# THE DEVELOPMENT OF IMAGE PROCESSING ALGORITHMS IN CRYO-EM

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Rui Yan

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

May 2019

Purdue University

West Lafayette, Indiana

# THE PURDUE UNIVERSITY GRADUATE SCHOOL
## STATEMENT OF DISSERTATION APPROVAL

Dr. Wen Jiang, Chair

    Department of Biological Sciences

Dr. Richard Kuhn

    Department of Biological Sciences

Dr. Daisuke Kihara

    Department of Computer Science and Department of Biological Sciences

Dr. Leifu Chang

    Department of Biological Sciences

**Approved by:**

    Dr. Janice Evans

        Head of the School Graduate Program

In dedication to my grandparents.

# ACKNOWLEDGMENTS

I am deeply grateful to my Ph.D advisor, Dr. Wen Jiang, for his support and guidance throughout my study. I would like to thank my committee members, Dr. Richard Kuhn, Dr. Daisuke Kihara, Dr. Leifu Chang, and my former committee member, Dr. Jason Lanman, for their patience and advice in my research projects. I would like to express my gratitude to my labmates and all faculty, staff and students in Hockmeyer Hall for Structural Biology, for their feedback, cooperation and of course friendship. Last but not the least, I would like to thank my family for supporting me spiritually throughout my life in general.

TABLE OF CONTENTS

LIST OF FIGURES

## ABSTRACT

Yan, Rui Ph.D., Purdue University, May 2019. The Development of Image Processing Algorithms in Cryo-EM. Major Professor: Wen Jiang. Professor.

Cryo-electron microscopy (cryo-EM) has been established as the leading imaging technique for structural studies from small proteins to whole cells at a molecular level. The great advances in cryo-EM have led to the ability to provide unique insights into a wide variety of biological processes in a close to native, hydrated state at near-atomic resolutions. The developments of computational approaches have significantly contributed to the exciting achievements of cryo-EM. This dissertation emphasizes new approaches to address image processing problems in cryo-EM, including tilt series alignment evaluation, simultaneous determination of sample thickness, tilt, and electron mean free path based on Beer-Lambert law, Model-Based Iterative Reconstruction (MBIR) on tomographic data, minimization of objective lens astigmatism in instrument alignment and defocus and magnification dependent astigmatism of TEM images. The final goal of these methodological developments is to improve the 3D reconstruction of cryo-EM and visualize more detailed characterization.

# 1. INTRODUCTION

Imaging techniques are essential research tools in many science and engineering disciplines. In recent years, cryo-electron microscopy (cryo-EM) has emerged as a mainstream technology for providing exciting insights into the architecture of whole cells, viruses and protein complexes which are difficult to elucidate using other structural approaches. Numerous technological and computational advancements for cryo-EM allow three-dimensional (3D) structures of biological specimens to be determined in fully frozen-hydrated and near-native state at near-atomic resolution. Enhanced image processing strategies, coupled with developments in microscope design and detector technology, further advance the effectiveness of cryo-EM methods. The improvement of powerful image processing algorithms has facilitated the calculation of more detailed structures from the noisy cryo-EM images and pushed the resolution limits towards the atomic level. This dissertation focuses on the development of cryo-EM image processing approaches, including quantitative evaluation of tilt series alignment, simultaneous determination of sample thickness, tilt, and electron mean free path based on Beer-Lambert law, Model-Based Iterative Reconstruction (MBIR), minimization of objective lens astigmatism, defocus and magnification dependent astigmatism of TEM images.

This dissertation is organized in the following manner.

In Chapter 2, we present a fast and accurate method, *tomoAlignEval*, based on the Beer-Lambert law, for the evaluation of alignment quality of tilt series in cryo-electron tomography (cryo-ET). Cryo-ET has been established as an emerging technique that can elucidate the architecture of macromolecular complexes and cellular ultrastructure in a near-native state. Similar to medical computed tomography, cryo-ET images the sample along multiple views by systematically tilting the sample and then computationally reconstructs the 3-D structures from these 2-D views. Accurate alignment

of tilt series is an essential step in attaining high-resolution 3D reconstructions. Nevertheless, quantitative assessment of alignment quality has remained a challenging issue, even though many alignment methods have been reported. Our method is able to globally estimate the alignment accuracy by measuring the goodness of log-linear relationship of the beam intensity attenuations at different tilt angles. Extensive tests with experimental data demonstrated its robust performance with stained and cryo samples. Our method is not only significantly faster but also more sensitive than measurements of tomogram resolution using Fourier shell correlation method ($FSC_{e/o}$). From these tests, we also conclude that while current alignment methods are sufficiently accurate for stained samples, inaccurate alignments remain a major limitation for high resolution cryo-ET.

In Chapter 3, we describe a computational method, *tomoThickness*, based on the Beer-Lambert law, to simultaneously determine the sample thickness, tilt and electron inelastic mean free path by solving an overdetermined nonlinear least square optimization problem utilizing the strong constraints of tilt relationships. In the field of cryo-ET, some important sample parameters, such as thickness and tilt, are needed for 3D reconstruction. However, these parameters can currently only be determined using trial 3D reconstructions. Accurate electron mean free path plays a significant role in modeling image formation process essential for simulation of electron microscopy images and MBIR method; however, their values are voltage and sample dependent and have only been experimentally measured for a limited number of sample conditions. Our method has been extensively tested with both stained and cryo datasets. The fitted electron mean free paths are consistent with reported experimental measurements. The accurate thickness estimation eliminates the need for a generous assignment of Z-dimension size of the tomogram. Interestingly, we have also found that nearly all samples are a few degrees tilted relative to the electron beam. Compensation of the intrinsic sample tilt can result in horizontal structure and reduced Z-dimension of tomograms. Our fast, pre-reconstruction method can

thus provide important sample parameters that can help improve the performance of tomographic reconstruction of a wide range of samples.

In Chapter 4, we introduce MBIR method to obtain tomographic 3D reconstructions of cryo-ET datasets. Due to radiation damage and the limited tilt range, cryo-ET suffers from low contrast and missing wedge artifacts, which limits the tomograms to low resolution and hinders further biological interpretation. To address these challenges of cryo-ET, we have applied MBIR to experimental data and demonstrated its advantages in contrast improvement, missing wedge artifacts reduction, and missing information restoration compared with other reconstruction approaches. Considering the outstanding reconstruction quality, MBIR has a great potential in the determination of high resolution biological structures with cryo-ET.

In Chapter 5, we develop an $s^2$ power spectra based method, $s^2 stigmator$, to perform fast and sensitive detection of astigmatism in TEM live images. Minimization of the astigmatism of the objective lens is a critical daily instrument alignment task essential for high resolution TEM imaging. Fast and sensitive detection of astigmatism is needed to provide real-time feedback and adjust the stigmators to efficiently reduce astigmatism. Currently the method used by many microscopists is to visually examine the roundness of a diffractogram (Thon rings) and iteratively adjust the stigmators to make the Thon rings circular. This subjective method is limited by poor sensitivity and potentially biased by the astigmatism of human eyes. Our method was implemented as a DigitalMicrograph script to allow fast and sensitive detection of the astigmatism of live images. The "radar"-style display provides real-time feedback to guide the adjustment of the objective lens stigmators. Such unique capability allowed us to discover the mapping of the two stigmators to the astigmatism amplitude and angle, which led us to develop a single-pass tuning strategy capable of significantly quicker minimization of the objective lens astigmatism.

In Chapter 6, we investigate how image astigmatism varies with the imaging conditions (e.g. defocus, magnification). Daily alignment of the microscope is a prerequisite to reaching optimal lens conditions for high resolution imaging in cryo-EM. We

have found that the large change of defocus/magnification between visual correction of astigmatism and subsequent data collection tasks, or during data collection, will inevitably result in undesirable astigmatism in the final images. The dependence of astigmatism on the imaging conditions varies significantly from time to time, so that it cannot be reliably compensated by pre-calibration of the microscope. Based on these findings, we recommend that the same magnification and the median defocus of the intended defocus range for final data collection are used in the objective lens astigmatism correction task during microscope alignment and in the focus mode of the iterative low-dose imaging. It is also desirable to develop a fast, accurate method that can perform dynamic correction of the astigmatism for different intended defocuses during automated imaging. Our findings also suggest that the slope of astigmatism changes caused by varying defocuses can be used as a convenient measurement of objective lens rotation symmetry and potentially an acceptance test of new electron microscopes.

# 2. A FAST CROSS-VALIDATION METHOD FOR ALIGNMENT OF ELECTRON TOMOGRAPHY IMAGES BASED ON BEER-LAMBERT LAW

## 2.1  Introduction

Electron tomography (ET) is an emerging technique capable of revealing the 3D structure of complex macromolecular architectures and cellular ultrastructure. A 3D tomogram of the specimen is reconstructed from a series of 2D projection images at different tilt angles. To correct the specimen drift due to instability of the instrument, accurate alignment of the tilt series is a prerequisite to obtaining a high-resolution tomographic reconstruction.

The alignment methods for ET tilt series can be classified into two categories: fiducial marker-based and marker-free alignment methods. Marker-based methods track the positions of high-contrast gold beads over the entire tilt series, and fit them into a projection model in order to determine the alignment parameters [1–4]. Marker-free alignment can be further divided into two types, correlation-based and feature-based methods, which rely on cross-correlations between adjacent image pairs within the tilt series [1] and identifying features in images as virtual markers [5], respectively. In addition, 3D model-based approaches were also developed employing projection matching which is commonly used in single particle analysis [5, 6].

Although there are multiple approaches to align the tilt series, quantitative assessment of alignment quality remains a critical issue for achieving high-resolution reconstruction. Most alignment methods use the root mean square deviation (RMSD) between the measured and expected positions of fiducials based on least-square metric to evaluate the alignment quality [4]. It is inevitable that these methods are sensitive to outliers from a few unreliably tracked fiducials, and can be arbitrarily minimized

by manual over-tuning of the bead positions. The most rigorous evaluation of alignment accuracy is to measure the resolution of the reconstructed tomogram [7] and use the resolution as an indication of alignment quality. However, it measures the compound effect of many factors, including data quality, alignment accuracy, CTF fitting/correction, and 3D reconstruction. Furthermore, it is also computationally intensive for large datasets [7] and is usually only done after the tomogram has already been reconstructed using the aligned images.

To remedy these problems, we have developed a fast and robust approach to estimate the quality of 2D alignment. Our approach is based on a mathematical model derived from the Beer-Lambert law and computes the total error between the image data and the image formation model, allowing us to globally assess the alignment quality with a merit figure.

## 2.2   Methods

### 2.2.1   Mathematical model

When the electron beam passes through the sample, the electrons will pass through without being scattered, or can be elastically or inelastically scattered. The inelastic scattering will reduce the electron wave amplitude and the beam intensity. The incident beam intensity $I_0$ and the transmitted beam intensity $I_{exit}$ follow the Beer-Lambert law, no matter whether inelastically scattered electrons are removed by energy filter. Based on the BeerLambert law, the beam intensity attenuates exponentially with increasing thickness of the sample.

This relationship can be described in Eq.2.1

$$\frac{I_{exit}}{I_0} = e^{-\frac{d_e}{\Lambda_{in}}} \tag{2.1}$$

where $d_e$ is the effective thickness representing the distance that the electron beam has traveled through in the specimen, $\Lambda_{in}$ is the mean free path for inelastic scattering, $I_0$ is the intensity of incident electron beam and $I_{exit}$ is the intensity of electron

beam exiting the specimen and hitting the detector. The Beer-Lambert law has been used in electron tomography to estimate sample thickness [8–10] or model the image formation process [11].

Note that the effective thickness $d_e$ in Eq.2.1 is not equal to the absolute geometric thickness $d_0$ due to tilt angle $\theta$. We can compute the effective thickness $d_e$ in Eq.2.2 using the geometric relationship:

$$d_e = \frac{d_0}{\cos \theta} \tag{2.2}$$

Furthermore, the exiting beam intensity $I_exit$ can be represented as Eq.2.3 to associate it with the pixel values we measure from images

$$I_{image} = A \cdot I_{exit} \tag{2.3}$$

where $I_{image}$ is the average pixel value of the selected area from the image, A is the gain factor of the detector.

Thus, we can transform our mathematical model for a tilted image to Eq.2.4 by combining Eqs.2.1-2.3

$$I_{image} = A \cdot I_0 \cdot e^{-\frac{d_0}{\Lambda_{in}} \cdot \frac{1}{\cos \theta}} \tag{2.4}$$

and alternatively on a log-scale

$$ln(I_{image}) = ln(A \cdot I_0) + (-\frac{d_0}{\Lambda_{in}}) \cdot \frac{1}{\cos \theta} \tag{2.5}$$

Assuming $A$, $I_0$, $\Lambda_{in}$ and $d_0$ are constant, there is a linear relationship between the independent variable $\frac{1}{\cos \theta}$ and the dependent variable $ln(I_{image})$ in the aligned image series.

## 2.2.2    Quantitative assessment of alignment quality

The approach consists of several steps as described below:

(1) Multiple regions (8 regions for the example shown in Fig.2.1) are selected in the untilted image and the corresponding patches in the entire tilt series are tracked according to the tilt angles and the geometric relationship. The patch size in the X-dimension ($N_x(\theta)$, perpendicular to the tilt axis along the Y-dimension) can either remain constant or adapt to the reduced sizes for tilted images($N_x(\theta) = N_x(0)\cos\theta$).
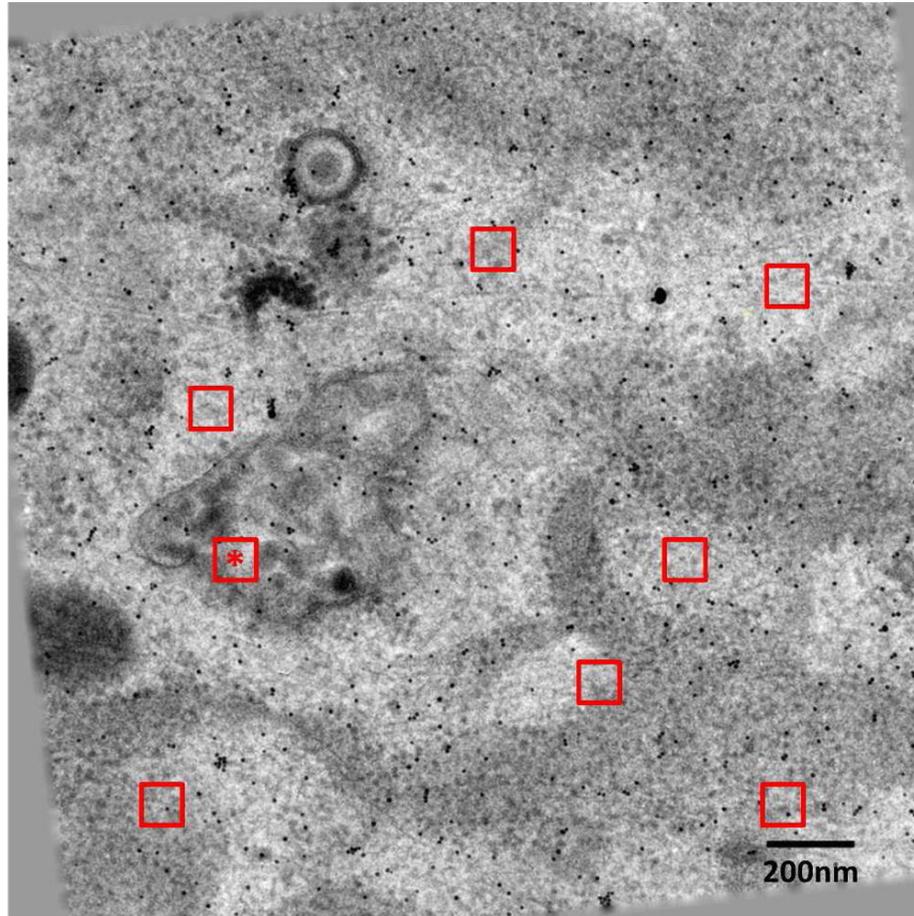


Fig. 2.1. Zero tilt image of an aligned tilt series of Sindbis virus infected BHK cell section. The infected cells were plastic embedded, sectioned, and then stained before being imaged. The tilt axis is vertical. The 8 red squares indicate regions selected and tracked through the tilt series. The region size is 100×100 pixels.

(2) For each selected region, the average pixel value of each tracked patch is used as the measured $I_{image}$ (Fig.2.2A).
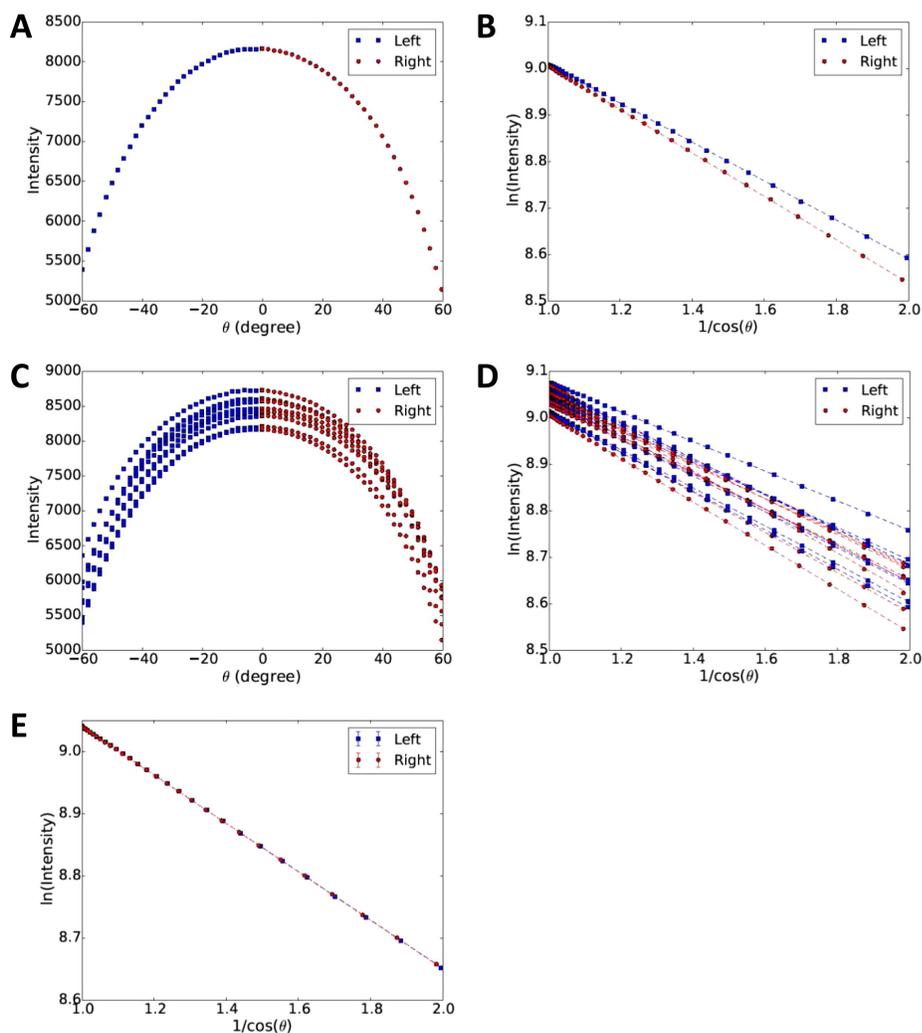
Fig. 2.2. Relationship between image intensity and tilt angles. (A) Profile of image intensity and tilt angle ($\theta$) for the red square region marked with symbol * shown in Fig.2.1. The blue squares and red circles indicate left and right branches, corresponding to negative and positive tilts, respectively. The images were taken in order of negative to positive tilt angles. (B) Replot of (A) using log of intensity (Y-axis) and $1/\cos\theta$ (X-axis). Colors are the same as in (A). The two dash lines indicate independent linear regression to points of two branches. (C and D) Expansion of plots (A) and (B) to include multiple regions. (E) Plots in (D) are rescaled and averaged using a common reference line with the median slope and the median intercept. The errors are too small to make the error bars visible.

(3)Based on Eq.2.5, it can be seen that $ln(I_{image})$ and $\frac{1}{\cos\theta}$ follow a linear trend if the tilt series is accurately aligned (Fig.2.2B). Each selected region would contribute two linear trends, one for negative tilts and one for positive tilts. When multiple regions (for example, 8 regions as shown by the red boxes in Fig.2.1) were selected, separate linear trends can be seen due to different image contents in different selected regions (Fig.2.2C and D). Among all linear trends, we found the median slope and the median intercept and used them as the reference to re-scale all linear trends to this common slope and intercept. For example, if we want to rescale the n-th linear trend $(y_n = k_n x + b_n)$ to the median slope $k_{median}$ and the median intercept $b_{median}$, the rescaled n-th linear trend will be $y'_n = k_{median}(\frac{y_n - b_n}{k_n}) + b_{median}$. We then computed the mean and standard deviation of the re-scaled data points of all regions at each tilt angle (Fig.2.2E). Thus, averaging of multiple regions was performed to reduce noise. In the final averaged plot (Fig.2.2E), the number of points is equal to the number of tilts, with every point $\bar{g}$ and its error bar representing the average pixel value and the standard deviation of different regions.

(4) Based on the mean pixel values for the different tilts obtained in the previous step, we generated a line fitting $f$ by linear regression in which $k_{fit}$ is the slope and $b_{fit}$ is the intercept.

$$f(i) = k_{fit} \cdot \frac{1}{\cos\theta(i)} + b_{fit} \tag{2.6}$$

where $i$ is the $i$th image in the tilt series.

(5) The squared error between the fitting line $f$ in Eq.2.6 and the averaged data $\bar{g}$ at each tilt angle is computed and denoted as residual $res(i)$.

$$res(i) = \left[\frac{\bar{g}(i) - f(i)}{f(i)}\right]^2 \tag{2.7}$$

The mean and standard deviation of all $res$ values from the entire tilt series (Fig.2.2E) serve as global quantitative indicators of goodness-of-alignment.

### 2.2.3  Implementation

The method described here has been implemented as a stand-alone python program *tomoAlignEval.py* for easy usage. Several parameters, such as, patch locations, patch size and patch type (constant or adaptive), were provided as command line options to give users the flexibilities to experiment for diverse samples. EMAN2 library functions [12] were used for image IO. *numpy/scipy* software [13, 14] was used to perform linear least square fitting and *matplotlib* software [15] was used for plotting. Although the program was only tested on Linux systems, it should also run on all major platforms (Linux, Windows and MacOS) on which the dependent software packages (python, numpy, scipy, matplotlib and EMAN2) are supported. The *tomoAlignEval.py* program is freely downloadable from the authors' Web site (http://jiang.bio.purdue.edu).

### 2.2.4  Test datasets

Both stained and cryo datasets were used to test the performance of our approach. Sindbis virus infected Baby Hamster Kidney fibroblasts (BHK) cells and Flock House virus infected *Drosophila* S2 cells were embedded in resins, sectioned, and then stained. Tilt series of both samples were obtained on a FEI Titan Krios at 300 kV and were recorded on a Gatan 4 K × 4 K CCD camera at room temperature. The samples were first pre-irradiated ($\sim 10^4 e/\text{Å}^2$) to stabilize the resin and minimize specimen shrinkage during data collection [16]. Tilt images were collected from -60° to +60° in 2° increments with constant dose for each tilt and a total dose of 4000 $e/\text{Å}^2$. The pixel size was 0.404 nm for the datasets used in Fig.2.5D, 0.51 nm for datasets of other stained samples. For cryo-datasets, the frozen-hydrated specimens, purified Sindbis virus and B. *burgdorferi*, were plunge frozen and imaged at liquid nitrogen temperature. The tilt series of purified Sindbis virus were acquired using a FEI Polara TEM at 300 kV on a Tietz 4 K × 4 K CCD and the tilt series of B. *burgdorferi* were acquired using a JEOL 2200 TEM with energy filter at 200 kV on

a Gatan 4 K × 4 K CCD camera. Tilt images were taken from -60° to +60° in 1°
increment and 1.5° increment for purified Sindbis virus and B. *burgdorferi*, respectively, with constant dose for each tilt and a total dose of 100 $e/Å^2$. The images of
purified Sindbis virus were 2× binned to 2048 × 2048 pixels. The pixel size was 0.45
nm for the datasets used in Fig.2.5H, 0.55 nm for other cryo datasets.

All tilt series were aligned using gold fiducial markers with the IMOD software
package [17]. For each tilt series, raw stack, pre-aligned stack and aligned stack were
all used to test if our method can effectively distinguish their alignment qualities.
Tomographic reconstruction of raw stack, pre-aligned stack and aligned stack were
performed using the back projection method implemented in IMOD. The $FSC_{e/o}$
curves, Fourier shell correlation of two tomograms built from even and odd subsets
of tilt images, were computed using ELECTRA [7].

## 2.3  Results

### 2.3.1  Experimental confirmation of the mathematical model

To confirm that the mathematical model (Eqs.2.4, 2.5), derived from the Beer-
Lambert law is indeed consistent with experimental tomographic tilt images, we plot-
ted the mean pixel values at different tilt angles. One region (100 × 100 pixels, the
red square marked with symbol * in Fig.2.1), was selected from the untilted image,
and tracked through the entire aligned tilt series according to its geometric relation-
ship. The average pixel values ($I_{image}$ in Eq.2.3) of all tracked patches were plotted
as a function of tilt angles in Fig.2.2A. When the tilt angles become larger, the effec-
tive thickness $d_e$ increases, leading to exponential attenuation of beam intensity. The
distribution of average pixel values depicts a parabolic curve which is qualitatively
consistent with the theoretical model shown in Eq.2.4, since the tilt angles at positive
and negative directions should produce the same effective thickness and the same
intensity.

To better illustrate the tilt-angle dependence for the beam intensity, Fig.2.2A was replotted in Fig.2.2B to show the log of intensity as a function of $\frac{1}{\cos\theta}$. It is now evident that there is a linear relationship (dash line) between $\frac{1}{\cos\theta}$ and $ln(I_{image})$ in the aligned image series, in excellent agreement with the Beer-Lambert law based mathematical model shown in Eq.2.5. For convenience, the curve in Fig.2.2A was divided into left (blue squares) and right (red circles) branches, corresponding to negative and positive tilts during data collection, respectively. The two branches were transformed to log scale in Fig.2.2B and marked with the corresponding colors and symbols used in Fig.2.2A. When multiple regions (8 red squares as shown in Fig.2.1) are plotted, multiple parabolic curves and straight lines are present, with one region for each set of curves (Fig.2.2C) or lines (Fig.2.2D). The curves/lines are offset from each other due to differences in the contents of different regions, which cause different attenuation rates of the beam and brightnesses in the tilt images. In Fig.2.2E, all linear trends were scaled to the median slope and the median intercept, and re-scaled data points at same tilt angle but from different regions were then averaged. The point and its error bar in Fig.2.2E represent the mean and standard deviation of intensities of different regions after rescaling. It can be seen that all the points form a nearly perfect line with no point visibly offset from the line. The error bar of each point in Fig.2.2E is too small to be visible. These results have confirmed that our theoretic model (Eq.2.5) is indeed consistent with the aligned experimental images.

### 2.3.2 Tests with computationally added alignment errors

To use the goodness of log-linear relationship as a quantitative measurement of the alignment quality of tilt series, a strong correlation of these two properties is required. We thus performed tests in which a series of perturbations [18], including shifts (Fig.2.3A), in-plane rotations (Fig.2.3B), and combinations of shifts and in-plane rotations (Fig.2.3C), were purposely applied to an aligned stack images shown

in Fig.2.1. In Fig.2.3, the points and error bars correspond to mean and standard deviation of the residuals defined in Eq.2.7. The mean residual of the aligned stack was set at 1 and all other values were shown as ratios to the mean residual of the aligned stack without additional alignment perturbations. The results (e.g. mean and standard deviation of residuals) showed a strong positive correlation with the applied perturbations to the alignments. Both the mean and standard deviation increase with larger applied perturbations, indicating that our method is able to distinguish the different amounts of alignment errors for a tilt series.

### 2.3.3 Quantitative assessment of alignment quality of stained samples

To assess the performance of the proposed method, we tested experimental datasets of both stained and cryo specimens. Fig.2.4 displays the intensities of the same regions ($100 \times 100$ pixels) in the raw stack (Fig.2.4A), the prealigned stack (Fig.2.4B) and the aligned stack (Fig.2.4C) of a dataset of stained sections of a Sindbis virus infected BHK cell.

Here, the prealigned stack and aligned stack refer to the tilt series produced by coarse alignment using cross-correlation method and fine alignment using fiducial marker based method provided in IMOD, respectively. As can be seen from Fig.2.4A, the data points of the raw stack usually have larger standard deviation on each data point and multiple outliers are significantly offset from the fitting line (dash line), representing a very poor alignment quality for the raw stack. In contrast, Fig.2.4B shows most of the data points of the prealigned stack follow the linear trend with smaller standard deviation, though a few of them still slightly deviate from the fitting line, suggesting that the prealigned stack has a better quality of alignment. In addition, the data points of the aligned stack present a linear relationship with invisible standard deviation and perfectly match the fitting line, demonstrating that marker-based alignment is able to significantly improve the alignment quality in this case. The comparison shown in Fig.2.4A-C confirms that the goodness of linear trend
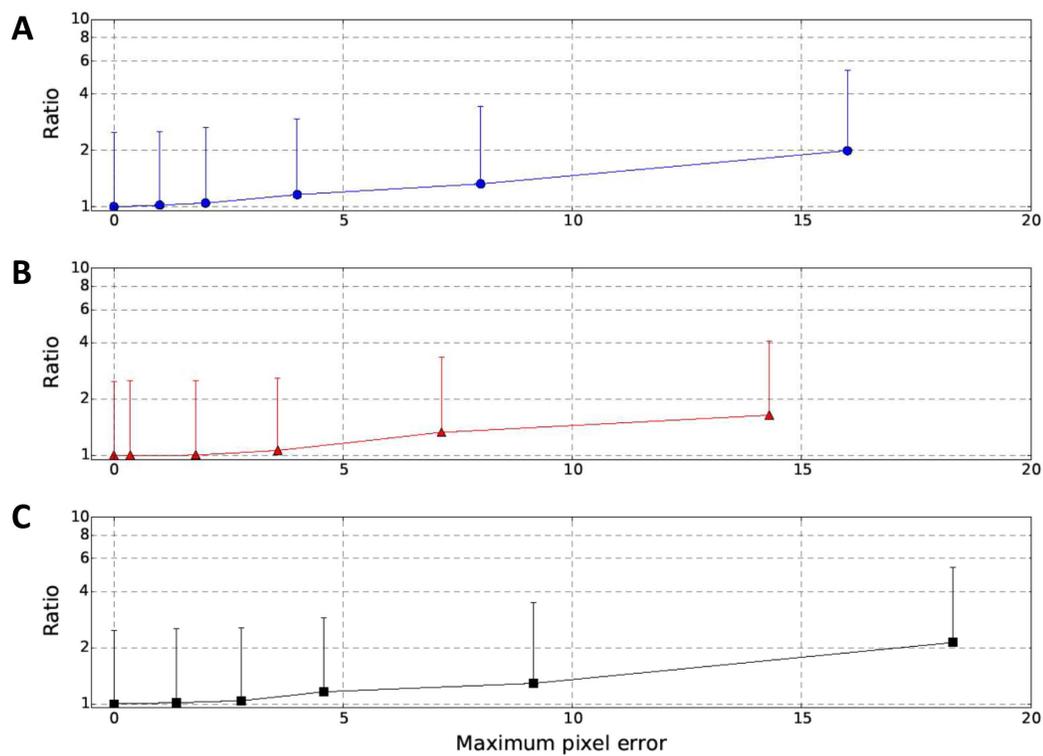
Fig. 2.3. Quantitative tests with computationally added alignment errors. A series of random errors, including shifts (A), in-plane rotations (B), and combination of shifts and in-plane rotations (C), were added to the aligned stack of stained cell section (same dataset shown in Fig.2.1, Fig.2.2). The X-axis values represent a maximum pixel error up to the value [18]. The points and error bars indicate mean and standard deviation of residuals, respectively. Note that the mean residual of the aligned stack is set as 1 here and all other mean and standard deviation are represented as ratios. A different random error was used for each image of the aligned tilt series.

of data points is correlated with the alignment quality, supporting the concept of our mathematical model.

To quantitatively analyze the alignment accuracy, we computed the residuals between data points and the fitting line (Eq.2.7) of each stack and compared them statistically. As shown in Fig.2.4D, we first sorted all residuals of one stack, set the mean residual of raw stack as 1, and used it as a reference to convert all residuals to ratios, aiming to intuitively show how much the residuals are reduced by coarse and fine alignment, respectively. We found those with largest residuals are mostly from high tilt angles. It is evident that, as expected, the raw stack has largest residuals (blue line in Fig.2.4D), while the aligned stack has the smallest residuals (red line in Fig.2.4D). These sorted plots also provide clear clues to the tilts with largest alignment errors that need to be further investigated. We also computed the mean and standard deviation of residuals of each stack and used them as a global figure of merit of alignment (Fig.2.4E). As shown in Fig.2.4E, the improvement of alignment quality from raw stack to aligned stack is accompanied with significant reduction of mean and standard deviation of residuals. The mean of residuals of final aligned stack is decreased to less than 2% of that of raw stack (Fig.2.4E). The marker based fine alignment is thus able to considerably boost the alignment accuracy over the correlation based prealignment. It is evident that our method is capable of evaluating and revealing the difference of alignment quality at different stages of alignment.

To further corroborate the effectiveness of our method, we validated our method with resolution criteria of reconstructed tomograms. Fig.2.4F depicts the $FSC_{e/o}$ curves of three tomograms, reconstructed from the raw stack, the prealigned stack and the aligned stack in this example, respectively. It can be seen that the resolution of tomograms is improved from raw stack to aligned stack, matching our analysis (Fig.2.4A-E) and confirming the effectiveness of our proposed method.
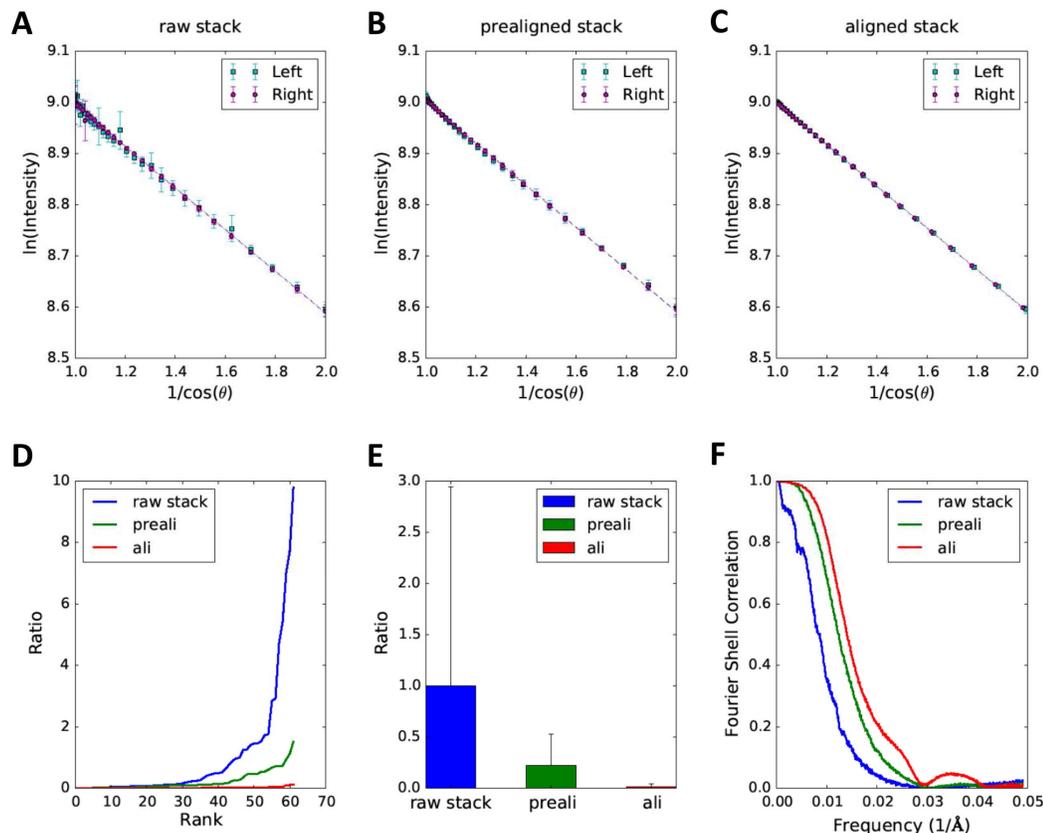
Fig. 2.4. Quantitative comparison of alignment quality for raw stack, pre-aligned stack and aligned stack of stained sample. The linear patterns of log intensity as a function of $1/\cos\theta$ are shown for the raw stack (A), prealigned stack (B) and aligned stack (C). (D) Sorted residual plot of the three stacks. Note that the mean residual of raw stack is set at 1 and all residuals are represented as ratios. (E) The bar graph illustrates the statistical analysis (mean $\pm$ standard deviation) of residuals from 10 selected regions. The column denotes mean and the error bar denotes standard deviation. The region size is $100 \times 100$ pixels. (F) The plot depicts resolution comparisons for tomographic reconstructions of raw stack, prealigned stack and aligned stack on the basis of $FSC_{e/o}$. The small dip at around 0.03 $\overset{\circ}{\text{A}}^{-1}$ in the curve of aligned stack (red curve) reflects the first CTF zero frequency of the mean defocus ($\sim 5.8\mu m$) for this dataset.

### 2.3.4 Quantitative assessment of alignment quality of cryo-ET images

We then tested if our method can also reliably work with cryo datasets, which are typically imaged with lower doses and are more noisy than stained sample images. Compared to the narrow spread of the data points around the fitting line for the stained example (Fig.2.4), the data points for the cryo datasets (region size $200 \times 200$ pixels) are widely scattered although a linear trend is still evident (Fig.2.5).

Due to the large spreads, it is difficult to visualize the difference of alignment quality directly from the log-linear plots of raw (Fig.2.5A) prealigned (Fig.2.5B) and aligned (Fig.2.5C) stack. However, the plot of residuals in Fig.2.5D indicates that the aligned stack attains the best alignment accuracy with the smallest residuals (red line in Fig.2.5D) among these three stacks, even though it is only marginally better than the prealigned stack (green line in Fig.2.5D). The statistical analysis in Fig.2.5E suggests that both coarse alignment and fine alignment are able to improve alignment quality of cryo datasets, and marker based fine alignment can achieve the best alignment accuracy in this case, even though it does not enhance the alignment accuracy as much as it does for the stained sample images (Fig.2.4E). The $FSC_{e/o}$ curves of prealigned and aligned stacks are indistinguishable in Fig.2.5F and fail to reveal the difference between their resolutions, which is likely due to the poor SNR of these images. We then visually examined the tomograms reconstructed from the prealigned stack (Fig.2.S1A, C, E) and final aligned stack (Fig.2.S1B, D, F). Based on the clearly worse missing wedge artifacts around the gold beads in the tomogram of prealigned stack (Fig.2.S1A, C, E) and the more symmetric artifacts in the tomogram of aligned stack (Fig.2.S1B, D, F), we conclude that the marker based alignment can indeed further improve the alignment quality of the coarse prealignment based on cross correlation although such minor improvements cannot be reliably detected by the $FSC_{e/o}$ resolution tests. However, such minor alignment improvements could still be detected by our method (Fig.2.5D and E).
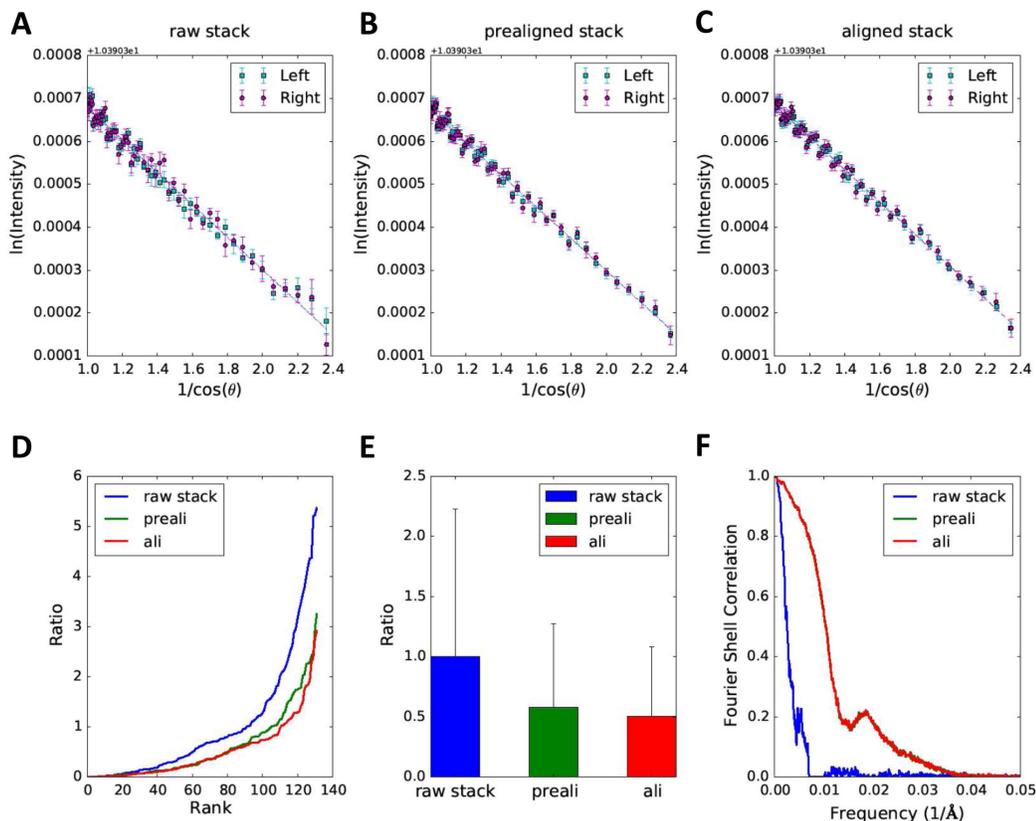
Fig. 2.5. Quantitative comparison of alignment quality for raw stack, pre-aligned stack and aligned stack of cryo dataset. The linear patterns of log intensity as a function of 1/cos are shown for the raw stack (A), pre-aligned stack (B) and aligned stack (C). (D) Sorted residual plot of the three stacks. Note that the mean residual of raw stack is set as 1 and all residuals are represented as ratios. (E) The bar graph illustrates the statistical analysis (mean $\pm$ standard deviation) of residuals from 16 selected regions. The region size is $200 \times 200$ pixels. (F) The plot depicts resolution comparisons for tomographic reconstructions of raw stack, prealigned stack and aligned stack on the basis of $FSC_{e/o}$.
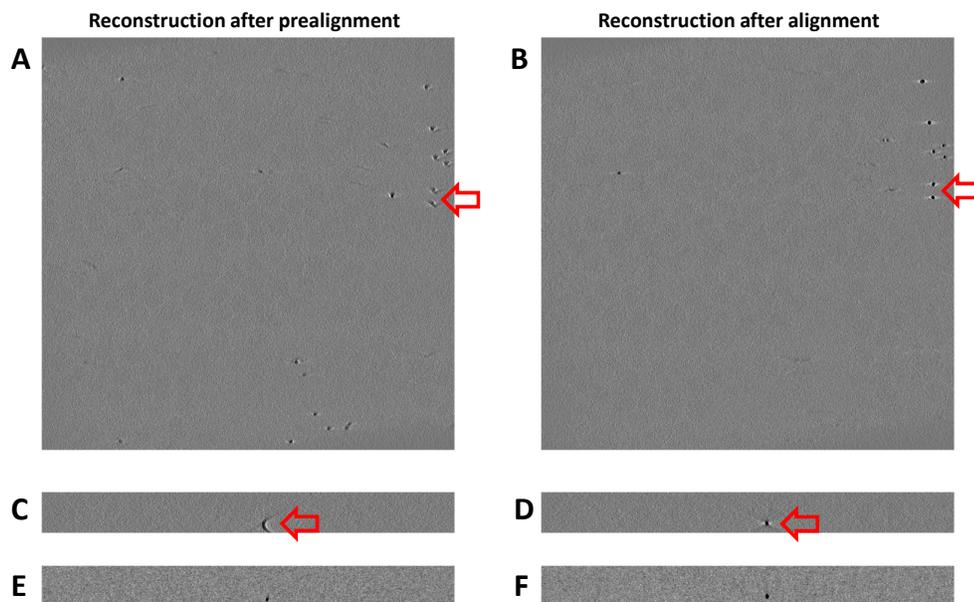
Fig. 2.S1. Comparison of reconstruction qualities of prealigned and aligned cryo tilt series shown in Fig.2.5. Sections (A, B for XY plane, C, D for XZ plane, and E, F for YZ plane) are shown for tomogram of prealigned (left column, A, C, E) and aligned (right column, B, D, F) stacks. Note that the reconstruction quality of gold beads in the prealigned stack reconstruction, as judged by the asymmetric missing-wedge artifacts around the beads, is worse than that of the beads in tomogram of aligned stack. This small quality improvement from marker-based alignment can be detected by our method (Fig.2.5D and E) but not by $FSC_{e/o}$ (Fig.2.5F).

### 2.3.5  Alignment inaccuracies

The above tests indicated excellent alignment quality for the stained sample (Fig.2.4) but significantly worse quality for the cryo sample (Fig.2.5). To investigate if such difference in alignment quality is a general phenomenon, we have further applied our method to additional stained sample datasets (Fig.2.6A-D) and cryo sample datasets (Fig.2.6E-H).

It can be seen that residuals of stained datasets were almost all minimized to very small values after pre and fine alignments. There were only occasional suboptimal alignments for a small number of tilt images (Fig.2.6D). In stark contrast, the alignment qualities for cryo datasets appear to be much poorer with large residuals after both pre and fine alignments for most datasets. For cryo datasets, most of the alignment improvements were achieved by correlation based prealignment while the marker based fine alignment could only provide small additional improvements. Compared to the apparently superior alignment quality for stained datasets, current methods appear to be inadequate in providing optimal alignment for cryo datasets.

To better understand the drastically different levels of residuals for stained and cryo datasets, we investigated the effect of several factors, number of regions (Figs.2.S2 and 2.S3), region sizes (Fig.2.S4) and adaptive change of region sizes at different tilt angles (Fig.2.S5). To test effects of noise and the efficacies of averaging of more regions in reducing the residuals (Figs.2.S2 and 2.S3), we applied different levels of random noise to the raw stack and then performed prealignment and marker-based alignment for each of the new stacks. For a fixed number of regions, increasing noise will increase, as expected, the residual of both prealigned and aligned stacks. For a fixed noise level, averaging of more regions can effectively reduce the residuals to negligible levels for stained samples (Fig.2.S2). However, significant levels of residuals remain even when large numbers of regions were used for cryo datasets (Fig.2.S3). These results suggest that the remaining residuals for aligned stacks of cryo datasets were not caused by the higher level of noises but more likely from systematic errors, for example, alignment

inaccuracies, that cannot not be reduced by averaging. To test the effect of different region sizes, we repeated the analysis for constant region sizes ranging from $50 \times 50$ to $300 \times 300$ pixels (Fig.2.S4). We found that the levels of residuals for each of the three stacks remain essentially the same for all these tested region sizes. Thus, region size is not a sensitive parameter for our method and an arbitrary size ($100 \times 100$ or $200 \times 200$ pixels) in this tested range should be fine in general. However, a user option was provided for our program to let the user specify a region size. To test if the extra contents from left/right side of the region in the tilted image would affect the performance of our method, we also tested variable region sizes by reducing the X-dimension according to the tilt angle ($N_x(\theta) = N_x(0)\cos\theta$) (Fig.2.S5). For all region sizes, the adaptive region size strategy (Fig.2.S5) returned very similar results as those by the simple constant region size strategy (Fig.2.S4). Thus, our method is robust for a wide range ($50 \times 50$ - $300 \times 300$ pixels) of both constant and adaptive region sizes. However, a user option was provided for our program to let user specify either constant or adaptive region size for the best performance of diverse samples.

### 2.3.6 Effects of defocus

In a tilted image, a defocus gradient exists in the direction perpendicular to the tilt axis. The overall defocus of different tilt images can also be different. These defocus variations can potentially cause the variation of pixel values and affect the performance of our method. To test the effect of defocus variation on the performance of our method, we selected multiple sets of regions with varying distances to the tilt axis. The first row in Fig.2.7 (Fig.2.7A-C) illustrates three different sets of locations: along the tilt axis, offset from tilt axis at halfway towards and close to the edges, which represent minimal, medium and maximal defocus variations, respectively. The alignment qualities for the stained (Fig.2.7D-F) and cryo (Fig.2.7G-I) samples were evaluated using our method for each of these three sets of locations. It is evident that the defocus variations have no significant effect on the evaluation of alignment
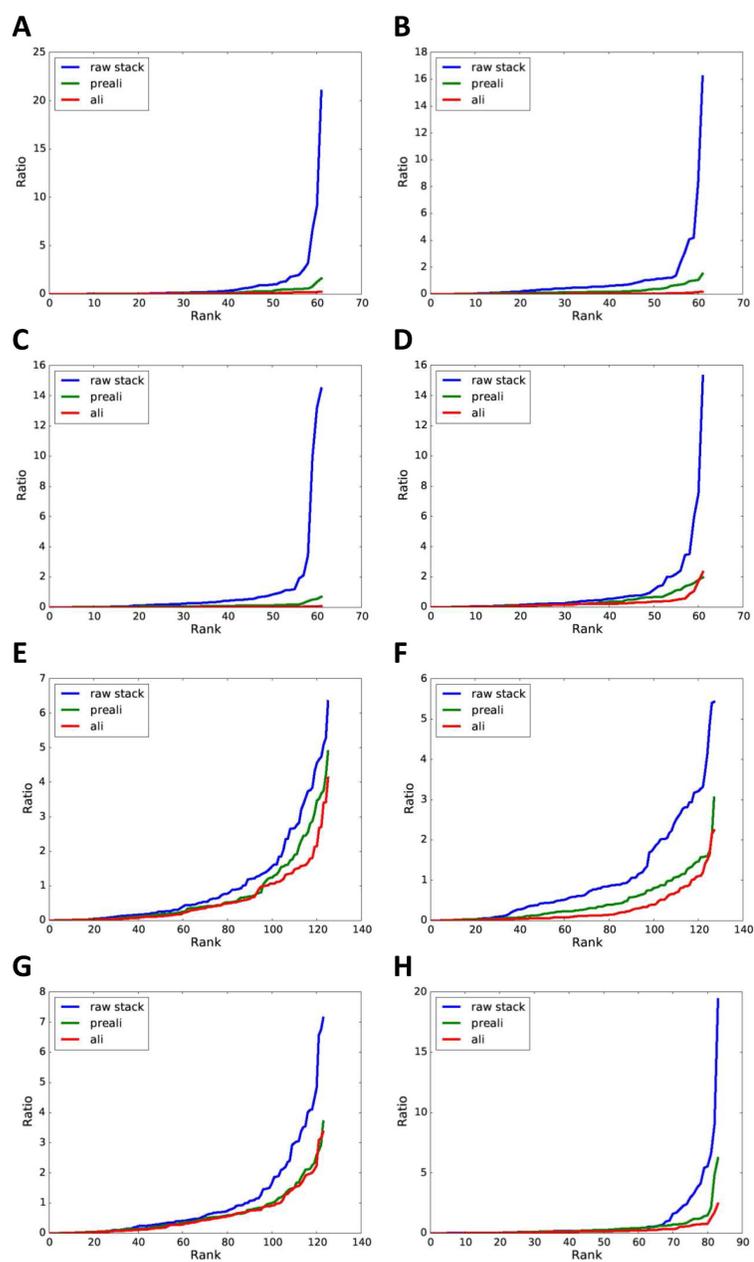
Fig. 2.6. Evaluations of alignment qualities of additional datasets. Shown are sorted residual plots of tilt series of additional stained (A-D) and cryo (E-H) samples. (A-C) Sindbis virus infected BHK cell sections embedded in resin and stained. (D) Flock House virus infected Drosophila S2 cell sections embedded in resin and stained. (E-G) Frozen-hydrated, purified Sindbis virus. (H) Frozen-hydrated B. *burgdorferi* cells.
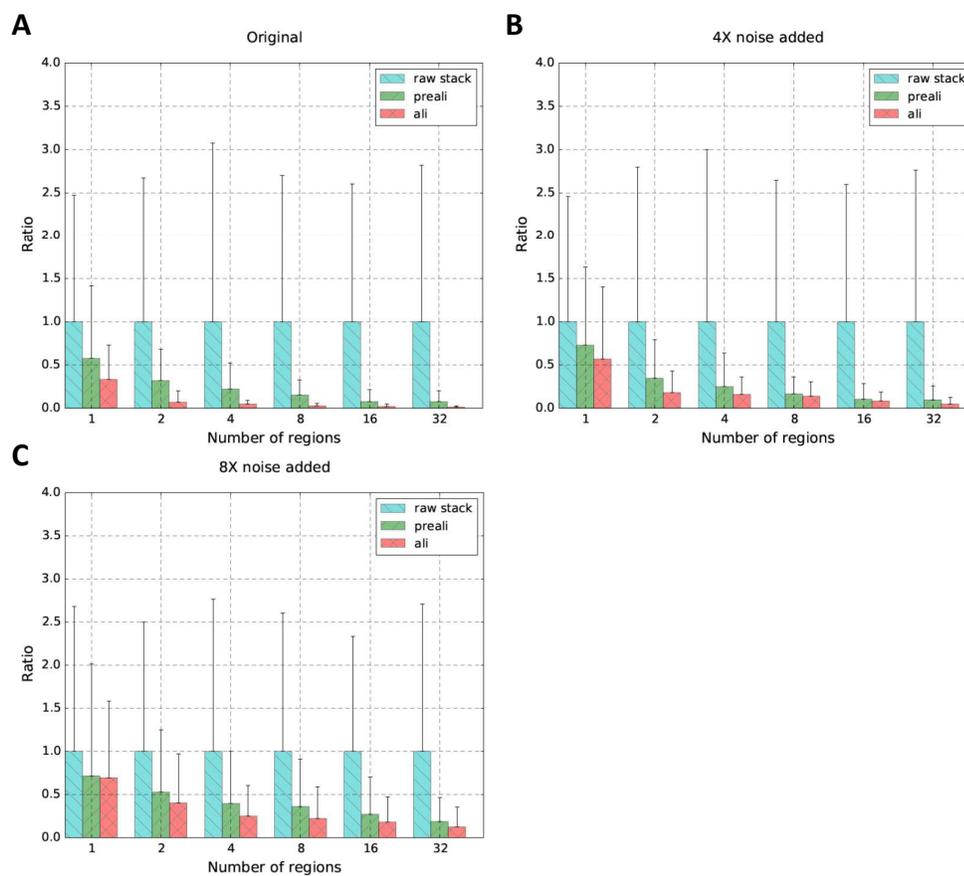
Fig. 2.S2. Performance with stained sample dataset at multiple levels of computationally added noise and different numbers of regions used for alignment quality evaluation. The region size is $100 \times 100$ pixels. We added different levels of random noise to raw stack and then aligned the noise-added raw stack. Here the random noise is Gaussian noise with zero mean and its standard deviation is 4 (B) or 8 (C) times of the standard deviation of the original image. The bar graphs compare alignment qualities of the original raw stack (A), stack with $4\times$ noise added (B), and stack with $8\times$ noise added (C).
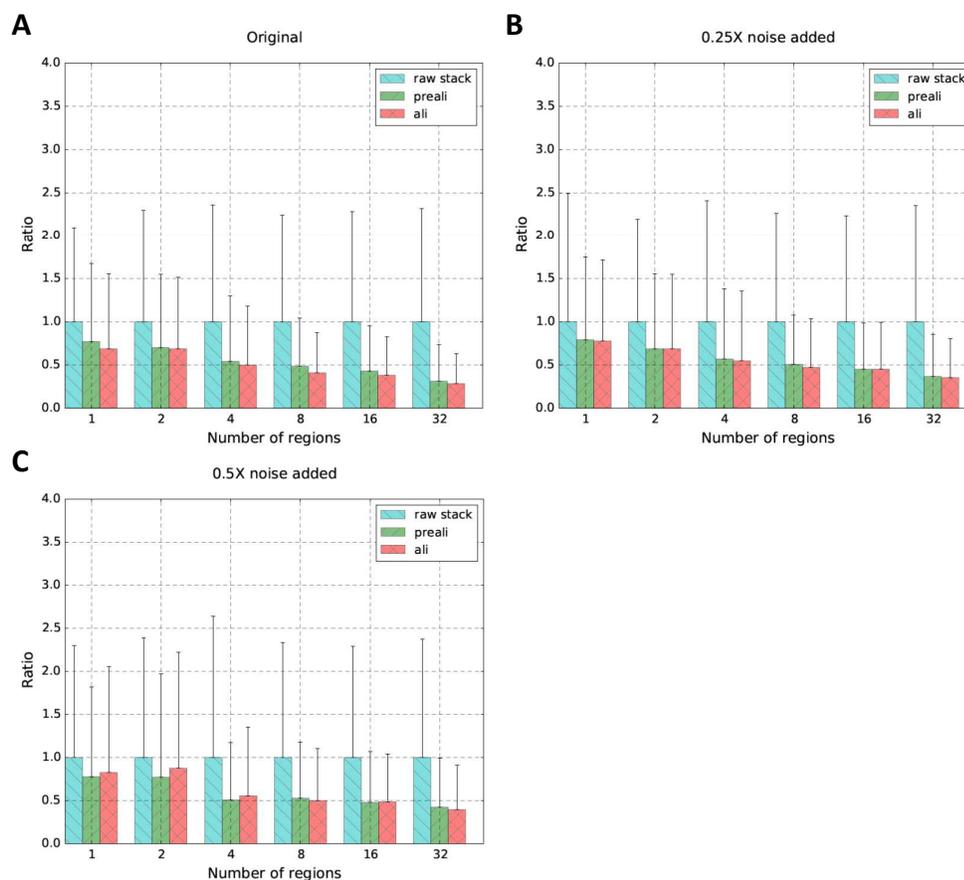
Fig. 2.S3. Performance with cryo dataset at multiple levels of computationally added noise and different numbers of regions used for alignment quality evaluation. The region size is $200 \times 200$ pixels. We added different levels of random noise to raw stack and then aligned the noise-added raw stack. Here the random noise is Gaussian noise with zero mean and its standard deviation is 0.25 (B) or 0.5 (C) times of the standard deviation of the original image. The bar graphs compare alignment qualities of the original raw stack (A), stack with $0.25\times$ noise added (B), and stack with $0.5\times$ noise added (C).
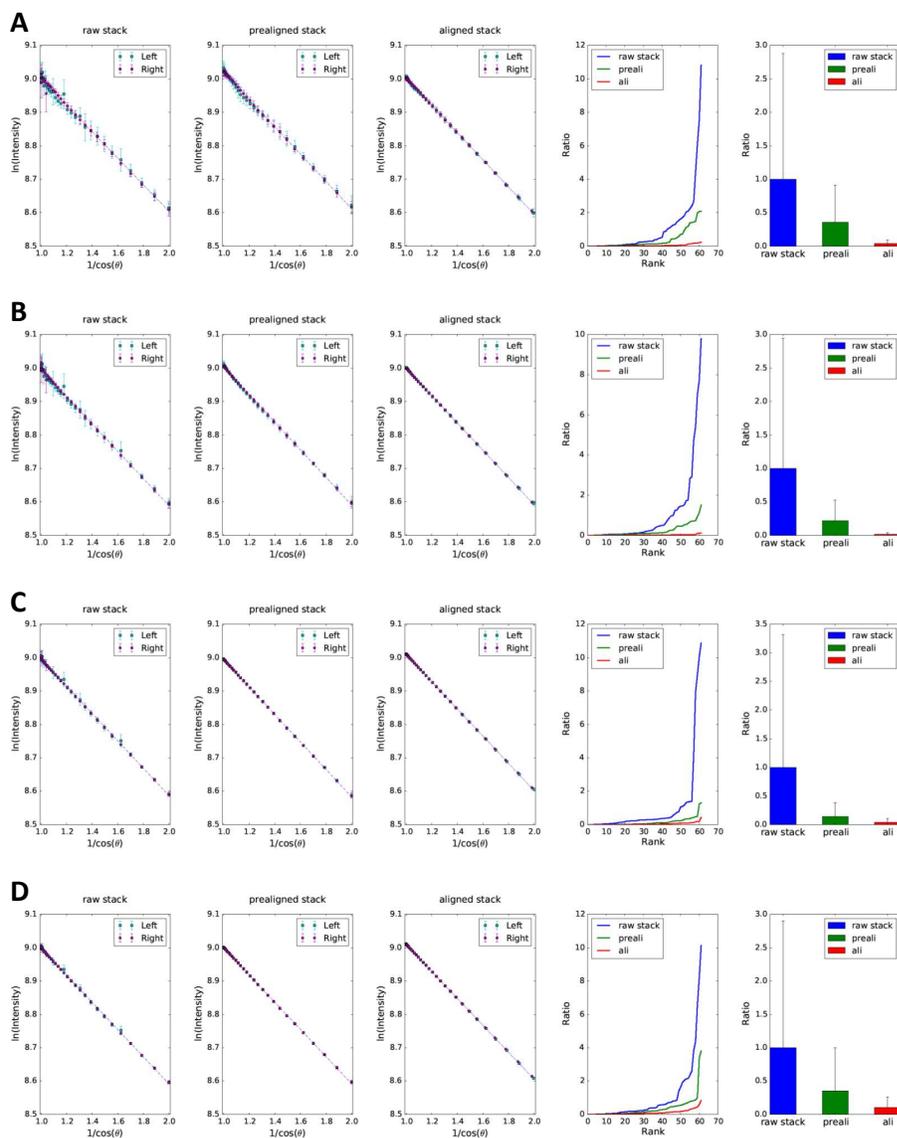
Fig. 2.S4. Performance with different constant region sizes. Using the same set of selected region locations in a dataset of stained sample (shown in Fig.2.1), we tested different region sizes, $50 \times 50$ (A), $100 \times 100$ (B), $200 \times 200$ (C), and $300 \times 300$ (D) pixels.
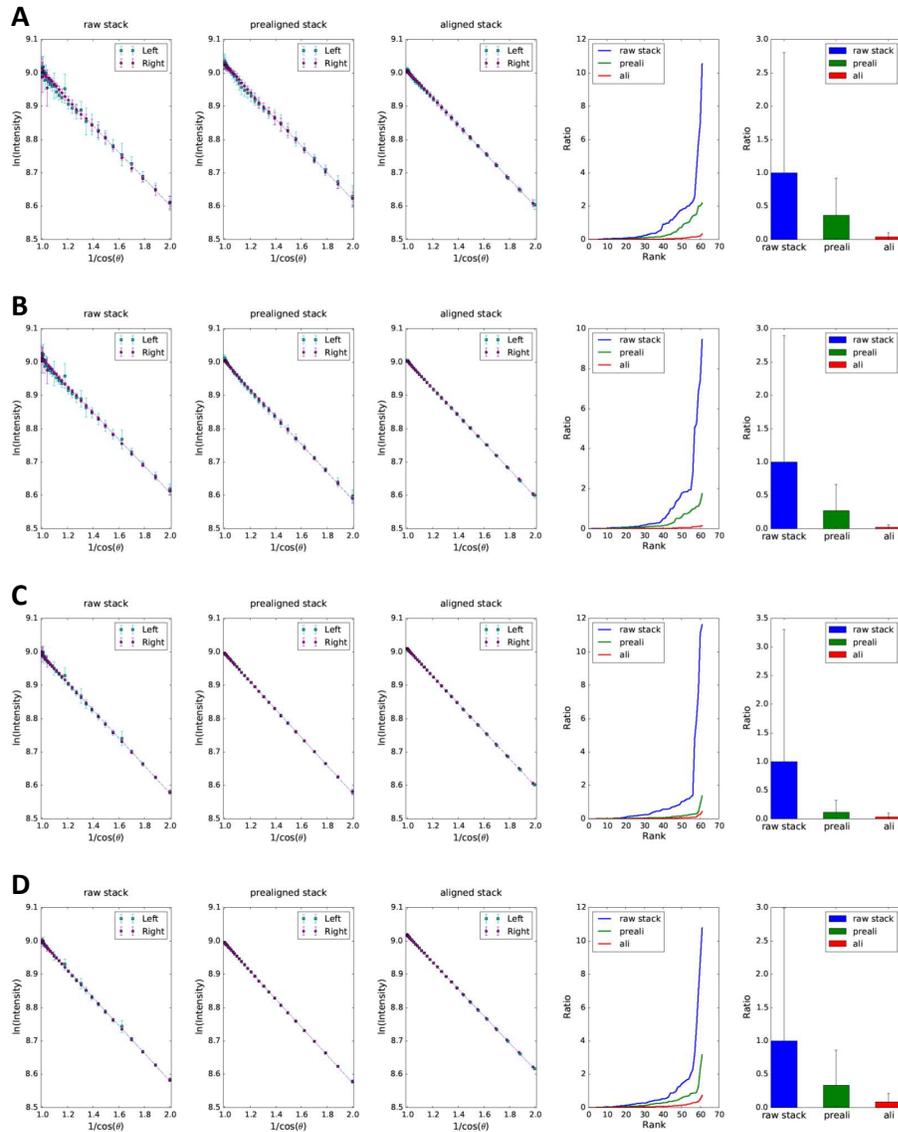
Fig. 2.S5. Performance with different variable region sizes adaptive to the tilt angles. Using the same set of selected region locations in a dataset of stained sample (shown in Fig.2.1), we tested different variable region sizes, $50 \times 50$ (A), $100 \times 100$ (B), $200 \times 200$ (C), and $300 \times 300$ (D) pixels. The X-dimension region sizes are varied according to the tilt angle $N_x(\theta) = N_x(0) \cos \theta$.

quality, since the method is able to correctly distinguish the alignment quality of raw stack, prealigned stack and aligned stack no matter where the regions are located. Hence, the results indicate that our method can perform robustly on the tilt series and the method is insensitive to defocus variations.

To understand the basis of our method's insensitivity to defocus variations, we used simulations to probe the effects of contrast transfer function (CTF) with varying defocuses on a simulated image with random noise of Gaussian distribution. We measured the mean and standard deviation of the CTF modulated images and plotted them as function of defocus in Fig.2.S6. The profiles of mean and standard deviation exhibit very different behaviors: the mean values remain nearly constant, while the standard deviation increases with defocus until reaching a plateau. The constant mean pixel value explains why our method is insensitive to defocus variations. Since both image contrast and image standard deviation measure the pixel value variations, we can use the standard deviation as a measurement of image contrast. The increased image contrast at large defocuses is thus consistent with the weak phase object image formation theory for TEM of thin biological specimens. The plateauing of image contrast also suggests that there is no need for excessively large defocus.

## 2.4  Discussion

In this chapter, we have introduced a new method, derived from the principle of the Beer-Lambert law (Eq.2.1), in order to establish a reliable cross-validation tool for alignment quality of tilt series. The underlying idea is to evaluate the alignment quality in terms of the goodness of linear relationship of data points calculated from tilt series (Eq.2.5 and Fig.2.2). The tests with both experimental stained and cryo samples (Figs.2.2-2.6) demonstrated that this method allows a robust and accurate assessment of alignment quality of electron tomographic tilt images. This method is fast and takes less than 1 min on a typical desktop computer for a tilt series. Compared to the $FSC_{e/o}$ resolution test [7], our new method is not only orders faster
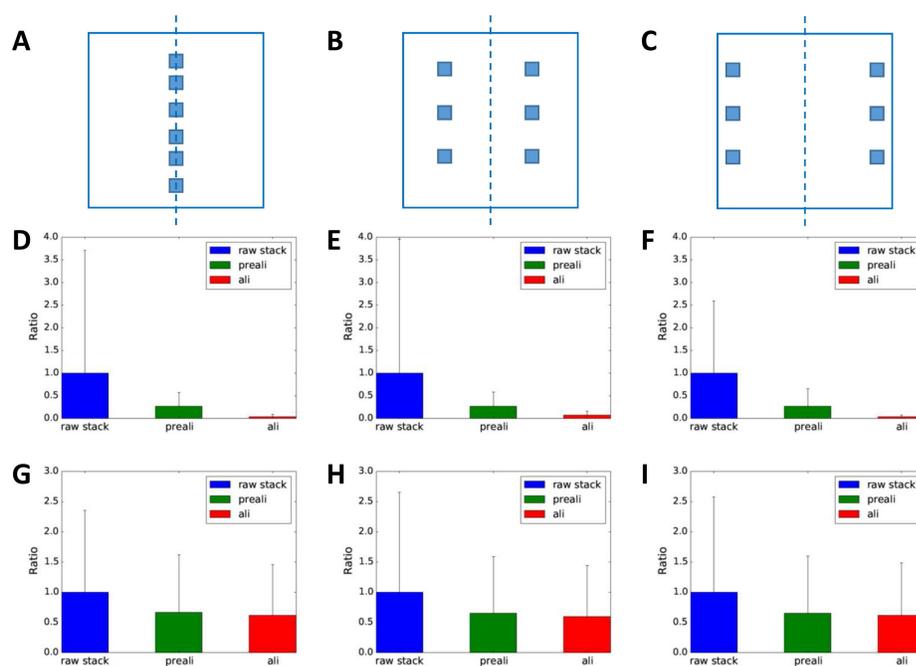
Fig. 2.7. Performances with varying defocuses. (A-C) The three panels represent three different sets of regions with different amounts of defocus gradient: minimum (A), intermediate (B) and large defocus gradient (C). The dash lines represent the tilt axis. (D-F) The three panels compare the residuals when multiple regions were selected as shown in (A-C) from one stained sample tilt series. (G-I) The three panels compare the residuals when multiple regions were selected as shown in (A-C) from one cryo sample tilt series.
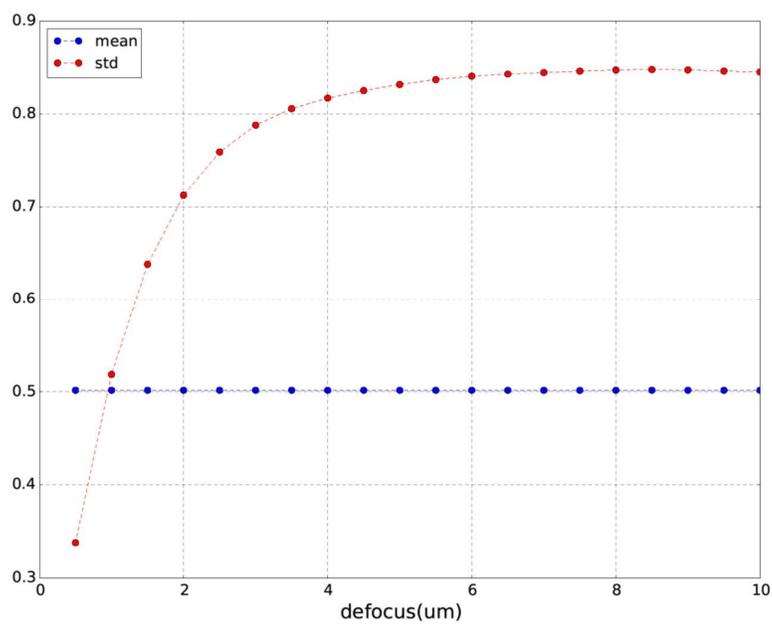
Fig. 2.S6. Defocus-dependence of image mean and standard deviation. CTFs with varying defocuses were applied to a simulated noise image. The CTFs used the following parameters: voltage = 300 kV, Cs = 2 mm, amplitude contrast = 0.1, B factor = 4000 $\text{Å}^2$, sampling = 5 Å/pixel, and image size of $256 \times 256$ pixels.

but also a direct method specifically for testing alignment instead of measuring the compound effects of image alignment and 3D reconstruction methods. This specificity for alignment might also explain its better sensitivity to alignment improvement than the $FSC_{e/o}$ resolution test (Fig.2.5). With these unique features, our new method can serve as an accurate and fast cross-validation of alignment quality and also as a guide for further optimizing the alignment before reconstructing 3D tomogram.

From our tests with both stained and cryo samples and the drastically different alignment qualities discovered for both types of images (Figs.2.2-2.6), we have also gained insights to the key factors for accurate alignments and the limitations of current alignment methods. If we assume that the residuals reported by our method consist of two sources, one from alignment errors and the other from the noise contribution to the pixel values, the minimal residuals for the stained samples suggest that both sources are negligible for these datasets. Further tests with larger number of regions and increased level of noises (Fig.2.S2) suggest that noise-related residuals can be effectively minimized by averaging of multiple regions used by our method. The high level of residuals for cryo datasets even after averaging a large number of regions thus suggests significant alignment errors for the cryo datasets. We think that inaccurate alignment might still be a major bottleneck for high resolution tomography. A similar conclusion was also independently reported based on other evidences [19, 20]. We hypothesize that the gold bead positions undergo a non-negligible amount of movement [21] during imaging when considering that vitreous ice undergoes melt/freeze cycles of phase changes induced by the electron beam [22] and the highly localized deposition of energy on the electron-dense gold. In contrast, images of stained section samples were taken at room temperature without fluidic phase and the samples were embedded in highly cross-linked resin. The gold beads are thus likely much more stable and allow much more accurate alignment for the stained samples. Based on these analyses, we suggest a new marker-free alignment method without relying on electron dense beads will be needed to provide better alignment accuracy for higher resolution tomograms of cryo samples. As shown in Fig.2.3, the residuals from the

linear fitting line are highly correlated with the magnitude of alignment errors. A direction to explore will be to use the residuals from the linear fitting line of tilt images not only as an evaluation criterion but also as a scoring function for optimization of the alignment parameters to minimize these residuals.

Our finding of the insensitivity of the mean pixel values to defocus variations (Figs.2.7 and 2.S6) is both satisfying and initially puzzling. This property removes a major complication related to defocus variations due to either defocus gradient within a single tilt image or change of mean defocus across different tilt images. We will understand it from the physical process of electrons passing through the sample and objective lens. While most electrons pass through the sample without being scattered, some of the electrons will be either elastically or inelastically scattered with ∼1:3 ratio in probability [23]. If we consider electrons as waves, the amplitude of the wave will be reduced by inelastic scattering but elastic scattering will only influence the phases. Since the beam intensity is only dependent on the wave amplitude but not the phase, the Beer-Lambert law used here thus primarily utilized the inelastically scattered electrons. On the contrary, image contrast is dominated by the defocus-dependent phase contrast resulted from phase modulations by objective lens, which do not change the wave amplitude and beam intensity. This property can also be explained using the CTF theory in image formation: the Fourier transform of the image is the Fourier transform of the sample multiplied by CTF and then corrupted by noises. In this Fourier formulation, the mean pixel value corresponds to the DC term (i.e. $F(0,0)$, the origin of the Fourier transform). Since varying defocus only changes the CTF oscillations at non-zero frequencies but not the DC term, the mean pixels values should stay constant at different defocuses.

# 3. SIMULTANEOUS DETERMINATION OF SAMPLE THICKNESS, TILT, AND ELECTRON MEAN FREE PATH USING TOMOGRAPHIC TILT IMAGES BASED ON BEER-LAMBERT LAW

## 3.1 Introduction

In recent years, cryo-electron tomography (cryo-ET) has emerged as a powerful imaging technique that allows 3D visualization of sub-cellular architecture and macromolecular organization in a near-native and fully frozen-hydrated state. Cryo-ET circumvents the deleterious effects from fixation, dehydration or staining [24, 25]. This technique bridges the gap of knowledge between cellular architecture revealed by low resolution light microscopy and high resolution structures of macromolecular complexes by single particle cryo-EM.

To prepare samples for cryo-ET, a commonly used technique is plunge freezing in which the sample solution is deposited onto a holey-carbon coated grid, blotted with filter paper, and vitrified by rapidly plunging into a cryogen (e.g. liquid ethane) cooled by liquid nitrogen [26, 27]. However, this method has poor control of thickness of the resulted vitreous ice. Another technique for cryo-sample preparation is cryo-sectioning in which the high pressure frozen thick specimen is trimmed using a diamond knife. Unfortunately, previous studies have shown that there is considerable variation in section thickness, especially for thin sections [28]. Focused ion beam (FIB) milling has gained considerable acceptance in recent years as a precision section preparation method. However, slice thickness by FIB still has unignorable variations due to differential thermal expansion [29], charging [30, 31], stage movements [32] or ion beam instabilities [30].

To obtain a high quality tomogram, it is essential to use volumes with sufficiently large Z-dimension to contain the entire sample. Due to lack of both thickness control during sample preparation and a method for reliable estimation of sample thickness, it is a common practice to use a large Z-dimension for 3D reconstruction. Alternatively, a trial reconstruction of a small region can be used to first determine the thickness before a full reconstruction is performed. However, the poor image contrast and the significant smearing along Z-axis due to missing wedge often render the sample boundaries hard to detect. It is thus desirable to have a method to reliably determine sample thickness to allow 3D reconstruction with optimal Z-dimension, which will not only eliminate the need for trial reconstruction but also improve the speed of reconstruction by reducing the tomogram volume to the smallest possible Z-dimension. Current methods for thickness measurement include measuring the shift of top/bottom surface projections of a small cylindrical hole through sample in a tilt pair image [33], image contrast between open and sample area [33], and electron energy loss spectroscopy [8, 9]. The first method is destructive to the sample while the last two methods rely on an accurately known electron mean free path [10]. Unfortunately, the mean free path is dependent on accelerating voltage and sample types, and can vary significantly as shown by drastically different values obtained by experimental measurements for a few sample conditions [34–36]. Different elemental compositions of the sample, for example, samples stained with heavy metals vs cryo biological samples mostly composed of low atomic number elements, can lead to a large change of mean free path [37, 38]. It is desirable to have a convenient method that can reliably estimate the mean free path of every sample.

It is a common assumption in cryo-EM that the specimen plane is horizontal and thus perpendicular to the electron beam. However, unintended sample tilts have often been observed due to multiple factors, including undulation of the carbon support film [39], instability of the sample holder mechanics, and insufficient reproducibility of goniometer [40]. These residual sample tilts were not detectable during data acquisition, which would result in systematic errors to the tilt angles assigned to all

images in a cryo-ET tilt series. The residual tilt, if not corrected, will lead to tilted structure in the 3D tomogram that requires larger Z-dimension to fully contain the structure and larger computing resource for reconstruction. For some specimens, for example, stained sections with gold beads coated on both surfaces, the 3D geometric model of the fiducial markers obtained from alignment of the whole tilt series can be used to determine/correct the residual sample tilt [17]. However, this approach will not be applicable to most cryo-ET samples without markers or with fiducial markers randomly distributed in the sample solution [6, 41].

In this study, we describe a computational approach that can simultaneously estimate sample thickness, tilt and inelastic mean free path using only the tilt images already collected for cryo-ET without need for additional data. This new approach employs a mathematical model derived from the Beer-Lambert law and estimates these parameters as a solution of a multi-variable overdetermined nonlinear least square problem with strong constraints provided by unique geometric relationships among the serial tilts of a common structure.

## 3.2   Methods

### 3.2.1   Mathematical model for thickness determination

The relationship between sample thickness and beam intensity can be represented by Eq.3.1 based on the Beer-Lambert law

$$\frac{d_e}{\lambda_{in}} = ln(\frac{I_0}{I_{exit}}) \tag{3.1}$$

where $d_e$ represents the effective thickness which is the distance that the electron beam travels through the specimen, $\lambda_{in}$ represents the mean free path for inelastic scattering, $I_0$ represents the intensity of the incident electron beam on the specimen and $I_{exit}$ represents the intensity of the electron beam exiting the specimen and hitting the detector.

It is obvious that the effective thickness $d_e$ in Eq.3.1 varies when the sample is tilted. Assuming the specimen is placed in an arbitrarily tilted plane in 3D space before serial tilting, we define the residual sample tilt $\gamma_0$ as the angle between the normal vector of this plane and the Z-axis (i.e. direction of electron beam). Here, the effective thickness $d_e$ for each tilt can be described in Eq.3.2 by taking the nominal tilt angles (i.e. intended tilt angles during data collection) and residual sample tilt into consideration

$$d_e = \frac{d_0 \cdot \cos\theta_0}{\cos\gamma_0 \cdot \cos(\theta_0 + \theta)} = d_0 \cdot \frac{\cos\theta_0 \cdot \sqrt{\tan^2\theta_0 + \tan^2\alpha_0 + 1}}{\cos(\theta_0 + \theta)} \qquad (3.2)$$

where $d_0$ represent the absolute geometric thickness of the specimen, $\theta$ represents the intended tilt angles around Y axis, $\theta_0$ and $\alpha_0$ represent the residual sample tilt around Y and X axis, respectively, and they can be measured via the corresponding side views of the 3D reconstruction.

Next, the exiting beam intensity $I_{exit}$ can be expressed as Eq.3.3 according to the linear relationship between $I_{exit}$ and pixel values of images

$$I_{image} = A \cdot I_{exit} + B \qquad (3.3)$$

where $I_{image}$ represents the average pixel value of the targeted area in the image. $A$ represents the gain factor of the detector. $B$ represents the average pixel value when no electron hits the detector. Although it should always be zero, we found some corner cases in which the detector is not properly gain-normalized or the image pixel values are shifted post imaging during alignment of the tilt series. To make our method robust for all datasets, this $B$ variable is included in our model as a nuisance parameter.

Hence, we can write our complete mathematical model for each tilt image as Eq.3.4 by combining Eqs.3.1-3.3.

$$\frac{d_0}{\lambda_{in}} \cdot \frac{\cos\theta_0 \cdot \sqrt{\tan^2\theta_0 + \tan^2\alpha_0 + 1}}{\cos(\theta_0 + \theta)} = ln(\frac{A \cdot I_0}{I_{image} - B}) \qquad (3.4)$$

In Eq.3.4, $I_{image}$ and $\theta$ are known from tilt series. The 7 unknown parameters are $I_0$, $d_0$, $\theta_0$, $\alpha_0$, $\lambda_{in}$, $A$ and $B$. The 4 parameters of interest in our study are $d_0$, $\theta_0$, $\alpha_0$ and $\lambda_{in}$ while the remaining three are nuisance parameters. In our case, this is a vastly overdetermined system since the number of equations (the number of selected regions in each image $\times$ the number of tilt images) is much larger than the number of unknowns (7) in the model.

### 3.2.2 Parameter determination as a constrained nonlinear least square problem

Parameter determination as a constrained nonlinear least square problem In order to obtain the solution of this overdetermined least square problem, we minimize the scoring function defined in Eq.3.5

$$f(I_0, d_0, \theta_0, \alpha_0, \lambda_{in}, A, B) =$$
$$\sum_{j=1}^{N} \sum_{i=1}^{M} \left[ ln(\frac{A \cdot I_0}{I_{image}(i,j) - B}) - \frac{d_0}{\lambda_{in}} \cdot \frac{\cos\theta_0 \cdot \sqrt{\tan^2\theta_0 + \tan^2\alpha_0 + 1}}{\cos(\theta_0 + \theta(i))} \right]^2 \quad (3.5)$$

where $i$ represents the $i$th image in the tilt series, $j$ represents the $j$th selected region, $M$ represents the total number of images in the tilt series, $N$ represents the total number of selected regions, and $\theta(i)$ represents the intended tilt angle of the $i$th image in the tilt series. We would first select $N$ regions of interest in the untilted image, and then track the corresponding areas in all tilts according to tilt geometry. The average pixel value of the $j$th region in the $i$th tilted image is denoted as $I_{image}(i,j)$.

Since each unknown variable has its own range that the solution must satisfy, we can estimate the solution of unknowns in Eq.3.5 by solving the following constrained nonlinear optimization problem

$$\min_{I_0,d_0,\theta_0,\alpha_0,\lambda_{in},A,B} f(I_0, d_0, \theta_0, \alpha_0, \lambda_{in}, A, B)$$

subject to

$$A \cdot I_0 > I_{image}(i,j) - B > 0 \qquad (3.6)$$

$$d_0 > 0$$

$$\lambda_{in} > 0$$

where $f(I_0, d_0, \theta_0, \alpha_0, \lambda_{in}, A, B)$ is defined in Eq.3.5.

### 3.2.3 Implementation

Our approach has been implemented into *tomoThickness.py*, a single python script for easy usage. We used EMAN2 library functions [12] for image IO. *numpy/scipy* software [13, 14] was used to perform nonlinear least square fitting and *matplotlib* software [15] was used for plotting. There are many well-developed algorithms and software packages available to solve the constrained nonlinear optimization problem. The global minimization method basin-hopping in *scipy.optimize* module was used to find solutions to the 7 unknown parameters in less than a minute for each tilt series. Although we have only tested this program on Linux systems, the program should run on all major computer platforms (Linux, Windows and MacOS) since the dependent software packages (python, numpy, scipy, matplotlib, and EMAN2) are all compatible with these platforms. The *tomoThickness.py* program is freely downloadable from the authors' Web site (http://jiang.bio.purdue.edu).

### 3.2.4 Test datasets

In this study, we used both stained and cryo tilt series to test the performance of our method. Sindbis virus infected Baby Hamster Kidney fibroblasts (BHK) cells and Flock House virus infected *Drosophila* S2 cells were embedded in resins, sectioned, and then stained. Tilt series of both samples were obtained on a FEI Titan Krios at

300 kV and were recorded on a Gatan 4 K × 4 K CCD camera at room temperature. Before automated tilt series collection using SerialEM [42], the samples were first pre-irradiated ($\sim 10^4 e/\text{Å}^2$) to stabilize the resin and minimize specimen shrinkage during data collection [16]. Tilt images were collected from -60° to +60° in 2° increments with constant dose for each tilt and a total dose of 4000 $e/\text{Å}^2$. The pixel size was 0.51 nm for the images of Sindbis virus infected Baby Hamster Kidney fibroblasts (BHK) cell sections and 0.404 nm for Flock House Virus infected *Drosophila* S2 cell sections. One group of cryo-datasets are the images of purified Sindbis virus plunge frozen and imaged at liquid nitrogen temperature. The cryo tilt series were obtained using a FEI Polara TEM at 300 kV, and were recorded on a Tietz 4K × 4K CCD. Tilt images were taken from -60° to +60° in 1° increment with constant dose for each tilt and a total dose of 100 $e/\text{Å}^2$. The images were 2× binned to 2048 × 2048 pixels with final pixel size of 0.55 nm. The other group of cryo-datasets are the tilt series of DNP-labeled liposomes and anti-DNP antibody downloaded from EMPIAR (EMPIAR-10009) [43]. This group of datasets were imaged on a FEI Tecnai 200 kV microscope.

The tilt series were aligned using gold fiducial markers with the IMOD software package [17]. IMOD was also used for visualization of the tilt series and tomograms [44]. The parameters of interest including sample thickness and tilt were verified by examining the corresponding side views of the reconstructed tomogram [11, 44].

## 3.3   Results

### 3.3.1   Tilt dependent variation of image intensity

To determine sample parameters using the tilt dependent variation of image intensity, we first selected a region in the zero-tilt image then tracked this region through the aligned stack of tilt series according to the tilt angles. We usually selected regions without extremely dark contents. Fig.3.1 shows a region selected for a stained

(Fig.3.1A) and a cryo (Fig.3.1B) tilt series, respectively. We calculated the average pixel values in the tracked patches and used them to estimate the sample parameters.
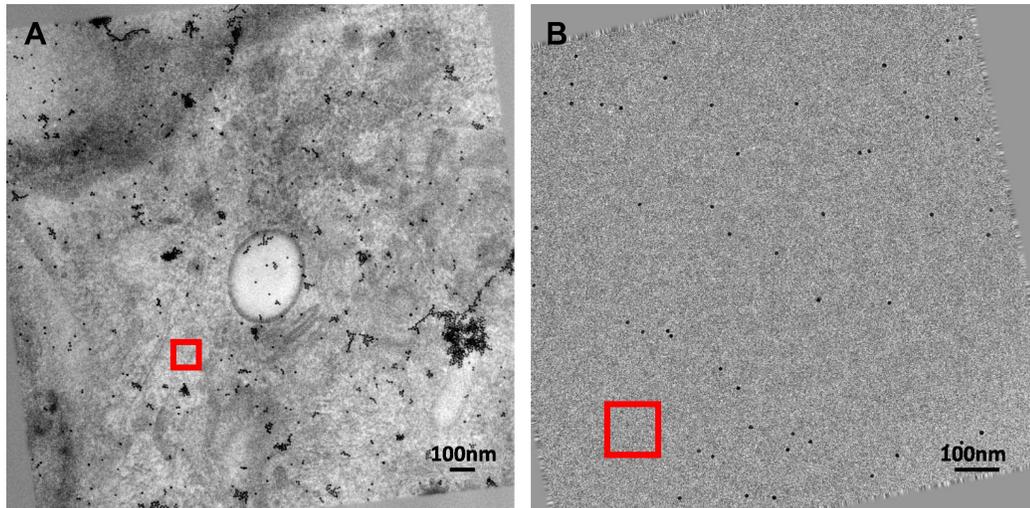


Fig. 3.1. Zero tilt image of aligned tilt series. (A) Stained Sindbis virus infected BHK cell section. (B) Plunge-frozen Sindbis virus. The tilt axis is vertical. The red squares indicate regions selected and tracked through the tilt series. The region size is $200 \times 200$ pixels.

The effective sample thickness is directly associated with the tilt angles. When the specimen is tilted toward larger angles, the effective sample thickness increases, leading to more attenuation of the electron beam passing through specimen and hitting the detector ($I_{exit}$ in Eq.3.1). The relationship between image intensity and tilt angle of a stained tilt series is shown in Fig.3.2A in which the gradual decrease of intensity (i.e. average pixel value, $I_{image}$ in Eq.3.3) with increasing tilt angle is qualitatively consistent with the theoretical prediction (Eq.3.1). However, close inspection of Fig.3.2A revealed that the curve was asymmetric about $0°$ tilt angle, inconsistent with the expected symmetric shape.

Based on Eqs.3.1-3.3 and assuming that residual tilt is $0°$ ($\theta_0 = 0°$, $\alpha_0 = 0°$) and that there is no linear scaling of pixel values ($A = 1$, $B = 0$), we can obtain Eq.3.7:
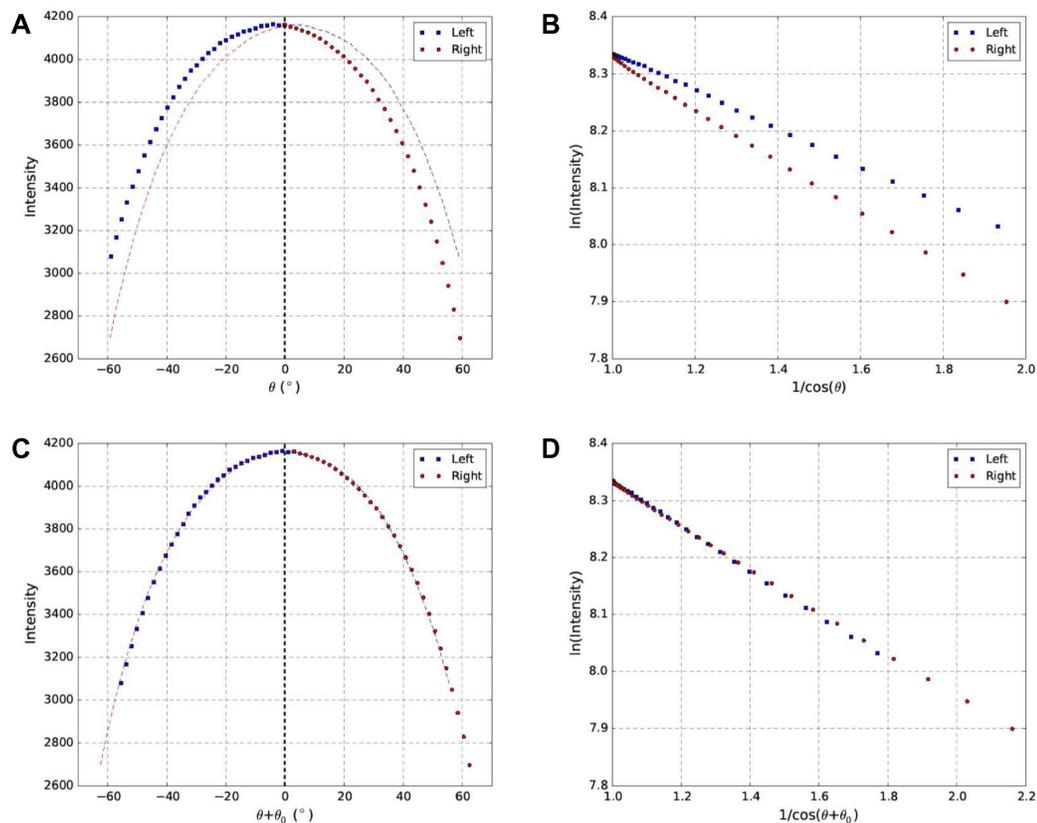
Fig. 3.2. Relationship between image intensity and tilt angles of a stained dataset before and after correction of residual tilt. (A) Profile of image intensity and intended tilt angle ($\theta$) for the red square shown in Fig.3.1A. (B) Replot of (A) using log of intensity (Y-axis) and $1/\cos\theta$ (X-axis). (C) Profile of image intensity and corrected tilt angle ($\theta + \theta_0$) after estimation of parameters of interest. (D) Replot of (C) using log of intensity (Y-axis) and $1/\cos(\theta + \theta_0)$ (X-axis). To provide visual guide for the asymmetric (A) or symmetric (C) distributions about 0° tilt angle, a mirror copy of both sides of 0° tilt angle was also plotted in the opposite side with dash curves in corresponding colors.

$$ln(I_{exit}) = ln(I_0) - \frac{d_0}{\lambda_{in}} \cdot \frac{1}{\cos\theta} \tag{3.7}$$

From Eq.3.7, we expect that log of image intensity will be a linear function of $1/\cos\theta$. To better illustrate the tilt-intensity relationship, we thus replotted Fig.3.2A as log of intensity (Y axis) with $1/\cos\theta$ (X axis) shown in Fig.3.2B. It was obvious that linear relationship did exist. However, the two straight lines representing the left half (blue squares) and right half (red circles) of the curve in Fig.3.2A were separated with different slopes. Similar to the curves in Fig.3.2A with unexpected asymmetric shape, this line separation was also puzzling as two opposite directions of tilting should result in an identical change of effective thickness and overlapping straight lines in the log-linear plots in Fig.3.2B and symmetric about 0° tilt in Fig.3.2A. Electron beam induced mass loss [16] cannot be the cause as smaller slope would be expected for the line corresponding to the right half tilts which were imaged last and would cause less attenuation of the beam after mass loss. We suspected that residual sample tilt caused these unexpected results, which led us to develop the nonlinear least square fitting method reported here to resolve these issues. As shown in results below, residual sample tilt was indeed found to be the cause.

### 3.3.2 Determination of sample thickness, tilt and electron mean free path

We applied our nonlinear least square fitting algorithm (Eq.3.6) to the tilt series in order to determine the parameters of interest (sample thickness, tilt and electron mean free path). As shown in Fig.3.2C, the plot of intensity became more symmetric about 0° tilt after correcting the tilt angles with residual sample tilt ($\theta_0 = 3.38°$). As shown in Fig.3.2D, the two separated lines now converged to a single line after taking the residual sample tilt ($\theta_0 = 3.38°$) into account. We thus have confirmed that residual sample tilt was indeed the cause for the asymmetric curve in Fig.3.2A and line separation in Fig.3.2B. Furthermore, it also demonstrated that our nonlinear

least square method could be used to correctly determine the residual sample tilt from a high contrast stained tilt series.

We then applied our method to analyze a cryo tilt series (Fig.3.3) which showed similar results with Fig.3.2. The cryo dataset was much noisier with lower contrast, as expected, since the signal-to-noise ratio (SNR) of cryo tilt series (Fig.3.3) imaged at much lower electron dose is poorer than that of the stained dataset (Fig.3.2). Nevertheless, the residual tilt could still be accurately determined to make the curve symmetric (Fig.3.3C) and the lines to converge (Fig.3.3D).

Our method could use arbitrary number of regions (Eq.3.5, $N \geq 1$) from the tilt series to estimate a single set of parameters. Fig.3.S1 and Fig.3.S2 extended those in Fig.3.2 and Fig.3.3 and showed curves/lines for two different regions in the corresponding stained and cryo tilt series, respectively. It was evident that the originally separate lines for each region converged after considering a common residual sample tilt parameter for both regions. The converged lines for the two regions remain separate due to the different local sample contents. Together, we showed that our method worked well for both stained and cryo tilt series.

### 3.3.3 Cross validation with measurements from 3D tomogram

To verify these fitted parameters with independent measurements, we reconstructed both tilt series shown in Fig.3.2, Fig.3.3 and manually measured the corresponding parameters from side views of the tomograms. Fig.3.4 shows the measured sample tilt $\theta_0 \approx 3.3°$ (Fig.3.4A) and $\alpha_0 \approx 0°$ (Fig.3.4B), and thickness $d_0 \approx 170$ nm, which agree well with the fitting results $\theta_0 = 3.38°$, $\alpha_0 = 0.17°$, $d_0 = 168$ nm, $\lambda_{in} = 230$ nm of this stained dataset used in Fig.3.2. To further demonstrate that the fitted sample tilt angle is accurate, we modified the tilt angles by adding the fitted $\theta_0$ and then performed reconstruction with the modified tilt angles. In Fig.3.4C the sample structure now become horizontal in XZ view, implying that the estimated $\theta_0$ is correct. Similarly, compensating the residual tilt $\alpha_0$ around the X-axis also resulted
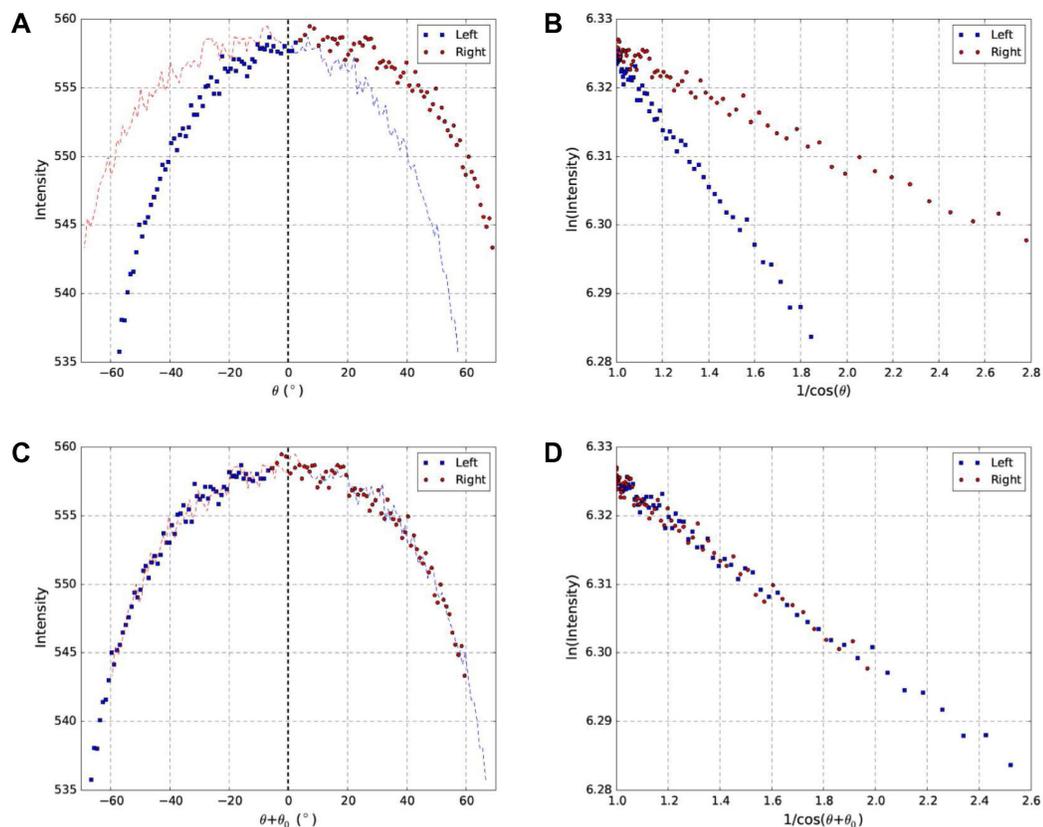
Fig. 3.3. Relationship between image intensity and tilt angles of a cryo dataset before and after correction of residual tilt. (A) Profile of image intensity and intended tilt angle ($\theta$) for the red square shown in Fig.3.1B. (B) Replot of (A) using log of intensity (Y-axis) and $1/\cos\theta$ (X-axis). (C) Profile of image intensity and corrected tilt angle ($\theta + \theta_0$) after estimation of parameters of interest. (D) Replot of (C) using log of intensity (Y-axis) and $1/\cos(\theta + \theta_0)$ (X-axis). To provide visual guide for the asymmetric (A) or symmetric (C) distributions about 0° tilt angle, a mirror copy of both sides of 0° tilt angle was also plotted in the opposite side with dash curves in corresponding colors.
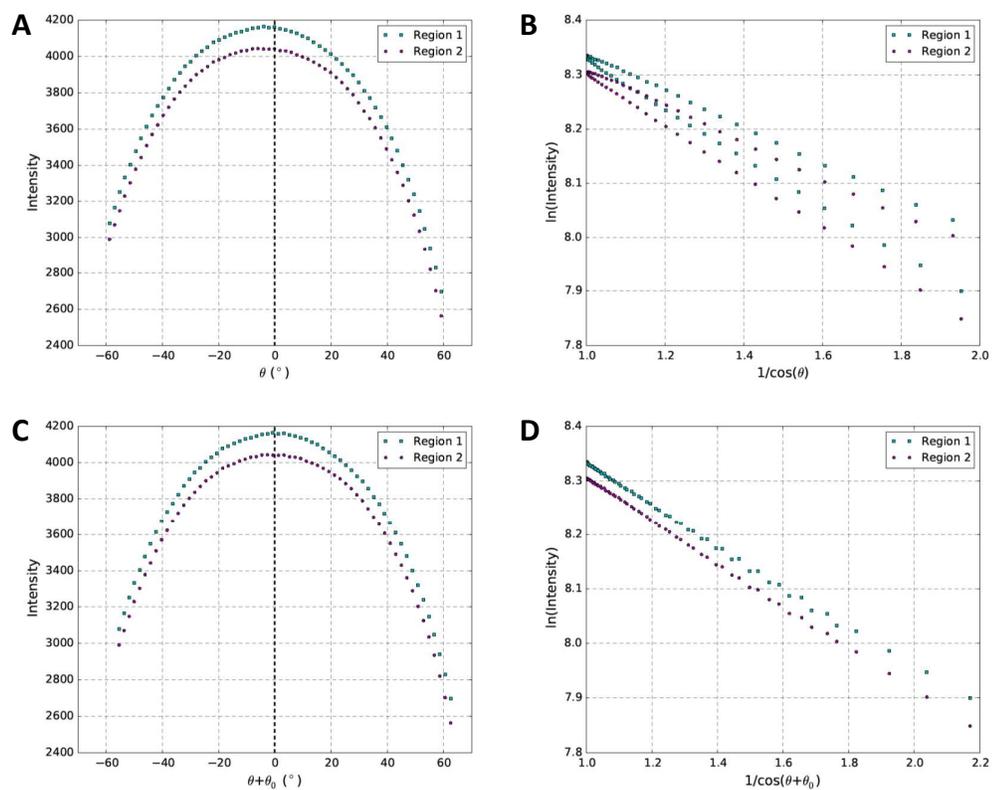
Fig. 3.S1. Expansion of plots in Fig.3.2 to include two regions for the stained tilt series. Note that the accurate fitted residual tilt can help converge the two separate lines with different slopes (B) into well-superimposed lines (D).
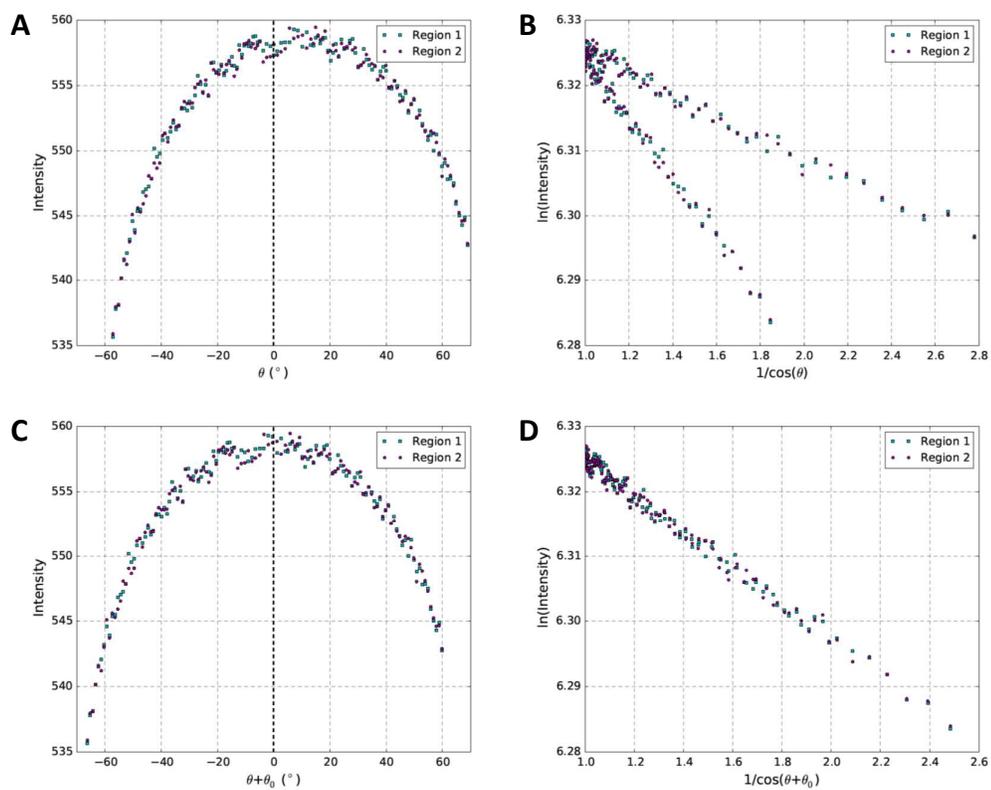
Fig. 3.S2. Expansion of plots in Fig.3.3 to include two regions for the stained tilt series. Note that the accurate fitted residual tilt can help converge the two separate lines with different slopes (B) into well-superimposed lines (D).

in horizontal structure in YZ view (Fig.3.4D), although $\alpha_0$ for this tilt series is small and the original tomogram without this correction (Fig.3.4B) was already visually horizontal.

Repeating these tests with the cryo dataset used in Fig.3.3 further confirmed that the fitted parameters were also correct. The parameters $\theta_0 \approx 8.5°$ (Fig.3.5A) and $\alpha_0 \approx 2°$ (Fig.3.5B), and $d_0 \approx 72$ nm measured from tomogram also matched well with the fitted parameters $\theta_0 = 8.7°$, $\alpha_0 = 2.12°$, $d_0 = 74.6$ nm, $\lambda_{in} = 356$ nm and the correction of sample tilt also made the specimen horizontal in XZ (Fig.3.5C) and YZ views (Fig.3.5D).

It is worth pointing out that the fitted inelastic mean free path for vitreous ice $\lambda_{in} = 356$ nm is very close to the predicted value 350 nm in the literature [36] although no experimental values have been published for $\lambda_{in}$ at 300 kV. In addition, the fitted $\lambda_{in}$ for stained dataset is 230 nm which is significantly smaller than that for cryo dataset. While there is no experimentally measured $\lambda_{in}$ for stained sections, the reduced value is consistent with expectation: electrons are scattered with much higher probability by the heavy metal materials in the stained sample than by the light atoms in cryo samples [37].

### 3.3.4 Robustness of basin-hopping optimization method

As described in Section 3.2.2, we used a basin-hopping method to solve the constrained nonlinear optimization problem. Basin-hopping is a stochastic global optimization algorithm that has been shown to be extremely useful for problems in physics and chemistry. During the iterative search, the basin-hopping algorithm introduces random perturbation of the search variables to escape local minima and find the global minimum efficiently. To test the robustness of our approach, we reran the script 100 times on both stained and cryo datasets. The results in Fig.3.S3 showed that sample thickness $d_0$ (Fig.3.S3A), mean free path $\lambda_{in}$ (Fig.3.S3B), sample tilt ($\theta_0$, $\alpha_0$ in Fig.3.S3C, $\gamma_0$ in Fig.3.S3D) could be correctly determined in all runs with little

48
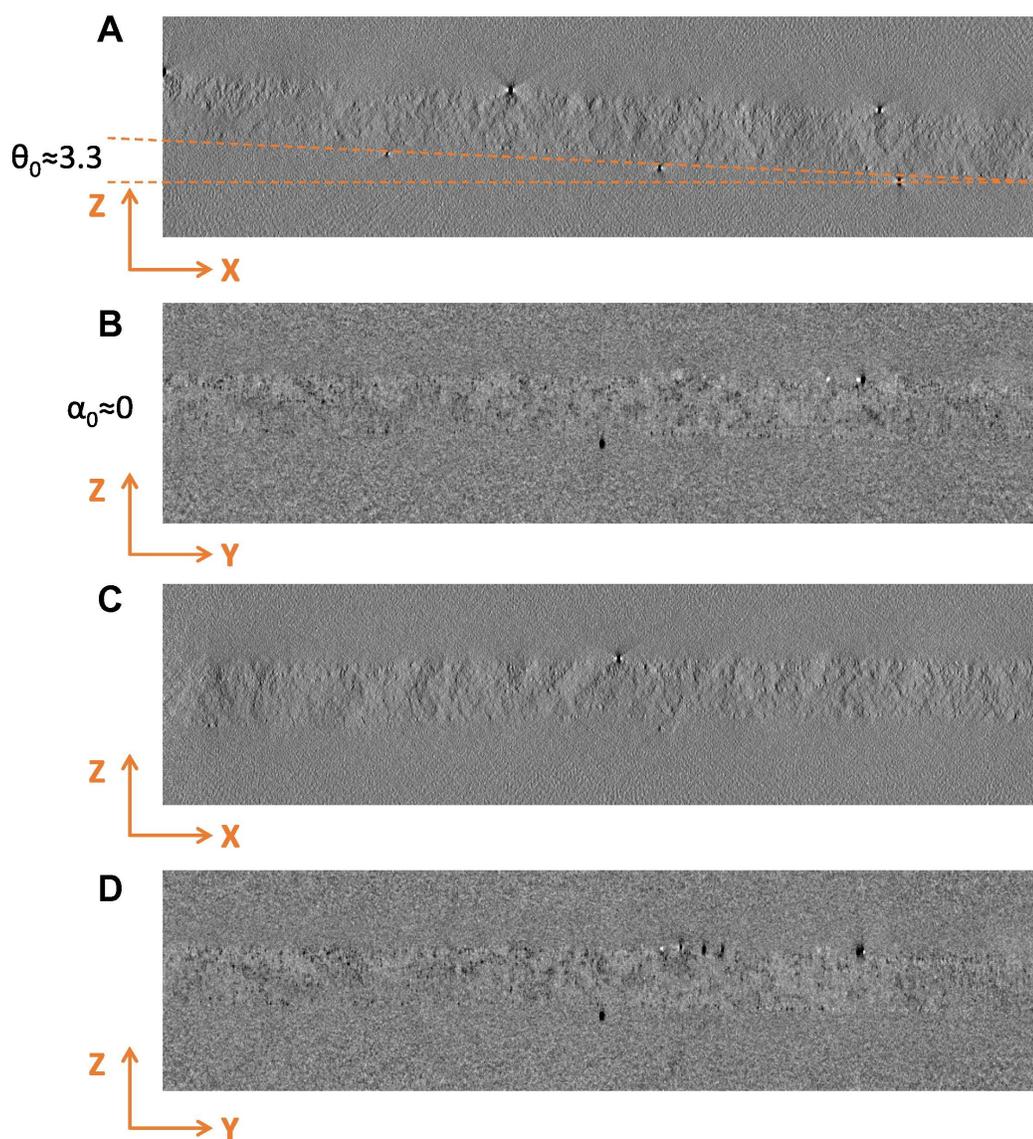


Fig. 3.4. Comparison of reconstruction of the stained tilt series shown in Fig.3.1A before and after correction of residual tilt. Sections (A) and (B) are XZ and YZ planes before tilt correction. Sections (C) and (D) are XZ and YZ planes after tilt correction. Note that the sample plane is horizontal after tilt correction (C, D) and the measured thickness is close to fitted value, indicating that the fitted parameters are reliable for stained datasets.
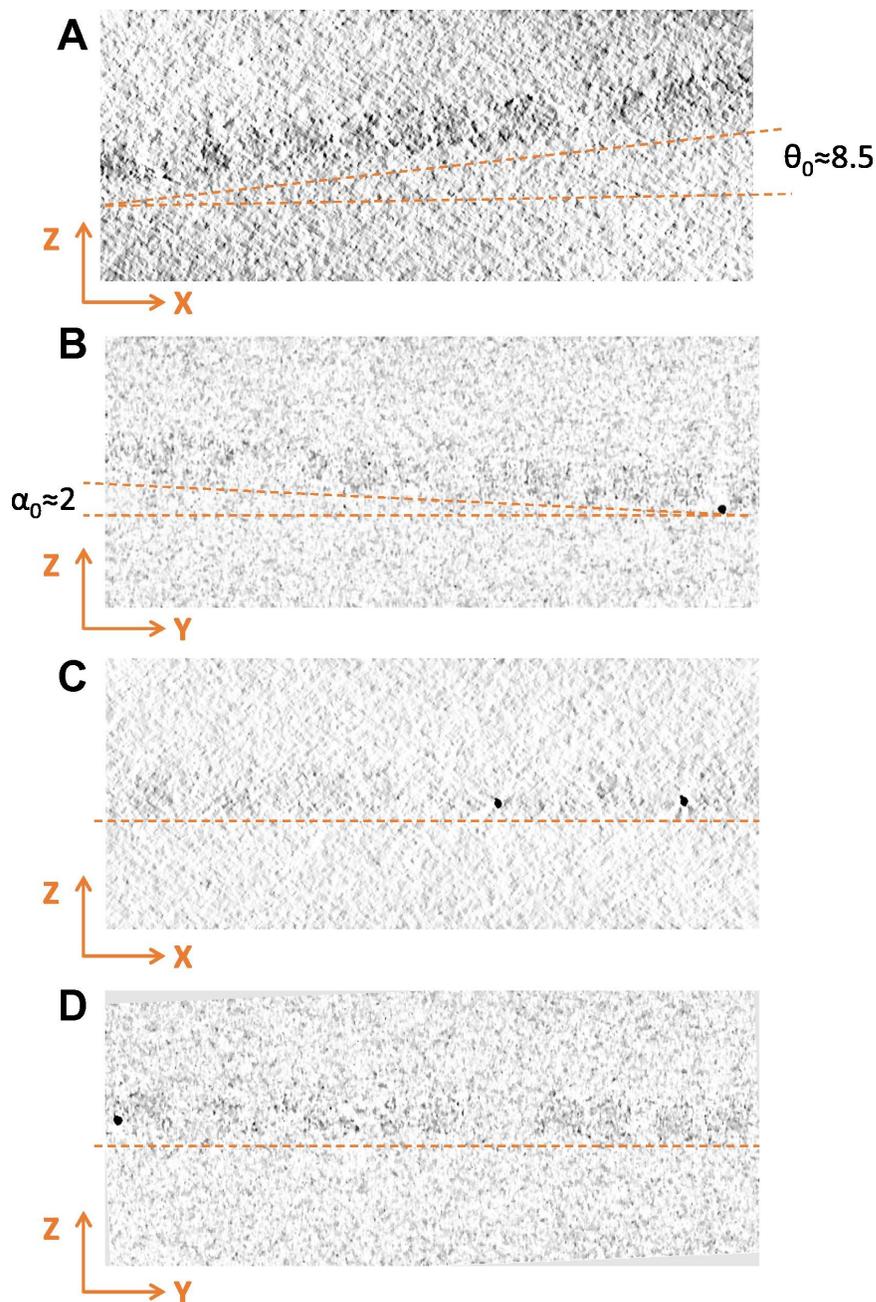
Fig. 3.5. Comparison of reconstruction of the cryo tilt series shown in Fig.3.1B before and after correction of residual tilt. Sections (A) and (B) are XZ and YZ planes before tilt correction. Sections (C) and (D) are XZ and YZ planes after tilt correction. Note that the sample plane is horizontal after tilt correction (C, D) and the measured thickness is close to fitted value, indicating that the fitted parameters are reliable for cryo datasets.

variations among different runs. These results implied that the basin-hopping algorithm employed in our approach can reliably find the global minimum and successfully determine the sample parameters.
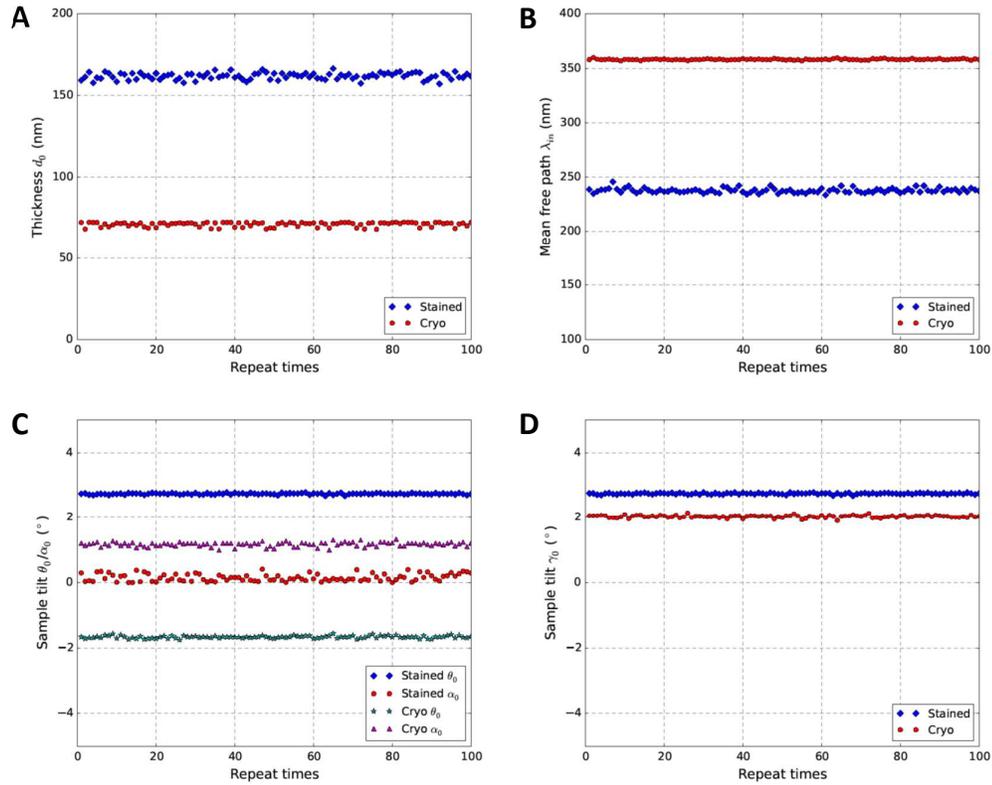


Fig. 3.S3. Robustness of basin-hopping global optimization method employed in our approach. We ran the script 100 times and plotted the fitted parameters, including $d_0$ (A), mean free path $\lambda_{in}$ (B) and sample tilt ($\theta_0$, $\alpha_0$ in C, $\gamma_0$ in D). The tests were performed for both a stained Sindbis virus infected BHK cell section dataset and a cryo dataset of purified Sindbis virus.

### 3.3.5 Tests with varying number of regions

We next attempted to examine the robustness of our method with different number sample regions used for fitting (Fig.3.S4). Different numbers of regions were chosen in the stained tilt series (blue symbols in Fig.3.S4) and the cryo tilt series (red symbols in Fig.3.S4) to determine the sample parameters. The fitted parameters, including thickness $d_0$ (Fig.3.S4A), mean free path $\lambda_{in}$ (Fig.3.S4B) and sample tilt ($\theta_0$, $\alpha_0$ in Fig.3.S4C, $\gamma_0$ in Fig.3.S4D), were plotted in Fig.3.S4 as a function of the number of regions, respectively. It is evident that our approach is stable with varying number of selected regions for both stained and cryo datasets. It is satisfying to observe that the method could correctly determine the parameters with just one region despite the low SNR of cryo tilt series.

To better understand the overall parameters determined from multiple regions, we investigated the results from individual regions and overall result from multiple regions. As shown in Fig.3.S5, we chose six regions with different amount of contents (i.e. darkness) from the stained tilt series and plotted the individual results (blue diamonds in Fig.3.6) determined independently for each region. The mean (red circles in Fig.3.6) of individual fitting results is consistent with the value (green triangle in Fig.3.6) determined using all regions simultaneously. The thickness and mean free path values independently determined from different parts of the sample are very consistent with only small scatter (Fig.3.6), which demonstrated the robustness of both the optimization algorithm and the non-linear least square formulation of our method. In contrast, the individually fitted sample tilts have larger scatter (Fig.3.6C), which might arise from cutting artifact wrinkles in section samples or the nonuniform density distribution in the regions leading to different amounts of apparent tilt added to the actual sample tilt. These results suggest that fitting with multiple regions will help average out local variations and obtain a set of parameters representing the overall sample properties.
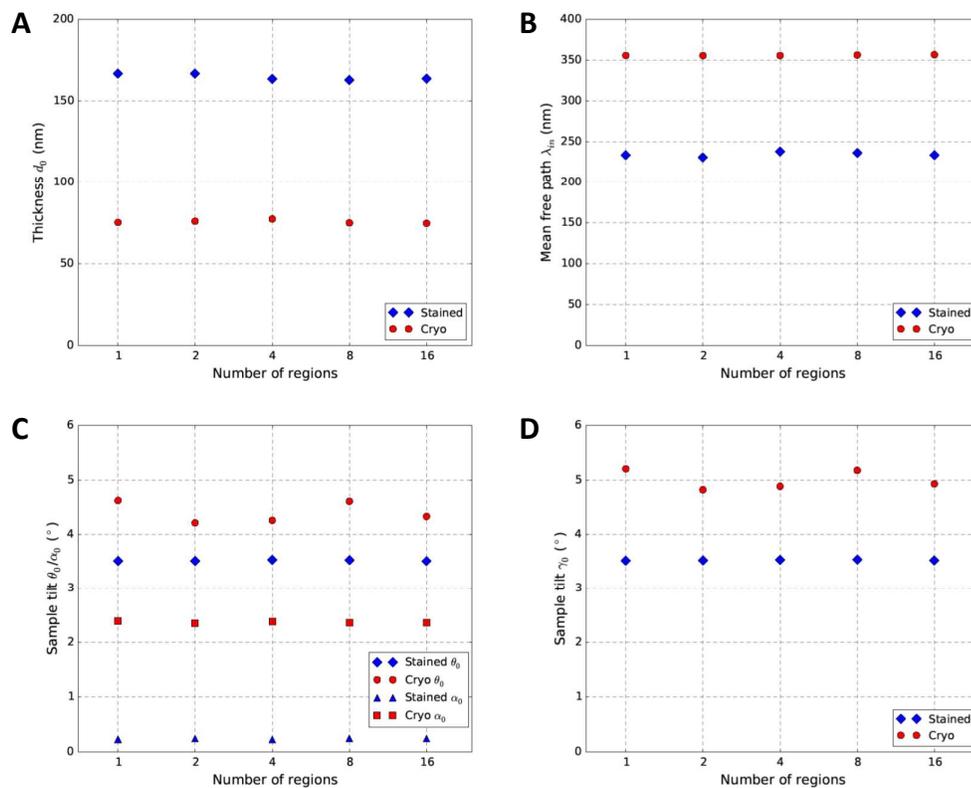
Fig. 3.S4. Performance with different numbers of regions for both stained (blue symbols) and cryo (red symbols) tilt series. The plots demonstrate the stability of thickness $d_0$ (A), mean free path $\lambda_{in}$ (B) and sample tilt ($\theta_0$, $\alpha_0$ in C, $\gamma_0$ in D) determined by our method with increasing number of regions used for fitting.

In order to further study the robustness of our method in the measurement of sample thickness, we tested one cryo dataset with ice thickness gradient (Fig.3.S6). In this dataset, the gradient of ice thickness could be observed in the tilt image (Fig.3.S6A) and in the side view of the tomographic reconstruction (Fig.3.S6B). We selected 6 ice regions (red boxes in Fig.3.S6A) along the direction of ice gradient and used our method to estimate the ice thickness of these regions. It is clear that the estimated thicknesses increase smoothly (Fig.3.S6C), which is consistent with the gradient visible in the 2-D image (Fig.3.S6A) and the side view of the tomogram (Fig.3.S6B).

### 3.3.6 Cross validation with more experimental tilt series

As can be seen from the above results, our approach could reliably determine the sample parameters of both stained and cryo tilt series. We further tested this approach with additional stained and cryo datasets. In Fig.3.7A-D, we could see that fitted parameters vs parameters measured from tomograms all tightly clustered on the diagonal lines, suggesting excellent agreement between the two sets of independently derived parameters. Since the mean free path cannot be measured from tomograms and there is no experimental measurement for comparable imaging/sample conditions in literature, we use three indirect evidences to support our fitted mean free paths (Fig.3.7E). First, the independently fitted values were self-consistent for different datasets within the same sample groups, stained (blue diamonds), cryo datasets collected at 300 kV (red circles) and 200 kV (red thin diamonds), respectively. Second, the relative ranking of the values was consistent with the expected larger mean free paths for cryo samples at 300 kV (red circles) than those of cryo samples at 200 kV (red thin diamonds) and stained samples (blue diamonds). Third, the fitted free path for cryo samples acquired at 300 kV (red circles) and 200 kV (red thin diamonds) was also consistent with the value 350 nm (green point-up triangle) and
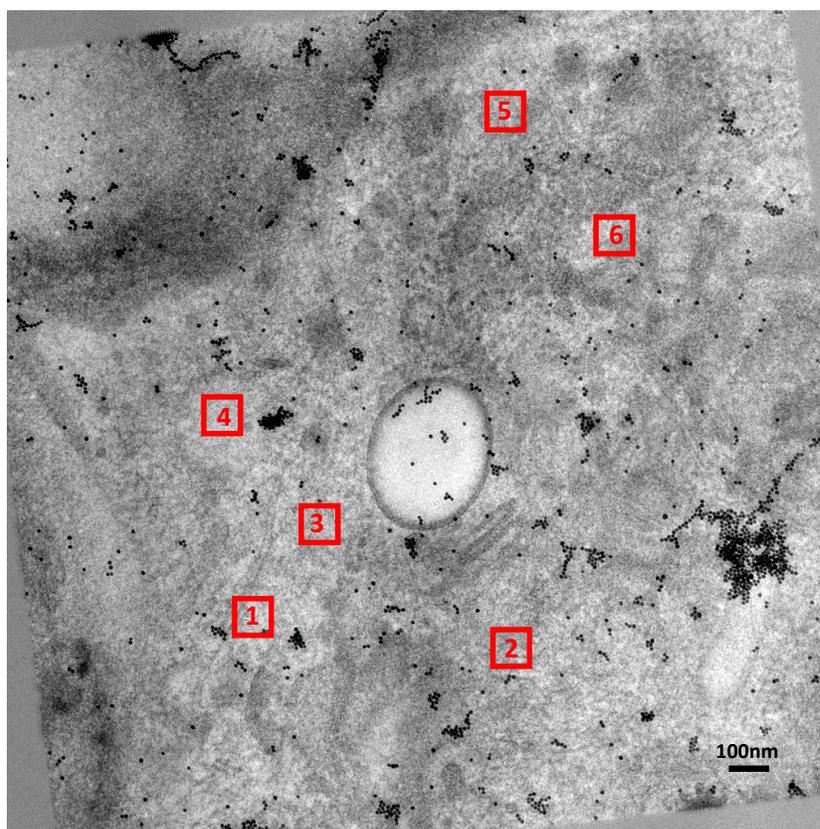
Fig. 3.S5. Multiple regions with different contents selected for fitting. The dataset is same as the one used in Fig.3.1A. The fitting results were shown in Fig.3.6.
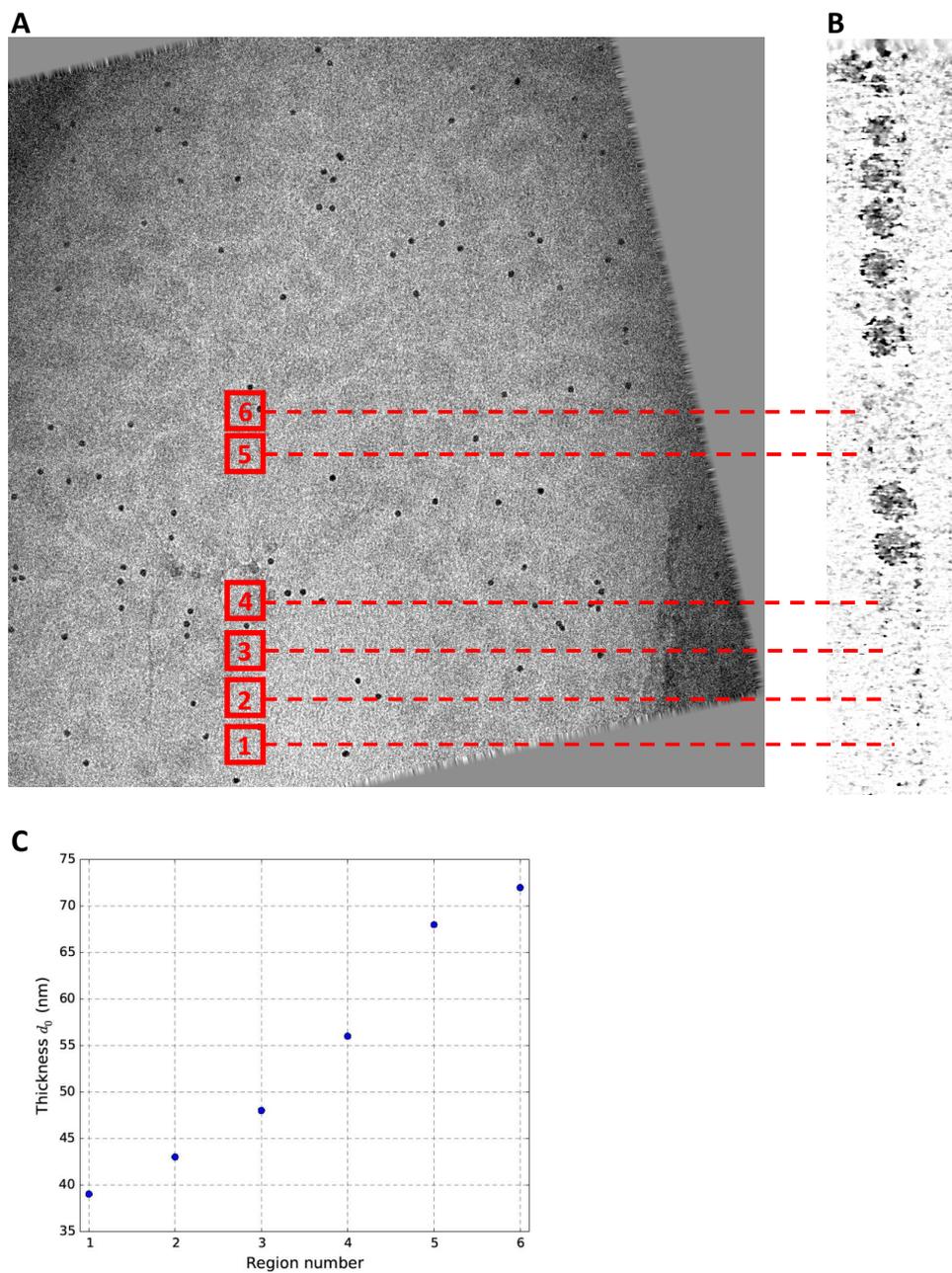
Fig. 3.S6. Performance with vitreous ice thickness gradient. (A) A -60°
tilt image of cryo dataset of purified Sindbis virus. The 6 marked regions
were used for the test. (B) Side view of the 3D tomogram with clearly
visible thickness gradient. (C) Plot of the estimated sample thickness of
the 6 regions shown in (A).
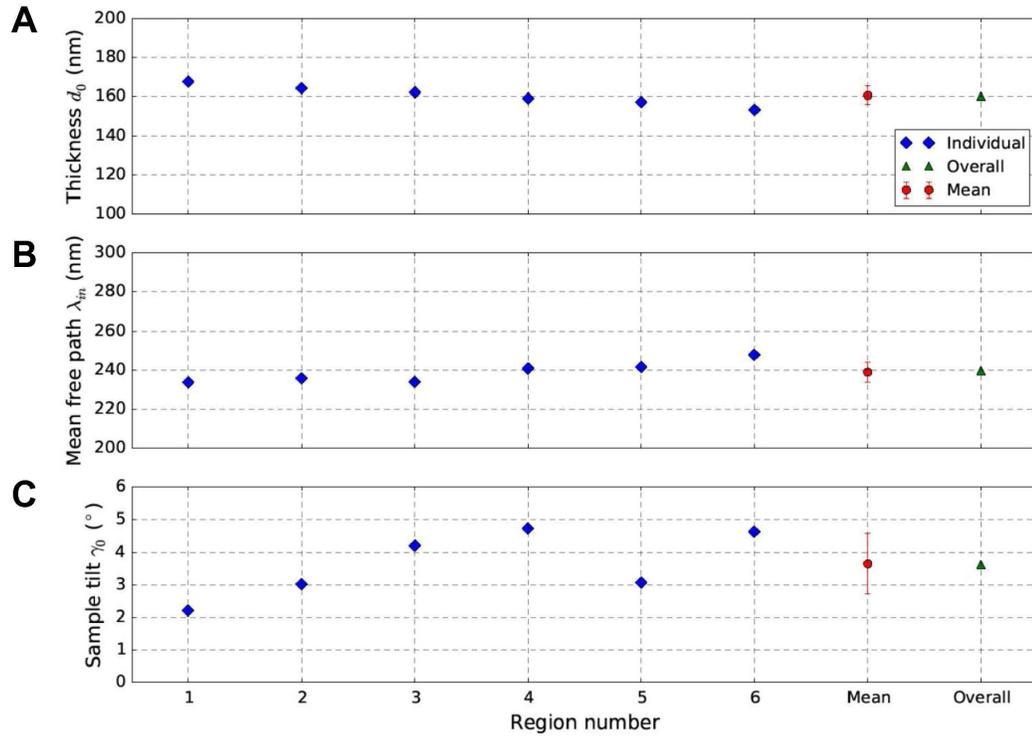
Fig. 3.6. Comparison of the parameters fitted from different regions shown in Fig.3.S5. The fitted parameters include thickness $d_0$ (A), mean free path $\lambda_{in}$ (B) and sample tilt $\gamma_0$ (C). The blue diamonds represent the parameters determined from individual regions. The red circle represents the mean of all blue diamonds. The green triangle represents the overall parameters determined by simultaneously fitting all selected regions. Note that the overall parameters (green triangles) are in agreement with the mean value (red circles), although the individually determined parameters vary slightly from region to region.

300 nm (green point-down triangle) extrapolated from experimental measurements of different imaging conditions [36].

## 3.4 Discussion

In this study, we developed a novel computational approach to determine sample thickness, tilt and mean free path for electron tomography by solving a constrained nonlinear least square problem of an overdetermined system. Due to the benefit of an overdetermined system, the resulting parameters from this approach are reliable. The systematic tests with good performance have shown that this approach can provide accurate and robust estimations of these parameters for both stained and cryo tilt series.

It is well-established that appropriate thickness of vitreous ice in cryo-sample is crucial for obtaining high quality images [45]. Thick ice will result in poor image contrast, while the embedded sample structure may be disturbed due to the contact with the air-water interface if the ice is too thin. Thus, reliable and convenient determination of ice thickness can help evaluation of image quality, selection of good images and improvement of reconstruction resolution. It will also provide feedback to help further improve sample preparation conditions. In addition, an optimal Z-dimension of cryo-tomogram can be predefined based on reliable estimation of the thickness of specimen, allowing a more efficient reconstruction process and obviate the need of visual examination of the thickness in trial reconstruction with poor contrast. Therefore, our method, which is able to reliably and quickly determine the thickness from 2D tilt series before 3D reconstruction, will be a useful tool for cryo-ET studies. By integrating this tool with existing alignment and reconstruction tools, it will help assemble a more automated and efficient high-throughput data processing pipeline for cryo-ET.

Based on our extensive tests with large numbers of stained and cryo samples (Fig.3.7B-D), it can be concluded that residual sample tilt occurs unintentionally but
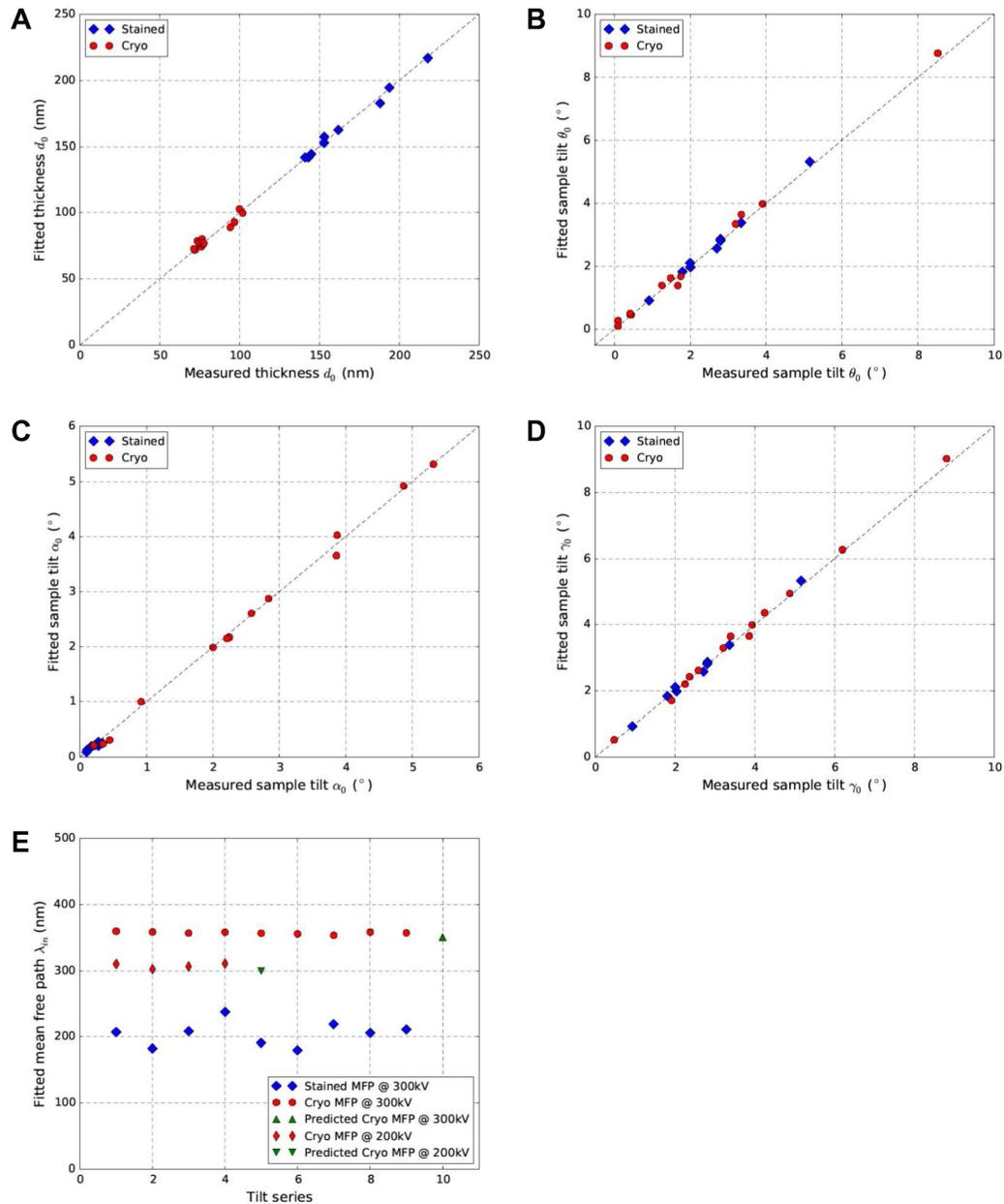
Fig. 3.7. Cross validation of the parameters fitted by our method (Y-axis) and measured from the 3D tomogram (X-axis) for both stained (blue diamonds) and cryo (red circles) tilt series. The cross-validated parameters include thickness $d_0$ (A), sample tilt $\theta_0$ (B), $\alpha_0$ (C) and $\gamma_0$ (D). (E) Plot of fitted mean free path $\lambda_{in}$. The green triangles in (E) represents the predicted mean free paths for 300 kV and 200 kV electrons through vitreous ice from literature [36].

commonly during data acquisition in ET. Similar observations were also reported based on iterative alignment of tilt images with projections of tomograms [6]. These residual sample tilts, if not correctly determined and corrected, will limit tomogram reconstruction in both computational performance and quality. Combined with the reliable thickness, correction of residual sample tilt can generate a 3D tomogram in a horizontal plane with optimal Z-dimension size, leading to faster speed and smaller 3D volume of reconstruction. For example, it took 3 min to reconstruct a $2048 \times 2048 \times 100$ tomogram from 137 tilt images using IMOD on a 3.2 GHz CPU computer. If the user has no clue about the correct sample thickness and might just blindly specify a very large number, 200 pixels, for the reconstruction. In this case, the reconstruction of tomogram would take 5 min. Since defocus gradient is one of the critical resolution limiting factors in ET [20], accurate tilt angle of each tilt image is essential for accurate determination and correction of defocus gradient, and eventually improvement of the resolution of the tomogram. Our approach is designed to determine residual sample tilt based on the Beer-Lambert law and it can work accurately and robustly without the limitation of current method based on gold beads distribution [41]. Furthermore, it also performs well on low dose cryo tilt series whose residual sample tilt is hard to determine either by fiducial marker model or by visual examination of 3D tomogram with faint top/bottom boundaries due to low contrast and smearing by missing wedge.

Experimental methods are too time-consuming to determine the electron mean free path at a wide range of incident energies of electron and sample conditions. In fact, most reported values from experiments were obtained at low acceleration voltage ($\leq$120 kV) and vary among different measurements [36]. It is thus highly desirable to have some means of predicting and verifying the mean free path. The mean free path is a critical parameter in accurate simulation of image formation process in electron microscopy. A model based iterative reconstruction (MBIR) was reported to significantly improve the reconstruction quality according to its applications on electron tomography of inorganic materials [11]. Since the physical foundation of image formation model in MBIR is the Beer-Lambert law, accurate mean free path

is essential for it to obtain an optimal reconstruction. Our method can thus help improve the quality of 3D reconstruction methods like MBIR by providing optimal mean free path value for every sample/imaging condition, instead of reusing a fixed value measured at different conditions.

# 4. MBIR: A CRYO-ELECTRON TOMOGRAPHY 3D RECONSTRUCTION METHOD THAT EFFECTIVELY MINIMIZES MISSING WEDGE ARTIFACTS AND RESTORES MISSING INFORMATION

## 4.1 Introduction

Cryo-electron tomography (cryo-ET) has emerged as a promising technique that allows us to comprehensively explore macromolecular complexes and cellular architecture in near-native states [46]. Using cryo-ET, the 3D tomogram of the biological sample can be reconstructed from a 2D tilt series collected by sequentially tilting the sample at different projection angles around a tilt axis [47]. In practice, the quality of reconstruction with cryo-ET remains limited by several challenges in the data acquisition and reconstruction process.

The extremely poor signal-to-noise ratio (SNR) of cryo-ET is the first major challenge in improving cryo-ET resolution [48]. To prevent significant radiation damage to biological samples by the electron beam, the total dose used for a cryo-ET tilt series is typically less than 100 $e/\mathring{A}^2$. This low-dose imaging strategy in combination with the increment of sample thickness during tilting results in very noisy, low contrast 2D projections, which poses a challenge in subsequent 2D tilt series alignments and deteriorates the resolution of cryo-ET 3D reconstruction [49, 50].

The second major challenge of cryo-ET is the missing wedge artifacts caused by the limited tilt angle range during data collection [51]. Since more electrons are lost to inelastic scattering as the effective sample thickness increases when the sample is tilted [48], the maximal tilt range of cryo-ET is typically restricted within $\pm 70°$ to ensure enough electrons can traverse through the sample, generate elastic scattering, and form reliable images [52]. Consequently, the absence of the high tilt angles

$(-90° \sim -70°$ and $+70° \sim +90°)$ becomes a missing wedge of unsampled information in Fourier space, leading to severe ray artifacts, structural elongation, and distortion effects in the final reconstruction [53]. The missing wedge artifacts dramatically weaken the interpretability of the reconstructed tomogram and limit the achievable resolution of cryo-ET [46].

To address these challenges of cryo-ET, we introduce the Model Based Iterative Reconstruction (MBIR) method [54] for tomographic reconstruction and benchmark the tomogram quality with the state-of-the art algorithms, including Back Projection (BP), Simultaneous Iterative Reconstruction Technique (SIRT) and Iterative Compressed-sensing Optimized Non-uniform fast Fourier transform reconstruction (ICON) [55]. In MBIR framework, the reconstruction is formulated as the maximum a posterior (MAP) estimate of the unknowns given the measurements

$$(\hat{f}, \hat{\varnothing}) = argmax_{f,\varnothing}\{log\ p(g)\} = argmax_{f,\varnothing}\{-log\ p(f,\varnothing) - log\ p(f)\} \qquad (4.1)$$

where $g$ represents the data obtained from an imaging system (e.g. cryo-ET tilt series), $f$ represents the unknown 3D structure to be discovered, $\varnothing$ represents the unknown nuisance parameters of the system such as beam intensity fluctuations and noise characteristics. $p(f, \varnothing)$ is the likelihood function that models how the observations are related to the unknowns, $p(f)$ is the assumed prior distribution of the unknown structure. Here $p(f, \varnothing)$ and $p(f)$ indicate the forward model of image formation and prior model of the tomogram in MBIR algorithm, respectively [54]. Currently, the forward model computes the decay of electron beam intensity following Beer-Lambert Law. The prior model uses a Gaussian Markov Random Field to account for diffuse or sharp interfaces between structural features and encourage smoothness in the solution. The goal of MBIR will be to compute a final estimate $\hat{f}$ that represents a balance between fitting the system forward model $p(f, \varnothing)$ Fig.4.1 illustrates a general framework of MBIR for solving inverse problems in imaging applications.

$$\left(\hat{f},\widehat{\emptyset}\right) = \underset{f,\emptyset}{\mathrm{argmax}}\{\log p(f,\emptyset|g)\} = \underset{f,\emptyset}{\mathrm{argmin}}\{-\log p(g|f,\emptyset) - \log p(f)\}$$
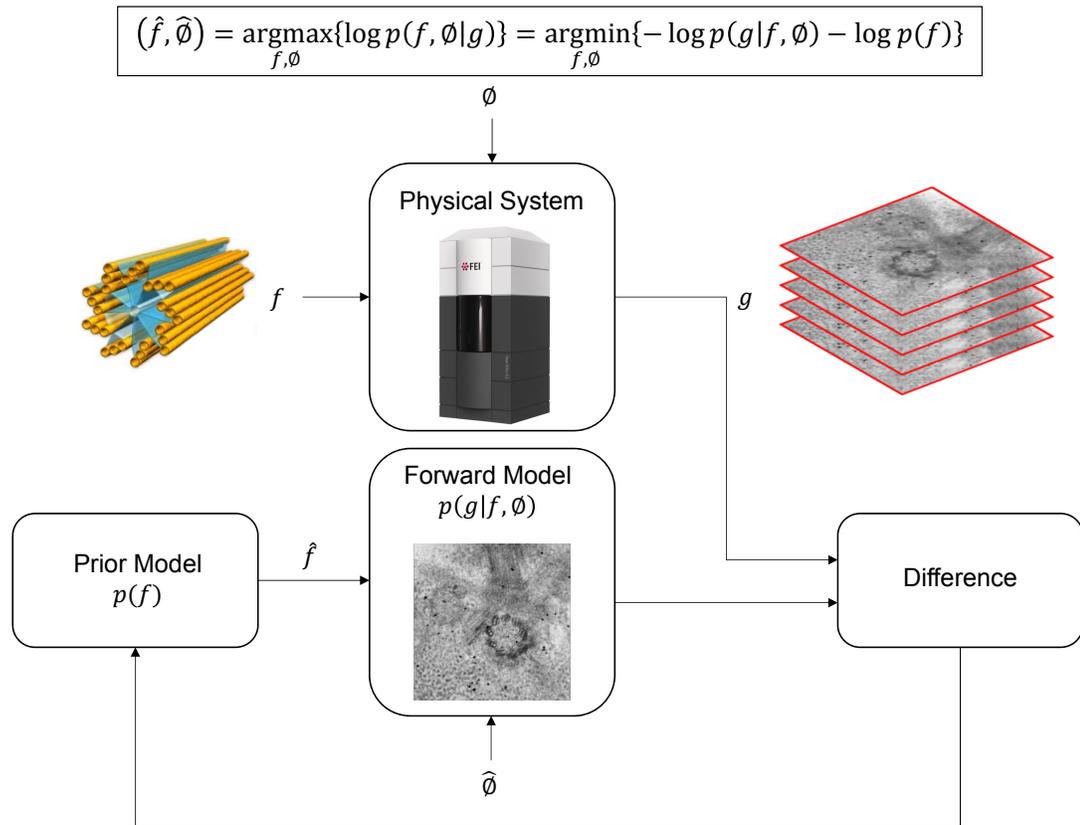
Fig. 4.1. Graphical scheme of the MBIR algorithm. $g$ denotes the tilt series from cryo-ET, $f$ denotes the unknown structure, and $\emptyset$ denotes unknown nuisance parameters of the system (e.g. noise characteristics) which needs to be determined in the inverse process. $p(.)$ denotes the probability density function and $p(f,\emptyset)$ and $p(f)$ are the forward model and prior model in the MBIR algorithm, respectively. $\hat{f}$ and $\widehat{\emptyset}$ denote the estimate of $f$ and $\emptyset$, respectively.

MBIR method has been previously shown to generate better quality tomograms when applied to tomography applications like CT scan, X-Ray tomography, positron emission tomography (PET), optical diffusion tomography (ODT), and atomic resolution high-dose electron tomography of radiation-resistant material specimens with heavy atoms [54, 56]. MBIR combines a forward model for image formation with a prior model for the unknown structure to reconstruct tomograms. In this study, tests with both plastic embedded ET dataset and ice embedded low-dose cryo-ET datasets of radiation-sensitive biological samples have shown that MBIR can significantly improve the reconstruction quality with enhanced contrast, reduced missing wedge artifacts, and partially restored information in the un-sampled angular region.

## 4.2   Methods

### 4.2.1   Implementation of MBIR

The MBIR algorithm was implemented into a standalone program with C++ language, it is cross-platform portable and works on Linux, Windows and Mac OS X operating systems. The MBIR software package used for ET is freely available in the form of binary executables and source codes from the authors' website (https://engineering.purdue.edu/ bouman/OpenMBIR/bf-em).

### 4.2.2   Test datasets

We evaluated the performance of MBIR method on both plastic embedded ET dataset and cryo-ET datasets by comparing its results with three reconstruction techniques used in the cryo-ET community, including Back Projection (BP) and Simultaneous Iterative Reconstruction Technique (SIRT) available in IMOD [17], and ICON [55]. The plastic embedded ET dataset obtained from IMOD tutorial website [17] was originally provided for dual axes reconstruction, but we only used the first tilt series (BBa.st) in our study. Four published experimental cryo-ET datasets

(EMPIAR-10037, EMPIAR-10045, EMPIAR-10064 and EMPIAR-10110) were downloaded from the public database EMPIAR [57]. EMPIAR-10064 dataset was collected with the Volta phase plate (VPP). These tilt series were aligned based on fiducial gold markers using IMOD and then reconstructed by the four reconstruction techniques, respectively. In this study, the images used for comparing the performance of different methods are contrast-normalized to avoid subjectivity of observations and to ensure the reliability of comparison.

## 4.3 Results

### 4.3.1 Missing wedge assessments using gold markers

We first evaluated MBIR using one cryo-ET dataset (EMPIAR-10045) by visually examining the missing wedge artifacts of gold markers in different slice views of the tomograms. Due to the missing wedge problem, the gold markers become elongated along the direction of the missing wedge and suffer from halos and streaking artifacts in the adjacent region. Fig.4.2 compares slice views of the reconstructions generated by the four methods using gold markers as an indicator of quality. In each block, three planes represent the XY-slice (middle plane), XZ-slice (top plane) and YZ-slice (right plane) of the tomogram, respectively, intersecting at the same gold marker. The zoomed-in view of the gold markers pointed by white arrows in the three planes are placed at the corner of the corresponding planes. From the XY-slices of tomograms, it is clear MBIR (XY-slice in Fig.4.2D) has eliminated the halos artifacts and displays more round, sharp-edged gold markers than other methods. In the XZ and YZ-slices, MBIR (Fig.4.2D) significantly reduced the elongation and ray artifacts of gold markers with improved contrast of the biological structures, compared with the tomograms reconstructed by other methods. Hence, MBIR-reconstructed tomograms show less artifacts from the missing wedge problem, better contrast in cryo specimen, and clearer background.

To further examine the performance of MBIR, we applied it to one cryo-ET dataset acquired with VPP (EMPIAR-10064 in Fig.4.3A), two cryo-ET datasets without VPP (EMPIAR-10037 and EMPIAR-10110 in Fig.4.3B and C), and one plastic embedded ET dataset (IMOD tutorial dataset in Fig.4.3D). Fig.4.3 shows the slice views of these four datasets in which each row represents the results of one dataset reconstructed by the four methods and each column represents the results of one method applied to different datasets. In Fig.4.3B and D, XY-slices are mainly used to reveal the reconstruction quality of sample areas without targeting at a gold marker because the sample and markers are not on the same XY plane. For a challenging dataset shown in Fig.4.3B, it is clear that BP reconstruction quality is too poor to make the biological sample visible. SIRT and ICON reconstructions contain phantoms of gold markers at the upper left corners in XY-slice (circled by dash lines in Fig.4.3B) which is caused by the missing wedge artifacts and should not appear here since XZ-slices in Fig.4.3B indicate this gold marker is located in different Z sections of the sample. In stark contrast, MBIR in Fig.4.3B is able to drastically reduce the missing wedge problem in XZ-slice and YZ-slice, completely suppress the gold marker phantoms in XY-slice and considerably enhance the sample contrast. In addition, MBIR provides better quality of tomogram in other datasets of Fig.4.3, which is in a good agreement with the results shown in Fig.4.2. In summary, the comparison of slice views among different methods in Fig.4.3 and Fig.4.2 gives a clear impression that MBIR has superior performance in boosting contrast of biological specimens, eliminating halos and streaking artifacts, retaining sharp features, and reducing noise. The superior performance of MBIR is evident in both cryo-ET (Fig.4.2 and Fig.4.3A-C) and plastic-embedded ET (Fig.4.3D) datasets.

### 4.3.2 Power spectra evaluation

To quantitatively evaluate MBIR's ability in restoring missing information, we calculated the log-scaled power spectrum of the central XZ-slice and used it as a
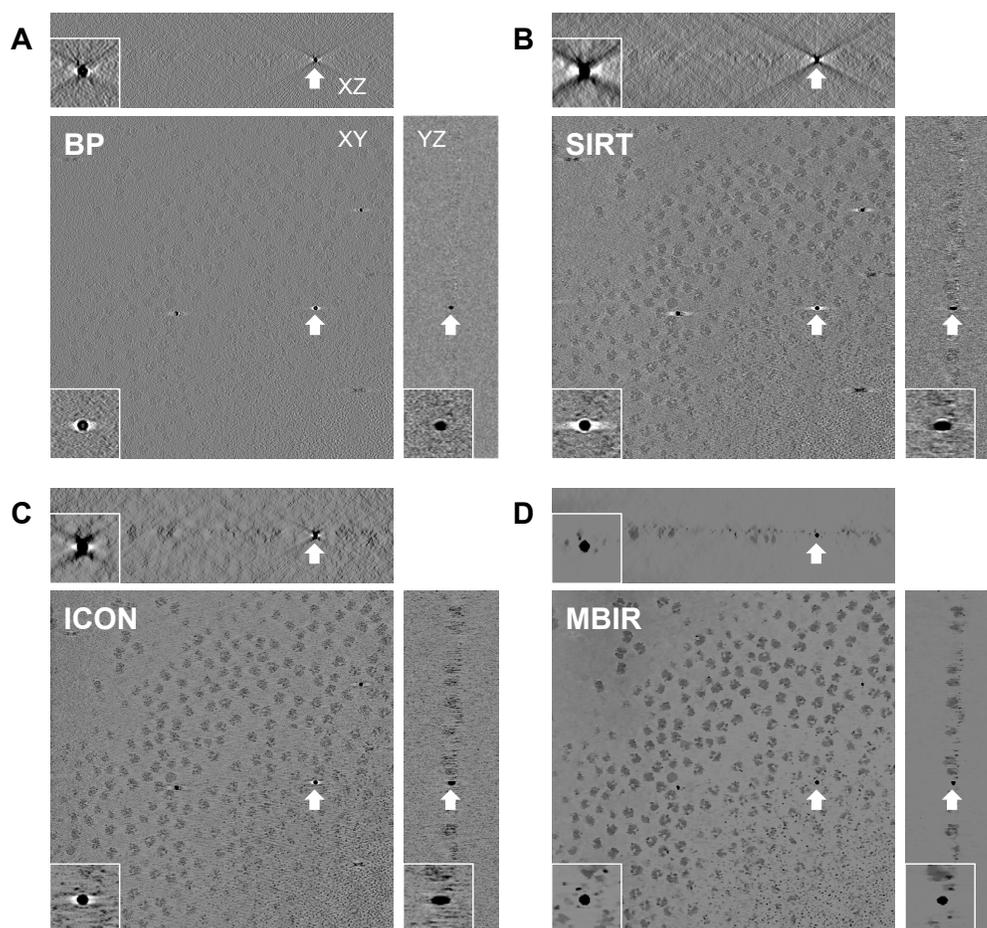
Fig. 4.2. Comparison of tomograms from an experimental cryo-ET dataset (EMPIAR-10045) reconstructed by BP(A), SIRT (B), ICON (C) and MBIR (D) methods. The three planes for each method represent the XY-slice (middle plane), XZ-slice (top plane) and YZ-slice (right plane) of the tomogram intersecting at the same gold marker. In each plane, the gold marker is indicted by a white arrow with corresponding zoomed-in view showing the missing wedge artifacts.
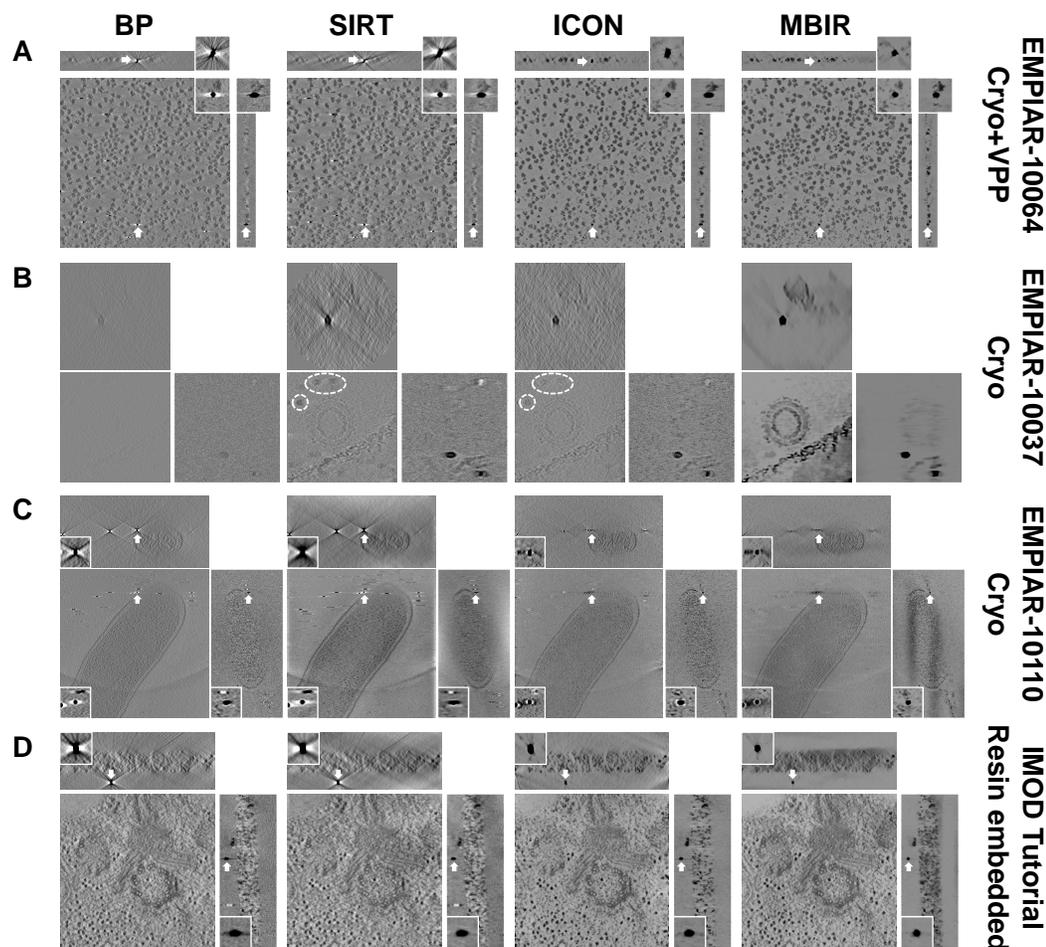
Fig. 4.3. Comparison of tomograms from multiple experimental ET datasets reconstructed by different reconstruction techniques. Each row indicates the reconstructions from the same dataset using different methods. Each column indicates the reconstructions from the same method applied to different datasets. The data type and EMPIAR ID are denoted at the right side of each row. The method of comparison in each dataset is the same as described in Fig.4.2. Note that the XY-slices of the dataset shown in (B) and (D) are used to show the biological sample area and not targeted at the gold markers since the sample and gold markers are located in different Z sections.

measurement of information restoration in 3D reconstruction. As depicted in Fig.4.4, four plots of power spectra correspond to the central XZ-slices of the tomograms reconstructed by the four methods shown in Fig.4.2. It is noted that MBIR can fill more un-sampled region in Fourier space than other methods, not only in the region of the missing wedge but also the empty space between two adjacent tilts, suggesting better performance of MBIR in restoring missing information. It is worth noting that the lines at the corners of BP (Fig.4.4A) and SIRT (Fig.4.4B) power spectra are due to the aliasing issue. To check if such aliasing issues are unique to our results, we downloaded another four tomograms from EMDB [58], calculated their central XZ-slices power spectra, and plotted them in Fig.4.S1. The results in Fig.4.S1 suggest that this aliasing issue is a general phenomenon in the cryo-ET field since it is observed in a variety of data, including data from multiple research groups, varying TEM facilities and imaging conditions, a diverse range of samples, and different reconstruction softwares.

We next examined the central XZ-slice power spectra of the datasets displayed in Fig.4.3 and compared them in Fig.4.S2. In general, MBIR and ICON yield more non-zero values in the missing wedge region than BP and SIRT, except for one challenging dataset (Fig.4.S2B). However, power spectrum may not be a reliable and complete assessment for the information restoration because it only conveys the amplitude information without considering the phase information. What's more, varying filters can be internally applied to tomograms in different methods to balance the non-uniform sampling in Fourier space [59]. As a result, further validation is still needed to confirm the advantage of MBIR in restoring not only amplitude but also phase information.

### 4.3.3 Cross validation of projections using the leave-one-out FRC method

We used the leave-one-out Fourier ring correlation (FRC) method [7] to explore the correctness of the information restored by MBIR and compare it with the performance

**BP**  **SIRT**  **ICON**  **MBIR**
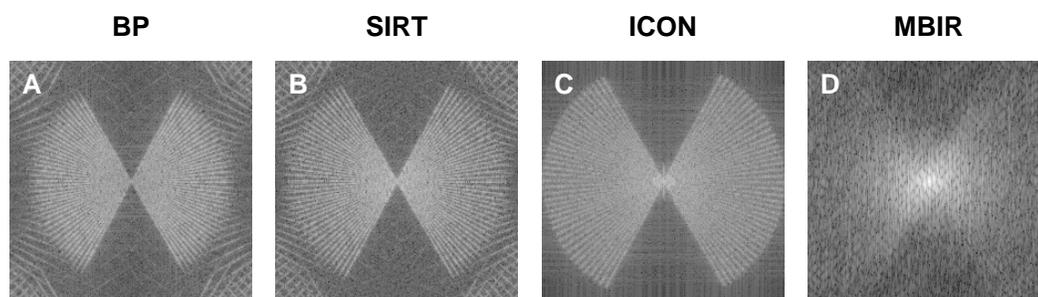


Fig. 4.4. Comparison of the central XZ-slice power spectra from the tomograms shown in Fig.4.2. The tomograms were reconstructed by BP (A), SIRT (B), ICON (C) and MBIR (D) methods, respectively.



**emd-3792**  **emd-3967**  **emd-3977**  **emd-5887**
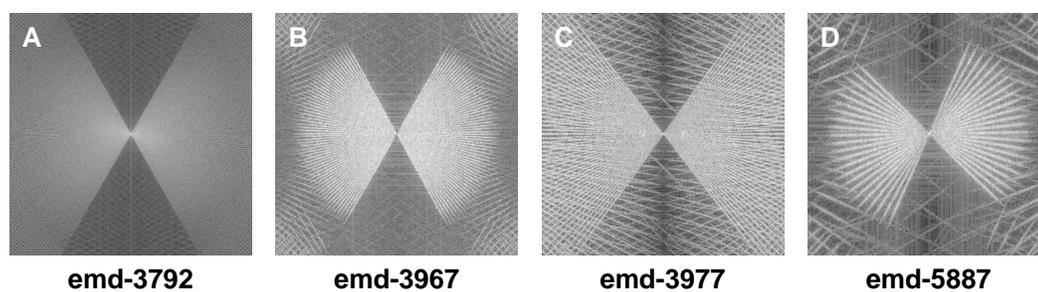
Fig. 4.S1. Observations of aliasing issue in the central XZ-slice power spectra of cryo-ET tomograms downloaded from EMDB. The EMDB ID is marked at the bottom of each images.
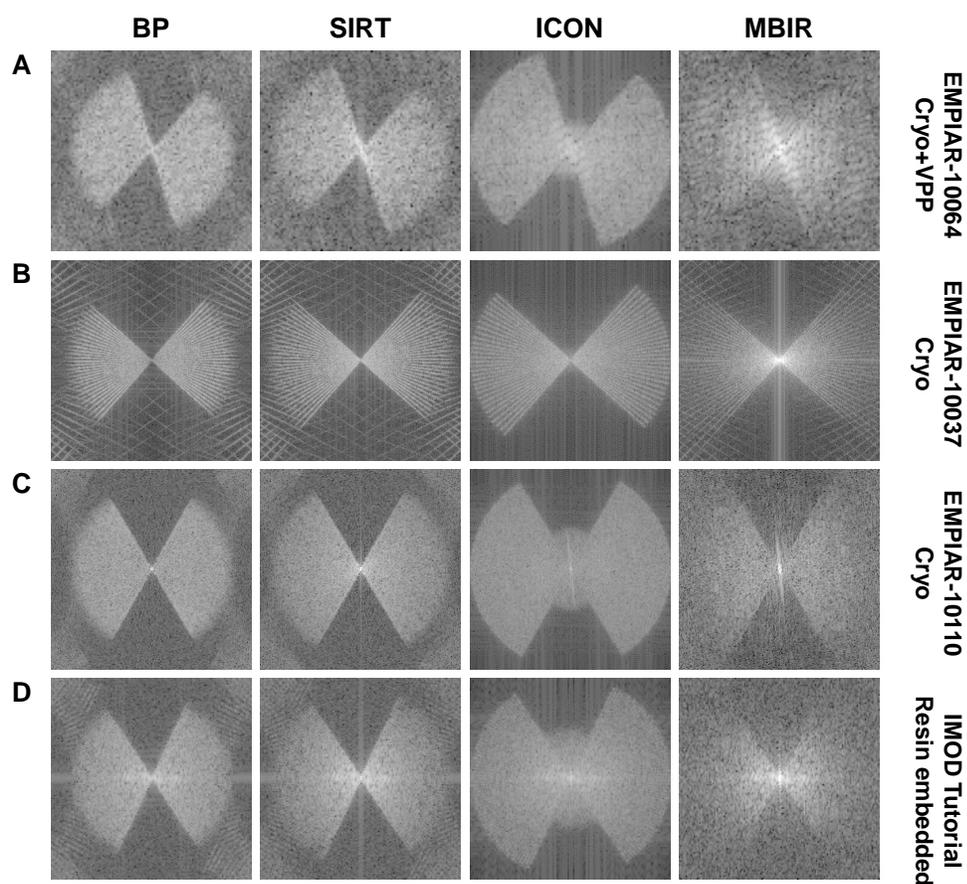
Fig. 4.S2. Comparison of the central XZ-slice power spectra from the tomograms reconstructed by different reconstruction techniques. Each row shows the power spectra of central XZ-slices from the same dataset using different methods. Each column shows the power spectra from the same method applied to different datasets. The data type and EMPIAR ID are denoted at the right side of each row.

of other reconstruction methods. In this test, the FRC is calculated for the raw tilt image $X$ and the corresponding reprojection $\tilde{X}^-$ from a tomogram computed from all other tilts without tilt X. Here the tilde sign represents the reprojection from a tomogram, and the minus sign represents the tomogram used for reprojection is calculated by omitting the tilt $X$ from the original tilt series to avoid bias. We first excluded a raw image $X$ at a certain tilt angle and utilized the rest images of the tilt series to generate a tomogram. Next, we re-projected this tomogram at the angle of tilt $X$ to obtain a reprojection $\tilde{X}^-$. Finally, we calculated the FRC curve between the excluded raw image $X$ and the reprojection $\tilde{X}^-$, and used this FRC curve as a quantitative evaluation of phase information recovery. As shown in Fig.4.5A, the first row and the second row are the raw images $X$ (the first image in each row) and the reprojections $\tilde{X}^-$ corresponding to different reconstruction methods at a smaller tilt angle 0° and a larger tilt angle 45°, respectively. The gold markers indicated by white arrows are zoomed in and placed at the lower left corners of each images. It is evident that the gold marker in MBIR reprojection is circular without discernible distortion or blurring, which is nearly identical to the original tilted image, even at a high tilt angle. In contrast, the gold markers in the reprojections of other methods clearly suffer from missing wedge artifacts including elongation, white halos, and blurring. Furthermore, such visual assessments are verified quantitatively by the FRC (Fig.4.5B and C) of the raw tilt images and reprojections shown in Fig.4.5A. As shown in Fig.4.5B and C, the quick drop of BP (blue curve), SIRT (red curve) and ICON (green curve) FRC curves implies that only low resolution information is reliably restored in the non-sampled angular regions. However, the FRC curve of MBIR exhibits a significantly higher correlation between the reprojection and the original tilt image, confirming the successful restoration of the missing information.

To further substantiate the capability of MBIR in restoring missing information, we performed the same analysis as described in Fig.4.5 on more datasets and summarized the comparisons of raw images and reprojections in Figs.4.S3 and 4.S4 and FRC comparisons in Fig.4.6. As can be seen from Figs.4.S3 and 4.S4, MBIR preserved the

round shape of gold markers in the leave-one-out reprojections at low (Fig.4.S3) and high (Fig.4.S4) tilt angles in both cryo (Fig.4.S3A-C, Fig.4.S4A-C) and plastic embedded datasets (Fig.4.S3D, Fig.4.S4D), which is consistent with the results shown in Fig.4.5A. Fig.4.6 shows the FRC comparisons of different methods when 0° (Fig.4.6A, C, E, G) and 45° (Fig.4.6B, D, F, H) tilts were excluded in the leave-one-out tests, respectively. The FRC curve of MBIR (yellow curve) in Fig.4.6 is typically higher than that of other methods, which suggests the superior quality of MBIR in recovering authentic information of biological samples in 3D tomographic reconstructions. As demonstrated in Fig.4.6A-B, VPP used in this dataset boosts the signal-to-noise ratio of cryo-ET images and improves the low frequency signal in FRC curve compared with Fig.4.5B-C, leading to a smaller difference among the results of the four reconstruction methods than the case shown in Fig.4.5. However, the local missing wedge artifacts remained in the tomograms reconstructed by the other three methods but not by MBIR for this cryo-ET dataset with VPP as shown in the corresponding slice views (Fig.4.3A) and reprojections (Fig.4.S3A and Fig.4.S4A), emphasizing the advantages of MBIR method. Therefore, all the analyses above validate MBIR's capability to partially restore the missing information in both cryo-ET and plastic embedded datasets.

## 4.4 Discussion

As a widely explored 2D/3D reconstruction method, MBIR has a growing impact on the medical, industrial, and scientific imaging fields. In the present work, we introduced the MBIR method into biological ET and corroborated the substantial advantages of MBIR over current, state-of-the-art reconstruction methods for both cryo and plastic embedded data. MBIR employs a model of the image formation process and combines it with a prior model of the 3D object to formulate a MAP estimation cost function which rejects measurements that do not fit the model. Results on experimental data have effectively demonstrated the excellent performance
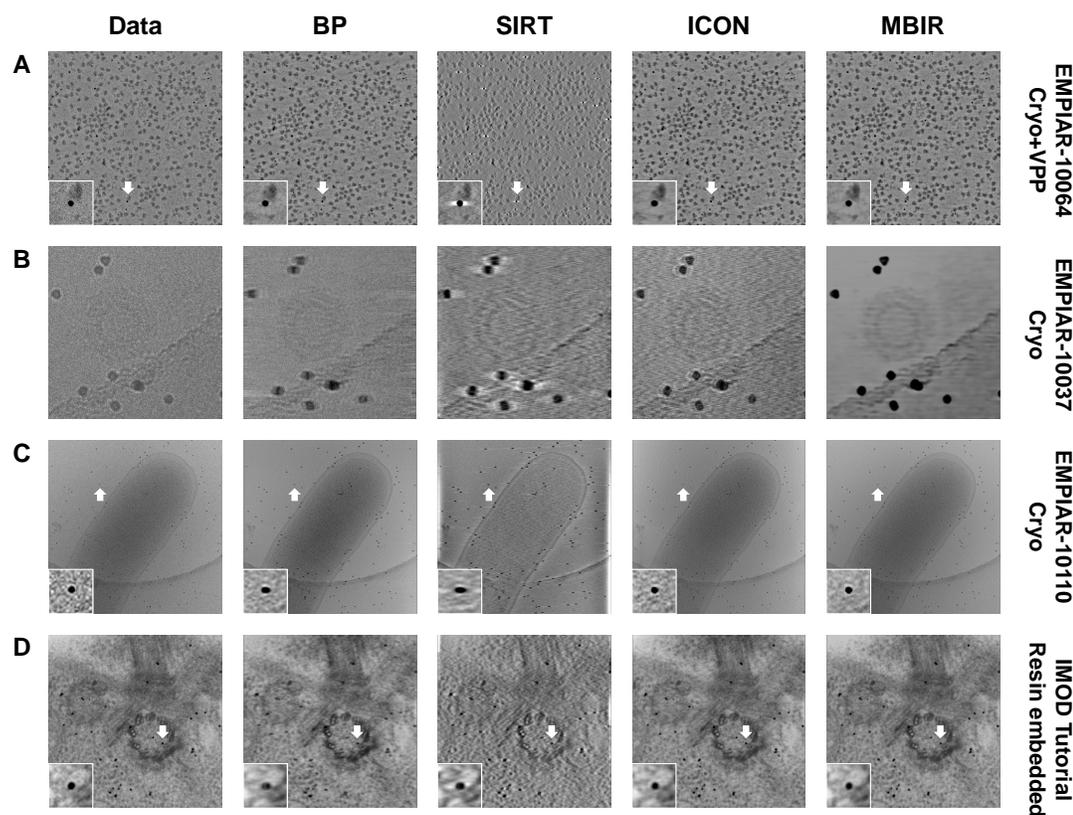
Fig. 4.S3. Comparison of reprojections at 0° when the tomograms are generated using different reconstruction techniques without the corresponding tilt. Each row shows the reprojections from the same dataset using different methods. Each column shows the reprojections from the same method applied to different datasets. The data type and EMPIAR ID are denoted at the right side of each row. In each image, the gold marker indicted by a white arrow is displayed with corresponding zoomed-in views.
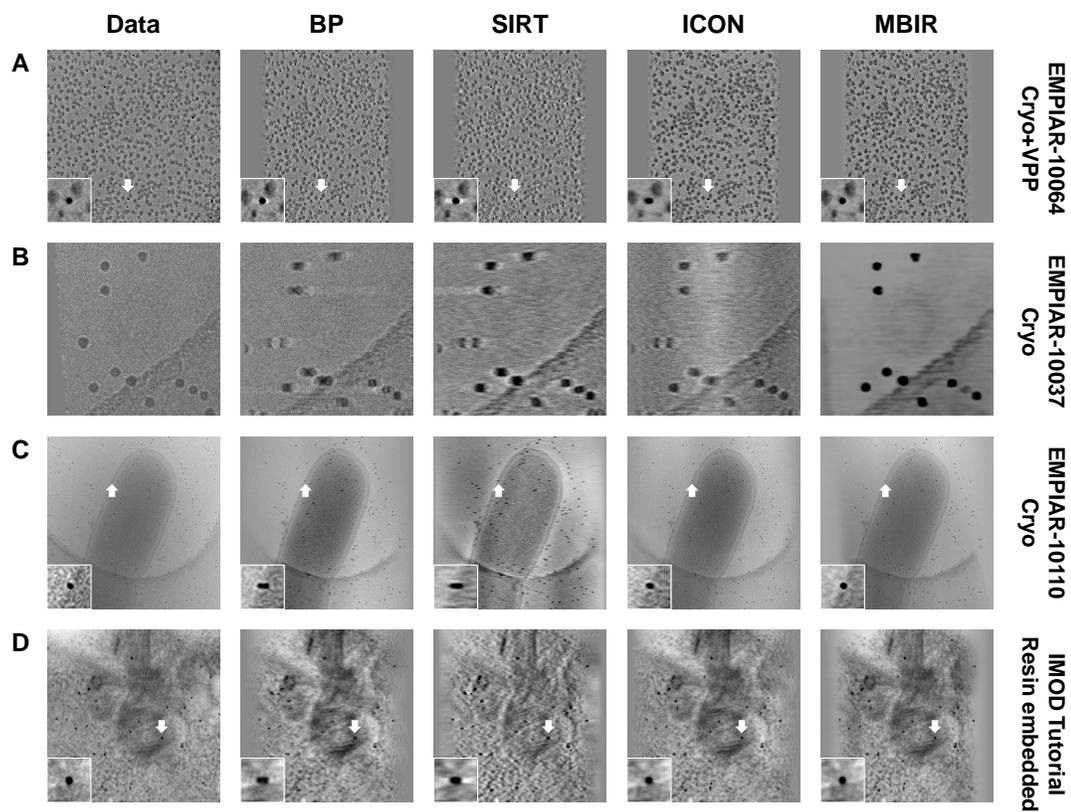
Fig. 4.S4. Comparison of reprojections at 45° when the tomograms are generated using different reconstruction techniques without the corresponding tilt. The details are the same as described in Fig.4.S3.
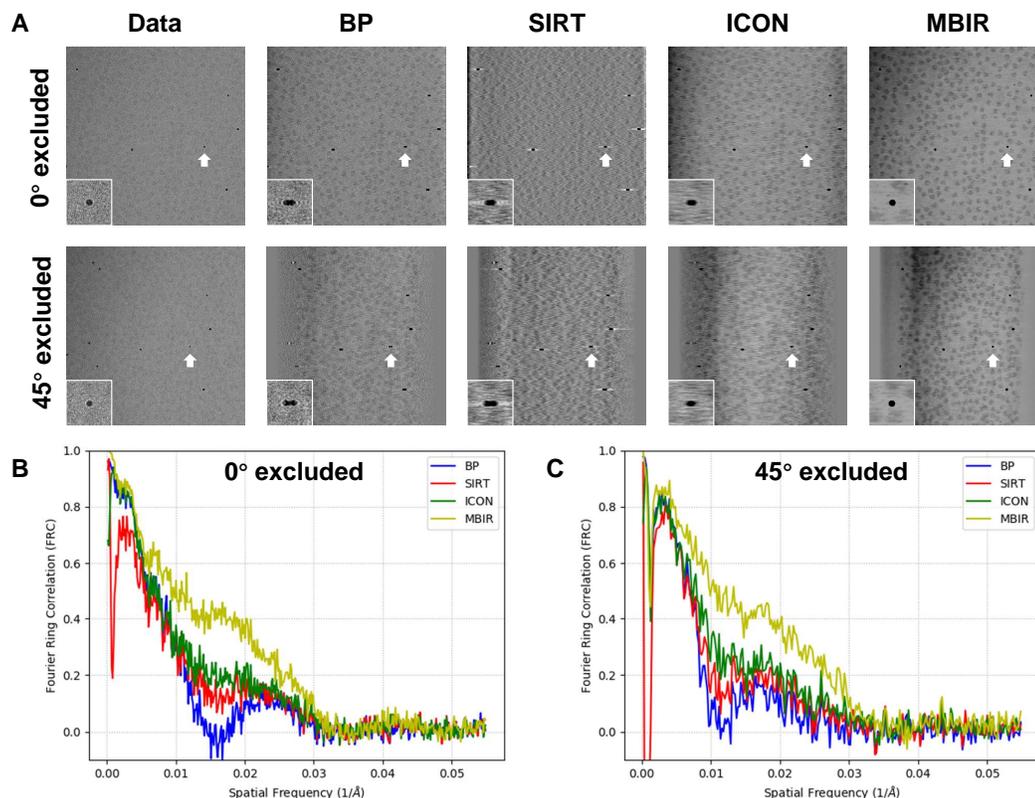
Fig. 4.5. Comparison of missing information restoration from an experimental cryo-ET data (EMPIAR-10045) reconstructed by different reconstruction techniques using the leave-one-out FRC method. (A) Comparison of reprojections at two tilt angles (0° in the first row and 45° in the second row) using the tomograms generated without the corresponding tilt. The images in the first column are extracted from the tilt series, serving as the ground truth for comparison. In each plane, the gold marker indicted by a white arrow is displayed with corresponding zoomed-in view. (B) and (C) are comparisons of the FRC curves of reprojections against the ground truth as depicted in (A) when 0° and 45° tilts are excluded in the leave-one-out test, respectively.
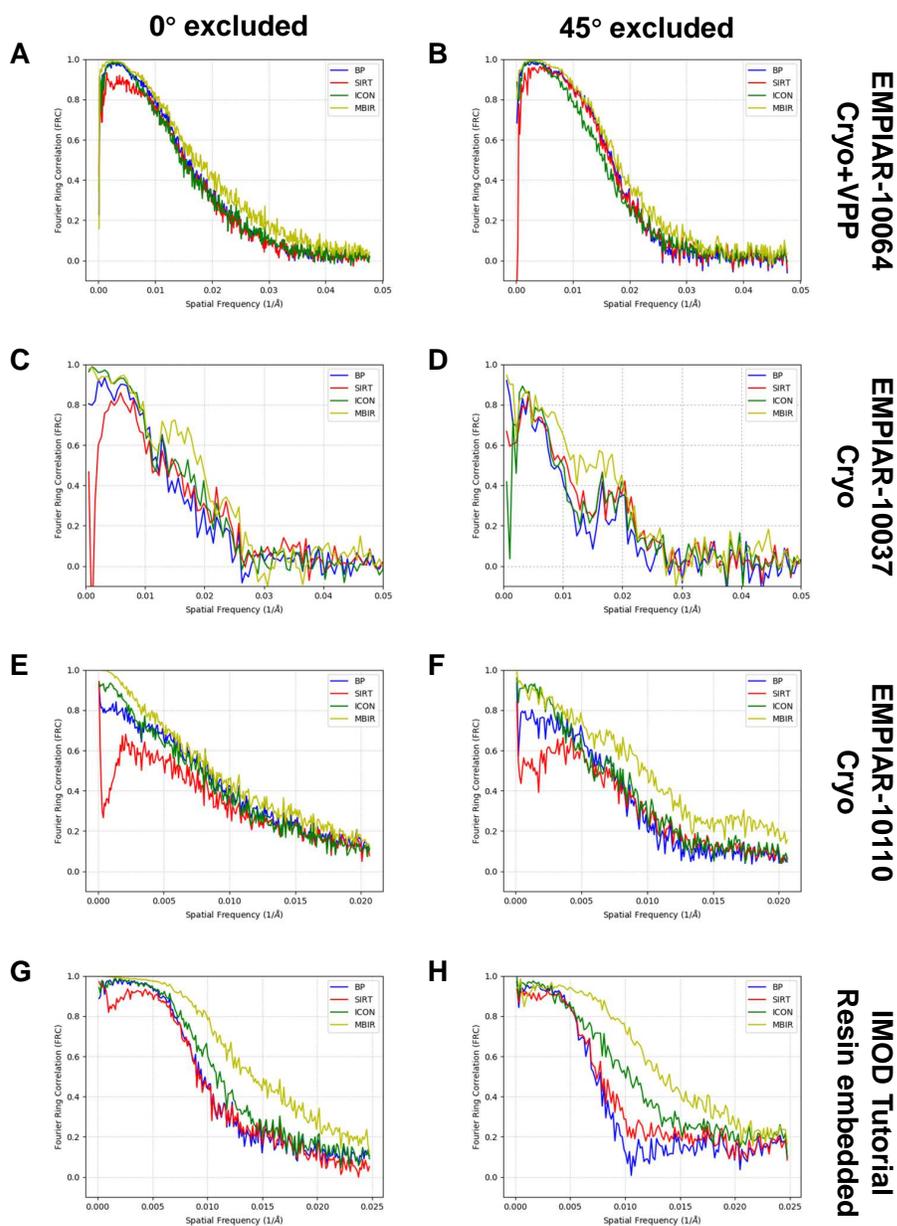
Fig. 4.6. Comparisons of FRC curves from multiple experimental ET datasets reconstructed by different reconstruction techniques using the leave-one-out FRC method. Each row represents the comparison of FRC curves from the same dataset when 0° (left plot) and 45° (right plot) tilts are excluded in the leave-one-out test, respectively. The details of the corresponding reprojections and ground truths are shown in Fig.4.S3 (0° excluded) and Fig.4.S4 (45° excluded), respectively.

of MBIR in contrast enhancement, missing wedge artifacts reduction, and missing information restoration, generating visually and quantitatively accurate tomograms.

Cryo-ET tomographic reconstruction usually suffers from problems such as high level of noise, poor contrast, artifacts caused by the missing wedge issue and unreliable restoration of missing information, which poses significant challenges to subsequent analysis of the tomograms. The clear benefits of MBIR will not only help achieve better quality reconstruction as shown in this work, but also facilitate further visualization and computational tasks, such as biological feature interpretation, structure segmentation, subtomogram averaging, and ultimately help advance cryo-ET to higher resolution.

While MBIR significantly improves tomography quality, the extensive computational load makes its speed slower compared to other approaches and restricts the application of MBIR to large datasets. Recently, a computationally optimized algorithm of Non-Uniform Parallel Super-Voxel (NU-PSV) has been developed for MBIR 2D reconstruction of CT images which enables massively parallel reconstruction and ensures fast convergence [60]. Thus, it is desirable to implement this powerful parallel algorithm into cryo-ET MBIR reconstruction in the future, using either GPU or multicore CPUs on multiple computer nodes. On the other hand, MBIR should be generalized to support tomographic reconstruction using double-tilt geometry and incorporate the objective lens contrast transfer function (e.g. defocus, astigmatism, Volta phase shift) into its forward image formation model during its iterative reconstruction process.

# 5. REAL-TIME DETECTION AND SINGLE-PASS MINIMIZATION OF TEM OBJECTIVE LENS ASTIGMATISM

## 5.1 Introduction

Transmission electron microscopy (TEM) has become a powerful technique for structural characterization of a wide range of materials including macromolecular complexes at near-atomic resolutions. In order to obtain high-quality images, it is crucial to carefully align the microscope to reach optimal illumination lens and imaging lens conditions. Minimization of the astigmatism of the objective lens is a critical task of the daily microscope alignment.

The astigmatism of TEM objective lens is defined as the angular dependency of defocus, which is primarily represented by the twofold astigmatism with elliptical Thon rings [61] in the power spectra of TEM images. The astigmatism is an essential component of the contrast transfer function (CTF) of TEM images. Many approaches have been proposed for the estimation of CTF parameters based on the power spectra of images [39, 62–69]. The CTF parameters are typically determined by iterative fitting simulated power spectra to the experimental power spectra by varying the parameters in the CTF model. However, due to the iterative nature of current methods and their availability limited to offline batch processing on Linux computers, current CTF fitting methods are not used for providing real-time feedback to guide instrument alignment. As a result, current method widely used by many microscopists is to visually examine the roundness of Thon rings from live images and simultaneously adjust objective lens stigmators to make the Thon rings as circular as possible. However, the drawbacks of this subjective method have been well recognized [70], such as its limited sensitivity for small astigmatism and potential bias caused by the astig-

matism of human eyes. It is not uncommon that a user spends significant amount of time (e.g. 30 min) to iteratively adjust the two stigmator knobs just to convince himself/herself that the objective lens astigmatism is indeed minimized. Therefore, a quantitative method capable of fast and sensitive estimation of astigmatism is desirable for improved TEM instrument alignment by providing real-time feedbacks for the adjustment of objective lens stigmators to minimize the objective lens astigmatism.

In this work, we have developed such a method, $s^2 stigmator$, that uses a direct, closed-form solution and comprehensively takes advantage of Fourier transform theory to fast calculate the mean defocus, astigmatism amplitude and astigmatism angle. The method was implemented as a DigitalMicrograph script to allow fast and sensitive detection of the astigmatism of live images. It thus can provide real-time feedback and user-friendly "radar"-style display to help guide the adjustment of objective lens stigmators and efficiently correct the astigmatism of the objective lens.

## 5.2 Methods

### 5.2.1 The $s^2$ power spectra based algorithm for astigmatism estimation

Fig.5.1 shows an overview of the $s^2 stigmator$ algorithm. Fig.5.1A is a synthetic image in which each particle is generated by projecting a 3D density map of herpesvirus (EMD-3358) into an arbitrary orientation. A CTF is applied to this image with defocus 2000 nm, astigmatism amplitude 100 nm and astigmatism angle 30°. The astigmatism angle is defined as the angle of the maximum defocus direction according to the conventions defined by the Electron Microscopy eXchange (EMX) initiative [71]. This image is used here to represent a "live" image to help illustrate the algorithm as detailed in the following paragraphs.

**Step 1:** Computation of the average 2D power spectra

In order to enhance the signal-to-noise ratio (SNR) and shorten the processing time in the subsequent steps, the 2D power spectra (Fig.5.1B) is calculated using periodogram averaging method [72,73]. We divided the image (Fig.5.1A) into multiple
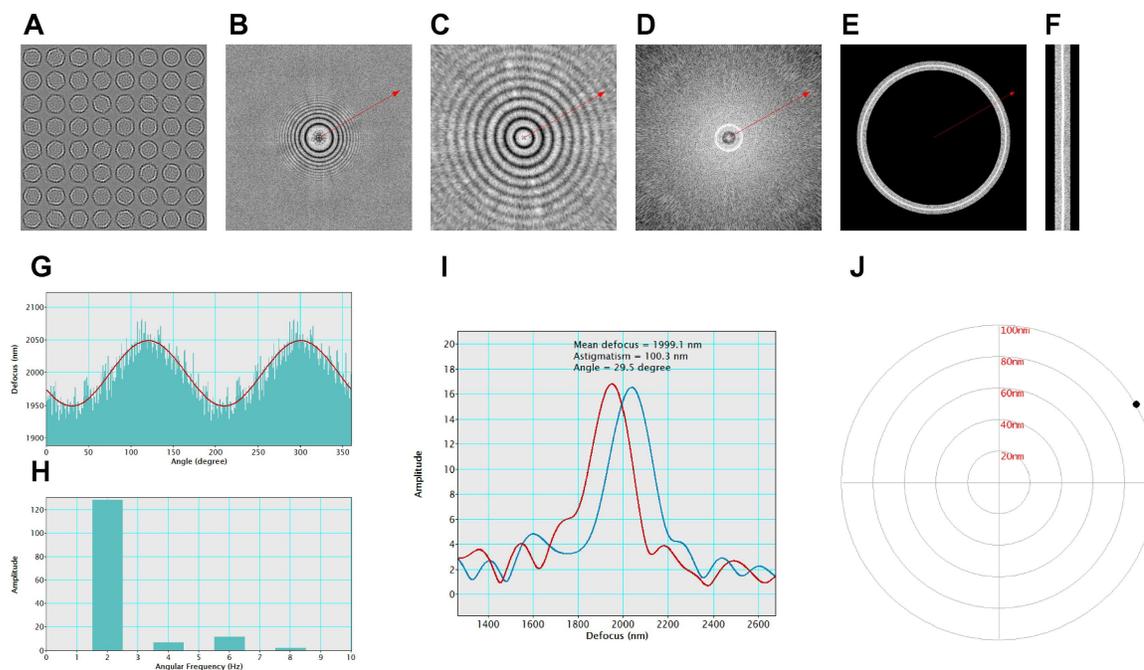
Fig. 5.1. An overview of the $s^2 stigmator$ algorithm for astigmatism determination and correction. (A) A synthetic image used for algorithm demonstration, defocus = 2000 nm, astigmatism amplitude = 100 nm, astigmatism angle = 30°, pixel size = 1 Å, B-factor = 150 Å$^2$, amplitude contrast = 0.1 and image size 2048 × 2048 pixels. (B) Regular 2D power spectra. (C) 2D $s^2$ power spectra truncated to 5.3 Å resolution at the edge. (D) A single ring generated by applying FFT to (C). (E) A shifted/enlarged single ring after performing the 2nd type Fourier shifting of (D). (F) Unwrapped ring in (E). (G) Data curve (greenish region) of the bright wave in (F) and its 2 Hz Fourier component curve (red curve). (H) FFT of data curve (greenish region). (I) Peak plot in final output. The two curves display the pixel values of rows in (F) with largest (blue) and smallest (red) defocus. (J) Radar plot in final output.

small square patches, then applied the Fast Fourier Transformation (FFT) on each patch to obtain the individual 2D power spectra, incoherently summed all of them to obtain the average 2D power spectra, and finally performed log-transform on the power spectra to minimize the magnitude difference from low to high resolution (i.e. center to edge). This method supports both square images (by most CCDs and Falcon direct electron detector) and rectangular images (by K2, DDD direct electron detectors).

**Step 2:** Conversion to 2D $s^2$ power spectra

After removing the intensity gradient by high pass filter, denoising by low pass filter, and truncating the 2D power spectra at a specific resolution (i.e. 5.3 Å in this example), 2D $s^2$ power spectra (Fig.5.1C) is generated by resampling using bilinear interpolation to make the radii represent the square of spatial frequency $s^2$ instead of spatial frequency $s$ [64]. Compared with the progressively increasing frequency of oscillations in regular 2D power spectra (Fig.5.1B), 2D $s^2$ power spectra (Fig.5.1C) consist of uniform oscillations from center to edge with equal spacing between adjacent Thon rings [64]. Furthermore, the Thon rings in $s^2$ power spectra appear more elliptic than those in regular power spectra when astigmatism exists in the image [64].

**Step 3:** Generation of defocus-dependent ring from 2D $s^2$ power spectra

Due to the uniform oscillations in the 2D $s^2$ power spectra, a single ring (Fig.5.1D) can be obtained after applying FFT on the 2D $s^2$ power spectra. The radius of the single ring along a specific direction is proportional to the CTF oscillation frequency or defocus in this direction since there is a positive linear relationship between the defocus and the frequency of oscillations in $s^2$ power spectra [64]. Consequently, the ellipticity of the single ring can be considered as a sensitive indicator to estimate the astigmatism in the image. The single ring is perfectly round for images without astigmatism. In contrast, it is elliptic for astigmatic images, suggesting angular variation of the defocus.

**Step 4:** Enlarging the ring with the 2nd type Fourier shifting Note that smaller defocus corresponds to fewer Thon rings in the 2D $s^2$ power spectra, resulting in the

single ring very close to the Fourier origin (i.e. center in Fig.5.1D) with very coarse sampling by a small number of pixels. The coarse sampling inevitably causes difficulties in ring-shape determination and astigmatism estimation in subsequent steps. To overcome this drawback, we took advantage of the 2nd Fourier shift property (Eq.5.1) to move the single ring outward by the same amount in all directions (Fig.5.1E) closer to the edge. The shifted/enlarged ring contains more pixels with finer sampling that will allow more accurate detection of the ellipticity of the ring.

The 2nd Fourier shift theorem is described in Eq.5.1:

$$e^{j2\pi s_0 t} f(t) \leftrightarrow F(s - s_0) \tag{5.1}$$

where $f$ represents the function in real space (e.g. 2D $s^2$ power spectra in Fig.5.1C), $F$ represents the Fourier transformed function (e.g. the shifted single ring in Fig.5.1E), $\leftrightarrow$ represents Fourier transform, $j$ is the imaginary unit, $t$ and $s$ represent the independent variables in real and Fourier space, respectively. The parameter $s_0$ represents the amount of shift in Fourier space. The amount of shift can be tailored to different images by first detecting the mean radius of the ring in the unshifted $s^2$ power spectra and then calculating its distance to the target radius (e.g. 4/5 of the patch size). This 2nd Fourier shift theorem is closely related to the more commonly used Fourier shift theorem that relates shift in real space with phase changes in Fourier space. However, the shift and phase changes occurred in opposite spaces for these two types of Fourier shift theorems.

In our algorithm, we multiplied the complex term $e^{j2\pi s_0 t} f(t)$ to the 2D $s^2$ power spectra (Fig.5.1C), then performed FFT and masked out the inner and outer areas, so that a shifted/enlarged single ring (Fig.5.1E) sampled by more pixels can be obtained.

**Step 5:** Representation of the shifted single ring in polar coordinates

The shifted single ring (Fig.5.1E) is transformed into polar coordinates (Fig.5.1F) using bilinear interpolation. In Fig.5.1F, the horizontal dimension represents the radius of the shifted single ring and the vertical dimension represents angles from 0 to $2\pi$.

**Step 6:** Direct, closed-form solution of the mean defocus, astigmatism amplitude and angle

We detected the peak position (radius of the shifted single ring) in each row (i.e. each direction) of the polar representation (Fig.5.1F) and displayed them in Fig.5.1G. Since the peak position corresponds to the defocus in a specific direction, we can convert the peak position to defocus according to Eq.5.2:

$$ds = 1.0/(apix \times b)$$

$$r_{max} = int(1.0/res/ds + 0.5)$$

$$s_{max} = r_{max} \times ds \tag{5.2}$$

$$\lambda = 12.2639/\sqrt{V \times 10^3 + V^2 \times 0.97845}$$

$$f = (p - \tau)/(s^2_{max} \times \lambda \times 10) + f_{s4} = r/(s^2_{max} \times \lambda \times 10) + f_{s4}$$

where $apix$ is the sampling (Å/pixel) of the original image (Fig.5.1A), $b$ is the size of the small square patch in periodogram averaging (see Step 1), $res$ is the specific resolution at which the 2D power spectra is truncated (i.e. 5.3 Å in this example, see Step 2), $\lambda$ is the electron wavelength, $V$ is the accelerating voltage (kV) of TEM, $f$ is the defocus, $p$ is the peak position (pixels) of each row in Fig.5.1F, $\tau$ is the amount of Fourier shifting (pixels) in Fig.5.1E and $r$ is the radius of the single ring along one direction in Fig.5.1D, $f_{s4}$ is a small adjustment factor to compensate the contribution of the $s^4$ term in the CTF model. In Fig.5.1G, the X-axis indicates the angles from 0 to $2\pi$ and Y-axis indicates the defocus (nm) after conversion from peak positions.

The wavy curve shown in Fig.5.1G represents the angular distribution of the defocuses. The average would equate the mean defocus. The apparent oscillations would indicate noticeable defocus variations and thus astigmatism in the image. To obtain the astigmatism amplitude and angle, we performed FFT of this 1D curve. Fig.5.1H shows the amplitude of the first ten elements in the FFT. It is well-established that the twofold astigmatism is the primary astigmatism of objective lens, which is consistent with the dominant second element in Fig.5.1H. To demonstrate the second

element is sufficient to depict the curve in Fig.5.1G, we only kept the second element in the complex array and zeroed all other complex numbers, then performed inverse FFT and produced a smooth 2-cycle sine wave (red curve in Fig.5.1G). As can be seen in Fig.5.1G, there is a good agreement between the curve of measured defocuses (greenish region) and the 2-fold astigmatism curve (red curve). Thus, we can directly estimate the astigmatism amplitude and angle from the amplitude and phase of the second complex number, respectively.

In addition, the astigmatism angle derived from the phase of the second complex number is marked in Figs.5.1B-E by red arrows. It is noted that the major axis direction of the single ring (red arrows in Fig.5.1D and E) is the maximum defocus direction since the distance to the center here is linearly proportional to the defocus. This direction corresponds to the minor axis direction in the Thon rings (red arrows in Fig.5.1B and C) along which the defocus is the largest and the Thon rings oscillate fastest.

**Step 7:** Final "radar"-style display to provide real-time feedback for the correction of astigmatism

Fig.5.1I and J the final plots of our script that serve as a visual guide for the adjustment of objective lens stigmators. As shown in Fig.5.1I, the intensity of pixels along the major and minor directions of the shifted single ring are plotted in blue and red, respectively. According to Eq.5.2, these 2 peaks are converted to defocus in the major and minor directions and the difference between them is the astigmatism amplitude. Defocus, astigmatism amplitude and astigmatism angle are displayed at the top of Fig.5.1I and all of them closely match the set values of the simulated image. When astigmatism is reduced, these two peaks will move closer and overlap at astigmatism-free conditions.

The radar plot in Fig.5.1J provides a big picture for the correction of astigmatism. The black dot in Fig.5.1J indicates the current astigmatism with the distance to the center proportional to the astigmatism amplitude and the angle from +X axis equal to the astigmatism angle. Based on this plot, it is intuitive to recognize the astigmatism

amplitude and angle and the trajectory of astigmatism in response to the tuning of objective lens stigmators. When astigmatism is perfectly corrected, the black dot will move to the center of the plot.

### 5.2.2 Implementation

The algorithm described here has been implemented as a stand-alone DigitalMicrograph script *s2stigmator.s*, allowing immediate processing of live images acquired by the computer controlling the microscope. The script was tested on DigitalMicrograph version 1.85 on our CM200 and version 2.32 on our Titan Krios microscopes. Fig.5.1B-J and similar plots in other figures were generated by this DigitalMicrograph script. An offline version of our algorithm has also been written as *s2ctf.py*, a Python script for batch determination of the CTF parameters of images post acquisition. Both scripts will be available on our web site (http://jiang.bio.purdue.edu).

### 5.2.3 Test datasets

Our method was first validated by simulated images generated using EMAN2 library functions [12]. Images were synthesized by projecting a herpesvirus density map (EMD-3358) into arbitrary orientations. A CTF function was applied to each synthetic image with a specific setting of defocus, astigmatism amplitude and angle.

Furthermore, four published experimental datasets (EMPIAR-10038, EMPIAR-10031, EMPIAR-10028 and EMPIAR-10002) of experimental datasets were downloaded from EMPIAR [57] to test the performance of our method using these real micrographs. The estimated defocus, astigmatism amplitude and angle values were cross-validated with CTFFIND3 method [39].

Finally, the real-time feedback was tested using live images of carbon film obtained on our CM200 at 200 kV and FEI Titan Krios at 300 kV, and recorded on a Gatan CCD and K2 Summit camera, respectively. In order to test the dose sensitivity of our script, 10 movies (22 frames per movie) were collected with a dose rate of 9

electron/pixel/s (eps) and 0.5 s exposure time for each frame using the Titan Krios microscope and K2 camera. Subsequently, motion correction [74] was performed on different numbers (e.g. the first two frames, the first four frames, etc) of frames in each movie to generate 11 groups of images with different total dose. These images were analyzed using our method and the root-mean-square deviation (RMSD) of the results in each dose group was calculated to find the appropriate dose range for our method. In addition, videos were recorded on CM200 and Titan Krios to demonstrate the performance of our method and illustrate the recommended strategy of astigmatism correction. The anisotropic magnification distortion was corrected from the live images according to the previously measured parameters [75].

## 5.3 Results

### 5.3.1 Validation of $s^2stigmator$ method using simulated images

To substantiate the performance of our $s^2stigmator$ method, we first tested four simulated images obtained as described in Section 5.2.3 with different combinations of CTF parameters. Fig.5.2 displays the results for the simulated images with different values of defocus, astigmatism amplitude and angle. The first row (Fig.5.2A-D) contains images of the single ring before Fourier shifting. It can be seen that the size of the ring increases as defocus increases, however, it is very hard to visually examine the ellipticity. It is worth pointing out that Fourier shifting (Step 4 in Section 5.2.1) is essential for the detection of ellipticity, especially at small defocus (Fig.5.2A). The second (Fig.5.2E-H) and third (Fig.5.2I-L) rows are the standard plots of our script. Apparently, the offsets between the red and blue peaks (Fig.5.2E-H) becomes larger with increased astigmatism. The black dots (Fig.5.2I-L) also become more distant to the origin in the radar plots. Both types of plots accurately reflect the defocus, astigmatism amplitude and angle of the corresponding images. The astigmatism angles are marked using red arrows in Fig.5.2A-D.

Next, we tested our method on more simulated images with varying defocus (500 nm $\sim$ 4000 nm), astigmatism amplitude (20 nm $\sim$ 120 nm) and angle (5° $\sim$ 175°). The tight clustering along the diagonal lines in Fig.5.3 confirms a good consistency between the determined defocus (Fig.5.3A), astigmatism amplitude (Fig.5.3B), astigmatism angle (Fig.5.3C) and their true values used in simulations, respectively. Hence, all these tests with simulated images fully support the accuracy and reliability of our method.

## 5.3.2 Cross-validation with experimental images

Our algorithm was also cross-validated using experimental datasets. Fig.5.4 shows the results of four images, one from each experimental datasets, EMPIAR-10038 (the first column, Fig.5.4A, E, I), EMPIAR-10031 (the second column, Fig.5.4B, F, J), EMPIAR-10028 (the third column, Fig.5.4C, G, K) and EMPIAR-10002 (the fourth column, Fig.5.4D, H, L). The astigmatism angles are also marked in the figures of single rings (Fig.5.4A-D) using red arrows. As can be seen from Fig.5.4B-D, the elliptical shape is not obvious due to the relatively small size of the single ring even though the astigmatism is large (>200 nm in Fig.5.4B), emphasizing the importance of Fourier shifting in the detection of ellipticity. Due to the varying combinations of $apix$, box size ($b$ in Eq.5.2) and the specific resolution ($res$ in Eq.5.2) at which the 2D power spectra is truncated, the radius of the single ring is widely variable. The Fourier shifting is able to deal with the variability of the ring size by shifting/enlarging the radius large enough (such as Fig.5.1E) to reliably measure the ellipticity. The distance between the two peaks (Fig.5.4E-H) clearly indicates the amount of astigmatism. These results were then cross-validated using CTFFIND3 of which the results are shown as red dots in the radar plots (Fig.5.4I-L). The superimposition of black and red dots indicates a good agreement between the results independently derived from these two approaches.
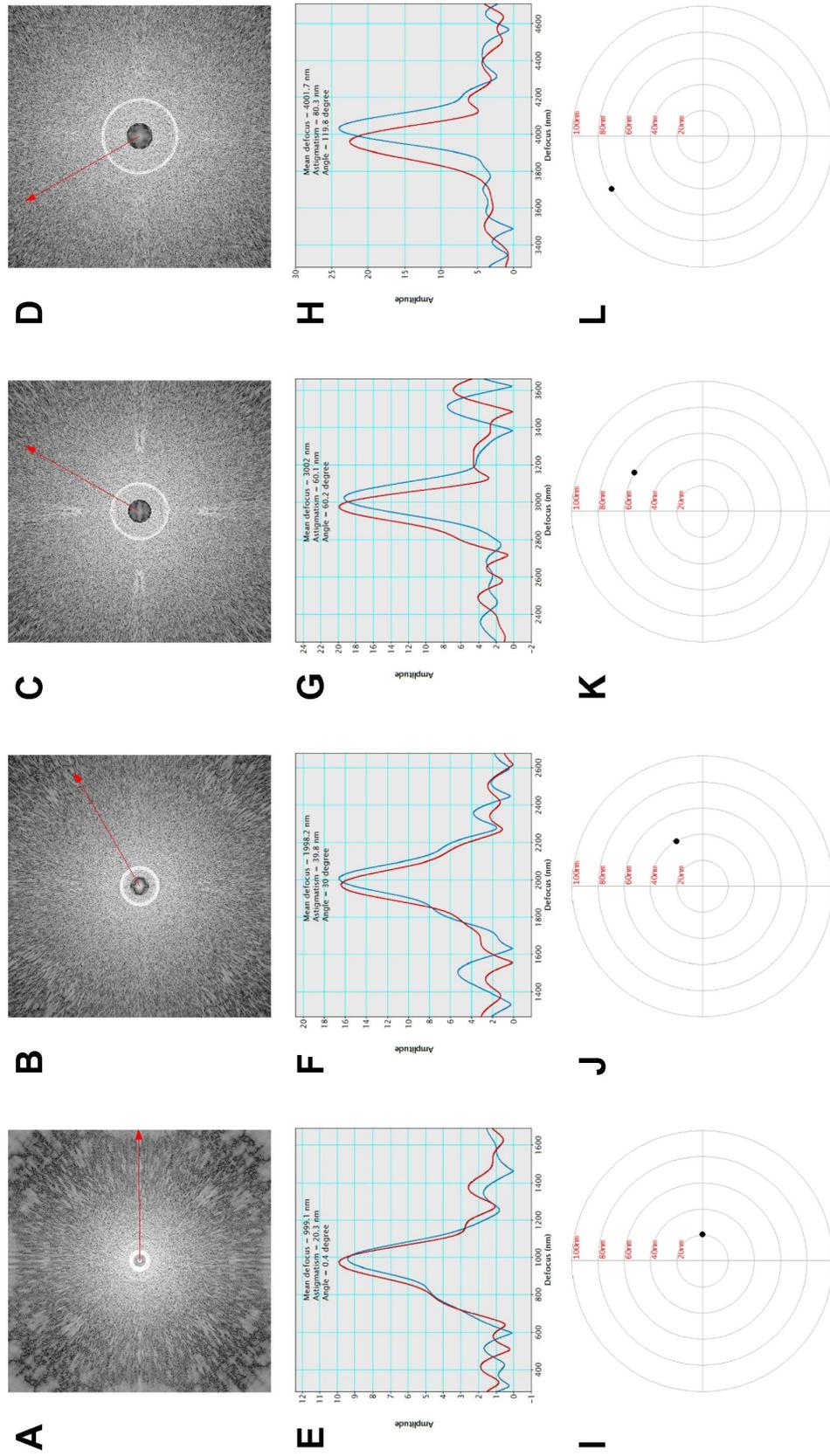
Fig. 5.2. Comparison of results of four representative simulated images with varying defocus, astigmatism amplitudes and angles. Columns from left to right are results for synthetic images with defocus 1000 nm, astigmatism amplitude 20 nm and angle 0° (A, E, I), defocus 2000 nm, astigmatism amplitude 40 nm and angle 30° (B, F, J), defocus 3000 nm, astigmatism amplitude 60 nm and angle 60° (C, G, K), and defocus 4000 nm, astigmatism amplitude 80 nm and angle 120°, respectively. For all simulated images, voltage = 300 kV, $C_s$ = 2 mm, pixel size = 1 Å, B-factor = 150 Å$^2$, amplitude contrast = 0.1 and image size is 2048 × 2048 pixels. The high resolution limit is 5.3 Å for 2D $s^2$ power spectra. The first row (AD) illustrates the images of the single rings before Fourier shifting, the second (EH) and third row (IL) are the standard outputs, representing peak plots and radar plots, respectively.
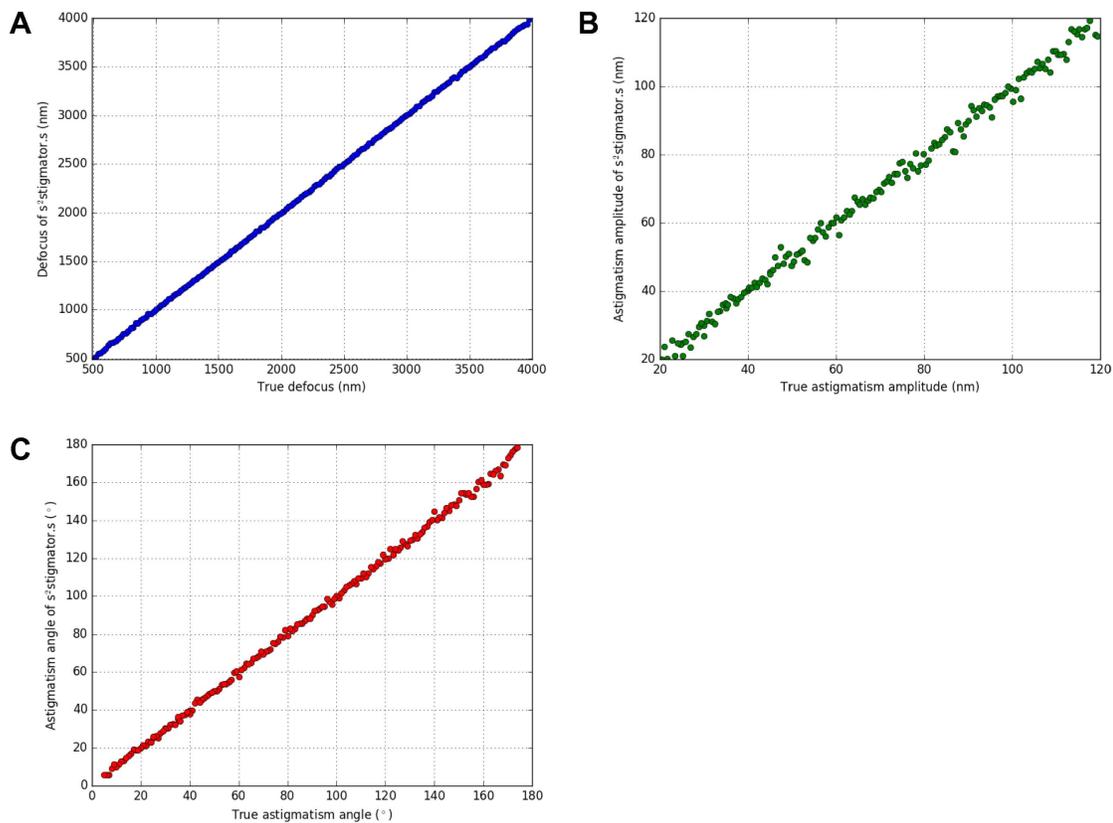
Fig. 5.3. Validation tests using simulated images with varying defocuses, astigmatism amplitudes and angles. The ranges of defocus, astigmatism amplitude and angle are 500 nm $\sim$ 4000 nm, 20 nm $\sim$ 120 nm, and $5° \sim 175°$, respectively. For all simulated images, voltage = 300 kV, $C_s$ = 2 mm, pixel size = 1 Å, B-factor = 150 Å$^2$, amplitude contrast = 0.1 and image size is 2048 × 2048 pixels. The high resolution limit ranges from 4 Å to 5.5 Å for 2D $s^2$ power spectra. The measured defocus (Y-axis in A), astigmatism amplitude (Y-axis in B) and astigmatism angle (Y-axis in C) are compared with the corresponding ground truth values (X-axis in A-C).

Based on the above results, our method is able to reliably determine the astigmatism and defocus of both simulated and experimental images. To further corroborate this conclusion, we used our method and the CTFFIND3 method to test larger number of experiment images (i.e. the four experimental datasets listed in Section 5.2.3, 100 micrographs in each). In Fig.5.5A, there is an excellent agreement between these two methods in the estimation of defocus. The results of astigmatism amplitude and angle are also clustered along the diagonal lines in Fig.5.5B and C although with small spread. Considering the challenges in estimating astigmatism and the large discrepancies among the different CTF fitting methods found by the CTF Challenge [76], such small spread shown in Fig.5.6B and C can be considered excellent agreement. As these two astigmatism determination methods are dramatically different in theory, overall consistent results shown in Fig.5.5 suggest that the results by both methods are correct although the ground truth is unknown.

### 5.3.3  Performance tests with live images

For a comprehensive evaluation of the real-time feedback of our $s^2 stigmator$ method, we tested our script with live images acquired as described in Section 5.2.3. We first analyzed the reproducibility (RMSD of astigmatism amplitude and angle) of our method with images acquired at different doses. As shown in Fig.5.6A, it is evident that the RMSD is quickly reduced as the dose increases. The RMSD decreases to less than 1 nm when the dose is higher than 10 $e/\text{Å}^2$, a dose level that is smaller than the dose used by typical single particle cryo-EM images. Fig.5.6B-D display the distribution of results from 10 measurements for total dose of 5.2 $e/\text{Å}^2$, 10.3 $e/\text{Å}^2$ and 56.9 $e/\text{Å}^2$, respectively, which correspond to the first, second and last points in Fig.5.6A. It is apparent that the points are tightly clustered in Fig.5.6C and D, demonstrating that the measurement error ($<1$ nm) of our method can be essentially ignored for doses larger than 10 $e/\text{Å}^2$.
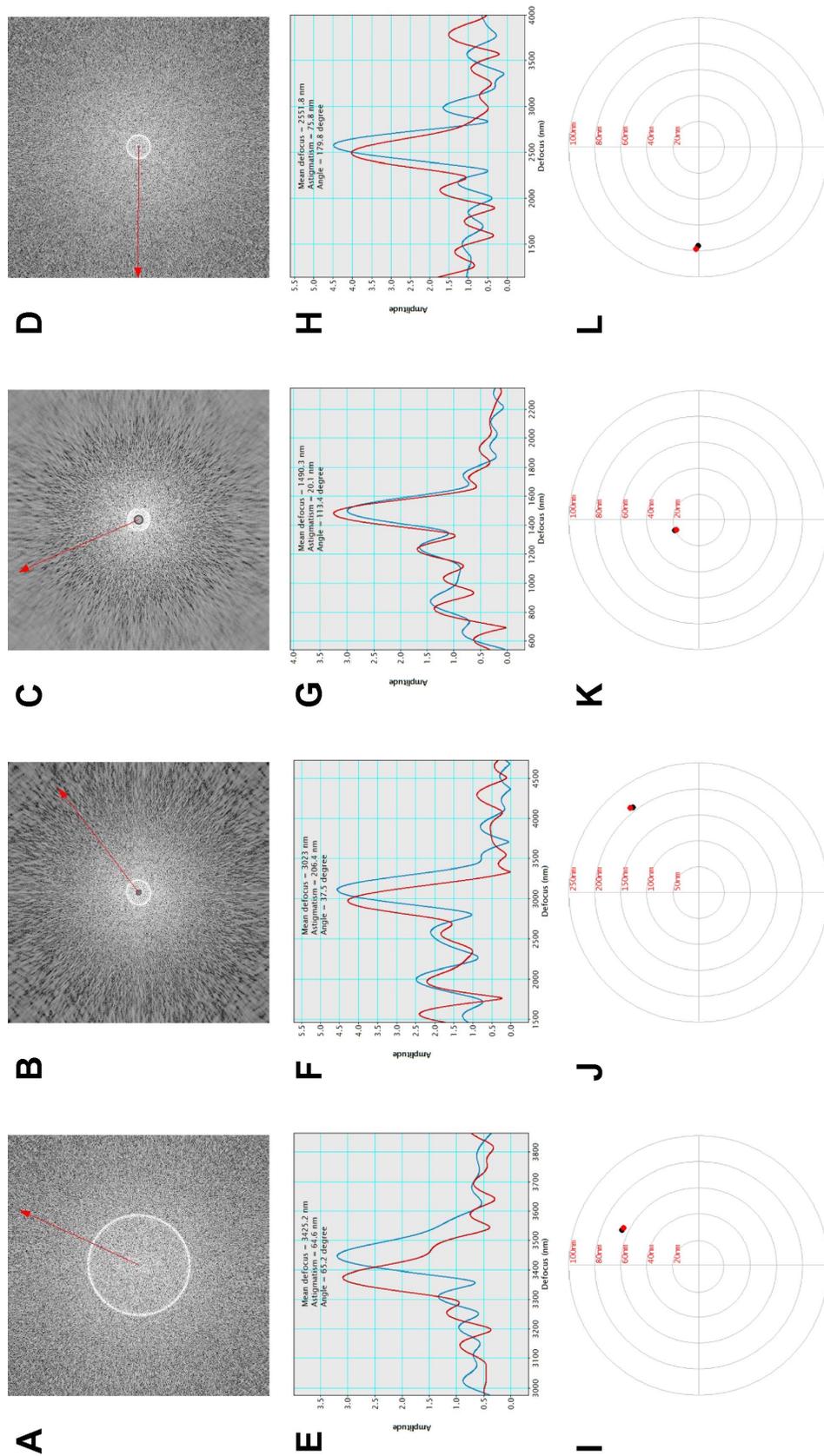
Fig. 5.4. Results for four representative experimental micrographs with varying defocuses, astigmatism amplitudes and angles. Columns from left to right are results for images downloaded from EMPIAR-10038 (A, E, I), EMPIAR-10031 (B, F, J), EMPIAR-10028 (C, G, K), and EMPIAR-10002 (D, H, L), respectively. In the top row (AD), the red arrows indicate the direction of the major axis (i.e. largest defocus) found by $s^2 stigmator$. In the radar plots of the last row, black dots represent the results of $s^2 stigmator$ and red dots representing the results of CTFFIND3 were manually added to the radar plots for comparison.
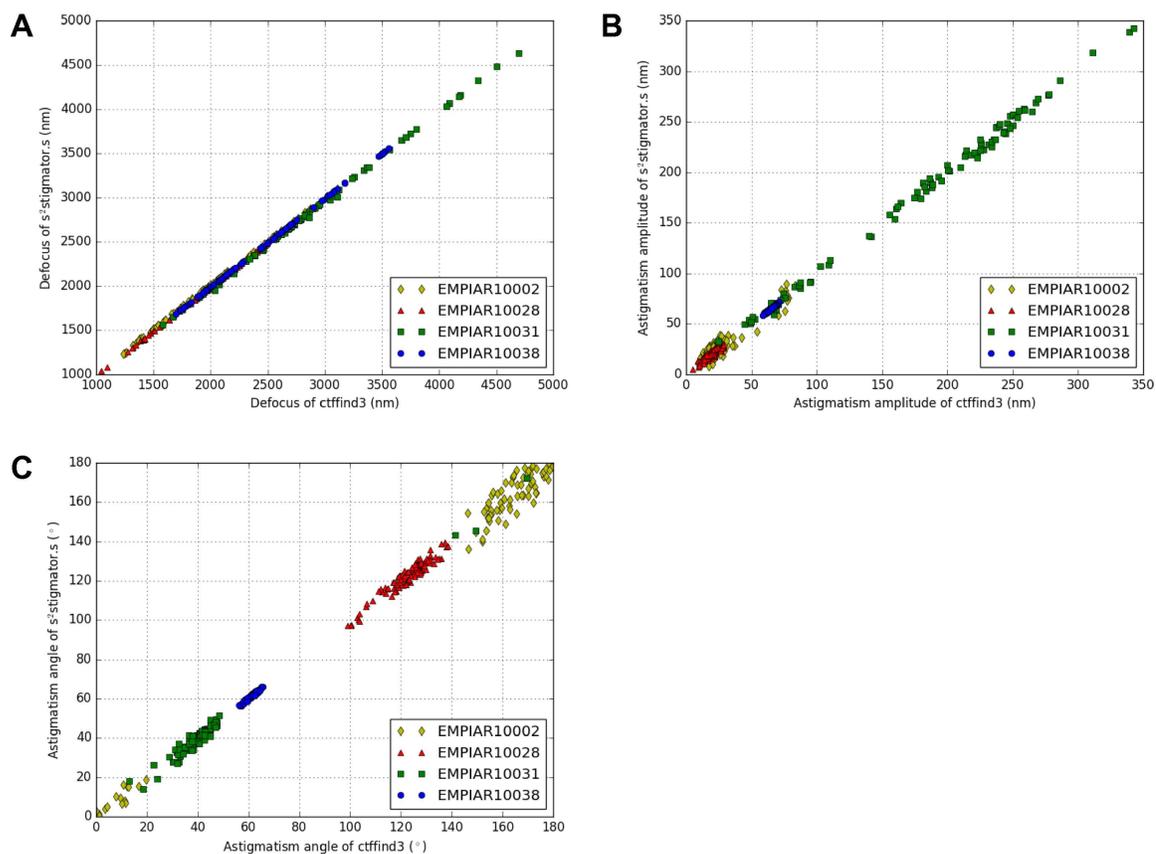
Fig. 5.5. Cross validations of $s^2stigmator$ and CTFFIND3 method using experimental images. Shown are the estimated defocus (A), astigmatism amplitude (B) and astigmatism angle (C) of experimental micrographs. 100 micrographs in each of the four datasets, EMPIAR-10038, EMPIAR-10031, EMPIAR-10028 and EMPIAR-10002, were used for the tests.
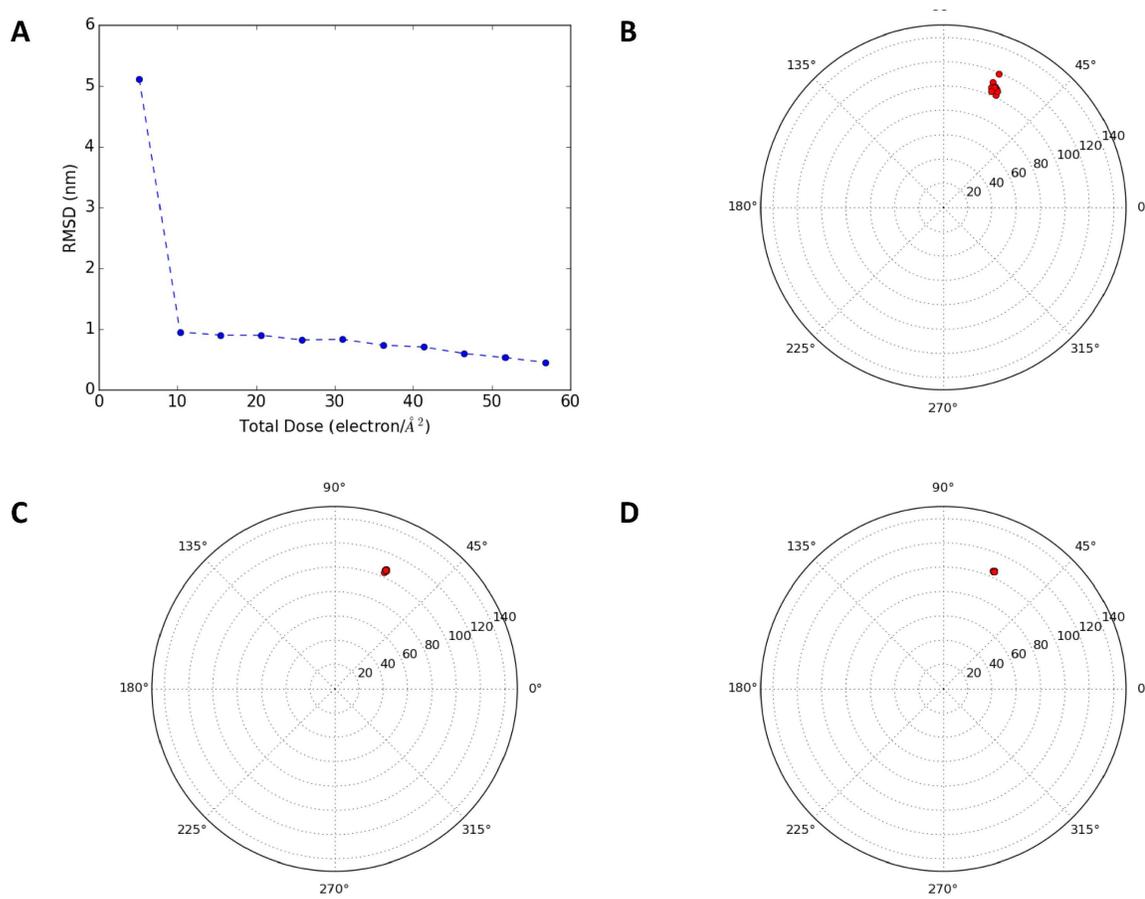
Fig. 5.6. Reproducibility tests with images of different electron doses. (A)
The RMSD of results for 10 images at each of the tested dose levels. (B-D)
The distribution of the 10 results for total dose of 5.2 $e/\text{Å}^2$ (B), 10.3 $e/\text{Å}^2$
(C), and 56.9 $e/\text{Å}^2$ (D), respectively. The corresponding RMSD of these
three distributions are shown as the first, second and last point in (A).

### 5.3.4 Single-pass tuning strategy for adjusting objective lens stigmators

The excellent reproducibility of our method for live images allowed us to use its radar plot (Fig.5.1J, Fig.5.2I-L, Fig.5.4I-L, Fig.5.6B-D) as a visual guide for real-time minimization of objective lens astigmatism by adjusting the two objective lens stigmators (i.e. M-X and M-Y knobs). By observing how the points move in the radar plot after adjustments of these two stigmators, we quickly realized that the two stigmators independently controls the astigmatism amplitude (by M-X) and angle (by M-Y), instead of the X or Y components of the astigmatism as one would infer from the labels of the two knobs. This observation has led us to understand the relationship of the objective lens astigmatism and the correction field generated by the stigmators (Fig.5.7A), and how the stigmators can be controlled to optimally compensate the astigmatism (Fig.5.7B and C).

In the vector diagram (Fig.5.7A), the residual astigmatism of the objective lens is represented by a solid vector. The correction field generated by the objective stigmators is represented by a dashed vector. The M-Y stigmator would simply rotate the dashed vector while the M-X stigmator would change the length of the dashed vector. The objective lens astigmatism would be perfectly corrected when the dashed vector is adjusted to have the same length but in opposition direction to the solid vector (Fig.5.7C). The challenge is thus how to arrive at the optimal correction state.

Based on this understanding, we designed a single-pass tuning strategy (Fig.5.7D). *Initial state.* The yellow star indicates the initial state in which the two vectors have an arbitrary relationship (length and angle) as shown Fig.5.7A. *Step 1.* The angular knob (M-Y) is continuously adjusted to find the optimal angle. In the radar plot, the points representing the residual astigmatism should exhibit an arc-like trajectory as represented by the black markers (black circles and black triangle) and the long black arrow in Fig.5.7D. The position in the arc closest to the origin (black triangle in Fig.5.7D) will be the optimal M-Y setting that makes the dashed vector in opposite

direction of the solid vector (Fig.5.7B). This effect of angular knob adjustment can be explained using vectors in Fig.5.7B in which the dashed vector is continuously rotated until it has sampled the optimal angle (i.e. opposite to the solid angle). *Step 2.* The angular knob is turned backward to bring the point back to the optimal point (black triangle). This part of trajectory can be represented by the short black arrow in Fig.5.7D. This will bring the dashed vector in opposite direction of the solid vector again (Fig.5.7B). *Step 3.* The amplitude knob (M-X) is continuously adjusted. In this process, the resulted points in the radar plot should directly move to the origin along radial direction as shown by red markers (red squares and red triangle) and red arrow in Fig.5.7D. In this step, turning M-X knob will simply change the length of the dashed vector until its length is equal to that of the solid vector (Fig.5.7B and C). The red triangle located at the origin of Fig.5.7D means that the astigmatism of objective lens has now been completely corrected (Fig.5.7C).

### 5.3.5   Validation of the single-pass tuning strategy

To validate the strategy illustrated in Fig.5.7, we have tested it on our CM200 and Titan Krios microscope using live images. Fig.5.8 and Fig.5.S1 show screenshots of the process collected from Titan Krios (Fig.5.8, Fig.5.S1A) and CM200 (Fig.5.S1B, C) in which the darkest circle is the current point and the gray circles are historical points. To make the sequence of the points more obvious, the greyness level of the circles is varied to make more recent ones darker. It is noted that the real-time trajectory in Fig.5.8 perfectly matches the theoretical prediction in Fig.5.7, thus confirming the validity and feasibility of the proposed strategy for correction of the astigmatism of the objective lens. In addition to the screenshot shown in Fig.5.8 and Fig.5.S1, we have also recorded videos (Video S1 and Video S2) that include the entire tuning process to demonstrate the strategy in its fullest details.

To further validate the single-pass tuning strategy, we tested our method with different initial astigmatism values by using different initial stigmator values. Fig.5.S2
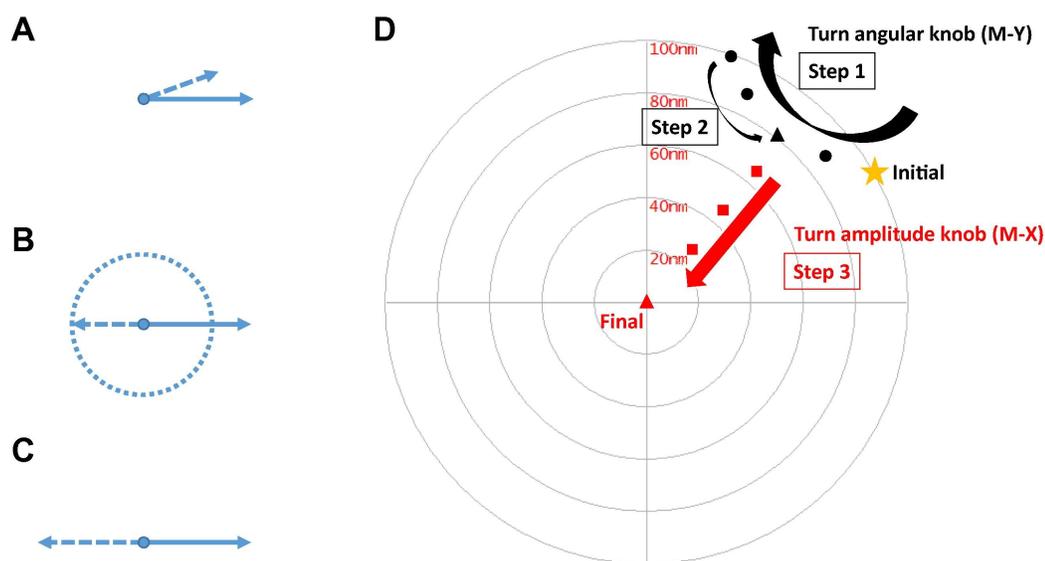
Fig. 5.7. A new single-pass strategy of astigmatism correction using $s^2$ *stigmator*. (AC) Vector diagrams to illustrate the principle of correction of astigmatism by the two objective lens stigmators. (A) Initial state. The correction vector (dashed) is at an arbitrary state (length and direction) relative to the residual astigmatism vector (solid). (B) Angle search. The angle for the correction vector (controlled by M-Y knob) is optimal when it is at the opposite direction of the astigmatism vector. (C) Amplitude minimization. The astigmatism is corrected when the length of the correction vector (controlled by M-X knob) is changed to the same length of the astigmatism vector. (D) A marked radar plot showing the expected trajectory of data points representing the astigmatism in the process of astigmatism correction using our new tuning strategy. The yellow star suggests the initial state (A) before correction. The black markers forming an arc-like trajectory (black arrows) indicate the change of data points if only the angular knob (M-Y) is turned (as explained in B). To find the optimal angle (shown as the black triangle) that results in minimal astigmatism in the black trajectory, the M-Y angle knob needs to cover sufficiently wide range to clearly reveal a data point (black triangle) in the arc that is closest to the origin. Starting from the black triangle (i.e. optimal M-Y setting), only the M-X amplitude knob is adjusted, which will result a series of red markers following a straight path represented by the red arrow. The red triangle located at the origin of the radar plot implies the complete elimination of objective lens astigmatism.

displays 8 screenshots of the entire trajectories acquired from CM200 microscope. The first row (Fig.5.S2A-D) are the trajectories with different initial astigmatism values in the first quadrant of the radar plot. It is noted that the direction of the straight trace segment does not change with varying initial astigmatism amplitudes and the angles are all around 62° (green dash lines in Fig.5.S2A-D). This can be explained by the vector diagram shown in Fig.5.7: the dashed vector (Fig.5.7B) controlled by the stigmators at optimal correction angle is always opposite to the solid vector (Fig.5.7B) representing the residual astigmatism of the objective lens, which is independent of the stigmator values. The plots in the second row (Fig.5.S2E-H) also have a similar property as the first row, however, the angle of the straight trace segment is around 152° (green dash lines in Fig.5.S2E-H), orthogonal to that in the first row. The 90° change of the orientation is due to the over-correction of the astigmatism which causes the switch of the major and minor axis in the power spectra (Fig.5.1B-C) and the single ring (Fig.5.1D).

## 5.4   Discussion

In this paper, we have introduced a new method, $s^2 stigmator$, for real-time detection of astigmatism in live images and established a reliable single-pass strategy for users to correct the astigmatism and improve image quality during TEM operation. Systematic tests have demonstrated that this method works accurately and robustly with both synthetic and experimental images. Owing to the closed-form algorithm without the need of iterative search of multiple parameters, the method is inherently fast and it takes only a few seconds on a desktop computer to measure the astigmatism of one image even as a script in DigitalMicrograph software. The speed can be significantly enhanced by converting to compiled code and using GPU acceleration in the future.

The $s^2$ power spectra provides the essential basis for our fast closed-form algorithm. Due to the uniform oscillation property of $s^2$ power spectra, the problem of
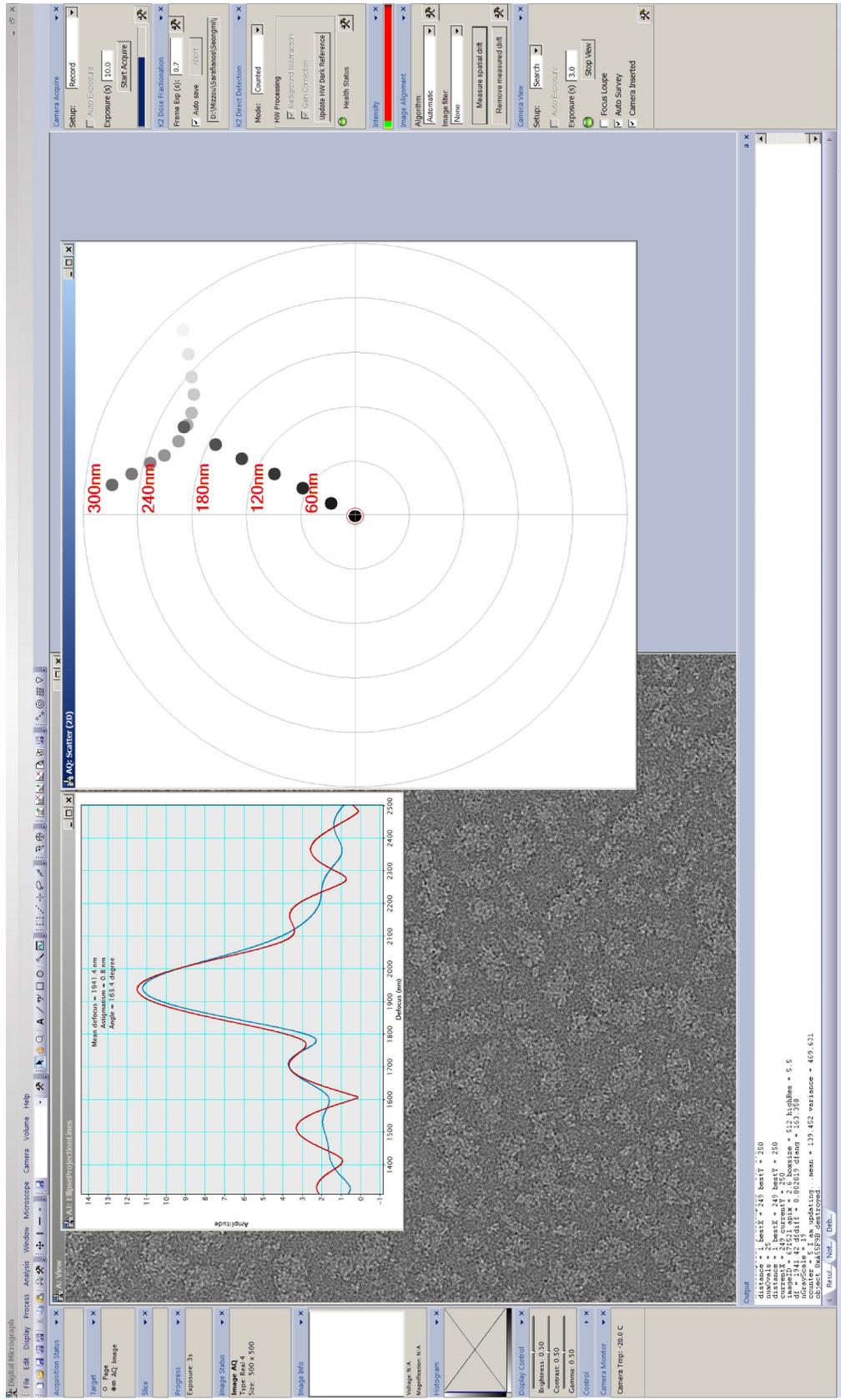
Fig. 5.8. The current point is marked by a black circle and all gray circles exhibit the trajectory of previous data points in the process of astigmatism correction. The grayness level of the circles is progressively reduced for data points further in the history of the correction process. A red circle (overlapped with the current point here) is used to indicate the currently best point with minimum astigmatism. Additional screenshots from Titan Krios and CM200 microscopes are shown in Fig.5.S1.
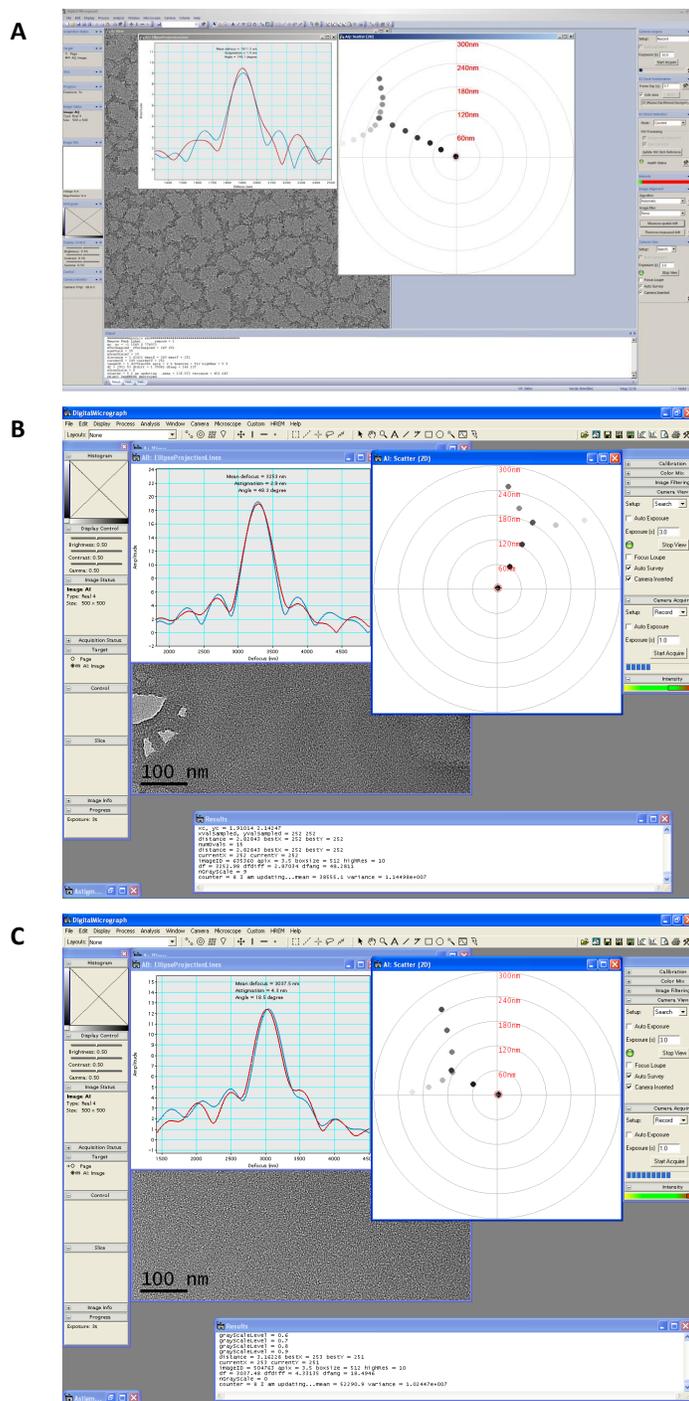
Fig. 5.S1.  Additional screenshots of the radar plot by $s^2stigmator$ as shown in Fig.5.8.  (A) Screenshot from Titan Krios microscope.  (B-C) Screenshots from CM200 microscope.
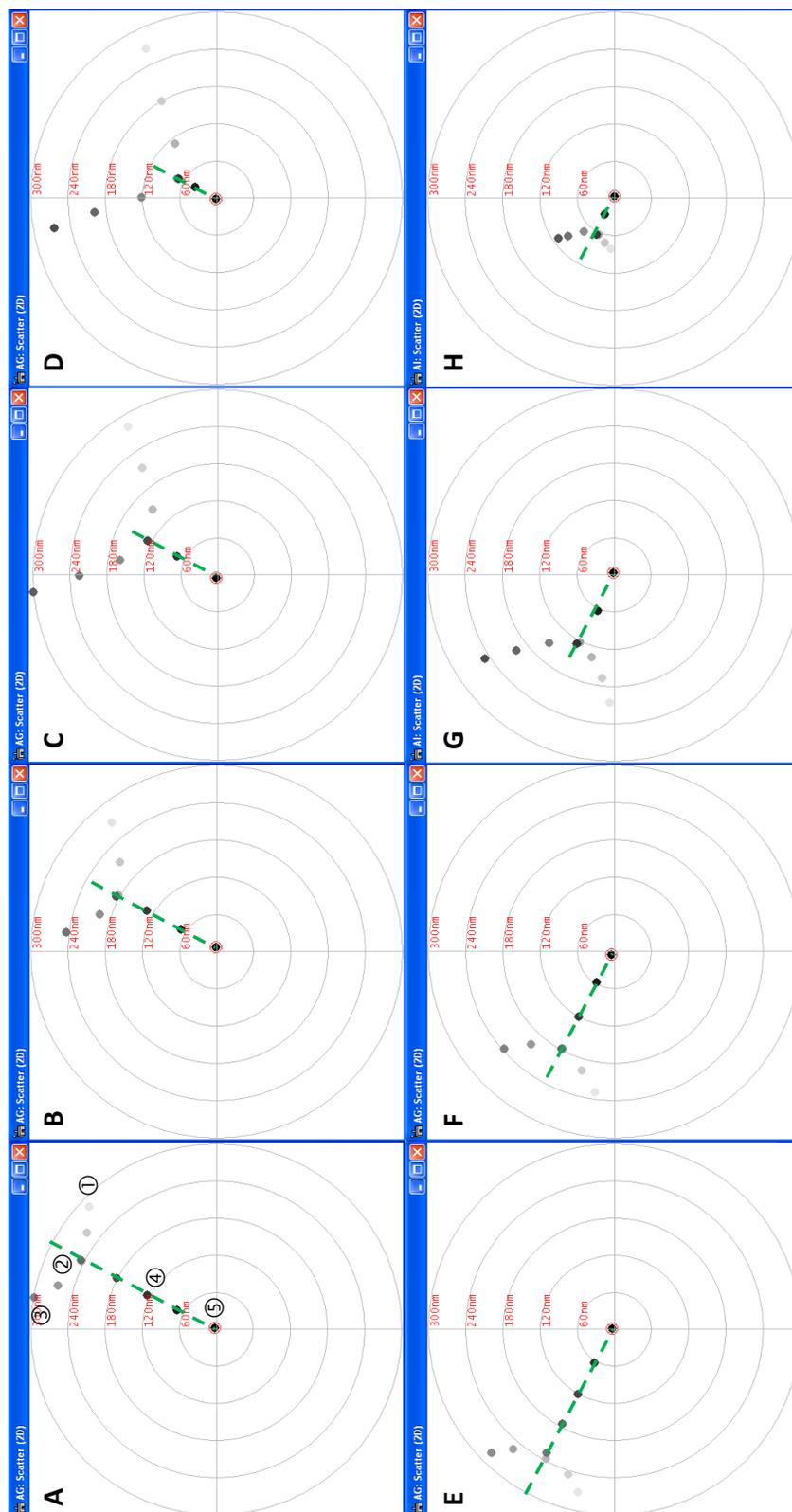
Fig. 5.S2. Performance of the single-pass tuning strategy with varying initial stigmatism. The 8 plots show the entire trajectories at a nominal magnification of 66,000× using CM200 microscope when the astigmatism is under-corrected (A-D) and over-corrected (E-H) at the initial point. The green dash line in each plot represents the angle of the straight trace segment in the trajectory. The points marked by the circled numbers in (A) correspond to the stigmator M-X/M-Y values of -2500/+350 (①), -2500/+50 (②), -2500/-250 (③), -2300/+50 (④), and -2100/+50 (⑤), respectively.

determining the defocus along all directions is drastically simplified to measuring the radii of a single ring (Fig.5.1D). This is in stark contrast to existing methods, such as dividing the 2D power spectra into sectors for angular rotational average [63, 69], applying edge detection method to approximate the ellipticity of the Thon rings [65], studying the mathematical relationship between the radial averages of TEM images with and without astigmatism [62], and performing a fully 2D power spectra optimization to fit CTF model [39, 67]. The single ring not only completely maintains the information of ellipticity but also becomes much easier and faster to evaluate.

To the best our knowledge, our method is the first to employ the 2nd Fourier shifting theorem (Eq.5.1) in TEM image processing, although the 1st Fourier shifting theory ($f(t - t_0) \leftrightarrow e^{-j2\pi st_0} F(s)$) has already been commonly used in the cryo-EM field. The application of Fourier shifting effectively overcomes the limitation of coarse sampling caused by small single ring at small defocus, and notably improves the accuracy of ellipticity detection. Although the size of the single ring is enlarged after Fourier shifting, the angular variation of the radius is unchanged due to the constant shift in all directions.

The single ring representation of the angular distribution of the defocuses also allows the use of 1D FFT to directly compute the amplitude and phase of a sine-like curve (red curve in Fig.5.1G), which correspond to the astigmatism amplitude and angle of the twofold objective lens astigmatism. The accurate direct solution of the astigmatism amplitude and angle not only eliminates the need to explicitly search for major/minor axes of the defocus ring but also avoids the errors due to noise (Fig.5.1G) in assigning the radii (i.e. defocus) and angles of the major/minor axes.

Most importantly, our script provides real-time radar-style visual feedback and quantitative values for the correction of objective lens astigmatism. Compared to current approach relying on visual examination of the roundness of Thon rings, our new method considerably enhances the sensitivity in ellipticity determination, and simultaneously prevents the bias and subjective results from operators. Furthermore, we have established an efficient single-pass tuning strategy (Fig.5.7), which allows

the users to rationally and sequentially adjust, instead of blindly playing with, the two stigmator knobs to correct the objective lens astigmatism. Combined with our results of real-time tests (Fig.5.8, Fig.5.S1, Video S1 and Video S2), we have also gained insights to the underlying principle of astigmatism correction (Fig.5.7).

Currently, most microscopists correct astigmatism at very high magnification and very small defocus, which is necessary due to the limited sensitivity of current approach relying on visual examination of the roundness of Thon rings. However, the imaging condition that astigmatism is corrected is drastically different from the conditions, magnification and defocus, used for data collection. In contrast, our method is capable of guiding users to correct astigmatism at any magnification and any defocus, and it thus allows the users to correct astigmatism at the same magnification and defocus level used for final imaging. We think that this is not only more convenient but also likely to further reduce the residual astigmatism in the final images.

# 6. DEFOCUS AND MAGNIFICATION DEPENDENT VARIATION OF TEM IMAGE ASTIGMATISM

## 6.1 Introduction

Cryo-electron microscopy (cryo-EM) has become a powerful technique for structural studies of macromolecular complexes and assemblies at near-atomic resolutions. Good alignment of the microscope, such as the gun, condensers, apertures, beam tilt, coma [77, 78], astigmatism, *etc.*, is a prerequisite to reaching optimal illumination lens and imaging lens conditions for high resolution TEM imaging. Currently, microscope alignments are still performed manually and rely on visual, qualitative feedback. Moreover, manual microscope alignment requires extensive training and experience but still often suffers from suboptimal efficiency and quality. Minimizing astigmatism of the objective lens is an indispensable daily instrument alignment task essential for high resolution TEM imaging.

Astigmatism of the objective lens represents the angular dependency of defocus. 2-fold astigmatism is the major type of astigmatism relevant to cryo-EM, which results in the elliptical elongation of Thon rings [61] in the power spectra of TEM images. Currently, many microscopists visually correct astigmatism at large magnifications and small defocuses, then switch to a lower magnification to collect data [79, 80] and intentionally vary defocus to sample the entire reciprocal space and even-out the zero-nodes of the contrast transfer function (CTF) [73, 81, 82]. Similarly, many microscopists also use a magnification in the focus mode larger than that of the exposure mode during low-dose imaging. The implicit assumption of these strategies is that the astigmatism is invariant to the change of magnification and defocus. Conversely, other microscopists prefer to use the same magnification for astigmatism correction and in the focus mode as that used in the exposure mode during data

collection. Unfortunately, there have been no systematic studies to test these two different opinions. In this study, our quantitative studies have conclusively shown the defocus/magnification dependence of astigmatism and validated the latter approach as the preferred method.

Due to the poor sensitivity of human eyes, microscopists have to rely on large magnifications and small defocuses when they visually examine the roundness of Thon rings in the 2D power spectra displayed on a computer screen and iteratively adjust the two objective lens stigmators (we use MX and MY to call the two stigmators as they are natural abbreviations of the Multi-functional knobs X and Y used to adjust the current of the two stigmators) to make the Thon rings as circular as possible. This tedious and subjective method is not only inaccurate and potentially biased by the astigmatism of human eyes, but also hampers the systematic quantification of the astigmatism variations for different magnifications/defocuses. To overcome this challenge, it is desirable to take advantage of a proper approach which can sensitively measure and accurately correct astigmatism at any imaging condition.

We have recently published a method, $s^2stigmator$ [83], with a single-pass tuning strategy, that allows rapid and sensitive detection of astigmatism using TEM live images and can reliably and efficiently guide the user to manually adjust the two stigmators to correct astigmatism. This new method opens up possibilities to minimize astigmatism with real-time feedback at a wide range of imaging conditions that are not available by visual examination. As shown in Fig.6.S1, the user would first adjust any one of the two stigmators (MX, MY) to find the optimal correction (points marked by blue arrows in Fig.6.S1) that gives rise to minimal astigmatism in the arc-shaped trajectory; then adjust the other stigmator to linearly approach the center at which the astigmatism is zero. Both sequences of stigmator adjustment (MY first, then MX or MX first, then MY) are able to minimize astigmatism with the similar shape of trajectories with 45° angular offset (Fig.6.S1), suggesting the two stigmators with 45° offset physically are functionally equivalent with the difference only in the direction of correction. In this article, we present systematic and quantitative inves-

tigations of astigmatism dependence on imaging conditions by employing $s^2stigmator$ to correct astigmatism and then measure the re-emerging of astigmatism after varying imaging conditions (defocus, magnification). Underlying physical principles are used to interpret the variability of astigmatism and its dependence on image conditions. Based on the findings of these studies, several recommendations are provided for instrument alignment and data acquisition to help maximally reduce astigmatism and improve high resolution imaging.

## 6.2  Methods

### 6.2.1  Experimental cryo-EM datasets for initial test

Our study started with nine datasets of experimental cryo-EM micrographs obtained from the public database EMPIAR [57]. For each dataset, we downloaded the particle parameter star file deposited by the authors (Fig.6.1A-H) or generated such a file using Gctf (Fig.6.1I) [84] and then examined the correlation between defocus and astigmatism using the values included in the star files.

### 6.2.2  Data collection for the study of defocus-dependent astigmatism

Next, the defocus-dependence of astigmatism was examined using live images of carbon film obtained on our CM200 microscope at 200 kV and FEI Titan Krios at 300 kV, and recorded on a Gatan US4000 CCD and K2 Summit camera, respectively. We minimized the objective lens astigmatism at small, medium and large defocuses using our $s^2stigmator$ tool, then increased or decreased defocus from the starting defocus used for astigmatism correction. At each defocus, ten images were collected and the defocuses, astigmatisms were calculated to obtain their mean and root-mean-square deviation (RMSD) values. The same experiment was repeated on different days to explore the reproducibility of defocus-dependence of astigmatism.
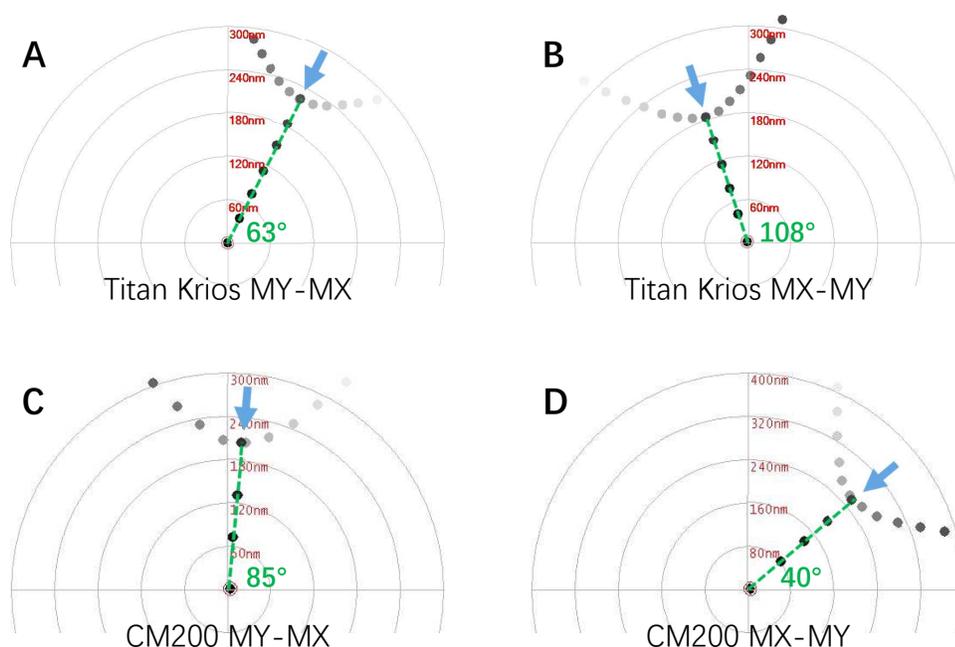
Fig. 6.S1. Performance of $s^2 stigmator$ method and the single-pass tuning strategy. (A, B) Two screenshots of the trajectory from Titan Krios microscope obtained at 1000 nm defocus, a nominal magnification of 22,500× and 25 $e/\mathring{A}^2$ dose with images recorded on a Gatan K2 Summit direct electron detector operated at counted mode using 15 e/pixel/second dose rate and 3s exposure time. Stigmator MY was adjusted first (arc-like segment) and then MX was adjusted (straight segment) in (A) while the opposite order (i.e. MX first and then MY) was used in (B). (C, D) Two screenshots of trajectory from CM200 microscope obtained at 1700 nm defocus, a nominal magnification of 115,000× and 40 $e/\mathring{A}^2$ dose with images recorded on a Gatan UltraScan 4k CCD with 3s exposure time. Stigmator MY was adjusted first (arc-like segment) and then MX was adjusted (straight segment) in (C) while the opposite order (i.e. MX first and then MY) was used in (D). The wide blue arrow indicates the optimal point with minimal astigmatism in the arc-like segment of each trajectory. The green dash line in each plot represents the angle of the straight trace segment in the trajectory, and the exact angle is marked in green next to the green dash line. Both sequences of stigmator adjustment (MY first, then MX or MX first, then MY) on the same instrument are able to minimize astigmatism with the similar shape of trajectories with 45° angular offset (63° in A v.s. 108° in B, and 85° in C v.s. 40° in D).

### 6.2.3 Data collection for the study of magnification-dependent astigmatism

Finally, the magnification-dependence of astigmatism was examined using the same sample and instrument as described above. To emulate the change of magnification used for astigmatism correction and data collection, we minimized the astigmatism at a high magnification using $s^2stigmator$, then successively lowered the magnifications. At each magnification, twenty images were collected, and their mean and RMSD of astigmatism and defocus were computed. In addition to the magnification-dependence of astigmatism, magnification-dependence of defocus was also simultaneously examined using the same set of data. The same measurement was also repeated multiple times to detect the stability of the relationship between astigmatism and magnification.

The magnifications were calibrated using polycrystalline gold sample grids. The anisotropic magnification distortion [75,85] was corrected from the live images according to the previously determined parameters before the astigmatism was calculated.

### 6.2.4 Data availability

All data generated or analyzed during this study are included in this published article.

### 6.3 Results

### 6.3.1 Observations of defocus-dependent astigmatism in experimental cryo-EM data

To examine the defocus-dependence of objective lens astigmatism, we first extracted the defocus and astigmatism parameters of the nine cryo-EM datasets as described in Section 6.2.1 and plotted the results in Fig.6.1. It can be clearly seen that there are positive, linear correlations between astigmatism and defocus, provid-
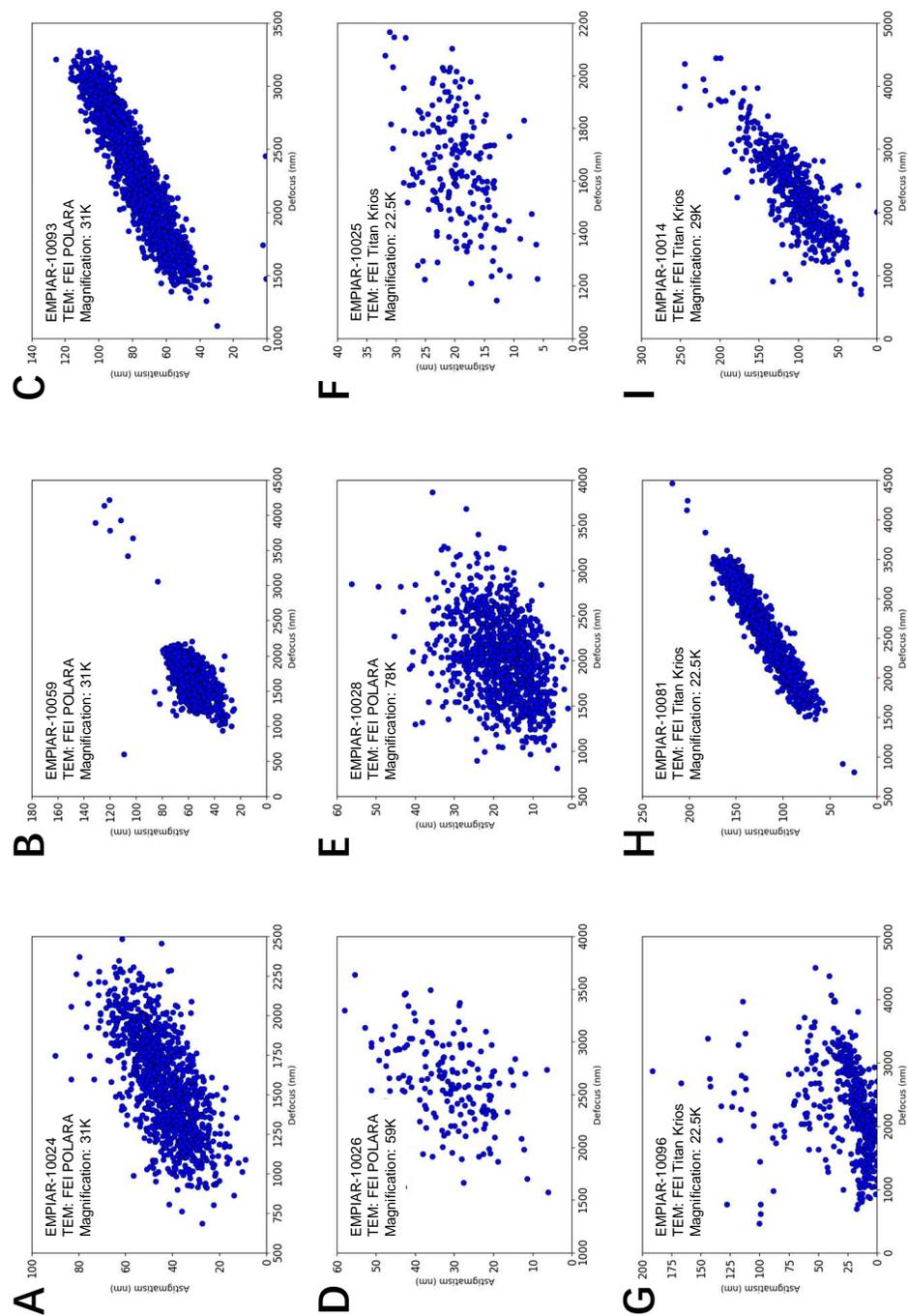
Fig. 6.1. Observations of the relationship between defocus and astigmatism in experimental cryo-EM datasets. The EMPIAR ID [57], instrument, and magnification are marked at the upper left corner of each plot. These datasets were collected by the cryo-EM facilities at the University of California, San Francisco (A-C), MRC Laboratory of Molecular Biology (D, E), Scripps Research Institute (F), New York Structural Biology Center (G), Rockefeller University (H), and HHMI Janelia Farm Research Campus (I), respectively.

ing evidence for defocus-dependent astigmatism in experimental cryo-EM data. It is worth pointing out that this correlation is a general phenomenon since it is observed in a wide variety of data, such as data from multiple research groups, microscopes in different facilities, a diversity of samples, and varying imaging conditions. As shown in Fig.6.1, the variations of astigmatism for different defocuses are significant (e.g. >100 nm) and distinct from dataset to dataset. Hence, it is desirable to comprehensively examine the dependence of astigmatism on imaging conditions (e.g. defocus, magnification) which are frequently changed in TEM alignment and during data acquisition.

### 6.3.2 Robust performance of $s^2 stigmator$ and the single-pass tuning strategy at different defocuses and magnifications

For quantitative measurement of astigmatism variations, astigmatism needs to be accurately corrected at a wide range of imaging conditions, which is very challenging for the current method relying on visual examination. We have recently published a closed-form algorithm, $s^2 stigmator$, with a single-pass tuning strategy [83], that allows fast and sensitive detection of astigmatism using TEM live images and guides the users to reliably and efficiently adjust the two objective lens stigmators to correct astigmatism. We tested the performance of our $s^2 stigmator$ method by correcting astigmatism at various imaging conditions, including defocus and magnification. Fig.6.S2 shows the screenshots of trajectories obtained from Titan Krios when correcting astigmatism at different defocuses (Fig.6.S2A-C) and different magnifications (Fig.6.S2D-F). In these six screenshots, the trajectories are very similar and all consistently led to correction of astigmatism at a wide range of defocuses and magnifications. The angle of the straight trace segment (63°, the angle marked in green) corresponds to the 2nd stigmator used in this single-pass strategy, i.e. stigmator MX here. This angle is determined by the angular position of the stigmators, e.g. octupole objective lens stigmator [86,87].

Fig.6.S3 shows the trajectories acquired at a variety of defocuses (Fig.6.S3A-C) and magnifications (Fig.6.S3D-F) on CM200. Similarly, the change of defocus on CM200 in the range of cryo-EM research does not have a significant influence on the shape of the trajectories (Fig.6.S3A-C). However, the switch of magnification does have an effect on the orientation of the trajectories (Fig.6.S3D-F). It is noted that the trajectories turn clockwise when magnification increases (Fig.6.S3D-F), which is consistent with the rotation of real images at the same set of magnifications (Fig.6.S3G-I). Moreover, the angle changes of the straight trace segments between two adjacent magnifications are 14° (65° in Fig.6.S3D v.s. 51° in Fig.6.S3E) and 26° (51° in Fig.6.S3E v.s. 25° in Fig.6.S3F), identical to the change of angles (marked by red dash lines in Fig.6.S3G-I) among the corresponding real images (14° between Fig.6.S3G and H, 26° between Fig.6.S3H and I). Thus, we attribute the rotation of trajectories at different magnifications to the imperfect implementation of the rotation-free imaging function on CM200. In contrast, the rotation-free imaging function on Titan Krios is excellent as shown by the absence of rotations of the trajectories in Fig.6.S2D-F for different magnifications.

### 6.3.3 Defocus-dependent astigmatism

After confirming that the astigmatism of objective lens could be accurately minimized using $s^2stigmator$ at a wide range of defocuses and magnifications (Fig.6.S2 and Fig.6.S3), we systematically investigated the dependence of astigmatism on defocus using live images of carbon film at room temperature. We first corrected the astigmatism at a specific defocus using $s^2stigmator$, then measured the astigmatism with all other instrument parameters remaining constant while only the defocus was gradually altered with a fixed step size (e.g. 100 nm). Fig.6.2 shows clear correlation between defocus and astigmatism for both Titan Krios (Fig.6.2A-D) and CM200 (Fig.6.2E-H) microscopes at a nominal magnification of 22,500× and 115,000×, respectively. At each defocus, the point and error bar represent the mean and RMSD

**A** Mag 22.5K defocus 500nm

**B** Mag 22.5K defocus 1500nm

**C** Mag 22.5K defocus 2500nm

**D** Mag 22.5K defocus 2500nm

**E** Mag 37K defocus 2500nm

**F** Mag 59K defocus 2500nm
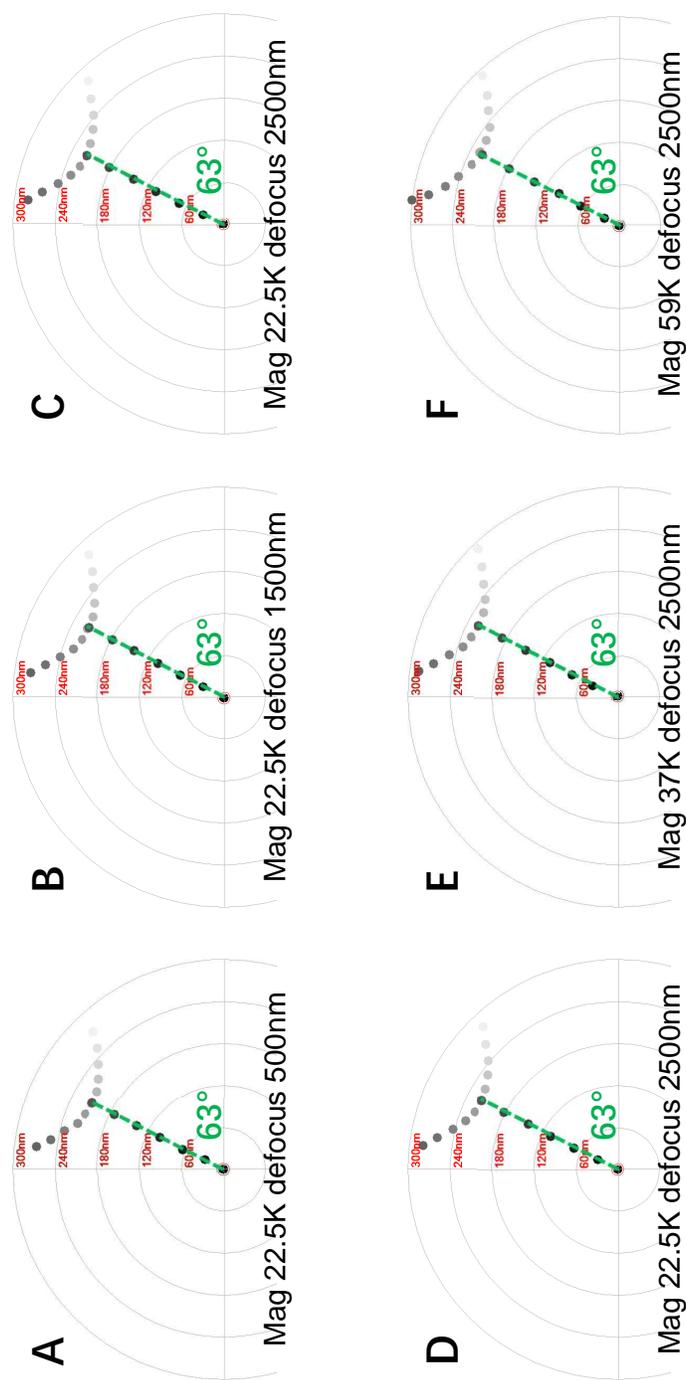
Fig. 6.S2. Representative trajectories of astigmatism correction at varying defocuses and magnifications on Titan Krios microscope. (A-C) The screenshots of the trajectories acquired at a nominal magnification of 22,500× and defocus 500 nm (A), 1500 nm (B) and 2500 nm (C), respectively. (D-F) The screenshots of the trajectories acquired at a nominal magnification of 22,500× (D), 37,000× (E) and 59,000× (F), respectively, and defocus 2500 nm. The green dash line in each plot represents the angle of the straight trace segment in the trajectory, and the exact angle is marked in green next to the green dash line.
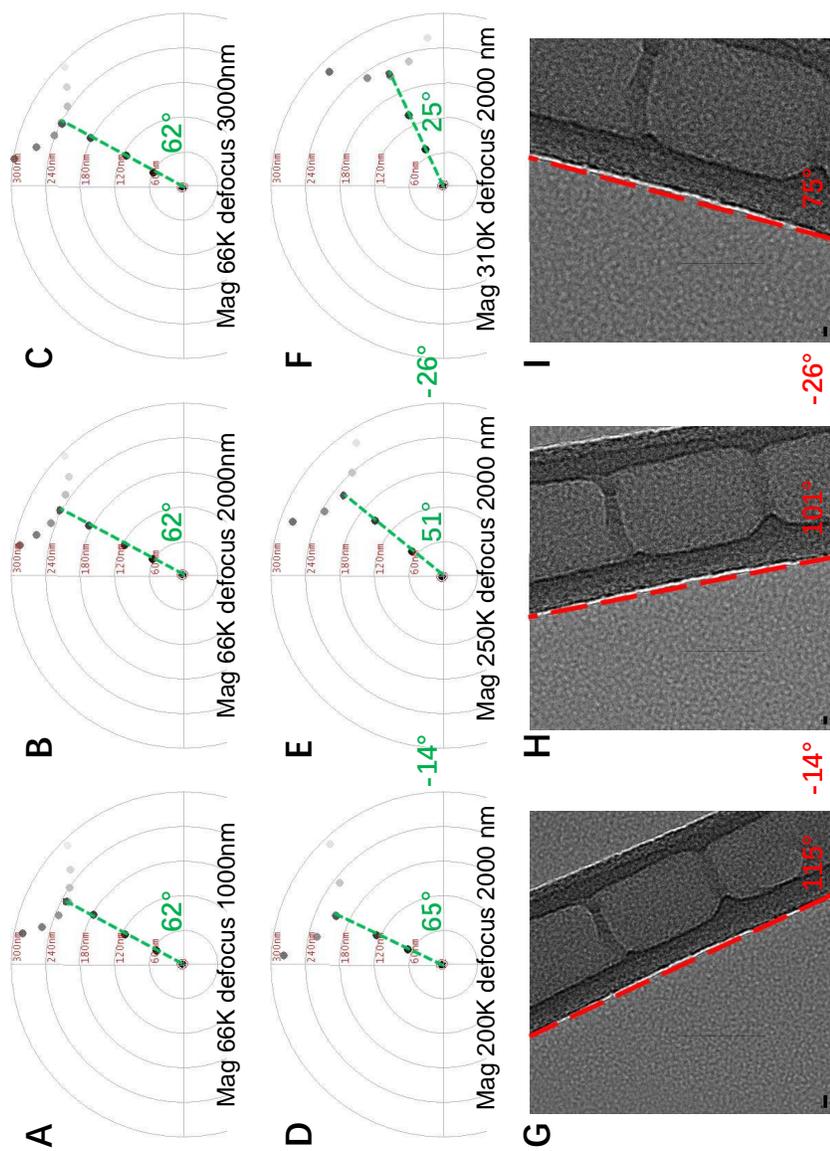
Fig. 6.S3. Representative trajectories of astigmatism correction at varying defocuses and magnifications on CM200 microscope. (A–C) The screenshots of the trajectories acquired at a nominal magnification of 66,000× and defocus 1000 nm (A), 2000 nm (B) and 3000 nm (C), respectively. The green dash line in each plot represents the angle of the straight trace segment in the trajectory, and the exact angle is marked in green next to the green dash line. The angles of the straight trace segments are the same (62° in A–C). (D–F) The screenshots of the trajectories acquired at a nominal magnification of 200,000× (D), 250,000× (E) and 310,000× (F), respectively, and defocus 2000 nm. The angles of the straight trace segments are 65°, 51° and 25°, respectively. And the difference between two adjacent trajectories are 14° and 26°. (G–I) The real space images collected at a nominal magnification of 200,000× (G), 250,000× (H) and 310,000× (I), respectively. The rotation of the red dash line represents the rotation of the image in real space with the change of magnifications. The angles of the red dash lines are 115°, 101° and 75°, respectively. And the angular difference between two adjacent images are 14° and 26°, in agreement with those of the trajectories shown in (D–F).

of astigmatism from ten images, respectively. When the astigmatism is minimized at small, medium, and large defocus (red, green and blue lines in Fig.6.2A and E), astigmatism linearly increases as the defocus is continuously increased/decreased from the starting defocus used for astigmatism correction. Apparently, the slopes of the lines from CM200 (Fig.6.2E) are much larger than those from Titan Krios (Fig.6.2A), implying the dependence of astigmatism on defocus for CM200 is much more severe than that for Titan Krios. Furthermore, polar plots were used to show the raw data distribution of the astigmatism used for the line graphs with the same colors. Fig.6.2B-D present the polar distribution of astigmatism with varying defocus when the astigmatism is corrected at small (Fig.6.2B), medium (Fig.6.2C), and large defocus (Fig.6.2D) on Titan Krios, corresponding to the red, green, and blue line in Fig.6.2A, respectively. It is evident that the astigmatism angle stably points to a certain direction as the astigmatism amplitude gradually increases due to the monotonically ascending (Fig.6.2B) or descending defocus (Fig.6.2D). Fig.6.2C shows that the astigmatism angle changes about 90° when defocus changes bi-directionally after astigmatism minimization. The 90 angle change corresponds to the swapping of the major and minor axes of astigmatism. Similar distributions can also be seen from the raw data acquired on CM200 (Fig.6.2F-H) but with larger increase of astigmatism than that on Titan Krios (Fig.6.2B-D). In addition, we observed that the distributions of astigmatism due to the bi-directional change of defocus have similar profiles whether the initial astigmatism is minimized (red points in Fig.6.3) or not (blue points in Fig.6.3). Therefore, all the analyses above demonstrate the general existence of defocus-dependent astigmatism in the objective lens of TEM.

For a comprehensive understanding of defocus-dependent astigmatism, we repeated our measurement on different days and compared the variation of astigmatism. As shown in Fig.6.4A and E, the slopes of lines are not identical even for the measurements made using the same conditions, demonstrating the amount of dependence varies from day to day on both Titan Krios (Fig.6.4A) and CM200 (Fig.6.4E) microscopes. What's more, much more pronounced differences in the astigmatism

angles were shown in the data collected on different days (Fig.6.4). The differences indicate that the distribution of defocus-dependent astigmatism cannot be exactly reproduced even though the defocus is adjusted in the same way. This irreproducibility can be observed from the data collected on both Titan Krios (Fig.6.4A-D) and CM200 (Fig.6.4E-H) whether the astigmatism was initially corrected at small defocus (Fig.6.4A-D) or large defocus (Fig.6.4E-H). Consequently, the comparison of the repetitive measurements confirms the variability of defocus-dependent astigmatism in the objective lens of TEM.

### 6.3.4   Magnification-dependent astigmatism

We also used $s^2stigmator$ to investigate the dependence of astigmatism on magnification and to test the implicit assumption of invariant astigmatism at different magnifications for the practice of using a higher magnification for correction of astigmatism than that for data acquisition. We first corrected astigmatism at a nominal magnification of 96,000× on Titan Krios and 250,000× on CM200 microscope, respectively, then measured the astigmatism as the magnification was stepwise reduced while keeping all other instrument parameters unchanged. Fig.6.5 shows the change of astigmatism/defocus with magnification for Titan Krios and the variability of this change across multiple measurements. At each magnification, we collected twenty images, plotted the distribution of astigmatism in polar coordinate (Fig.6.5A-C), and then calculated their mean and RMSD of astigmatism (blue line)/defocus (red line) represented as points and error bars (may be too small to be visible) in Fig.6.5D-F. When repeating the measurements on Titan Krios, the changes of defocus (red lines in Fig.6.5D-F) followed a similar pattern, but the profiles of both astigmatism amplitude (blues lines in Fig.6.5D-F) and astigmatism angle (Fig.6.5A-C) were irreproducible. Similar results were also found for CM200 (Fig.6.6) in which the defocus tends to increase with lower magnifications (red lines in Fig.6.6D-F), rather than decrease as shown for Titan Krios (red lines in Fig.6.5D-F). All these data demonstrate
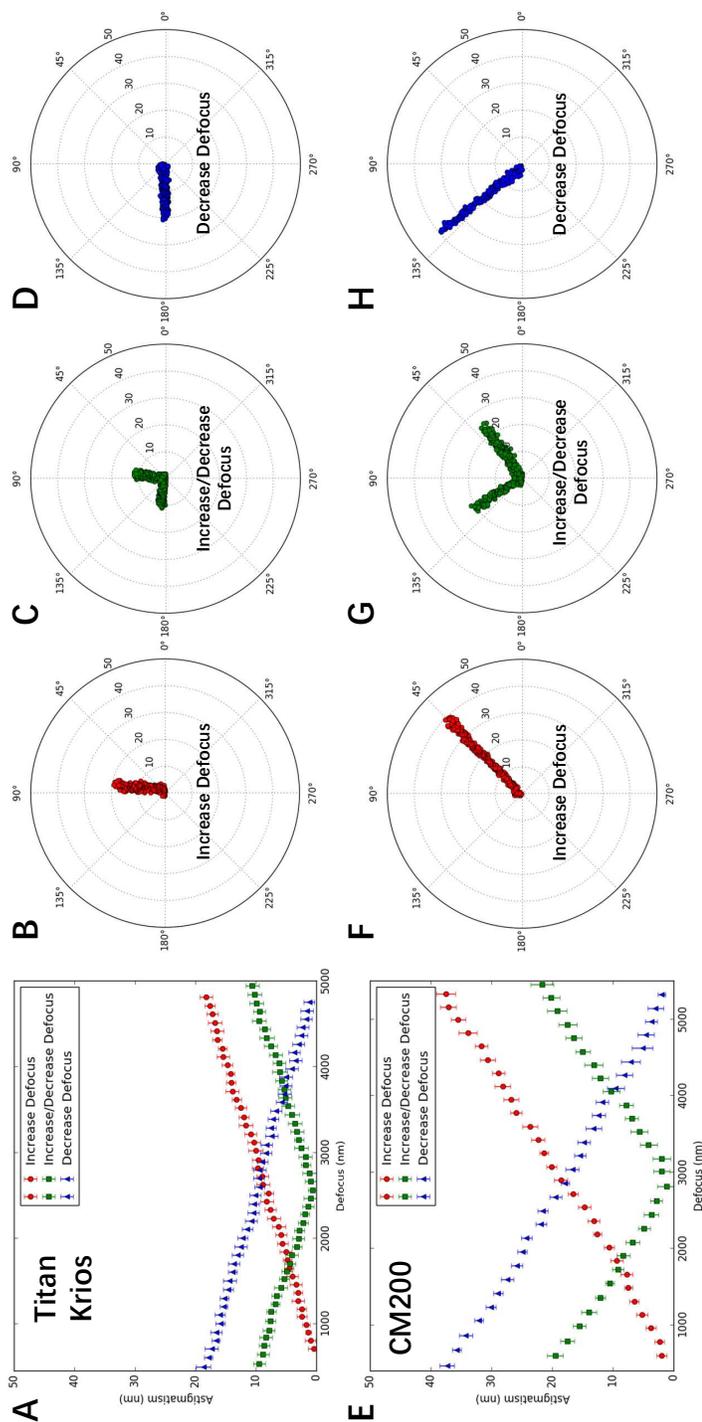
Fig. 6.2. Defocus-dependent astigmatism. (A) The increment of astigmatism with the change of defocus on Titan Krios microscope when the astigmatism is corrected at small (red), medium (green) and large (blue) defocus, respectively. (B-D) The polar distribution of all data obtained from Titan Krios microscope when the astigmatism is corrected at small (B), medium (C) and large (D) defocus, corresponding to the line with the same color in (A), respectively. (E) The increment of astigmatism with the change of defocus on CM200 microscope obtained from the same experiment described in (A). The correlation between (E) and (F-H) is the same as that between (A) and (B-D). In the line plots (A, E), the point and error bar at each defocus represent the mean and RMSD of astigmatism from ten images, respectively.
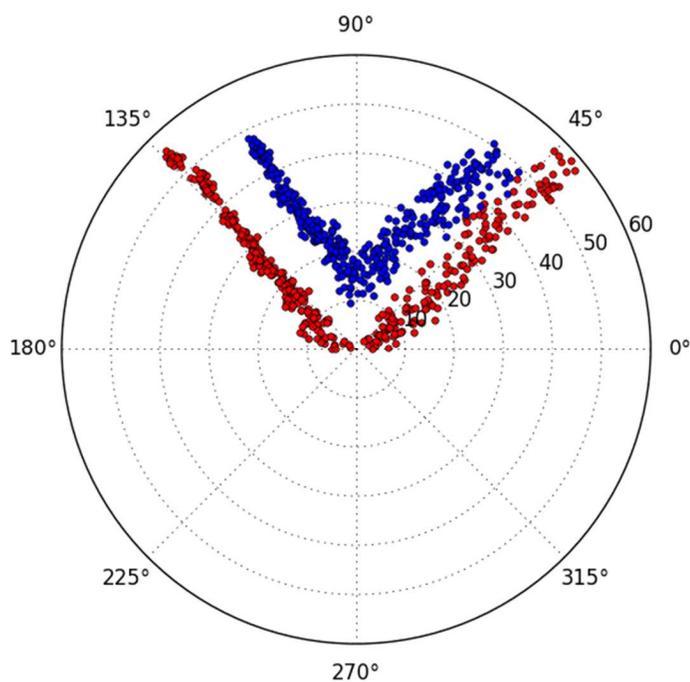
Fig. 6.3. Comparison of defocus-dependent astigmatism with the initial astigmatism minimized or with residual initial astigmatism. The increases of astigmatism with the bi-directional change of defocus on CM200 microscope are shown when the starting astigmatism is minimized (red points) or not minimized (blue points). The initial defocus is around 8000nm and the range of defocus variation is around 7000nm in both directions. The image collection strategy is the same as the data shown in Fig.6.2G.
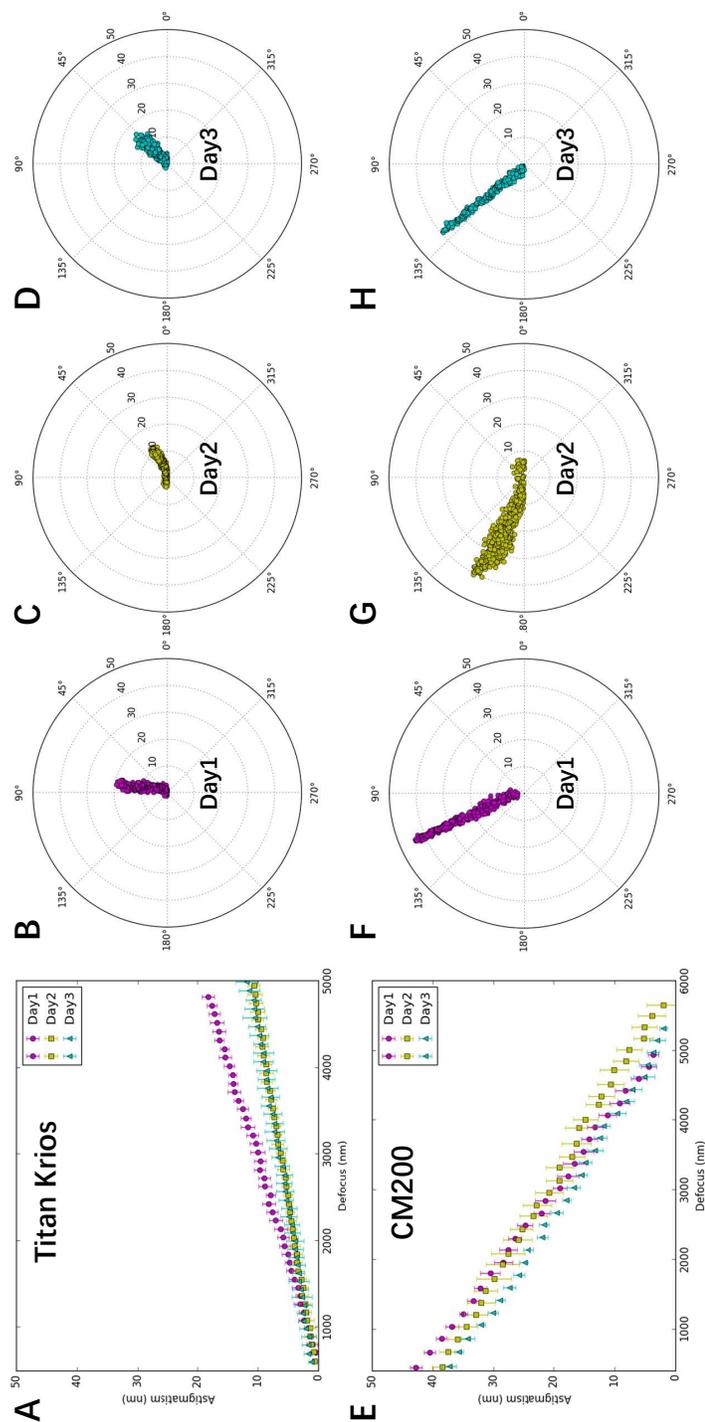
Fig. 6.4. Variability of defocus-dependent astigmatism. (A) The profile of astigmatism increment as defocus increases on Titan Krios microscope when the astigmatism is minimized at small defocus on three different days. (B-D) The polar distribution of all data from the repeated experiments on Titan Krios microscope described in (A). Each polar distribution corresponds to the line with the same color as in (A). (E) The profile of astigmatism increment as defocus decreases on CM200 microscope when the astigmatism is minimized at large defocus on three different days. The correlation between (E) and (F-H) is the same as that between (A) and (B-D). In the line plots (A, E), the point and error bar at each defocus represents the mean and RMSD of astigmatism from ten images, respectively.

the existence of magnification-dependent astigmatism and its stochastic fluctuations for both high-end and low-end TEM.

## 6.4 Discussion

In contrast to the dramatic progress in the automated cryo-EM data acquisition and image processing methods, little has changed for the microscope alignment tasks before data acquisition. In this paper, inspired by the observations of defocus-astigmatism correlations in experimental cryo-EM datasets (Fig.6.1), we have discovered the defocus/magnification-dependence of objective lens astigmatism and their stochastic variability using our recently published tool $s^2 stigmator$. These findings have essentially invalidated a basic assumption of a current cryo-EM imaging strategy that assumes constant astigmatism for the significantly different defocuses/magnifications used in microscope alignment stage and final data acquisition stage.

### 6.4.1 Vector summation model of the net astigmatism and the single pass tuning strategy

As shown here (Fig.6.S1-Fig.6.S3) and in our previous work [83], $s^2 stigmator$ can help achieve accurate correction of objective lens astigmatism at any imaging condition using a single pass tuning strategy. Understanding the principle of this single-pass tuning strategy will help further unveil the underlying theory of astigmatism variations with imaging parameters. Fig.6.7 illustrates the vector diagrams (Fig.6.7B-D) of three key points in a typical trajectory (Fig.6.7A), including the initial point (⓪ in Fig.6.7A) the optimal point (① in Fig.6.7A) in the arc-like segment when tuning the first stigmator (e.g. MY), and the final point at the center (② in Fig.6.7A) after tuning the other stigmator (e.g. MX). The corresponding stigmator MX/MY values are labelled in parentheses next to the circled numbers. In Fig.6.7B, $\vec{V}_{obj}^0$, $\vec{V}_{MX}^0$, $\vec{V}_{MY}^0$ represents the initial state of the astigmatism of the objective lens, and the correction
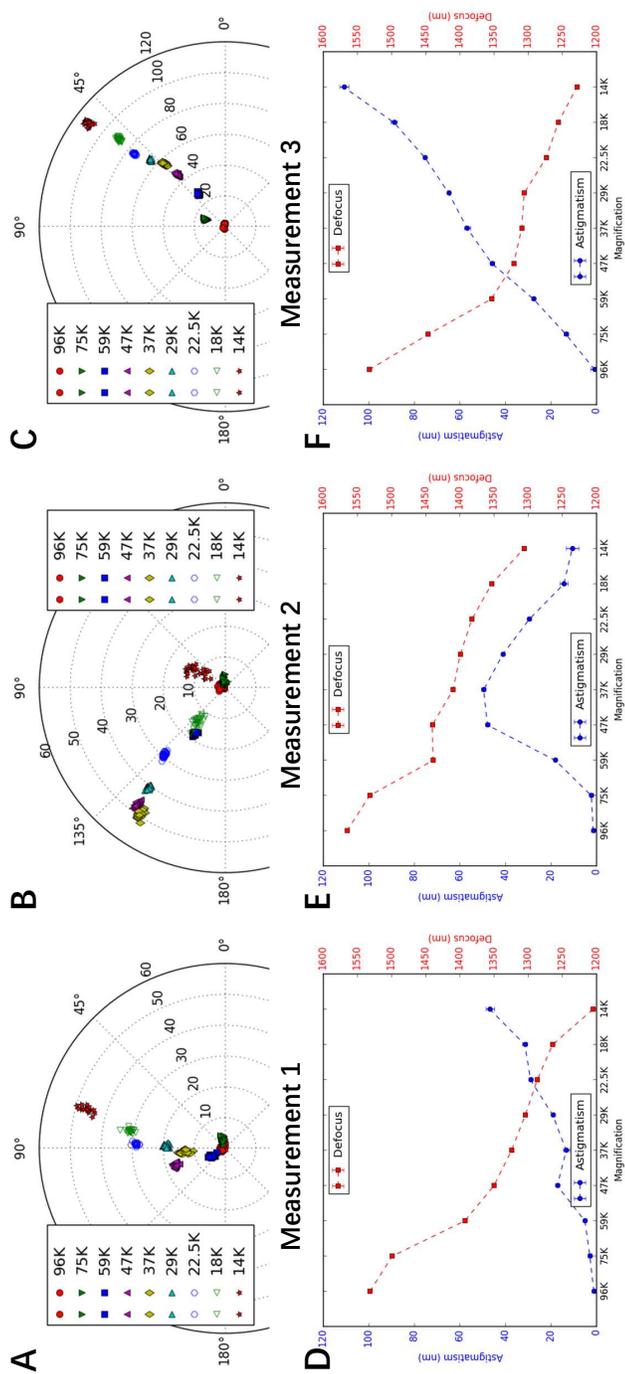
Fig. 6.5. Magnification-dependent astigmatism detected on Titan Krios microscope. (A-C) Plots of astigmatism polar distribution with stepwise reduction of magnifications in three repeated measurements in a single day when the astigmatism is corrected at a nominal magnification of 96,000×. (D-F) The profiles of the variations of astigmatism (blue line) or defocus (red line) with the change of magnifications, corresponding to the measurements in (A-C). At each magnification, twenty images were collected and their astigmatisms and defocuses were calculated. The point and error bar represent the mean and RMSD of astigmatism in the blue line or defocus in the red line, respectively. The error bar may be too small to be visible in the line plot.
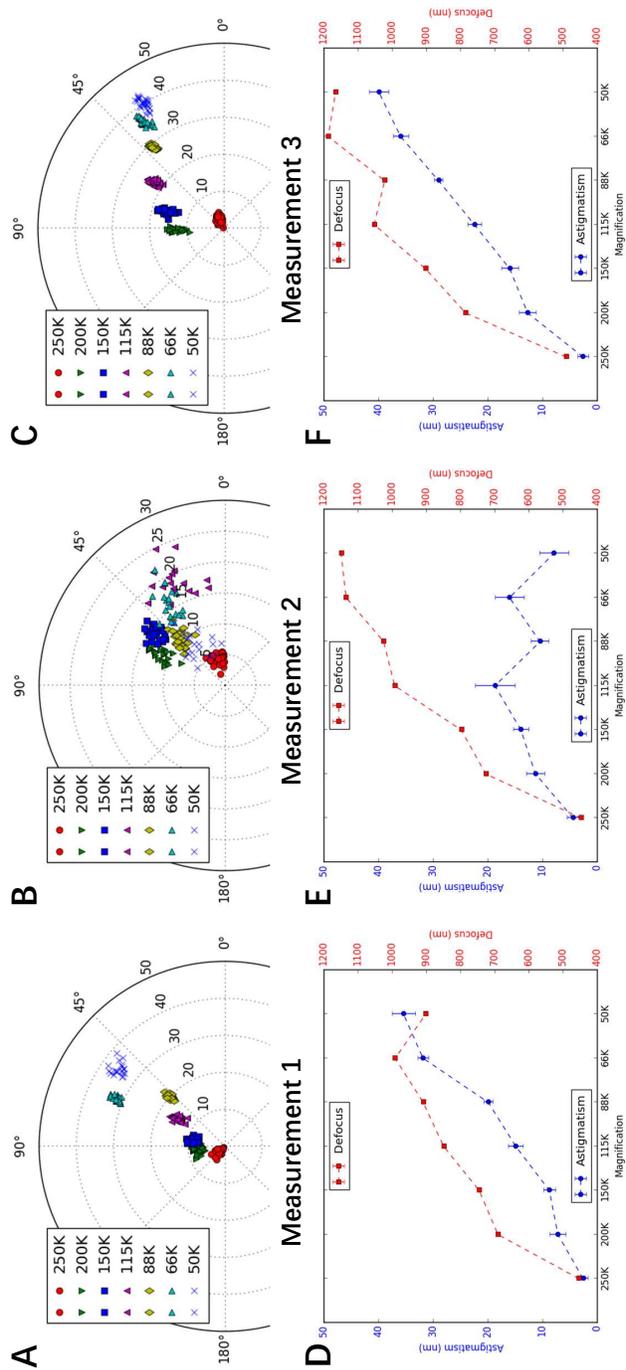
Fig. 6.6. Magnification-dependent astigmatism detected on CM200 microscope. The measurements are the same as described in Fig.6.5. The only difference being that the astigmatism is corrected at a nominal magnification of 250,000× on the CM200.

field of the stigmator MX and MY, respectively. $\vec{V}_{sum}^0$ is the summation of these three vectors and corresponds to the point ⓪ in Fig.6.7A. Here $\vec{V}_{obj}^0$ can be considered as a constant vector in a short period, e.g. during astigmatism correction, while the correction field $\vec{V}_{MX}$ and $\vec{V}_{MY}$ will be varied to cancel $\vec{V}_{obj}^0$ in order to minimize astigmatism. However, the orientations of $\vec{V}_{MX}^0$ and $\vec{V}_{MY}^0$ are fixed and the angle between them is also fixed at $45°$, which are determined by the design of the octupole objective lens stigmator [86, 87] assembly containing two interdigitated quadruple stigmators with $45°$ offset. When the stigmators are adjusted, the $\vec{V}_{MX}$ and $\vec{V}_{MY}$ vectors will change lengths without turning. The astigmatism correction task is to find the optimal lengths for both stigmator vectors to make the sum of the two correction vectors exactly inverse of the objective astigmatism vector $\vec{V}_{obj}^0$ (i.e. same length but opposite direction). In Fig.6.7C, only the stigmator MY is adjusted (red line in Fig.6.7C) until it reaches the optimal length ($\vec{V}_{MY}^1$) at which its vector sum with $\vec{V}_{obj}^0$ is along the direction of $\vec{V}_{MX}^1$. In this process, the resulted points (i.e. the net astigmatism, or the sum of the three vectors) of the trajectory shown in Fig.6.7A exhibit an arc-like segment. Here the stigmator MX does not change ($\vec{V}_{MX}^1 = \vec{V}_{MX}^0$) and $\vec{V}_{sum}^1$ is along the direction of $\vec{V}_{MX}^1$, corresponding to the point ① in Fig.6.7A. In Fig.6.7D, only the stigmator MX is adjusted to cause its correction field vector $\vec{V}_{MX}^0$ to change length ($\vec{V}_{MX}^2$, green line in Fig.6.7D) until the overall summation of vectors is 0 ($\vec{V}_{sum}^2 = 0$, ② in Fig.6.7A). In this part of trajectory, the resulted points directly move to the origin, forming a straight trace segment. As demonstrated in Fig.6.7, the orientation of the straight trace segment is determined by the manufacturer's setting of the stigmator MX's direction. This finding can explain why the orientations of trajectories on the Titan Krios microscope shown in Fig.6.S2 are the same, independent of defocus and magnification, when the two stigmators are adjusted in the order of MY first then MX. On the contrary, if the order of stigmators is switched during adjustment (MX first, then MY), the trajectory will rotate $45°$ and the straight trace segment will represent the direction of stigmator MY (Fig.6.S1). This vector summation model is also validated on CM200 when the imperfect rotation-free function is considered for

the rotation of trajectories between different magnifications. The analysis described above explains the relationship of the objective lens astigmatism and the correction fields generated by the stigmators, and how the stigmators can be controlled to optimally compensate the objective lens astigmatism. The 3-vector summation model described here further refines our previous model [83].

### 6.4.2 Defocus-dependent astigmatism

During single particle cryo-EM image acquisition, the defocuses for different images are intentionally varied to average out the effect of zero-nodes of CTF and obtain signals at all frequencies [73, 81, 82]. Nevertheless, the discovery of defocus-dependent astigmatism in published experiment datasets (Fig.6.1) and tests here (Fig.6.2-Fig.6.3) implies a significant problem of re-emerging astigmatism in this imaging strategy. Fig.6.8 uses vector diagrams to explain the case of corrected astigmatism (Fig.6.8A), and the re-emerged astigmatism after increasing (Fig.6.8B) and decreasing (Fig.6.8C) defocus. As can been seen from Fig.6.8A, the total summation vector $\vec{V}_{sum} = 0$ when $\vec{V}_{obj}^0$ is canceled by $\vec{V}_{stigmator}$. Here $\vec{V}_{obj}^0$ and $\vec{V}_{stigmator}$ represent the initial astigmatism of objective lens and the combined correction field of the two stigmators, respectively. As the astigmatic fields of the objective lens and the stigmators are generated by the current, the field strength will be proportional to the current. We can thus express the total astigmatism ($\vec{V}_{sum}$) as

$$\begin{aligned} \vec{V}_{sum} &= \vec{V}_{obj}^0 + \vec{V}_{MX} + \vec{V}_{MY} + \Delta\vec{V}_{obj}^{df} \\ &= k_{obj} \cdot I_{obj}^0 \cdot \vec{e}_{obj} + k_{MX} \cdot I_{MX} \cdot \vec{e}_{MX} + k_{MY} \cdot I_{MY} \cdot \vec{e}_{MY} + k_{obj} \cdot (I_{obj} - I_{obj}^0) \cdot \vec{e}_{obj} \end{aligned}$$

$$(6.1)$$

where $\vec{V}_{obj}^0$, $\vec{V}_{MX}$, $\vec{V}_{MY}$ are the initial astigmatism of the objective lens, and the correction field of the stigmator MX and MY, respectively. $\Delta\vec{V}_{obj}^{df}$ is the change of astigmatism of the objective lens induced by the variation of defocus. $\vec{e}_{obj}$, $\vec{e}_{MX}$, $\vec{e}_{MY}$ are unit vectors representing the direction of the objective lens astigmatism, the
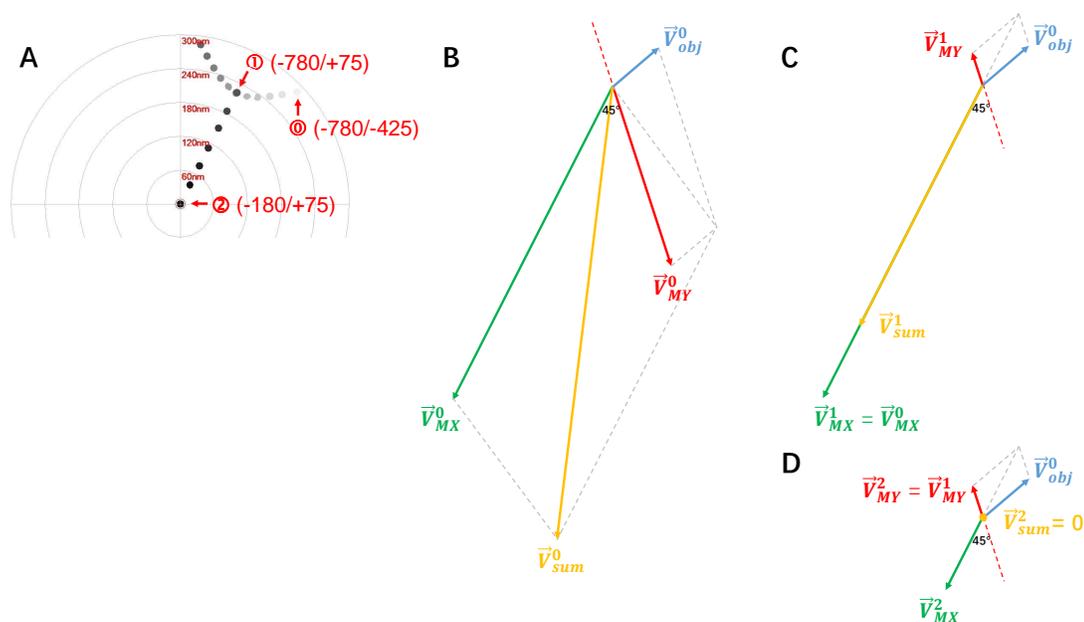
Fig. 6.7. Vector diagrams to illustrate the principle of single-pass tuning strategy for astigmatism correction. (A) A screenshot of the trajectory on the Titan Krios microscope in which three key points are marked by red circled numbers ⓪, ①, ②, corresponding to the vector diagrams in (B-D), respectively. The corresponding stigmator MX/MY values are labelled in parentheses next to the circled numbers. (B) Initial point ⓪. The astigmatism of the objective lens, the correction fields of stigmator MX and MY are represented by $\vec{V}_{obj}^0$, $\vec{V}_{MX}^0$, $\vec{V}_{MY}^0$, and their summation is represented by $\vec{V}_{sum}^0$. It is noted that $\vec{V}_{obj}^0$ is assumed as a constant vector here, the directions of $\vec{V}_{MX}^0$ and $\vec{V}_{MY}^0$ are fixed and the angle between them is 45°. (C) The optimal point ① in the arc-like segment. The stigmator MY first reaches its optimal value ($\vec{V}_{MY}^1$) after adjusting along its own direction (red line) until $\vec{V}_{sum}^1$ is located in the direction of the stigmator MX ($\vec{V}_{MX}^1$). Here $\vec{V}_{sum}^1$ corresponds to the point ① in (A). (D) The final point ② of astigmatism correction. The stigmator MX is now adjusted along the direction of green line until $\vec{V}_{sum}^2$ is zero, corresponding to the point ② in (A). Consequently, the orientation of the straight trace segment in the trajectory (A) is determined by the orientation of stigmator MX.

correction field direction of stigmator MX and MY, respectively. $k_{obj}$, $k_{MX}$, $k_{MY}$ are scaling factors representing how strong the dependence is between the field and the corresponding current of the objective lens, stigmator MX, and stigmator MY, respectively. The first three terms in the second line of Eq.6.1 represent the initial astigmatism correction corresponding to the starting objective lens current ($I_{obj}^0$), $I_{MX}$ and $I_{MY}$, and the sum of these three terms should be zero if the astigmatism is fully corrected. The last term in the second line of Eq.6.1 represents the defocus-dependent astigmatism when the defocus is increased or decreased by changing the objective lens current ($I_{obj}$) from its starting point ($I_{obj}^0$). In Fig.6.8A, the astigmatism was corrected completely and the shape of Thon rings is perfectly circular (Fig.6.8D). However, to increase defocus the objective lens current needs to be reduced to weaken the lens bending power, leading to a smaller objective lens astigmatism (the sum of blue arrows, $\vec{V}_{obj}^0 + \Delta \vec{V}_{obj}^{df\uparrow}$ in Fig.6.8B). The correction fields by the two stigmator ($\vec{V}_{stigmator}$), which have not been changed from previous values optimized for a larger amount of objective lens astigmatism, now over-corrects the new, reduced objective lens astigmatism. A non-zero $\vec{V}_{sum}$ (Fig.6.8B) now appears and the Thon rings (Fig.6.8E) become elongated. The analysis in Fig.6.8B agrees with the observations from the Titan Krios (red line in Fig.6.2A and B) and CM200 (red line in Fig.6.2E and F) microscopes in the case of increasing defocus. Similarly, another non-zero $\vec{V}_{sum}$ (Fig.6.8C) appears in the opposite direction when defocus decreases and a larger objective lens astigmatism is generated (the sum of blue arrows, $\vec{V}_{obj}^0 + \Delta \vec{V}_{obj}^{df\downarrow}$ in Fig.6.8C), resulting in the Thon rings (Fig.6.8F) becoming elongated along the perpendicular direction of Fig.6.8E, since the switch of $\vec{V}_{sum}$ direction is equivalent to change the ellipticity by 90°. The analysis in Fig.6.8C also agrees with the observed defocus-dependence of astigmatism as defocus decreases (blue line in Fig.6.2A and D for Titan Krios; blue line in Fig.6.2E and H for CM200). Combining the vector diagrams in both Fig.6.8B and C, the bi-directional increment of astigmatism can also be clearly understood, as well as the 90° angle between two branches in the polar plots (green line in Fig.6.2A and C for Titan Krios; green line in Fig.6.2E and G, and

Fig.6.3 for CM200) when the astigmatism is minimized at the middle point of the defocus range. This finding of defocus-dependent astigmatism is also consistent with theoretic predictions based on Zernike polynomial expression of lens aberrations [88].

### 6.4.3 Quantification of objective lens asymmetry

In the line plots from the Titan Krios (Fig.6.2A) and CM200 (Fig.6.2E), it is evident that the slopes of the linear trends are different for the two instruments. The slope measures how strong the dependence of astigmatism on defocus is and should be proportional to the scaling factor $k_{obj}$ in Eq.6.1. For a perfectly round lens $k_{obj}$ is equal to 0 and as asymmetry of the lens increases, the $k_{obj}$ becomes larger. Therefore, we can use $k_{obj}$ as a parameter to quantify the quality of a TEM magnetic lens in terms of its cylindrical symmetry. A lens with smaller $k_{obj}$ will be desirable. Using this criterion, the objective lens of the Titan Krios microscope is more cylindrical than that of CM200 microscope. This is consistent with the common understanding of current generation Titan Krios as a higher quality TEM than the CM200 microscope which was produced more than two decades ago. We propose that the defocus-dependent plots of astigmatism as shown in Fig.6.2A and E are convenient measurements of the asymmetry level of the objective lens of a TEM instrument. Such quantitative measurements can be useful in several applications. For example, it can be used as one of the acceptance tests after the installation of a new TEM instrument. It can also be used to monitor the performance of the objective lens and to detect potential deterioration, for example, caused by a large contamination in the objective lens area.

### 6.4.4 Stochastic variations of defocus-dependent astigmatism

The defocus-dependent astigmatism, including both the slope and the direction, was found to vary in our data (Fig.6.4) when the same measurement was repeated after two-weeks. While the astigmatism of objective lens ($\vec{V}_{obj}$) is considered stable within a short period time (a few hours to one or two days), it is also well-known
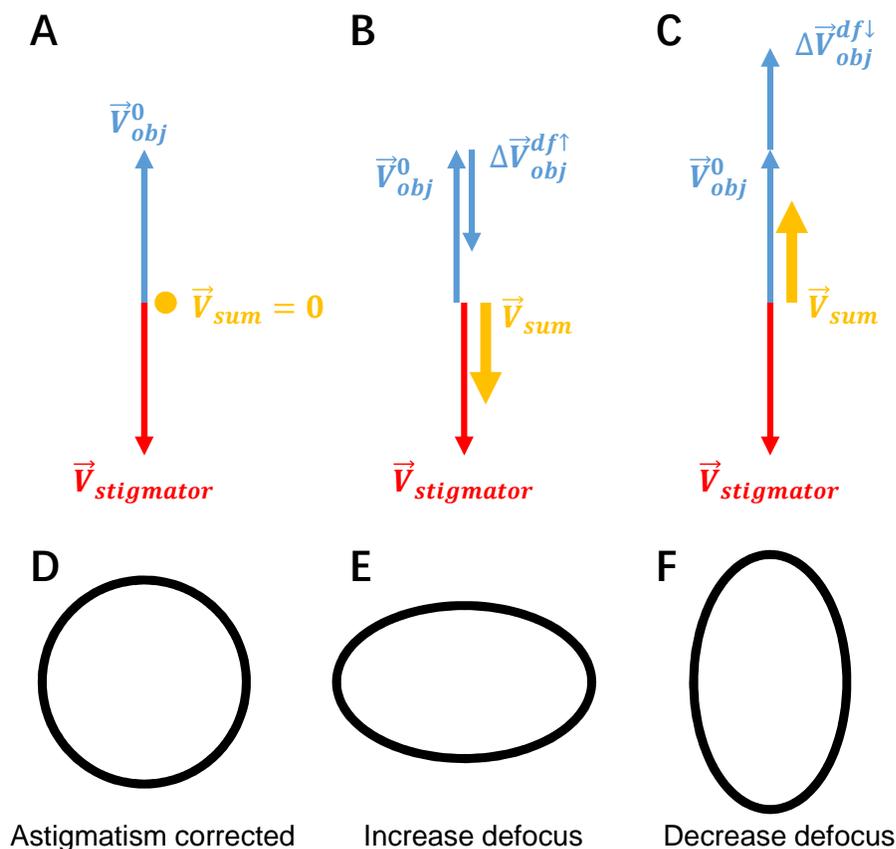
Fig. 6.8. Vector diagram to interpret the defocus-dependent astigmatism shown in Fig.6.2. (A-C) Vector diagrams illuminate the state of astigmatism fully corrected at a defocus (A), increased defocus after correction (B), decreased defocus after correction (C). Here the red arrows represent the combined correction field of the two stigmators $\vec{V}_{stigmator}$ and the blue arrows represent the astigmatism of the objective lens which is proportional to the strength of objective lens current. When defocus increases or decreases, the associated objective lens astigmatism (the sum of blue arrows) becomes smaller ($\vec{V}^0_{obj}+\Delta\vec{V}^{df\uparrow}_{obj}$ in B) or larger ($\vec{V}^0_{obj}+\Delta\vec{V}^{df\downarrow}_{obj}$ in C) than the original $\vec{V}^0_{obj}$ and can no longer be fully compensated by $\vec{V}_{stigmator}$. (D-F) The representative sketches of the shape of Thon rings in the astigmatism states corresponding to the vector diagrams depicted in (A-C), respectively.

that astigmatism tends to vary. As a result, it is a common practice to check and re-correct astigmatism as one of the daily-instrument alignment tasks. Our measurements (Fig.6.4) have thus quantitatively verified the variability and validated the need for daily correction of astigmatism. Such variability can also be explained using the vector summation model (Fig.6.9). $\vec{V}_{obj}$ can vary due to the change of either the unit vector $\vec{e}_{obj}$ direction or the amplitude of the scaling factor $k_{obj}$ (Eq.6.1) at different times. In the vector diagram (Fig.6.9), the varying $\vec{V}_{obj}$ (the objective lens astigmatism, blue arrows) is compensated by corresponding $\vec{V}_{stigmator}$ (stigmator values, red arrows) that needs to be updated from day to day. There is a wide variety of reasons for the stochastic changes of $\vec{e}_{obj}$ and $k_{obj}$, such as objective lens asymmetry due to imperfect manufacturing processes, electronic instability of the voltage and power supplies, column contaminations, temperature fluctuations of the objective lens chilling water, *etc.* [89].

### 6.4.5 Magnification-dependent astigmatism

In addition to the defocus-dependent astigmatism, magnification-dependent astigmatism was also observed (Fig.6.5 and Fig.6.6). This implies that noticeable astigmatism would re-emerge during data collection if a different magnification is used for the correction of astigmatism during instrument alignment. Compared with the variability of astigmatism due to defocus, the astigmatism dependence on magnification is even more variable. When the same test was repeated three times on Titan Krios on the same day, the distribution of astigmatism at different magnifications is considerably different in both amplitude and angle (Fig.6.5A-C). In the line plot for each measurement, the profile of astigmatism variation appears random in different tests (blue lines in Fig.6.5D-F) while the profile of defocus variation is much more reproducible (red lines in Fig.6.5D-F). Similar observations were obtained for both Titan Krios (Fig.6.5) and CM200 (Fig.6.6), which indicates that the defocus change is stable but the astigmatism change is unpredictable.
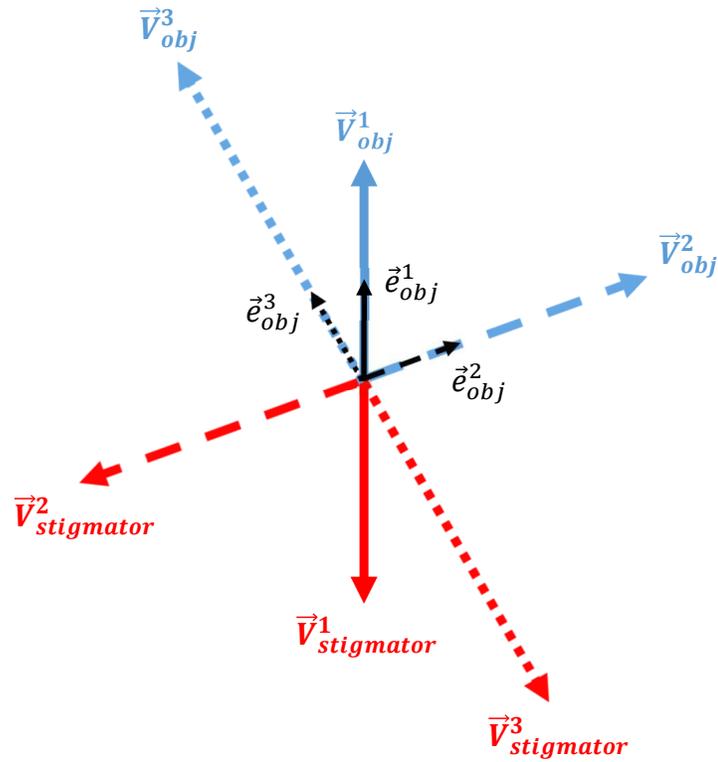
Fig. 6.9. Vector diagram to interpret the variability of defocus-dependent astigmatism shown in Fig.6.4. Here the vector representing the astigmatism of the objective lens ($\vec{V}_{obj}$, blue arrows) varies from day to day. The randomness of $\vec{V}_{obj}$ is determined by the orientation of the unit vector in black, $\vec{e}_{obj}^{\,1}$, $\vec{e}_{obj}^{\,2}$ and $\vec{e}_{obj}^{\,3}$, coupled with different scaling factors, $k_{obj}$ (Eq.6.1). The combined effect of two stigmators ($\vec{V}_{stigmator}$), represented by red arrows, needs to be varied accordingly to cancel $\vec{V}_{obj}$ on different days.

Modern TEM instruments usually use multiple imaging lenses, including an objective lens, a diffraction lens, an intermediate lens, and two projector lenses, to provide a wide range of magnifications. The astigmatism measured in the TEM image is a combined result of the astigmatism of all these lenses. The two stigmators actually correct the combined astigmatism of all these imaging lenses. When the magnification is changed, the current of all or a subset of these lenses would change, which leads to the change of individual lens astigmatism (Eq.6.1) and the combined astigmatism. As the stigmators have been tuned to correct the combined astigmatism at a particular magnification, the change of magnification will thus lead to re-emerging of astigmatism in the image at a different magnification. Since the currents of these lenses need to be changed in a non-linear pattern to achieve rotation-free imaging at multiple total magnifications, the combined astigmatism thus also varies in a non-linear pattern (Fig.6.5 and Fig.6.6). The irreproducibility of the profile of magnification-dependent astigmatism are likely caused by some random factors, such as column contaminations. Since the change of any one of the five lenses will change the combined astigmatism, it is thus not surprising the irreproducibility of the profile of magnification-dependent astigmatism is significantly worse than the irreproducibility of the profile of defocus-dependent astigmatism that is only affected by a single lens, the objective lens. In contrast, the profile of magnification-dependent defocus is more reproducible than that of magnification-dependent astigmatism as the pattern of current change is the same and the focus length of the lenses is more resistant to the random factors affecting the lens astigmatism.

## 6.4.6 Recommendations for optimal TEM operations

The astigmatism of TEM images has been shown here to vary with changes in imaging conditions (e.g. defocus, magnification), indicating that correction of astigmatism at high magnification and near-focus conditions by the current approach will not be optimal after switching to different conditions for data acquisition. What's

worse, the dependence of astigmatism on the imaging conditions varies from time to time, so that astigmatism cannot be reliably compensated by pre-calibration of the instrument. Although post-imaging computational CTF correction has now become a common practice, the community-wide CTF Challenge [76] has found that it is still more challenging to reliably and accurately fit the astigmatism than the average defocus. To aim at best quality results for each step and to avoid accumulation of issues in early steps that must be rescued by later steps, it is thus desirable to minimize the astigmatism during data collection. Based on our systematic measurements and analyses in this work, we suggest that (1) the magnification used for astigmatism correction during instrument alignment should be the same as the one used for data collection; (2) the defocus used for correction of astigmatism during instrument alignment should be set at the median defocus of the defocus range intended for subsequent data collection; (3) the focus-mode in the search-focus-exposure iterations of low-dose imaging should use the same magnification that is used for the exposure mode. Additionally, there are other factors such as stage and sample grid position that also lead to the variation of astigmatism, especially when the conductivity of materials in the imaging area is poor and local charging is induced. To optimally correct the astigmatism for images taken at different defocuses for different sample areas, a fast, accurate and automated method needs to be developed to avoid the defocus-dependent astigmatism by adaptively correcting the astigmatism at all focuses.

# 7. SUMMARY

Imaging techniques are essential research tools in many science and engineering disciplines. Cryo-EM has become a powerful technique for structural studies of macromolecular complexes at near-atomic resolutions. The advancements of computational methods have significantly contributed to the exciting achievements of cryo-EM. This dissertation emphasizes new approaches to address image processing problems in cryo-EM, including tilt series alignment evaluation, simultaneous determination of sample thickness, tilt, and electron mean free path based on Beer-Lambert law, MBIR on tomographic data, minimization of objective lens astigmatism in instrument alignment, and defocus and magnification dependent astigmatism of TEM images. The final goal of these methodological developments is to improve cryo-EM data quality and the 3D reconstructions of macromolecular structures, which will visualize more detailed characterization and help reveal the structural basis of biological processes.

## REFERENCES

[1] S. Brandt and U. Ziese, "Automatic tem image alignment by trifocal geometry," *Journal of microscopy*, vol. 222, no. 1, pp. 1–14, 2006.

[2] R. Han, F. Zhang, X. Wan, J.-J. Fernández, F. Sun, and Z. Liu, "A marker-free automatic alignment method based on scale-invariant features," *Journal of structural biology*, vol. 186, no. 1, pp. 167–180, 2014.

[3] S. Jones and M. Härting, "A new correlation based alignment technique for use in electron tomography," *Ultramicroscopy*, vol. 135, pp. 56–63, 2013.

[4] D. N. Mastronarde, "Fiducial marker and hybrid alignment methods for single- and double-axis tomography," in *Electron Tomography*. Springer, 2007, pp. 163–185.

[5] S. S. Brandt, "Markerless alignment in electron tomography," in *Electron tomography*. Springer, 2007, pp. 187–215.

[6] H. Winkler and K. A. Taylor, "Accurate marker-free alignment with simultaneous geometry determination and reconstruction of tilt series in electron tomography," *Ultramicroscopy*, vol. 106, no. 3, pp. 240–254, 2006.

[7] G. Cardone, K. Grünewald, and A. C. Steven, "A resolution criterion for electron tomography based on cross-validation," *Journal of structural biology*, vol. 151, no. 2, pp. 117–129, 2005.

[8] H.-J. Cho, J.-K. Hyun, J.-G. Kim, H. S. Jeong, H. N. Park, D.-J. You, and H. S. Jung, "Measurement of ice thickness on vitreous ice embedded cryo-em grids: investigation of optimizing condition for visualizing macromolecules," *Journal of Analytical Science and Technology*, vol. 4, no. 1, p. 7, 2013.

[9] T. Malis, S. Cheng, and R. Egerton, "Eels log-ratio technique for specimen-thickness measurement in the tem," *Journal of electron microscopy technique*, vol. 8, no. 2, pp. 193–200, 1988.

[10] H.-R. Zhang, R. F. Egerton, and M. Malac, "Local thickness measurement through scattering contrast and electron energy-loss spectroscopy," *Micron*, vol. 43, no. 1, pp. 8–15, 2012.

[11] S. V. Venkatakrishnan, L. F. Drummy, M. De Graef, J. P. Simmons, and C. A. Bouman, "Model based iterative reconstruction for bright field electron tomography," in *Computational Imaging XI*, vol. 8657. International Society for Optics and Photonics, 2013, p. 86570A.

[12] G. Tang, L. Peng, P. R. Baldwin, D. S. Mann, W. Jiang, I. Rees, and S. J. Ludtke, "Eman2: an extensible image processing suite for electron microscopy," *Journal of structural biology*, vol. 157, no. 1, pp. 38–46, 2007.

[13] T. E. Oliphant, "Python for scientific computing," *Computing in Science & Engineering*, vol. 9, no. 3, 2007.

[14] S. v. d. Walt, S. C. Colbert, and G. Varoquaux, "The numpy array: a structure for efficient numerical computation," *Computing in Science & Engineering*, vol. 13, no. 2, pp. 22–30, 2011.

[15] J. D. Hunter, "Matplotlib: A 2d graphics environment," *Computing in science & engineering*, vol. 9, no. 3, pp. 90–95, 2007.

[16] M. Braunfeld, A. Koster, J. Sedat, and D. Agard, "Cryo automated electron tomography: Towards high-resolution reconstructions of plastic-embedded structures," *Journal of microscopy*, vol. 174, no. 2, pp. 75–84, 1994.

[17] J. R. Kremer, D. N. Mastronarde, and J. R. McIntosh, "Computer visualization of three-dimensional image data using imod," *Journal of structural biology*, vol. 116, no. 1, pp. 71–76, 1996.

[18] L. Joyeux and P. A. Penczek, "Efficiency of 2d alignment methods," *Ultramicroscopy*, vol. 92, no. 2, pp. 33–46, 2002.

[19] A. Bartesaghi, F. Lecumberry, G. Sapiro, and S. Subramaniam, "Protein secondary structure determination by constrained single-particle cryo-electron tomography," *Structure*, vol. 20, no. 12, pp. 2003–2013, 2012.

[20] L. M. Voortman, M. Vulović, M. Maletta, A. Voigt, E. M. Franken, A. Simonetti, P. J. Peters, L. J. van Vliet, and B. Rieger, "Quantifying resolution limiting factors in subtomogram averaged cryo-electron tomography using simulations," *Journal of structural biology*, vol. 187, no. 2, pp. 103–111, 2014.

[21] J. Z. Chen, C. Sachse, C. Xu, T. Mielke, C. M. Spahn, and N. Grigorieff, "A dose-rate effect in single-particle electron microscopy," *Journal of structural biology*, vol. 161, no. 1, pp. 92–100, 2008.

[22] G. McMullan, K. Vinothkumar, and R. Henderson, "Thon rings from amorphous ice and implications of beam-induced brownian motion in single particle electron cryo-microscopy," *Ultramicroscopy*, vol. 158, pp. 26–32, 2015.

[23] R. Henderson, "The potential and limitations of neutrons, electrons and x-rays for atomic resolution microscopy of unstained biological molecules," *Quarterly reviews of biophysics*, vol. 28, no. 2, pp. 171–193, 1995.

[24] J. Frank, *Three-dimensional electron microscopy of macromolecular assemblies: visualization of biological molecules in their native state.* Oxford University Press, 2006.

[25] M. van Heel, B. Gowen, R. Matadeen, E. V. Orlova, R. Finn, T. Pape, D. Cohen, H. Stark, R. Schmidt, M. Schatz *et al.*, "Single-particle electron cryo-microscopy: towards atomic resolution," *Quarterly reviews of biophysics*, vol. 33, no. 4, pp. 307–369, 2000.

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

[27] J. Dubochet, M. Adrian, J.-J. Chang, J.-C. Homo, J. Lepault, A. W. McDowall, and P. Schultz, "Cryo-electron microscopy of vitrified specimens," *Quarterly reviews of biophysics*, vol. 21, no. 2, pp. 129–228, 1988.

[28] P. K. Luther, "Sample shrinkage and radiation damage of plastic sections," in *Electron Tomography*. Springer, 2007, pp. 17–48.

[29] K. Boergens and W. Denk, "Controlling fib-sbem slice thickness by monitoring the transmitted ion beam," *Journal of microscopy*, vol. 252, no. 3, pp. 258–262, 2013.

[30] H. Jones, K. Mingard, and D. Cox, "Investigation of slice thickness and shape milled by a focused ion beam for three-dimensional reconstruction of microstructures," *Ultramicroscopy*, vol. 139, pp. 20–28, 2014.

[31] M. Schaffer, J. Wagner, B. Schaffer, M. Schmied, and H. Mulders, "Automated three-dimensional x-ray analysis using a dual-beam fib," *Ultramicroscopy*, vol. 107, no. 8, pp. 587–597, 2007.

[32] S. Korte, M. Ritter, C. Jiao, P. Midgley, and W. Clegg, "Three-dimensional electron backscattered diffraction analysis of deformation in mgo micropillars," *Acta Materialia*, vol. 59, no. 19, pp. 7241–7254, 2011.

[33] A. Cheng, D. Fellmann, J. Pulokas, C. S. Potter, and B. Carragher, "Does contamination buildup limit throughput for automated cryoem?" *Journal of structural biology*, vol. 154, no. 3, pp. 303–311, 2006.

[34] B. Feja and U. Aebi, "Determination of the inelastic mean free path of electrons in vitrified ice layers for on-line thickness measurements by zero-loss imaging," *Journal of microscopy*, vol. 193, no. 1, pp. 15–19, 1999.

[35] R. Grimm, D. Typke, M. Bärmann, and W. Baumeister, "Determination of the inelastic mean free path in ice by examination of tilted vesicles and automated most probable loss imaging," *Ultramicroscopy*, vol. 63, no. 3-4, pp. 169–179, 1996.

[36] M. Vulović, R. B. Ravelli, L. J. van Vliet, A. J. Koster, I. Lazić, U. Lücken, H. Rullgård, O. Öktem, and B. Rieger, "Image formation modeling in cryo-electron microscopy," *Journal of structural biology*, vol. 183, no. 1, pp. 19–32, 2013.

[37] J. P. Langmore and M. F. Smith, "Quantitative energy-filtered electron microscopy of biological molecules in ice," *Ultramicroscopy*, vol. 46, no. 1-4, pp. 349–373, 1992.

[38] J. Wall, J. Langmore, M. Isaacson, and A. Crewe, "Scanning transmission electron microscopy at high resolution," *Proceedings of the National Academy of Sciences*, vol. 71, no. 1, pp. 1–5, 1974.

[39] J. A. Mindell and N. Grigorieff, "Accurate determination of local defocus and specimen tilt in electron microscopy," *Journal of structural biology*, vol. 142, no. 3, pp. 334–347, 2003.

[40] L. Houben and M. B. Sadan, "Refinement procedure for the image alignment in high-resolution electron tomography," *Ultramicroscopy*, vol. 111, no. 9-10, pp. 1512–1520, 2011.

[41] M. Hayashida, M. Malac, M. Bergen, R. F. Egerton, and P. Li, "Accurate measurement of relative tilt and azimuth angles in electron tomography: A comparison of fiducial marker method with electron diffraction," *Review of Scientific Instruments*, vol. 85, no. 8, p. 083704, 2014.

[42] D. N. Mastronarde, "Automated electron microscope tomography using robust prediction of specimen movements," *Journal of structural biology*, vol. 152, no. 1, pp. 36–51, 2005.

[43] C. A. Diebolder, F. J. Beurskens, R. N. De Jong, R. I. Koning, K. Strumane, M. A. Lindorfer, M. Voorhorst, D. Ugurlar, S. Rosati, A. J. Heck *et al.*, "Complement is activated by igg hexamers assembled at the cell surface," *Science*, vol. 343, no. 6176, pp. 1260–1263, 2014.

[44] D. N. Mastronarde, "Dual-axis tomography: an approach with alignment methods that preserve resolution," *Journal of structural biology*, vol. 120, no. 3, pp. 343–352, 1997.

[45] J. Quispe, J. Damiano, S. E. Mick, D. P. Nackashi, D. Fellmann, T. G. Ajero, B. Carragher, and C. S. Potter, "An improved holey carbon film for cryo-electron microscopy," *Microscopy and microanalysis*, vol. 13, no. 5, pp. 365–371, 2007.

[46] V. Lučić, F. Förster, and W. Baumeister, "Structural studies by electron tomography: from cells to molecules," *Annu. Rev. Biochem.*, vol. 74, pp. 833–865, 2005.

[47] G. T. Herman, *Fundamentals of computerized tomography: image reconstruction from projections.* Springer Science & Business Media, 2009.

[48] M. Pilhofer, M. S. Ladinsky, A. W. McDowall, and G. J. Jensen, "Bacterial tem: new insights from cryo-microscopy," in *Methods in cell biology.* Elsevier, 2010, vol. 96, pp. 21–45.

[49] B. Turoňová, L. Marsalek, T. Davidovič, and P. Slusallek, "Progressive stochastic reconstruction technique (psrt) for cryo electron tomography," *Journal of structural biology*, vol. 189, no. 3, pp. 195–206, 2015.

[50] D. Vanhecke, S. Asano, Z. Kochovski, R. Fernandez-Busnadiego, N. Schrod, W. Baumeister, and V. Lučić, "Cryo-electron tomography: methodology, developments and biological applications," *Journal of microscopy*, vol. 242, no. 3, pp. 221–227, 2011.

[51] E. Orlova and H. R. Saibil, "Structural analysis of macromolecular assemblies by electron microscopy," *Chemical reviews*, vol. 111, no. 12, pp. 7710–7748, 2011.

[52] P. Penczek, M. Marko, K. Buttle, and J. Frank, "Double-tilt electron tomography," *Ultramicroscopy*, vol. 60, no. 3, pp. 393–410, 1995.

[53] Y. Chen, Y. Zhang, K. Zhang, Y. Deng, S. Wang, F. Zhang, and F. Sun, "Firt: filtered iterative reconstruction technique with information restoration," *Journal of structural biology*, vol. 195, no. 1, pp. 49–61, 2016.

[54] S. V. Venkatakrishnan, L. F. Drummy, M. Jackson, M. De Graef, J. Simmons, and C. A. Bouman, "Model-based iterative reconstruction for bright-field electron tomography," *IEEE Transactions on Computational Imaging*, vol. 1, no. 1, pp. 1–15, 2015.

[55] Y. Deng, Y. Chen, Y. Zhang, S. Wang, F. Zhang, and F. Sun, "Icon: 3d reconstruction with missing-informationrestoration in biological electron tomography," *Journal of structural biology*, vol. 195, no. 1, pp. 100–112, 2016.

[56] S. V. Venkatakrishnan, L. F. Drummy, M. A. Jackson, M. De Graef, J. Simmons, and C. A. Bouman, "A model based iterative reconstruction algorithm for high angle annular dark field-scanning transmission electron microscope (haadf-stem) tomography," *IEEE Transactions on Image Processing*, vol. 22, no. 11, pp. 4532–4544, 2013.

[57] A. Iudin, P. K. Korir, J. Salavert-Torres, G. J. Kleywegt, and A. Patwardhan, "Empiar: a public archive for raw electron microscopy image data," *Nature methods*, vol. 13, no. 5, p. 387, 2016.

[58] M. Tagari, R. Newman, M. Chagoyen, J.-M. Carazo, and K. Henrick, "New electron microscopy database and deposition system," *Trends in biochemical sciences*, vol. 27, no. 11, p. 589, 2002.

[59] Y. Wei, G. Wang, and J. Hsieh, "An intuitive discussion on the ideal ramp filter in computed tomography (i)," *Computers & Mathematics with Applications*, vol. 49, no. 5-6, pp. 731–740, 2005.

[60] X. Wang, A. Sabne, P. Sakdhnagool, S. J. Kisner, C. A. Bouman, and S. P. Midkiff, "Massively parallel 3d image reconstruction," in *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*. ACM, 2017, p. 3.

[61] F. Thon, "Phase contrast electron microscopy," *Electron Microscopy in Materials Science*, 1971.

[62] K. V. Fernando, "Radial averages of astigmatic tem images," *Journal of structural biology*, vol. 164, no. 1, pp. 49–59, 2008.

[63] Z. Huang, P. R. Baldwin, S. Mullapudi, and P. A. Penczek, "Automated determination of parameters describing power spectra of micrograph images in electron microscopy," *Journal of structural biology*, vol. 144, no. 1-2, pp. 79–94, 2003.

[64] W. Jiang, F. Guo, and Z. Liu, "A graph theory method for determination of cryo-em image focuses," *Journal of structural biology*, vol. 180, no. 2, pp. 343–351, 2012.

[65] S. P. Mallick, B. Carragher, C. S. Potter, and D. J. Kriegman, "Ace: automated ctf estimation," *Ultramicroscopy*, vol. 104, no. 1, pp. 8–29, 2005.

[66] B. Sander, M. Golas, and H. Stark, "Automatic ctf correction for single particles based upon multivariate statistical analysis of individual power spectra," *Journal of structural biology*, vol. 142, no. 3, pp. 392–401, 2003.

[67] C. Sorzano, S. Jonic, R. Núñez-Ramírez, N. Boisset, and J. Carazo, "Fast, robust, and accurate determination of transmission electron microscopy contrast transfer function," *Journal of structural biology*, vol. 160, no. 2, pp. 249–262, 2007.

[68] M. Vulović, E. Franken, R. B. Ravelli, L. J. van Vliet, and B. Rieger, "Precise and unbiased estimation of astigmatism and defocus in transmission electron microscopy," *Ultramicroscopy*, vol. 116, pp. 115–134, 2012.

[69] C. Yang, W. Jiang, D.-H. Chen, U. Adiga, E. G. Ng, and W. Chiu, "Estimating contrast transfer function and associated parameters by constrained non-linear optimization," *Journal of microscopy*, vol. 233, no. 3, pp. 391–403, 2009.

[70] W. Saxton, "A new way of measuring microscope aberrations," *Ultramicroscopy*, vol. 81, no. 2, pp. 41–45, 2000.

[71] R. Marabini, S. J. Ludtke, S. C. Murray, W. Chiu, M. Jose, A. Patwardhan, J. B. Heymann, and J. M. Carazo, "The electron microscopy exchange (emx) initiative," *Journal of structural biology*, vol. 194, no. 2, pp. 156–163, 2016.

[72] J.-J. Fernández, J. Sanjurjo, and J.-M. Carazo, "A spectral estimation approach to contrast transfer function detection in electron microscopy," *Ultramicroscopy*, vol. 68, no. 4, pp. 267–295, 1997.

[73] J. Zhu, P. A. Penczek, R. Schröder, and J. Frank, "Three-dimensional reconstruction with contrast transfer function correction from energy-filtered cryoelectron micrographs: Procedure and application to the 70sescherichia coliribosome," *Journal of structural biology*, vol. 118, no. 3, pp. 197–219, 1997.

[74] X. Li, P. Mooney, S. Zheng, C. R. Booth, M. B. Braunfeld, S. Gubbens, D. A. Agard, and Y. Cheng, "Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-em," *Nature methods*, vol. 10, no. 6, p. 584, 2013.

[75] G. Yu, K. Li, Y. Liu, Z. Chen, Z. Wang, R. Yan, T. Klose, L. Tang, and W. Jiang, "An algorithm for estimation and correction of anisotropic magnification distortion of cryo-em images without need of pre-calibration," *Journal of structural biology*, vol. 195, no. 2, pp. 207–215, 2016.

[76] R. Marabini, B. Carragher, S. Chen, J. Chen, A. Cheng, K. H. Downing, J. Frank, R. A. Grassucci, J. B. Heymann, W. Jiang *et al.*, "Ctf challenge: result summary," *Journal of structural biology*, vol. 190, no. 3, pp. 348–359, 2015.

[77] K. Ishizuka, "Coma-free alignment of a high-resolution electron microscope with three-fold astigmatism," *Ultramicroscopy*, vol. 55, no. 4, pp. 407–418, 1994.

[78] F. Zemlin, K. Weiss, P. Schiske, W. Kunath, and K.-H. Herrmann, "Coma-free alignment of high resolution electron microscopes with the aid of optical diffractograms," *Ultramicroscopy*, vol. 3, pp. 49–60, 1978.

[79] J. Sun and H. Li, "How to operate a cryo-electron microscope," in *Methods in enzymology*. Elsevier, 2010, vol. 481, pp. 231–249.

[80] R. A. Grassucci, D. Taylor, and J. Frank, "Visualization of macromolecular complexes using cryo-electron microscopy with fei tecnai transmission electron microscopes," *Nature protocols*, vol. 3, no. 2, p. 330, 2008.

[81] Y. Cheng, N. Grigorieff, P. A. Penczek, and T. Walz, "A primer to single-particle cryo-electron microscopy," *Cell*, vol. 161, no. 3, pp. 438–449, 2015.

[82] P. A. Penczek, "Image restoration in cryo-electron microscopy," in *Methods in enzymology*. Elsevier, 2010, vol. 482, pp. 35–72.

[83] R. Yan, K. Li, and W. Jiang, "Real-time detection and single-pass minimization of tem objective lens astigmatism," *Journal of structural biology*, vol. 197, no. 3, pp. 210–219, 2017.

[84] K. Zhang, "Gctf: Real-time ctf determination and correction," *Journal of structural biology*, vol. 193, no. 1, pp. 1–12, 2016.

[85] T. Grant and N. Grigorieff, "Automatic estimation and correction of anisotropic magnification distortion in electron microscopes," *Journal of Structural Biology*, vol. 192, no. 2, pp. 204–208, 2015.

[86] P. W. Hawkes, *The beginnings of electron microscopy*. Academic Press, 2013, vol. 16.

[87] P. Rai-Choudhury, *Handbook of microlithography, micromachining, and microfabrication: microlithography*. Iet, 1997, vol. 1.

[88] J. Vargas, J. Otón, R. Marabini, S. Jonic, J. De La Rosa-Trevín, J. Carazo, and C. Sorzano, "Fastdef: fast defocus and astigmatism estimation for high-throughput transmission electron microscopy," *Journal of structural biology*, vol. 181, no. 2, pp. 136–148, 2013.

[89] J. Barthel and A. Thust, "On the optical stability of high-resolution transmission electron microscopes," *Ultramicroscopy*, vol. 134, pp. 6–17, 2013.

<div style="text-align:center">VITA</div>

EDUCATION

Ph.D., Structural and Computational Biology and Biophysics, Computational Life Sciences, Purdue University, West Lafayette, Indiana, USA

M.S., Basic Medical Sciences, Purdue University, West Lafayette, Indiana, USA

M.S., Bioinformatics, Nankai University, Tianjin, China

B.E., Mechanical Engineering, Tianjin University, Tianjin, China

PUBLICATIONS

**Yan, R.**, Venkatakrishnan, S., Liu, J., Bouman, C., Jiang, W., 2019. MBIR: A Cryo-electron Tomography 3D Reconstruction Method that Effectively Minimizes Missing Wedge Artifacts and Restores Missing Information. J. Struct. Biol. (In Press)

**Yan, R.**, Li, K., Jiang, W., 2018. Defocus and magnification dependent variation of TEM image astigmatism. Sci Rep 8 (1), 344.

**Yan, R.**, Li, K., Jiang, W., 2017. Real-time detection and single-pass minimization of TEM objective lens astigmatism. J. Struct. Biol. 197, 210219

**Yan, R.**, Jiang, W., 2016. A Digital Micrograph Script for Detection of Astigmatism in TEM Images. Microsc. Microanal. 22, 20722073.

**Yan, R.**, Edwards, T.J., Pankratz, L.M., Kuhn, R.J., Lanman, J.K., Liu, J., Jiang, W., 2015. A fast cross-validation method for alignment of electron tomography images based on Beer-Lambert law. J. Struct. Biol. 192, 297306.

**Yan, R.**, Edwards, T.J., Pankratz, L.M., Kuhn, R.J., Lanman, J.K., Liu, J., Jiang, W., 2015. Simultaneous determination of sample thickness, tilt, and electron mean free path using tomographic tilt images based on Beer-Lambert law. J. Struct. Biol. 192, 287296.

Yu, G., **Yan, R.**, Zhang, C., Mao, C., Jiang, W., 2015. Single-Particle Cryo-EM and

3D Reconstruction of Hybrid Nanoparticles with Electron-Dense Components. Small 11, 51575163.

Yu, G., Vago, F., Zhang, D., Snyder, J.E., **Yan, R.**, Zhang, C., Benjamin, C., Jiang, X., Kuhn, R.J., Serwer,P., Thompson, D.H., Jiang, W., 2014. Single-step antibody-based affinity cryo-electron microscopy for imaging and structural analysis of macromolecular assemblies. J. Struct. Biol. 187, 19.

**Yan, R.**, Page, J.C., Shi, R., 2015. Acrolein-mediated conduction loss is partially restored by K channel blockers. J. Neurophysiol. 115, 701710.

## SELECTED TALKS & POSTERS

**Purdue Cryo-EM Symposium**, Purdue University, November 6-8th, 2019.

Presentation: $s^2$ Power Spectra Based CTF Determination and Applications

**The Hitchhiker's Guide to the Biomolecular Galaxy**, Purdue University, May 9-10th, 2018.

Poster: MBIR: A Cryo-electron Tomography 3D Reconstruction Method that Effectively Minimizes Missing Wedge Artifacts and Restores Missing Information

**The Hitchhiker's Guide to the Biomolecular Galaxy**, Purdue University, May 10-11th, 2017.

Presentation: $s^2 stigmator$: real-time detection and single-pass minimization of TEM objective lens astigmatism

**Computational Science and Engineering Student Conference**, Purdue University, April 14th, 2017.

Poster: A Digital Micrograph script for real-time detection and single-pass minimization of TEM objective lens astigmatism

**2016 Microscopy & Microanalysis Meeting**, Columbus, Ohio, July 24-28th, 2016.

Abstract: A Digital Micrograph Script for Detection of Astigmatism in TEM images

**The Hitchhiker's Guide to the Protein Galaxy**, Purdue University, May 13-14th, 2015.

Poster: A fast cross-validation method for alignment of electron tomography images based on Beer-Lambert law

HONORS & AWARDS

H.E. Umbarger Outstanding Graduate Student in Research Award, April 25th, 2017

Computational Science and Engineering Student Conference Poster Award, April 14th, 2017

College of Science Cagiantas Fellowship, April 5th, 2017

Microscopy & Microanalysis Meeting Student Poster Award, July 26th, 2016

Travel Grant Recipient, December 2015

Graduate School Summer Research Grant, April 2015