(\boldsymbol{t_2} - \boldsymbol{R_p}) \psi(\boldsymbol{t_1}) \psi^*(\boldsymbol{t_2})\, exp\left[2\pi i (\boldsymbol{t_1} \cdot \boldsymbol{K_f} - \boldsymbol{t_2} \cdot \boldsymbol{K_f})\right] d\boldsymbol{t_1} d\boldsymbol{t_2} \tag{3-38}$$

In Equation (3-38), $\boldsymbol{t_1}$ and $\boldsymbol{t_2}$ are dummy variables.

Alternatively, Equation (3-38) can be further rewritten using the convolution theorem of Fourier transforms of $P(\boldsymbol{R} - \boldsymbol{R_p})$ and $\psi(\boldsymbol{R})$, which can also be derived from the convolution in Equation (3-37) [12]:

$$\left|M(\boldsymbol{K_f}, \boldsymbol{R_p})\right|^2 =$$
$$\iint A(\boldsymbol{t_1}) A^*(\boldsymbol{t_2}) \Psi(\boldsymbol{K_f} - \boldsymbol{t_1}) \Psi^*(\boldsymbol{K_f} - \boldsymbol{t_2})\, exp\left[2\pi i \boldsymbol{R_p} \cdot (\boldsymbol{t_1} - \boldsymbol{t_2})\right] d\boldsymbol{t_1} d\boldsymbol{t_2} \tag{3-39}$$

The intensity distribution $\left|M(\boldsymbol{K_f}, \boldsymbol{R_p})\right|^2$ is the actual signal recorded the detector at the diffraction plane. Therefore, $\left|M(\boldsymbol{K_f}, \boldsymbol{R_p})\right|^2$ can be regarded as the 4DSTEM dataset recorded in an experiment. In the following sections, several algorithms to reconstruct ψ from $\left|M(\boldsymbol{K_f}, \boldsymbol{R_p})\right|^2$ will be discussed.

$\left|M(\boldsymbol{K_f}, \boldsymbol{R_p})\right|^2$ can be Fourier transformed with respect to $\boldsymbol{R_p}$. This results in a 4D data array defined as $G(\boldsymbol{K_f}, \boldsymbol{Q_p})$ [13].

$$\begin{aligned}\mathrm{G}(\boldsymbol{K_f}, \boldsymbol{Q_p}) &= \mathcal{F}_{\boldsymbol{R_p}}\left\{\left|M(\boldsymbol{K_f}, \boldsymbol{R_p})\right|^2\right\} \\ &= A(\boldsymbol{K_f})A^*(\boldsymbol{K_f}+\boldsymbol{Q_p}) \otimes_{\boldsymbol{K_f}} \Psi(\boldsymbol{K_f})\Psi^*(\boldsymbol{K_f}-\boldsymbol{Q_p})\end{aligned} \tag{3-40}$$

$\mathrm{G}(\boldsymbol{K_f}, \boldsymbol{Q_p})$ in Equation (3-40) plays an important role in various reconstruction methods, including Wigner Distribution Deconvolution (WDD) and Single-side band (SSB), which will be detailed in Section 3.5, 3.6, as well as the SSBm-COM method that will be discussed in Chapter 5.

3.3.2 Phase Information in a 4DSTEM dataset

The phase information in a 4DSTEM dataset can be described using the same process as in Section 3.2.3. The only difference is that the probe is scanning. It scans on the sample as the new variable $\boldsymbol{R_p}$ is introduced. The introduction of $\boldsymbol{R_p}$ changes the convolution form of M from Equation (3-20) to Equation (3-37), which is an additional $exp(-2\pi i \boldsymbol{K_f} \cdot \boldsymbol{R_p})$ factor multiplied by $A(\boldsymbol{K_f})$. As a result, the same derivation process as in Equations (3-20) to (3-32) can be extended to the 4DSTEM dataset with the additional factor on $A(\boldsymbol{K_f})$. Therefore, the fringes in the overlapping area of centre disc and a diffraction disc corresponding to a spatial frequency $\boldsymbol{g}$ can be expressed based on Equation (3-30) as:

$$\begin{aligned}&A(\boldsymbol{K_f})\, exp(-2\pi i \boldsymbol{K_f} \cdot \boldsymbol{R_p})\, A^*(\boldsymbol{K_f}-\boldsymbol{g})\, exp(2\pi i(\boldsymbol{K_f}-\boldsymbol{g}) \cdot \boldsymbol{R_p})\, \Psi(\boldsymbol{0})\Psi^*(\boldsymbol{g}) \\ &+A(\boldsymbol{K_f}-\boldsymbol{g})\, exp(-2\pi i(\boldsymbol{K_f}-\boldsymbol{g}) \cdot \boldsymbol{R_p})\, A^*(\boldsymbol{K_f})\, exp(2\pi i \boldsymbol{K_f} \cdot \boldsymbol{R_p})\, \Psi(\boldsymbol{g})\Psi^*(\boldsymbol{0}) \\ &= 2|\Psi(\boldsymbol{0})||\Psi(\boldsymbol{g})|A_t(\boldsymbol{K_f})A_t(\boldsymbol{K_f}-\boldsymbol{g}) \\ &\qquad \cdot cos(\angle\Psi(\boldsymbol{0}) - \angle\Psi(\boldsymbol{g}) + \chi(\boldsymbol{K_f}) - \chi(\boldsymbol{K_f}-\boldsymbol{g}) - 2\pi \boldsymbol{g} \cdot \boldsymbol{R_p})\end{aligned} \tag{3-41}$$

Comparing Equation (3-41) to the corresponding term in Equation (3-30), introducing $\boldsymbol{R_p}$ does not change the period of the interference fringes with respect to $\boldsymbol{K_f}$, but adds an extra phase shift depending on the dot product of $\boldsymbol{g}$ and $\boldsymbol{R_p}$. The results discussed in Section 3.2.3, Equation (3-32), (3-34) and (3-35), can also be applied to the 4DSTEM condition after a similar correction as in Equation (3-41).

3.4 Algorithm of Wigner Distribution Deconvolution

In Chapter 2, the application and features of the non-iterative ptychography were introduced. In this section, one of the non-iterative ptychography algorithm, Wigner Distribution Deconvolution (WDD) method, will be discussed in detail.

The first step of WDD is to perform an inverse Fourier transforms on $\mathrm{G}(\boldsymbol{K_f}, \boldsymbol{Q_p})$ with respect to $\boldsymbol{K_f}$. The result is denoted $H(\boldsymbol{R}, \boldsymbol{Q_p})$ [14, 15]:

$$\begin{aligned} H(\boldsymbol{R}, \boldsymbol{Q_p}) &= \mathcal{F}^{-1}_{\boldsymbol{K_f}}\{\mathrm{G}(\boldsymbol{K_f}, \boldsymbol{Q_p})\} \\ &= \int P^*(\boldsymbol{t_1}) P(\boldsymbol{t_1} + \boldsymbol{R}) \exp(-2\pi i \boldsymbol{Q_p} \cdot \boldsymbol{t_1})\, d\boldsymbol{t_1} \\ &\cdot \int \psi^*(\boldsymbol{t_2}) \psi(\boldsymbol{t_2} + \boldsymbol{R}) \exp(2\pi i \boldsymbol{Q_p} \cdot \boldsymbol{t_2})\, d\boldsymbol{t_2} \end{aligned} \tag{3-42}$$

In Equation (3-42), $H(\boldsymbol{R}, \boldsymbol{Q_p})$ is expressed as a product of two integrals. It is important to note that, after this transformation, neither the integral term includes cross terms of P and ψ; each integral term is related to P or ψ respectively. These two integral terms can be regarded as Wigner distribution functions which take the form:

$$\chi_q(\boldsymbol{a}, \boldsymbol{b}) = \int q^*(\boldsymbol{c}) q(\boldsymbol{c} + \boldsymbol{a}) \exp(2\pi i \boldsymbol{c} \cdot \boldsymbol{b})\, d\boldsymbol{c} \tag{3-43}$$

Combining Equation (3-42) and (3-43), the $H(\boldsymbol{R}, \boldsymbol{Q_p})$ can be rewritten as:

$$H(\boldsymbol{R}, \boldsymbol{Q_p}) = \chi_P(\boldsymbol{R}, -\boldsymbol{Q_p}) \cdot \chi_\psi(\boldsymbol{R}, \boldsymbol{Q_p}) \tag{3-44}$$

In Equation (3-44), the $H(\boldsymbol{R}, \boldsymbol{Q_p})$, representing the measurement, is calculated from the 4DSTEM dataset. $\chi_P(\boldsymbol{R}, -\boldsymbol{Q_p})$ is based on the probe function. Assuming the probe function is known (including aberrations), $\chi_\psi(\boldsymbol{R}, \boldsymbol{Q_p})$ can be determined from $H(\boldsymbol{R}, \boldsymbol{Q_p})$ and $\chi_P(\boldsymbol{R}, -\boldsymbol{Q_p})$ using Wiener deconvolution [14]:

$$\chi_{\psi}(\boldsymbol{R}, \boldsymbol{Q_p}) = \frac{\chi_P^*(\boldsymbol{R}, -\boldsymbol{Q_p}) H(\boldsymbol{R}, \boldsymbol{Q_p})}{\left|\chi_P(\boldsymbol{R}, -\boldsymbol{Q_p})\right|^2 + \epsilon} \tag{3-45}$$

The parameter ϵ in Equation (3-45) is included to prevent the denominator from being zero. It is usually set by trial and error as a relatively small value compared to $\left|\chi_P(\boldsymbol{R}, -\boldsymbol{Q_p})\right|^2$.

By equation (3-45), $\chi_{\psi}(\boldsymbol{R}, \boldsymbol{Q_p})$ is calculated. Theoretically, the aberration in the dataset is compensated by $\chi_P(\boldsymbol{R}, -\boldsymbol{Q_p})$ as well. Furthermore, $\chi_{\psi}(\boldsymbol{R}, \boldsymbol{Q_p})$ can be transformed as follows [14, 15]:

$$\Phi(\boldsymbol{K_f}, \boldsymbol{Q_p}) = \Psi(\boldsymbol{K_f}) \Psi^*(\boldsymbol{K_f} - \boldsymbol{Q_p}) = \mathcal{F}_{\boldsymbol{K_f}}\{\chi_{\psi}(\boldsymbol{R}, \boldsymbol{Q_p})\} \tag{3-46}$$

Here, $\Phi(\boldsymbol{K_f}, \boldsymbol{Q_p})$ is a complex 4D dataset based on Ψ, the Fourier transform of the transmission function. To solve for Ψ, a straightforward method is to extract a slice at $\boldsymbol{K_f} = \boldsymbol{0}$ from the 4D dataset $\Phi(\boldsymbol{K_f}, \boldsymbol{Q_p})$. Then $\Psi_{WDD}(\boldsymbol{Q_p})$ can be expressed as:

$$\Psi_{WDD}(\boldsymbol{Q_p}) = \frac{\Phi^*(\boldsymbol{0}, \boldsymbol{Q_p})}{\sqrt{\Phi(\boldsymbol{0}, \boldsymbol{0})}} \tag{3-47}$$

The transmission function $\psi_{WDD}(\boldsymbol{R_p})$ can be then calculated by performing inverse Fourier transform on $\Psi_{WDD}(\boldsymbol{Q_p})$, as in Equation (3-48).

$$\psi_{WDD}(\boldsymbol{R_p}) = \mathcal{F}^{-1}\{\Psi_{WDD}(\boldsymbol{Q_p})\} \tag{3-48}$$

In the traditional WDD algorithm, the resolution of $\psi_{WDD}(\boldsymbol{R_p})$ is limited to 2α [12, 15]. To achieve higher resolution, A strategy called “stepping out” is proposed, which further utilizes not only the $\boldsymbol{K_f} = \boldsymbol{0}$ in $\Phi(\boldsymbol{K_f}, \boldsymbol{Q_p})$ as in Equation (3-47) [12].

To calculate $H(\boldsymbol{R}, \boldsymbol{Q_p})$ in Equation (3-44), probe function P is assumed to be known. However, in actual experiments, it is often difficult to directly obtain precise aberration information, including defocus. To address this issue, the redundancy in the phase of the

4DSTEM data can be used to estimate the phase and defocus [16], thereby constructing the probe function P from the data.

In the WDD algorithm, sample is considered as a single-layer object function interacted with probe function under multiplicative object approximation. It is not required for the sample to satisfy POA or WPOA condition. In the next section, another ptychography algorithm, SSB, which is simpler than WDD but relies on WPOA, will be discussed.

3.5 Algorithm of Single-side Band

As discussed in Chapter 2, the Single-side Band method is another non-iterative ptychography method based on the same 4DSTEM dataset as WDD. However, it requires the sample to satisfy the WPOA condition.

The SSB algorithm starts from $\mathrm{G}(\boldsymbol{K_f}, \boldsymbol{Q_p})$ as shown in Equation (3-40). Furthermore, if a WPOA condition, *i.e.* Equation (3-7) and (3-8), is applied on Equation (3-40), the convolution in $\mathrm{G}(\boldsymbol{K_f}, \boldsymbol{Q_p})$ can be rewritten as three terms [17]:

$$\begin{aligned}\mathrm{G}(\boldsymbol{K_f}, \boldsymbol{Q_p}) \approx |A(\boldsymbol{K_f})|^2 \delta(\boldsymbol{Q_p}) + A(\boldsymbol{K_f}) A^*(\boldsymbol{K_f} + \boldsymbol{Q_p}) \Psi_s^*(-\boldsymbol{Q_p}) \\ + A^*(\boldsymbol{K_f}) A(\boldsymbol{K_f} - \boldsymbol{Q_p}) \Psi_s(\boldsymbol{Q_p})\end{aligned} \tag{3-49}$$

In Equation (3-49), the first term contains a Dirac delta function, which results in no contribution to $\mathrm{G}(\boldsymbol{K_f}, \boldsymbol{Q_p})$ when $\boldsymbol{Q_p}$ is not zero. Correspondingly, the second and third terms directly provide information about Ψ_s.

Taking term $A^*(\boldsymbol{K_f}) A(\boldsymbol{K_f} - \boldsymbol{Q_p}) \Psi_s(\boldsymbol{Q_p})$ as an example. $A^*(\boldsymbol{K_f})$ and $A(\boldsymbol{K_f} - \boldsymbol{Q_p})$ can be considered as translations in $\boldsymbol{K_f}$ space. After translation, the region where their product is non-zero is called the overlap region. When only 2 translated aperture functions (or its conjugate) overlap, this region is referred to as the double overlap (DO), as shown by the hatched area in Figure 3-3. However, under different $\boldsymbol{Q_p}$ conditions, there may be regions where $A(\boldsymbol{K_f})$, $A(\boldsymbol{K_f} + \boldsymbol{Q_p})$, and $A(\boldsymbol{K_f} - \boldsymbol{Q_p})$, or their conjugates, all overlap. Such regions are called triple overlap (TO) region, as indicated by the grid-shaded area in Figure 3-3.

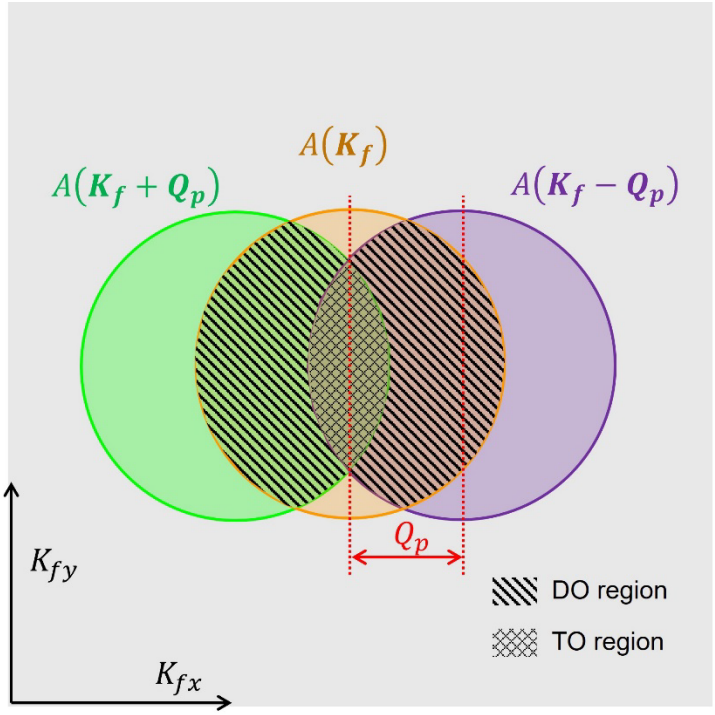

Figure 3-3 | Schematic diagram of TO and DO regions. The yellow circle indicates $A(\boldsymbol{K_f})$. Purple and green circles indicate $A(\boldsymbol{K_f}+\boldsymbol{Q_p})$, and $A(\boldsymbol{K_f}-\boldsymbol{Q_p})$ that are translated $+\boldsymbol{Q_p}$ and $-\boldsymbol{Q_p}$ from $A(\boldsymbol{K_f})$. The DO and TO regions are marked as hatched areas and grid-shaded area.

Obviously, the data in the DO region contributes to Equation (3-49). Due to the WPOA condition, Ψ_s satisfies conjugate antisymmetry, as shown in Equation (3-8). Therefore, the information in the two DO regions is essentially equivalent. For the TO region, the second and third terms in Equation (3-49) can be simplified as one term: $[A^*(\boldsymbol{K_f})A(\boldsymbol{K_f}-\boldsymbol{Q_p}) - A(\boldsymbol{K_f})A^*(\boldsymbol{K_f}+\boldsymbol{Q_p})]\cdot\Psi_s(\boldsymbol{Q_p})$. When the 4DSTEM data is free of aberrations, a simplified expression for TO region gives exact cancellation, resulting the TO region being zero. Consequently, in the SSB method, only one side of the DO region is used to calculate Ψ_s, which is the origin of the term “single-side.”

In Figure 3-3, when $\boldsymbol{Q_p} > 2\alpha$, there will be no overlapping area, DO or TO regions, between $A(\boldsymbol{K_f})$ and $A(\boldsymbol{K_f}+\boldsymbol{Q_p})$ or $A(\boldsymbol{K_f}-\boldsymbol{Q_p})$. As a result, the factor of $\Psi_s(\boldsymbol{Q_p})$ in $\mathrm{G}(\boldsymbol{K_f},\boldsymbol{Q_p})$ will be zero, meaning no information will be transferred. This limits the resolution of SSB.

After calculating $\mathrm{G}(\boldsymbol{K_f}, \boldsymbol{Q_p})$, masks $m_{+\boldsymbol{Q_p}}(\boldsymbol{K_f}, \boldsymbol{Q_p})$ and $m_{-\boldsymbol{Q_p}}(\boldsymbol{K_f}, \boldsymbol{Q_p})$, as in Equation (3-50) can be computed based on the known aperture function $A(\boldsymbol{K_f})$. This mask corresponds to the conjugate of the aperture function and its translated terms that form the DO and TO regions.

$$
\begin{aligned}
m_{+\boldsymbol{Q_p}}(\boldsymbol{K_f}, \boldsymbol{Q_p}) &= \begin{cases} A^*(\boldsymbol{K_f})A(\boldsymbol{K_f} + \boldsymbol{Q_p}), & within\ the + \boldsymbol{Q_p}\ beam\ DO\ region \\ 0, & outside\ the + \boldsymbol{Q_p}\ beam\ DO\ region \end{cases} \\
m_{-\boldsymbol{Q_p}}(\boldsymbol{K_f}, \boldsymbol{Q_p}) &= \begin{cases} A(\boldsymbol{K_f})A^*(\boldsymbol{K_f} - \boldsymbol{Q_p}), & within\ the - \boldsymbol{Q_p}\ beam\ DO\ region \\ 0, & outside\ the - \boldsymbol{Q_p}\ beam\ DO\ region \end{cases}
\end{aligned}
\tag{3-50}
$$

After multiplying $\mathrm{G}(\boldsymbol{K_f}, \boldsymbol{Q_p})$ and $m(\boldsymbol{K_f}, \boldsymbol{Q_p})$, since the DO regions on two sides do not overlap, only one side is considered for now. Taking the $-\boldsymbol{Q_p}$ term as an example, the $\mathrm{G}(\boldsymbol{K_f}, \boldsymbol{Q_p})$ overlap with $m_{-\boldsymbol{Q_p}}$ is denoted as $G_{s,-\boldsymbol{Q_p}}$. Obviously, the product of $G_{s,-\boldsymbol{Q_p}}$ and $m_{-\boldsymbol{Q_p}}$ is not zero only if $\boldsymbol{K_f}$ is in the $-\boldsymbol{Q_p}$ beam DO region (the right DO region in Figure 3-3). This condition is represented as $\boldsymbol{K_f} \in -\boldsymbol{Q_p}\ DO$.

$$
\begin{aligned}
&G_{s,-\boldsymbol{Q_p}}(\boldsymbol{K_f}, \boldsymbol{Q_p}) \cdot m_{-\boldsymbol{Q_p}}(\boldsymbol{K_f}, \boldsymbol{Q_p}) \\
&\qquad = A^*(\boldsymbol{K_f})A(\boldsymbol{K_f} - \boldsymbol{Q_p})\Psi_s(\boldsymbol{Q_p}) \cdot A(\boldsymbol{K_f})A^*(\boldsymbol{K_f} - \boldsymbol{Q_p}) \\
&\qquad = \left|A(\boldsymbol{K_f})\right|^2 \left|A(\boldsymbol{K_f} - \boldsymbol{Q_p})\right|^2 \Psi_s(\boldsymbol{Q_p}) \\
&\qquad\qquad where\ \boldsymbol{K_f} \in -\boldsymbol{Q_p}\ DO
\end{aligned}
\tag{3-51}
$$

It is worth noting that aberration information (including defocus) is included in the aperture function $A(\boldsymbol{K_f})$, as shown in Equations (3-10) and (3-11). This allows $G(\boldsymbol{K_f}, \boldsymbol{Q_p})$ to be expressed as the product of a real coefficient and $\Psi_s(\boldsymbol{Q_p})$, rather than as a product of two complex functions. As a result, both sides of the equation can be divided by the same real coefficient to calculate $\Psi_s(\boldsymbol{Q_p})$. Therefore, this masking step also serves the purpose of aberration correction.

The result of Equation (3-51) is then summed with respect to $\boldsymbol{K_f}$, the summation is denoted as $\Psi_{SSB}(\boldsymbol{Q_p})$.

$$\begin{aligned}\Psi_{SSB}(\boldsymbol{Q_p}) &= \int G_{s,-\boldsymbol{Q_p}}(\boldsymbol{K_f},\boldsymbol{Q_p})\cdot m_{-\boldsymbol{Q_p}}(\boldsymbol{K_f},\boldsymbol{Q_p})\cdot d\boldsymbol{K_f} \\ &= \Psi_s(\boldsymbol{Q_p}) \int\limits_{\boldsymbol{K_f}\in -\boldsymbol{Q_p}\,DO} \left|A(\boldsymbol{K_f})\right|^2 \left|A(\boldsymbol{K_f}-\boldsymbol{Q_p})\right|^2 d\boldsymbol{K_f}\end{aligned} \tag{3-52}$$

The integral in Equation (3-52) is taken over $\boldsymbol{K_f}$ that satisfying the condition $\boldsymbol{K_f} \in -\boldsymbol{Q_p}\,DO$. Finally, as in the WDD method, an inverse Fourier transform is applied to $\Psi_{SSB}(\boldsymbol{Q_p})$ to obtain $\psi_{SSB}(\boldsymbol{R_p})$.

$$\psi_{SSB}(\boldsymbol{R_p}) = \mathcal{F}^{-1}\{\Psi_{SSB}(\boldsymbol{Q_p})\} \tag{3-53}$$

The main difference between SSB and WDD is that SSB simplifies subsequent processing by adding the WPOA condition.

The contrast transfer function (CTF) of SSB can be derived from Equation (3-52), which expresses the relationship between $\Psi_{SSB}(\boldsymbol{Q_p})$ and $\Psi_s(\boldsymbol{Q_p})$ with respect to $\boldsymbol{Q_p}$. For each $\boldsymbol{Q_p}$, the integral within the $-\boldsymbol{Q_p}$ DO region is the weighting factor of $\Psi_{SSB}(\boldsymbol{Q_p})$. A term of $\int\left|A(\boldsymbol{K_f})\right|^2 d\boldsymbol{K_f}$ is used as denominator to normalise the CTF below 1. Therefore, following to Equation (3-52), the CTF of SSB can be written as:

$$CTF_{SSB}(\boldsymbol{Q_p}) = \frac{\int_{\boldsymbol{K_f}\in -\boldsymbol{Q_p}\,DO}\left|A(\boldsymbol{K_f})\right|^2\left|A(\boldsymbol{K_f}-\boldsymbol{Q_p})\right|^2 d\boldsymbol{K_f}}{\int\left|A(\boldsymbol{K_f})\right|^2 d\boldsymbol{K_f}} \tag{3-54}$$

3.6 Algorithm of Centre-of-mass

Another method to retrieve phase information from 4DSTEM dataset is Centre-of-mass (COM), as well as related integrated Centre-of-mass (iCOM) and differential Centre-of-mass (dCOM). The COM method assumes that the sample meets the POA conditions (discussed in Section 3.2) [18].

Like the previously introduced SSB and WDD methods, the COM method also starts its calculations from the 4DSTEM dataset $\left|M(\boldsymbol{K_f},\boldsymbol{R_p})\right|^2$. However, an important difference is

that the COM method does not solve for the transmission function ψ directly, but instead calculates the gradient of ψ first. Since ψ is proportional to the sample projected potential, its gradient is proportional to the electric field on the sample plane. The gradient of ψ calculated using the COM method, as in Equation (3-55), is represented as $\boldsymbol{COM} = (COM_x, COM_y)$, which is a 2D vector field [18].

$$\begin{aligned} COM_x(\boldsymbol{R_p}) &= \int K_{fx} \cdot \left|M(\boldsymbol{K_f}, \boldsymbol{R_p})\right|^2 dK_{fx} \\ COM_y(\boldsymbol{R_p}) &= \int K_{fy} \cdot \left|M(\boldsymbol{K_f}, \boldsymbol{R_p})\right|^2 dK_{fy} \\ \boldsymbol{COM}(\boldsymbol{R_p}) &= (COM_x, COM_y) \end{aligned} \tag{3-55}$$

In Equation (3-55), K_{fx} and K_{fy} are the components of $\boldsymbol{K_f}$ in x and y directions.

As mentioned before, COM is related to the projected electric field distribution ($\boldsymbol{E_p}$) of the sample. However, COM_x and COM_y are not directly proportional to the electric field distribution. Similar to how imaging the transmission function ψ requires considering the contrast transfer function, imaging the electric field with COM is also modulated by the probe function [19].

$$\boldsymbol{COM}(\boldsymbol{R_p}) = \int \boldsymbol{E_p}(\boldsymbol{r}) \cdot \left|P(\boldsymbol{R_p} - \boldsymbol{r})\right|^2 \cdot d\boldsymbol{r} = \boldsymbol{E_p}(\boldsymbol{R_p}) \otimes \left|P(\boldsymbol{R_p})\right|^2 \tag{3-56}$$

In Equation (3-56), the imaging result of COM is the convolution of the projected electric field $\boldsymbol{E_p}$ and the intensity of the probe function $\left|P(\boldsymbol{R_p} - \boldsymbol{r})\right|^2$.

After obtaining COM, iCOM and dCOM reconstructions can be calculated from COM. The relationships between iCOM, dCOM, and COM are as follows:

$$\begin{aligned} \boldsymbol{COM}(\boldsymbol{R_p}) &= \left(\frac{\partial iCOM(\boldsymbol{R_p})}{\partial R_{px}}, \frac{\partial iCOM(\boldsymbol{R_p})}{\partial R_{py}}\right) = \boldsymbol{\nabla} iCOM(\boldsymbol{R_p}) \\ dCOM(\boldsymbol{R_p}) &= \frac{\partial COM_x(\boldsymbol{R_p})}{\partial R_{px}} + \frac{\partial COM_y(\boldsymbol{R_p})}{\partial R_{py}} = \boldsymbol{\nabla} \cdot \boldsymbol{COM}(\boldsymbol{R_p}) \end{aligned} \tag{3-57}$$

In Equation (3-57), R_{px} and R_{py} are the components of $\boldsymbol{R_p}$ in x and y directions. The

gradient of the projected electric potential (iCOM) is the projected electric field (COM), which is the relationship between the projected electric field and the projected electric potential. Similarly, the divergence of the projected electric field is proportional to the projected charge density, meaning that dCOM is the imaging of the (modulated) projected charge density.

For dCOM, the divergence can be directly calculated from COM as in Equation (3-57). However, calculating iCOM from COM usually uses Equation (3-58) [18].

$$iCOM(\boldsymbol{R_p}) = \mathcal{F}^{-1}\left\{\frac{\boldsymbol{Q_p} \cdot \mathcal{F}\{\boldsymbol{COM}(\boldsymbol{R_p})\}}{2\pi i \boldsymbol{Q_p^2}}\right\} \tag{3-58}$$

In Equation (3-58), $\boldsymbol{Q_p}$ is the frequency domain variable with respect to $\boldsymbol{R_p}$.

The frequency transfer of COM, iCOM and dCOM can be expressed as follows [18]:

$$\begin{aligned}
\mathcal{F}\{iCOM(\boldsymbol{R_p})\} &= \boldsymbol{\mathcal{F}}\left\{\left|P(\boldsymbol{R_p})\right|^2\right\} \cdot \mathcal{F}\{\psi(\boldsymbol{R_p})\} = \boldsymbol{\mathcal{F}}\left\{\left|P(\boldsymbol{R_p})\right|^2\right\} \cdot \Psi(\boldsymbol{Q_p}) \\
\mathcal{F}\{\boldsymbol{COM}(\boldsymbol{R_p})\} &= \boldsymbol{\mathcal{F}}\left\{\left|P(\boldsymbol{R_p})\right|^2\right\} \cdot \mathcal{F}\{\boldsymbol{\nabla}\psi(\boldsymbol{R_p})\} \\
\mathcal{F}\{dCOM(\boldsymbol{R_p})\} &= \boldsymbol{\mathcal{F}}\left\{\left|P(\boldsymbol{R_p})\right|^2\right\} \cdot \mathcal{F}\{\boldsymbol{\nabla}^2\psi(\boldsymbol{R_p})\}
\end{aligned} \tag{3-59}$$

The COM method indirectly images the transmission function ψ by first calculating the gradient field of ψ, which are COMx and COMy. In addition, COMx and COMy have clear physical interpretations—as the projected electric field after being processed by the transfer function—which facilitates the direct mapping of the electric field. However, unlike the ptychography method, the current COM method cannot perform aberration correction. COM requires the intensity distribution of probe function P to be as focused as possible; otherwise, its intensity distribution will be convolved with the imaging object (transmission function ψ or its gradient field, etc.), causing the image to blur. In Chapter 5, an improvement to the COM method to perform aberration correction will be discussed.

3.7 Algorithm of Extended Ptychographic Iterative Engine

In addition to the non-iterative methods discussed above, the extended Ptychographic Iterative Engine (ePIE) [20] is another common algorithm used for phase reconstruction in TEM. In this thesis, ePIE is used in Chapter 4 for comparison with SSB and COM reconstructions. In this section ePIE algorithm used in Chapter 4 is provided.

As an iterative method, an initial probe function $P_0(\boldsymbol{R})$ and an initial transmission function $\psi_0(\boldsymbol{R}+\boldsymbol{R_p})$ are defined before the initial iteration (0th iteration). The initial probe function is calculated based on estimated aberrations, while the initial transmission function is uniformly set to unity. In this thesis, the inverse-multislice ePIE as mentioned in Chapter 2 is not used. The sample is assumed to satisfy POA for single slice ePIE, but it can also reflect amplitude change due to sample absorption.

After the initialization, the algorithm proceeds with iterations. For the probe position $\boldsymbol{R_p}$ in the nth iteration, an expected exit wave $EW_{E,n}(\boldsymbol{R},\boldsymbol{R_p})$ can be calculated based on $P_n(\boldsymbol{R})$ and $\psi_n(\boldsymbol{R}+\boldsymbol{R_p})$:

$$EW_{E,n}(\boldsymbol{R},\boldsymbol{R_p}) = P_n(\boldsymbol{R})\cdot\psi_n(\boldsymbol{R}+\boldsymbol{R_p}) \tag{3-60}$$

Then, the expected exit wave is propagated to the diffraction plane via a Fourier transform, denoted as $M_{E,n}(\boldsymbol{K_f},\boldsymbol{R_p})$. After that, $M_{E,n}(\boldsymbol{K_f},\boldsymbol{R_p})$ can be represented by its amplitude $\left|M_{E,n}(\boldsymbol{K_f},\boldsymbol{R_p})\right|$ and phase $\angle\left[M_{E,n}(\boldsymbol{K_f},\boldsymbol{R_p})\right]$, as shown in Equation (3-61).

$$\begin{aligned} M_{E,n}(\boldsymbol{K_f},\boldsymbol{R_p}) &= \mathcal{F}\{EW_{E,n}(\boldsymbol{R},\boldsymbol{R_p})\} \\ &= \left|M_{E,n}(\boldsymbol{K_f},\boldsymbol{R_p})\right|\cdot\exp\{i\cdot\angle\left[M_{E,n}(\boldsymbol{K_f},\boldsymbol{R_p})\right]\} \end{aligned} \tag{3-61}$$

On the other hand, the experimental measurements of a 4DSTEM dataset, *i.e.* the intensity distribution $\left|M(\boldsymbol{K_f},\boldsymbol{R_p})\right|^2$, are square-rooted and replace the $\left|M_{E,n}(\boldsymbol{K_f},\boldsymbol{R_p})\right|$ in Equation (3-60) to construct a new wave function $M_{R,n}(\boldsymbol{K_f},\boldsymbol{R_p})$.

$$M_{R,n}(\boldsymbol{K_f}, \boldsymbol{R_p}) = \left|M(\boldsymbol{K_f}, \boldsymbol{R_p})\right| \cdot \exp\{i \cdot \angle[M_{E,n}(\boldsymbol{K_f}, \boldsymbol{R_p})]\} \tag{3-62}$$

The new wave function, after amplitude replacement, is then backpropagated to the sample plane, which is denoted as $EW_{R,n}(\boldsymbol{R}, \boldsymbol{R_p})$.

$$EW_{R,n}(\boldsymbol{R}, \boldsymbol{R_p}) = \mathcal{F}^{-1}\{M_{R,n}(\boldsymbol{K_f}, \boldsymbol{R_p})\} \tag{3-63}$$

Then, the probe function $P_n(\boldsymbol{R})$ and the transmission function $\psi_n(\boldsymbol{R} + \boldsymbol{R_p})$ are updated based on the new exit wave $EW_{R,n}(\boldsymbol{R}, \boldsymbol{R_p})$:

$$\begin{aligned} \psi_{n+1}(\boldsymbol{R} + \boldsymbol{R_p}) &= \psi_n(\boldsymbol{R} + \boldsymbol{R_p}) + \alpha_\psi \cdot \frac{P_n^*(\boldsymbol{R})}{|P_n(\boldsymbol{R})|^2_{max}} \\ &\quad \cdot \left[EW_{R,n}(\boldsymbol{R}, \boldsymbol{R_p}) - EW_{E,n}(\boldsymbol{R}, \boldsymbol{R_p})\right] \\ P_{n+1}(\boldsymbol{R}) &= P_n(\boldsymbol{R}) + \alpha_P \cdot \frac{\psi_n^*(\boldsymbol{R} + \boldsymbol{R_p})}{\left|\psi_n(\boldsymbol{R} + \boldsymbol{R_p})\right|^2_{max}} \cdot \left[EW_{R,n}(\boldsymbol{R}, \boldsymbol{R_p}) - EW_{E,n}(\boldsymbol{R}, \boldsymbol{R_p})\right] \end{aligned} \tag{3-64}$$

Where $|P_n(\boldsymbol{R})|^2_{max}$ and $\left|\psi_n(\boldsymbol{R} + \boldsymbol{R_p})\right|^2_{max}$ represents the maximum values in $|P_n(\boldsymbol{R})|^2$ and $\left|\psi_n(\boldsymbol{R} + \boldsymbol{R_p})\right|^2$, respectively. α_ψ and α_P represents two update factors for transmission function and probe function, respectively. After $\psi_{n+1}(\boldsymbol{R} + \boldsymbol{R_p})$ and $P_{n+1}(\boldsymbol{R})$ are calculated and updated, a new iteration with next probe position (next $\boldsymbol{R_p}$) begins as described in Equation (3-60).

After the Nth iteration, the algorithm converges. Then $\psi_N(\boldsymbol{R} + \boldsymbol{R_p})$ and $P_N(\boldsymbol{R})$ are regarded as the final reconstruction results.

When ePIE is used for defocused dataset, an estimated defocus value is used to calculate the initial probe function $P_0(\boldsymbol{R})$. The iteration starts with a defocused condition and the aberration corrected transmission function can be reconstructed.

3.8 Conclusion

In this chapter, the theoretical background relevant to the following chapters is introduced. This includes the constraints on the interaction between the sample and the electron beam,

the mathematical description of the sample, and the approximation conditions such as POA and WPOA, the probe function, and the simplified description of aberrations. After defining these conditions for the sample and the electron beam, the relation between the sample information and the interference fringes on the diffraction discs is derived.

Next, the mathematical description of 4DSTEM datasets is introduced. Additionally, the expression of the interference fringes under 4DSTEM conditions is discussed.

After that, four methods for recovering phase information from the 4DSTEM dataset are introduced. These include three ptychography methods (2 non-iterative and 1 iterative methods) capable of aberration correction: WDD, SSB and ePIE. WDD assumes that electrons are only elastic scattered. At the same time the sample thickness can be neglected and satisfy POA. SSB, in addition to these assumptions, requires the sample to satisfy the WPOA condition. ePIE with single slice transmission function is the only iterative ptychography method used in this thesis. The sample condition for ePIE is the same as for WDD. The other method, COM, which requires a POA condition, first calculates the gradient of the transmission function and then images the transmission function based on this gradient. However, the current COM method cannot be used to correct aberrations and therefore requires an in-focus condition.

Reference

[1] C. M. O'Leary, G. T. Martinez, E. Liberti, M. J. Humphry, A. I. Kirkland, and P. D. Nellist, "Contrast transfer and noise considerations in focused-probe electron ptychography," *Ultramicroscopy,* vol. 221, p. 113189, 2021.

[2] C. M. O'Leary *et al.*, "Phase reconstruction using fast binary 4D STEM data," *Applied Physics Letters,* vol. 116, no. 12, 2020.

[3] J. G. Lozano, G. T. Martinez, L. Jin, P. D. Nellist, and P. G. Bruce, "Low-dose aberration-free imaging of Li-rich cathode materials at various states of charge using electron ptychography," *Nano letters,* vol. 18, no. 11, pp. 6850-6855, 2018.

[4] M. J. Zachman, Z. Yang, Y. Du, and M. Chi, "Robust Atomic-Resolution Imaging of Lithium in Battery Materials by Center-of-Mass Scanning Transmission Electron

Microscopy," *ACS Nano,* vol. 16, no. 1, pp. 1358-1367, 2022.

[5] H. Bethe, "Theorie der Beugung von Elektronen an Kristallen," *Annalen der Physik,* vol. 392, no. 17, pp. 55-129, 1928.

[6] P. Goodman and A. F. Moodie, "Numerical evaluations of N-beam wave functions in electron scattering by the multi-slice method," *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography,* vol. 30, no. 2, pp. 280-290, 1974.

[7] Y. Aharonov and D. Bohm, "Significance of Electromagnetic Potentials in the Quantum Theory," *Physical Review,* vol. 115, no. 3, pp. 485-491, 1959.

[8] D. B. Williams and C. B. Carter, "High-resolution TEM," in *Transmission Electron Microscopy: A Textbook for Materials Science. Imaging. III*, vol. 3: Springer Science & Business Media, 2008, ch. 28, pp. 483-509.

[9] O. L. Krivanek, N. Dellby, and A. R. Lupini, "Towards sub-Å electron beams," *Ultramicroscopy,* vol. 78, no. 1, pp. 1-11, 1999.

[10] M. Haider, S. Uhlemann, and J. Zach, "Upper limits for the residual aberrations of a high-resolution aberration-corrected STEM," *Ultramicroscopy,* vol. 81, no. 3-4, pp. 163-175, 2000.

[11] C. Ophus, "Four-dimensional scanning transmission electron microscopy (4D-STEM): From scanning nanodiffraction to ptychography and beyond," *Microscopy and Microanalysis,* vol. 25, no. 3, pp. 563-582, 2019.

[12] J. M. Rodenburg and R. H. T. Bates, "The theory of super-resolution electron microscopy via Wigner-distribution deconvolution," *Philosophical Transactions of the Royal Society of London. Series A: Physical and Engineering Sciences,* vol. 339, no. 1655, pp. 521-553, 1992.

[13] P. D. Nellist and J. M. Rodenburg, "Beyond the conventional information limit: the relevant coherence function," *Ultramicroscopy,* vol. 54, no. 1, pp. 61-74, 1994.

[14] H. Yang *et al.*, "Electron ptychographic phase imaging of light elements in crystalline materials using Wigner distribution deconvolution," *Ultramicroscopy,* vol. 180, pp. 173-179, 2017.

[15] P. Li, T. B. Edo, and J. M. Rodenburg, "Ptychographic inversion via Wigner distribution deconvolution: noise suppression and probe design," *Ultramicroscopy,* vol. 147, pp. 106-113, 2014.

[16] H. Yang *et al.*, "Simultaneous atomic-resolution electron ptychography and Z-contrast imaging of light and heavy elements in complex nanostructures," *Nature Communications,* vol. 7, no. 1, p. 12532, 2016.

[17] J. M. Rodenburg, B. C. McCallum, and P. D. Nellist, "Experimental tests on double-resolution coherent imaging via STEM," *Ultramicroscopy,* vol. 48, no. 3, pp. 304-314, 1993.

[18] I. Lazić, E. G. T. Bosch, and S. Lazar, "Phase contrast STEM for thin samples: Integrated differential phase contrast," *Ultramicroscopy,* vol. 160, pp. 265-280, 2016.

[19] K. Müller *et al.*, "Atomic electric fields revealed by a quantum mechanical approach to electron picodiffraction," *Nature Communications,* vol. 5, no. 1, p. 5653, 2014/12/15 2014.

[20] A. M. Maiden and J. M. Rodenburg, "An improved ptychographical phase retrieval algorithm for diffractive imaging," *Ultramicroscopy,* vol. 109, no. 10, pp. 1256-1262, 2009.

Chapter 4:
Influence of Inelastically Scattered Electrons on Electron Ptychography

4.1 Introduction

In this chapter, the influence of inelastically scattered electrons on electron ptychography is examined. As noted in Chapter 3, these electrons have lost a measurable amount of energy during their interactions with the sample. They provide insights into plasmon excitations, chemical composition, bonding, and valence, and can be measured in an electron energy loss spectrum (EELS) or energy-filtered transmission electron microscope (EFTEM) configuration. Numerous applications in material science and biology leverage inelastic scattering [1-5]. However, the inherent chromatic aberrations of electromagnetic lenses limits resolution [6], and thus the effect of inelastic scattering is to reduce the contrast and resolution in conventional TEM images [7-9]. The effects of inelastically scattered electrons on TEM imaging also depends on the nature of the specimen. Generally, as the thickness of the sample increases, the probability of inelastic scattering increases [10]. Therefore in thicker samples, the influence of inelastic scattering electrons is more significant.

A reduction in the effect of inelastic scattering of electrons is most commonly achieved by use of an energy filter [11-13] to select only the elastic scattering. However, this leads to an overall reduction in the number of electrons contributing to the image; an effect which is important for thicker samples where inelastic scattering is significant and for low fluence

imaging.

The same energy filter can also be used in electron ptychography by recording an energy-filtered 4DSTEM dataset. However, in terms of the reconstruction algorithm, as discussed in Section 3.2, only elastically scattered electrons are considered. Experimentally, it has been shown that plasmon-excited electrons can display the coherent interference features necessary for ptychography [14], but a reconstruction has not been demonstrated.

In this chapter, the impact of inelastically scattered electrons is examined and compared using ptychographic reconstructions, employing both iterative and non-iterative methods, using filtered and unfiltered 4D STEM datasets. In Section 4.2, the influence of plasmon-excited electrons is explored. Factors and settings that influence plasmon-excited electrons are described in Section 4.2.3. Following this, in Section 4.3, I discuss high-loss electrons and energy-filtered diffraction patterns within high-loss ranges. Finally, in Section 4.4, the effects of thermal diffused scattering (TDS) are examined, including the distribution of TDS electrons on diffraction patterns and their corresponding reconstructions. By understanding the influence of inelastic scattered electrons, I hope to optimize the experimental parameters, utilizing the information each electron carries, and thereby reducing the electron fluence required for imaging.

4.2 Influence of Plasmon-excited Electrons

In this section, the influence of plasmon-excited electrons on 4DSTEM datasets and ptychography reconstructions is explored. Plasmon-excited electrons are located at energy loss 10-50 eV and usually forms a plasmon peak on EELS spectrums. An energy-filtered 4D STEM dataset was recorded (Section 4.2.1), and ptychography reconstructions were performed and compared using these energy-filtered datasets (Section 4.2.2). However, several key factors control the influence of plasmon-excited electrons on ptychography, which are discussed in Section 4.2.3.

4.2.1 Energy-filtered 4DSTEM

Energy-filtered 4DSTEM data was acquired using a JEOL ARM200F (S)TEM with a GATAN Imaging Filter. An Ultrascan camera positioned after the energy filter recorded 4DSTEM datasets, both with and without energy filtering. Figure 4-1 illustrates the configuration used. The sample consisted of gold nanoparticles on a carbon film, covered by a layer of amorphous carbon. As shown in Figure 4-2, the integrated EELS spectrum of the sample exhibits a plasmon peak. The diameter of the gold nanoparticles ranged from 5-10 nm and all 4DSTEM datasets were collected from the edges of the gold nanoparticles.

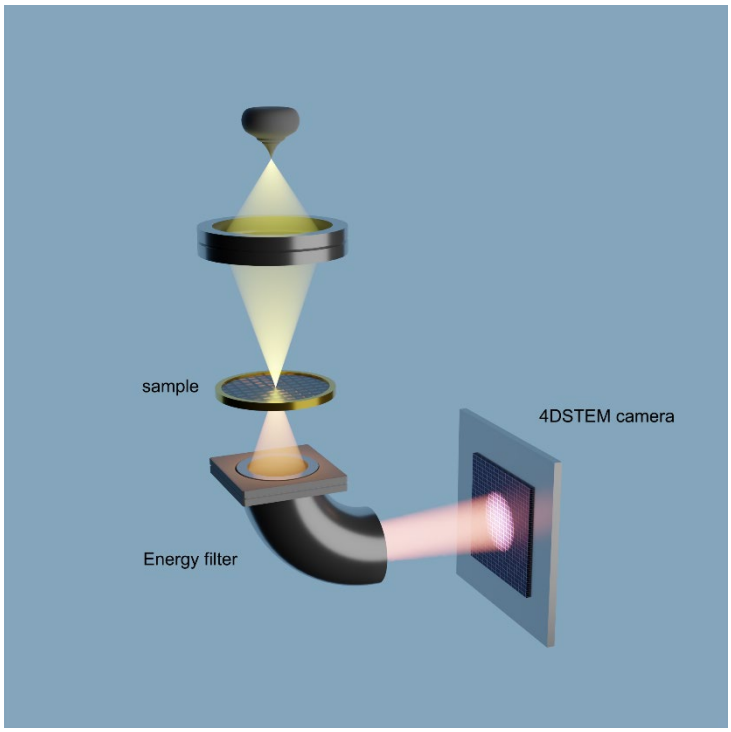

Figure 4-1 | Schematic electron optical configuration for the energy-filtered 4DSTEM collection. A post-column filter (Gatan GIF) is used to obtain energy-filtered CBEDs on a 4DSTEM camera.

The diffraction patterns shown in Figure 4-2 were recorded under typical optical settings used for high-resolution ptychographic reconstructions, with a convergence angle of 22.5 mrad at 200 kV. The energy filter was tuned in EFTEM mode, and 400 diffraction patterns (20×20 probe positions) were recorded at different energy loss windows, covering the zero-

loss peak or the plasmon peak. The 4DSTEM datasets were recorded using the Ultrascan CCD camera installed after the energy filter with a Gatan DigiScan system. Due to the slower speed of this camera compared to typical fast direct detection cameras designed for 4DSTEM, only 20×20 scanning positions were recorded. The camera was binned to 256×256, resulting in a pixel size corresponding to 0.37 mrad in scattering angle. The slit used for the zero-loss peak had a width of -5 to 5 eV, and for the plasmon peak had a width of 20 to 30 eV. 4DSTEM datasets using in-focus and defocused (-17 nm) probes were recorded with the same beam settings but different scan steps (0.329 Å for the in-focus datasets and 2.5 Å for the defocused dataset). The probe size of the 2 datasets is 0.558 Å for the in-focus datasets and 7.65 Å for defocused dataset. The scanning area of the 4DSTEM dataset is in a thin area at the top edge of the particle in Figure 4-2(a). These 4DSTEM datasets were then used for ptychographic reconstructions.

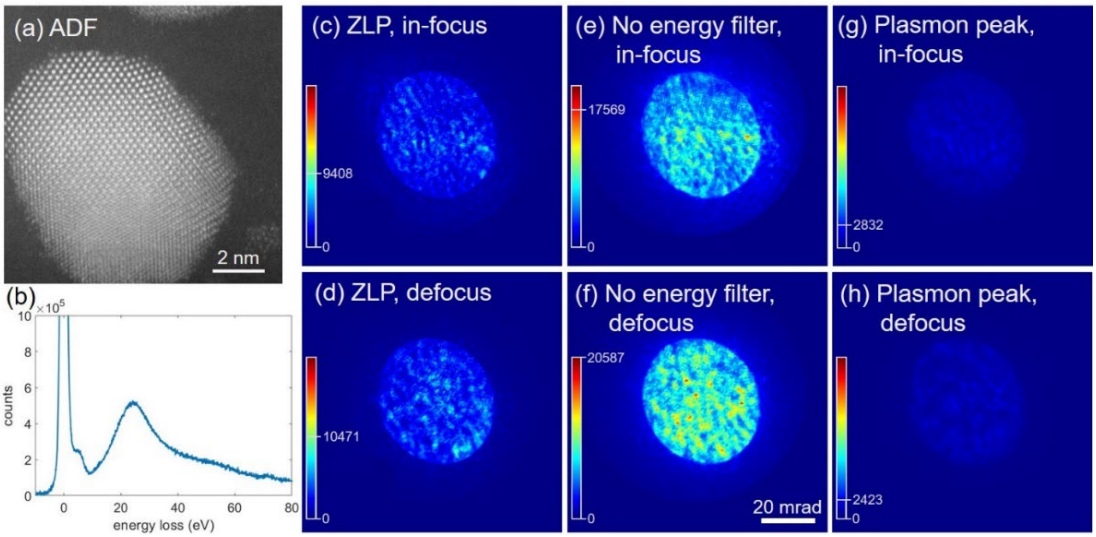

Figure 4-2 | (a) ADF image of a gold nanoparticle. (b) EELS spectrum collected over a larger region (approximately 200x200 nm) that includes the particle shown in (a). The spectrum in (b) represents the integral of all sampling points within the collection area (c-h) Example diffraction patterns from the zero-loss peak, plasmon peak and unfiltered datasets for both in-focus and defocused (-17 nm) acquisition. Maximum values for each diffraction pattern are marked on the colourbars. The diffraction patterns deviate from a perfect circle due to residual diffraction astigmatism.

In Figure 4-2, examples of diffraction patterns at the zero-loss peak, plasmon peak, and without energy filtering demonstrate the differences in intensity. To obtain sufficient counts for ptychography reconstruction in the plasmon peak, a high electron fluence (1.65×10^{12} $e^-/Å^2$ for the in-focus dataset and 2.85×10^{10} $e^-/Å^2$ for defocused dataset) was used in the experiment.

The thickness of the sample can be estimated by inelastic mean free path of the sample using log-ratio method [15]. The relation between inelastic mean free path and sample thickness can be written as:

$$\frac{t}{\lambda_{iMFP}} = \ln\left(\frac{I_t}{I_{ZLP}}\right) \tag{4-1}$$

In Equation (4-1), t represents sample thickness, λ_{iMFP} represents inelastic mean free path of the sample, I_t represents integrated intensity of the whole spectrum, I_{ZLP} represents integrated intensity of zero-loss peak.

The inelastic mean free path for different accelerating voltages and materials can be readily calculated using standard methods [10, 16]. The inelastic mean free path to 200 kV accelerated electrons is 160 nm in amorphous carbon, and 84 nm in FCC gold [17]. In this experiment, since the thickness ratio between gold particles and amorphous carbon was not measured in advance, the above formula cannot be directly used to estimate the thickness. However, it is still possible to roughly estimate the upper and lower limits of the sample thickness by assuming that the mean free path of the sample is entirely that of either amorphous carbon or gold particles. If the mean free path is assumed to be that of amorphous carbon, then according to the formula, the estimated upper limit of the thickness is approximately 110 nm. Conversely, if it is assumed to be that of gold particles, the estimated lower limit is about 58 nm. If the actual thickness of the gold particles is assumed to be close to their diameter in the X/Y direction, approximately 10 nm, then the thickness of the amorphous carbon should be less than 110 nm.

In the above estimation, I_{ZLP} is integrated from the minimum value of energy loss obtained

from the experimental data, up to the first local minimum located between the zero-loss peak and the plasmon peak. At the same time, $\frac{I_t}{I_{ZLP}}$ can also be estimated, which indicates that approximately 50% of the electrons have undergone inelastic scattering.

Besides, as mentioned earlier, this experiment used an Ultrascan CCD. If a higher-performance direct detection camera were used, with improved MTF and DQE [18], the experimental 4DSTEM datasets and thickness estimation can be more representative.

4.2.2 Energy-filtered Ptychography

The in-focus 4DSTEM datasets were reconstructed using both single-side band (SSB) [19] and ePIE [20] algorithms. I also show simultaneous virtual ADF (vADF, virtual collection angle 25 ~ 50 mrad) and corresponding integrated centre-of-mass (iCOM) [21] images reconstructed from the same dataset for comparison (Figure 4-3). For the in-focus reconstructions, the field of view (FOV) is approximately 6×6 Å, given that only 400 (20×20) diffraction patterns were recorded. This means there are only 20×20 pixels for vADF, iCOM, and the SSB reconstructions, with a pixel size of 0.329 Å if no upsampling is applied. The ePIE algorithm was also used for the same dataset. For the in-focus datasets (Figure 4-3, last column), the ePIE reconstructions have a pixel size of 0.265 Å. For defocused 4DSTEM datasets (Figure 4-4, defocus of -17 nm), a larger field of view, about 50×50 Å, can be reconstructed using the ePIE algorithm. The defocused probe diameter is about 7.5 Å, and the overlap ratio is approximately 59%. The reconstructed phase from the plasmon loss (energy loss of 20~30 eV) still shows atomic resolution in the defocused data. All ptychographic phases (SSB and ePIE, in-focus and defocus) were reconstructed with an electron wavelength corresponding to a 200 kV acceleration voltage. Although the number of pixels is limited in all cases, atomic resolution information is still conveyed in all the datasets (no filter, ZLP, and plasmon-loss). The lattice is distorted because of the slow data acquisition coupled with some microscope instabilities.

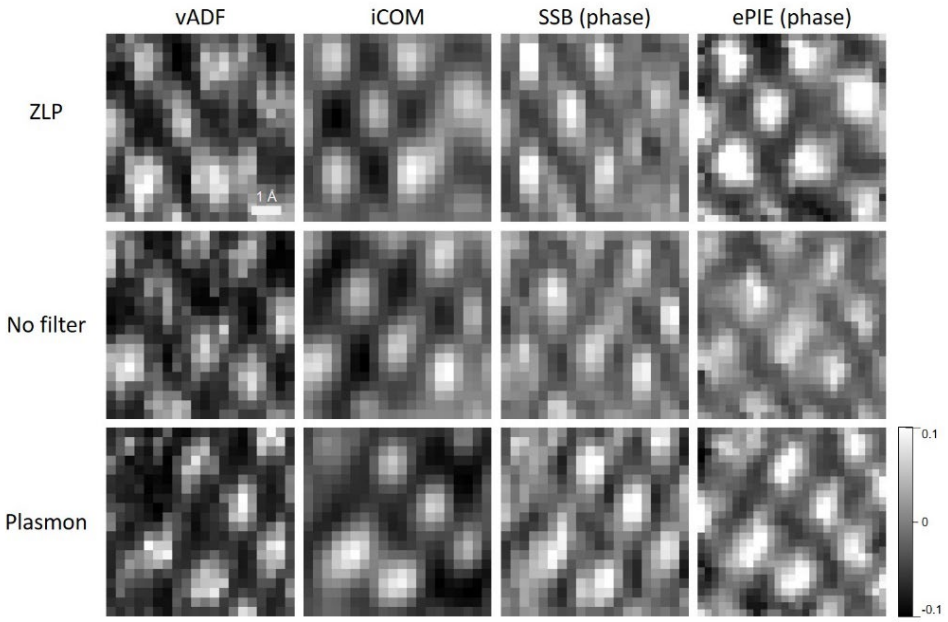

Figure 4-3 | Virtual ADF (vADF), integrated center-of-mass (iCOM), Single-side Band (SSB) and ePIE reconstructions from in-focus 4DSTEM datasets using the zero-loss peak (ZLP), no energy filter and the plasmon peak. Reconstructed phase using SSB and ePIE are plotted with the same colourbar. iCOM is plotted with a separated greyscale because the image is not normalized to rad as SSB and ePIE. The absolute value of iCOM image depends on intensity of the incident beam as shown in Equation (3-60).

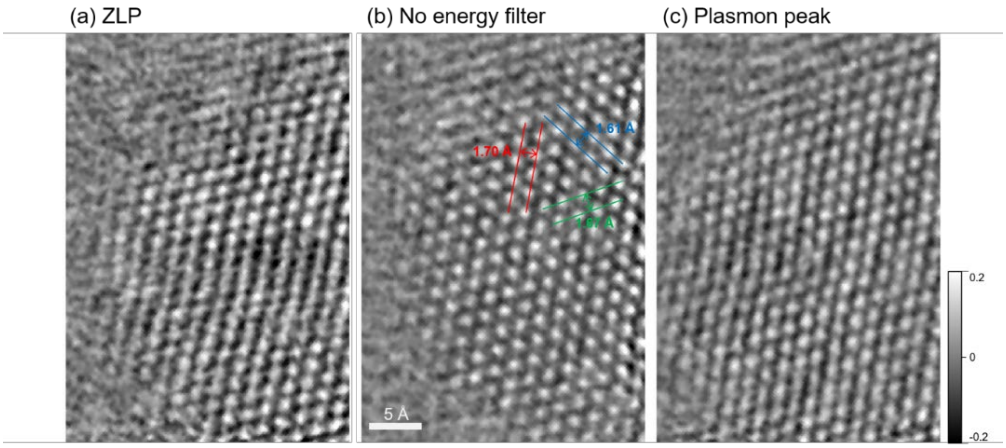

Figure 4-4 | Defocused ePIE phase reconstructions for (a) zero-loss peak (ZLP), (b) no filter and (c) plasmon 4DSTEM datasets. The defocus is -17 nm. Phases are plotted using the same greyscale. The d-spacings marked on (b) are based on an average of 8~10 planes for each direction.

As shown in Figures 4-3 and 4-4, the ptychography reconstructions show gold atom columns, even when only diffraction patterns recorded from the plasmon peak are used. Furthermore, in Figure 4-3, both iCOM and vADF also show gold atom columns. In Figure 4-4, all reconstructions (from the zero-loss peak, plasmon loss, or without energy filtering) show a similar resolution (approximately 1.6 Å, measured by the average spacing of lattice planes in the phase reconstructions). However, reconstructions from the zero-loss peak signals (a) and plasmon peak signals (c) in Figure 4-4 exhibit a loss of some high-frequency details. This is due to microscope stability issues during the long acquisition time per pixel (0.1 second) and decreased signal.

In this section, the energy-filtered ptychography experiments demonstrate that, under the specified beam settings (200 kV, 22.5 mrad, 0.37 mrad pixel size on the detector, in-focus to -17 nm defocus), plasmon inelastic scattered electrons (with an energy loss of 20-30 eV) do not have a detectable impact on the resolution of ptychographic phase reconstruction.

4.2.3 Factors Affecting Energy-filtered Ptychography

At the experimental conditions described in previous section, all the ptychographic reconstructions can resolve atomic columns using ZLP electron dataset, plasmon-excited electron dataset or unfiltered dataset. This demonstrated that, at the experimental condition used in previous section, plasmon-excited electrons do not have significant influence on atomic-resolution ptychography reconstruction. In this section, I will further explore the reasons behind this observation. Additionally, I will discuss key factors related to energy-filtered ptychography and calculate the range of parameters where the impact of inelastically scattered electrons can be neglected.

For traditional energy-filtered imaging; the concept of preservation of elastic contrast where elastic contrast remains in an image or diffraction pattern formed by inelastically scattered electrons has been widely discussed [22, 23]. A similar preservation of elastic contrast can occur at the diffraction plane, where the 4DSTEM dataset is recorded. In Figure 4-3 and 4-4, both ptychography reconstructions with or without plasmon-excited electrons show

lattice in both in-focus and defocus conditions. To explain this, we consider the distribution of inelastically scattered electrons at the diffraction plane. In the following sections, we will discuss inelastic characteristic scattering angle, and its relation to defocus setting and detector sampling.

4.2.3.1 Characteristic Inelastic Scattering Angle

At the diffraction plane, inelastically scattered electrons exhibit an additional Lorentzian distribution in addition to the elastically scattered electrons [10]. The inelastic characteristic scattering angle (θ_E) is used to describe this spread, and is related to the energy of the incident beam and the energy lost during the inelastic scattering process. This relationship can be expressed as [1]:

$$\theta_E \approx \frac{\Delta E}{2E_0} \tag{4-2}$$

where, $2\theta_E$, represents the full width at half maximum (FWHM) of the peak of a Lorentzian distribution. The corresponding energy loss is denoted as ΔE, and E_0 is the primary beam energy, approximated without considering relativistic effects [10].

Consequently, the unfiltered diffraction patterns can be approximated as a summation of the elastically scattered diffraction pattern and a series of additional diffraction patterns, each convolved with a Lorentzian distribution corresponding to an energy loss. The Lorentzian distribution is determined by the characteristic scattering angle θ_E.

For the experimental parameters used here, the characteristic scattering angle of the plasmon-excited electrons (with an energy loss of 20 eV) is approximately 0.05 mrad.

This inelastic characteristic scattering angle controls the blurring of features in the diffraction pattern caused by inelastically scattered electrons. In the following two sections, we will discuss its relationship with other parameters.

4.2.3.2 Defocus and Interference Fringes

One of the features that can be blurred by inelastically scattered electrons are the

interference fringes, which are the origin of information in ptychography, as described in Chapter 3. Ptychography relies on interference between scattered beams and is often dominated by interference between the direct beam and a diffracted beam or diffracted beams which gives rise to fringes that are responsible for contrast in the reconstructed phase. Hence, if the inelastic characteristic scattering angle is comparable to the fringe size, they will be blurred with a consequent loss of information. The size of interference fringes is calculated in Section 3.2. The period of the interference fringes with only defocus considered, following Equation (3-33) in Chapter 3, can be written as:

$$\frac{1}{\lambda C_1 |\boldsymbol{g}| \cos\left[\angle\left(\boldsymbol{K_f}, \boldsymbol{g}\right)\right]} \tag{4-3}$$

Where λ represents the wavelength of incident beam, C_1 is a defocus value and $\angle$ represents the angle between the vectors $\boldsymbol{K_f}$ and $\boldsymbol{g}$. The vector $\boldsymbol{g}$, as a spatial frequency, is the vector between the centres of the bright field disc and a diffraction disc. $\boldsymbol{K_f}$ denotes the coordinate in the diffraction plane.

Equation (4-3) was derived in Chapter 3 without considering aberrations other than defocus. The influence of other aberrations on Equation (4-3) can be calculated, as described in Section 3.2.3, Chapter 3, by including more aberrations terms in phase distortion function $\chi(\boldsymbol{K_f})$. In most experimental cases, usually other aberrations are minimized during the beam alignment, only defocus is sometimes adjusted to specific values to control the probe size on the sample. Thus, in this section only defocus is considered as a parameter in the experimental setting.

In addition, incoherence of the incident beam is not considered in Equation (4-3). Energy spread and chromatic aberration can cause interference fringes with different periods to overlap incoherently. However, since the energy spread of the electron gun is very small relative to the electron wavelength, for example, an energy spread of 0.3 eV at an accelerating voltage of 200 kV corresponds to less than one part per million, the resulting variation of λ in the Equation (4-3) for interference fringe period is negligible.

A limitation for Equation (4-3) is it that multiple scattering is not considered. If a small amount of multiple scattering occurs, it may reduce the intensity of the interference fringes. However, it will not cause a significant change in their period. If the sample is thick enough for extensive multiple elastic scattering to occur, these interference fringes may be covered.

Based on the above equation, if it assumed that the direction of the vector ($\boldsymbol{g}$) is the same as the direction of the K_{fx} or K_{fy} component of $\boldsymbol{K_f}$, then the equation can be simplified from two dimensions to one dimension as:

$$\frac{1}{\lambda C_1 g} \tag{4-4}$$

Considering g is a frequency, the corresponding d-spacing of the lattice can be defined as:

$$d = \frac{1}{g} \tag{4-5}$$

When θ_d (the scattering angle corresponding to d-spacing d) is small, $sin(\theta_d)$ can be approximated as θ_d and applying Braggs law the relationship between λ and θ_d is:

$$d \cdot \sin(\theta_d) = \lambda \approx \frac{\theta_d}{g} \tag{4-6}$$

Using Equation (4-6), Equation (4-4) can be transformed to express the period of fringes in angular units (mrad) rather than in frequency (m^{-1}). We define θ as the period of fringes in mrad, which represents the size of the fringes, by multiplying Equation (4-4) and λ:

$$\theta = \frac{\lambda}{\theta_d C_1} = \frac{d}{C_1} \tag{4-7}$$

Hence, with an increase in defocus, the size of the features in the CBED patterns (θ) decreases.

Based on Equation (4-7), θ with respect to C_1 is plotted for different λ and θ_d conditions, as shown in Figure 4-5. It should be noted that θ_d must be smaller than 2α in

an experiment, otherwise there will be no interference between the bright field disc and a diffracted disc, as described in Section 3.2. However, in Figure 4-5 this 2α limitation is not introduced, in other words, 2α here is assumed to be sufficiently large. The extreme condition of $\theta_d = 2\alpha$ is discussed later.

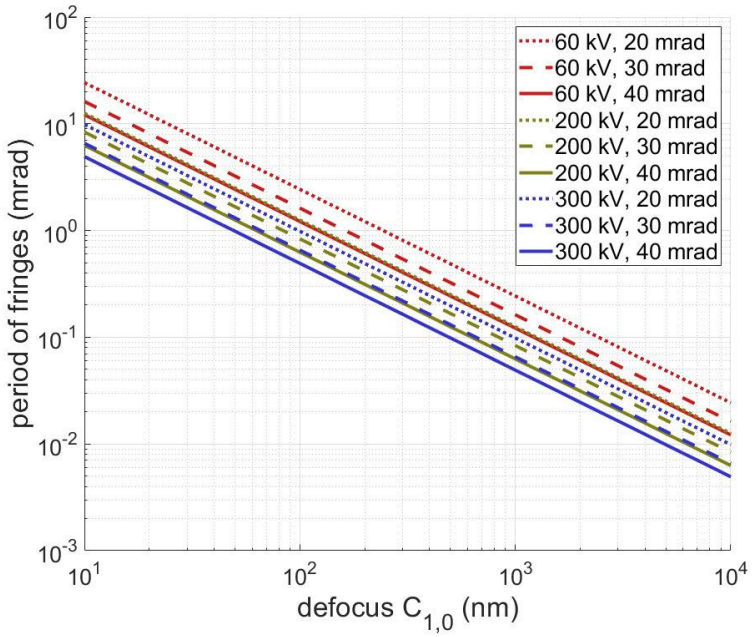

Figure 4-5 | Period of fringes (θ) with respect to defocus at different voltages (60, 200, 300 kV) and diffraction angles (θ_d) (20, 30, 40 mrad). θ_d must be smaller than 2α in an experiment and here 2α is assumed to be large enough for all plotted θ_d.

If the fringe size is comparable to the characteristic scattering angle of plasmon excited electrons, the fringes become blurred, leading to information loss. We consider a maximum defocus value (df_{max}) that makes the fringes of a specific lattice spacing equal to the inelastic characteristic scattering angle as:

$$df_{max} = \frac{d}{\theta_E} \tag{4-8}$$

Where df_{max} is the maximum defocus. d is the spacing of features that lead to interference fringes in the CBED pattern and θ_E is the inelastic characteristic scattering angle (typically 0.05 mrad for electrons undergoing plasmon losses). Using Equation (4-8),

we can calculate a typical limiting defocus for specific lattice structures and energy loss.

Considering interference features in the bright-field disc of the CBED pattern, the maximum scattering angle at which a diffracted disc can interfere with the bright-field disc is 2α, where α is the semi-angle of convergence. This gives the smallest lattice spacing which can give rise to interference within the bright-field disc as:

$$d = \frac{\lambda}{2\alpha} \tag{4-9}$$

A defocus limit for the preservation of elastic contrast for a specific energy loss can hence be calculated by combining Equations (4-8) and (4-9) to give:

$$df_{MAX} = \frac{\lambda}{2\alpha\theta_E} \tag{4-10}$$

Although the above discussion only considered interference within the bright-field disc for simplicity, high spatial resolution can be achieved by using higher order diffracted discs [24-26]. Following a similar logic, Equation (4-10) also applies to higher-order diffraction and the angular distance between two high-order diffraction discs still needs to be less than 2α.

According to Equation (4-11), for the typical atomic resolution ptychography setup described in this section, df_{MAX} is approximately 1.1 μm. However, in literatures, defocus values used is for atomic resolution ptychography are usually less than ±100 nm [25-29], Therefore, a micrometre-scale defocus is experimentally sufficient to cover the range commonly used in such setups.

For a larger energy loss, such as the Carbon K-edge at 284 eV, θ_E is approximately 0.7 mrad at 200 kV. This results in a df_{MAX} of approximately 80 nm, which is still a significant defocus range for atomic resolution ptychographic reconstructions. The effects due to high-loss electrons are discussed in Section 4.3.

4.2.3.3 The Limiting Defocus

I now further explore a wider range of experimental conditions for defocus. The value of λ in Equation (4-10) is also determined by E_0, or the accelerating voltage V_0 since $E_0 = eV_0$ where e is charge of an electron.

If relativistic effects are not considered, λ can be calculated as:

$$\lambda = \frac{h}{\sqrt{2m_0 eV_0\left(1 + \frac{eV_0}{2m_0c^2}\right)}} \tag{4-11}$$

Where h is Planck's constant, m_0 is rest mass of an electron, and c is the speed of light in vacuum. If relativistic effects are considered, then Equation (4-2) can be rewritten as [10]:

$$\theta_E = \frac{\Delta E}{\frac{v^2}{c^2}(eV_0 + m_0c^2)} = \frac{\Delta E}{(eV_0 + m_0c^2) - \frac{m_0^2c^4}{eV_0 + m_0c^2}} \tag{4-12}$$

where v is the speed of an electron.

This gives an expression for df_{MAX} with variables$(V_0, \alpha, \Delta E)$ as:

$$df_{MAX} = \frac{h\left[(eV_0 + m_0c^2) - \frac{m_0^2c^4}{eV_0 + m_0c^2}\right]}{2\alpha\Delta E\sqrt{2m_0 eV_0\left(1 + \frac{eV_0}{2m_0c^2}\right)}} \tag{4-13}$$

The expression for df_{MAX}, given by Equation (4-13), under various experimental conditions, is plotted in Figure 4-6. We consider a series of typical convergent angles, from 1.5 mrad to 22.5 mrad, and acceleration voltages of 60, 200, and 300 kV. In Figure 4-6, for a given voltage and convergent angle, defocus values below the curve can be regarded as the limiting defocus setting for the corresponding energy loss. The limiting defocus is where the characteristic angle of inelastically scattered electrons for a specific energy loss is smaller than the angular distribution of the interference fringes at a 2α resolution limit.

Below the limiting defocus, inelastic scattered electrons can be ignored in typical ptychography reconstructions. In addition, in an energy-filtered reconstruction at the corresponding energy loss ΔE, the preservation of elastic contrast is the dominant factor. It does not however mean θ_E will have no influence on the diffraction patterns as the inelastic scattering distribution is still convolved with the diffraction patterns. However, the influence of this can be ignored in the ptychographic reconstruction.

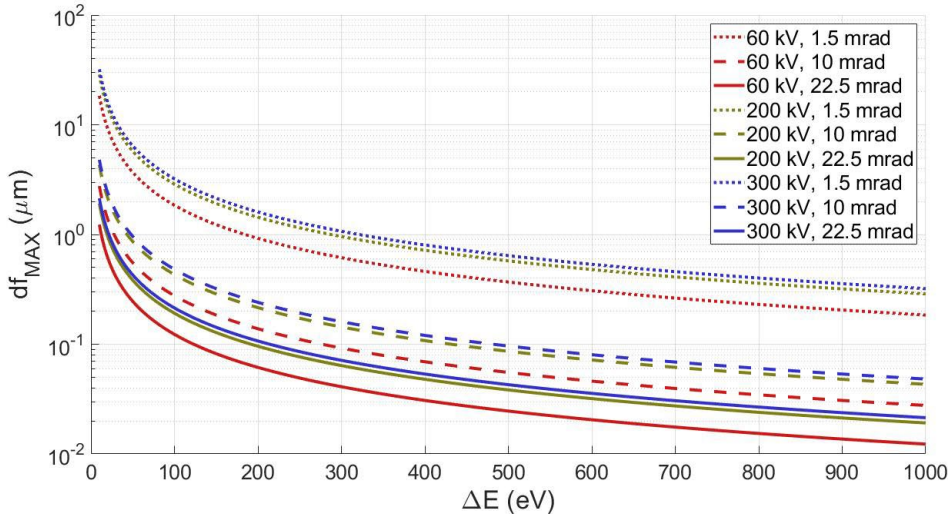

Figure 4-6 | The relation between the limiting defocus and energy loss (df_{MAX}-ΔE) for different accelerating voltages (V_0) and convergent angles (α). Calculated based on Equation (4-13). The vertical axis is in logarithmic scale.

In Figure 4-6, within the plasmon range, *i.e*. for energy losses less than 50 eV, a defocus setting of less than 200 nm is below the limiting defocus for ptychography with convergence angle of 22.5 mrad, provided the microscope is operating at an accelerating voltage of at least 60 kV. Furthermore, at low convergence angles (*e.g.*, 1.5 mrad), the limiting defocus for plasmon-excited electrons can reach up to 3 micrometres. At higher voltages, such as 300 kV, the limiting defocus can reach 8.0 μm at 40 eV energy loss. A high limiting defocus allows for the use of a larger probe size when scanning a sample with a small convergence angle. This helps to reduce the number of scanning points and the data size for large field of view experiments. This can be advantageous when using defocused probe ptychography for the reconstruction of large fields of view.

4.2.3.4 Detector Sampling

As mentioned in previous section, if the experiment setting satisfies the limiting defocus condition, the influence of inelastically scattered electron on ptychography reconstruction can be ignored. However, it does change the intensity distribution on detector. In this section, we will compare the inelastically characteristic scattering angle and detector sampling, or pixel size at the diffraction plane, to quantify to what extent inelastically scattered electrons will change diffraction patterns.

To quantify the extent to which inelastically scattered electrons alter diffraction patterns, the probability (P_s) that an inelastically scattered electron with a specific energy loss still lands on the same pixel can be estimated. This is done by convolving a normalized 2D Lorentzian distribution with a square window function (representing the pixel area). This convolutional estimation disregards multiple scattering, electron interactions, and pixel edge effects, and assumes a constant inelastic cross section across the sample. For typical 4DSTEM detectors with number of pixel from 128×128 to 512×512, the diameter of bright disc is estimated to equal 2/3 of the side length of the detector. The ratio of bright disc diameter and side length of detector can be adjusted by camera length, but here only the typical condition is considered. Therefore, a probe with a convergence angle of 22 mrad, leads to a pixel size from 0.52 mrad (on 128×128 detector) to 0.13 mrad (on 512×512 detector).

The curves for P_s under different pixel size up to 1 mrad is plotted in Figure 4-7. For example, a 20 eV energy-loss electron (accelerated at 200 kV), with a pixel size of approximately 0.5 mrad, P_s equals 50%.

For high resolution ptychography, a pixel size of around 0.5 mrad, which means the diameter of a 22 mrad bright disk is around 44 pixels. As discussed in Figure 4-5, the period of interference fringes is higher than 0.5 mrad when accelerate voltage is higher than 60 kV and defocus is less than 100 nm. Therefore, pixel size of 0.5 mrad is sufficient to satisfy the Nyquist sampling condition for the interference fringes under the condition that

accelerate voltage is higher than 60 kV and defocus is less than 100 nm. Under these conditions, statistically, about 50% of the inelastically scattered electrons in the plasmon range will still land in the same area defined by the bright field disc. Therefore, the intensity distribution of recorded diffraction patterns, with or without plasmon-excited inelastic scattered electrons, is almost the same as the total signal of zero-loss and plasmon-loss is incoherently summed on the detector.

With an increment in energy loss, for example, at 150 eV and 283.8 eV (Carbon K-edge), the P_s curves are below the 20 eV curve. This means higher energy-loss electrons may cause more significant blurring in diffraction patterns due to a higher characteristic scattering angle, as shown in Equation (4-2). However, for thin samples, their low signal counts compared to zero-loss peak and plasmon peak makes their influence on ptychography reconstruction results negligible.

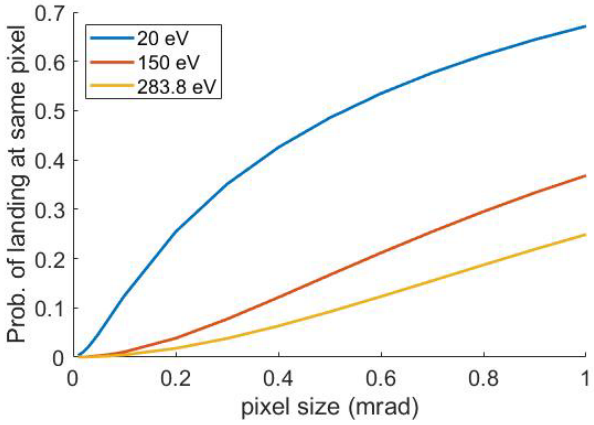

Figure 4-7 | Estimated probability (P_s) of an inelastic scattered electron (energy loss: 20, 150 and 283.8 eV) landing on the same pixel as elastic scattered. Incident beam is 200 kV. Horizontal axis is the size of pixel up to 1 mrad.

In the most experimental conditions, to ensure that the interference fringes in the diffraction pattern are adequately sampled, the period of the interference fringes is at least several times the sampling of the detector. Otherwise, the interference fringes cannot be recorded with insufficient sampling. Thus, even if inelastically scattered electrons do alter the intensity distribution of the electron beam in the detector plane, for example landed on an

adjacent pixel, their impact on the reconstruction can be negligible. This is because the partial blurring on adjacent one or two pixels cannot fully cover the interference fringes. Therefore, the size of the interference fringes in the diffraction pattern compared to the pixel size of the detector is the more critical factor.

4.3 High-loss CBED patterns

In addition to low-loss electrons, primarily consisting of plasmon-excited electrons, high-loss electrons in 4DSTEM are also considered. An energy-filtered diffraction pattern series centred at the carbon K-edge (283.8 eV), as shown in Figure 4-8(a), was recorded using the same optical settings as the energy-filtered 4DSTEM dataset in previous section. The energy loss scan step for the series was 4 eV. The counts in each diffraction pattern at different energy losses were then summed for plotting. The blue curve in Figure 4-8(b) represents the sum of diffraction patterns with respect to energy loss, plotted alongside a reference EELS spectrum shown as an orange line. A carbon K-edge peak (283.8 eV) is visible on the summed energy-filtered diffraction patterns curve in Figure 4-8(b), demonstrating that the core-loss can be recorded. However, the energy window used here does not allow the near-edge fine structure to be resolved, as seen in Figure 4-8(b). This suggests a potential method for collecting 4DSTEM datasets using core loss peaks, but it was not pursued further due to the low edge intensity leading to high levels of noise. The asymmetry in the bright-field disc is due to the dataset being slightly out of focus.

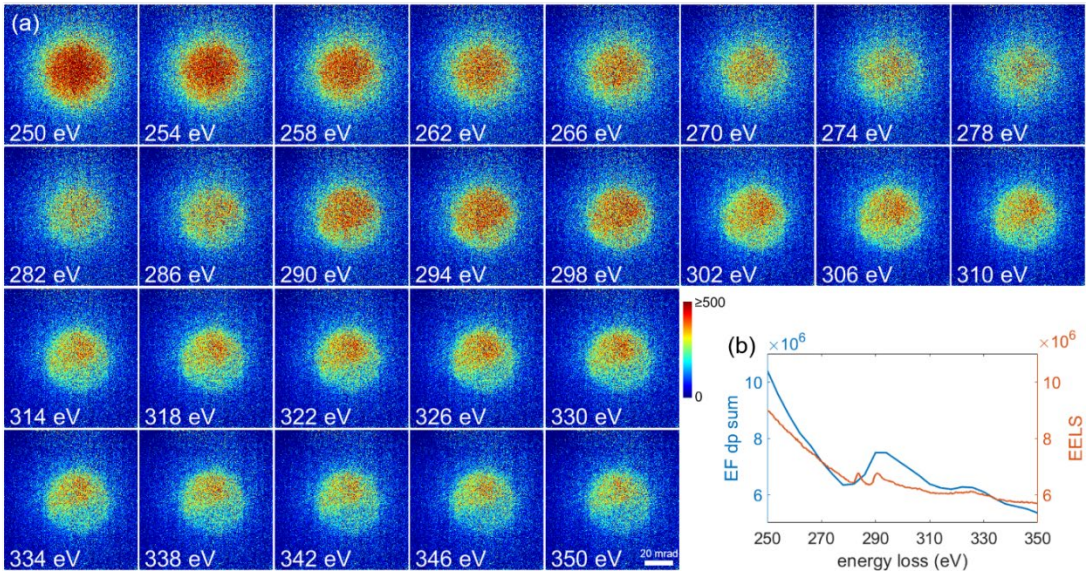

Figure 4-8 | Energy-filtered diffraction pattern series and EELS spectrum. (a) Energy-filtered CBED patterns for energy loss between 250 ~ 350 eV. The colourmap used for all CBED patterns are the same (0~500 counts), as shown in the figure. (b) EELS spectrum (orange) and sum of energy-filtered CBED patterns (blue).

In most cases, the core-loss energy range can be ignored in unfiltered ptychographic reconstruction since its intensity is extremely low compared to zero-loss and plasmon losses. Currently, we are unable to perform a ptychographic reconstruction on a 4DSTEM dataset in the core-loss range. However, it is predicted that a ptychographic reconstruction at a specific energy window in the core-loss range should be possible if the experimental conditions meet the limiting defocus condition. It has been demonstrated that low fluence ptychographic reconstructions [29] can be carried out using extremely noisy datasets. Therefore, in principle ptychographic reconstruction using core losses is feasible. However, the contrast reconstructed at a high-loss peak is still likely to be just preserving elastic contrast. Inelastic scattered electrons are still coherent with themselves [30]. This is because the inelastic and elastic scattering happens independently, and inelastic scattered CBEDs can be regarded as an extra distribution convolved on elastic scattering CBEDs, without any additional information. Unlike element mapping in EELS, the reconstructed image at a specific energy loss cannot reflect the intensity of the energy loss peak.

4.4 Influence of Thermal Diffuse Scattered Electrons

Another inelastic scattering process that occurs in a TEM is Thermal Diffuse Scattering (TDS), also known as inelastic phonon excitation. TDS electrons are commonly collected for HAADF imaging in STEM mode and also influence conventional TEM imaging [31]. However, the energy loss of phonon excitations (less than 500 meV) is much smaller than that of plasmon or single-electron excitations [32, 33]. This means that TDS electrons are usually not experimentally filtered from zero-loss datasets for ptychography reconstructions. Nevertheless, the distribution of TDS electrons contributes to increased signal at higher scattering angles compared to elastically scattered electrons [34]. The frozen-phonon multislice method is commonly used to simulate TDS electrons in TEM/STEM imaging [35-37]. The frozen-phonon approximation captures the vibrations by assuming that, at a given moment, the positions of the atoms are "frozen". A number of positions of atoms are then calculated and incoherent summed in multislice simulation independently. To test the influence of TDS electrons on ptychography reconstructions, multislice simulations as input for ptychographic reconstructions were carried out.

4.4.1 Simulation of 4DSTEM with TDS

An atomic model of monolayer graphene was used for simulation. Diffraction patterns from elastically scattered electrons and TDS electrons were simulated using the Quantum Excitation of Phonons (QEP) model in the μSTEM package [38]. Simulations were performed at an accelerating voltage of 60 kV with an aberration-free electron probe, a convergence angle of 25 mrad, and a scanning step size of 0.1 Å over a field of view (FOV) of approximately 15×15 Å². Diffraction patterns were calculated using a 256×256 pixel array. For TDS, 100 Monte Carlo simulations are conducted. The number of Monte Carlo simulations was limited by calculation time. The mean squared displacement of carbon atoms for the TDS calculations was set at 44.8 pm² (310 K), 163.5 pm² (770 K), and 236.5 pm² (1270 K), based on experimental measurements at these temperatures [39, 40].

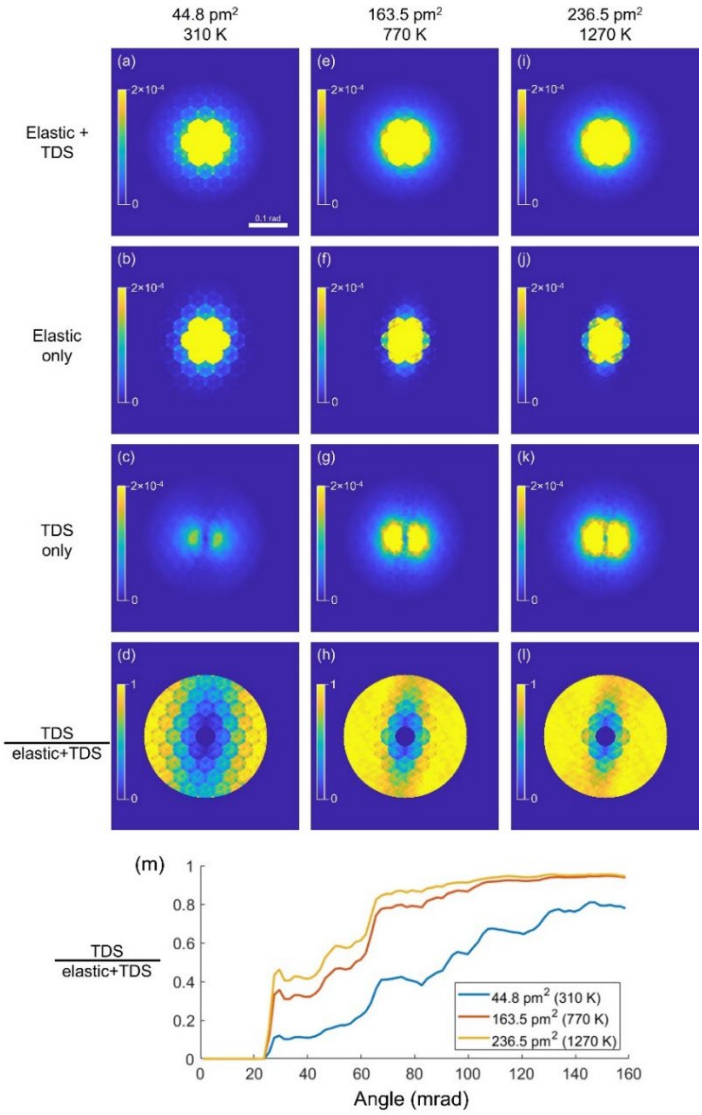

Figure 4-9 | Simulated diffraction patterns for elastic-scattered electrons and TDS electrons at different temperatures. (a, e, i) Diffraction patterns of both elastically scattered electrons and TDS electrons, with contrast boosted to emphasize high-angle features. The colormap scale is 0 ~ 2×10^{-4}, and the intensity of the brightfield disk is normalized to 1. Scale bar is 0.1 rad. (b, f, j) Diffraction patterns of only elastically scattered electrons (excluding TDS electrons). (c, g, k) Distributions of TDS electrons in the diffraction pattern. The electron beam was not focused on a high-symmetry special position within the lattice when these diffraction patterns are simulated. This resulted in the lack of circular symmetry. (d, h, l) Ratio of TDS to total electrons, *i.e.* $\frac{TDS}{TDS+elastic}$. (m) Radial averaged line profiles from (d), (h) and (l).

The diffraction patterns from all electrons, elastically scattered electrons, and TDS electrons are shown in Figure 4-9. The proportion of TDS electrons in the collected electrons is given as the ratio of TDS electrons to total electrons, *i.e.*, $\frac{TDS}{TDS+elastic}$, and is shown in Figure 4-9(d), (h) and (l). Although the distribution of TDS electrons tends to concentrate in the low-angle area, as shown in Figures 4-9(c), (g) and (k), TDS electrons contribute less than 1% of the total signal inside the bright-field disk. Averaged radial line profiles from Figures 4-9(d), (h), and (l) are plotted in Figure 4-9(m), showing the proportion of TDS electrons increasing with scattering angle and temperature. For the monolayer graphene specimen used in this simulation, the ratio of TDS electrons exceeds 50% when the scattering angle is higher than approximately 100 mrad (310 K), 60 mrad (770 K), and 45 mrad (1270 K). Simulation results are plotted in Figure 4-9.

4.4.2 Simulations of Ptychography with TDS

The simulated diffraction patterns in the previous section and Figure 4-9 are used as inputs for ptychographic reconstruction. The ePIE and SSB algorithms, introduced in Chapter 3, are applied to 4DSTEM datasets simulated using only elastic scattered electrons and elastic scattered electrons + TDS electrons. For the ePIE reconstruction, α_{ψ} (update factor for object) and α_{P} (update factor for probe) are set to 0.1 and 50 iterations are performed.

In the ePIE reconstruction results, due to elastically scattered electrons dominating the low-angle scattering, TDS electrons can be ignored if only the signals in the bright-field disk are used, as shown in Figure 4-10. In Figure 4-10, the third column shows the line profiles of reconstructed adjacent atoms under two conditions: reconstruction using only elastic scattered electrons and using elastic scattered electrons + TDS electrons. Under both conditions, the line profiles almost overlap. There is only negligible effect on the resolution of the resolution or the maximum phase value.

However, TDS electrons affect high-resolution ptychographic reconstructions that include electrons scattered to high angles. When dark-field signals are used in the reconstruction,

the reconstruction of the ePIE reconstruction is affected by the presence of TDS electrons compared to reconstructions using only elastically scattered electrons, as shown in Figure 4-11. At higher temperatures, 770 K and 1270 K, as shown in the second (770 K) and third rows (1270 K) of Figures 4-11, this effect becomes more significant. This limitation on resolution due to lattice vibration is consistent with the experiment reported by Chen *et al*. [25]. The value of mean square displacement will vary with different samples and temperature.

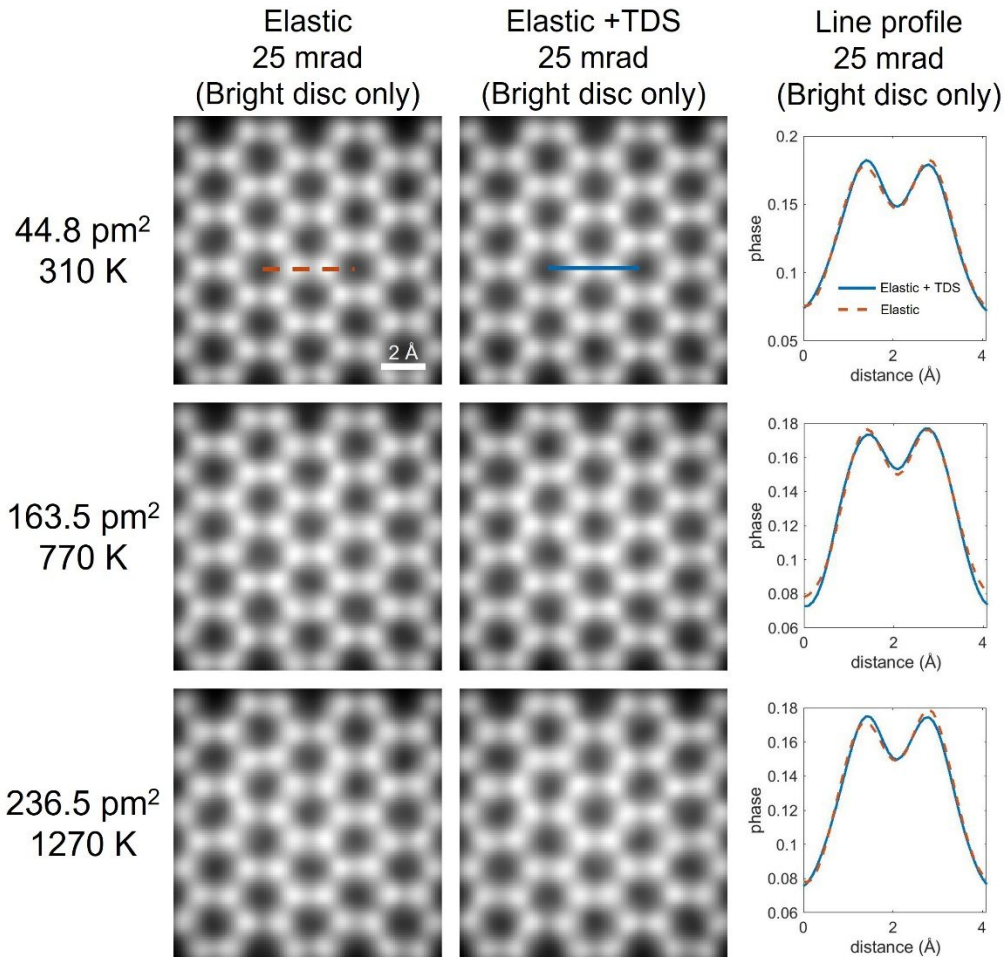

Figure 4-10 | ePIE reconstructions using simulated diffraction patterns of only the bright field disc (25 mrad), reconstructed using only elastically scattered electrons (column 1) and elastic + TDS electrons (column 2). Positions of line profiles (column 3) are marked on the first and second images in the first row. The scale bar is 2 Å.

In Figure 4-11, as the temperature increases and the amplitude of lattice vibrations grows,

the simulated reconstruction using only elastically scattered electrons (first column) is more significantly affected compared to using both elastic scattered electrons and TDS electrons (second column). In the line profile shown in Figure 4-11, it can be observed that after 50 iterations of ePIE, the reconstruction results using elastic + TDS electrons achieve similar resolution, peak width, and maximum reconstructed phase across different temperatures. In contrast, the maximum phase value reconstructed using only elastic scattered electrons is affected by temperature, showing a decrease at 770 K and 1270 K.

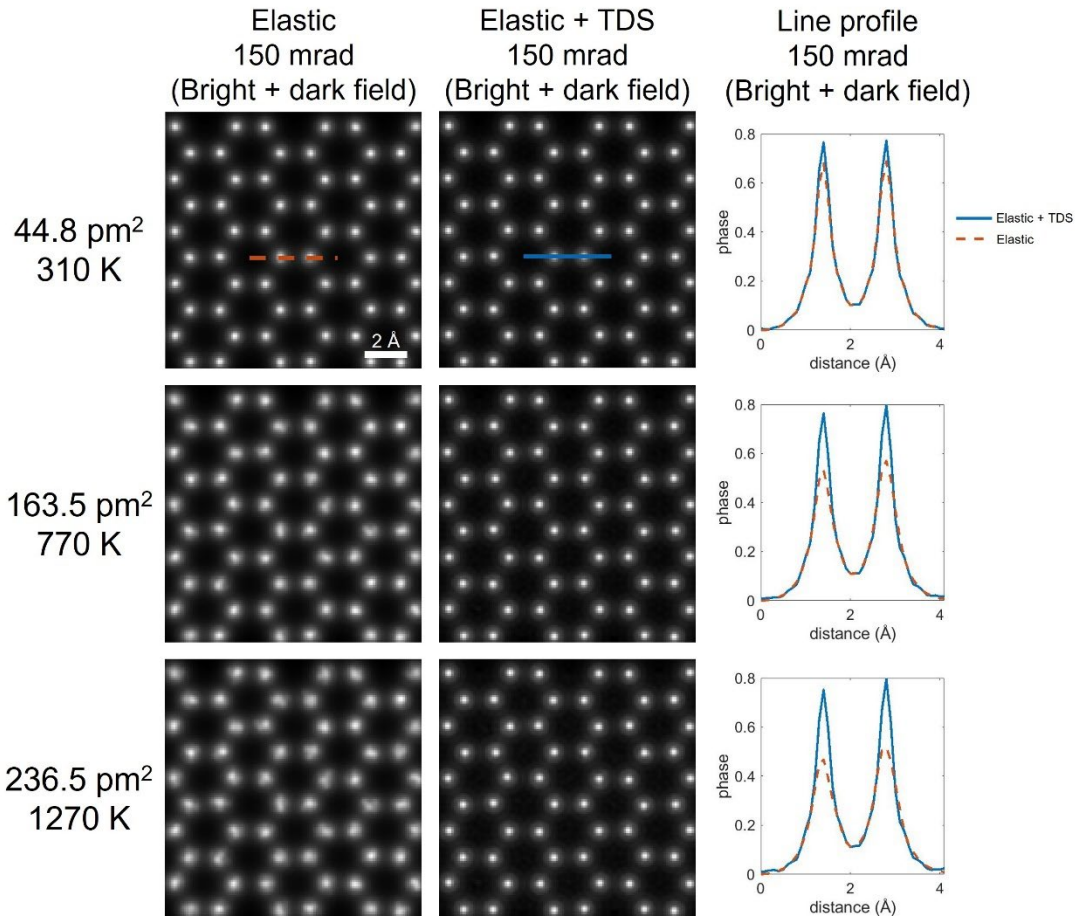

Figure 4-11 | ePIE reconstructions using simulated diffraction patterns including up to 150 mrad dark-field information, reconstructed using only elastically scattered electrons (column 1) and elastic + TDS electrons (column 2). Positions of line profiles (column 3) are marked on the first and second images in the first row. The scale bar is 2 Å.

Further simulations and experimental data are needed to confirm this phenomenon. One

possible source of error is that in the simulations described above, the Monte Carlo simulation used in the frozen phonon method only involved 100 iterations. For ePIE reconstructions that consider high-angle (up to 150 mrad) dark-field signals, this number of Monte Carlo iterations is insufficient. From the ePIE reconstruction results, at temperatures of 770 K and 1270 K, the reconstructed images using only elastic scattered electrons show that most atomic images are not circularly symmetric. This may be due to the limited number of Monte Carlo iterations, which causes the spatial positions of atomic vibrations to be represented in a discretized manner. To eliminate this error, the further simulations should use a significantly higher number of iterations than 100.

For the SSB algorithm, as discussed in Chapter 3, only double overlap areas involving the direct beam and diffraction beam are considered, so the maximum scattering angle that needs to be considered equals the convergence angle. Therefore, even at high temperature (e.g. 1270 K), TDS signals in the dark field will not affect the SSB reconstruction in contrast to the ePIE reconstruction results, as shown in Figure 4-12.

The influence of TDS needs to be considered only if high-angle dark-field signals are used in ePIE. However, under low fluence conditions, the counts in dark-field signals can be very low. Therefore, dark-field TDS signals are generally not important under such conditions. As bright-field signals are mainly used for low-fluence conditions, the TDS in the bright-field can be ignored due to the low intensity at low angle. Thus, for these two reasons, in general, the influence of TDS electrons under low-fluence conditions is negligible.

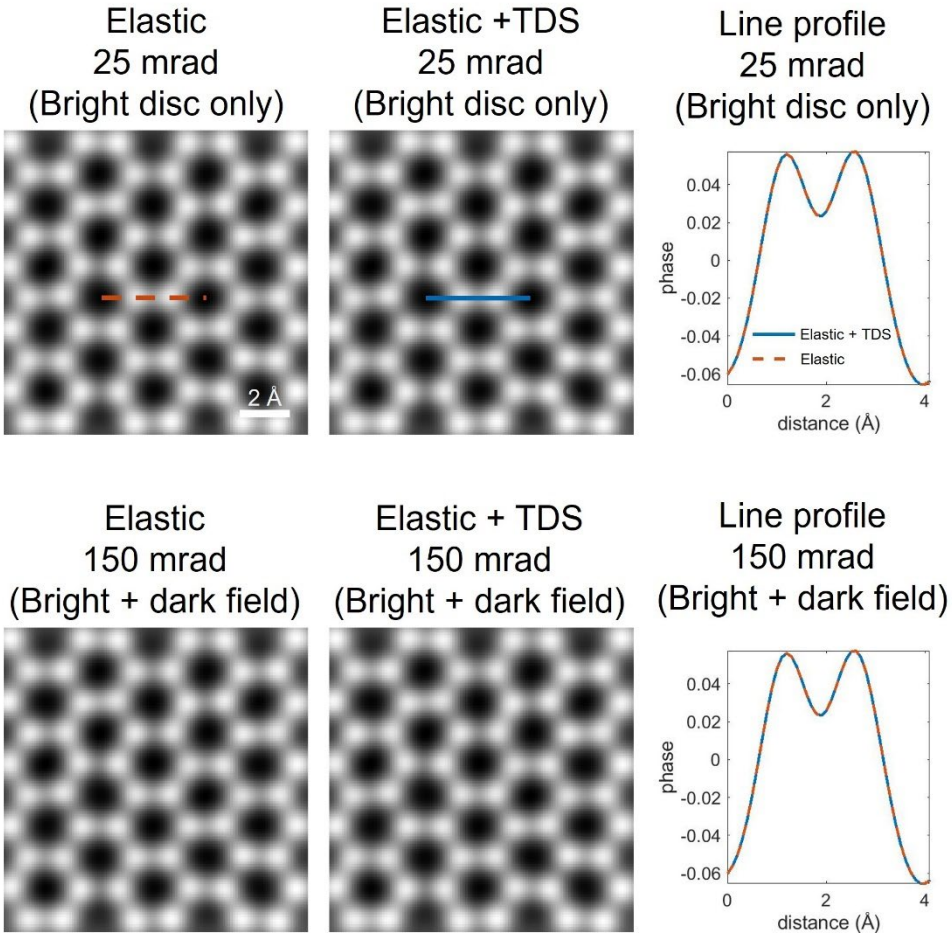

Figure 4-12 | SSB reconstructions using simulated diffraction patterns, including only the bright field disc (25 mrad) or up to 150 mrad dark field information. Reconstructions using only elastically scattered electrons are shown in the first column, while those using both elastic and TDS electrons are shown in the second column. The mean square displacement is 236.5 pm² (1270 K). Positions of line profiles (column 3) are marked on the first and second images in the first row. The scale bar is 2 Å.

4.5 Conclusion

In this Chapter, I have described the effects of inelastic electron scattering on ptychographic reconstruction, particularly with regards to interference fringes which are the source of information in ptychographic reconstruction. Inelastically scattered electrons have a negligible influence for typical high resolution (less than 2 Å) experimental settings. I have specifically discussed the influence of defocus, which controls the size of the

interference fringes on CBED patterns and predicted a limiting defocus condition for different accelerating voltages and convergent angles.

Furthermore, it has been shown that TDS electrons do not significantly influence ptychographic reconstructions that use only the signal in the bright field disc. However, they can reduce resolution in ePIE reconstructions if dark field signals are considered, particularly at higher temperatures.

In summary, if it is necessary to control the electron fluence used for imaging, not using an energy filter can allow more electrons to be detected, thereby reducing the effect of Poisson noise. This approach does not significantly affect ptychographic imaging, provided the limiting defocus condition is met. Additionally, under low-fluence conditions, TDS signals can usually be ignored because their contribution to the intensity in the bright-field disc is very small and usually only bright field is used for reconstruction due to the weak signal in dark field.

Reference

[1] R. F. Egerton, "TEM applications of EELS," in Electron Energy-Loss Spectroscopy in the Electron Microscope: Springer, 2011, pp. 293-397.

[2] M. Weyland and P. A. Midgley, "Extending energy-filtered transmission electron microscopy (EFTEM) into three dimensions using electron tomography," Microscopy and Microanalysis, vol. 9, no. 6, pp. 542-555, 2003.

[3] M. Hättestrand and H. O. Andren, "Evaluation of particle size distributions of precipitates in a 9% chromium steel using energy filtered transmission electron microscopy," Micron, vol. 32, no. 8, pp. 789-797, 2001.

[4] R. D. Leapman, E. Kocsis, G. Zhang, T. L. Talbot, and P. Laquerriere, "Three-dimensional distributions of elements in biological samples by energy-filtered electron tomography," Ultramicroscopy, vol. 100, no. 1-2, pp. 115-125, 2004.

[5] D. P. Bazett-Jones, B. Leblanc, M. Herfort, and T. Moss, "Short-range DNA looping by the Xenopus HMG-box transcription factor, xUBF," Science, vol. 264, no. 5162, pp. 1134-1137, 1994.

[6] M. Linck et al., "Chromatic aberration correction for atomic resolution TEM imaging from 20 to 80 kV," Physical review letters, vol. 117, no. 7, p. 076101, 2016.

[7] W. M. Stobbs and W. O. Saxton, "Quantitative high resolution transmission electron microscopy: the need for energy filtering and the advantages of energy-loss imaging," Journal of Microscopy, vol. 151, no. 2, pp. 171-184, 1988.

[8] O. L. Krivanek, C. C. Ahn, and G. J. Wood, "The inelastic contribution to high resolution images of defects," Ultramicroscopy, vol. 33, no. 3, pp. 177-185, 1990.

[9] J. Verbeeck, D. Van Dyck, and G. Van Tendeloo, "Energy-filtered transmission electron microscopy: an overview," Spectrochimica Acta Part B: Atomic Spectroscopy, vol. 59, no. 10-11, pp. 1529-1534, 2004.

[10] R. F. Egerton, "Physics of electron scattering," in Electron Energy-Loss Spectroscopy in the Electron Microscope, 2011, pp. 111-229.

[11] R. R. Schröder, W. Hofmann, and J. F. Ménétret, "Zero-loss energy filtering as improved imaging mode in cryoelectronmicroscopy of frozen-hydrated specimens," Journal of structural biology, vol. 105, no. 1-3, pp. 28-34, 1990.

[12] Z. L. Wang, D. Van Heerden, D. Josell, and A. J. Shapiro, "Energy-filtered high-resolution electron microscopy for quantitative solid state structure determination," Journal of research of the National Institute of Standards and Technology, vol. 102, no. 1, p. 1, 1997.

[13] K. Yonekura, M. B. Braunfeld, S. Maki-Yonekura, and D. A. Agard, "Electron energy filtering significantly improves amplitude contrast of frozen-hydrated protein at 300 kV," Journal of structural biology, vol. 156, no. 3, pp. 524-536, 2006.

[14] P. D. Nellist and J. M. Rodenburg, "Electron ptychography. I. Experimental demonstration beyond the conventional resolution limits," Acta Crystallographica Section A: Foundations of Crystallography, vol. 54, no. 1, pp. 49-60, 1998.

[15] R. F. Egerton, "Chapter 5: TEM applications of EELS," in Electron Energy-Loss Spectroscopy in the Electron Microscope, 2011, ch. 5, pp. 293-397.

[16] K. Iakoubovskii, K. Mitsuishi, Y. Nakayama, and K. Furuya, "Mean free path of inelastic electron scattering in elemental solids and oxides using transmission electron microscopy: Atomic number dependent oscillatory behavior," Physical Review B, vol. 77, no. 10, p. 104102, 2008.

[17] R. F. Egerton, "Appendix C: Plasmon Energies and Inelastic Mean Free Paths," in Electron Energy-Loss Spectroscopy in the Electron Microscope, 2011, pp. 419-422.

[18] S. L. Chang, C. Dwyer, J. Barthel, C. B. Boothroyd, and R. E. Dunin-Borkowski, "Performance of a direct detection camera for off-axis electron holography," Ultramicroscopy, vol. 161, pp. 90-97, 2016.

[19] J. M. Rodenburg, B. C. McCallum, and P. D. Nellist, "Experimental tests on double-resolution coherent imaging via STEM," Ultramicroscopy, vol. 48, no. 3, pp. 304-314, 1993.

[20] A. M. Maiden and J. M. Rodenburg, "An improved ptychographical phase retrieval

algorithm for diffractive imaging," Ultramicroscopy, vol. 109, no. 10, pp. 1256-1262, 2009.

[21] I. Lazić, E. G. T. Bosch, and S. Lazar, "Phase contrast STEM for thin samples: Integrated differential phase contrast," Ultramicroscopy, vol. 160, pp. 265-280, 2016.

[22] Y. Zhu, A. Soukiassian, D. G. Schlom, D. Muller, and C. Dwyer, "Towards artifact-free atomic-resolution elemental mapping with electron energy-loss spectroscopy," Applied physics letters, vol. 103, no. 14, 2013.

[23] H. G. Brown, A. J. D'Alfonso, B. D. Forbes, and L. J. Allen, "Addressing preservation of elastic contrast in energy-filtered transmission electron microscopy," Ultramicroscopy, vol. 160, pp. 90-97, Jan 2016.

[24] Y. Jiang et al., "Electron ptychography of 2D materials to deep sub-ångström resolution," Nature, vol. 559, no. 7714, pp. 343-349, 2018.

[25] Z. Chen et al., "Electron ptychography achieves atomic-resolution limits set by lattice vibrations," Science, vol. 372, no. 6544, pp. 826-831, 2021.

[26] K. X. Nguyen et al., "Achieving sub-0.5-angstrom–resolution ptychography in an uncorrected electron microscope," Science, vol. 383, no. 6685, pp. 865-870, 2024.

[27] P. Wang, F. Zhang, S. Gao, M. Zhang, and A. I. Kirkland, "Electron ptychographic diffractive imaging of boron atoms in LaB6 crystals," Scientific Reports, vol. 7, no. 1, pp. 1-8, 2017.

[28] H. Yang et al., "Electron ptychographic phase imaging of light elements in crystalline materials using Wigner distribution deconvolution," Ultramicroscopy, vol. 180, pp. 173-179, 2017.

[29] J. Song et al., "Atomic resolution defocused electron ptychography at low dose with a fast, direct electron detector," Scientific reports, vol. 9, no. 1, pp. 1-8, 2019.

[30] A. Howie, "Inelastic scattering of electrons by crystals. I. The theory of small-angle in elastic scattering," Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences, vol. 271, no. 1345, pp. 268-287, 1963.

[31] B. D. Forbes et al., "Thermal diffuse scattering in transmission electron microscopy," Ultramicroscopy, vol. 111, no. 12, pp. 1670-1680, 2011.

[32] C. A. Gadre et al., "Nanoscale imaging of phonon dynamics by electron microscopy," Nature, vol. 606, no. 7913, pp. 292-297, 2022.

[33] R. Qi et al., "Measuring phonon dispersion at an interface," Nature, vol. 599, no. 7885, pp. 399-403, 2021.

[34] Z. L. Wang, "Thermal diffuse scattering in sub-angstrom quantitative electron microscopy—phenomenon, effects and approaches," Micron, vol. 34, no. 3-5, pp. 141-155, 2003.

[35] E. J. Kirkland, "Multislice Applications and Examples," in Advanced computing in electron microscopy, vol. 12: Springer, 1998, ch. 7, pp. 163-197.

[36] R. F. Loane, P. Xu, and J. Silcox, "Thermal vibrations in convergent-beam electron diffraction," Acta Crystallographica Section A: Foundations of Crystallography, vol. 47, no. 3, pp. 267-278, 1991.

[37] S. Hillyard and J. Silcox, "Detector geometry, thermal diffuse scattering and strain effects in ADF STEM imaging," Ultramicroscopy, vol. 58, no. 1, pp. 6-17, 1995.

[38] L. J. Allen and S. D. Findlay, "Modelling the inelastic scattering of fast electrons," Ultramicroscopy, vol. 151, pp. 11-22, 2015.

[39] C. S. Allen et al., "Temperature dependence of atomic vibrations in mono-layer graphene," Journal of Applied Physics, vol. 118, no. 7, 2015.

[40] B. Shevitski et al., "Dark-field transmission electron microscopy and the Debye-Waller factor of graphene," Physical Review B, vol. 87, no. 4, p. 045417, 2013.

Chapter 5:
Single-side Band Masked Centre-of-Mass Imaging

5.1 Introduction

Centre-of-mass (COM) is a phase-contrast imaging method in scanning transmission electron microscopy (STEM). The contrast in COM and integrated COM (iCOM) arises from the movement of the centre of mass (COM) of a convergent beam electron diffraction (CBED) pattern. Related scanning phase-contrast methods were first introduced to transmission electron microscopes by H. Rose [1], N. Dekkers and H. de Lang in 1974 [2]. Later, E. Waddell and J. Chapman discussed a developed method which is the prototype of centre-of-mass (COM) [3]. Following this, differential phase contrast (DPC) methods, which utilize segmented detectors, have been reported [4, 5] and can be considered an approximation of COM methods for pixelated detectors. Depending on the processing algorithm, COM can be further divided into COMx/y (centre-of-mass shift in specific directions) and iCOM (integrated centre-of-mass) [4]. The linearity of DPC/COM contrast has been discussed [6], showing that for a phase object, the contrast of COMx/y and iCOM images are proportional to the projected electric field and potential distribution (phase change of the exit wave resulting from the projected potential of the sample) respectively [4].

With the development of 4DSTEM techniques, COM/iCOM and related DPC methods are used in 4DSTEM datasets to obtain fast (or real-time) phase-contrast images [7]. As

described in Chapter 3, COM/iCOM methods are alternative form of phase imaging, allowing COM/iCOM or DPC/iDPC (integrated Differential phase contrast) images to map projected electric field and potential distributions [8-11].

However, in principle, the COM/iCOM method requires an in-focus probe to achieve high-resolution imaging. A defocused probe will blur COM/iCOM images due to the convolution between the probe function and the transmission function (object function) [4]. This presents a limitation when minimizing the fluence to control beam damage on sensitive samples, as fine-tuning the in-focus probe necessitates additional fluence on the sample. This does not contribute to the final image but can cause damage.

In this chapter, Single-Side Band masked COM (SSBm-COM) will be discussed within the Weak Phase Object Approximation (WPOA). SSBm-COM can correct aberrations (including defocus) after 4DSTEM data collection and reduce high-frequency noise. In this method, a Fourier transform with respect to the probe scanning positions is calculated to transform to reciprocal space, and a mask similar to the one used in Single-Side Band (SSB) ptychography [12, 13] is applied. This mask optimally filters the original dataset, removing pixels that lack phase-contrast information and contain only noise. Therefore, in addition to aberration correction, this method effectively reduces the high-frequency noise present in traditional iCOM images.

5.2 SSBm-COM Algorithm

In this section, I discuss algorithm for SSBm-COM/iCOM and their theoretical background.

As discussed in Chapter 3, a 4DSTEM dataset can be considered as an electron intensity distribution in 4 dimensions, denoted as $\left|M\left(\boldsymbol{K_f}, \boldsymbol{R_p}\right)\right|^2$, where $M\left(\boldsymbol{K_f}, \boldsymbol{R_p}\right)$ is:

$$M\left(\boldsymbol{K_f}, \boldsymbol{R_p}\right) = \left[A\left(\boldsymbol{K_f}\right) \cdot exp\left(i2\pi \boldsymbol{K_f} \cdot \boldsymbol{R_p}\right)\right] \otimes_{\boldsymbol{K_f}} \Psi\left(\boldsymbol{K_f}\right) \tag{5-1}$$

In Equation (5-1), $\boldsymbol{K_f}$, noted in bold style as a vector $\left(K_{fx}, K_{fy}\right)$, represents the coordinates at the diffraction plane (or detector plane), while $\boldsymbol{R_p}$ represents coordinates at the object

plane (scanning coordinates on the sample plane). $A(\boldsymbol{K_f})$ is the complex aperture function, and $\Psi(\boldsymbol{K_f})$ is the Fourier transform of the sample complex transmission function and $\otimes_{\boldsymbol{K_f}}$ denotes a convolution with respect to $\boldsymbol{K_f}$.

For the traditional COM/iCOM algorithm, $COM_x(\boldsymbol{R_p})$, $COM_y(\boldsymbol{R_p})$ and $\boldsymbol{COM}(\boldsymbol{R_p})$ are calculated as:

$$\begin{aligned} COM_x(\boldsymbol{R_p}) &= \iint K_{fx} \cdot \left|M(\boldsymbol{K_f}, \boldsymbol{R_p})\right|^2 d\boldsymbol{K_f} \\ COM_y(\boldsymbol{R_p}) &= \iint K_{fy} \cdot \left|M(\boldsymbol{K_f}, \boldsymbol{R_p})\right|^2 d\boldsymbol{K_f} \\ \boldsymbol{COM}(\boldsymbol{R_p}) &= \left(COM_x(\boldsymbol{R_p}), COM_y(\boldsymbol{R_p})\right) \end{aligned} \tag{5-2}$$

Subsequently, $iCOM(\boldsymbol{R_p})$ is calculated from $\boldsymbol{COM}(\boldsymbol{R_p})$ [4] as :

$$iCOM(\boldsymbol{R_p}) = \mathcal{F}^{-1}\left\{\frac{\boldsymbol{Q_p} \cdot \mathcal{F}\{\boldsymbol{COM}(\boldsymbol{R_p})\}}{2\pi i \boldsymbol{Q_p^2}}\right\} \tag{5-3}$$

$\mathcal{F}$ and $\mathcal{F}^{-1}$ are Fourier and inverse Fourier transform operators. Equation (5-3) is the traditional method for calculating iCOM image.

For SSBm-COM, a Fourier transform is applied to $COM_x(\boldsymbol{R_p})$ and $COM_y(\boldsymbol{R_p})$. Since the variable $\boldsymbol{R_p}$ is independent with respect to an integration variable $\boldsymbol{K_f}$, the order of the Fourier transform and the integral calculation can be interchanged. Therefore, a Fourier transform of $COM_x(\boldsymbol{R_p})$ and $COM_y(\boldsymbol{R_p})$, which is denoted as $com_x(\boldsymbol{Q_p})$ and $com_y(\boldsymbol{Q_p})$, can be calculated as in Equation (5-4). The function $\mathrm{G}(\boldsymbol{K_f}, \boldsymbol{Q_p})$ the Fourier transform of $\left|M(\boldsymbol{K_f}, \boldsymbol{R_p})\right|^2$ with respect to $\boldsymbol{R_p}$, as used in ptychography [13, 14].

$$
\begin{aligned}
com_x(\boldsymbol{Q_p}) &= \mathcal{F}_{\boldsymbol{R_p}}\{COM_x(\boldsymbol{R_p})\} = \mathcal{F}_{\boldsymbol{R_p}}\left\{\iint K_{fx}\cdot\left|M(\boldsymbol{K_f},\boldsymbol{R_p})\right|^2 d\boldsymbol{K_f}\right\} \\
&= \iint K_{fx}\mathrm{G}(\boldsymbol{K_f},\boldsymbol{Q_p})d\boldsymbol{K_f} \\
com_y(\boldsymbol{Q_p}) &= \mathcal{F}_{\boldsymbol{R_p}}\{COM_y(\boldsymbol{R_p})\} = \mathcal{F}_{\boldsymbol{R_p}}\left\{\iint K_{fy}\cdot\left|M(\boldsymbol{K_f},\boldsymbol{R_p})\right|^2 d\boldsymbol{K_f}\right\} \\
&= \iint K_{fy}\mathrm{G}(\boldsymbol{K_f},\boldsymbol{Q_p})d\boldsymbol{K_f}
\end{aligned}
\tag{5-4}
$$

$\mathrm{G}(\boldsymbol{K_f},\boldsymbol{Q_p})$ can be written as a convolution, as given in Equation (5-5) [15, 16], and as discussed in Chapter 3.

$$
\begin{aligned}
\mathrm{G}(\boldsymbol{K_f},\boldsymbol{Q_p}) &= \mathcal{F}_{\boldsymbol{R_p}}\left\{\left|M(\boldsymbol{K_f},\boldsymbol{R_p})\right|^2\right\} \\
&= A(\boldsymbol{K_f})A^*(\boldsymbol{K_f}+\boldsymbol{Q_p}) \otimes_{\boldsymbol{K_f}} \Psi(\boldsymbol{K_f})\Psi^*(\boldsymbol{K_f}-\boldsymbol{Q_p})
\end{aligned}
\tag{5-5}
$$

Here, the Weak Phase Object Approximation (WPOA) is incorporated into Equation (5-5). The WPOA was discussed in Chapter 3 and the same notation is used in this chapter. ψ_s is used to represent the transmission function of a weak phase object, and Ψ_s is its Fourier transform. The relation of the original transmission function ψ and ψ_s is defined as follows:

$$
\psi(\boldsymbol{R_p}) = 1 + \psi_s(\boldsymbol{R_p}) \tag{5-6}
$$

The Fourier transform of ψ_s is given as Ψ_s in Equation (5-7):

$$
\Psi_s(\boldsymbol{Q_p}) = \mathcal{F}\{\psi_s(\boldsymbol{R_p})\} \tag{5-7}
$$

Under the WPOA, Equation (5-5) can be rewritten using a replacement of Ψ with Ψ_s. As a result, $\mathrm{G}(\boldsymbol{K_f},\boldsymbol{Q_p})$ can be further simplified [13] as in Equation (5-8).

$$
\begin{aligned}
\mathrm{G}(\boldsymbol{K_f},\boldsymbol{Q_p}) \approx{}& \left|A(\boldsymbol{K_f})\right|^2\delta(\boldsymbol{Q_p}) + A(\boldsymbol{K_f})A^*(\boldsymbol{K_f}+\boldsymbol{Q_p})\Psi_s^*(-\boldsymbol{Q_p}) \\
&+ A^*(\boldsymbol{K_f})A(\boldsymbol{K_f}-\boldsymbol{Q_p})\Psi_s(\boldsymbol{Q_p})
\end{aligned}
\tag{5-8}
$$

In Equation (5-8), the first item is 0 when $\boldsymbol{Q_p} \neq 0$, and has no influence on non-zero

frequencies. Alternatively, the first item can be considered as a constant background and does not affect the contrast of the reconstruction.

The second and third items are used for SSB ptychography reconstruction [13] and were discussed in Chapter 3. For a specific $\boldsymbol{Q_p}$, $A(\boldsymbol{K_f})A^*(\boldsymbol{K_f}+\boldsymbol{Q_p})$ or $A^*(\boldsymbol{K_f})A(\boldsymbol{K_f}-\boldsymbol{Q_p})$ represents the overlapping region of an aperture function (or zero-beam), $A(\boldsymbol{K_f})$ or $A^*(\boldsymbol{K_f})$, and the aperture function shifted by $\boldsymbol{Q_p}$ or $-\boldsymbol{Q_p}$ in the $\boldsymbol{K_f}$ plane, as shown in Figure 1 (a). The region overlapped only by the zero-shift aperture function ($A(\boldsymbol{K_f})$ or $A^*(\boldsymbol{K_f})$), and a $-\boldsymbol{Q_p}$ or $\boldsymbol{Q_p}$ beam is denoted as a "double-overlap (DO) region". The region overlapped by $A(\boldsymbol{K_f})$ or $A^*(\boldsymbol{K_f})$ and both $-\boldsymbol{Q_p}$ and $\boldsymbol{Q_p}$ beam is denoted as the "triple-overlap (TO) region". A schematic diagram showing the DO and TO regions is given in Figure 1 (a). The aperture function $A(\boldsymbol{K_f})$ is a top-hat function at $|\boldsymbol{K_f}| \leq \alpha$, where α is the convergence angle, with a phase modified by the objective lens aberrations. The phase information in $\Psi_s(\boldsymbol{Q_p})$ can be analysed in the DO regions, where only one diffracted beam interferes with the non-shifted aperture function. An example of a simulated $\mathrm{G}(\boldsymbol{K_f}, \boldsymbol{Q_p})$ slice with defocus is shown in Figure 5-1 (b) and (c).

In Figure 5-1 (a), the highest $|\boldsymbol{Q_p}|$ value for which $\mathrm{G}(\boldsymbol{K_f}, \boldsymbol{Q_p})$ in Equation (5-8) remains non-zero is $|\boldsymbol{Q_p}| = 2\alpha$, where α is the convergence angle of the probe. The DO and TO regions discussed in the previous paragraph are defined based on condition $|\boldsymbol{Q_p}| \leq 2\alpha$. When $|\boldsymbol{Q_p}| > 2\alpha$, $A(\boldsymbol{K_f})$ and $A(\boldsymbol{K_f}+\boldsymbol{Q_p})$ have no overlapping area, neither DO or TO regions, resulting in $\mathrm{G}(\boldsymbol{K_f}, \boldsymbol{Q_p}) = 0$. This condition will be referenced in the next section to discuss frequency transfer.

In Equation (5-8), it is shown that sample information (Ψ_s) is only contained in the two DO regions. Outside the DO region, the terms $A(\boldsymbol{K_f})A^*(\boldsymbol{K_f}+\boldsymbol{Q_p})$ and $A^*(\boldsymbol{K_f})A(\boldsymbol{K_f}-\boldsymbol{Q_p})$ are zero, blocking information transfer outside DO regions as described in Equation (5-8). The TO region will be zero if 4DSTEM dataset is in-focus (as discussed in Chapter 3), or a combination of Ψ_s and Ψ_s^* when out of focus, from which it is difficult to separate information in Ψ_s. Equation (5-8) provides a simple solution for Ψ_s or Ψ_s^* by creating masks for DO regions and simplify this to 2 separate terms containing object information.

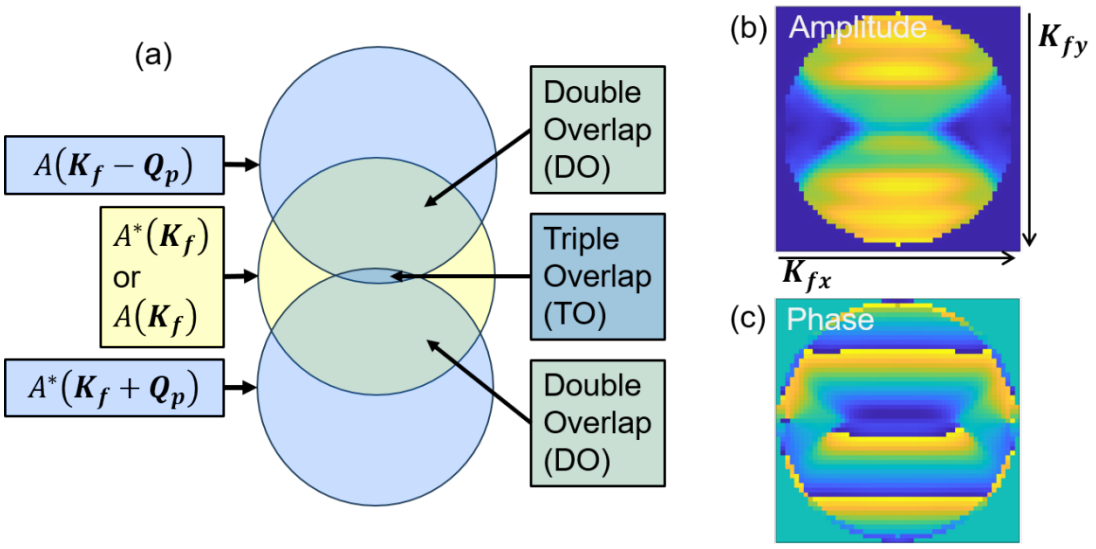

Figure 5-1 | Schematic showing the G matrix and DO/TO regions (a). (b) and (c) a slice along G from the DO region. The structures caused by defocus can be seen (b) in the amplitude and (c) phase of $\mathrm{G}(\boldsymbol{K_f}, \boldsymbol{Q_p})$. Acceleration voltage: 80kV, defocus 10 nm.

Aberration correction can be achieved by modifying the DO mask. The mask is generated by multiplying a conjugate term, $A^*(\boldsymbol{K_f})A(\boldsymbol{K_f}+\boldsymbol{Q_p})$ and $A(\boldsymbol{K_f})A^*(\boldsymbol{K_f}-\boldsymbol{Q_p})$, with the DO term in $\mathrm{G}(\boldsymbol{K_f}, \boldsymbol{Q_p})$. Consequently, a mask function, $m(\boldsymbol{K_f}, \boldsymbol{Q_p})$, including aberrations can be defined as:

$$m(\boldsymbol{K_f}, \boldsymbol{Q_p}) = \begin{cases} A^*(\boldsymbol{K_f})A(\boldsymbol{K_f}+\boldsymbol{Q_p}), & in\ \boldsymbol{Q_p}\ beam\ DO\ region \\ A(\boldsymbol{K_f})A^*(\boldsymbol{K_f}-\boldsymbol{Q_p}), & in -\boldsymbol{Q_p}\ beam\ DO\ region \\ 0, & out\ of\ DO\ region \end{cases} \quad (5\text{-}9)$$

The mask for DO regions is applied in the SSBm-COM, to $\mathrm{G}(\boldsymbol{K_f}, \boldsymbol{Q_p})$ in Equation (5-4). For comparison, the traditional COM method, can be regarded as having a mask function $m(\boldsymbol{K_f}, \boldsymbol{Q_p}) = 1$ at the same position in Equation (5-4). As a result, the new masked expressions for $com_x(\boldsymbol{Q_p})$ and $com_x(\boldsymbol{Q_p})$ are denoted as $com_x^f(\boldsymbol{Q_p})$ and $com_y^f(\boldsymbol{Q_p})$, respectively. Here, the superscript "f" indicates "filtered", referring to the filtering of pixels outside the DO region.

$$com_x^f(\boldsymbol{Q_p}) = \iint K_{fx}\mathrm{G}(\boldsymbol{K_f}, \boldsymbol{Q_p})m(\boldsymbol{K_f}, \boldsymbol{Q_p})d\boldsymbol{K_f}$$
$$com_y^f(\boldsymbol{Q_p}) = \iint K_{fy}\mathrm{G}(\boldsymbol{K_f}, \boldsymbol{Q_p})m(\boldsymbol{K_f}, \boldsymbol{Q_p})d\boldsymbol{K_f} \quad (5\text{-}10)$$

In filtered $com_x^f(\boldsymbol{Q_p})$ and $com_y^f(\boldsymbol{Q_p})$, there is an intrinsic filtering effect based on different $\boldsymbol{Q_p}$, as the area of mask varies when changing $\boldsymbol{Q_p}$. This acts as a bandpass filter, suppressing high-frequency noise near the limit at 2α. The filtering is also incorporated into the contrast transfer function of SSBm-COM, which will be discussed in next section.

After applying the mask, filtered COM images can be calculated by performing an inverse Fourier transform on $com_x^f(\boldsymbol{Q_p})$ and $com_y^f(\boldsymbol{Q_p})$:

$$COM_x^f(\boldsymbol{R_p}) = \mathcal{F}^{-1}\{com_x^f(\boldsymbol{Q_p})\}$$
$$COM_y^f(\boldsymbol{R_p}) = \mathcal{F}^{-1}\{com_y^f(\boldsymbol{Q_p})\} \quad (5\text{-}11)$$
$$\boldsymbol{COM^f}(\boldsymbol{R_p}) = \left(COM_x^f(\boldsymbol{R_p}), COM_y^f(\boldsymbol{R_p})\right)$$

Filtered $iCOM^f(\boldsymbol{R_p})$ can be calculated using the traditional algorithm with a replaced $\boldsymbol{com^f}(\boldsymbol{Q_p})$:

$$iCOM^f(\boldsymbol{R_p}) = \boldsymbol{\mathcal{F}}^{-1}\left\{\frac{\boldsymbol{Q_p} \cdot \boldsymbol{com^f}(\boldsymbol{Q_p})}{2\pi i \boldsymbol{Q_p^2}}\right\} \quad (5\text{-}12)$$

Importantly, the filtered $\boldsymbol{COM^f}$, $iCOM^f$, $dCOM^f$ are aberration corrected.

5.3 Transfer Function

The information transfer function of SSBm-COM and SSBm-iCOM is calculated and discussed in this section. As given in equation (5-7), the sample in the previous derivation is assumed to be a weak phase object. Therefore, although $\psi_s(\boldsymbol{R_p})$ is a complex function only the imaginary part, applies. Consequently, the Fourier transform of $\psi_s(\boldsymbol{R_p})$, $\Psi_s(\boldsymbol{Q_p})$ is conjugate antisymmetric [13].

$$\Psi_s(\boldsymbol{Q_p}) = -\Psi_s^*(-\boldsymbol{Q_p}) \tag{5-13}$$

Therefore, the second and third items in $\mathrm{G}(\boldsymbol{K_f}, \boldsymbol{Q_p})$ can be written as:

$$\mathrm{G}(\boldsymbol{K_f}, \boldsymbol{Q_p}) = \left|A(\boldsymbol{K_f})\right|^2 \delta(\boldsymbol{Q_p}) + \Gamma(\boldsymbol{K_f}, \boldsymbol{Q_p})\Psi_s(\boldsymbol{Q_p}) \tag{5-14}$$

Where a $\Gamma(\boldsymbol{Kf}, \boldsymbol{Qp})$ is introduced as in [17]:

$$\Gamma(\boldsymbol{K_f}, \boldsymbol{Q_p}) = A^*(\boldsymbol{K_f})A(\boldsymbol{K_f} - \boldsymbol{Q_p}) - A(\boldsymbol{K_f})A^*(\boldsymbol{K_f} + \boldsymbol{Q_p}) \tag{5-15}$$

Similar to $\mathrm{G}(\boldsymbol{K_f}, \boldsymbol{Q_p})$ in Equation (5-8), $\boldsymbol{Q_p}$ in Equation (5-15) must satisfy the condition $\boldsymbol{Q_p} \leq 2\alpha$ for $\Gamma(\boldsymbol{K_f}, \boldsymbol{Q_p})$ to be non-zero.

Here, taking SSBm-COMx, *i.e.* $COM_x^f(\boldsymbol{R_p})$, as an example, the Fourier transform of $COM_x^f(\boldsymbol{R_p})$ can be represented using $\Gamma(\boldsymbol{K_f}, \boldsymbol{Q_p})$ as follows:

$$\begin{aligned} \mathcal{F}\{COM_x^f(\boldsymbol{R_p})\} = com_x^f(\boldsymbol{Q_p}) &= \iint K_{fx}\mathrm{G}(\boldsymbol{K_f}, \boldsymbol{Q_p})m(\boldsymbol{K_f}, \boldsymbol{Q_p})d\boldsymbol{K_f} \\ &= \iint K_{fx}\left[\left|A(\boldsymbol{K_f})\right|^2 \delta(\boldsymbol{Q_p}) + \Gamma(\boldsymbol{K_f}, \boldsymbol{Q_p})\Psi_s(\boldsymbol{Q_p})\right] m(\boldsymbol{K_f}, \boldsymbol{Q_p})d\boldsymbol{K_f} \end{aligned} \tag{5-16}$$

If $\boldsymbol{Q_p} \neq 0$, then the Dirac function $\delta(\boldsymbol{Q_p})$ can be ignored and equation (5-16) can be simplified to:

$$\begin{aligned} \mathcal{F}\{COM_x^f(\boldsymbol{R_p})\} &= \iint K_{fx}\Gamma(\boldsymbol{K_f}, \boldsymbol{Q_p})\Psi_s(\boldsymbol{Q_p})m(\boldsymbol{K_f}, \boldsymbol{Q_p})d\boldsymbol{K_f} \\ &= \Psi_s(\boldsymbol{Q_p}) \iint K_{fx}\Gamma(\boldsymbol{K_f}, \boldsymbol{Q_p})m(\boldsymbol{K_f}, \boldsymbol{Q_p})d\boldsymbol{K_f} \end{aligned} \tag{5-17}$$

Consequently, when $\psi_s(\boldsymbol{R_p})$ is considered as the source signal, the transfer function (TF) for $COM_x^f(\boldsymbol{R_p})$ is:

$$TF_x^f(\boldsymbol{Q_p}) = \iint K_{fx}\Gamma(\boldsymbol{K_f}, \boldsymbol{Q_p})m(\boldsymbol{K_f}, \boldsymbol{Q_p})d\boldsymbol{K_f} \tag{5-18}$$

Following the same calculations for $COM_y^f(\boldsymbol{R_p})$, the transfer function is in the same format

but in K_{fy} direction.

$$TF_y^f(\boldsymbol{Q_p}) = \iint K_{fy}\Gamma(\boldsymbol{K_f}, \boldsymbol{Q_p}) m\big(\boldsymbol{K_f}, \boldsymbol{Q_p}\big) d\boldsymbol{K_f} \tag{5-19}$$

In Equation (5-18) and (5-19), since $\Gamma\big(\boldsymbol{K_f}, \boldsymbol{Q_p}\big) = 0$ when $\boldsymbol{Q_p} > 2\alpha$, the highest frequency transferable for SSBm-COMx/y is corresponding to 2α.

Based on Equation (5-12), the Fourier transform of SSBm-iCOM can be written as:

$$\begin{aligned}\mathcal{F}\{iCOM^f(\boldsymbol{R_p})\} &= \frac{\boldsymbol{Q_p} \cdot \boldsymbol{com^f}(\boldsymbol{Q_p})}{2\pi i \boldsymbol{Q_p^2}} = \frac{Q_{px} com_x^f(\boldsymbol{Q_p}) + Q_{py} com_y^f(\boldsymbol{Q_p})}{2\pi i \boldsymbol{Q_p^2}} \\ &= \Psi_s(\boldsymbol{Q_p}) \cdot \frac{Q_{px} TF_x^f(\boldsymbol{Q_p}) + Q_{py} TF_y^f(\boldsymbol{Q_p})}{2\pi i \boldsymbol{Q_p^2}}\end{aligned} \tag{5-20}$$

Therefore, the transfer function of SSBm-iCOM is given as:

$$TF_{iCOM}^f = \frac{Q_{px} TF_x^f(\boldsymbol{Q_p}) + Q_{py} TF_y^f(\boldsymbol{Q_p})}{2\pi i \boldsymbol{Q_p^2}} \tag{5-21}$$

In Equation (5-21), the numerator becomes zero when $\boldsymbol{Q_p} > 2\alpha$, meaning the highest transferable frequency for SSBm-iCOM is 2α, or equivalently, the resolution is $\frac{\lambda}{2\alpha}$. This resolution is the same as that of SSB (see Chapter 3), as both are limited by $\boldsymbol{Q_p} \leq 2\alpha$ in Equation (5-8).

Using Equations (5-18), (5-19) and (5-21), simulated 2D TFs for SSBm-COMx,y and SSBm-iCOM are shown in Figure 5-2. The simulation was calculated at 200 kV based on an aberration-free probe with a convergence angle of 20 mrad.

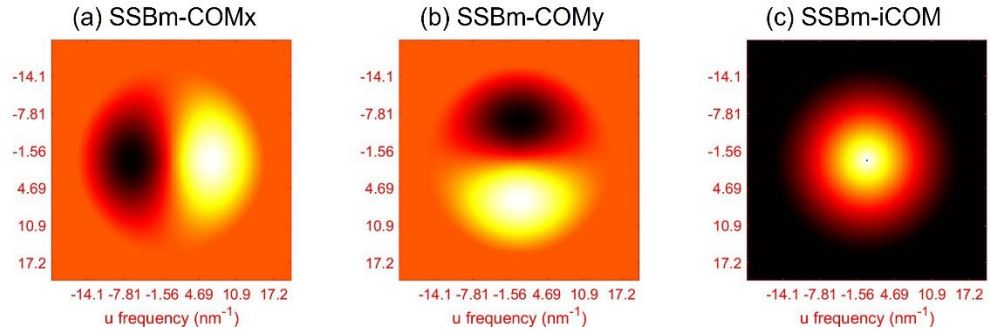

Figure 5-2 | 2D TF of (a) SSBm-COMx, (b) SSBm-COMy and (c) SSBm-iCOM. Acceleration voltage: 200 kV, convergence angle: 20 mrad, aberration free.

At the high frequency end, SSBm-iCOM has the same resolution ($\frac{\lambda}{2\alpha}$) as SSB. However, the low-frequency component (for example $|Q_p| < \frac{1}{2}\alpha$) differs, as SSBm-iCOM exhibits higher transfer near zero frequency. The transfer function at low frequencies is boosted due to the smaller denominator the at high frequency range in Equation (5-21). This is advantageous when low-frequency information is required, as shown in Section 5.5.2. However, in cases where the sample contains noise, the low-frequency noise may be amplified by transfer function, which will be discussed in Section 5.5.2.

It is worth noting that although the CTFs of SSBm-iCOM and SSB ptychography are different, they merely apply different modulations to the same original information $\psi_s(\boldsymbol{R_p})$. In an ideal noise-free condition, these can be corrected. Therefore, in principle, SSBm-iCOM does not provide more information than SSB ptychography; it simply displays the original signal with a different weight. This will be further discussed in Section 5.5.

5.4 Experiments on Defocused samples

To demonstrate aberration correction feature using SSBm-COM, an experiment was conducted using a JEOL ARM300F microscope with a monolayer graphene sample at 80 kV. 4DSTEM datasets at different defocus conditions were recorded using a MerlinEM detector from Quantum Detectors. The convergence angle was 24.8 mrad and the scanning step was 0.22 Å. The experimental fluence for each 4DSTEM dataset was 3.19×10^4 $e^-/Å^2$. 4DSTEM focal series were recorded at defoci of 0, -7.5 and -22 nm.

The focal series datasets were collected from the same scanning area and reconstructed separately using traditional COM/iCOM, SSBm-COM/iCOM and SSB (with aberration correction) algorithms. The focal series dataset was collected by moving the defocus by changing the current in the objective lens and recording the same sample area for each different defocus setting. Reconstructions are shown in Figure 5-3 (iCOM, SSBm-iCOM and SSB ptychography) and Figure 5-4 (COMx/y and SSBm-COMx/y). As described in the last section, experimental iCOM images can amplify low frequency noise. To avoid the amplified low-frequency covering lattice, in this experiment, a high-pass filter (hard threshold) was applied to traditional iCOM images in Figure 5-3 to reduce low frequency noise.

In Figure 5-3 and 5-4, with increasing defocus, at -7.5 and -22 nm, SSBm-COMx/y and SSBm-iCOM significantly improve the reconstructions compared to traditional COMx/y and iCOM. For traditional iCOM, in Figure 5-3, with a defocus of -22 nm, the reconstruction shows no lattice information, whereas at the same defocus SSBm-COM shows atomic resolution.

As shown in Figure 5-4, uncorrected COMx/y contains no useful information about the sample at a defocus of -22 nm. As a comparison, SSBm-COMx/y contains lattice information at the same defocus condition. The high-frequency noise in the in-focus COMx/y results are reduced in the SSBm-COMx/y results.

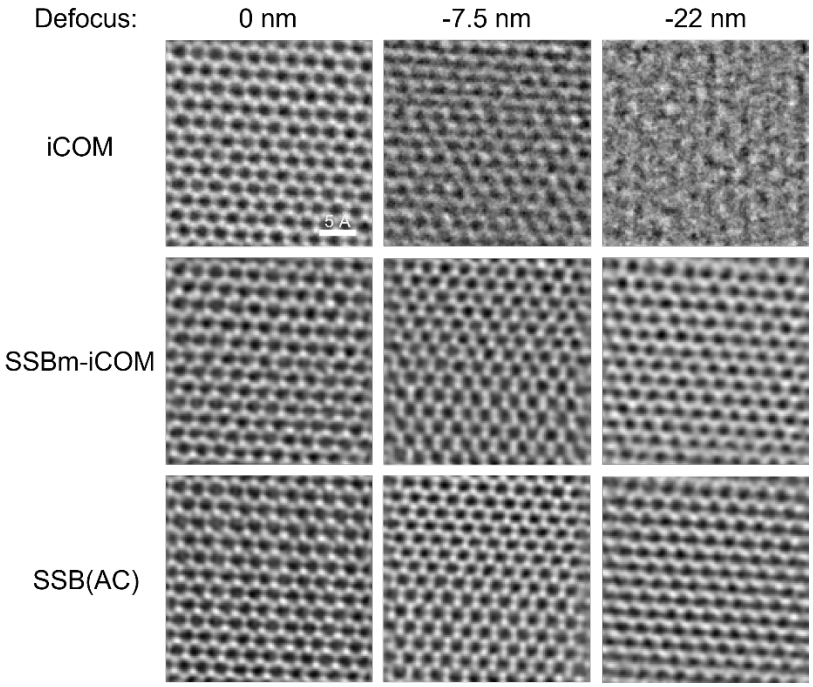

Figure 5-3 | Experimental reconstructions of iCOM, SSBm-iCOM and SSB for a monolayer graphene sample with defocus of 0, -7.5 and -22 nm. SSB with aberration correction and SSBm-iCOM compensates defocus in the reconstructions under defocus of -7.5 and -22 nm. Acceleration voltage: 80 kV, convergence angle: 24.8 mrad, pixel size: 0.22 Å.

This experiment demonstrated that the SSBm-COMx/y and SSBm-iCOM algorithms have capability to correct defocus. In Figure 5-3, the aberration-corrected SSBm-iCOM demonstrates a capability comparable to that of the aberration-corrected SSB. Theoretically, the same algorithm can also be used to correct other aberrations, for example astigmatism, coma and spherical aberration. For experimental datasets, correcting aberrations beyond defocus depends on accurate aberration measurements, which can be obtained from an aberration corrector or calculated from a 4D STEM dataset [18, 19]. The effectiveness of aberration correction is constrained by the precision of the aberration values provided to the SSBm-COM algorithm.

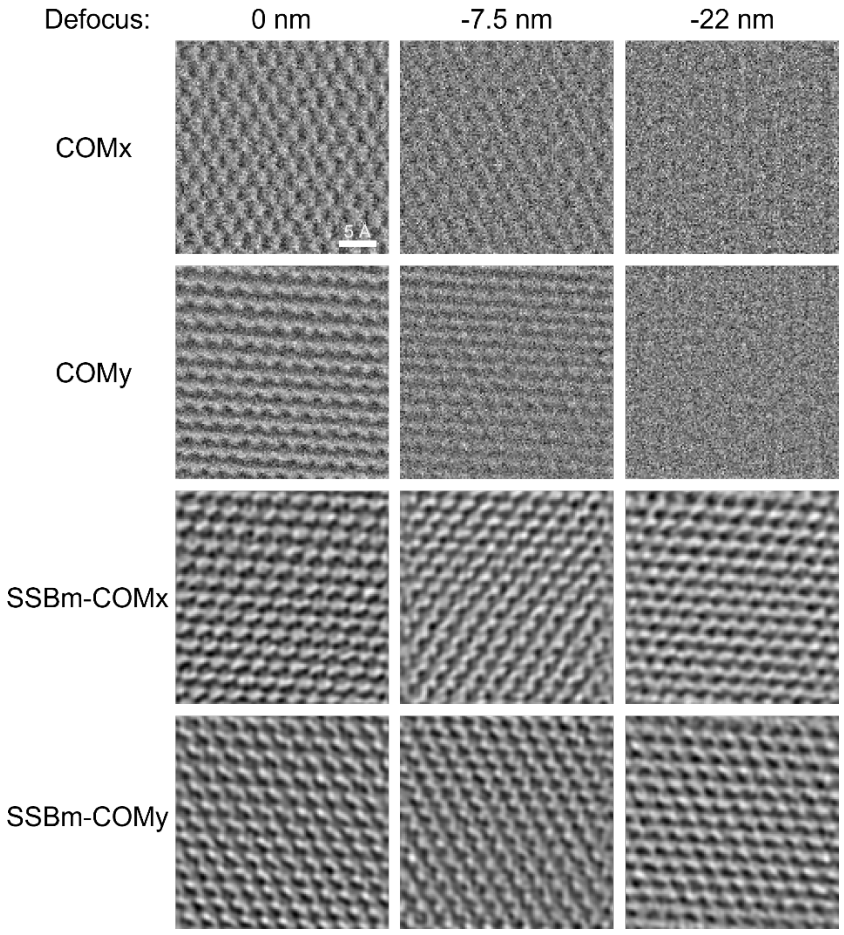

Figure 5-4 | Experimental reconstructions of COMx/y and SSBm-COMx/y for monolayer graphene. SSBm-COMx/y compensates defocus and reconstructs lattice information under defocus of -7.5 and -22 nm. Acceleration voltage: 80 kV, convergence angle: 24.8 mrad, pixel size: 0.22 Å. Defocus: 0, -7.5, -22 nm

This makes the SSBm-COMx/y and SSBm-iCOM algorithms more adaptable to defocus probe conditions than traditional COMx/y and iCOM. Experimentally, this also helps minimize the electron fluence on the sample during defocus adjustments prior to data acquisition, thereby reducing beam damage.

5.5 Simulations

5.5.1 SSBm-iCOM under Different Fluence Conditions

SSBm-COM/iCOM offers two advantages over traditional COM/iCOM images. In addition

to aberration correction, the SSB mask also filters out data that introduces noise without providing useful information, *i.e*. data points outside the DO area. Theoretically, if the sample can be regarded as a weak phase object and an aberration-free probe is assumed, all phase-contrast information is included in the SSB mask as shown in Equation (5-8) and reference [20]. As a result, in SSBm-COM, filtering the area outside the SSB mask improves the signal-to-noise ratio of data used for reconstruction. As $\boldsymbol{Q_p}$ increases from a frequency corresponding to α, the SSB mask area decreases and eventually reaches zero at a frequency corresponding to 2α, inherently suppressing high-frequency noise.

To demonstrate this denoising effect, Figure 5-5 shows a series of in-focus simulations at different electron fluence for iCOM and SSBm-iCOM on the same sample (monolayer graphene with a silicon atom replacement). Compared to traditional iCOM, SSBm-iCOM reconstructs smoother images, especially at fluences ranging from 5,000 to 20,000 $e^-/Å^2$, due to the intrinsic filtering effect.

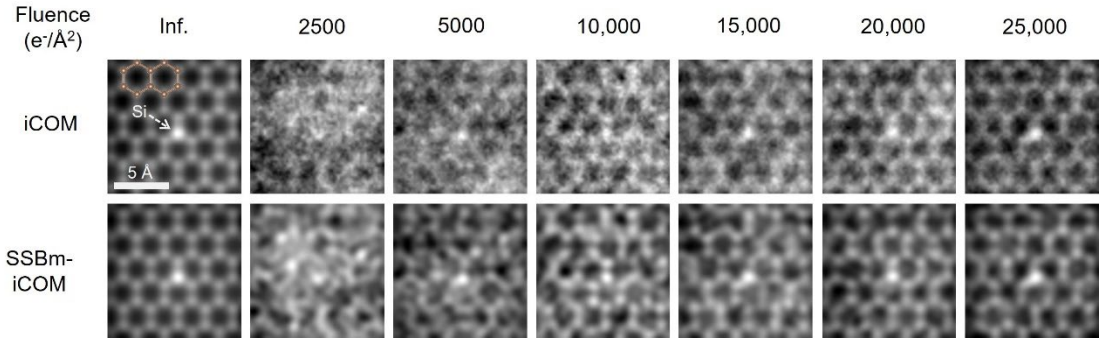

Figure 5-5 | Reconstructions of simulated silicon-replaced monolayer graphene at different fluence conditions, including 2.5k, 5k, 10k, 15k, 20k, 25k and infinite fluence. SSBm-iCOM shows a suppression of high-frequency noise under low-fluence condition compared to iCOM. Voltage: 80 kV. Convergence angle: 22.5 mrad. Pixel size: 0.2 Å.

5.5.2 Low-frequency transfer

In addition to aberration correction and suppression of noise, as discussed in Section 5.3, SSBm-iCOM reconstructs the same sample phase information with WPOA as SSB ptychography. Their contrast transfer functions, as shown in Figure 5-2, share similarities,

such as tending to 0 at the frequency corresponding to 2α. However, at low frequencies, the transfer function for SSBm-iCOM, as for the traditional iCOM method, is significantly higher than SSB. The $\boldsymbol{Q_p^2}$ term in the denominator of Equation (5-21) enhances low-frequency transfer (excluding zero frequency, which is lost). 1-dimension transfer functions of SSBm-iCOM and SSB ptychography are plotted in Figure 5-6. This results in different reconstruction outcomes from SSBm-iCOM and SSB when the sample contains low-frequency information, even though this low-frequency information may not always be of interest.

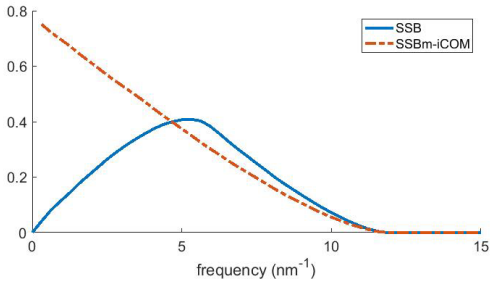

Figure 5-6 |Transfer functions for SSB and SSBm-iCOM. SSBm-iCOM exhibits amplified low-frequency information transfer (below 5 nm^{-1}), whereas information within the same frequency range is attenuated by SSB. Calculations are at 80 kV. Convergence angle is 25 mrad.

To demonstrate the different outcome of low-frequency transfer between SSBm-iCOM and SSB ptychography, a simulation of a perfect (defect-free) monolayer MoS_2 crystal on an amorphous carbon film was performed. The amorphous carbon film was generated by randomly distributing carbon atoms within a defined volume, ensuring consistency with the density of amorphous carbon (2.1 g/cm^3 [21]). This model does not account for the radial distribution function of amorphous carbon since the short-range order is not considered. But for the purpose of this simulation (introduce low-frequency noise on the sample), the effects caused by the radial distribution function can be ignored. The thickness of the carbon film was adjusted based on the thickness map shown in Figure 5-7 (a) to introduce low-frequency noise. This non-uniform carbon film was then combined with the monolayer MoS_2 as depicted in Figure 5-7 (b).

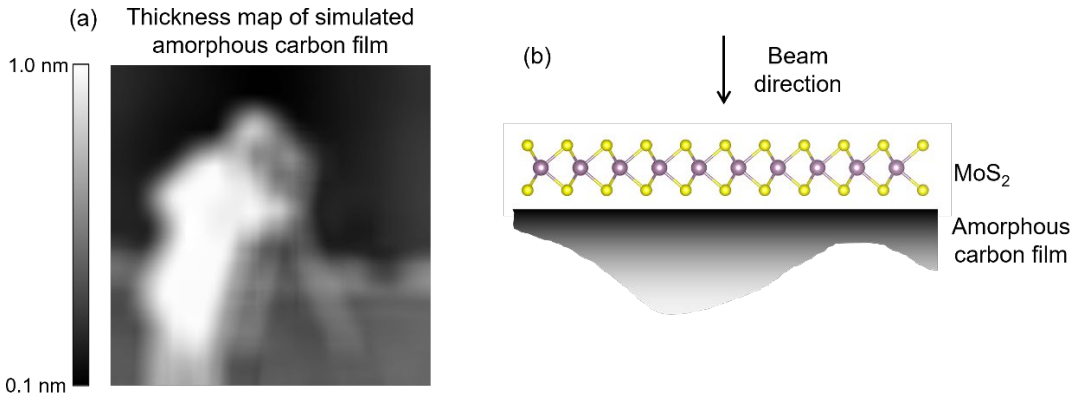

Figure 5-7 | Sample used in the simulation. (a) thickness map of the non-uniform amorphous carbon film. (b) the schematic side view of the sample. The electron beam first goes through monolayer MoS2 then the non-uniform amorphous carbon film.

The simulation code is based on multislice simulation [22, 23] and modified MULTEM package [24]. Simulation parameters are listed in Table 5-1. Reconstructions are shown in Figure 5-8. Traditional iCOM, SSBm-iCOM and SSB ptychography are used for reconstructions.

Table 5-1 | Simulation parameters

Sample	perfect monolayer MoS_2 (defect-free)
Acceleration voltage	80 kV
Convergence angle	25 mrad
Scanning step	0.18 Å
Size of detector	64×64 pixels
Angular sampling on detector	1.2 mrad/pixel
Defocus	0 and 20 nm
Fluence	30864 and 6173 $e^-/Å^2$

In Figure 5-8, with an in-focus probe, traditional iCOM, SSBm-iCOM and SSB all provide good contrast for the MoS_2 lattice. However, in the SSB result, the carbon film background is filtered out due to the suppression of low frequency information. For the traditional iCOM and SSBm-iCOM results, the amorphous carbon background appears faint but is still recognisable, duo to its boosted low-frequency transfer as shown in Figure 5-6. At 20 nm defocus, the aberration corrected methods, *i.e*. SSBm-iCOM and SSB, still successfully reconstruct the MoS_2 lattice. In contrast, traditional iCOM fails to reconstruct the lattice image.

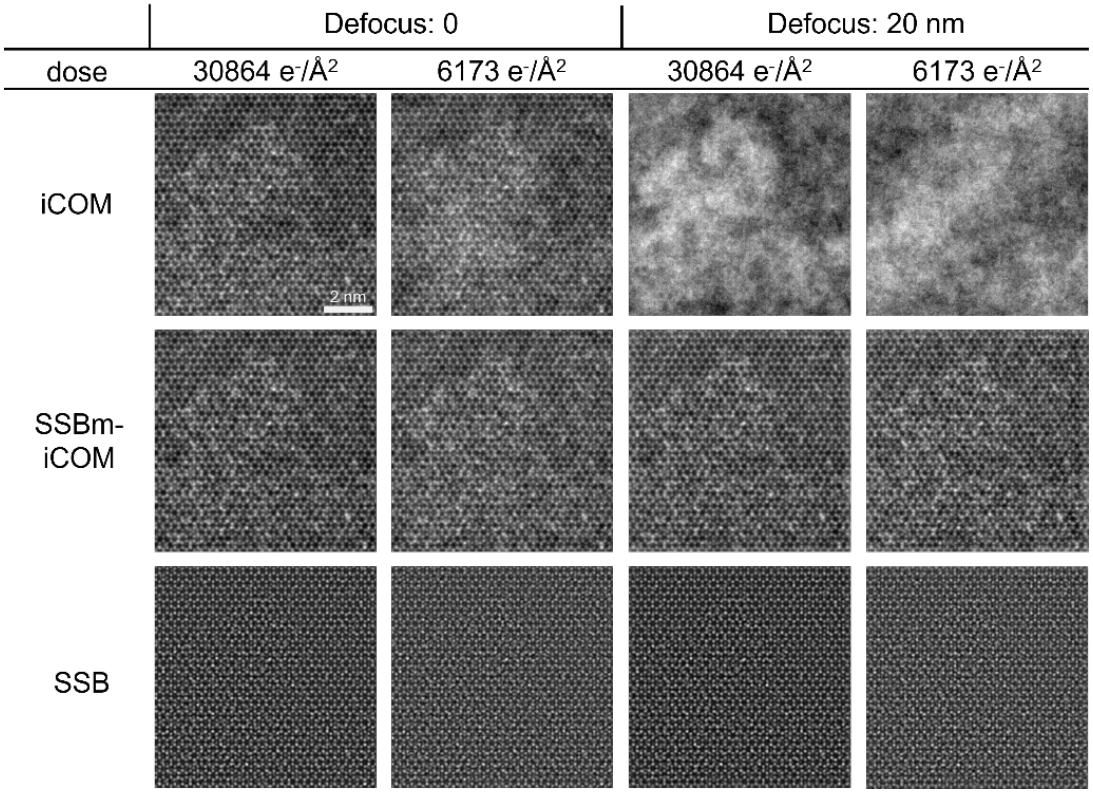

Figure 5-8 | Simulated reconstructions for MoS_2 on an amorphous carbon film. The thickness of carbon film is modified as shown in Figure 5-7. In both iCOM and SSBm-iCOM, the distribution of amorphous carbon can be recognised, whereas in SSB, it is filtered out by the transfer function as low-frequency noise. Under a defocus of 20 nm, SSB and SSBm-iCOM reconstruct lattice information, which is absent in iCOM. Acceleration voltage: 80 kV, convergence angle: 25 mrad, pixel size: 0.18 Å.

For 4D STEM imaging, low-frequency sensitivity can be advantageous in specific situations. In Figure 5-8, without prior knowledge of the carbon film, it is challenging to assess the reliability of SSB reconstructions. When low-frequency information, which includes details about the carbon film, is partially filtered by SSB ptychography, some remaining noise may be misleading and could be mistaken for lattice defects, as shown in Figure 5-9. The in-focus reconstructions at 6173 $e^-/Å^2$ are used as examples. Blue circles in Figure 5-9 indicate positions that might be interpreted as “additional atoms in the lattice”.

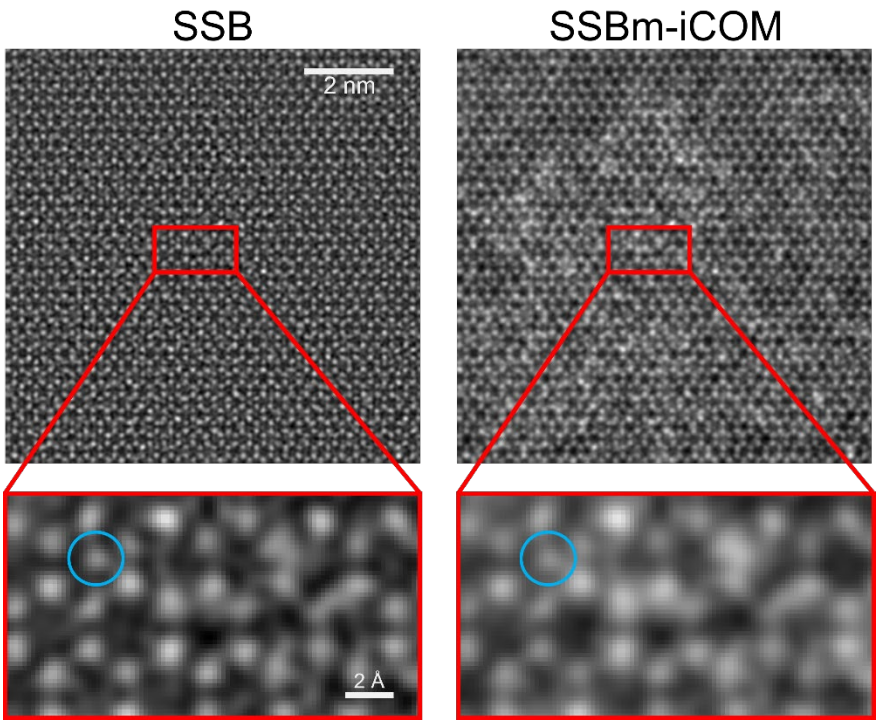

Figure 5-9 | Subareas in SSB and SSBm-iCOM reconstruction from the simulation of MoS_2 on an amorphous carbon film. Two subareas are extracted from the in-focus reconstructions at 6173 $e^-/Å^2$ in Figure 5-8. Blue circles indicate misleading positions. Acceleration voltage: 80 kV, convergence angle: 25 mrad, pixel size: 0.18 Å.

In Figure 5-9, blue circles indicate misleading positions where a false “additional atom” may be interpretated on both SSB and SSBm-iCOM reconstruction. But large field-of-view SSBm-iCOM reconstructions provide additional low-frequency information about the possible presence of non-uniform carbon film or contamination, which is absent in the SSB reconstruction. Therefore, if the sample contrast is affected by substrates or unknown

material with low-frequency information, such as non-uniform carbon films or contamination, the low-frequency information provided by SSBm-iCOM can be essential for assessing the reliability of the information transferred.

However, it must be emphasized that the enhancement of low-frequency information transfer in SSBm-iCOM does not mean that more information can be transmitted; it merely increases the weight of low-frequency information in the reconstruction.

SSB ptychography can also achieve the same low-frequency transfer by redistributing or compensating for the intensity of different frequencies in the reconstruction. To achieve this, $\Psi_{SSB}(\boldsymbol{Q_p})$, the Fourier transform of SSB reconstruction ψ_{SSB} described in Chapter 3, can be processed as shown in Equation (5-22):

$$\Psi'_{SSB}(\boldsymbol{Q_p}) = \Psi_{SSB}(\boldsymbol{Q_p}) \cdot \frac{TF^{f}_{iCOM}(\boldsymbol{Q_p})}{TF_{SSB}(\boldsymbol{Q_p})} \qquad (5\text{-}22)$$

$$where\ |\boldsymbol{Q_p}| > 0$$

In Equation (5-22), $\Psi'_{SSB}(\boldsymbol{Q_p})$ represents the processed $\Psi_{SSB}(\boldsymbol{Q_p})$,processed using a frequency weight redistribution factor $\frac{TF^{f}_{iCOM}(\boldsymbol{Q_p})}{TF_{SSB}(\boldsymbol{Q_p})}$. The numerator of the factor is from Equation (5-21), while the denominator corresponds to the TF of SSB, as described in Chapter 3. The processed SSB reconstruction can be obtained by applying an inverse Fourier transform to $\Psi'_{SSB}(\boldsymbol{Q_p})$. A similar approach to CTF normalisation for ptychography is also reported to adjust the reconstruction based on the noise distribution [12].

To demonstrate this, the SSB simulation results for the same sample and simulation conditions as listed in Table 5-1 (except for fluence, where an infinite fluence is used) are processed according to Equation (5-22). The original and processed SSB phase reconstructions are presented in Figure 5-10.

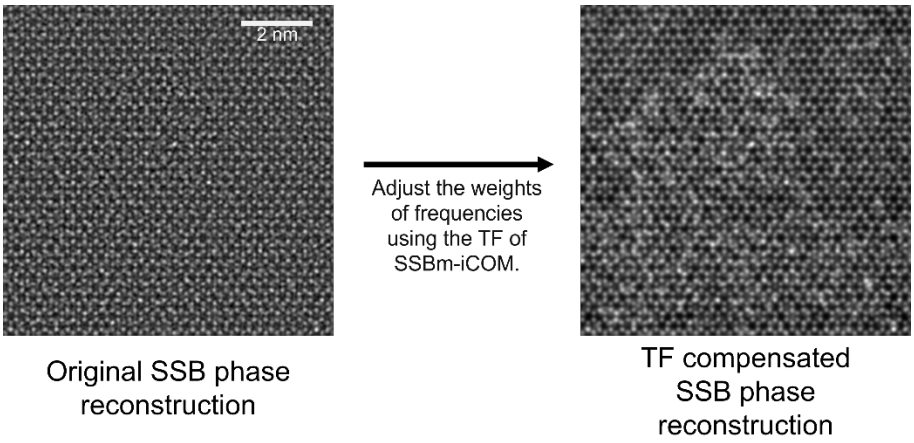

Figure 5-10 | Simulated TF compensated SSB phase reconstruction. Sample and conditions are the same as used in Table 5-1 except an infinite fluence is used here. The original SSB phase reconstruction is processed as described in Equation (5-22). The TF compensated SSB phase reconstruction shows low-frequency information.

In Figure 5-10, the processed phase reconstruction (right side) exhibits more low-frequency information (corresponding to the thickness map of the carbon film in Figure 5-7) compared to the original reconstruction (left side). This is due to the low-frequency information being enhanced by the factor $\frac{TF^{f}_{iCOM}(\boldsymbol{Q_p})}{TF_{SSB}(\boldsymbol{Q_p})}$ in Equation (5-22). This demonstrates that low-frequency information is not lost in SSB, but rather suppressed by the SSB transfer function. In theory, the SSB reconstruction contains the same information as SSBm-iCOM. The differences observed in Figure 5-8 between the SSB and SSBm-iCOM reconstructions arise from their distinct transfer functions, rather than from differences in the reconstructed information.

In conclusion, fundamentally, the reconstruction results of SSBm-iCOM and SSB ptychography do not differ in terms of information. However, SSBm-iCOM, compared to SSB ptychography, provides intrinsic low-frequency enhancement without requiring additional post-processing, such as the method described in Equation (5-22). This feature is advantageous for samples with low-frequency details but also increases sensitivity to contamination.

5.6 Conclusion

In this chapter, a phase reconstruction method Single-side Band masked Centre-of-mass, (SSBm-COM), which combines SSB ptychography and Centre-of-mass (COM) imaging, is discussed. In its derivation, the weak phase object approximation (WPOA) is assumed. SSBm-COM can compensate aberrations in 4DSTEM dataset as in SSB ptychography. This makes it not limited to an in-focus condition as is required for traditional COM methods. For sensitive samples this can be an advantage since total electron fluence at the sample can be reduced by reducing the time for fine-tuning the defocus before data collection.

An experimental dataset is shown in this chapter to illustrate that SSBm-iCOM can reconstruct an atomic resolution lattice with residual defocus. In addition, because of the SSB-like masking process in the SSBm-iCOM method, high frequency noise is significantly reduced compared to traditional COM methods.

Furthermore, simulations are used to illustrate the frequency contrast transfer functions for SSB ptychography and SSBm-iCOM. Theoretically, SSBm-iCOM shows a low-frequency enhancement compared to SSB ptychography, although overall they transfer the same information. In the simulations, a carbon film with a low-frequency thickness variation as a substrate is shown to misleading results for both SSB ptychography and SSBm-iCOM. SSBm-iCOM provides additional low-frequency information about the sample (SSB with a modified CTF can achieve similar reconstructions). In conclusion, SSBm-COM methods (SSBm-COMx/y and SSBm-iCOM) have the advantage of aberration-correction and high-frequency noise filtering as for SSB ptychography and low-frequency transfer as for traditional COM.

SSBm-COM methods still have limitations. Firstly, they rely on a 4DSTEM dataset recorded using a pixelated detector, similar to that used for ptychography. Pixelated detectors can be expensive and are often slower than multi-channel BF/DF detectors. Differential phase contrast (DPC), as mentioned in the introduction, approximates COM imaging using segmented multi-channel detectors. Further developments on SSBm-COM related

methods to adapt to arbitrary detectors will be discussed in the following chapters. Secondly, traditional COM only constrains the sample as a phase object, but SSB and SSBm-COM requires sample to meet the weak phase object approximation. This approximation theoretically limits the applications of these methods. Therefore, new reconstruction methods which do not require the WPOA will be discussed in Chapter 7. Moreover, SSBm-iCOM, like ptychography, assumes a coherent probe. This assumption imposes additional experimental constraints and instrument requirements to ensure high probe coherence.

Reference

[1] H. Rose, "Phase Contrast in Scanning Transmission Electron Microscopy," *Optik,* vol. 39, no. 4, pp. 416-436, 1974.

[2] N. H. Dekkers and H. de Lang, "Differential Phase Contrast in a STEM," *Optik,* vol. 41, no. 4, pp. 452-456, 1974.

[3] E. M. Waddell, J. N. Chapman, and R. P. Ferrier, "Linear Imaging of Strong Phase Objects by Differential Phase Contrast," in *Inst Phys Conf Ser*, 1977, 36 ed., pp. 267-270.

[4] I. Lazić, E. G. T. Bosch, and S. Lazar, "Phase contrast STEM for thin samples: Integrated differential phase contrast," *Ultramicroscopy,* vol. 160, pp. 265-280, 2016.

[5] N. Shibata, S. D. Findlay, Y. Kohno, H. Sawada, Y. Kondo, and Y. Ikuhara, "Differential phase-contrast microscopy at atomic resolution," *Nature Physics,* vol. 8, no. 8, pp. 611-615, 2012.

[6] E. M. Waddell and J. N. Chapman, "Linear Imaging of Strong Phase Objects Using Asymmetrical Detectors in STEM," *Optik,* vol. 54, no. 2, pp. 83-96, 1979.

[7] C. Yu, T. Friedrich, D. Jannis, S. Van Aert, and J. Verbeeck, "Real-time integration center of mass (riCOM) reconstruction for 4D STEM," *Microscopy and Microanalysis,* vol. 28, no. 5, pp. 1526-1537, 2022.

[8] T. Grieb *et al.*, "4D-STEM at interfaces to GaN: Centre-of-mass approach & NBED-disc detection," *Ultramicroscopy,* vol. 228, p. 113321, Sep 2021.

[9] M. J. Zachman, Z. Yang, Y. Du, and M. Chi, "Robust Atomic-Resolution Imaging of Lithium in Battery Materials by Center-of-Mass Scanning Transmission Electron Microscopy," *ACS Nano,* vol. 16, no. 1, pp. 1358-1367, 2022.

[10] K. Müller *et al.*, "Atomic electric fields revealed by a quantum mechanical approach to electron picodiffraction," *Nature communications,* vol. 5, no. 1, p. 5653, 2014.

[11] N. Shibata *et al.*, "Electric field imaging of single atoms," *Nature communications,* vol. 8, no. 1, p. 15631, 2017.

[12] C. M. O'Leary, G. T. Martinez, E. Liberti, M. J. Humphry, A. I. Kirkland, and P. D. Nellist, "Contrast transfer and noise considerations in focused-probe electron ptychography," *Ultramicroscopy,* vol. 221, p. 113189, 2021.

[13] J. M. Rodenburg, B. C. McCallum, and P. D. Nellist, "Experimental tests on double-resolution coherent imaging via STEM," *Ultramicroscopy,* vol. 48, no. 3, pp. 304-314, 1993.

[14] H. Yang *et al.*, "Electron ptychographic phase imaging of light elements in crystalline materials using Wigner distribution deconvolution," *Ultramicroscopy,* vol. 180, pp. 173-179, 2017.

[15] J. M. Rodenburg and R. H. T. Bates, "The theory of super-resolution electron microscopy via Wigner-distribution deconvolution," *Philosophical Transactions of the Royal Society of London. Series A: Physical and Engineering Sciences,* vol. 339, no. 1655, pp. 521-553, 1992.

[16] P. D. Nellist and J. M. Rodenburg, "Beyond the conventional information limit: the relevant coherence function," *Ultramicroscopy,* vol. 54, no. 1, pp. 61-74, 1994.

[17] H. Yang, P. Ercius, P. D. Nellist, and C. Ophus, "Enhanced phase contrast transfer using ptychography combined with a pre-specimen phase plate in a scanning transmission electron microscope," *Ultramicroscopy,* vol. 171, pp. 117-125, 2016.

[18] H. Yang *et al.*, "Simultaneous atomic-resolution electron ptychography and Z-contrast imaging of light and heavy elements in complex nanostructures," *Nature Communications,* vol. 7, no. 1, p. 12532, 2016.

[19] A. R. Lupini, P. Wang, P. D. Nellist, A. I. Kirkland, and S. J. Pennycook, "Aberration measurement using the Ronchigram contrast transfer function," *Ultramicroscopy,* vol. 110, no. 7, pp. 891-898, 2010.

[20] T. J. Pennycook, A. R. Lupini, H. Yang, M. F. Murfitt, L. Jones, and P. D. Nellist, "Efficient phase contrast imaging in STEM using a pixelated detector. Part 1: Experimental demonstration at atomic resolution," *Ultramicroscopy,* vol. 151, pp. 160-167, 2015.

[21] M. Iwaki, "Estimation of the atomic density of amorphous carbon using ion implantation, SIMS and RBS," *Surface and Coatings Technology,* vol. 158, pp. 377-381, 2002.

[22] P. Goodman and A. F. Moodie, "Numerical evaluations of N-beam wave functions in electron scattering by the multi-slice method," *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography,* vol. 30, no.

2, pp. 280-290, 1974.

[23] J. M. Cowley and A. F. Moodie, "The scattering of electrons by atoms and crystals. I. A new theoretical approach," *Acta Crystallographica,* vol. 10, no. 10, pp. 609-619, 1957.

[24] I. Lobato and D. Van Dyck, "MULTEM: A new multislice program to perform accurate and fast electron diffraction and imaging simulations using Graphics Processing Units with CUDA," *Ultramicroscopy,* vol. 156, pp. 9-17, 2015.

Chapter 6: Ptychography in the Weak Phase Object Approximation using an Arbitrary Detector

6.1 Introduction

We have discussed SSBm-COM methods in the previous chapter, which can be regarded as a hybrid method of phase-contrast ptychography and centre-of-mass imaging. These have the advantage of aberration correction taken from ptychography and low-frequency sensitivity from COM. However, they both require a pixelated detector to record the 4DSTEM dataset. A potential alternative is using segmented detectors and the Differential Phase Contrast (DPC) method [1, 2] as an approximation to the centre-of-mass method. Another advantage of segmented detectors (or multi-channel ADF detectors) is that they are usually faster than pixelated detectors since they have fewer output channels. Faster recording speed leads to a shorter beam time on the sample and this can be a useful feature for controlling beam damage on sensitive samples.

Naturally, combining ptychography methods with segmented detectors [3-5] is an ideal choice to balance recording speed and experimental flexibility (such as defocus imaging). On the basis of those theories and experiments, in this chapter, a ptychography method named AD-ptycho is developed, which can be implemented on any multi-channel detector, including segmented detectors. AD-ptycho introduces a detector function $D(\boldsymbol{K_f})$ to

describe the shape of detector channels or the detector layout. In theory, AD-ptycho can compensate for aberration in the dataset as it is a ptychographic method. However, the ability to compensate for aberrations depends on the shape of the detector function and the number of channels. With all these features, AD-ptycho is a more general framework of ptychography with weak phase objects on arbitrary detectors.

6.2 The AD-ptycho Method

Following the weak phase object approximation (WPOA) as used in the previous chapters (see Section 5.2), the function $\mathrm{G}(\boldsymbol{K_f}, \boldsymbol{Q_p})$ which represents the Fourier transform with respect to the scanning coordinate ($\boldsymbol{R_p}$) of a 4DSTEM dataset $\left|M(\boldsymbol{K_f}, \boldsymbol{R_p})\right|^2$, can be expressed as [6]:

$$\begin{aligned}\mathrm{G}(\boldsymbol{K_f}, \boldsymbol{Q_p}) &= \mathcal{F}_{Rp\to Q_p}\left\{\left|M(\boldsymbol{K_f}, \boldsymbol{R_p})\right|^2\right\} \\ &= \left|A(\boldsymbol{K_f})\right|^2\delta(\boldsymbol{Q_p}) + A(\boldsymbol{K_f})A^*(\boldsymbol{K_f}+\boldsymbol{Q_p})\Psi_s^*(-\boldsymbol{Q_p}) \\ &+ A^*(\boldsymbol{K_f})A(\boldsymbol{K_f}-\boldsymbol{Q_p})\Psi_s(\boldsymbol{Q_p})\end{aligned} \tag{6-1}$$

Furthermore, we consider a detector function $D_n(\boldsymbol{K_f})$ in the diffraction plane. This function represents the area of the nth detector channel. The shape of $D_n(\boldsymbol{K_f})$ can be arbitrary in theory, including pixelated detectors. Experimentally, the shape of $D_n(\boldsymbol{K_f})$ is related to the actual detector layout. As an example, multiple $D_n(\boldsymbol{K_f})$, $D_1(\boldsymbol{K_f}), D_2(\boldsymbol{K_f}), \cdots D_N(\boldsymbol{K_f})$, can also form a set of concentric segmented detectors. The value of $D_n(\boldsymbol{K_f})$ represents the response of the detector channel to a single electron hitting the detector area. For simplicity, a detector channel in a multi-channel detector array is typically assumed to have $D_n(\boldsymbol{K_f})$ values of 1 (inside the detection area) or 0 (outside the detection area).

For a detector function $D_n(\boldsymbol{K_f})$, the signal $C_n(\boldsymbol{R_p})$ detected by corresponding detector channel n is:

$$C_n(\boldsymbol{R_p}) = \iint D_n(\boldsymbol{K_f}) \cdot \left|M(\boldsymbol{K_f}, \boldsymbol{R_p})\right|^2 \cdot d\boldsymbol{K_f} \tag{6-2}$$

In Equation (6-2), $\boldsymbol{R_p}$ is independent of the integral variable $\boldsymbol{K_f}$. Therefore, the Fourier transform of Equation (6-2) with respect to $\boldsymbol{R_p}$ can be written as:

$$c_n(\boldsymbol{Q_p}) = \mathcal{F}_{\boldsymbol{R_p}\to\boldsymbol{Q_p}}\{C_n(\boldsymbol{R_p})\} = \iint D_n(\boldsymbol{K_f}) \cdot \mathrm{G}(\boldsymbol{K_f}, \boldsymbol{Q_p}) \cdot d\boldsymbol{K_f} \tag{6-3}$$

In Equation (6-3), $c_n(\boldsymbol{Q_p})$ is very similar to the $com_x(\boldsymbol{Q_p})$ and $com_y(\boldsymbol{Q_p})$ in Equation (5-4), discussed in Chapter 5. If the diffraction coordinate term K_{fx} or K_{fy} is regarded as an extension of $D_n(\boldsymbol{K_f})$, then COM imaging can also be incorporated into the equation. This implies that the centre-of-mass can be regarded as a special case within the AD-ptycho framework, which is discussed in Section 6.7.

For the three terms of $\mathrm{G}(\boldsymbol{K_f}, \boldsymbol{Q_p})$ under the WPOA, as described in Equation (6-1), $c_n(\boldsymbol{Q_p})$ can be separated into three integrals:

$$\begin{aligned} c_n(\boldsymbol{Q_p}) = &\iint D_n(\boldsymbol{K_f}) \cdot \left|A(\boldsymbol{K_f})\right|^2 \delta(\boldsymbol{Q_p}) \cdot d\boldsymbol{K_f} \\ &+ \iint D_n(\boldsymbol{K_f}) \cdot A(\boldsymbol{K_f}) A^*(\boldsymbol{K_f} + \boldsymbol{Q_p}) \Psi_s^*(-\boldsymbol{Q_p}) \cdot d\boldsymbol{K_f} \\ &+ \iint D_n(\boldsymbol{K_f}) \cdot A^*(\boldsymbol{K_f}) A(\boldsymbol{K_f} - \boldsymbol{Q_p}) \Psi_s(\boldsymbol{Q_p}) \cdot d\boldsymbol{K_f} \end{aligned} \tag{6-4}$$

In Equation (6-4), the first term contains a Dirac function $\delta(\boldsymbol{Q_p})$, which is ignored when $\boldsymbol{Q_p} \neq \boldsymbol{0}$. Consequently, Equation (6-4) can be simplified by moving the Ψ_s terms out of the integrals, as they are independent of $\boldsymbol{K_f}$ to give:

$$\begin{aligned} c_n(\boldsymbol{Q_p}) = &\Psi_s^*(-\boldsymbol{Q_p}) \iint D_n(\boldsymbol{K_f}) \cdot A(\boldsymbol{K_f}) A^*(\boldsymbol{K_f} + \boldsymbol{Q_p}) \cdot d\boldsymbol{K_f} \\ &+ \Psi_s(\boldsymbol{Q_p}) \iint D_n(\boldsymbol{K_f}) \cdot A^*(\boldsymbol{K_f}) A(\boldsymbol{K_f} - \boldsymbol{Q_p}) \cdot d\boldsymbol{K_f} \end{aligned} \tag{6-5}$$

In Equation (6-5), the integral terms can be replaced by $T_{Pos,n}(\boldsymbol{Q_p})$ and $T_{Neg,n}(\boldsymbol{Q_p})$, which are defined in Equation (6-6).

$$\begin{cases} T_{Neg,n}(\boldsymbol{Q_p}) = \iint D_n(\boldsymbol{K_f}) \cdot A(\boldsymbol{K_f}) A^*(\boldsymbol{K_f} + \boldsymbol{Q_p}) \cdot d\boldsymbol{K_f} \\ T_{Pos,n}(\boldsymbol{Q_p}) = \iint D_n(\boldsymbol{K_f}) \cdot A^*(\boldsymbol{K_f}) A(\boldsymbol{K_f} - \boldsymbol{Q_p}) \cdot d\boldsymbol{K_f} \end{cases} \tag{6-6}$$

After this replacement, $c_n(\boldsymbol{Q_p})$ for detector channel n can be written in a more concise form as:

$$c_n(\boldsymbol{Q_p}) = \Psi_s^*(-\boldsymbol{Q_p}) T_{Neg,n}(\boldsymbol{Q_p}) + \Psi_s(\boldsymbol{Q_p}) T_{Pos,n}(\boldsymbol{Q_p}) \tag{6-7}$$

Equation (6-7) is obtained under the WPOA. As discussed in Chapter 3, under the WPOA conditions, Ψ_s exhibits conjugate antisymmetry, as shown in Equation (6-8):

$$\Psi_s(\boldsymbol{Q_p}) = -\Psi_s^*(-\boldsymbol{Q_p}) \tag{6-8}$$

Considering the conjugate antisymmetry of Ψ_s, Equation (6-7) can be simplified to [7, 8]:

$$c_n(\boldsymbol{Q_p}) = \Psi_s(\boldsymbol{Q_p}) \cdot [T_{Pos,n}(\boldsymbol{Q_p}) - T_{Neg,n}(\boldsymbol{Q_p})] = \Psi_s(\boldsymbol{Q_p}) \cdot \Gamma_{T,n}(\boldsymbol{Q_p})$$

Where $\Gamma_{T,n}$ is defined as: (6-9)

$$\Gamma_{T,n}(\boldsymbol{Q_p}) = T_{Pos,n}(\boldsymbol{Q_p}) - T_{Neg,n}(\boldsymbol{Q_p})$$

Given the presence of multiple detector channels (channel $1,2,\cdots N$), whether they are concentric sectors or part of a pixelated array, $c_1(\boldsymbol{Q_p})$ to $c_N(\boldsymbol{Q_p})$ can now be expressed as a series of equations:

$$\begin{cases} c_1(\boldsymbol{Q_p}) = \Psi_s(\boldsymbol{Q_p}) \cdot \Gamma_{T,1}(\boldsymbol{Q_p}) \\ \vdots \\ c_N(\boldsymbol{Q_p}) = \Psi_s(\boldsymbol{Q_p}) \cdot \Gamma_{T,N}(\boldsymbol{Q_p}) \end{cases} \tag{6-10}$$

In Equation (6-10), the number of equations in the series is determined by the number of detector channels (or the number of pixels for pixelated detectors). However, in Equation (6-10), both $\Psi_s(\boldsymbol{Q_p})$ and $\Gamma_{T,n}(\boldsymbol{Q_p})$ are complex functions, and their multiplication results in cross terms between the real and imaginary parts. To avoid cross terms while compensating the aberrations in $\Gamma_{T,n}(\boldsymbol{Q_p})$, both sides of all equations are multiplied by the

conjugate of $\Gamma_{T,n}(\boldsymbol{Q_p})$. This transforms one side of the equation into the form of $\Psi_s(\boldsymbol{Q_p})$ multiplied by a real function ($\left|\Gamma_{T,n}(\boldsymbol{Q_p})\right|^2$), thus avoiding cross terms. Consequently, in the following calculation, each equation can be simply decomposed into operations on the real and imaginary parts separately.

$$\begin{cases} c_1(\boldsymbol{Q_p}) \cdot \Gamma_{T,1}(\boldsymbol{Q_p})^* = \Psi_s(\boldsymbol{Q_p}) \cdot \left|\Gamma_{T,n}(\boldsymbol{Q_p})\right|^2 \\ \vdots \\ c_N(\boldsymbol{Q_p}) \cdot \Gamma_{T,N}(\boldsymbol{Q_p})^* = \Psi_s(\boldsymbol{Q_p}) \cdot \left|\Gamma_{T,n}(\boldsymbol{Q_p})\right|^2 \end{cases} \tag{6-11}$$

Furthermore, Equation (6-11) can be represented as a linear system:

$$\underbrace{\begin{bmatrix} \left|\Gamma_{T,n}(\boldsymbol{Q_p})\right|^2 \\ \vdots \\ \left|\Gamma_{T,n}(\boldsymbol{Q_p})\right|^2 \end{bmatrix}}_{Matrix:\, T_m\ (N\times 1)} \times \underbrace{\left[\Psi_s(\boldsymbol{Q_p})\right]}_{Vector:\, \Psi_{sv}\ (1\times 1)} = \underbrace{\begin{bmatrix} c_1(\boldsymbol{Q_p})\Gamma_{T,1}(\boldsymbol{Q_p})^* \\ \vdots \\ c_N(\boldsymbol{Q_p})\Gamma_{T,N}(\boldsymbol{Q_p})^* \end{bmatrix}}_{Vector:\, c_v\ (N\times 1)} \tag{6-12}$$

or:

$$T_m \Psi_{sv} = c_v$$

In equation (6-12), since there is only one unknown, $\Psi_s(\boldsymbol{Q_p})$, this system of equations appears to be redundant. However, in the case where D_n is rotationally symmetric, it is possible for $T_{Pos,n}$ and $T_{Neg,n}$ to exactly cancel and thus $\Gamma_{T,n}(\boldsymbol{Q_p}) = 0$. The simplest example is to assume that D_n is a detector occupying the entire $\boldsymbol{K_f}$ plane, in which $\Gamma_{T,n}(\boldsymbol{Q_p})$ will equal zero for all $\boldsymbol{Q_p}$. In this situation, the above equations become unsolvable. In Section 6.3, the solvable conditions are described.

When $\Psi_s(\boldsymbol{Q_p})$ is solvable, any method developed for solving overdetermined linear systems can be applied, such as the least squares method or singular value decomposition. Since the linear system is overdetermined, the solution is influenced by different normalisation and error estimation methods. For example, an iterative method for optimisation can be used [5]. For simplicity, the least squares method (LSM), which is the simplest approach, is chosen here.

The solution using LSM is:

$$\Psi_{sv} = (T_m^T T_m)^{-1} T_m^T c_v \tag{6-13}$$

In theory, the solved Ψ_{sv}, or $\Psi_s(\boldsymbol{Q_p})$, can be then inverse Fourier transformed to form an image of the transmission function $\psi_s(\boldsymbol{R_p})$. However, $\Psi_s(\boldsymbol{Q_p})$ solved using Equation (6-12) and (6-13) is not modulated in frequency, unlike traditional SSB which weights different spatial frequencies based on the area of the DO (double overlap) region, as discussed in Chapter 3. This can lead to excessive amplification of noise at certain frequencies, especially those close to the frequency corresponding to $\frac{2\alpha}{\lambda}$. This can result in some information being obscured by noise. Therefore, it is necessary to weight the solved $\Psi_s(\boldsymbol{Q_p})$ in frequency domain, as is done in SSB. Theoretically, there are several methods for modulation [9]. In this section, a modulation method based on $\Gamma_{T,n}(\boldsymbol{Q_p})$ as weights is adopted, as shown in Equation (6-13). The reconstruction results with and without this weighting step are discussed in Section 6.4.

$$\Psi_{AD-ptycho}(\boldsymbol{Q_p}) = \Psi_s(\boldsymbol{Q_p}) \cdot \sum_{n=1}^{N} \left|\Gamma_{T,n}(\boldsymbol{Q_p})\right| \tag{6-13}$$

Subsequently, $\psi_{AD-ptycho}(\boldsymbol{R_p})$ can be calculated by inverse Fourier transform of $\Psi_{AD-ptycho}(\boldsymbol{Q_p})$ with respect to $\boldsymbol{Q_p}$:

$$\psi_{AD-ptycho}(\boldsymbol{R_p}) = \mathcal{F}^{-1}\{\Psi_{AD-ptycho}(\boldsymbol{Q_p})\} \tag{6-14}$$

6.3 Solvable Conditions and the Shape of $\boldsymbol{D_n}$

Rodenburg *et. al*. have demonstrated that a minimum of three detector sectors is necessary for a 4DSTEM dataset to convey information in all directions of ψ_s [3]. Building on this work, Brown *et. al*. have experimentally shown that a three-sector detector can achieve phase reconstruction using an iterative algorithm [5]. In this section, we will discuss why fewer than three sectors cannot reconstruct information in all directions, as well as scenarios involving three or more sectors.

In Equation (6-11), the region where $\Gamma_{T,n}(\boldsymbol{Q_p})$ is non-zero is relevant for solving Ψ_s. If at a certain $\boldsymbol{Q_p}$, this term is zero for all detectors, then Ψ_s at this $\boldsymbol{Q_p}$ cannot be solved. Therefore, the shape of $D_n(\boldsymbol{K_f})$ becomes important for its influence on $\Gamma_{T,n}(\boldsymbol{Q_p})$. In the previous section, we discussed the simplest case, without aberrations, $D_n(\boldsymbol{K_f}) = 1$. Under this condition, due to the symmetry of $A(\boldsymbol{K_f})A^*(\boldsymbol{K_f} + \boldsymbol{Q_p})$ and $A^*(\boldsymbol{K_f})A(\boldsymbol{K_f} - \boldsymbol{Q_p})$, $\Gamma_{T,n}(\boldsymbol{Q_p})$ is completely cancelled. Similarly, when $D_n(\boldsymbol{K_f})$ is a circle or ring in the $\boldsymbol{K_f}$ plane and concentric with the diffraction pattern, this symmetry still exists, hence $D_n(\boldsymbol{K_f})$ cannot be centrosymmetric.

If $D_n(\boldsymbol{K_f})$ is symmetrically divided into two equal regions, Ψ_s can be solved for most $\boldsymbol{Q_p}$. However, there is still some loss of information at the junction of the two halves of the detector region because at the edge of the detector, $\Gamma_{T,n}(\boldsymbol{Q_p})$ is cancelled. To demonstrate this a circular pattern with radial stripes inside was used as the transmission function to illustrate the flaws of this $D_n(\boldsymbol{K_f})$ layout. As shown in Figure 6-1, in the reconstruction, the horizontal stripe boundaries become blurred. This is because $\Gamma_{T,n}(\boldsymbol{Q_p})$ gradually approaches zero in the vertical region and is zero along the centre vertical axis, as shown in Figure 6-1. This leads to missing horizontal information in the reconstruction resulting in blurred horizontal stripes.

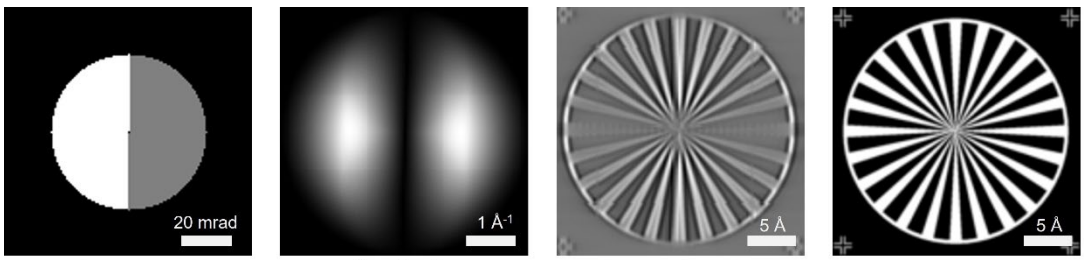

Figure 6-1 | Simulation of 2-half $D_n(\boldsymbol{K_f})$, $|\Gamma_{T,n}(\boldsymbol{Q_p})|$ and corresponding reconstruction. Accelerating voltage: 300 kV, convergence angle: 25 mrad, scan step size: 0.2 Å.

When $D_n(\boldsymbol{K_f})$ is evenly divided into three sector-shaped detectors, $|\Gamma_{T,n}(\boldsymbol{Q_p})|$ corresponding to the three sectors are still axially symmetric, as shown in the second row of Figure 6-2. However, when $|\Gamma_{T,n}(\boldsymbol{Q_p})|$ of the three sectors are summed, they cover all

directions (as shown in Figure 6-2), forming an annular non-zero region. This means that the reconstruction has information transfer in all directions, as shown in Figure 6-2.

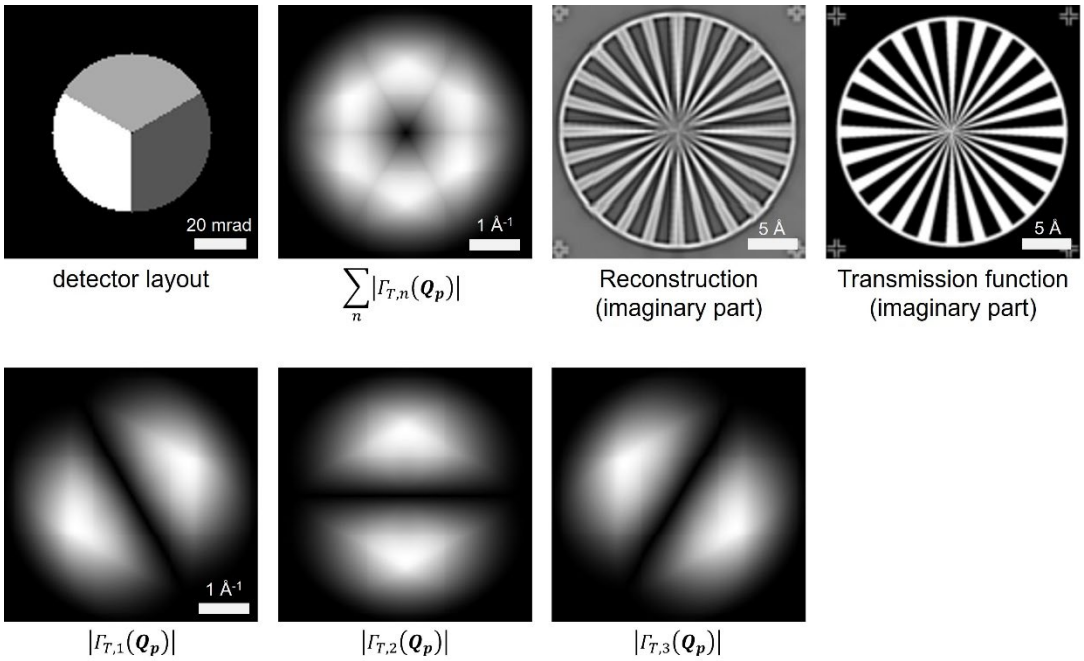

Figure 6-2 | Simulation of 3-sector $D_n(\boldsymbol{K_f})$, $|\Gamma_{T,n}(\boldsymbol{Q_p})|$ and corresponding reconstruction. Accelerating voltage: 300 kV, convergence angle: 25 mrad, scan step size: 0.2 Å.

If only one sector is considered, regardless of how $D_n(\boldsymbol{K_f})$ is divided into sectors, $A(\boldsymbol{K_f})A^*(\boldsymbol{K_f}+\boldsymbol{Q_p})$ and $A^*(\boldsymbol{K_f})A(\boldsymbol{K_f}-\boldsymbol{Q_p})$ remain 2-fold symmetric, and this symmetry does not change with the division of $D_n(\boldsymbol{K_f})$.

To explain the relation between symmetry of D_n and $|\Gamma_{T,n}(\boldsymbol{Q_p})|$, Figures 6-3 and 6-4 are plotted. In this simulation, for a specific detector sector, if there exists a $\boldsymbol{Q_p}$ such that $T_{Pos,n} > 0$, since the symmetry of $A(\boldsymbol{K_f})A^*(\boldsymbol{K_f}+\boldsymbol{Q_p})$ and $A^*(\boldsymbol{K_f})A(\boldsymbol{K_f}-\boldsymbol{Q_p})$, as shown in Figure 6-3 (b), then $T_{Neg,n}$ can be zero at $-\boldsymbol{Q_p}$ and results in $\Gamma_{T,n}(\boldsymbol{Q_p}) > 0$. If the vector $\boldsymbol{Q_p}$ is then rotated 180 degrees based on the origin, the condition of $T_{Pos,n}$ and $T_{Neg,n}$ will be exactly symmetric. But due to the additional negative sign in $T_{Neg,n}$ in Equation (6-9), $\Gamma_{T,n}(-\boldsymbol{Q_p})$ will be below zero, $\Gamma_{T,n}(-\boldsymbol{Q_p}) < 0$ and $\Gamma_{T,n}(\boldsymbol{Q_p}) = -\Gamma_{T,n}(-\boldsymbol{Q_p})$. This means $\Gamma_{T,n}(\boldsymbol{Q_p})$ exhibit central antisymmetry with respect to $\boldsymbol{Q_p}$. When taking the modulus of the complex $\Gamma_{T,n}(\boldsymbol{Q_p})$, $|\Gamma_{T,n}(\boldsymbol{Q_p})|$ will necessarily exhibit central symmetry.

Meanwhile, when $\boldsymbol{Q_p}$ takes values such that both $T_{Pos,n}$ and $T_{Neg,n}$ are nonzero, as shown in Figure 6-3 (c), $T_{Pos,n}$ and $T_{Neg,n}$ will cancel each other out in Equation (6-9), leading to a decrease in $\left|\Gamma_{T,n}(\boldsymbol{Q_p})\right|$. Under the segmented detector layout used in this simulation, each sector is axisymmetric, and the phase of Aperture function A, when considering only defocus, satisfies central symmetry about the origin. As a result, $\left|\Gamma_{T,n}(\boldsymbol{Q_p})\right|$ will reach zero at those values of $\boldsymbol{Q_p}$ where $T_{Pos,n}$ and $T_{Neg,n}$ exactly cancel each other out. These $\boldsymbol{Q_p}$ values are distributed along a line perpendicular to the symmetry axis of the corresponding sector and passing through the origin. Consequently, $\left|\Gamma_{T,n}(\boldsymbol{Q_p})\right|$ must simultaneously satisfy central symmetry and axisymmetry with a symmetry axis perpendicular to the symmetry axis of the corresponding sector. On the symmetry axis, $\left|\Gamma_{T,n}(\boldsymbol{Q_p})\right|$ equals zero.

In this simulation, no other $\boldsymbol{Q_p}$ values cause $\left|\Gamma_{T,n}(\boldsymbol{Q_p})\right|$ to be zero, therefore a two-fold symmetry arises. This symmetry is independent of how many sectors the detector is divided into within the 360° circular layout, as long as each detector sector itself satisfies axisymmetry and the symmetry axis passes through the origin.

However, since the symmetry axis of $\left|\Gamma_{T,n}(\boldsymbol{Q_p})\right|$ is related to the relative orientation of the sector and the origin, the symmetry axes of $\left|\Gamma_{T,n}(\boldsymbol{Q_p})\right|$ for each sector are not overlapping. When summing $\left|\Gamma_{T,n}(\boldsymbol{Q_p})\right|$ over all sectors (*i.e.* with respect to n) in Equation (6-13), $\sum_n\left|\Gamma_{T,n}(\boldsymbol{Q_p})\right|$ will reflect the same symmetry as the detector layout.

In this simulation, after aligning the three non-overlapping symmetry axes of the 3-sector layout, 6-fold symmetry is achieved, as demonstrated by $\sum_n\left|\Gamma_{T,n}(\boldsymbol{Q_p})\right|$ in Figure 6-2.

Generally, when $D_n(\boldsymbol{K_f})$ is divided into 3 or more sectors, it ensures that there is no direction in which $\left|\Gamma_{T,n}(\boldsymbol{Q_p})\right|$ is entirely zero (since those symmetry axes of $\left|\Gamma_{T,n}(\boldsymbol{Q_p})\right|$, where $\left|\Gamma_{T,n}(\boldsymbol{Q_p})\right| = 0$, are not overlapping on every n). This implies that as long as $D_n(\boldsymbol{K_f})$ is divided into 3 or more sectors, it is possible to recover information about Ψ_s in all directions. This conclusion is consistent with the literature [3, 5].

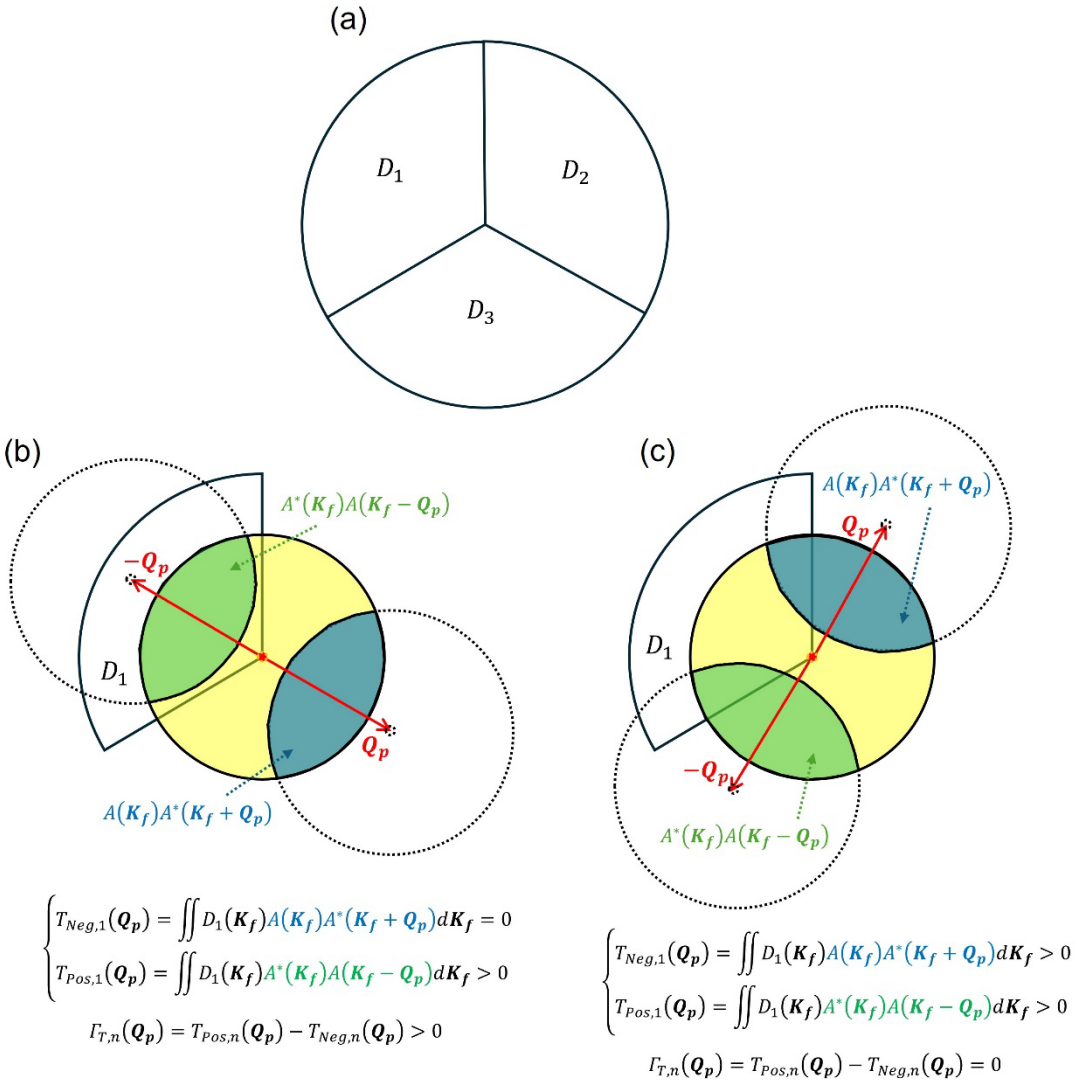

Figure 6-3 | Schematic figure of the symmetry of D_n and $\Gamma_{T,n}$. (a) 3-sector detector layout. (b) The situation of $T_{Pos,1}(\boldsymbol{Q_p})>0$ *and* $T_{Neg,1}(\boldsymbol{Q_p})=0$. (c) The situation of $T_{Pos,1}(\boldsymbol{Q_p})$ and $T_{Neg,1}(\boldsymbol{Q_p})$ cancelling out each other.

However, this does not mean that $D_n(\boldsymbol{K_f})$ can be divided arbitrarily. For example, an extreme case is considered as detector is divided into two 1° sectors and one 358° sector (as shown in Figure 6-4). In Figure 6-4 (b) $T_{Pos,1}$ and $T_{Neg,1}$ cancel each other. The only chance that the 358° sector has a non-zero $\Gamma_{T,n}(\boldsymbol{Q_p})$ is as Figure 6-4 (c). Only if either $A(\boldsymbol{K_f})A^*(\boldsymbol{K_f}+\boldsymbol{Q_p})$ or $A^*(\boldsymbol{K_f})A(\boldsymbol{K_f}-\boldsymbol{Q_p})$ covers those two 1° sectors, $\Gamma_{T,n}(\boldsymbol{Q_p})$ can be non-zero because of the incomplete cancelling out of $T_{Pos,1}(\boldsymbol{Q_p})$ and $T_{Neg,1}(\boldsymbol{Q_p})$. Therefore, the 358° sector is equivalent to a small 2° sector around its symmetric axis. So

this layout is approximately equivalent to a 2-sector layout with two 1°~2° symmetric sectors at the opposite side of the origin. There will still be a direction in which $|\Gamma_{T,n}(\boldsymbol{Q_p})|$ is approximately zero. Therefore, in general, sectors are usually divided evenly, so that each sector has the same area. This approach also ensures that the data output from each sector has a similar signal-to-noise ratio.

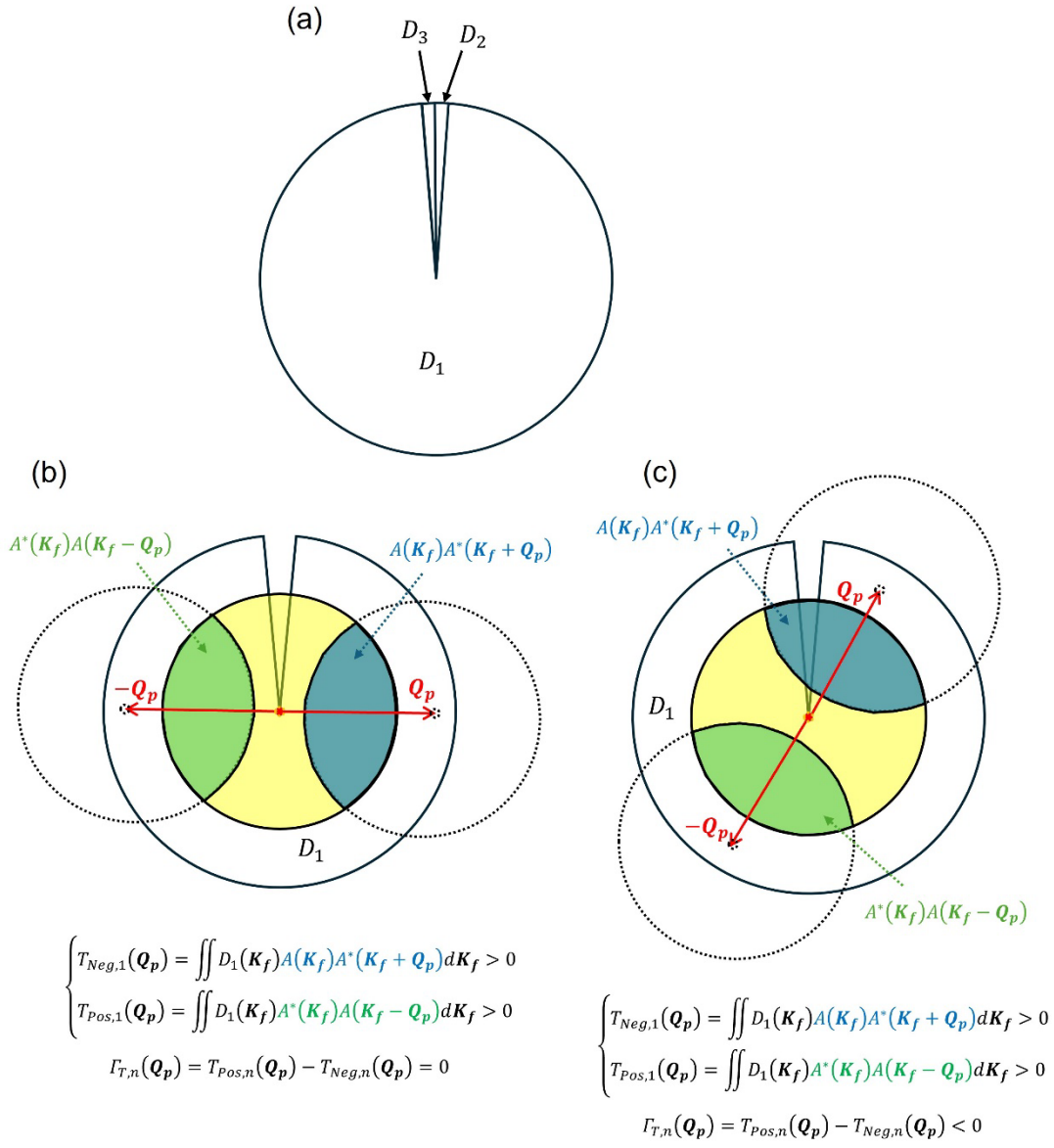

Figure 6-4 | Schematic figure of the symmetry of D_n and $\Gamma_{T,n}$ on an unevenly divided detector. (a) Unevenly divided 3-sector detector layout. (b) The situation of $T_{Pos,1}(\boldsymbol{Q_p})$ and $T_{Neg,1}(\boldsymbol{Q_p})$ cancelling out each other. (c) The situation of $T_{Pos,1}(\boldsymbol{Q_p}) > 0$, $T_{Neg,1}(\boldsymbol{Q_p}) > 0$ but $\Gamma_{T,n} < 0$.

As shown in Figures 6-3 and 6-4, if $|\boldsymbol{Q_p}| > 2\alpha$, then both $T_{Neg,n}(\boldsymbol{Q_p})$ and $T_{Pos,n}(\boldsymbol{Q_p})$ will be zero, regardless of the detector layout. This results in an all-zero linear system in Equation (6-10), rendering Ψ_s unsolvable. In other words, information beyond $|\boldsymbol{Q_p}| > 2\alpha$ cannot be reconstructed using AD-ptycho. The resolution of AD-ptychography is limited to 2α, which is the same as that of the SSB and COM methods.

6.4 Weighting of $\boldsymbol{\Psi_s}$ and Information Transfer

As discussed in Section 6.2, Equation (6-13) assigns different weights to information at different frequencies on Ψ_s based on the distribution of $D_n(\boldsymbol{K_f})$. This is equivalent to modulating a transfer function. This modulation is similar to noise reduction by changing the frequency transfer in SSB ptychography [9]. Without this modulation, there will be increased noise at high (up to $\frac{2\alpha}{\lambda}$) or low frequencies. A transfer function that damps both low and high frequencies help to suppress noise, resulting in a reconstruction that emphasises information with a high signal-to-noise ratio. In other words, the contrast transfer function prevents an AD-ptycho (or SSB) reconstruction from being dominated by high or low frequency noise, in regions where the sample information has weaker transmission.

Since the unweighted AD-ptycho result lacks filtering for low and high frequency noise, the result appears noisy, and high frequency artifacts may obscure the sample information. A simulated reconstruction of monolayer graphene with a silicon atom replacement was calculated to compare the AD-ptycho with and without the weighting process described in Equation (6-12). These results are plotted in Figure 6-5. For three evenly distributed sectors, the unweighted $|\Psi_s|$ shows high-intensity positions at frequencies close to $\frac{2\alpha}{\lambda}$. In comparison, the weighted $|\Psi_{AD-ptycho}|$ reflects the frequency information of the sample, and the reconstruction is closer to the SSB reconstruction.

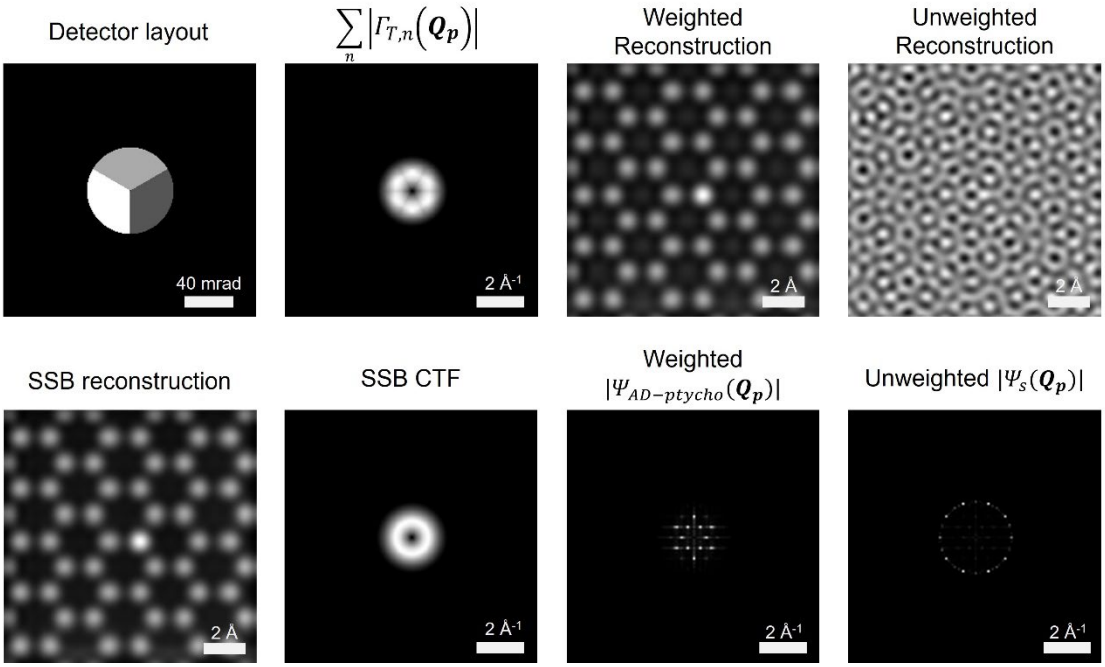

Figure 6-5 | Simulations of monolayer graphene with a Silicon atom replacement, reconstructed by: AD-ptycho with/without weighting on Ψ_s, and SSB. Voltage: 80 kV, convergence angle: 30 mrad, scan step: 0.1 Å, aberration free. The weighted AD-ptycho reconstruction shows the same graphene lattice as in SSB reconstruction. Weighted $|\Psi_{AD-ptycho}|$ shows high-intensity positions corresponding to the periodic lattice structure. In the Unweighted AD-ptycho reconstruction, lattice is covered by high-frequency artefacts, which can be seen in unweighted $|\Psi_s|$ as high-intensity positions at frequencies close to $\frac{2\alpha}{\lambda}$.

The simulations in Figure 6-5 demonstrate the necessity of the weighting step in Equation (6-13). However, it is worth noting that this weighting step is also consistent with the modulation of the contrast transfer function. For AD-ptycho, the weight distribution $\sum|\Gamma_{T,n}(\boldsymbol{Q_p})|$ at different frequencies in Equation (6-13) is related to the shape of $D_n(\boldsymbol{K_f})$, as shown in Equation (6-6) and (6-9).

However, the weighting method for AD-ptycho is not unique and other weighting methods as alternatives to Equation (6-13) can yield different results. If the transfer function of SSB or centre-of-mass is used to weight Ψ_s, results similar to the corresponding methods can

be obtained. Specifically, for SSB, applying the transfer function for SSB (see Chapter 3) to modulate the solved Ψ_s leads to results that are identical to the SSB algorithm if the detector condition is the same.

6.5 Aberration Correction

As a method based on non-iterative ptychography, AD-ptycho can also compensate for known aberrations. For data requiring aberration correction, when calculating $\Gamma_{T,n}(\boldsymbol{Q_p})$, the effect of aberrations needs to be considered in the phase of the aperture function A. In this section, only the correction of defocus is simulated.

To demonstrate aberration correction in AD-ptycho using segmented detectors, a series of simulations were generated using Python. The simulations utilised a sample of monolayer graphene with a silicon atom replacement. For simplicity, only defocus was considered in these simulations. However, the same algorithm and workflow can be applied for correction of other aberrations.

Simulations used 80 kV acceleration voltage and a 30 mrad convergence angle. The probe scan step was set to 0.1 Å. At the detector plane, a virtual hollow 4-sector detector was used, as shown in Figure 6-6. The radius of the centre hole was 10 mrad, and the maximum collection angle was 50 mrad. The weighting function $\sum|\Gamma_{T,n}(\boldsymbol{Q_p})|$ was varied for different defocus conditions, as shown in Figure 6-6. The AD-ptycho reconstruction results are shown in Figure 6-7 with corresponding iDPC images and ABF images for comparison.

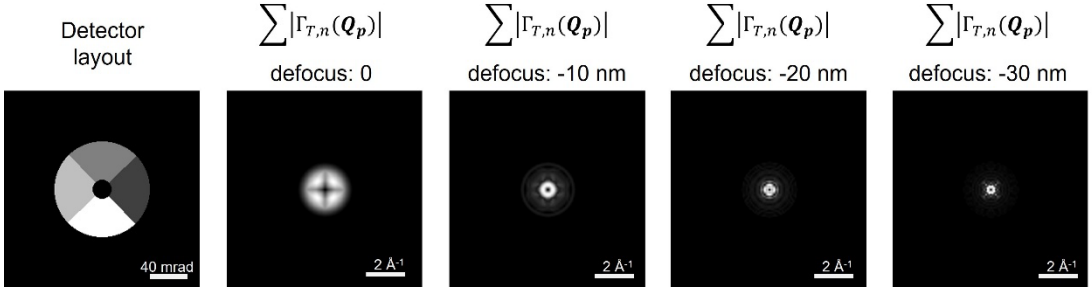

Figure 6-6 | 4-Fold detector geometry and $\sum|\Gamma_{T,n}(\boldsymbol{Q_p})|$ for various defocus conditions.

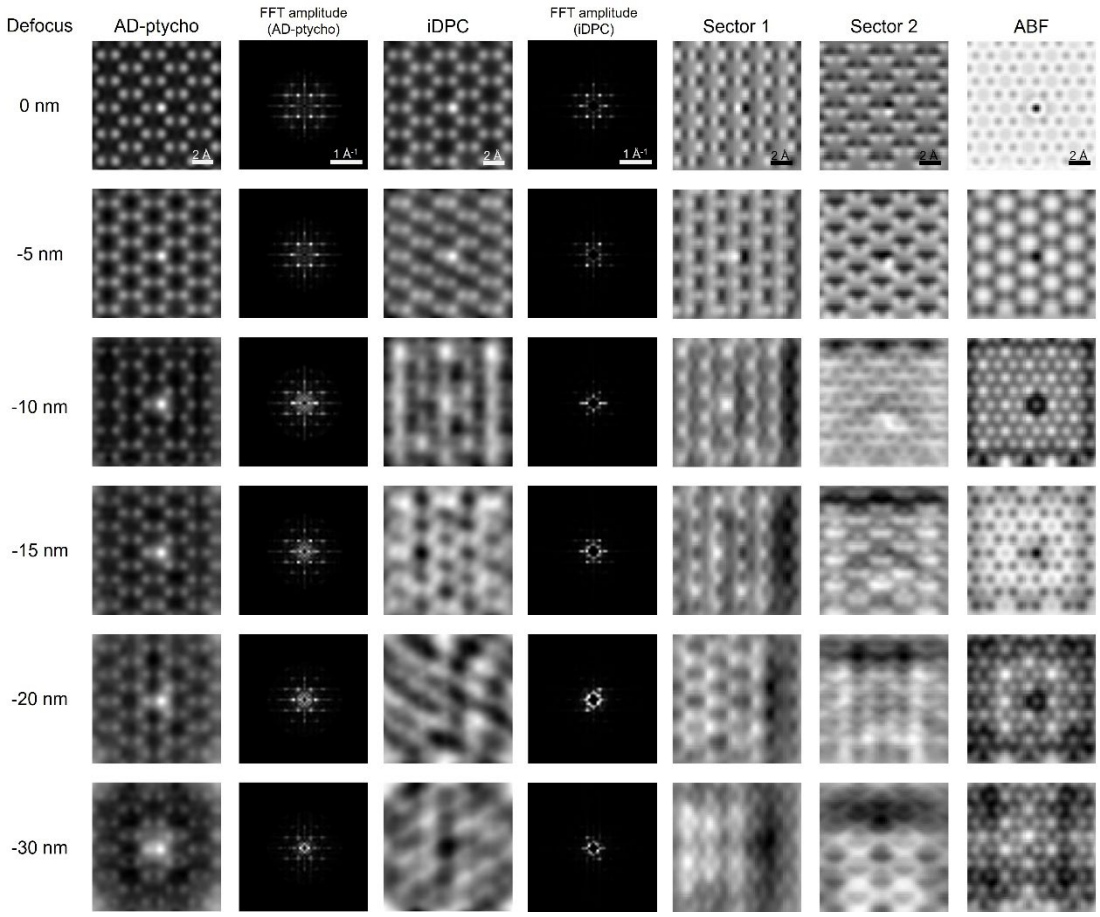

Figure 6-7 | Simulation of AD-ptycho reconstructions under different defocus conditions on a sample of monolayer graphene with a silicon atom replacement. iDPC images are reconstructed from the same datasets. A high-pass filter (cut-off at 0.25 $Å^{-1}$) is applied to the iDPC images. The dark areas in the centre of iDPC FFT amplitude images indicates the frequency be filtered. Sector 1 and Sector 2 images are generated directly from the right-side and upper-side sectors shown in Figure 6-6. ABF images are the summation of all 4 sectors. AD-ptycho reconstructions shows lattice structure and high intensity positions in the corresponding FFT amplitude images up to a -30 nm defocus. iDPC, ABD and image for a single channel are blurred when defocus is over -10 nm.

In Figure 6-7, according to the real space reconstruction and FFT amplitude images, AD-ptycho reconstructs lattice information for the graphene sample up to a -30 nm defocus, with the silicon atom substitution clearly identifiable. As the defocus varies, the weighting function used in the algorithm also changes, leading to progressively brighter low-

frequency components (below 0.5 $Å^{-1}$, corresponding to within 6 pixels of the centre) in the FFT amplitude images. Due to these low-frequency enhancements, the contrast in real-space images appears to have changed. More low-frequency noise emerges near the silicon atom, altering the relative contrast of the carbon atoms. For example, at -10 nm defocus, the vacuum between carbon atoms tends to appear darker than at -5 nm defocus. However, this is a result of the modulation by the weighting function in the frequency domain. As can be seen from the FFT amplitude images, the highest resolvable frequency is the same in both cases.

In comparison, for iDPC and single sector images, defoci over -10 nm result in blurring, failing to reveal the lattice structure. The highest resolvable frequency in FFT amplitude decreases with increasing defocus. ABF images, which are the summation of all four sectors, exhibit multiple contrast reversals as the defocus changes, making it difficult to extract information from a defocused ABF image.

In Figure 6-7, the AD-ptycho reconstructions with defocus values of -20 and -30 nm exhibit bright regions in the four corners. This is caused also by the weighting function discussed in the previous section. The weighting functions $\sum|\Gamma_{T,n}(\boldsymbol{Q_p})|$ for 0, -10, -20 and -30 nm are plotted in Figure 6-6. Compared to those at smaller defocus values, the -20 nm and -30 nm weighting functions assign greater emphasis to low-frequency regions. As a result, low frequency artefacts are more obvious in the reconstruction results at -20 and -30 nm defocus.

These simulations demonstrates that 4-sector AD-ptycho can compensate for defocus. For example, under the simulation condition in this section, it can correct the defocus below -30 nm for images of a graphene lattice and point defects as shown in Figure 6-7.

6.6 Experimental data

The AD-ptycho algorithm has also been validated on experimental data. In this section, 3 experimental datasets are presented. The first dataset is collected using a 4-sector annular detector from a sample of gold nanoparticles with an in-focus probe. The second and third

datasets are 4DSTEM datasets of monolayer graphene, recorded with in-focus and defocus settings (defocus of 0 and -7.5 nm). The 4DSTEM datasets were masked with virtual detector areas and then integrated to form either 4-sector or 16-sector detectors.

The first dataset was recorded on a JEM-ARM300F at a voltage of 300 kV and in-focus condition, with the results shown in Figure 6-8. In addition to the AD-ptycho reconstruction, images from the four sectors (detector channels) were summed to create an ABF image, where the lattice details are obscured. The AD-ptycho reconstruction, however, is able to recover lattice detail. An aberration-free probe is used in the AD-ptycho algorithm for reconstruction of this dataset.

As comparison, iDPC reconstruction using the same detector images is also plotted in Figure 6-8. iDPC reconstruction, without weighting in frequency, shows the same lattice as AD-ptycho reconstruction. However, with the weighting of Ψ_s, AD-ptycho reconstruction weaken the low-frequency background.

The second and third datasets were recorded at 80 kV, with a convergence angle of 24.5 mrad. The second dataset is recorded with an in-focus probe, while the third dataset is recorded with a defocused probe (-7.5 nm). Both datasets are processed by virtual 4-sector and 16-sector detector functions and then reconstructed using the AD-ptycho algorithm.

The results of the in-focus dataset are shown in Figure 6-9. For comparison, the SSB reconstruction of the same dataset is shown in Figure 6-10 (in-focus dataset). It shows that in-focus AD-ptycho reconstructs the lattice information as predicted by simulations, demonstrating that it can serve as a replacement for SSB for in-focus conditions.

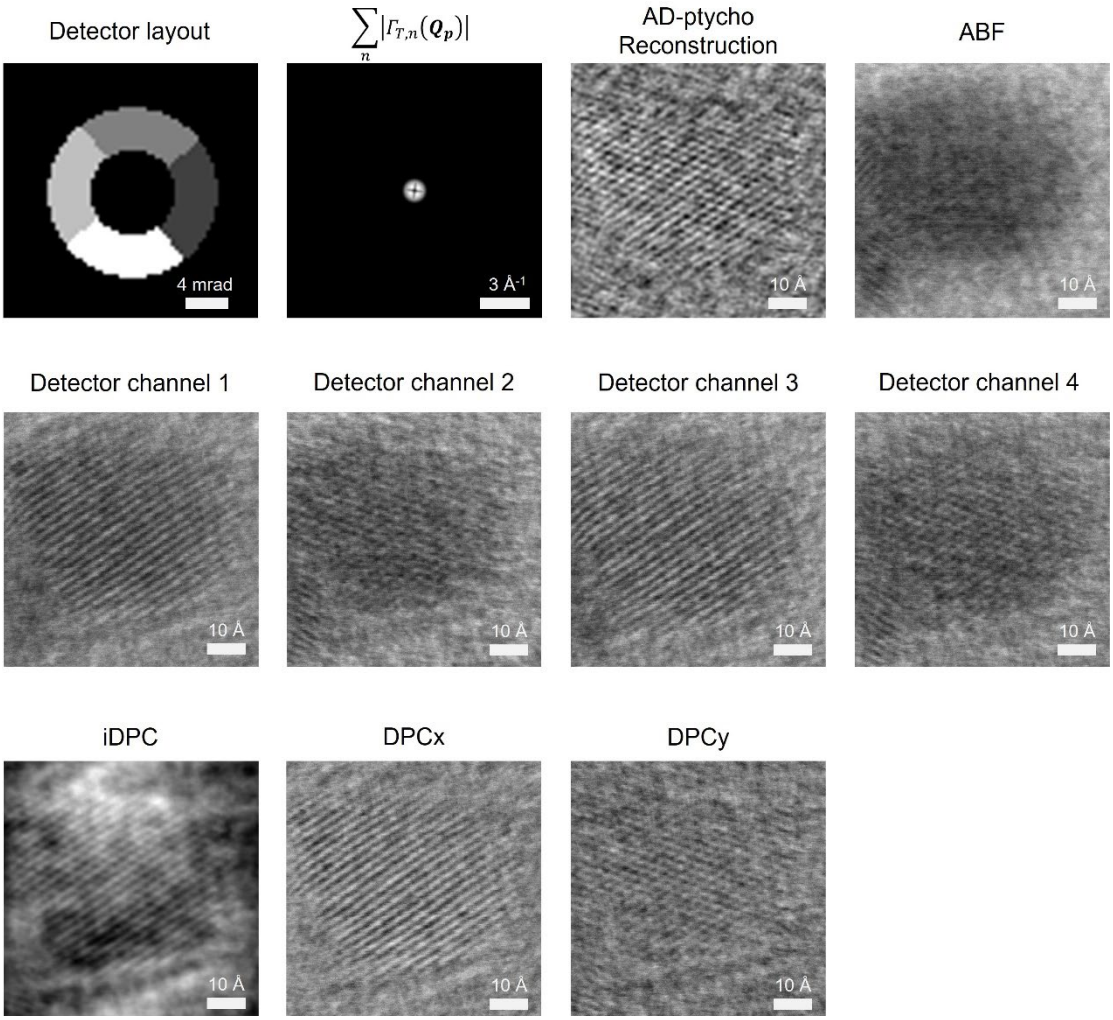

Figure 6-8 | AD-ptycho and iDPC reconstruction of a gold nanoparticle. The ABF image is the sum of detector channels 1 to 4. DPCx/y and iDPC images are reconstructed using the same detector channel images. Voltage: 300 kV, convergence angle: 8 mrad, scanning position: 1024×1024. The lattice of the nanoparticle is shown in the 4-segmented AD-ptycho reconstruction which no lattice can be recognised in the summed ADF image.

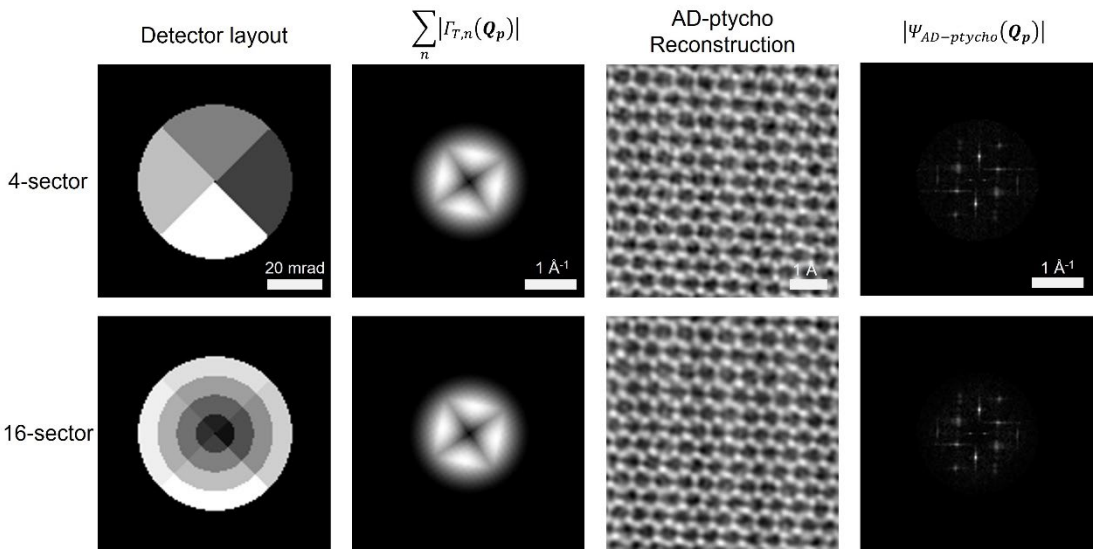

Figure 6-9 | In-focus AD-ptycho reconstruction of monolayer graphene with 4-sector and 16-sector virtual detector functions. The $|\Psi_{AD-ptycho}(\boldsymbol{Q_p})|$, which shows the intensity distribution of AD-ptycho reconstructions on frequency domain, are plotted with reconstructions.

In the in-focus condition, the reconstruction results of the 4-sector and 16-sector detectors, using the same rotation angle and collection angle range (0–28 mrad in Figure 6-9), are very similar. This occurs because, for in-focus datasets, $A(\boldsymbol{K_f})A^*(\boldsymbol{K_f}+\boldsymbol{Q_p})$ and $A^*(\boldsymbol{K_f})A(\boldsymbol{K_f}-\boldsymbol{Q_p})$ are real and exhibit a uniform intensity distribution within non-zero regions. Consequently, the integrated intensity on a detector channel $D_n(\boldsymbol{K_f})$ only depends on its area. As a result, the summation of 4 sectors in a direction of the 16-sector layout conveys the same information as 1 sector in the 4-sector layout. Therefore, when the rotation angle and collection angles (minimum and maximum) are identical, the reconstruction is theoretically equivalent. As shown in Figure 6-9, the AD-ptycho reconstructions for the 4-sector and 16-sector detector layouts are nearly identical, with the same number of high-intensity positions on $|\Psi_{AD-ptycho}(\boldsymbol{Q_p})|$ resolved. The experimental reconstruction is consistent with theoretical prediction.

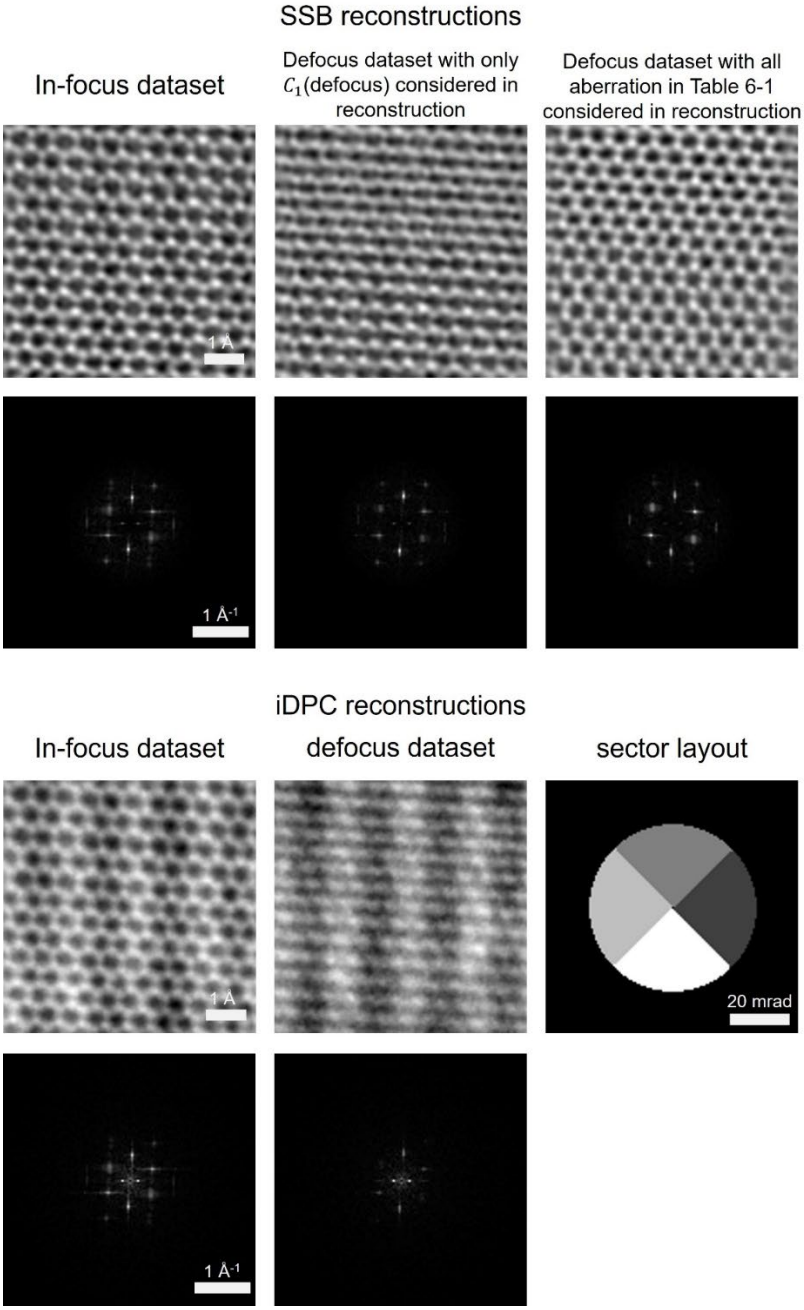

Figure 6-10 | SSB and iDPC reconstructions of monolayer graphene for in-focus and defocused datasets and corresponding power spectrums. The distortion of lattice may be caused by sample drifting during data collection. iDPC reconstructions are neither weighted nor filtered in the frequency domain. The vertical stripes observed in iDPC images originate from scanning noise present in the 4D-STEM dataset. This same noise also appears in SSB reconstructions, but it is suppressed by the intrinsic transfer function of SSB, which reduces low-frequency components where the noise is most prominent.

For the defocused dataset, aberrations were measured from the 4DSTEM dataset using an iterative SVD method [10] and are listed in Table 6-1. AD-ptycho reconstructions are then carried out with only C_1 (defocus) considered or using all the aberrations listed in Table 6-1.

Table 6-1 | aberrations measured from the defocused dataset.

C_1	-7.51 nm
$C_{12} - x, y$	1.13, 3.74 nm
$C_{21} - x, y$	-212, -64.4 nm
$C_{23} - x, y$	58.8, -101 nm
C_3	66 nm
$C_{32} - x, y$	-3.96, 1.20 μm
$C_{34} - x, y$	7.26, 0.826 μm

The AD-ptycho reconstruction with only C_1 considered is shown in Figure 6-11, while the results with all aberrations in Table 6-1 considered is shown in Figure 6-12. In Figures 6-11 and 6-12, the reconstruction results for the 16-sector detector resolve more high-intensity positions (red and green marks in Figure 6-11 and red mark in Figure 6-12) than those for 4-sectors. This indicates for reconstructions with a defocused probe, more sectors allow more information to be reconstructed. This is because, for dataset with aberration, $A(\boldsymbol{K_f})A^*(\boldsymbol{K_f} + \boldsymbol{Q_p})$ and $A^*(\boldsymbol{K_f})A(\boldsymbol{K_f} - \boldsymbol{Q_p})$ are complex arrays. The phase shift in $A(\boldsymbol{K_f})$ makes $A(\boldsymbol{K_f})A^*(\boldsymbol{K_f} + \boldsymbol{Q_p})$ and $A^*(\boldsymbol{K_f})A(\boldsymbol{K_f} - \boldsymbol{Q_p})$ not flatly distributed in their non-zero areas. So, the intensity integrated on a detector channel $D_n(\boldsymbol{K_f})$ is no longer determined by its area only. As a result, from Equation 6-12, more sectors provide more independent equations in the system, which helps to obtain a more accurate solution of Ψ_s.

In principle, in terms of aberration compensation capability, more independent channels (sectors) on the detector, are desirable. However, the reconstruction results of the 16-sector condition still missed information compared to the SSB reconstruction results (Figure 6-10), especially at the horizontal direction on $|\Psi_{AD-ptycho}(\boldsymbol{Q_p})|$. The two high-intensity positions along the horizontal direction in the power spectrum of the SSB shown in Figure 6-10 cannot be identified in either the 4-sector or 16-sector AD-ptycho reconstructions in Figures 6-11 and 6-12. This is because the SSB uses a pixelated 4DSTEM dataset, which can be considered as for example 128*128 independent detector channels. In the next section, the relationship between SSB and AD-ptycho is discussed.

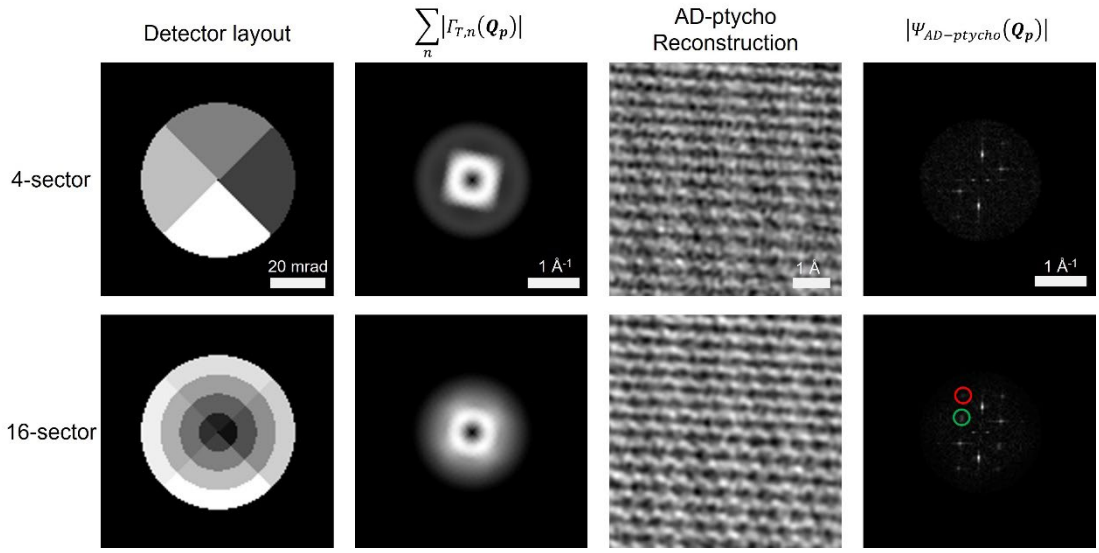

Figure 6-11 | AD-ptycho reconstructions with only C_1 (defocus) considered. The $|\Psi_{AD-ptycho}(\boldsymbol{Q_p})|$, which shows the intensity distribution of AD-ptycho reconstructions on frequency domain, are plotted with reconstructions. For 16-sector reconstruction, 2 additional high-intensity positions on $|\Psi_{AD-ptycho}(\boldsymbol{Q_p})|$ (red and green mark) can be recognised compared to 4-sector reconstruction. This indicates that reconstruction with 16-sector detector has more information reconstructed than 4-sector detector layout when with only C_1 (defocus) considered.

Comparing Figures 6-11 (only C_1 in the reconstruction) and 6-12 (all aberrations in Table 6-1 considered), the phase of the aperture function A differs due to the consideration of

higher order aberrations. From an algorithmic perspective, this changes $\Gamma_{T,n}(\boldsymbol{Q_p})$, altering the solution to Equation (6-12) and consequently the weighting function $\sum|\Gamma_{T,n}(\boldsymbol{Q_p})|$. In Figure 6-12, after considering higher-order aberrations, the reconstruction results better reveal the hexagonal crystal structure. In 4-sector reconstruction, the green marked high-intensity position can be identified after considering high-order aberrations.

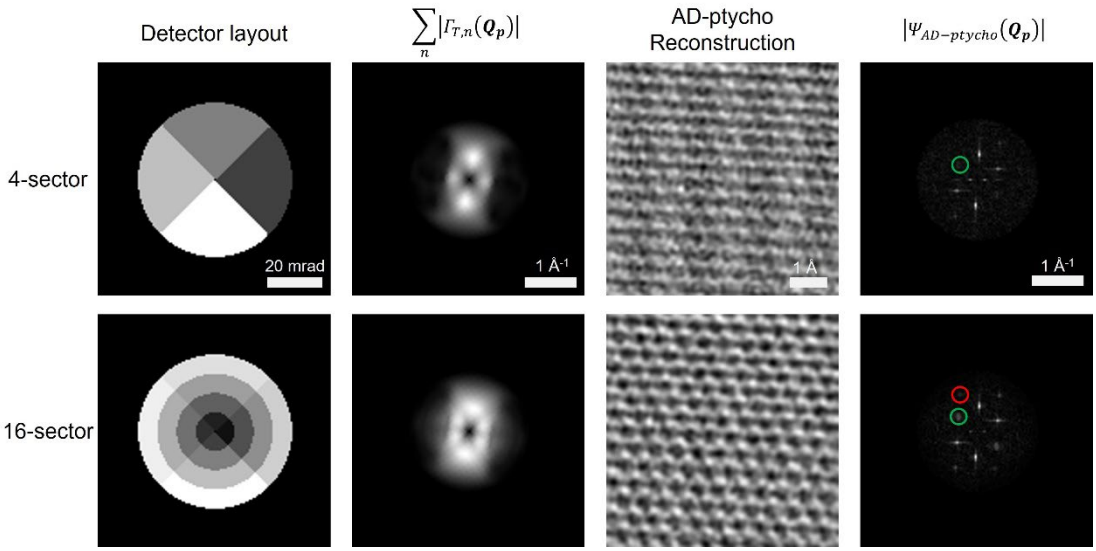

Figure 6-12 | AD-ptycho reconstructions with all aberrations in Table 6-1 considered. The $|\Psi_{AD-ptycho}(\boldsymbol{Q_p})|$, which shows the intensity distribution of AD-ptycho reconstructions on frequency domain, are plotted with reconstructions. On the 4-sector reconstruction, compared to reconstruction only C_1 (defocus) considered in Figure 6-11, one additional high-intensity position on $|\Psi_{AD-ptycho}(\boldsymbol{Q_p})|$ (green mark) can be recognised. This indicates considering more aberrations improves the quality of reconstruction. For 16-sector reconstruction, high-intensity position marked by red circle can be recognised. But this position is still missing on 4-sector reconstruction.

In experimental data processing, obtaining more accurate aberration estimations is also important. This is a disadvantage of AD-ptycho based on segmented detectors compared to SSB based on 4DSTEM. Aberrations can be estimated through the 4DSTEM dataset [10]. However, the same method to estimate aberrations cannot be applied to dataset recorded by segmented detectors with fewer channels. Therefore, for such segmented

detectors, it relies on aberration corrector to have aberration estimations.

Comparing to the iDPC reconstructions shown in Figure 6-10, 16-sector AD-ptycho reconstructions demonstrate improved resolution under defocused probe condition, as the red marked high-intensity position shows on $|\Psi_{AD-ptycho}(\boldsymbol{Q_p})|$ in Figures 6-11 and 6-12. However, for 4-sector AD-ptycho reconstructions and iDPC at a defocus of -7.5 nm, the experimental results do not show the clear differences predicted by the simulations (Figure 6-7). This discrepancy may require further investigation. One possible explanation could be calibration inaccuracies, such as inaccuracy in the pixel size of the original 4D-STEM dataset in the diffraction plane or in the aberration parameters used in the AD-ptycho algorithm.

In this section, AD-ptycho has been demonstrated for in-focus and defocused probe conditions using three experimental datasets with different detector geometries. Overall, AD-ptycho can reconstruct phase information from samples and compensate for aberrations in experimental data. However, this depends on the specific detector layout and the pre-estimation of aberrations. Beyond 2D imaging, with a well-aligned STEM, defocus or z-direction information may be obtained by combing AD-ptycho with defocus optical sectioning as in ptychography [10].

6.7 Connections to Other 4DSTEM methods

For AD-ptycho, the choice of D_n can be arbitrary. Different choices of D_n show the correlation between AD-ptycho and other 4DSTEM imaging methods. In this section, the relationship between AD-ptycho, SSB, and traditional centre-of-mass is discussed for specific D_n and frequency weights.

6.7.1 Connection to Single-side Band

Single-side band (SSB) ptychography can be regarded as a special case and extension of AD-ptycho. If a pixelated detector is considered as an array of multi-channel detectors formed by N pixels, which are positioned at $\boldsymbol{K_{f,n}}$ in the diffraction plane, and each pixel

has the same $D_{n,SSB}(\boldsymbol{K_f})$:

$$D_{n,SSB}(\boldsymbol{K_f}) = \begin{cases} 1, & where\ \boldsymbol{K_f} = \boldsymbol{K_{f,n}} \\ 0, & where\ \boldsymbol{K_f} \neq \boldsymbol{K_{f,n}} \end{cases} \tag{6-15}$$

Under this $D_{n,SSB}(\boldsymbol{K_f})$ condition, the integral with respect to $\boldsymbol{K_f}$ in Equation (6-2) and (6-4) can be simplified since $D_{n,SSB}(\boldsymbol{K_f})$ is non-zero only at a specific pixel, the nth pixel on the detector. Therefore, $C_n(\boldsymbol{R_p})$ and $c_n(\boldsymbol{Q_p})$ are:

$$\begin{aligned} C_n(\boldsymbol{R_p}) &= \left|M(\boldsymbol{K_{f,n}}, \boldsymbol{R_p})\right|^2 \\ c_n(\boldsymbol{Q_p}) &= \Psi_s^*(-\boldsymbol{Q_p})A(\boldsymbol{K_{f,n}})A^*(\boldsymbol{K_{f,n}} + \boldsymbol{Q_p}) + \Psi_s(\boldsymbol{Q_p})A^*(\boldsymbol{K_{f,n}})A(\boldsymbol{K_{f,n}} - \boldsymbol{Q_p}) \end{aligned} \tag{6-16}$$

In Equation (6-16), $C_n(\boldsymbol{R_p})$ represents an image of the same pixel $\boldsymbol{K_{f,n}}$ for all probe positions in a scan. Consequently, $c_n(\boldsymbol{Q_p})$ in Equation (6-16) is a simplified form of Equation (6-5). $\Psi_s(\boldsymbol{Q_p})$ can be solved by substituting Equation (6-15) into Equations (6-6) and (6-11), as shown in Equation (6-17).

$$\begin{aligned} &\begin{cases} c_1(\boldsymbol{Q_p}) \cdot \Gamma_{T,1,SSB}(\boldsymbol{Q_p})^* = \Psi_s(\boldsymbol{Q_p}) \cdot \left|\Gamma_{T,n,SSB}(\boldsymbol{Q_p})\right|^2 \\ \vdots \\ c_N(\boldsymbol{Q_p}) \cdot \Gamma_{T,N,SSB}(\boldsymbol{Q_p})^* = \Psi_s(\boldsymbol{Q_p}) \cdot \left|\Gamma_{T,n,SSB}(\boldsymbol{Q_p})\right|^2 \end{cases} \\ &where\ \Gamma_{T,n,SSB}(\boldsymbol{Q_p}) \\ &\qquad = \iint D_{n,SSB}(\boldsymbol{K_f}) \cdot A^*(\boldsymbol{K_f})A(\boldsymbol{K_f} - \boldsymbol{Q_p}) \cdot d\boldsymbol{K_f} \\ &\qquad - \iint D_{n,SSB}(\boldsymbol{K_f}) \cdot A(\boldsymbol{K_f})A^*(\boldsymbol{K_f} + \boldsymbol{Q_p}) \cdot d\boldsymbol{K_f} \end{aligned} \tag{6-17}$$

In Single-side Band ptychography, as discussed in Chapter 3, $G(\boldsymbol{K_f}, \boldsymbol{Q_p})$ is calculated by a Fourier transform with respect to the scanning coordinate $\boldsymbol{R_p}$, which is the same way that $c_n(\boldsymbol{Q_p})$ is calculated from $C_n(\boldsymbol{R_p})$. The function $G(\boldsymbol{K_f}, \boldsymbol{Q_p})$ with the delta function at 0 frequency ignored, is expressed as:

$$\mathrm{G}(\boldsymbol{K_f}, \boldsymbol{Q_p}) = \Psi_s^*(-\boldsymbol{Q_p})A(\boldsymbol{K_f})A^*(\boldsymbol{K_f} + \boldsymbol{Q_p}) + \Psi_s(\boldsymbol{Q_p})A^*(\boldsymbol{K_f})A(\boldsymbol{K_f} - \boldsymbol{Q_p}) \tag{6-18}$$

The functions $c_n(\boldsymbol{Q_p})$ and $\mathrm{G}(\boldsymbol{K_f}, \boldsymbol{Q_p})$ in Equations (6-16) and (6-18) are actually the same

array using a different notation.

In SSB ptychography (introduced in Chapter 3), two double-overlapping areas (DO areas) are processed with corresponding mask functions. Taking the $-\boldsymbol{Q_p}$ term as an example, the complex function in the DO area is masked using a conjunction of $A^*(\boldsymbol{K_f})A(\boldsymbol{K_f}-\boldsymbol{Q_p})$. It is then summed with respect to $\boldsymbol{K_f}$ to calculate $\Psi_{s.SSB}(\boldsymbol{Q_p})$, *i.e.* $\Psi_s(\boldsymbol{Q_p})$ reconstructed by SSB, as:

$$\Psi_{s,SSB}(\boldsymbol{Q_p}) = \Psi_s(\boldsymbol{Q_p}) \iint \left|A(\boldsymbol{K_f})\right|^2 \left|A(\boldsymbol{K_f}-\boldsymbol{Q_p})\right|^2 d\boldsymbol{K_f} \tag{6-19}$$

However, in AD-ptycho, $\Psi_s(\boldsymbol{Q_p})$ is solved without the factor $\iint \left|A(\boldsymbol{K_f})\right|^2 \left|A(\boldsymbol{K_f}-\boldsymbol{Q_p})\right|^2 d\boldsymbol{K_f}$, which is the transfer function for SSB ptychography. If in the weighting step, *i.e.* Equation (6-13), a weighting factor of $\sum_{n=1}^{N}\left|\Gamma_{T,n}(\boldsymbol{Q_p})\right|$ is replaced by $\iint \left|A(\boldsymbol{K_f})\right|^2 \left|A(\boldsymbol{K_f}-\boldsymbol{Q_p})\right|^2 d\boldsymbol{K_f}$, the results of SSB and AD-ptycho with $D_n(\boldsymbol{K_f})$ defined by Equation (6-15) should be identical. Therefore, SSB ptychography can be regarded as an extension of AD-ptycho with a specific $D_n(\boldsymbol{K_f})$ function and frequency weighting.

6.7.2 Connection to COM

In AD-ptycho, if we consider a weighted pixelated detector function as:

$$D_{n,COM}(\boldsymbol{K_f}) = \begin{cases} K_{fx,n}, & where\ \boldsymbol{K_f} = \boldsymbol{K_{f,n}} \\ 0, & where\ \boldsymbol{K_f} \neq \boldsymbol{K_{f,n}} \end{cases} \tag{6-20}$$

In this case, there are N detector channels, and each detector channel is weighted by its projected angular distance to the x (or y) axis from the centre point. $C_n(\boldsymbol{R_p})$ and $c_n(\boldsymbol{Q_p})$ can thus be written as:

$$\begin{aligned} C_{n,comx}(\boldsymbol{R_p}) &= K_{fx} \cdot \left|M(\boldsymbol{K_{f,n}}, \boldsymbol{R_p})\right|^2 \\ c_{n,comx}(\boldsymbol{Q_p}) &= K_{fx,n}\left|A(\boldsymbol{K_{f,n}})\right|^2 \delta(\boldsymbol{Q_p}) + K_{fx,n}A(\boldsymbol{K_{f,n}})A^*(\boldsymbol{K_{f,n}}+\boldsymbol{Q_p})\Psi_s^*(-\boldsymbol{Q_p}) \\ &\quad + K_{fx,n}A^*(\boldsymbol{K_{f,n}})A(\boldsymbol{K_{f,n}}-\boldsymbol{Q_p})\Psi_s(\boldsymbol{Q_p}) \end{aligned} \tag{6-21}$$

The $C_{n,comx}(\boldsymbol{R_p})$ here is the same as the traditional Centre-of-mass algorithm. Based on the process above we can rewrite the factors in Equation (6-6) as:

$$\begin{cases} T_{Neg,n}(\boldsymbol{Q_p}) = K_{fx,n}A(\boldsymbol{K_{f,n}})A^*(\boldsymbol{K_{f,n}} + \boldsymbol{Q_p}) \\ T_{Pos,n}(\boldsymbol{Q_p}) = K_{fx,n}A^*(\boldsymbol{K_{f,n}})A(\boldsymbol{K_{f,n}} - \boldsymbol{Q_p}) \end{cases} \tag{6-22}$$

By applying Equation (6-22) to Equation (6-11) and (6-12), Ψ_s can be solved. In Chapter 5, it was noted that in SSBm-COMx, $\Psi_s(\boldsymbol{Q_p})$ and the Fourier transform of the reconstruction have the following relationship:

$$\begin{aligned} \mathcal{F}\{COM_x^f(\boldsymbol{R_p})\} = \Psi_s(\boldsymbol{Q_p}) \iint K_{fx}[&A^*(\boldsymbol{K_f})A(\boldsymbol{K_f} - \boldsymbol{Q_p}) \\ &- A(\boldsymbol{K_f})A^*(\boldsymbol{K_f} + \boldsymbol{Q_p})]m(\boldsymbol{K_f}, \boldsymbol{Q_p})d\boldsymbol{K_f} \end{aligned} \tag{6-23}$$

For traditional COM methods, $m(\boldsymbol{K_f}, \boldsymbol{Q_p})$ equals 1 at all $\boldsymbol{K_f}$ position.

$$\mathcal{F}\{COM_x(\boldsymbol{R_p})\} = \Psi_s(\boldsymbol{Q_p}) \iint K_{fx}[A^*(\boldsymbol{K_f})A(\boldsymbol{K_f} - \boldsymbol{Q_p}) - A(\boldsymbol{K_f})A^*(\boldsymbol{K_f} + \boldsymbol{Q_p})]d\boldsymbol{K_f} \tag{6-24}$$

Therefore, as for Single-side band ptychography. Centre-of-mass can be seen as an extension of AD-ptycho with a weighted $D_n(\boldsymbol{K_f})$ and a frequency weighting applied to Equation (6-24).

6.8 Conclusion

In this chapter, a ptychography method (AD-ptycho) that utilises an arbitrary multi-channel detector under the weak phase object approximation (WPOA) has been described. This method can be applied to multi-channel detectors (no less than 3 sectors) with various geometries in the diffraction plane. For instance, detector channels can be arranged as traditional segmented DPC detectors or pixelated detectors. It has also been shown that SSB and the traditional centre-of-mass can be considered as extensions of AD-ptycho when specific detector functions and weighting functions are applied.

However, the Ψ_s solved by AD-ptycho without modification in the frequency domain results in a very noisy reconstruction, especially in the high-frequency range. To solve this problem,

Ψ_s is weighted in frequency domain by a factor of $\sum|\Gamma_{T,n}(\boldsymbol{Q_p})|$ to suppress noise, as described in Section 6.4. In Section 6.5, simulated 4DSTEM datasets are reconstructed by AD-ptycho (with and without frequency weighting) and SSB as a comparison.

AD-ptycho as a ptychographic method, can be used for aberration compensation. Its ability to correct aberrations on experimental datasets has been tested. However, aberration compensation in AD-ptycho is also limited by the detector geometry and the estimation of aberrations.

For AD-ptycho, there remains room for improvement. In AD-ptycho, the solution to Ψ_s can be converted to an overdetermined linear system. In this Chapter, the simplest least square method (LSM) is used to solve the linear system. But it has the potential to utilise various mathematical conditions and methods to solve the overdetermined linear system, for example, the iterative CGLS algorithm [5]. Other solving methods with more constraints on the linear system may slow the processing speed but have better tolerance to low-fluence conditions.

However, AD-ptycho still relies on the WPOA condition, which is a theoretical limitation of this method. Due to the WPOA, the results here cannot be compared those from the WDD method. Therefore, in the next chapter, a more general method that does not rely on the WPOA condition is discussed, within the context of analysing a 4DSTEM dataset as a linear system.

Reference

[1] N. Shibata, S. D. Findlay, Y. Kohno, H. Sawada, Y. Kondo, and Y. Ikuhara, "Differential phase-contrast microscopy at atomic resolution," *Nature Physics,* vol. 8, no. 8, pp. 611-615, 2012.

[2] I. Lazić, E. G. T. Bosch, and S. Lazar, "Phase contrast STEM for thin samples: Integrated differential phase contrast," *Ultramicroscopy,* vol. 160, pp. 265-280, 2016.

[3] B. C. McCallum, M. N. Landauer, and J. M. Rodenburg, "Complex image

reconstruction of weak specimens from a three-sector detector in the STEM," *Optik (Stuttgart),* vol. 101, no. 2, pp. 53-62, 1995.

[4] H. Yang, T. J. Pennycook, and P. D. Nellist, "Efficient phase contrast imaging in STEM using a pixelated detector. Part II: Optimisation of imaging conditions," *Ultramicroscopy,* vol. 151, pp. 232-239, 2015.

[5] H. Brown *et al.*, "Structure retrieval with fast electrons using segmented detectors," *Physical Review B,* vol. 93, no. 13, p. 134116, 2016.

[6] J. M. Rodenburg, B. C. McCallum, and P. D. Nellist, "Experimental tests on double-resolution coherent imaging via STEM," *Ultramicroscopy,* vol. 48, no. 3, pp. 304-314, 1993.

[7] K. Ooe, T. Seki, Y. Ikuhara, and N. Shibata, "Ultra-high contrast STEM imaging for segmented/pixelated detectors by maximizing the signal-to-noise ratio," *Ultramicroscopy,* vol. 220, p. 113133, 2021.

[8] H. Yang, P. Ercius, P. D. Nellist, and C. Ophus, "Enhanced phase contrast transfer using ptychography combined with a pre-specimen phase plate in a scanning transmission electron microscope," *Ultramicroscopy,* vol. 171, pp. 117-125, 2016.

[9] C. M. O'Leary, G. T. Martinez, E. Liberti, M. J. Humphry, A. I. Kirkland, and P. D. Nellist, "Contrast transfer and noise considerations in focused-probe electron ptychography," *Ultramicroscopy,* vol. 221, p. 113189, 2021.

[10] H. Yang *et al.*, "Simultaneous atomic-resolution electron ptychography and Z-contrast imaging of light and heavy elements in complex nanostructures," *Nature Communications,* vol. 7, no. 1, p. 12532, 2016.

Chapter 7: Ptychographic Linear System Solution Method without WPOA

7.1 Introduction

In Chapter 6, a framework (AD-ptycho) based on the Weak Phase Object Approximation (WPOA) for analysing 4DSTEM datasets was discussed. In this chapter, a similar approach by solving linear systems is applied, but without the WPOA constraint. A single-layer phase object with electron absorption, meaning the amplitude is not uniformly 1 for the transmission function, is considered. However, no dynamical scattering is considered, as electrons are only scattered with a single-layer transmission function. As an analytical framework without WPOA, this linear system solution (LSS) method is compared to a Wigner Distribution Deconvolution (WDD) [1] in Sections 7.3 and 7.6. Additionally, numerical simulations are conducted, and the reconstructions of simulations are plotted and discussed.

The LSS method can be regarded as an alternative approach to WDD for solving the deconvolution of 4DSTEM datasets. Both methods start from the same equation (Equation 7-2), and use related but different processes to solve for a transmission function that is not constrained by the WPOA, but only by single-layer sample. The algorithm will be discussed in Sections 7.2 and 7.3. The relationship between LSS, WDD, and AD-ptycho will be explored in Sections 7.3, 7.4, and 7.6. Following this, simulations under different sample conditions and parameters are discussed in Section 7.7.

7.2 Linear system of 4DSTEM dataset without WPOA

In this section, the sample transmission function (ψ) is considered as a single slice complex function with an amplitude distribution (absorption, including high angle scattering that is not measured by the detector) and phase variation. Dynamical scattering is ignored in this chapter.

A detector function $D_n(\boldsymbol{K_f})$ was introduced in Chapter 6, where the intensity integrated at nth detector channel and its Fourier transform with respect to scanning coordinate $\boldsymbol{R_p}$ can be written as:

$$\begin{aligned} C_n(\boldsymbol{R_p}) &= \iint D_n(\boldsymbol{K_f}) \cdot \left|M(\boldsymbol{K_f}, \boldsymbol{R_p})\right|^2 \cdot d\boldsymbol{K_f} \\ c_n(\boldsymbol{Q_p}) &= \mathcal{F}_{R_p \to Q_p}\{C_n(\boldsymbol{R_p})\} = \iint D_n(\boldsymbol{K_f}) \cdot \mathrm{G}(\boldsymbol{K_f}, \boldsymbol{Q_p}) \cdot d\boldsymbol{K_f} \end{aligned} \tag{7-1}$$

In Chapter 5 and 6, $\mathrm{G}(\boldsymbol{K_f}, \boldsymbol{Q_p})$ in Equation (5-1) was separated into 3 terms under the WPOA condition. For SSB and SSBm-COM, the double-overlap (DO) areas were used to calculate reconstructions. Here, without the WPOA constraint, the convolution form of $\mathrm{G}(\boldsymbol{K_f}, \boldsymbol{Q_p})$ [1, 2] is considered which can be expressed as:

$$\mathrm{G}(\boldsymbol{K_f}, \boldsymbol{Q_p}) = A(\boldsymbol{K_f})A^*(\boldsymbol{K_f} + \boldsymbol{Q_p}) \otimes_{\boldsymbol{K_f}} \Psi(\boldsymbol{K_f})\Psi^*(\boldsymbol{K_f} - \boldsymbol{Q_p})$$

or: (7-2)

$$\mathrm{G}(\boldsymbol{K_f}, \boldsymbol{Q_p}) = \iint \Psi(\boldsymbol{K'_f})\Psi^*(\boldsymbol{K'_f} - \boldsymbol{Q_p})A(\boldsymbol{K_f} - \boldsymbol{K'_f})A^*(\boldsymbol{K_f} - \boldsymbol{K'_f} + \boldsymbol{Q_p})d\boldsymbol{K'_f}$$

In Equation (7-2), $\mathrm{G}(\boldsymbol{K_f}, \boldsymbol{Q_p})$ can be written as an integral. $\boldsymbol{K'_f}$ is a dummy variable. Theoretically, $\boldsymbol{K'_f}$ has a range of $(-\infty, +\infty)$. Following this, the integral form of Equation (7-2) is combined with Equation (7-1):

$$c_n(\boldsymbol{Q_p}) = \iint D_n(\boldsymbol{K_f}) \cdot \left[\iint \Psi(\boldsymbol{K'_f})\Psi^*(\boldsymbol{K'_f} - \boldsymbol{Q_p})A(\boldsymbol{K_f} - \boldsymbol{K'_f})A^*(\boldsymbol{K_f} - \boldsymbol{K'_f} + \boldsymbol{Q_p})d\boldsymbol{K'_f}\right] \cdot d\boldsymbol{K_f} \tag{7-3}$$

Since $\boldsymbol{K'_f}$ and $\boldsymbol{K_f}$ are independent, the order of the integrals can be switched.

$$\begin{aligned} c_n(\boldsymbol{Q_p}) &= \iint \iint D_n(\boldsymbol{K_f}) \cdot \Psi(\boldsymbol{K'_f})\Psi^*(\boldsymbol{K'_f} - \boldsymbol{Q_p})A(\boldsymbol{K_f} - \boldsymbol{K'_f})A^*(\boldsymbol{K_f} - \boldsymbol{K'_f} + \boldsymbol{Q_p}) \cdot d\boldsymbol{K'_f}d\boldsymbol{K_f} \\ &= \iint \Psi(\boldsymbol{K'_f})\Psi^*(\boldsymbol{K'_f} - \boldsymbol{Q_p}) \cdot \left[\iint D_n(\boldsymbol{K_f}) \cdot A(\boldsymbol{K_f} - \boldsymbol{K'_f})A^*(\boldsymbol{K_f} - \boldsymbol{K'_f} + \boldsymbol{Q_p})d\boldsymbol{K_f}\right] d\boldsymbol{K'_f} \end{aligned} \tag{7-4}$$

In equation (7-4), the unknown variable $\Psi(\boldsymbol{K'_f})$ is separated from $D_n(\boldsymbol{K_f})$ and $A(\boldsymbol{K_f} - \boldsymbol{K'_f})$. Therefore, the integral in the square bracket can be replaced by a new variable $T_n(\boldsymbol{K'_f}, \boldsymbol{Q_p})$. For simplicity, $\Psi(\boldsymbol{K'_f})\Psi^*(\boldsymbol{K'_f} - \boldsymbol{Q_p})$ can also be replaced by a variable $\Phi\,(\boldsymbol{K'_f}, \boldsymbol{Q_p})$.

$$\begin{aligned} T_n(\boldsymbol{K'_f}, \boldsymbol{Q_p}) &= \iint D_n(\boldsymbol{K_f}) \cdot A(\boldsymbol{K_f} - \boldsymbol{K'_f})A^*(\boldsymbol{K_f} - \boldsymbol{K'_f} + \boldsymbol{Q_p})d\boldsymbol{K_f} \\ \Phi\,(\boldsymbol{K'_f}, \boldsymbol{Q_p}) &= \Psi(\boldsymbol{K'_f})\Psi^*(\boldsymbol{K'_f} - \boldsymbol{Q_p}) \end{aligned} \tag{7-5}$$

After replacement, $c_n(\boldsymbol{Q_p})$ can be expressed in a simplified form as:

$$c_n(\boldsymbol{Q_p}) = \iint \Phi\,(\boldsymbol{K'_f}, \boldsymbol{Q_p})T_n(\boldsymbol{K'_f}, \boldsymbol{Q_p})d\boldsymbol{K'_f} \tag{7-6}$$

In Equation (7-6), $\boldsymbol{K'_f}$ is integrated over the interval $K'_{fx}, K'_{fy} \in (-\infty, +\infty)$. However, $T_n(\boldsymbol{K'_f}, \boldsymbol{Q_p})$ will be 0 when $|\boldsymbol{K'_f}| \to \infty$. This is because $D_n(\boldsymbol{K_f})$ is limited by the physical size of the detector, and $\boldsymbol{K'_f}$ controls a shift of the term $A(\boldsymbol{K_f} - \boldsymbol{K'_f})A^*(\boldsymbol{K_f} - \boldsymbol{K'_f} + \boldsymbol{Q_p})$. When $|\boldsymbol{K'_f}|$ is significantly larger than the size of detector, $A(\boldsymbol{K_f} - \boldsymbol{K'_f})A^*(\boldsymbol{K_f} - \boldsymbol{K'_f} + \boldsymbol{Q_p})$ is shifted out of the detector area (D_n), resulting in $T_n(\boldsymbol{K'_f}, \boldsymbol{Q_p}) = 0$. Therefore, the range of $\boldsymbol{K'_f}$ that contributes to the integrated $c_n(\boldsymbol{Q_p})$ in Equation (7-6) is limited. This limitation will

also be discussed in Section 7.5.

The integral is over a limited area of $\boldsymbol{K_f'}$, where $\boldsymbol{K_f'}$ ensures that there exists at least one n such that $T_n(\boldsymbol{K_f'}, \boldsymbol{Q_p}) \neq 0$. If the $\boldsymbol{K_f'}$ area is sampled discretely, the integral can be approximated by a summation with a finite number of terms, as shown in Equation (7-7), which can be more easily programmed.

$$c_n(\boldsymbol{Q_p}) = \sum_{\boldsymbol{K_{f'}}} \Phi\left(\boldsymbol{K_f'}, \boldsymbol{Q_p}\right) T_n(\boldsymbol{K_f'}, \boldsymbol{Q_p}) \tag{7-7}$$

Theoretically, finer sampling conditions will make Equation (7-7) approximate the integral form, Equation (7-6). However, finer sampling of $\boldsymbol{K_f'}$ increases the computational load and extends the solving time for Equation (7-7). So, in the following simulation sections, $\boldsymbol{K_f'}$ sampling is determined based on the sampling of $\boldsymbol{Q_p}$ and $\boldsymbol{K_f}$.

In equation (7-7), $c_n(\boldsymbol{Q_p})$ is derived from all detector measurements while $T_n(\boldsymbol{K_f'}, \boldsymbol{Q_p})$ is determined from known conditions, such as aperture and detector settings. The only unknown variable is $\Phi\left(\boldsymbol{K_f'}, \boldsymbol{Q_p}\right)$. Therefore, Equation (7-7) can be expressed as a linear system, as in Equation (7-8). The terms $c_1(\boldsymbol{Q_p}) \cdots c_N(\boldsymbol{Q_p})$ can be considered as a vector with N elements. To distinguish different $\boldsymbol{K_f'}$ positions in the summation, a subscript after $\boldsymbol{K_f'}$ is used, such as $K_{f1}', K_{f2}', \cdots, K_{fM}'$, where M is the number of $\boldsymbol{K_f'}$ sampling positions in Equation (7-7). The linear system based on Equation (7-7) can be written as follows:

$$\underbrace{\begin{bmatrix} T_1(\boldsymbol{K_{f1}'}, \boldsymbol{Q_p}) & \cdots & T_1(\boldsymbol{K_{fM}'}, \boldsymbol{Q_p}) \\ \vdots & \ddots & \vdots \\ T_N(\boldsymbol{K_{f1}'}, \boldsymbol{Q_p}) & \cdots & T_N(\boldsymbol{K_{fM}'}, \boldsymbol{Q_p}) \end{bmatrix}}_{Matrix: T_m\ (N \times M)} \times \underbrace{\begin{bmatrix} \Phi\left(\boldsymbol{K_{f1}'}, \boldsymbol{Q_p}\right) \\ \vdots \\ \Phi\left(\boldsymbol{K_{fM}'}, \boldsymbol{Q_p}\right) \end{bmatrix}}_{Vector: \Phi_v\ (M \times 1)} = \underbrace{\begin{bmatrix} c_1(\boldsymbol{Q_p}) \\ \vdots \\ c_N(\boldsymbol{Q_p}) \end{bmatrix}}_{Vector: c_v\ (N \times 1)} \tag{7-8}$$

or:

$$T_m \Phi_v = c_v$$

In Equation (7-8), N represents the number of detector channels. T_m, Φ_v and c_v in Equation (7-8) are complex matrices or vectors. Their elements, $T_n(\boldsymbol{K_f'}, \boldsymbol{Q_p})$, $\Phi(\boldsymbol{K_f'}, \boldsymbol{Q_p})$ and $c_n(\boldsymbol{Q_p})$ can be represented by real matrices and vectors as given in Equation (7-9):

$$T_n(\boldsymbol{K'_f}, \boldsymbol{Q_p}) = T_{n,real}(\boldsymbol{K'_f}, \boldsymbol{Q_p}) + iT_{n,imag}(\boldsymbol{K'_f}, \boldsymbol{Q_p})$$
$$\Phi(\boldsymbol{K'_f}, \boldsymbol{Q_p}) = \Phi_{real}(\boldsymbol{K'_f}, \boldsymbol{Q_p}) + i\Phi_{imag}(\boldsymbol{K'_f}, \boldsymbol{Q_p}) \quad (7\text{-}9)$$
$$c_n(\boldsymbol{Q_p}) = c_{n,real}(\boldsymbol{Q_p}) + c_{n,imag}(\boldsymbol{Q_p})$$

In Equation (7-9), the terms on the right-hand side with subscripts "$real$" and "$imag$" represent the real and imaginary parts of $T_n(\boldsymbol{K'_f}, \boldsymbol{Q_p})$, $\Phi(\boldsymbol{K'_f}, \boldsymbol{Q_p})$ and $c_n(\boldsymbol{Q_p})$. By combining Equations (7-9), the complex linear system in Equation (7-8) can be converted into a real linear system, as given in Equation (7-10):

$$\underbrace{\begin{bmatrix} T_{1,real}(\boldsymbol{K'_{f1}}, \boldsymbol{Q_p}) & \cdots & T_{1,real}(\boldsymbol{K'_{fM}}, \boldsymbol{Q_p}) & -T_{1,imag}(\boldsymbol{K'_{f1}}, \boldsymbol{Q_p}) & \cdots & -T_{1,imag}(\boldsymbol{K'_{fM}}, \boldsymbol{Q_p}) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ T_{N,real}(\boldsymbol{K'_{f1}}, \boldsymbol{Q_p}) & \cdots & T_{N,real}(\boldsymbol{K'_{fM}}, \boldsymbol{Q_p}) & -T_{N,imag}(\boldsymbol{K'_{f1}}, \boldsymbol{Q_p}) & \cdots & -T_{N,imag}(\boldsymbol{K'_{fM}}, \boldsymbol{Q_p}) \\ T_{1,imag}(\boldsymbol{K'_{f1}}, \boldsymbol{Q_p}) & \cdots & T_{1,imag}(\boldsymbol{K'_{fM}}, \boldsymbol{Q_p}) & T_{1,real}(\boldsymbol{K'_{f1}}, \boldsymbol{Q_p}) & \cdots & T_{1,real}(\boldsymbol{K'_{fM}}, \boldsymbol{Q_p}) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ T_{N,imag}(\boldsymbol{K'_{f1}}, \boldsymbol{Q_p}) & \cdots & T_{N,imag}(\boldsymbol{K'_{fM}}, \boldsymbol{Q_p}) & T_{N,real}(\boldsymbol{K'_{f1}}, \boldsymbol{Q_p}) & \cdots & T_{N,real}(\boldsymbol{K'_{fM}}, \boldsymbol{Q_p}) \end{bmatrix}}_{Matrix:\, T_{mr}\ (2N\times 2M)}$$
$$\times \underbrace{\begin{bmatrix} \Phi_{real}(\boldsymbol{K'_{f1}}, \boldsymbol{Q_p}) \\ \vdots \\ \Phi_{real}(\boldsymbol{K'_{fM}}, \boldsymbol{Q_p}) \\ \Phi_{imag}(\boldsymbol{K'_{f1}}, \boldsymbol{Q_p}) \\ \vdots \\ \Phi_{imag}(\boldsymbol{K'_{fM}}, \boldsymbol{Q_p}) \end{bmatrix}}_{Vector:\, \Phi_{vr}\ (2M\times 1)} = \underbrace{\begin{bmatrix} c_{1,real}(\boldsymbol{Q_p}) \\ \vdots \\ c_{N,real}(\boldsymbol{Q_p}) \\ c_{1,imag}(\boldsymbol{Q_p}) \\ \vdots \\ c_{N,imag}(\boldsymbol{Q_p}) \end{bmatrix}}_{Vector:c_{vr}\ (2N\times 1)} \quad (7\text{-}10)$$

or:

$$T_{mr}\Phi_{vr} = c_{vr}$$

The solution to Equation (7-10), Φ_{vr}, can be used to construct the complex $\Phi(\boldsymbol{K'_f}, \boldsymbol{Q_p})$, which is equivalent to solving Equation (7-8).

As noted, the number of rows and columns is determined by the number of detector channel (N) and the number of $\boldsymbol{K'_f}$ sampling positions (M). Ideally, if $N \geq M$, Equation (7-8) is solvable. However, at some $\boldsymbol{Q_p}$ positions, Equation (7-10) may be underdetermined. To avoid this, the selection of $\boldsymbol{K'_f}$ sampling positions in Section 7.7 and relation between $\boldsymbol{Q_p}$ and $T_n(\boldsymbol{K'_f}, \boldsymbol{Q_p})$ will be further discussed in Section 7.5. Once Equation (7-10) forms

an overdetermined linear system, various methods can be used to solve Φ_{vr}. In this Chapter, for simplicity, the least square method (LSM) as applied in AD-ptycho is used. For the LSM, the linear system in Equation (7-10) must satisfy the condition to calculate the inverse matrix $(T_{mr}^T T_{mr})^{-1}$. Otherwise, $\Phi\left(\boldsymbol{K}_{\boldsymbol{f}}', \boldsymbol{Q}_{\boldsymbol{p}}\right)$ may not be solvable at some $(\boldsymbol{K}_{\boldsymbol{f}}', \boldsymbol{Q}_{\boldsymbol{p}})$ positions. Once this condition is satisfied, the Φ_{vr} can be solved as in Equation (7-11):

$$\Phi_{vr} = (T_{mr}^T T_{mr})^{-1} T_{mr}^T c_{vr} \tag{7-11}$$

For each $\boldsymbol{Q}_{\boldsymbol{p}}$, Φ_{vr} can then be calculate and $\Phi\left(\boldsymbol{K}_{\boldsymbol{f}}', \boldsymbol{Q}_{\boldsymbol{p}}\right)$ can be constructed from $\Phi_{real}\left(\boldsymbol{K}_{\boldsymbol{f}}', \boldsymbol{Q}_{\boldsymbol{p}}\right)$ and $\Phi_{imag}\left(\boldsymbol{K}_{\boldsymbol{f}}', \boldsymbol{Q}_{\boldsymbol{p}}\right)$, where $\Phi\left(\boldsymbol{K}_{\boldsymbol{f}}', \boldsymbol{Q}_{\boldsymbol{p}}\right)$ is the 4D matrix in Equation (7-5) represented as $\Phi\left(\boldsymbol{K}_{\boldsymbol{f}}', \boldsymbol{Q}_{\boldsymbol{p}}\right) = \Psi\left(\boldsymbol{K}_{\boldsymbol{f}}'\right)\Psi^*\left(\boldsymbol{K}_{\boldsymbol{f}}' - \boldsymbol{Q}_{\boldsymbol{p}}\right)$. The final step is to separate $\Psi\left(\boldsymbol{K}_{\boldsymbol{f}}'\right)$ and $\Psi^*\left(\boldsymbol{K}_{\boldsymbol{f}}' - \boldsymbol{Q}_{\boldsymbol{p}}\right)$ from $\Phi\left(\boldsymbol{K}_{\boldsymbol{f}}', \boldsymbol{Q}_{\boldsymbol{p}}\right)$, as discussed in the next section.

From Equation (7-1) to (7-11), the $\Phi\left(\boldsymbol{K}_{\boldsymbol{f}}', \boldsymbol{Q}_{\boldsymbol{p}}\right)$ related to sample information is calculated by solving a linear system. Therefore, this method is referred to as the “linear system solution (LSS)” for ptychography.

7.3 Solving the Transmission function

In the previous section, $\Phi\left(\boldsymbol{K}_{\boldsymbol{f}}', \boldsymbol{Q}_{\boldsymbol{p}}\right)$ was defined and solved using Equation (7-10). However, $\Phi\left(\boldsymbol{K}_{\boldsymbol{f}}', \boldsymbol{Q}_{\boldsymbol{p}}\right)$, as defined in Equation (7-5), does not directly provide information about the sample. To obtain the transmission function ψ, additional processing is required, which will be discussed in this section.

Before considering how Ψ can be calculated from Φ, it is important to discuss the connection between Equation (7-8) and the WDD method, which was introduced in Chapter 3. Equations (7-8) and (7-10) can be seen as an alternative approach to WDD for calculating the deconvolution from the same dataset. Consequently, both the WDD method and Equation (7-10) yield solutions in a very similar form:

$$\begin{aligned} &Linear\ system\ solution\colon \Phi\left(\boldsymbol{K}_{\boldsymbol{f}}', \boldsymbol{Q}_{\boldsymbol{p}}\right) = \Psi\left(\boldsymbol{K}_{\boldsymbol{f}}'\right)\Psi^*\left(\boldsymbol{K}_{\boldsymbol{f}}' - \boldsymbol{Q}_{\boldsymbol{p}}\right) \\ &WDD\ solution\colon \Phi_{WDD}\left(\boldsymbol{K}_{\boldsymbol{f}}, \boldsymbol{Q}_{\boldsymbol{p}}\right) = \Psi\left(\boldsymbol{K}_{\boldsymbol{f}}\right)\Psi^*\left(\boldsymbol{K}_{\boldsymbol{f}} - \boldsymbol{Q}_{\boldsymbol{p}}\right) \end{aligned} \tag{7-12}$$

In this section, for comparison, a subscript "WDD" is added to the WDD result $\Phi_{WDD}\left(\boldsymbol{K_f}, \boldsymbol{Q_p}\right)$ as in Equation (7-12).

Due the similarity between $\Phi\left(\boldsymbol{K'_f}, \boldsymbol{Q_p}\right)$ and $\Phi_{WDD}\left(\boldsymbol{K_f}, \boldsymbol{Q_p}\right)$, the transformation of $\Phi_{WDD}\left(\boldsymbol{K_f}, \boldsymbol{Q_p}\right)$ to $\Psi\left(\boldsymbol{Q_p}\right)$ can naturally be transferred to the LSS method. Similar to the WDD, consider a slice at $\boldsymbol{K'_f} = \boldsymbol{0}$ in the 4D array $\Phi\left(\boldsymbol{K'_f}, \boldsymbol{Q_p}\right)$, then, an approximate solution for $\Psi_{WDD}(-\boldsymbol{Q_p})$ is [3]:

$$\Psi_{WDD}\left(-\boldsymbol{Q_p}\right) = \frac{\Phi^*\left(\boldsymbol{0}, \boldsymbol{Q_p}\right)}{\sqrt{\Phi(\boldsymbol{0},\boldsymbol{0})}} = \frac{\Psi^*(\boldsymbol{0})}{|\Psi(\boldsymbol{0},\boldsymbol{0})|} \cdot \Psi(-\boldsymbol{Q_p}) \tag{7-13}$$

Equation (7-13) can be used to calculate an approximate solution $\Psi_{WDD}(-\boldsymbol{Q_p})$ using the linear system solution method (LSS). Equation (7-13) also limited the resolution of LSS, which will be discussed in Section 7.6.

As noted earlier, the LSS method can be considered as an alternative approach to solving the deconvolution of Equation (7-2). Unlike the WDD method, LSS does not require a super parameter ϵ. However, there is a possibility that linear system $T_m\Phi_v = c_v$ and $T_{mr}\Phi_{vr} = c_{vr}$ in Equation (7-8) and (7-10) may have no solution at certain $\boldsymbol{Q_p}$ positions. These situations will be discussed in Section 7.5.

Having solved $\Psi_{WDD}(-\boldsymbol{Q_p})$, an inverse Fourier transform can be performed to calculate sample transmission function $\psi(\boldsymbol{R_p})$, which is the final reconstruction result.

7.4 WPOA on LSS and Connection to AD-ptycho

In previous sections, solving the linear system in Equation (7-10) for $\Phi\left(\boldsymbol{K'_f}, \boldsymbol{Q_p}\right)$ and calculating Ψ_{WDD} from $\Phi\left(\boldsymbol{K'_f}, \boldsymbol{Q_p}\right)$ are discussed. In this section, the connection between LSS and WPOA condition will be discussed by comparing LSS and AD-ptycho method described in Chapter 6.

WPOA can be applied to LSS by considering extra constraints on the Fourier transform of transmission function Ψ. If the sample is a weak-phase object, as introduced in previous

chapters, a WPOA transmission function can be written as:

$$\psi(\boldsymbol{r}) \xrightarrow{POA} exp[\psi_s(\boldsymbol{r})] \xrightarrow{WPOA} 1 + \psi_s(\boldsymbol{r}) \tag{7-14}$$

While ψ_s is a complex function with an all-zero real part. The Fourier transform of equation (7-14) is:

$$\Psi(\boldsymbol{R}) \xrightarrow{WPOA} \mathcal{F}\{1 + \psi_s(\boldsymbol{r})\} = \delta(\boldsymbol{R}) + \Psi_s(\boldsymbol{R}) \tag{7-15}$$

Equation (7-15) expresses Ψ under the WPOA condition, with δ as the Dirac function. Furthermore, Φ in Equation (7-5) an be replaced by this WPOA formation, noted as Φ_{WPOA}.

$$\begin{aligned}\Phi_{WPOA}\left(\boldsymbol{K_f'}, \boldsymbol{Q_p}\right) &= \left[\delta\left(\boldsymbol{K_f'}\right) + \Psi_s\left(\boldsymbol{K_f'}\right)\right] \cdot \left[\delta\left(\boldsymbol{K_f'} - \boldsymbol{Q_p}\right) - \Psi_s\left(\boldsymbol{K_f'} - \boldsymbol{Q_p}\right)\right] \\ &= \delta\left(\boldsymbol{K_f'}\right)\delta\left(\boldsymbol{K_f'} - \boldsymbol{Q_p}\right) - \delta\left(\boldsymbol{K_f'}\right)\Psi_s\left(\boldsymbol{K_f'} - \boldsymbol{Q_p}\right) \\ &+\delta\left(\boldsymbol{K_f'} - \boldsymbol{Q_p}\right)\Psi_s\left(\boldsymbol{K_f'}\right) - \Psi_s\left(\boldsymbol{K_f'}\right)\Psi_s\left(\boldsymbol{K_f'} - \boldsymbol{Q_p}\right)\end{aligned} \tag{7-16}$$

In Equation (7-16), $\Phi_{\mathrm{WPOA}}\left(\boldsymbol{K_f'}, \boldsymbol{Q_p}\right)$ has been separated into 4 terms. When $\boldsymbol{Q_p} \neq \boldsymbol{0}$, the first term $\delta\left(\boldsymbol{K_f'}\right)\delta\left(\boldsymbol{K_f'} - \boldsymbol{Q_p}\right)$ is zero. Hence, this term is ignored. When $\boldsymbol{K_f'} \neq \boldsymbol{0}\ \boldsymbol{or}\ \boldsymbol{Q_p}$, the only non-zero term is the last one, $\Psi_s\left(\boldsymbol{K_f'}\right)\Psi_s\left(\boldsymbol{K_f'} - \boldsymbol{Q_p}\right)$. However, under the WPOA, where the phase information is assumed to be weak, Ψ_s can be considered as a small quantity. This means $\Psi_s\left(\boldsymbol{K_f'}\right)\Psi_s\left(\boldsymbol{K_f'} - \boldsymbol{Q_p}\right)$ has a limited contribution to the integral in equation (7-4) with respect to $\boldsymbol{K_f'}$. When $\boldsymbol{K_f'} = \boldsymbol{0}$, $\delta\left(\boldsymbol{K_f'}\right)\Psi_s\left(\boldsymbol{K_f'} - \boldsymbol{Q_p}\right)$ has a more significant influence on the integral, as does $\delta\left(\boldsymbol{K_f'} - \boldsymbol{Q_p}\right)\Psi_s\left(\boldsymbol{K_f'}\right)$ when $\boldsymbol{K_f'} = \boldsymbol{Q_p}$. Therefore, under the WPOA, the integral with respect to $\boldsymbol{K_f'}$ in Equation (7-4) can be approximated as the sum of the terms for $\boldsymbol{K_f'} = \boldsymbol{0}$ and $\boldsymbol{K_f'} = \boldsymbol{Q_p}$, as given by equation (7-17):

$$\Phi_{WPOA}\left(\boldsymbol{K_f'}, \boldsymbol{Q_p}\right) \approx \delta\left(\boldsymbol{K_f'} - \boldsymbol{Q_p}\right)\Psi_s\left(\boldsymbol{K_f'}\right) + \delta\left(\boldsymbol{K_f'}\right)\Psi_s^*\left(\boldsymbol{K_f'} - \boldsymbol{Q_p}\right) \tag{7-17}$$

At the same time, if only $\boldsymbol{K_f'} = \boldsymbol{0}$ and $\boldsymbol{K_f'} = \boldsymbol{Q_p}$ are considered, T_n can also be simplified. The simplified T_n is equivalent to $T_{Neg,n}\left(\boldsymbol{Q_p}\right)$ and $T_{Pos,n}\left(\boldsymbol{Q_p}\right)$ in Equation (6-5) for the

AD-ptycho algorithm, as shown in equation (7-18), since LSS and AD-ptycho share the same detector function $D_n(\boldsymbol{K_f})$.

$$\begin{cases} T_{Neg,n}(\boldsymbol{Q_p}) = T_n(\boldsymbol{0}, \boldsymbol{Q_p}) = \iint D_n(\boldsymbol{K_f}) \cdot A(\boldsymbol{K_f}) A^*(\boldsymbol{K_f} + \boldsymbol{Q_p}) \cdot d\boldsymbol{K_f} \\ T_{Pos,n}(\boldsymbol{Q_p}) = T_n(\boldsymbol{Q_p}, \boldsymbol{Q_p}) = \iint D_n(\boldsymbol{K_f}) \cdot A^*(\boldsymbol{K_f}) A(\boldsymbol{K_f} - \boldsymbol{Q_p}) \cdot d\boldsymbol{K_f} \end{cases} \tag{7-18}$$

Using the substitution in Equation (7-17), Equation (7-6) can be written as follows:

$$\begin{aligned} c_n(\boldsymbol{Q_p}) &= \iint \Phi_{\mathrm{WPOA}}\left(\boldsymbol{K_f'}, \boldsymbol{Q_p}\right) T_n(\boldsymbol{K_f'}, \boldsymbol{Q_p}) d\boldsymbol{K_f'} \\ &\approx \Psi_s^*(-\boldsymbol{Q_p}) T_n(\boldsymbol{0}, \boldsymbol{Q_p}) + \Psi_s(\boldsymbol{Q_p}) T_n(\boldsymbol{Q_p}, \boldsymbol{Q_p}) \end{aligned} \tag{7-19}$$

Equation (7-19) is the same as Equation (6-7) in the AD-ptycho algorithm. As noted, ψ_s is a complex function with only an imaginary part, so its Fourier transform Ψ_s is conjugate antisymmetric. This means $\Psi_s(\boldsymbol{Q_p}) + \Psi_s^*(-\boldsymbol{Q_p})$ are 0. This condition is considered in the AD-ptycho method to further simplify Equation (6-7). However, in the LSS method, as two different $\boldsymbol{K_f'}$ positions, $\Psi_s(\boldsymbol{Q_p})$ and $\Psi_s^*(-\boldsymbol{Q_p})$ are specified as two different terms in Φ_v for two different $\boldsymbol{K_f'}$ positions this gives a simplified Equation (7-8) with $M = 2$:

$$\underbrace{\begin{bmatrix} T_1(\boldsymbol{0}, \boldsymbol{Q_p}) & T_1(\boldsymbol{Q_p}, \boldsymbol{Q_p}) \\ \vdots & \vdots \\ T_N(\boldsymbol{0}, \boldsymbol{Q_p}) & T_N(\boldsymbol{Q_p}, \boldsymbol{Q_p}) \end{bmatrix}}_{Matrix: T_m\ (N \times 2)} \times \underbrace{\begin{bmatrix} \Psi_s^*(-\boldsymbol{Q_p}) \\ \Psi_s(\boldsymbol{Q_p}) \end{bmatrix}}_{Vector: \Phi_v\ (2 \times 1)} = \underbrace{\begin{bmatrix} c_1(\boldsymbol{Q_p}) \\ \vdots \\ c_N(\boldsymbol{Q_p}) \end{bmatrix}}_{Vector: c_v\ (N \times 1)} \tag{7-20}$$

The T_m in Equation (7-20) corresponds to selecting only $\boldsymbol{K_f'} = \boldsymbol{0}$ and $\boldsymbol{K_f'} = \boldsymbol{Q_p}$ in Equation (7-8). This indicates that applying the WPOA is related to a specific $\boldsymbol{K_f'}$ selection in the LSS. However, when solving Equation (7-20), the conjugate antisymmetry of $\Psi_s(\boldsymbol{Q_p})$ is not considered (Because in the LSS, WPOA is not considered as a predefined condition). As a result, the reconstruction of $\Psi_s(\boldsymbol{Q_p})$ and $-\Psi_s^*(-\boldsymbol{Q_p})$ may differ due to noise or misalignment in the dataset.

Furthermore, for a WPOA sample, if the noise and misalignments can be ignored (for example in simulation conditions which will be discussed in Section 7.7.3), considering

only $\boldsymbol{K_f'} = \mathbf{0}$ can results in identical reconstruction results to Equation (7-20), which considers $\boldsymbol{K_f'} = \mathbf{0}$ and $\boldsymbol{K_f'} = \boldsymbol{Q_p}$. This equivalence arises from the conjugate antisymmetry of $\Psi_s(\boldsymbol{Q_p})$.

In summary, under the WPOA condition and further approximations made in Equation (7-17), LSS and AD-ptycho give similar expressions for $c_n(\boldsymbol{Q_p})$. However, AD-ptycho applies additional simplification to Equation (7-20), resulting in differences between the AD-ptycho and the LSS reconstruction for specific $\boldsymbol{K_f'}$ selections of $\boldsymbol{K_f'} = \mathbf{0}\ \boldsymbol{and}\ \boldsymbol{Q_p}$. Simulations to illustrate this are discussed in Section 7.7.6.

7.5 Visualization of Matrix Elements (T_n)

In the previous section, how the selection of $\boldsymbol{K_f'}$ in Equation (7-7) relate to the WPOA condition was discussed. To better understand the significance of these $\boldsymbol{K_f'}$ positions and their interrelations, this section will focus on visualizing T_n.

For simplicity, cases where $\boldsymbol{K_f'} = \mathbf{0}\ or\ \boldsymbol{Q_p}$ are discussed first. As described in the previous section, $T_n(\boldsymbol{K_f'}, \boldsymbol{Q_p})$ with $\boldsymbol{K_f'} = \mathbf{0}\ or\ \boldsymbol{Q_p}$ can be expressed by equation (7-18). These two scenarios are illustrated in Figure 7-1.

In Figure 7-1, $T_n(\boldsymbol{K_f'}, \boldsymbol{Q_p})$ is plotted at the $\boldsymbol{K_f}$ plane, which is divided into multiple detector channels $D_1 \cdots D_{12}$. The aperture function $A(\boldsymbol{K_f})$ or shifted $A(\boldsymbol{K_f} - \boldsymbol{Q_p})$ is represented by the yellow circle. The conjugate of aperture function $A^*(\boldsymbol{K_f})$ or shifted $A^*(\boldsymbol{K_f} + \boldsymbol{Q_p})$ is represented by the blue circle. The overlapping areas, where $A(\boldsymbol{K_f})A^*(\boldsymbol{K_f} + \boldsymbol{Q_p})$ and $A^*(\boldsymbol{K_f})A(\boldsymbol{K_f} - \boldsymbol{Q_p})$ are non-zero, are summed on detector channel D_n, resulting in $T_n(\boldsymbol{K_f'}, \boldsymbol{Q_p})$ for each $\boldsymbol{Q_p}$. In Figure 7-1, these overlapping areas are the same as those in SSB ptychography and SSBm-COM (see Chapter 5).

Except for $\boldsymbol{K_f'} = \mathbf{0}$ and $\boldsymbol{K_f'} = \boldsymbol{Q_p}$, an arbitrary $\boldsymbol{K_f'}$ results in an overall shift of T_n at the $\boldsymbol{K_f}$ plane, as illustrated in Figure 7-2.

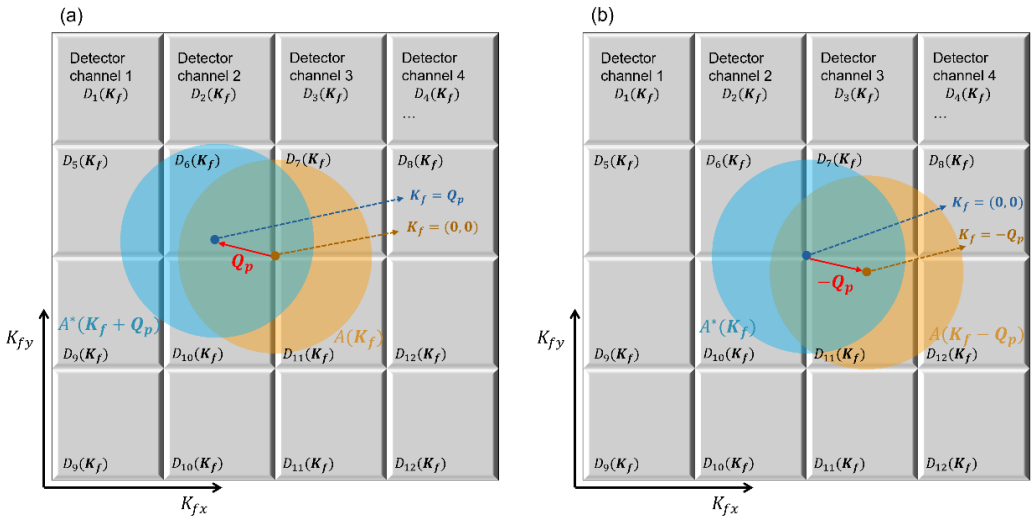

Figure 7-1 | $T_n(\boldsymbol{K'_f}, \boldsymbol{Q_p})$ visualized for (a) $\boldsymbol{K'_f} = \boldsymbol{0}$ and (b) $\boldsymbol{K'_f} = \boldsymbol{Q_p}$. Overlap areas of blue and yellow circle are none-zero area for $T_n(\boldsymbol{0}, \boldsymbol{Q_p})$ and $T_n(\boldsymbol{Q_p}, \boldsymbol{Q_p})$. the $\boldsymbol{K_f}$ plane is divided into multiple detector channels $D_1 \cdots D_{12}$. A yellow circle represents the aperture function $A(\boldsymbol{K_f})$ or its shifted version $A(\boldsymbol{K_f} - \boldsymbol{Q_p})$. A blue circle represents the conjugate of the aperture function $A^*(\boldsymbol{K_f})$ or its shifted $A^*(\boldsymbol{K_f} + \boldsymbol{Q_p})$. In the overlapping areas of $A(\boldsymbol{K_f})$ and $A^*(\boldsymbol{K_f} + \boldsymbol{Q_p})$ as well as $A^*(\boldsymbol{K_f})$ and $A(\boldsymbol{K_f} - \boldsymbol{Q_p})$, T_n is non-zero. T_n values are summed on corresponding detector channel D_n.

As shown in Figure 7-2, the blue circle $A^*(\boldsymbol{K_f} - \boldsymbol{K'_f} + \boldsymbol{Q_p})$ and yellow circle $A(\boldsymbol{K_f} - \boldsymbol{K'_f})$ are shifted together. A light-yellow dashed line circle indicates the original aperture function $A(\boldsymbol{K_f})$ when $\boldsymbol{K'_f} = \boldsymbol{0}$. It is worth noting that part of the overlap area of the yellow and blue circles extends beyond the light-yellow dashed circle. This suggests that, under these conditions (when $\boldsymbol{K'_f} \neq \boldsymbol{0}\ \boldsymbol{or}\ \boldsymbol{Q_p}$), information transfer may exceed the 2α limit. This resolution limit will be discussed in the following section and compared with the Wigner distribution deconvolution (WDD) method.

As Noted in Section 7.2, Figure 7-2 also shows that if $\boldsymbol{K'_f}$ continues to increase, it will eventually shift the overlap area outside the detector array. Consequently, $T_n(\boldsymbol{K'_f}, \boldsymbol{Q_p})$ will be zero if $|\boldsymbol{K'_f}|$ is sufficiently large.

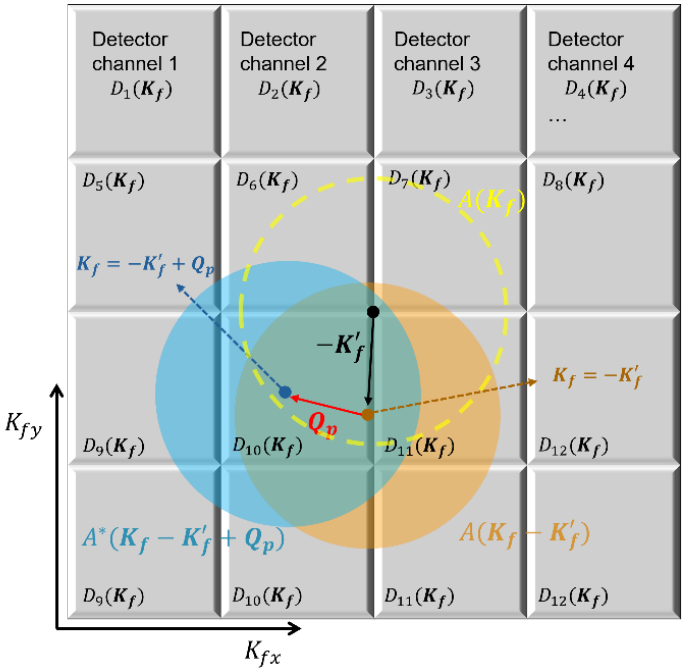

Figure 7-2 | $T_n(\boldsymbol{K'_f}, \boldsymbol{Q_p})$ visualized for $\boldsymbol{K'_f \neq 0}$ ***or*** $\boldsymbol{Q_p}$. $A^*(\boldsymbol{K_f - K'_f + Q_p})$ (blue circle) and $A(\boldsymbol{K_f - K'_f})$ (yellow circle) are shifted together to the direction of $\boldsymbol{K'_f}$. The original aperture function $A(\boldsymbol{K_f})$ (when $\boldsymbol{K'_f \neq 0}$) is represented by a light-yellow dashed line circle indicates. The position of overlapping area is not limited by $A(\boldsymbol{K_f})$ (when $\boldsymbol{K'_f \neq 0}$), indicating information beyond 2α limit.

7.6 Resolution

As described in Chapter 3, the traditional WDD, using Equation (7-13) to calculate Ψ_{WDD}, gives a resolution limit at a spatial frequency corresponding to 2α. This resolution is constrained by the highest $\boldsymbol{Q_p}$ that generates an overlap area between $A(\boldsymbol{K_f})$ and $A(\boldsymbol{K_f - Q_p})$. LSS, if only $\boldsymbol{K'_f = 0}$ is considered when solving Ψ and the transmission function ψ, has the same resolution limit. This limitation is also explained in Figure 7-2. When $\boldsymbol{K'_f = 0}$, the yellow circle $A(\boldsymbol{K_f - K'_f})$ is fixed at the centre of $\boldsymbol{K_f}$ plane, as indicated by the light-yellow dashed circle $A(\boldsymbol{K_f})$. As $\boldsymbol{Q_p}$ changes, the blue circle $A^*(\boldsymbol{K_f - K'_f + Q_p})$ moves and the overlap area between the yellow circle and the blue circle changes with $\boldsymbol{Q_p}$. When $\boldsymbol{Q_p} > \frac{2\alpha}{\lambda}$, an overlap area no longer exists at the $\boldsymbol{K_f}$ plane,

so $T_n(\mathbf{0}, \boldsymbol{Q_p})$ will be an all-zero column in the T_m matrix. As a result, $\Phi(\mathbf{0}, \boldsymbol{Q_p})$ will be unsolvable and Equation (7-13) does not hold. In other words, there is no information transfer when $\boldsymbol{Q_p} > \frac{2\alpha}{\lambda}$ if only $\boldsymbol{K'_f} = \mathbf{0}$ is considered when solving for Ψ.

Experimentally, a resolution up to $\frac{\lambda}{2\alpha}$ is usually sufficient for typical atomic resolution STEM or 4DSTEM. Many 4DSTEM analytical methods, such as SSB ptychography, Centre-of-Mass (COM) or Differential Phase Contrast (DPC), and SSBm-COM (Chapter 5) share this resolution limit due to the same mathematical structure of $A(\boldsymbol{K_f})A^*(\boldsymbol{K_f} + \boldsymbol{Q_p})$. However, the WDD method can theoretically achieve a resolution beyond $\frac{\lambda}{2\alpha}$. The strategy of "stepping out" to extent information transfer was proposed by Rodenburg and Bates [1], which is a similar concept as in ePIE to break the 2α resolution limit [4, 5]. Theoretically, the same "stepping out" process and noise suppression [6] can be applied to the LSS to achieve super resolution.

7.7 Simulations

In this section, comparative simulations of the LSS method with other methods such as SSB and WDD are discussed, under different sample conditions and parameter settings. An algorithm prototype has been implemented in Python using the PyTorch framework to accelerate linear algebra calculations on the GPU. The detail of the workstation used for simulation and comparison of calculation time with different LSS settings will be discussed in Section 7.7.4.

The selection of $\boldsymbol{K'_f}$ is a key factor in LSS reconstruction. In Sections 7.7.1 to 7.7.3 and 7.7.5, the influence of different $\boldsymbol{K'_f}$ conditions on the reconstruction results will be tested. Here, different $\boldsymbol{K'_f}$ conditions are achieved by selecting a subset of $\boldsymbol{K'_f}$ values rather than all $\boldsymbol{K'_f}$ spots for the summation limit in Equation (7-7). This is equivalent to removing some rows in the linear system of Equations (7-8) and (7-10). $\boldsymbol{K'_f}$ selection is also related to reconstructed phase scale, which will be discussed in Section 7.7.2 and 7.7.3.

Furthermore, for sensitive samples, aberration correction and the ability to adapt to non-

pixelated detectors are beneficial. Aberration correction during reconstruction allows for errors in focusing when collecting data, which can reduce the time the sample is illuminated during focusing and improve fluence efficiency. Segmented detectors are generally faster than pixelated 4DSTEM detectors, allowing for shorter beam times to save fluence. To test the potential of LSS on sensitive samples, simulations of LSS with a defocused probe and segmented detectors are described in Sections 7.7.6 and 7.7.7.

7.7.1 $\boldsymbol{K'_f = 0}$ and Simulation of a WPOA Sample

The simplest $\boldsymbol{K'_f}$ selection is $\boldsymbol{K'_f = 0}$. In this simulation, a monolayer graphene with one atom replaced by silicon is used as a test WPOA sample. Its transmission amplitude is 1, with all information contained in the phase. The phase of the transmission function $\angle\psi$ is less than 0.15π. This simulation was calculated at 80 kV with a convergence angle of 30 mrad and an aberration-free probe. The pixel size on the detector was 0.16 mrad/pixel. A noise-free 4DSTEM dataset was simulated and SSB, WDD, and LSS results were reconstructed from the same dataset. WDD is run with $\epsilon = 1$, and LSS is run with only $\boldsymbol{K'_f = 0}$. To use the 4DSTEM dataset directly in LSS reconstruction, $D_n(\boldsymbol{K_f})$ for a pixelated detector is defined as:

$$D_n(\boldsymbol{K_f}) = \begin{cases} 1, & where\ \boldsymbol{K_f} = \boldsymbol{K_{f,n}} \\ 0, & where\ \boldsymbol{K_f} \neq \boldsymbol{K_{f,n}} \end{cases} \quad (7\text{-}21)$$

Where $\boldsymbol{K_{f,n}}$ is coordinator of nth pixel on the detector.

As shown in Figure 7-3, the LSS ($\boldsymbol{K'_f = 0}$) results demonstrate graphene lattice and the silicon atom (the atom brighter than carbon atoms), with the replaced silicon atoms easily distinguishable, similar to traditional SSB and WDD methods for WPOA samples. Non-WPOA simulations and various $\boldsymbol{K'_f}$ conditions will be discussed in the following sections.

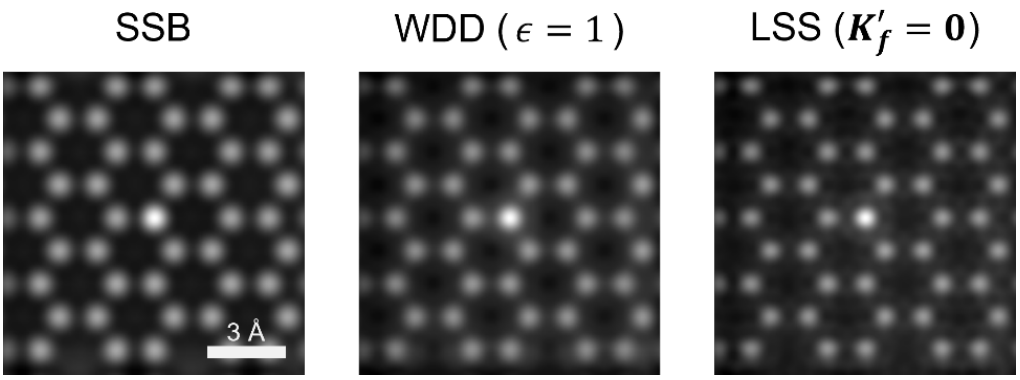

Figure 7-3 | Simulated phase image of a WPOA sample (monolayer graphene with Silicon replacement). All reconstructions of SSB, WDD ($\epsilon = 1$) and LSS ($\boldsymbol{K}'_f = 0$) shows lattice of graphene and the replaced silicon atom which is brighter than carbon atoms. Voltage: 80 kV, convergence angle: 30 mrad, aberration-free, pixel size on detector: 0.16 mrad/pixel.

7.7.2 Simulation of a non-WPOA Sample

For non-WPOA simulations, a sample with distinct amplitude and phase components in the exit wavefunction is ideal for observing how amplitude and phase information are transferred separately. However, the atomic projected potential typically exhibits very similar lattice patterns in both amplitude and phase. Therefore, in this simulation, a single-slice transmission function constructed from two normalised images, rather than an atomic projected potential, is used to separate amplitude and phase information. In Figure 7-4, the phase and amplitude of the constructed transmission function are shown. The normalised range of the phase is $[-0.5\pi, 0.5\pi]$, and the amplitude is $[0.5,1.0]$, as shown in the colorbars.

The simulation is based on this transmission function at an accelerating voltage of 300 kV and a convergence angle of 25 mrad, with an aberration-free probe. The real space scanning step is 0.2 Å, and 128×128 scan positions were used. The corresponding reciprocal space pixel size of $\boldsymbol{Q_p}$ is 0.039 Å^{-1}. A virtual detector with 128×128 pixels was used, with a pixel size of 0.75 mrad. SSB, WDD ($\epsilon = 1 \times 10^{-4}$), and LSS ($\left|\boldsymbol{K}'_f\right| \leq 2\ mrad$) were used for reconstruction, with the results are shown in Figure 7-5. The pixel size of $\boldsymbol{K}'_f$

is 0.75 mrad, which equals 0.039 Å^{-1}, the same as the real space sampling.

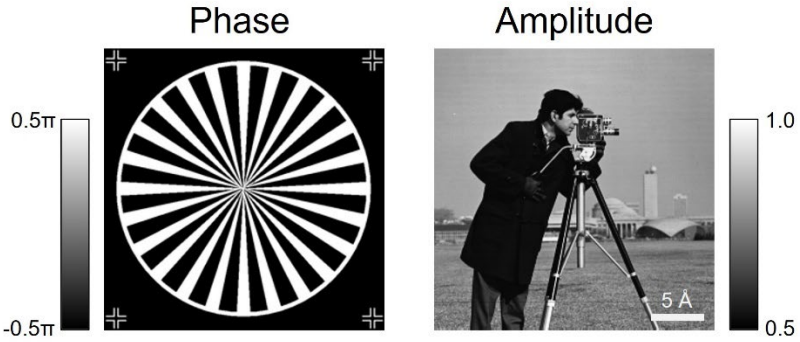

Figure 7-4 | Ground truth of constructed transmission function, phase and amplitude. A transmission function used in the following simulations is constructed by combining the phase and amplitude shown here. The range of the phase is from -0.5π to 0.5π and amplitude from 0.5 to 1. Scale bar: 5 Å

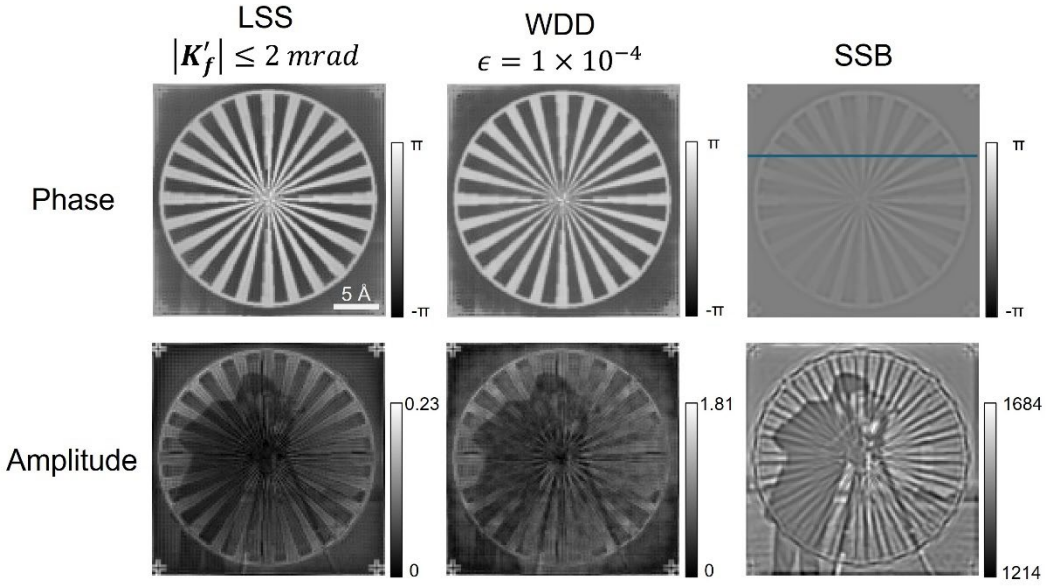

Figure 7-5 | Reconstruction results of LSS ($|K'_f| \leq 2\ mrad$), WDD ($\epsilon = 1 \times 10^{-4}$) and SSB. Colormaps are attached. The blue line on the SSB phase reconstruction indicates the position of line profile in Figure 7-6. Acceleration voltage: 300 kV, convergence angle: 25 mrad, aberration-free, scanning step: 0.2 Å, scan positions: 128×128.

The condition $|K'_f| \leq 2\ mrad$ for LSS and $\epsilon = 1 \times 10^{-4}$ for WDD reconstructions are selected based on Root Mean Square Error (RMSE) of the phase reconstructions and the

phase of transmission function shown in Figure 7-4. The RMSE results calculated for LSS and WDD under different $\boldsymbol{K}_{\boldsymbol{f}}'$ or ϵ conditions are shown in Figure A1-1, Appendix 1. $\left|\boldsymbol{K}_{\boldsymbol{f}}'\right| \leq 2\ mrad$ results in a LSS phase reconstruction with lowest RMSE among: $\boldsymbol{K}_{\boldsymbol{f}}' = \boldsymbol{0}$, $\left|\boldsymbol{K}_{\boldsymbol{f}}'\right| \leq 1mrad$, $\left|\boldsymbol{K}_{\boldsymbol{f}}'\right| \leq 2mrad$, $\left|\boldsymbol{K}_{\boldsymbol{f}}'\right| \leq 5mrad$, $\left|\boldsymbol{K}_{\boldsymbol{f}}'\right| \leq 10mrad$, $\left|\boldsymbol{K}_{\boldsymbol{f}}'\right| \leq 15mrad$. LSS reconstruction with these $\boldsymbol{K}_{\boldsymbol{f}}'$ conditions will be discussed in Section 7.7.3. For WDD, $\epsilon = 1 \times 10^{-4}$ results in a phase reconstruction with lowest RMSE among: $\epsilon = 1 \times 10^{-7}$, 1×10^{-6}, 1×10^{-5}, 1×10^{-4}, 1×10^{-3}, 1×10^{-2}, 1×10^{-1}, 1×10^{0}, 1×10^{1}, 1×10^{2}. These WDD reconstructions are shown in Figure A1-2 and A1-3 in Appendix 1.

In Figure 7-5, the noise-free simulated dataset allows for the comparison of algorithms and parameters without the influence of experimental differences in fluence. For phase reconstruction, all three methods in Figure 7-5 successfully reconstruct the phase information. The RMSE values for the ground truth of three-phase reconstructions are presented in Figure 7-6 (a). The LSS ($\left|\boldsymbol{K}_{\boldsymbol{f}}'\right| \leq 2\ mrad$) and WDD ($\epsilon = 1 \times 10^{-4}$) yield similar RMSE values, 0.665 and 0.667, respectively. In contrast, the SSB result has a higher RMSE of 1.29, indicating more error in its reconstruction for this simulation. In addition, on the line profiles shown in Figure 7-6 (b), LSS ($\left|\boldsymbol{K}_{\boldsymbol{f}}'\right| \leq 2\ mrad$) and WDD ($\epsilon = 1 \times 10^{-4}$) phase reconstructions align with the scale of the ground truth, while SSB phase scale is supressed, as shown on the line profiles in Figure 7-6 (b).

A potential reason that SSB cannot reconstruct a correct phase scale is the transmission function shown in Figure 7-4 not satisfying WPOA. As discussed in Chapter 2 and 3, the relationship between transmission functions and the WPOA is:

$$\psi(\boldsymbol{R}) \xrightarrow{POA} exp[\psi_s(\boldsymbol{R})] \xrightarrow{WPOA} 1 + \psi_s(\boldsymbol{R}) \qquad (7\text{-}22)$$

$$Where\ \psi_s(\boldsymbol{R})\ is\ an\ imaginary\ function$$

In this simulation, transmission function shown in Figure 7-4 is $\psi(\boldsymbol{R})$. The phase range of $\angle\psi(\boldsymbol{R})$ was set to $[-0.5\pi, 0.5\pi]$, which resulted in a $\psi_s(\boldsymbol{R}) \in [-0.5\pi i, 0.5\pi i]$ using POA. However, in WPOA condition, the phase of $1 + \psi_s(\boldsymbol{R})$ never reaches $\pm 0.5\pi$ even as

$\psi_s(\boldsymbol{R})$ approaches positive or negative infinity. Therefore, the reliance on the WPOA condition can be the reason why the SSB fails to reconstruct the accurate phase range in this simulation.

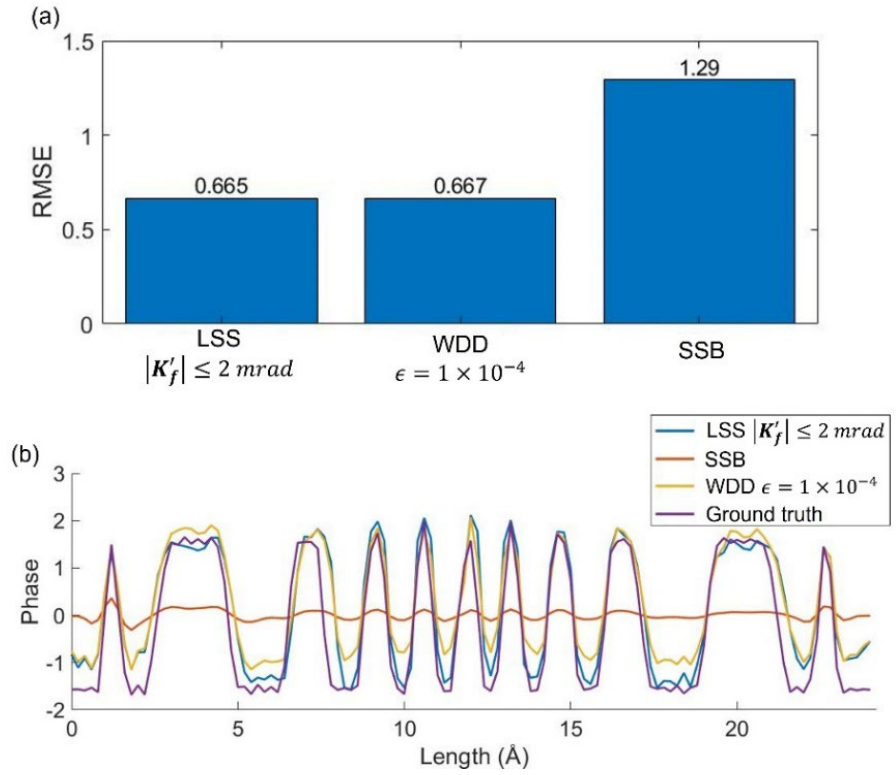

Figure 7-6 | RMSE and line profiles of LSS ($|\boldsymbol{K}_f'| \leq 2\ mrad$), WDD ($\epsilon = 1 \times 10^{-4}$) and SSB phase reconstructions shown in Figure 7-5. (a) RMSE. (b) Line profiles (position indicated by the blue line on the SSB reconstructed phase in Figure 7-5) of LSS ($|\boldsymbol{K}_f'| \leq 2\ mrad$), WDD ($\epsilon = 1 \times 10^{-4}$) and SSB phase reconstructions and phase of ground truth shown in Figure 7-4. LSS ($|\boldsymbol{K}_f'| \leq 2\ mrad$) and WDD ($\epsilon = 1 \times 10^{-4}$) phase reconstructions show similar scale compared to ground truth, while SSB phase reconstructions are supressed.

None of the reconstructed amplitudes in Figure 7-5 completely separates phase information from amplitude. For the SSB result, this may be due to a violation of the WPOA condition. However, the residual phase information observed in the amplitudes for LSS and WDD requires further investigation. A possible explanation is that it may be related to the choice of ϵ in WDD and $|K_f'|$ constraints is LSS.

In Figure 7-5 and 7-6, LSS demonstrates its comparability to the WDD method for

reconstructing phase sample with amplitude variation, provided that $\boldsymbol{K}_{\boldsymbol{f}}'$ is appropriately chosen. In the following sections, how different $\boldsymbol{K}_{\boldsymbol{f}}'$ conditions affect LSS reconstruction with be discussed through additional simulations.

7.7.3 $\left|\boldsymbol{K}_{\boldsymbol{f}}'\right|$ Threshold

As discussed in Section 7.2, ideally, $\boldsymbol{K}_{\boldsymbol{f}}' \in (-\infty, +\infty)$ should be used to achieve the best precision for solving Φ. However, solving a large linear system with many $\boldsymbol{K}_{\boldsymbol{f}}'$ positions require more computation time, which will be compared in next section. As discussed in Section 7.5, very high angle $\boldsymbol{K}_{\boldsymbol{f}}'$ positions (out of the detector area) do not contribute to Equation (7-8) and (7-10). Therefore, experimentally, it is not necessary to integrate Equation (7-6) with respect to $\boldsymbol{K}_{\boldsymbol{f}}' \in (-\infty, +\infty)$. There are hence various strategies for reducing the number of $\boldsymbol{K}_{\boldsymbol{f}}'$ sampling positions. In this section, the simplest strategy, limiting the range of $\left|\boldsymbol{K}_{\boldsymbol{f}}'\right|$, or called "$\left|\boldsymbol{K}_{\boldsymbol{f}}'\right|$ threshold", is considered and simulated.

$\left|\boldsymbol{K}_{\boldsymbol{f}}'\right|$ thresholding is equivalent to defining a hard aperture function on the $\boldsymbol{K}_{\boldsymbol{f}}'$ plane. The radius of the aperture (t) determines how many $\boldsymbol{K}_{\boldsymbol{f}}'$ positions are considered in Equation (7-8). A series of simulations under different $\left|\boldsymbol{K}_{\boldsymbol{f}}'\right|$ thresholding conditions on the same non-WPOA sample is performed in this section.

The simulations were run with the same parameters as used in Section 7.7.2. An accelerating voltage of 300 kV and a convergence angle of 25 mrad were used, with an aberration-free probe. In real space, 128×128 scan positions were recorded with a scanning step of 0.2 Å. The corresponding reciprocal space pixel size ($\boldsymbol{Q}_{\boldsymbol{p}}$) is 0.039 Å^{-1}. A virtual detector with 128×128 pixels was used, with a pixel size ($\boldsymbol{K}_{\boldsymbol{f}}$) of 0.75 mrad. The detector pixel size corresponds to 0.039 Å^{-1} equal to the pixel size of $\boldsymbol{Q}_{\boldsymbol{p}}$ and $\boldsymbol{K}_{\boldsymbol{f}}$, a pixel size of $\boldsymbol{K}_{\boldsymbol{f}}'$ used in simulations was also 0.039 Å^{-1} as for the reconstruction.

The reconstructed phase and amplitude under various $\boldsymbol{K}_{\boldsymbol{f}}'$ conditions ($\boldsymbol{K}_{\boldsymbol{f}}' = \boldsymbol{0}$, $\left|\boldsymbol{K}_{\boldsymbol{f}}'\right| \leq 1\ mrad$, $\left|\boldsymbol{K}_{\boldsymbol{f}}'\right| \leq 2\ mrad$, $\left|\boldsymbol{K}_{\boldsymbol{f}}'\right| \leq 5\ mrad$, $\left|\boldsymbol{K}_{\boldsymbol{f}}'\right| \leq 10\ mrad$, $\left|\boldsymbol{K}_{\boldsymbol{f}}'\right| \leq 15\ mrad$) are shown in Figure 7-7. The reconstructed phase and amplitude use consistent colour bars: $[-0.5\pi, 0.5\pi]$ for phase and $[0,0.3]$ for amplitude, as indicated in the first row of Figure 7-

7. A blue line on the phase reconstruction of $\boldsymbol{K_f' = 0}$ marks the position of the line profiles shown in Figure 7-8 (b). The corresponding RMSE results, calculated against the transmission function shown in Figure 7-4, are presented in Figure 7-8 (a) as a bar figure.

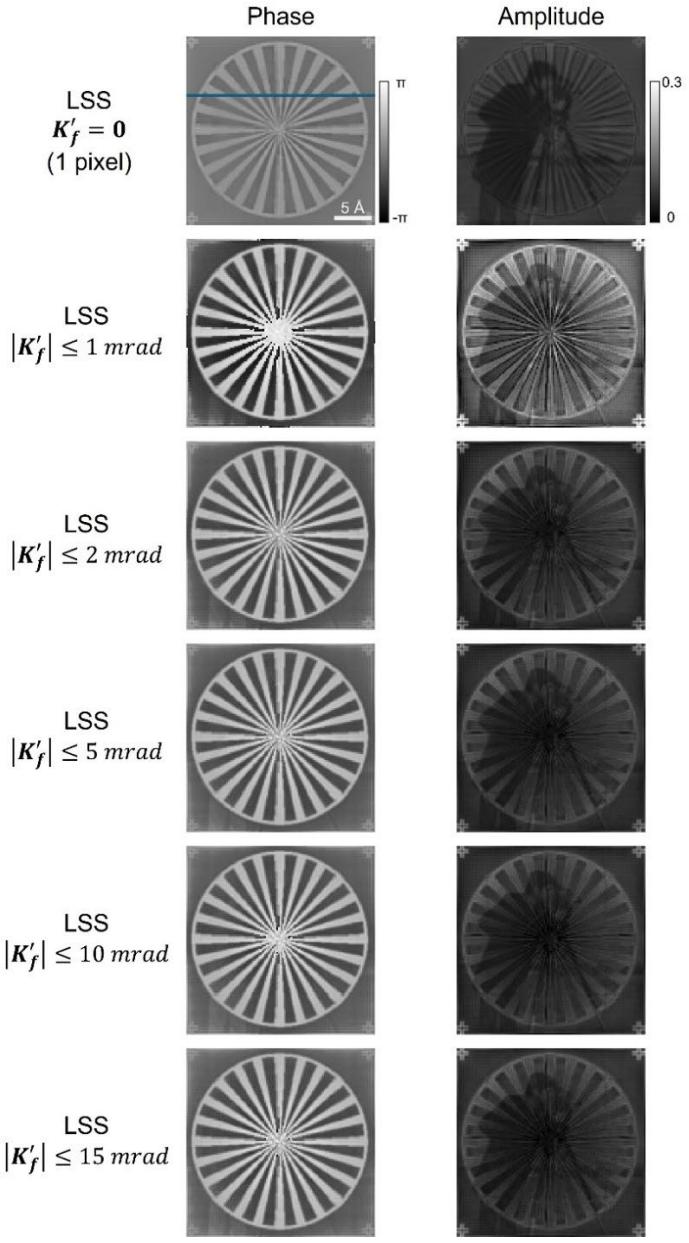

Figure 7-7 | LSS reconstructions with different $|\boldsymbol{K_f'}|$ thresholds: $\boldsymbol{K_f' = 0}$, $|\boldsymbol{K_f'}| \leq 1\ mrad$, $|\boldsymbol{K_f'}| \leq 2\ mrad$, $|\boldsymbol{K_f'}| \leq 5\ mrad$, $|\boldsymbol{K_f'}| \leq 10\ mrad$, $|\boldsymbol{K_f'}| \leq 15\ mrad$. The colour bar for reconstructed phase (-0.5π to 0.5π) and amplitude (0 to 0.3) are shown in the first row. Accelerating voltage: 300 kV, convergence angle: 25 mrad, aberration-free probe, scan positions: 128×128, scanning step: 0.2 Å.

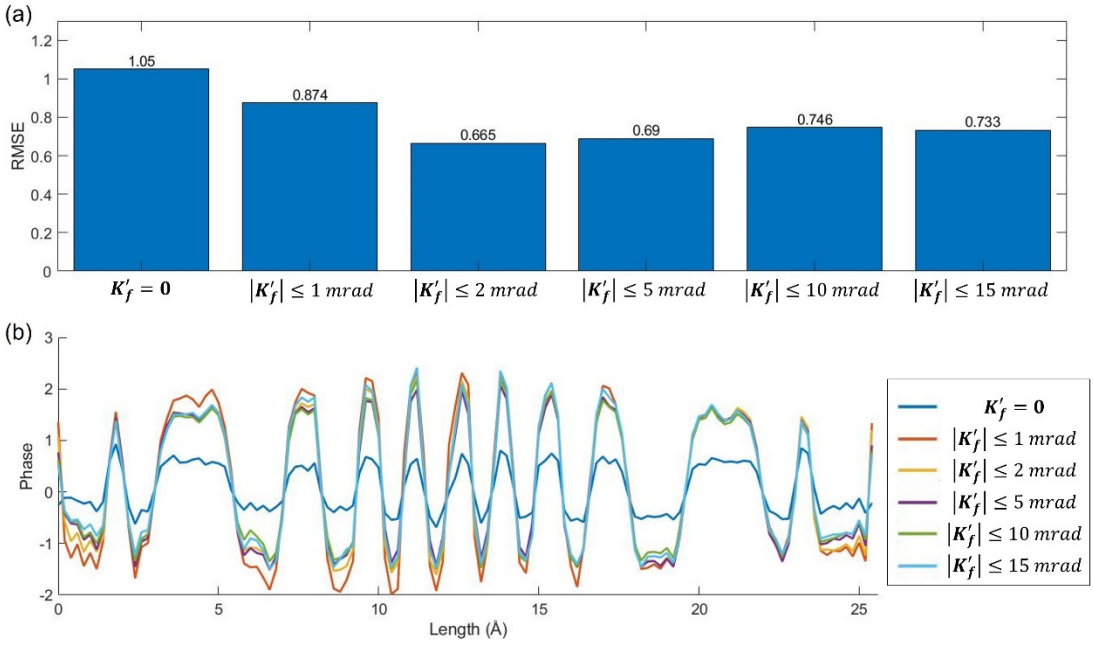

Figure 7-8 | RMSE and line profiles of LSS phase reconstructions shown in Figure 7-7. (a) RMSE. (b) Line profiles of the same position on all phase reconstructions in Figure 7-7 (position indicated by the blue line on the reconstructed phase of $\boldsymbol{K'_f = 0}$ in Figure 7-7). LSS phase reconstructions with a $\boldsymbol{K'_f}$ threshold higher than 1 mrad show similar scale (approximately -0.5π to 0.5π), while LSS phase reconstruction with $\boldsymbol{K'_f = 0}$ shows a smaller phase scale.

In the phase reconstructions of Figure 7-7, reconstructed phase of $\boldsymbol{K'_f = 0}$ shows a smaller phase scale than those of other conditions, and shows a weaker contrast under the same grey scale colour map. This weak contrast in the $\boldsymbol{K'_f = 0}$ reconstruction is more evident in the line profiles shown in Figure 7-8 (b), where the phase range for $\boldsymbol{K'_f = 0}$ is approximately half that of other conditions. This is potentially related to the violation of the WPOA condition, similar to the issue discussed regarding the SSB phase scale in the previous section.

As shown in Figure 7-8 (a), phase reconstruction of $\boldsymbol{K'_f = 0}$ shows the highest RMSE (1.05) among simulated conditions. Then the RMSE goes down to 0.665 with the condition $\boldsymbol{|K'_f| \leq 2\ mrad}$. Then slightly increase and become stable when the $\boldsymbol{|K'_f|}$ threshold is larger than 2 mrad. The change in RMSE proofs that consider positions that $\boldsymbol{|K'_f| \geq 0}$ can

result in phase reconstructions with less error than only considering $\boldsymbol{K}_f' = \mathbf{0}$. However, $|\boldsymbol{K}_f'| > 2\ mrad$ does not bring further improvements measurable by RMSE. On the contrast, high angle $\boldsymbol{K}_f'$ (up to 15 mrad) introduces phase wrapping to 3~4 pixels in the reconstruction, which can be the reason for the higher RMSE compared to $|\boldsymbol{K}_f'| \leq 2\ mrad$.

In conclusion, a proper $\boldsymbol{K}_f'$ threshold can minimize the RMSE between phase reconstruction and the phase of transmission function. For this simulation, $|\boldsymbol{K}_f'| \leq 2\ mrad$ is the optimal $\boldsymbol{K}_f'$ threshold among the conditions used. In the next section, calculation time of these simulations will be compared.

7.7.4 Calculation Time

In this section, the calculation times of LSS with various $\boldsymbol{K}_f'$ thresholds will be compared. However, it is important to note that calculation time is heavily influenced by several factors, including data size, code optimization, programming language, computer performance, computational device (CPU or GPU), and the number of cores or threads, among others. The comparison presented here is based solely on the simulation datasets used in Sections 7.7.2 and 7.7.3. Consequently, the calculation times provided should be considered as a limited reference for performance under different conditions.

The computer (workstation) used for these simulations is equipped with an Intel Core i9-10980XE CPU featuring 18 cores and 36 logical processors, 256 GB of RAM, and an NVIDIA Quadro RTX 8000 GPU. The program prototype is developed in Python using the PyTorch framework to leverage GPU acceleration. Linear algebra calculations are performed using the torch.linalg module in PyTorch.

The calculation times for simulations on the same sample and settings as in Section 7.7.2 and 7.7.3, using LSS (with different $\boldsymbol{K}_f'$ thresholds), SSB and WDD are summarized in Table 7-1. The calculation times of WDD with varying ϵ (shown in Appendix 1) are not listed, as the difference is negligible.

Table 7-1 | Calculation time of LSS, SSB and WDD

LSS	$\boldsymbol{K}_f' = \mathbf{0}$	50 sec
	$\lvert\boldsymbol{K}_f'\rvert \leq 1\ mrad$	2 min 4 sec (124 sec)
	$\lvert\boldsymbol{K}_f'\rvert \leq 2\ mrad$	7 min 39 sec (459 sec)
	$\lvert\boldsymbol{K}_f'\rvert \leq 5\ mrad$	50 min 49 sec (3049 sec)
	$\lvert\boldsymbol{K}_f'\rvert \leq 10\ mrad$	2 hr 50 min 27 sec (10027 sec)
	$\lvert\boldsymbol{K}_f'\rvert \leq 15\ mrad$	6 hr 2 min 20 sec (21740 sec)
SSB		19 sec
WDD		22 sec

The relationship between the LSS calculation times and $\boldsymbol{K}_f'$ thresholds square (t^2) is also plotted in Figure 7-9. The t^2 corresponds to the $\boldsymbol{K}_f'$ threshold defined as $\lvert\boldsymbol{K}_f'\rvert \leq t$. For example, $\boldsymbol{K}_f' = \mathbf{0}$ corresponds to $t^2 = 0$, $\lvert\boldsymbol{K}_f'\rvert \leq 1\ mrad$ corresponds to $t^2 = 1$, $\lvert\boldsymbol{K}_f'\rvert \leq 2\ mrad$ corresponds to $t^2 = 4$.

Figure 7-9 shows that the calculation time is approximately proportional to t^2, there t^2 is proportional to the area that satisfies $\boldsymbol{K}_f'$ threshold in LSS algorithm. This area, in turn, is proportional to the number of $\boldsymbol{K}_f'$ sampling positions, denoted as M, corresponding to the number of columns of matrix T_m and the number of rows of vector Φ_v in Equation (7-8). In conclusion, the calculation time for the code used in this chapter is approximately proportional to the number of $\boldsymbol{K}_f'$ sampling positions.

The current prototype program for LSS, which operates as a single thread, has potential for optimization and acceleration. One improvement could be enabling multi-threading. In the LSS algorithm, the linear systems at each $\boldsymbol{Q}_p$ position, as described in Equations (7-8) and (7-10), are independent and can be computed in parallel. This approach would

consume a manageable amount of RAM (or GPU RAM) relative to the dataset size. Implementing multi-threading could make the program several times faster, depending on the number of threads utilized.

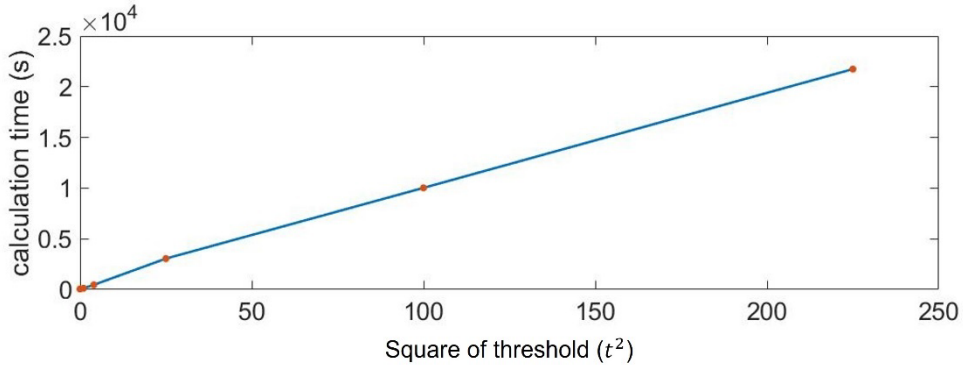

Figure 7-9 | Relation between square of $\boldsymbol{K}_f'$ threshold (t^2) and calculation time (seconds). The horizontal axis is t^2 that corresponding the $\boldsymbol{K}_f'$ threshold $|\boldsymbol{K}_f'| \leq t$. For example, $\boldsymbol{K}_f' = \boldsymbol{0}$ corresponding to a square of threshold $t^2 = 0$, $|\boldsymbol{K}_f'| \leq 1\ mrad$ corresponding to $t^2 = 1$, $|\boldsymbol{K}_f'| \leq 2\ mrad$ corresponding to $t^2 = 4$. The curve shows that the calculation time is approximately proportional to t^2.

7.7.5 $\boldsymbol{K}_f'$ Selection of Bragg Spots for Crystalline Samples

Another strategy for crystal samples is to select $\boldsymbol{K}_f'$ corresponding to specific Bragg reflections. For crystal samples, high-intensity bright spots are at the frequency corresponding to Bragg spots because they are the centres of diffraction discs. This means the $\boldsymbol{K}_f'$ on those spots can have a higher weight due to the high intensity. In this section, a simulation was carried out to test this strategy.

The simulation was carried out with a 200 kV acceleration voltage and a convergence angle of 25 mrad. Aberrations were not considered. The sample was FCC gold [100] film with a thickness of 100 Å. The scanning step was 0.1 Å, and 128×128 spots were recorded. A pixelated detector was used in this reconstruction and $D_n(\boldsymbol{K}_f)$ was set as defined by Equation (7-21).

For the $\boldsymbol{K}_f'$ selection, Bragg spots can be observed in the summation of $c_n(\boldsymbol{Q}_p)$ with respect to n, as shown in Figure 7-10 (a). The first circle of Bragg spots (marked in red circles) and the centre spot ($\boldsymbol{K}_f' = \boldsymbol{0}$, marked in green circles) were selected, and the corresponding $\boldsymbol{K}_f'$ positions were considered in Equation (7-8) and (7-10). The reconstruction using only $\boldsymbol{K}_f' = \boldsymbol{0}$ is shown in Figure 7-10 (b), while the reconstruction using all red and green circles marked spots as $\boldsymbol{K}_f'$ is plotted in Figure 7-10 (c).

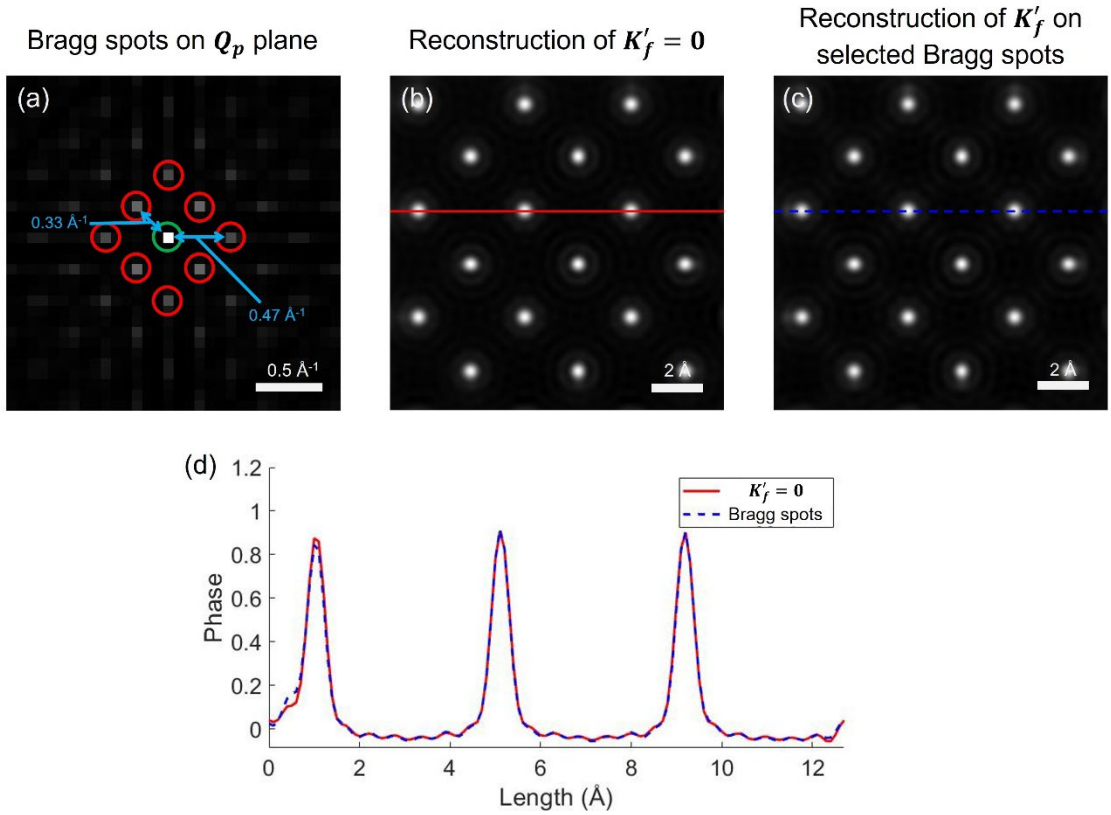

Figure 7-10 | Simulation of gold [100] film (thickness 100 Å), with $\boldsymbol{K}_f'$ selection on $\boldsymbol{K}_f' = \boldsymbol{0}$ or Bragg spots. (a) Selected Bragg spots (red and green circles marked) on the $\boldsymbol{Q}_p$ plane. (b) Phase reconstruction with $\boldsymbol{K}_f' = \boldsymbol{0}$ (green circle only). The red line indicates the line profile plotted in (d). (c) Phase reconstruction with $\boldsymbol{K}_f'$ on Bragg spots (red and green circles). The blue dashed line indicates the line profile plotted in (d). (d) Line profiles from (b) and (c), showing that the reconstructed phases are very similar. Voltage: 200 kV, convergence angle: 25 mrad, aberration-free probe, scanning step: 0.1 Å, scanning spots: 128×128.

In Figure 7-10, there is no obvious difference between reconstruction (b) and (c). The line profiles shown in Figure 7-10 (d) are almost identical except for small section that before 2 Å and after 11 Å on the horizontal axis. A potential reason for this similarity is that only $\boldsymbol{K}_{\boldsymbol{f}}' = \boldsymbol{0}$ is considered when solving ψ from $\Phi(\boldsymbol{K}_{\boldsymbol{f}}', \boldsymbol{Q}_{\boldsymbol{p}})$, as described in Section 7.3. The first circle of Bragg spots in Figure 7-10 (a) corresponds to 21 mrad in the diffraction plane. As noted in Section 7.6, a "stepping out" strategy [1, 6] has the potential to achieve a resolution of up to 0.51 Å (corresponding to 81 mrad, $2\alpha + 21$), if those Bragg spots are considered.

7.7.6 Aberration correction

In the LSS algorithm, a phase distortion of an aperture function A is considered in $T_n(\boldsymbol{K}_{\boldsymbol{f}}', \boldsymbol{Q}_{\boldsymbol{p}})$ and included in Equation (7-10). Therefore, LSS can compensate for aberration in the 4DSTEM dataset. In this section, a simulation using a defocused probe was carried out and reconstructed on a pixelated detector to illustrate this.

The results of the simulation are shown in Figure 7-11. The sample and simulation parameters are the same as those used in Section 7.7.3, with the only difference being a defocus of 15 nm added to the probe.

As shown in Figure 7-11, for LSS ($|\boldsymbol{K}_{\boldsymbol{f}}'| \leq 2\ mrad$), the effect of defocus is concentrated in the edge regions, specifically within 3 pixels of the edges. In these edge areas, the phase is wrapped and overlaps with the background. To mitigate the influence of phase wrapping at the edges, the RMSE calculations in Figure 7-12(a) consider only the central 121×121 pixels for all four phase reconstructions.

For the central 121 × 121 pixels, the RMSE values of the LSS ($|\boldsymbol{K}_{\boldsymbol{f}}'| \leq 2\ mrad$) reconstructions, calculated against the phase of the transmission function, are very close: 0.66496 for in-focus and 0.66530 for a 10 nm defocus, as shown in Figure 7-12 (a). In the central region, the line profiles of LSS ($|\boldsymbol{K}_{\boldsymbol{f}}'| \leq 2\ mrad$) under in-focus and defocus of 10 nm, shown in Figure 7-12 (b), align closely with each other and keep phase scales approximately from -0.5π to 0.5π. This indicates that LSS ($|\boldsymbol{K}_{\boldsymbol{f}}'| \leq 2\ mrad$) can

compensate for the influence of defocus in the central region, confining the major effects to the edge areas.

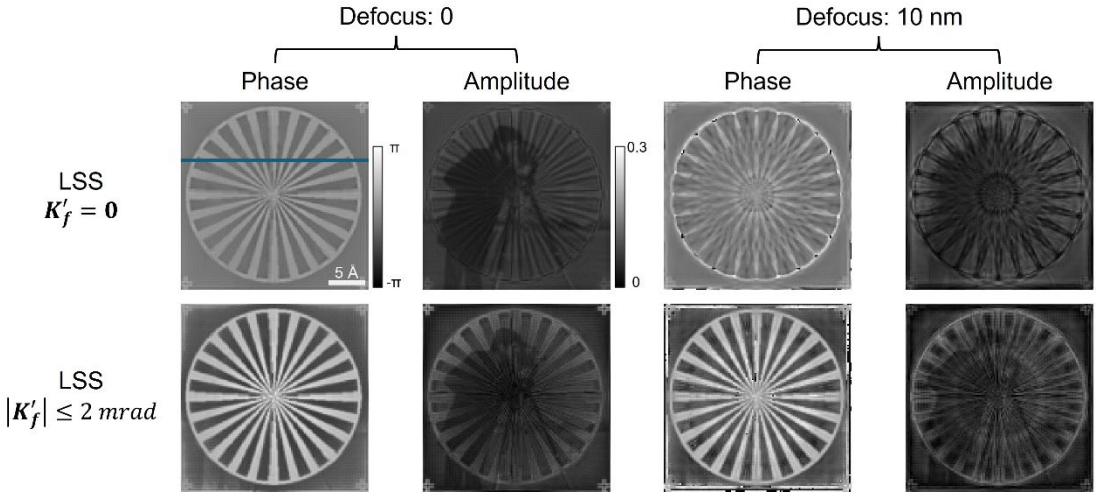

Figure 7-11 | LSS ($\boldsymbol{K'_f = 0}$ and $|\boldsymbol{K'_f}| \leq 2\ mrad$) reconstructions under in-focus and defocus conditions. The transmission function is shown in Figure 7-4. The reconstructed phase and amplitude use consistent colour bars: $[-0.5\pi, 0.5\pi]$ for phase and $[0,0.3]$ for amplitude. The blue line on the top left phase reconstruction shows the position of line profiles in Figure 7-12. Accelerating voltage: 300 kV, convergence angle: 25 mrad, scan positions: 128×128, scanning step: 0.2 Å.

However, for LSS ($\boldsymbol{K'_f = 0}$), a defocused probe influences the reconstructed phase. For the LSS ($\boldsymbol{K'_f = 0}$) reconstructions, the RMSE on the central regions is 1.0491 for in-focus 1.1897 for defocus of 10 nm, as shown in Figure 7-12 (a). On the Figure 7-12 (b), line profiles of LSS ($\boldsymbol{K'_f = 0}$) under in-focus and defocus of 10 nm do not fit as well as those of LSS ($|\boldsymbol{K'_f}| \leq 2\ mrad$). This suggests that the LSS method cannot fully compensate for affects by defocus when only $\boldsymbol{K'_f = 0}$ is considered.

These simulation results indicate that defocus may affect the LSS reconstruction of a strong phase sample. However, using different $\boldsymbol{K'_f}$ selections and sample conditions, the ability of LSS to compensate for aberrations, especially defocus, requires further experiments and simulations to be confirmed.

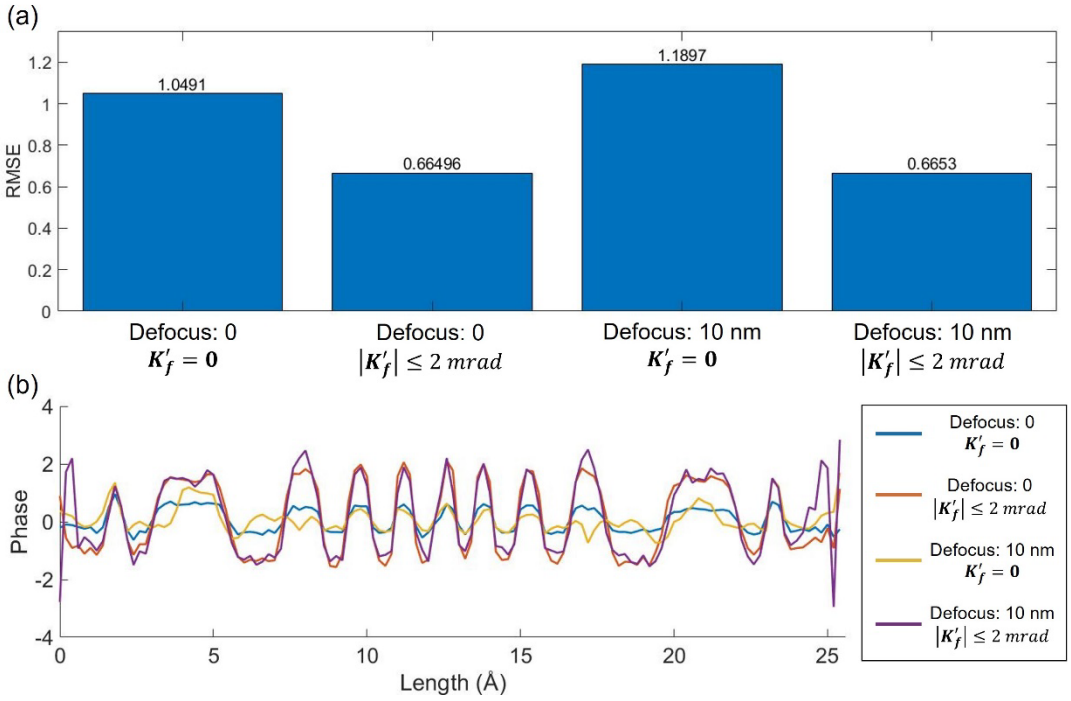

Figure 7-12 | RMSE on the centre 121 × 121 pixels and line profiles of LSS phase reconstructions shown in Figure 7-11. (a) RMSE on the centre 121 × 121 pixels of the phase reconstructions. (b) Line profiles of the same position on all phase reconstructions in Figure 7-11 (position indicated by the blue line on the reconstructed phase of $\boldsymbol{K'_f = 0}$ under in-focus condition).

7.7.7 LSS using Segmented detectors

LSS can also be applied to segmented detectors. The D_n is the same as used in AD-ptycho, as described in Chapter 6. A LSS simulation on a segmented detector is conducted and compared with AD-ptycho reconstruction on the same sample (the constructed transmission function shown in Figure 7-4). The simulation parameters are the same as used in Section 7.7.3 but only $\boldsymbol{K'_f = 0}$ is considered in this simulation. LSS ($\boldsymbol{K'_f = 0}$) reconstructed phase (a), amplitude (b) and AD-ptycho reconstruction (c) and the virtual detector layout (d) is shown in Figure 7-13.

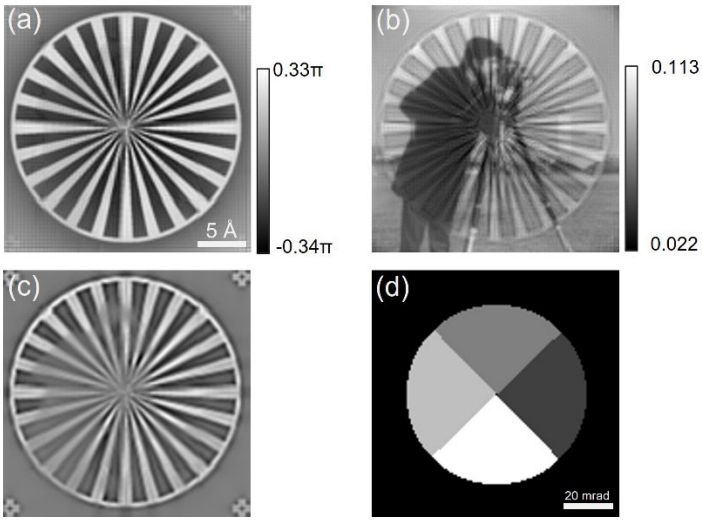

Figure 7-13 | LSS reconstruction on segmented detector compared to AD-ptycho. (a) Phase and (b) amplitude of LSS ($\boldsymbol{K}_{\boldsymbol{f}}' = \mathbf{0}$) reconstructions. (c) AD-ptycho phase reconstruction. (d) Segmented detector layout, segmented detector collection angle: 0-35 mrad. Accelerating voltage: 300 kV, convergence angle: 25 mrad, aberration-free probe, scan positions: 128×128, scanning step: 0.2 Å.

In Figure 7-13, both LSS ($\boldsymbol{K}_{\boldsymbol{f}}' = \mathbf{0}$) and AD-ptycho provide phase reconstructions even on a strong-phase sample with absorption. However, since AD-ptycho assumes that amplitude of transmission function is 1 (see Chapter 6), it only gives a phase reconstruction that mixes the phase and amplitude information from the transmission function. In Figure 7-13 (c), a shadow of the camera man from the amplitude information appears in the phase reconstruction. In contrast, LSS ($\boldsymbol{K}_{\boldsymbol{f}}' = \mathbf{0}$) separates phase and amplitude information, as shown in Figure 7-13 (a) and (b). This shows that LSS ($\boldsymbol{K}_{\boldsymbol{f}}' = \mathbf{0}$) can be used on segmented detectors as AD-ptycho and can provide better reconstruction on strong phase samples with absorption.

7.8 Conclusion

In this chapter, a linear system solution method (LSS) for analysing 4DSTEM datasets which does not require the WPOA is discussed. Similar to AD-ptycho (described in Chapter 6), LSS can reconstruct phase information on non-pixelated multichannel detectors. Its

connections to the AD-ptycho method under the WPOA condition and WDD in terms of resolution limits are also discussed. Furthermore, a series of simulations were conducted on both WPOA and non-WPOA samples to compare LSS with SSB, WDD and AD-ptycho.

Since LSS is a method based on solving linear systems, many algorithms could be applied to solve $\Phi(\boldsymbol{K}_f', \boldsymbol{Q}_p)$ beyond the least squares method used in this chapter. Importantly additional normalisation techniques or iterative methods may reduce errors or accelerate computation times.

Second, several simulations for different $\boldsymbol{K}_f'$ selections are described and it is demonstrated that there could be an optimized $\boldsymbol{K}_f'$ threshold that minimises the RMSE between phase reconstruction and ground truth. However, further investigation is needed to quantitatively determine the most efficient $\boldsymbol{K}_f'$ selection for specific experimental conditions.

Thirdly, LSS has not be tested under low-fluence conditions. However, similar to AD-ptycho, the ability to adapt to fast segmented detectors is beneficial for controlling the total fluence on the sample. In addition, aberration correction based on reconstruction algorithms can save time when focusing on the sample, reducing the electron fluence. However, LSS reconstruction on noisy datasets is not discussed in this chapter and all theoretical frameworks and simulations are based on noise-free conditions. Noise in the dataset may influence the solution of the linear system in LSS. Therefore, different strategies, beyond a simple least squares method, may perform differently under various noise conditions. Identifying an appropriate solving strategy is crucial for improving robustness in high-noise conditions, which is important for many sensitive specimens and low-fluence experiments [7, 8].

Fourthly, the resolution of LSS is limited by the process of solving ψ from $\Phi(\boldsymbol{K}_f', \boldsymbol{Q}_p)$, which is the same as in WDD. This resolution limit may be improved by applying the “stepping out” strategy or other super-resolution methods.

Reference

[1] J. Rodenburg and R. Bates, "The theory of super-resolution electron microscopy via Wigner-distribution deconvolution," *Philosophical Transactions of the Royal Society of London. Series A: Physical and Engineering Sciences,* vol. 339, no. 1655, pp. 521-553, 1992.

[2] P. Nellist and J. Rodenburg, "Beyond the conventional information limit: the relevant coherence function," *Ultramicroscopy,* vol. 54, no. 1, pp. 61-74, 1994.

[3] H. Yang *et al.*, "Electron ptychographic phase imaging of light elements in crystalline materials using Wigner distribution deconvolution," *Ultramicroscopy,* vol. 180, pp. 173-179, 2017.

[4] Y. Jiang *et al.*, "Electron ptychography of 2D materials to deep sub-ångström resolution," *Nature,* vol. 559, no. 7714, pp. 343-349, 2018.

[5] Z. Chen *et al.*, "Electron ptychography achieves atomic-resolution limits set by lattice vibrations," *Science,* vol. 372, no. 6544, pp. 826-831, 2021.

[6] P. Li, T. B. Edo, and J. M. Rodenburg, "Ptychographic inversion via Wigner distribution deconvolution: noise suppression and probe design," *Ultramicroscopy,* vol. 147, pp. 106-113, 2014.

[7] Z. Zhan, Y. Liu, W. Wang, G. Du, S. Cai, and P. Wang, "Atomic-level imaging of beam-sensitive COFs and MOFs by low-dose electron microscopy," *Nanoscale Horizons,* 2024.

[8] R. Egerton, P. Li, and M. Malac, "Radiation damage in the TEM and SEM," *Micron,* vol. 35, no. 6, pp. 399-409, 2004.

Chapter 8:
Conclusions and Future Work

8.1 Conclusions

Electron ptychography has been a widely used as an alternative phase retrieval method in TEM due to its high resolution [1, 2] and adaptability to various imaging conditions, including aberration correction [3] and low-fluence [4, 5]. Electron ptychography excels in achieving high resolution and fluence efficiency simultaneously, making it ideal for imaging sensitive samples [6, 7]. This thesis has explored the development and application of electron ptychography for fast imaging under low-fluence conditions, offering insights into its potential and limitations. The main chapters have provided an overview of the literature, detailed methodologies related to this thesis, and several developments of non-iterative ptychography methods towards fast and fluence-efficient ptychography dataset collection and imaging.

In Chapters 2 and 3, Current TEM phase-contrast imaging methods were reviewed, highlighting their advantages, disadvantages, and limitations regarding detector settings and fluence requirements. The mechanism for information transfer from the sample transmission function to diffraction patterns was discussed and detailed algorithms for existing non-iterative ptychography and related phase imaging methods were provided.

In Chapter 4, The impact of inelastic electron scattering on ptychographic reconstruction was discussed. Experiments and calculations indicated that inelastically scattered electrons have a negligible effect on high resolution experiments. The role of defocus,

which influences the size of interference fringes in CBED patterns, was also analysed and a limiting defocus condition for various accelerating voltages and convergent angles was calculated. Additionally, it was shown that TDS electrons do not significantly affect ptychographic reconstructions, although they may degrade resolution in ePIE reconstructions if dark field signals are included at high temperatures.

In Chapter 5, The Single-side Band masked Centre-of-mass (SSBm-COM) method, which integrates SSB ptychography and Centre-of-mass (COM) imaging, was proposed. This method applies the weak phase object approximation (WPOA) and is able to compensate for aberrations in 4DSTEM datasets, unlike traditional COM methods that require in-focus conditions during sample scanning. This feature is particularly advantageous for sensitive samples, as it reduces the total electron fluence by minimising the need for fine-tuning defocus before data collection. An experimental dataset showed that SSBm-iCOM can reconstruct graphene lattices with residual defoci (-7.5 and -22 nm). Additionally, the SSB-like masking process in SSBm-iCOM significantly reduces high-frequency noise compared to traditional COM methods. Simulations comparing the transfer function between SSB ptychography and SSBm-iCOM revealed that SSBm-iCOM enhances low-frequency information while transferring the same information from sample.

In Chapter 6, a non-iterative ptychography method, AD-ptycho, which can reconstruct data recorded on arbitrary multi-channel detectors under the weak phase object approximation (WPOA), was discussed. The flexibility of this method allows it to be applied to various detector layouts, including traditional segmented DPC and pixelated detectors. Despite its potential, AD-ptycho faces challenges such as high-frequency noise in reconstructions and limitations in aberration compensation related to the detector layout. An additional weighting in the frequency domain was applied to reduce high-frequency noise. By converting the problem of solving the transfer function into an overdetermined linear system, there are opportunities to enhance AD-ptycho using different methods to solve the linear system. However, the reliance on WPOA remains a theoretical constraint.

In Chapter 7, the linear system solution method (LSS) was introduced for analysing

4DSTEM datasets without relying on the WPOA, building on the AD-ptycho method. The selection of $\boldsymbol{K}_f'$, as a parameter for LSS, influences its reconstructions. In simulations, various settings for $\boldsymbol{K}_f'$ were tested and compared as well as computation time. LSS can compensate defocus and reconstruct phase information using non-pixelated multichannel detectors, same as AD-ptycho. Using simulations of both WPOA and non-WPOA samples, LSS with SSB, WDD, and AD-ptycho, were compared and the potential of LSS for imaging non-WPOA samples with pixelated or segmented detectors.

8.2 Future Work

In Chapter 4, only defocus was considered when calculating the “limiting defocus” for ptychography with inelastically scattered electrons, neglecting other aberrations. However, in real experimental conditions, more complex combinations of aberrations are common. Optimising experimental parameters to account for these additional aberrations, is a valuable future research direction.

For the SSBm-COM and AD-ptycho methods described in Chapters 5 and 6, experimental applications have so far been limited to a few material systems, primarily gold nano particles and graphene. To thoroughly test their robustness and stability, it is essential to gather more experimental datasets involving a variety of materials under different conditions, particularly low fluence conditions.

The linear system solution (LSS) method discussed in Chapter 7 presents several promising directions for further development. Since LSS is based on solving linear systems, exploring various algorithms beyond the least square method (LSM), such as iterative methods or additional normalisation process, could reduce errors and speed up computations. Further investigation into the optimal selection of $\boldsymbol{K}_f'$ positions is needed to enhance reconstruction quality for specific experimental conditions. Testing LSS under low-fluence conditions is also important, as its adaptability to fast segmented detectors can help control the pixel time and total electron fluence on samples. Therefore, more simulations and experiments are needed for addressing the impact of noise on LSS

reconstructions is essential. Furthermore, exploring super-resolution methods to enhance resolution and implementing normalisation steps to correct amplitude discrepancies are valuable pursuits for improving LSS. These efforts aim to make LSS a more robust and versatile tool for analysing 4DSTEM datasets in various experimental settings.

Reference

[1] Z. Chen *et al.*, "Electron ptychography achieves atomic-resolution limits set by lattice vibrations," *Science,* vol. 372, no. 6544, pp. 826-831, 2021.

[2] Y. Jiang *et al.*, "Electron ptychography of 2D materials to deep sub-ångström resolution," *Nature,* vol. 559, no. 7714, pp. 343-349, 2018.

[3] H. Yang *et al.*, "Simultaneous atomic-resolution electron ptychography and Z-contrast imaging of light and heavy elements in complex nanostructures," *Nature Communications,* vol. 7, no. 1, p. 12532, 2016.

[4] B. Küçükoğlu *et al.*, "Low-dose cryo-electron ptychography of proteins at sub-nanometer resolution," *Nature Communications,* vol. 15, no. 1, p. 8062, 2024.

[5] J. G. Lozano, G. T. Martinez, L. Jin, P. D. Nellist, and P. G. Bruce, "Low-dose aberration-free imaging of Li-rich cathode materials at various states of charge using electron ptychography," *Nano letters,* vol. 18, no. 11, pp. 6850-6855, 2018.

[6] J. Song *et al.*, "Atomic resolution defocused electron ptychography at low dose with a fast, direct electron detector," *Scientific reports,* vol. 9, no. 1, p. 3919, 2019.

[7] L. Zhou *et al.*, "Low-dose phase retrieval of biological specimens using cryo-electron ptychography," *Nature communications,* vol. 11, no. 1, p. 2773, 2020.

Appendix 1:
WDD reconstructions

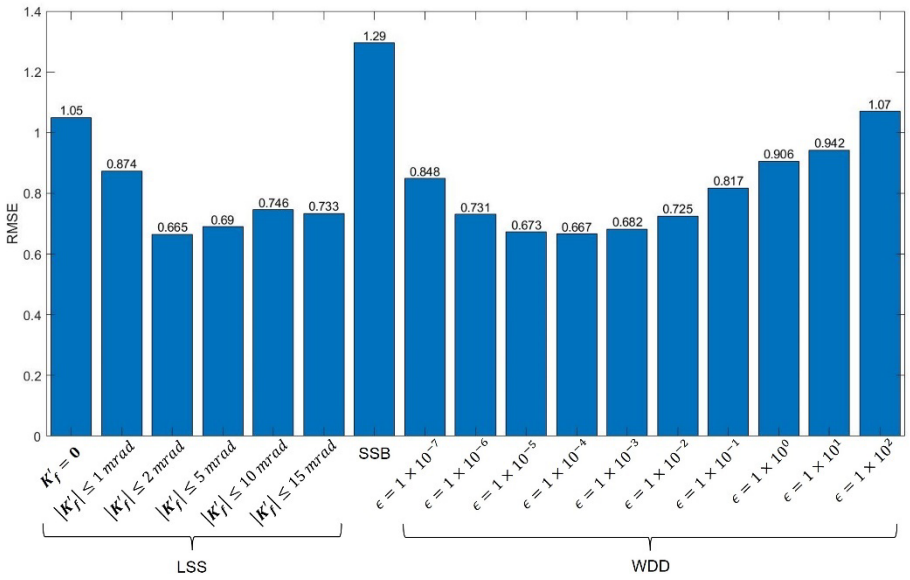

Figure A1-1 | RMSE results calculated based on SSB, LSS and WDD under various $\boldsymbol{K_f'}$ or ϵ conditions. For LSS, the $\boldsymbol{K_f'}$ conditions are: $\boldsymbol{K_f'} = \mathbf{0}$, $|\boldsymbol{K_f'}| \leq 1mrad$, $|\boldsymbol{K_f'}| \leq 2mrad$, $|\boldsymbol{K_f'}| \leq 5mrad$, $|\boldsymbol{K_f'}| \leq 10mrad$, $|\boldsymbol{K_f'}| \leq 15mrad$. For WDD, the ϵ conditions are: $\epsilon = 1\times10^{-7}$, 1×10^{-6}, 1×10^{-5}, 1×10^{-4}, 1×10^{-3}, 1×10^{-2}, 1×10^{-1}, 1×10^{0}, 1×10^{1}, 1×10^{2}. The conditions leading to lowest RMSE, $|\boldsymbol{K_f'}| \leq 2mrad$ for LSS and $\epsilon = 1\times10^{-4}$ for WDD, are used in Figure 7-5. The LSS phase and amplitude reconstructions are shown in Figure 7-7. The WDD phase and amplitude reconstructions are shown in Figure A1-2 and A1-3.

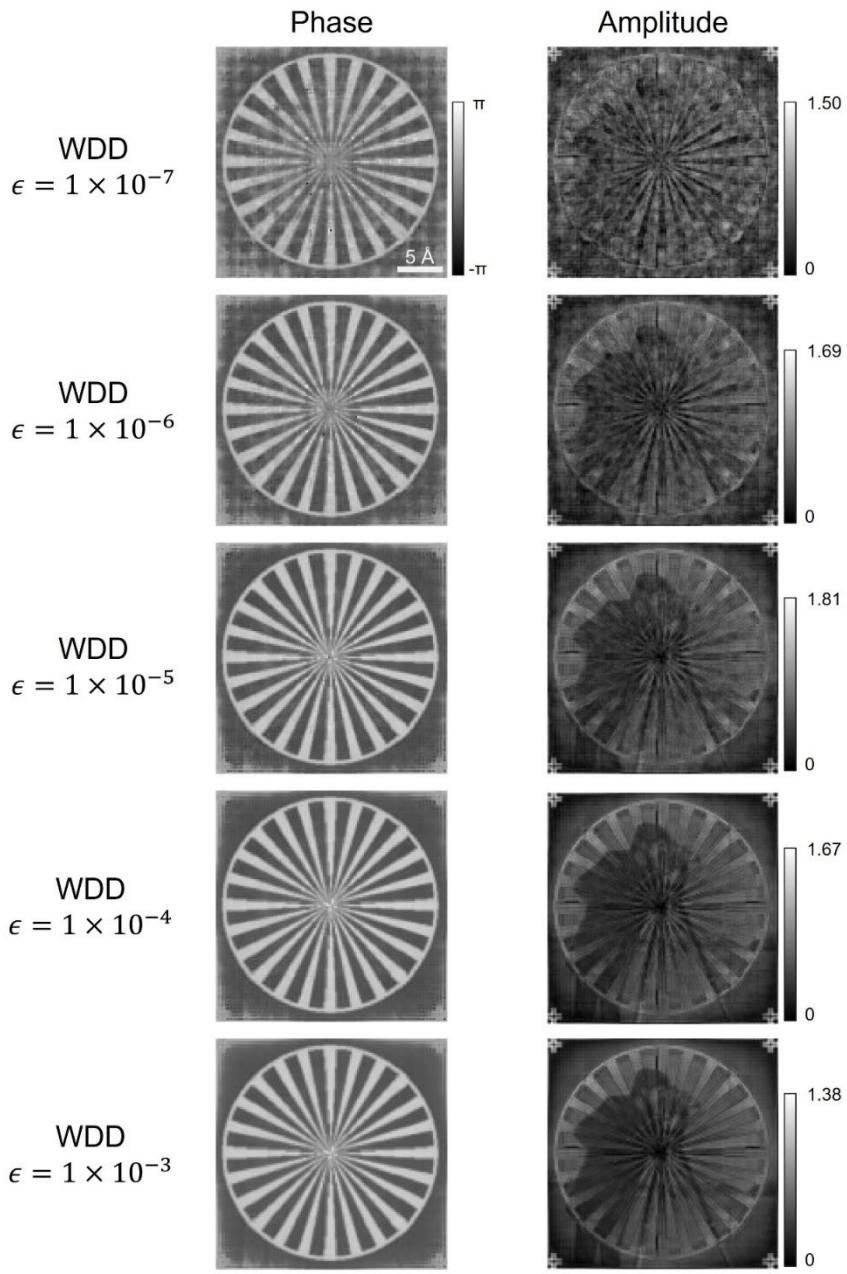

Figure A1-2 | WDD phase and amplitude reconstructions for the constructed transmission function shown in Figure 7-4 under various ϵ conditions: $\epsilon = 1 \times 10^{-7}$, 1×10^{-6}, 1×10^{-5}, 1×10^{-4}, 1×10^{-3}. The colour bar for phase reconstructions ranges from -0.5π to 0.5π, while the colour bar for amplitude reconstructions is displayed alongside each reconstruction. The RMSE values for the phase reconstructions, compared with the transmission function phase, are presented in Figure A1-1.

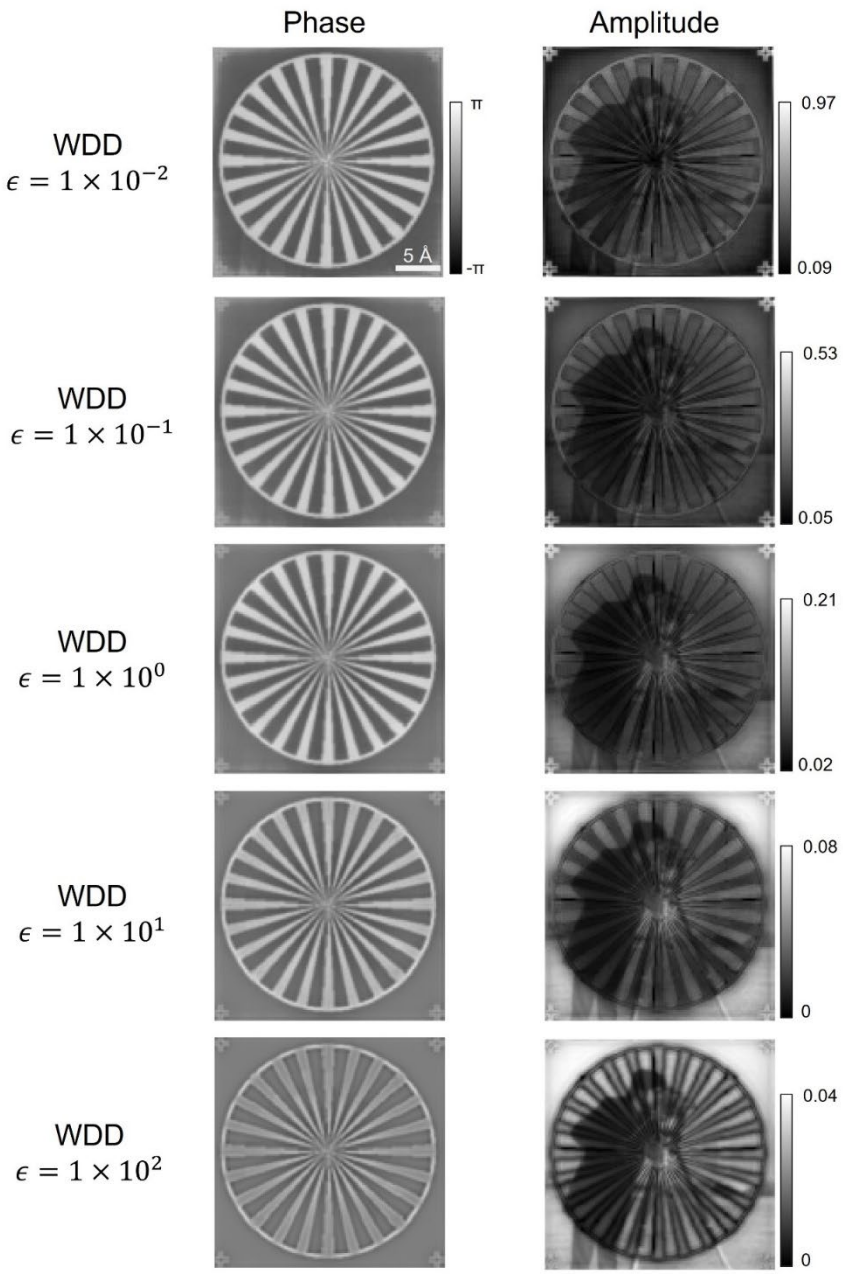

Figure A1-3 | WDD phase and amplitude reconstructions for the constructed transmission function shown in Figure 7-4 under various ϵ conditions: $\epsilon = 1 \times 10^{-2},\ 1 \times 10^{-1},\ 1 \times 10^{0},\ 1 \times 10^{1},\ 1 \times 10^{2}$. The colour bar for phase reconstructions ranges from -0.5π to 0.5π, while the colour bar for amplitude reconstructions is displayed alongside each reconstruction. The RMSE values for the phase reconstructions, compared with the transmission function phase, are presented in Figure A1-1.