INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.
Image Analysis and Image Reconstruction Improvements in the Determination of Protein Three-Dimensional Structures by Electron Microscopy

by

Yongyi Mao

A thesis submitted in conformity with the requirements for the degree of Master of Science Graduate Department of Medical Biophysics University of Toronto

© Copyright by Yongyi Mao (July, 1998)
The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

0-612-40785-3
Abstract

Image analysis and image reconstruction processes are critical in obtaining protein three-dimensional structures after their two-dimensional images are acquired by electron microscopy. This thesis reports my work on improving several aspects in this field.

The angular distribution dependent (ADD) filtering algorithm is developed for an improved accuracy in three-dimensional reconstruction by the filtered backprojection method. The phase difference analysis (PDA) is employed to enhance the peaks in the sinogram correlation functions for a more accurate determination of projection directions for the randomly orientated protein particles. These improvements will result in a higher achievable resolution in determining protein structures by electron microscopy.

A "shape factor" is defined for an investigation of automated molecular segmentation from electron micrographs and the potential feasibility has been shown. A novel three-dimensional reconstruction method, the finer voxel method, is proposed, and a theoretical discussion suggests that, with this approach, atomic resolution is potentially obtainable.
Acknowledgment to My Supervisor

I believe my graduate study in the University of Toronto with my supervisor, Dr. F. Peter Ottensmeyer, is one of the most significant things that have ever happened to my life. In the past three years with Peter I have been offered the best education that I have ever experienced, and I believe this is also the best education one could possibly experience.

I feel thankful to Peter, for his encouragement. It was he who encouraged me to have initiative, to be independent, and ambitious. He told me not to take anything for granted. He reminded me that I thought I was smarter than anybody else. He encouraged me to challenge current canonical methods. He made me think we are able to struggle a new way out of conventions. He said that likes graduate students, because they are so ignorant; that it is this ignorance that makes them think of something new, something different. All those words impressed me so much and they gave me the courage to try my new silly ideas, even if they failed.

I feel thankful to Peter, for his teaching. I do not want to mention his teaching that filled me with wide knowledge in science. The knowledge of human beings is always so insignificant to the universe. But there is something that once you have, you have your telescope to comprehend the universe. This is the philosophy of science. And this is what I feel grateful for. Peter taught me how to appreciate research, how to enjoy science, and how to step out of
the details of a problem and look at the problem from outside. I like his opinion on “truth”. He thinks that one can not ever get an absolute truth in science, but that what one can get is only the currently best approximation to the truth. All his philosophy will be an eternal guidance in my future pursuit for the truth of science.

I feel thankful to Peter, for his toughness on me. He embarrassed me on every mistake I made. In private he stopped me anytime in my talk when there was a pronunciation mistake. He let me practice for a presentation again and again as long as he thought it unsatisfactory; he even made me tone down the statements in this acknowledgement. Eventually my logic became clearer, my English became better, and I could get an “A” for my student seminar. All these are given by Peter! On finishing the program, Peter said to me, “You still need to practice your English. You need a nice person who can be tough on you.”

The years spent with Peter is the most enjoyable time for me. I believe what I have learned in these years will benefit my whole upcoming life. Some day when I am happy for my accomplishment, the first thing I shall say will be “Thank you, Peter!”

Yongyi Mao

July, 1998
Acknowledgment

I would like to thank the members in my supervisory committee, Dr. F. Peter Ottensmeyer, Dr. Emil Pai, and Dr. Stuart Foster for their scientific insights from different points of view, for their supervision on my progress, and for their guidance in this project.

I would like to thank Dr. George Harauz from University of Guelph, for his helpful discussion in exact filters, Fourier analysis, correlation methods and many other aspects in the field during this project.

I would like to thank Brenda Rutherford for her kind help in specimen preparation, operation of microscopes, and setting up experiments.

I would also like to thank Yew Meng Heng, Allan Fernandes, Michael Burke, David McAlDuff, Melanie Barfels, Gregory Czarnota, and Wen Li from University of Guelph, for their technical assistance, inspiring discussion, and noble friendship.

Yongyi Mao

July, 1998
**Table of Contents**

Chapter 1 Introduction 1

1.1 Electron microscopy in molecular and structural biology 1

1.2 Scanning transmission electron microscope 5

1.3 Specimen preparation 8

1.4 Image analysis and image reconstruction in electron microscopy 9

1.5 Thesis outline 14

Chapter 2 An Angular Distribution Dependent Filtering Method in Three-dimensional Reconstruction by Filtered Backprojection 17

2.1 Introduction 17

2.2 Filtered backprojection 19

2.3 ADD filtering for three-dimensional reconstruction 28

2.4 Comparison of ADD filtering with other filtering methods 31

2.5 Discussion and conclusion 40

Chapter 3 Sinogram Correlation Function by Phase Difference Analysis 44

3.1 Introduction 44

3.2 The SCF using the cross correlation coefficient 47

3.3 SCF using phase difference analysis 49

3.4 Discussion 54

3.4.1 Peak enhancement 54

3.4.2 Flexibility in the manipulating phases at different frequencies 55

3.4.3 Robustness to shift error 56

3.4.4 Phase wraparound 56

3.4.5 Computational speed 57
3.5 Conclusion

Chapter 4 Summary and Future Directions

4.1 Summary

4.2 Future directions

4.2.1 Introduction

4.2.2 Automated molecular segmentation in electron micrographs using a shape factor calculation

4.2.2.1 Introduction

4.2.2.2 Shape factor and compactness-etching plot

4.2.2.3 Conclusion

4.2.3 Determination of protein structure at atomic resolution with STEM: Three-dimensional reconstruction by a finer voxel method

4.2.3.1 Introduction

4.2.3.2 Finer voxel method

4.2.3.3 Underlying principles

4.2.3.3.1 Superposition of multiple measurements

4.2.3.3.2 Deconvolution of beam profile function

4.2.3.3.3 Principles of finer voxel method

4.2.3.4 Discussion

4.3 Conclusion

Appendix 1 Fourier analysis on superposition of multiple measurements

Appendix 2 Superposition of multiple measurements and deconvolution of beam profile function by solving linear equations

Appendix 3 Estimation the error in the calculation of center of mass
Appendix 4 Some formulae used in the calculation of the tables

References
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2D</td>
<td>two-dimensional</td>
</tr>
<tr>
<td>3D</td>
<td>three-dimensional</td>
</tr>
<tr>
<td>ADD filtering</td>
<td>angular distribution dependent filtering</td>
</tr>
<tr>
<td>FSC</td>
<td>Fourier shell correlation</td>
</tr>
<tr>
<td>FT</td>
<td>Fourier transform</td>
</tr>
<tr>
<td>PDA</td>
<td>phase difference analysis</td>
</tr>
<tr>
<td>SCF</td>
<td>sinogram correlation function</td>
</tr>
<tr>
<td>STEM</td>
<td>scanning transmission electron microscope</td>
</tr>
<tr>
<td>TEM</td>
<td>transmission electron microscope</td>
</tr>
</tbody>
</table>
List of Figures

Figure 1-1 ....................................................................................................................................... Page 6
Figure 1-2 ....................................................................................................................................... Page 11
Figure 1-3 ....................................................................................................................................... Page 12
Figure 2-1ab .................................................................................................................................. Page 21
Figure 2-1c ....................................................................................................................................... Page 22
Figure 2-2ab ................................................................................................................................... Page 26
Figure 2-2c ....................................................................................................................................... Page 27
Figure 2-3 ....................................................................................................................................... Page 32
Figure 2-4ab ................................................................................................................................... Page 33
Figure 2-4c ....................................................................................................................................... Page 34
Figure 2-5a ....................................................................................................................................... Page 36
Figure 2-5b ....................................................................................................................................... Page 37
Figure 2-5c ....................................................................................................................................... Page 38
Figure 2-6 ....................................................................................................................................... Page 39
Figure 3-1 ....................................................................................................................................... Page 48
Figure 3-2ab .................................................................................................................................... Page 52
Figure 3-2cd .................................................................................................................................... Page 53
Figure 3-3a ....................................................................................................................................... Page 58
Figure 3-3b ....................................................................................................................................... Page 59
Figure 3-3c ....................................................................................................................................... Page 61
Figure 4-1 ....................................................................................................................................... Page 68
Figure 4-2 ....................................................................................................................................... Page 70
Figure 4-3 ....................................................................................................................................... Page 71
Figure 4-4 ....................................................................................................................................... Page 73
Figure 4-5 ....................................................................................................................................... Page 77
Chapter 1 Introduction

1.1 Electron microscopy in molecular and structural biology

As biology has advanced to the molecular level, the biological behavior of proteins is of increasing interest to researchers. Biologists have established the concept that structures of proteins are closely related to their functions. In order to study the biological function of a protein and the interaction between the protein and its biological environment, it is usually necessary to perform a structural study on the protein. This study can either provide a structural basis of the protein functions or help to further the understanding of the biological behavior of the molecule.

Electron microscopy has played an important role in biology. Most structural information of biological organisms at subcellular level was obtained by electron microscopy. Since Finch and Klug started quantitative analysis on electron micrographs of viruses to study their three-dimensional structures thirty years ago (Finch and Klug, 1965), electron microscopy has been used widely to elucidate biological processes at molecular level. A broad spectrum of proteins have been studied by electron microscopy, from large protein complexes, like the ribosome, with molecular weight of a few thousand kD (Stark et al. 1997), to small polypeptides, like vasopressin-8-lysine I, composed of only nine amino acids (Ottensmeyer et al. 1977).
Utilizing electrons as the source of imaging, electron microscopy can potentially attain a resolution of 0.1 Å or better (based on the wavelength of electron). However, because of the spherical aberration and chromatic aberration of the imaging system, as well as the radiation damage from which biological specimens suffer, the current achievable resolution in the study of protein structures is at the nanometre to subnanometre level. This is often considered as the major disadvantage of electron microscopy, in comparison with the atomic resolution that can be obtained by the other two conventional techniques in this field, X-ray crystallography and NMR spectroscopy.

Nevertheless, electron microscopy has some advantages over these two techniques. First, electron microscopy can be performed with very little amounts of protein material, usually of nanomole quantities, while the amount of protein required in an X-ray crystallography or NMR spectroscopy experiment is often thousands times greater. Secondly, electron microscopy does not require crystallization or labeling of proteins as are required by X-ray crystallography and NMR spectroscopy. Thirdly, proteins of various sizes can be studied by electron microscopy. In contrast, NMR spectroscopy presently is not capable of studying protein bigger than 30 kD, and in the study of big proteins by X-ray crystallography there are either difficulties in crystallization or problems in solving phases. This makes electron microscopy effectively a unique approach to solve the structures of large proteins or protein complexes. Finally, even compared with X-ray crystallography, in which crystals may be surrounded by a physiological "mother liquor", in cryo-
electron microscopy experiments the proteins can be closer to their physiological condition, since there is no concern of crystallization-induced structure change.

For the reasons above, electron microscopy, although limited by resolution, still plays a significant role in the determination of protein structures. The highest resolution structures in electron microscopy have been obtained using two dimensional crystalline array of proteins, e.g., bacteriorhodopsin, the light harvesting complex (Henderson et al., 1976, 1990; Kühlerbrandt et al., 1994). However, even for single protein particle structures, in the last thirty years, the resolution of the structures determined by electron microscopy has been improved greatly by progress either in instrumentation, computation, or specimen preparation. For example, the use of the field emission gun, which generates highly coherent electron beams, enables microscopes to record high quality phase contrast images; multivariate statistical analysis, class averaging, single particle alignment, and three-dimensional reconstruction enhance the signal-to-noise ratio in images and allow the three-dimensional structure of randomly oriented protein particles to be obtained (De Rosier and Klug, 1968, van Heel, 1985, 1987, 1989); imaging and specimen preparation at liquid nitrogen temperature (cryo-techniques) preserve the native conformation of proteins and their high-resolution structural details. A review of some accomplishments in protein structure determination by electron microscopy can be used as an indication to the achieved resolution by this technique. The structure of the calcium-release channel (molecular weight 2600 kD, four-fold symmetry) was determined by Chiu and van Heel’s groups at 30 Å resolution.
(Serysheva et al., 1995); a 54 kD asymmetrical small protein, SRP54, was studied by Ottensmeyer’s group and 11 Å resolution was obtained (Czarnota et al., 1994); more recently a 9 Å structure and a 7.4 Å structure of hepatitis B virus capsid protein have been obtained by Steven’s group and Crowther’s group (Conway et al., 1997, Böttcher et al., 1997).

Due to the rapid progress of electron microscopy, sometimes the technique is considered to be competing with X-ray crystallography and NMR spectroscopy. Actually, electron microscopy is complementary to the other two techniques. Amongst the three techniques, each has certain advantages and disadvantages. The combination of their advantages sometimes can help to solve structures that are difficult for each technique separately. An elegant example of this combination is the determination of rhinovirus-Fab complex structure at 4 Å resolution by a combination of X-ray crystallography and electron microscopy (Smith et al., 1996). The atomic models of rhinovirus and Fab fragment were fit into a low resolution map of their complex solved by electron microscopy and they were used as a phase model for X-ray crystallography to determine the structure. Furthermore, another method of combining X-ray crystallography and electron microscopy has been proposed, where amplitudes from X-ray diffraction data and low resolution phases from structure solved by electron microscopy are combined to locate the heavy atoms in the crystal or to extend phases based on high order of symmetry (Mao et al., 1997). Modeling of this process has shown a potential feasibility.
In the rest of this introductory chapter, the imaging device used in this project, a scanning transmission electron microscope, is introduced in section 1.2. Section 1.3 describes the process of specimen preparation. Section 1.4 briefly illustrates the routine procedures of image analysis and image reconstruction used in electron microscopy, which form the background of the remaining chapters in the thesis. The organization of the thesis is outlined in section 1.5.

1.2 Scanning transmission electron microscope

The instrument used in this study is a scanning transmission electron microscope (STEM: VG601UX; Vacuum Generators; England). Its schematic layout is shown in Figure 1-1.

The electron source is a field emission gun. This is a very fine pointed single crystal of tungsten in a field strong enough to draw electrons from the tip by field emission. Such a gun can generate a very fine electron beam with very high brightness. The electron beam is then focused into a spot with a diameter of 3 \( \text{Å} \). This fine electron probe is scanned in a raster over the specimen surface by appropriate scan coils. Electrons transmitted by the specimen include elastically scattered electrons, scattered through relatively large angles, and inelastically scattered electrons, contained within the central cone of the beam. Most of the elastically scattered electrons are collected by an annular detector, and the inelastically scattered electrons, which are predominantly forward scattered, pass the central aperture of the detector and are energy-analyzed by an electron
Figure 1-1 Schematic layout of an STEM
spectrometer\(^1\) (PEELS, GATAN Inc., Pleasanton, CA). In STEM, the electron collection efficiency is improved greatly, since the collection angle is no longer limited by the lens spherical aberration as in common transmission electron microscopes (TEM) (there are no electron lenses after the specimen).

Compared with a TEM operating at room temperature, our STEM decreases electron dose on the specimen 200-fold for the same signal-to-noise ratio by detecting both elastically and inelastically scattered electrons and by imaging the specimen at liquid nitrogen temperature. This is important in reducing the radiation damage to protein specimens, and the fine electron probe makes the STEM a high resolution imaging device.

\(^1\) In the STEM in this lab, the conventional linear diode array detector on the spectrometer has been replaced by an integrating, faster, and more efficient photomultiplier to collect the total inelastically scattered electron signal (not shown in Figure1-1).
1.3 Specimen preparation

To image proteins in a near-native structural configuration and environment, protein specimens for electron microscopy are prepared by cryo-techniques.

Specimen grids are used to support the protein particles. Grids are coated by a layer of holey plastics, and then they are covered with a very thin layer of carbon film (20~30 Å). Protein solution is diluted to a proper concentration, for which the protein particles will spread uniformly on the carbon film with a proper spacing between molecules. This adjustment of concentration is carried out by trial and error.

A small drop of protein buffer is placed onto a specimen grid, then a drop of protein solution at the proper concentration is injected into the buffer bubble. After a few minutes of incubation to let the protein particles attach to the carbon film, the extra protein solution is sucked off by a filter paper. The still moist specimen is then quickly frozen by plunging it into liquid ethane (at -170 °C). The purpose of this process is to preserve the native conformation of protein particles in a glass-like aqueous solid.

After transferring the prepared specimen into microscope through liquid nitrogen environment, the specimen is then freeze-dried and imaged at a temperature lower than -130 °C. It can also be imaged in the vitreous ice layer directly if this is sufficiently thin.
1.4 **Image analysis and image reconstruction in electron microscopy**

The STEM images are two-dimensional presentations of the protein molecules, reflecting projections of the three-dimensional electron scattering density distribution. To obtain the three-dimensional structure, computational approaches are required after the images are obtained. This section briefly introduces the image analysis and image reconstruction procedures.

Individual molecular images are selected and cropped from the collected images. Each molecular image is normalized to correct for the difference in the intensity of the incident beams that give rise to the images. This normalization is carried out by dividing the brightness, or density, of each pixel in an image by the averaged carbon background brightness in that image. Then the carbon backgrounds are subtracted from all the images. Subsequently some noise suppression algorithms, such as low-pass filtering or box convolution, are used to reduce the noise present in the images (Jain, 1989).

Molecular boundaries are determined, and each molecule in an image is segmented from the surrounding background and shifted to align its center of mass to the center of the image.

Since the protein particles are randomly orientated on the carbon film, each particle is potentially imaged along a different orientation direction. These orientations are not known *a priori* and can only be calculated after the images are obtained. This determination of protein particle orientations
*a posteriori* is part of the process referred to as angular reconstitution (van Heel, 1987, Farrow and Ottensmeyer, 1992, 1993). The orientation determination is based on the Central Section Theorem, which states that the 2D Fourier transform of a 2D projection is a central section of the 3D Fourier transform of the 3D object that gives rise to the 2D projection. Thus the 2D Fourier transforms of any two projections of the same object will have a common line, *i.e.*, the intersection line, in the 3D Fourier space. If the common lines between any three projections from different orientations are known, the orientations, or the projections directions, of these three projections can be geometrically calculated (van Heel, 1987). This can be explained graphically in Figure 1-2. The convention of Euler angles used to describe projection orientations is illustrated in Figure 1-3. When there are more than three projections, the inconsistency of common line data due to noise and error makes the determination of orientations an optimization problem. A fast and elegant method, the quaternion assisted approach, is used in this lab to solve the problem (Farrow and Ottensmeyer, 1992).

The search for the common lines is the core of the angle determination processes. A so-called sinogram correlation function (SCF) is calculated for this purpose and the maxima in the SCFs are located. The SCF between any two images determines the similarity or correlation between any possible two line projections, one each from the pair of images being considered. The global maxima in the SCFs between any pair of images provide the angles of common lines between them.
Figure 1–2 Determination of the orientations of three projections. In each projection, the common lines between this projection and the other two projections have been determined. $C_{ij}$ refers to the common line between projection $i$ and projection $j$ in the plane of projection $i$. To align the three projections in space, an appropriate rotation has to be applied to each image so that the common line $C_{ij}$ overlaps with $C_{ji}$ in space.
Figure 1–3 The convention of Euler angles used in this thesis. Alpha, a rotation about z axis; beta, a rotation about the intermediate y axis; gamma, a rotation about the new z' axis.
After the projection directions are determined for all the molecular images, a 3D reconstruction can be performed using a filtered backprojection method (Gilbert, 1972, Shepp and Logan, 1974, Tanaka, 1979, Harauz and van Heel, 1986). The 3D structures obtained are then refined by iteratively reprojecting the structure into 2D projections at certain orientations and using them as new reference to recalculate the orientations of the images. The self consistency error in the fitting of common lines is minimized by this iterative process (Serysheva et al., 1995, Ottensmeyer et al., 1998).

When a reconstruction has been obtained on a protein molecule with unknown structure, the measure of the resolution achieved employs a self consistency criterion. Images are split into two groups, and from each group a reconstruction is performed separately. Fourier analysis such as Fourier shell correlation or the use of differential phase residuals is used to determine a "consistency resolution", or, the reciprocal of a cutoff frequency below which the two reconstructions are well correlated (Frank, 1996).
1.5 Thesis outline

As a recent technique in structural and molecular biology, electron microscopy is still being developed. Currently electron microscopy is not capable of determining protein structures at atomic resolution. This is due to the instrumental limitations, such as spherical aberration and chromatic aberration, which limit the acquisition resolution (Meek, 1970). The fragility of protein specimens is another limiting factor (Ottensmeyer et al., 1978). It limits the electron irradiation dose to levels low enough to avoid serious disruption of the protein structure, and so results in a fairly low signal-to-noise ratio in the image. The noise in turn produces inaccuracies in the orientation determination and thus lowers the achievable resolution. The work presented in this thesis is focused on improving several aspects of the image analysis and image reconstruction techniques in order to increase the robustness of the analysis and the accuracy of the reconstruction.

Chapter 2 describes a reconstruction algorithm, the angular distribution dependent (ADD) filtered backprojection algorithm, that is an improvement upon the conventional filtered backprojection method. The theory of 3D reconstruction from 2D projections by the filtered backprojection method has been well established for projections with uniform and quasi-continuous distribution of orientations in medical imaging (Gilbert, 1972). However, the reconstruction problem in electron microscopy is different from medical applications above in that the orientations are in general sparsely and nonuniformly distributed. In this case the canonical filtered backprojection method
does not produce a satisfactory reconstruction. In recognition of this problem, an “exact” filtering method that addresses the actual orientation distribution has been proposed for filtered backprojection to improve the accuracy of 3D reconstruction (Harauz and van Heel, 1986). In this thesis, a reexamination of backprojection reconstruction led to the development of a new filtering algorithm, the angular distribution dependent filtering method, which improves still further upon current filtering methods for backprojection.

Chapter 3 presents a new method of calculating the sinogram correlation functions (SCFs). The conventional method of calculating the SCFs in angular reconstitution employs cross correlation coefficients as the measure of similarity between line projections. Peaks in the resultant SCFs tend to be broad and flat-topped. This makes it difficult to locate the maxima, especially for noisy experimental data. In this work, a new measure of correlation, a modified phase difference analysis, is used to take the place of the cross correlation coefficient in the calculation of the SCFs. It enhances peaks in the SCFs, making them more distinctively defined. It can be implemented as a fast computational algorithm, and is also more robust to shift errors.

Chapter 4 summarizes the finished work in this project (those presented in Chapter 2 and Chapter 3 of the thesis) and describes some future directions in this area that have been investigated experimentally or theoretically during this project. These future directions include the employment of the “shape factor” for automated molecular segmentation and the “finer voxel” method for reconstruction of STEM images to atomic resolution. Preliminary results have been obtained in
the investigation of the "shape factor", suggesting that the "shape factor" calculation is a promising approach for molecular segmentation. The "finer voxel" method is discussed theoretically and it is shown capable of reconstructing a protein structure at a higher resolution than the imaging resolution determined by the sampling interval of the STEM. Using this method, protein structures at atomic resolution are potentially obtainable by electron microscopy.
Chapter 2  An Angular Distribution Dependent Filtering Method in Three-dimensional Reconstruction by Filtered Backprojection

2.1 Introduction

Two-dimensional or three-dimensional image reconstruction has been used in many fields, such as diagnostic radiology, radio astronomy, and electron microscopy. Among different reconstruction methods, filtered backprojection is one of the most established. To perform a reconstruction by this method, projections are filtered, or weighted at each frequency, before they are backprojected. The filtering serves to eliminate the blurring effect in reconstructions caused by the backprojection operation. When this method is applied, the reconstruction problems can be divided into two classes, depending on whether the projection directions are distributed uniformly or not. When the projection directions are uniformly distributed, as for instance in x-ray CT reconstruction, an identical filtering function is applied to each projection before the backprojection operation is performed. In the situation where the projections were generated with a set of nonuniformly distributed orientations, the filtering function should in principle be individualized for each projection. The typical example for this situation is electron microscopic image reconstruction from randomly oriented protein particles. Here the protein orientations are not known a priori, but
can be calculated by angular reconstitution techniques. Generally, the orientations of such protein particles are not uniformly distributed.

For uniformly distributed orientations, a ramp filtering function and its modified derivatives for noise reduction have been proposed (Gilbert, 1972, Shepp and Logan, 1974, Tanaka, 1979). However in the derivation of the ramp filtering function, the implicit assumption is made that an infinite number of different projections are available for reconstruction. This assumption never holds in practice. For the nonuniform case, ramp filtering is not capable of producing a good approximation for the reconstruction, in that it results in an “elongation” artifact along any “clustering” direction. An individualized filtering method, “exact filtering”, has been proposed to correct for the overweighted projections in the clustering directions (Harauz and van Heel, 1986). The exact filtering method is able to reduce the artifacts and produce more accurate reconstructions.

We present here a new numerical filtering method for reconstruction, the angular distribution dependent filtering method (referred to as ADD filtering below), which, like “exact filtering”, also corrects for clustering of projection orientations. This method is shown to produce results that improve upon those achieved by ramp filtering and exact filtering, both for discrete uniform distribution and for nonuniform angular distributions.
To provide context, the principle of filtered backprojection is described first. Then the ADD filtering method is introduced and compared to ramp filtering and to "exact" filtering using model data.

2.2 Filtered backprojection

For a two-dimensional density distribution function $O(x, y)$, its one-dimensional projection function along y direction $p(x)$ is:

$$p(x) = \int_{-\infty}^{\infty} O(x, y) dy$$  \hspace{1cm} (2-1-1)

For an arbitrary projection direction $\theta$, the one-dimensional projection function is:

$$p_{\theta}(x) = \int_{-\infty}^{\infty} R_{\theta} \{ O(x, y) \} dy$$  \hspace{1cm} (2-1-2)

where $R_{\theta}$ refers to a rotation operation about the origin with angle $\theta$.

By direct backprojection, the reconstructed two-dimensional density function is:

$$O'(x, y) = \sum_{\theta} R_{\theta}^{-1} \{ p_{\theta}(x) \}$$  \hspace{1cm} (2-1-3)
where $p_n(x)$ is now treated as a two-dimensional function, varying in $x$ and constant in $y$. $\mathbf{R}^{-1}$ is the inverse rotation $\theta$ about the origin. The reconstruction of function $O(x, y)$, $O'(x, y)$, is not identical to $O(x, y)$. This can be understood by considering $O(x, y)$ as a point (Figure 2-1). When a finite number of projections are used for backprojection, the reconstructed two-dimensional function is a star-like density distribution. When an infinite number of different projections are backprojected, the reconstructed two-dimensional function is a continuous isotropic point spread function (Gilbert, 1972). It is easy to formulate this function using a polar coordinate system as:

$$s(r, \theta) = \sqrt{\frac{1}{\pi r}}$$  \hspace{1cm} (2-1-4)

where $r$ is the radius in the polar coordinate system. It has been shown (Tanaka, 1979) that this point spread function has a Fourier transform

$$\mathfrak{F}\{s(r, \theta)\} = \frac{1}{\pi R}$$  \hspace{1cm} (2-1-5)

where $\mathfrak{F}$ represents the Fourier transform operation and $R$ is frequency in Fourier space. This point spread function has to be deconvolved from $O'(x, y)$ to restore the original density distribution $O(x, y)$. The deconvolution is performed by suitably filtering each one-dimensional projection before the backprojection operation. This filtering can be carried out either by
Figure 2-1c

Figure 2-1 Backprojection of a point. (a) A point and its projections. (b) Backprojecting a finite number of the projections produces a "star-like" density distribution. (c) When an infinite number of the projections are backprojected, the density distribution profile is a continuous point spread function.
convolving each projection in real space with a filter function (the so-called convolution backprojection

\[ h(r) = \frac{1}{\sqrt{2\pi r^2}} \quad (2-1-6) \]

or by multiplying the Fourier transform of the projection with the Fourier transform of the filtering function.

\[ H(R) = |\pi R| \quad (2-1-7) \]

For a three-dimensional density distribution function \( O(x, y, z) \), its two-dimensional projection along the \( z \) direction \( p(x, y) \) is:

\[ p(x, y) = \int_{-\infty}^{\infty} O(x, y, z) \, dz \quad (2-2-1) \]

For an arbitrary projection direction along an Euler angle \( \Omega \), the projection function is:

\[ p_{\Omega}(x, y) = \int_{-\infty}^{\infty} R_{\Omega}\{O(x, y, z)\} \, dz \quad (2-2-2) \]

where \( R_{\Omega} \) is a rotation operator about the origin with Euler angle \( \Omega \).
By direct backprojection the reconstructed three-dimensional function is:

\[ O'(x, y, z) = \sum_{\alpha} \mathcal{R}_{\alpha}^{-1} \{ p_{\alpha}(x, y) \} \]  \hspace{1cm} (2-2-3)

where \( p_{\alpha}(x, y) \) is now treated as a three-dimensional function varying in \( x \) and \( y \), and constant in \( z \). \( \mathcal{R}_{\alpha}^{-1} \) is the inverse rotation operator with Euler angle \( \Omega \). When assuming the number of different projections is infinite, \( O'(x, y, z) \) is equal to \( O(x, y, z) \) convolved with a point spread function

\[ s(r, \Omega) = \frac{1}{\sqrt{2\pi r^2}} \]  \hspace{1cm} (2-2-4)

and which has the Fourier transform (Tanaka, 1979)

\[ \mathcal{F}\{ s(r, \Omega) \} = \frac{1}{2R} \]  \hspace{1cm} (2-2-5)

To restore the original density function \( O(x, y, z) \), it is necessary to convolve the projections with

\[ h(r) = \frac{-1}{\sqrt{2\pi^2 r^3}} \]  \hspace{1cm} (2-2-6)

or filter the projections in Fourier domain by multiplying with
Equation (2-1-7) and (2-2-7) are the filter functions for the ideal situation of an infinite number of projections, distributed uniformly in all orientations. They are referred to as ramp filters hereafter. In reconstructions from noisy data, they are usually modified by multiplication with some windowing function, such as a Hamming window or a sinc function, to reduce the noise in the reconstruction by attenuating higher frequency components (Jain, 1989). In this paper we shall only focus on the use of noise free projections.

The ramp filtering functions are based on the assumption that the projection directions are evenly and continuously distributed. This assumption can not hold in practice, since only a finite number of projections are available for reconstruction. In particular, when the projection directions have a severely nonuniform distribution, the reconstruction will be "elongated", or less constrained, along the clustering direction. The reconstruction of a model letter "F" is used to show this artifact (Figure 2-2).

Harauz and van Heel have proposed an "exact filtering" method for backprojection to individualize the Fourier filter for each projection according to its contribution in the orientation distribution, to improve the results of reconstructions (Harauz and van Heel, 1986). We show the effect of exact filtering below, and also that ADD filtering improves upon this effect still further.
Figure 2-2a

Figure 2-2b
Figure 2-2 Ramp-filtered backprojection produces reconstructions with "elongation" artifacts. (a) A 3D model of the letter "F"; two views are shown. (b) An example of projection orientations of protein images. It consists of 292 nonuniformly distributed orientations. Alpha and beta are two of the three components of Euler angles using the convention in Figure 1-3. In this case the circled region includes the majority of the orientations. (c) The construction of the model using ramp-filtered backprojection: 292 projections are made from the model with orientation set in (b), then ramp-filtered backprojection is performed. Two views of the reconstruction are shown. The stretching or elongation artifacts of the letter "F" are in the direction of the major clustering of the projection orientation distribution shown in (b).
2.3 ADD filtering for three-dimensional reconstruction

The Central Section Theorem is the theoretical basis for all filtered backprojection methods. In relation to three-dimensional reconstruction, it states that the Fourier transform of a two-dimensional projection is a section passing through the origin (central section) of the Fourier transform of the three-dimensional density distribution function giving rise to the projection.

The proposed numerical filtering method for backprojection, designated ADD filtering, like exact filtering is angular distribution dependent. It, too, is based on the Central Section Theorem. When two-dimensional projections are backprojected into a three-dimensional volume, the process is equivalent to inserting the two-dimensional Fourier transforms of those projections into the three-dimensional Fourier space. For a digital projection, its Fourier transform can be considered as a plane composed of a set of two-dimensional “Fourier grid points”. The three-dimensional Fourier transform of the reconstruction can be considered a set of three-dimensional “Fourier grid points”. When a number of Fourier planes are inserted into the three-dimensional Fourier space from different orientations, the grid points in the planes in general do not fall on the three-dimensional Fourier grid points but contribute their “weights” to the local three-dimensional volumes or voxels near the 3D grid points. The modification of the combined weights for each three-dimensional grid point to produce a uniform weighting in three dimensions then becomes the purpose of the filtering operation in the two-dimensional planes before backprojection. When the angular distribution of the planes is uniform and continuous, the three-dimensional Fourier grid points get
more weight in a well behaved pattern as the center of the Fourier space is approached. When the angular distribution of the planes is nonuniform, the distribution of weights to each three-dimensional Fourier grid point is more complex. In ADD filtering, the weights on the three-dimensional Fourier grid points are determined directly by the number of the two-dimensional Fourier grid points that fall into the 3D voxel associated with any given 3D grid point.

To produce the 2D ADD filters we divide the 3D Fourier space into voxels with dimensions $l/D$, where $D$ is the dimension of the reconstruction volume, with each voxel centered on a three-dimensional Fourier grid point. When a Fourier plane is inserted into Fourier space, the grid points in the plane will fall into various voxels. Every grid point in the plane that falls into a particular voxel is considered to contribute its entire weight to the three-dimensional grid point centering the voxel. If $k$ points fall into a voxel, the centering grid point is weighted by a factor of $k$. To make each three-dimensional grid point have a unit weight, all the grid points on the two-dimensional planes that fall into that voxel are corrected by a weighting factor $l/k$. This weighting factor then is the coefficient for the corresponding frequency in the 2D filter for that two-dimensional projection contributing to the voxel weight.

The filtering method can be described mathematically as follows.
Assume the volume for reconstruction is a cubic box with size $D$. Then the three-dimensional Fourier transform consists of grid points spaced by $L/D$. We can define each voxel centered at $(h, k, l)$ in Fourier space as

$$V_{hkl} = \{(x, y, z) | h - \frac{1}{2}D \leq x < h + \frac{1}{2}D, k - \frac{1}{2}D \leq y < k + \frac{1}{2}D, l - \frac{1}{2}D \leq z < l + \frac{1}{2}D\}$$

(2-3-1)

For a two-dimensional Fourier plane $m$, we denote its grid point at $(i, j)$ as $P^{(m)}_{ij}$. We define the set $W_{hkl}$ as:

$$W_{hkl} = \{ P^{(m)}_{ij} | \Omega \cap \{ P^{(m)}_{ij} \} \in V_{hkl} \} \quad (2-3-2)$$

Then the filter coefficient for $P^{(m)}_{ij}$ or for the Fourier component at frequency $(i, j)$ for image $m$ is:

$$H^{(m)}_{ij} = \frac{1}{Card\{W_k\}}$$

(2-3-3)

where $Card$ stands for the cardinality of a set.
Backprojection after ADD filtering will result in the frequency components in all projection directions being approximately uniformly weighted, especially when there are enough central slices sampling the Fourier space. This eliminates the “elongation” artifacts along the clustering directions. Nevertheless, an improvement in signal-to-noise ratio along with a correspondingly better resolution will still occur in directions perpendicular to angular clustering (Rose, 1973).

2.4 Comparison of ADD filtering with other filtering methods

To compare ADD filtering with other filtering methods, a three-dimensional model was used. This was a letter “F”, masked by a sphere (Figure 2-3). Within the volume of the “F”, the density distribution has a gradient perpendicular to the flat faces of the letter, which makes the volume asymmetrical along all directions. This model contains low frequency components, as well as high frequency components from sharp edges. Projections were made from the model for three sets of orientations (Figure 2-4). The first set of orientations consisted of 59 relatively uniformly distributed projection directions. The second set consisted of 400 uniformly distributed orientations. The third set consisted of 400 grossly nonuniformly distributed orientations. The projections were backprojected after being processed by ramp filtering, exact filtering, or ADD filtering. The reconstructions are compared with the original model by Fourier shell correlation (FSC) (Harauz and van Heel, 1986). Fourier shell correlation is defined as
Figure 2-3

Test model used to compare ramp filtering, exact filtering, and ADD filtering. A 3D model of the letter “F” masked by a sphere. The density within the volume of the letter “F” has a gradient along the direction perpendicular to the flat faces.
Figure 2-4a

Figure 2-4b
Figure 2-4 Orientation sets for comparing filtering methods. Alpha and beta are two components of Euler angles using the convention in Figure 1-3. (a) Orientation set 1. It consists of 59 orientations with relatively uniform distribution. (b) Orientation set 2. It consists of 400 uniformly distributed orientations. (c) Orientation set 3. It consists of 400 nonuniformly distributed orientations.
where $F_1$ and $F_2$ are the Fourier components for the reconstructed volume and the model respectively, and $\Delta R$ is the thickness of a shell of radius $R$ in Fourier space. For a perfect reconstruction $FSC(R) = 1$ for all spatial frequencies.

The results of the comparison between reconstructions are shown in Figure 2-5. For a uniform distribution of 59 orientations, exact filtering provides slightly better reconstruction than ramp filtering, while the ADD filtering generates reconstructions which show the best correlation to the original model. For 400 orientations either uniformly or nonuniformly distributed, ADD filtering also provides the best reconstruction among the three filtering methods.

Reconstructions produced by the three methods using orientation set 1 are shown in Figure 2-6. Due to the nonuniform distribution of projection directions, fairly severe artifacts are present in the reconstruction by ramp filtering. These artifacts are significantly reduced by exact filtering and are further decreased by ADD filtering. Moreover, high resolution information contained in sharp edges is better reconstructed by ADD filtering.
Figure 2-5a
Figure 2-5b
Figure 2-5 Comparison between ramp filtering, exact filtering, and ADD filtering. Reconstructions of the model letter “F” (shown in figure 2-3a) by the three methods are used for this comparison. (a) Orientation set 1 (sparse but uniform distribution); (b) orientation set 2 (400 orientations, uniform distribution); (c) orientation set 3 (400 nonuniform orientations). fc is the Nyquist frequency.
Figure 2–6 Reconstructions of the model letter "F" using orientation set 1. (a) ramp-filtered reconstruction; (b) exact filtered reconstruction; (c) ADD-filtered reconstruction. The same threshold, corresponding to the whole volume of the "F" in the original model, is applied to the three reconstructions.
Another model, whose frequency spectrum more closely resembled that of a real protein density distribution, was also used for the comparison. This was a current working model reconstruction for an unknown protein structure, mannosidase II. The same experiment was performed using this model, and it showed similar results by Fourier Shell Correlation as obtained for the model letter “F”. Qualitatively similar artifacts were also seen in the reconstruction. However, the effect had a less visual impact due to the already complex underlying protein structure (data not shown).

2.5 Discussion and conclusion

The results show that ADD-filtered backprojection produces a better correlation with the original models than the other backprojection methods. As the number of projections increases, all filtering methods produce superior reconstructions, with the ADD approach each time providing reconstructions with the highest correlations. As expected for nonuniform orientation distributions, both orientation-dependent methods produce improved results compared to ramp filtering, but they also do so for relative uniform distributions.

The ramp filtering method employs the filtering function derived for the case of infinite number of projections, and it ignores the nonuniformity of the orientation distribution. For these reasons the ramp filter performance is worse than ADD filtering and "exact" filtering, and it is not surprising the "elongation" artifacts in ramp-filtered reconstructions are severe for nonuniform distribution of the orientations. In the "exact filtering" method, the central sections in Fourier space are considered as slices with a certain thickness. Each Fourier grid point in a central section is effected
by the distance from another central section, if it is within the thickness range of that section. Adaptively taking care of the orientation distribution, exact filtering produces better reconstructions than ramp filtering. The shortcoming of exact filtering resides in two aspects. First, this method concentrates on normalizing the weight for each Fourier grid point in the **two-dimensional** central sections, while the weights for the Fourier components in the resultant **three-dimensional** transform of the reconstruction are not necessarily normalized. Secondly, the weighting effects from neighbouring grid points within the same central section are not considered in estimating the overall weights. In contrast ADD filtering is designed to normalize the weights for the Fourier grid points in the **three-dimensional** Fourier space, rather than the Fourier grid points in the central sections. Moreover, the weight for each **three-dimensional** Fourier grid point is estimated from all the neighbouring grid points in the central sections that fall into the voxel centered at the three-dimensional grid point, which could include more than one neighbouring points from the same central section.

A current further advantage of ADD filtering is its implementation by a fast computational algorithm. Initially each “Fourier voxel” is allocated memory. When a “Fourier grid point” in a central section falls into a voxel, a pointer to the voxel is established for that Fourier grid point. The voxel keeps the count of the number of two-dimensional Fourier grid points falling into it. After looping over all the central sections, each Fourier grid point in the central sections is still linked to its corresponding Fourier voxel that then contains the appropriate weight for the 2D grid point. The elegance of this approach is that all the Fourier planes, or central sections, are examined
only once. So the computational complexity is $O(N)$, in comparison with $O(N^2)$ for exact filtering, where $N$ is the number of projections. Therefore, where the speed for filtering is of importance, the ADD filtering approach is advantageous.

However, it is worth noting that the determination of the ADD filter for each image is slightly dependent on the choice of the orientation of the 3D Fourier coordinate system. When the 3D Fourier grid is built based on an orthogonal coordinate system at a slightly different orientation, a few grid points in the central sections near the sides of 3D Fourier voxels may fall into a neighbouring voxel and thus result in slightly different ADD filters. Tests have been performed using the model data above with a coordinate system at a different orientation, and the difference in the resultant reconstruction is insignificant at the current level of analysis (data not shown). This is due to the fact that most ADD filter coefficients stay the same even if a different coordinate system is used.

In the discussion above, the filtering methods are studied in absence of noise. When noise is present in images, the performance of all the three filtering algorithms will be degraded, especially for the high frequency components in the reconstructions. Given the loss of such components, the reconstruction can be improved by modifying the filtering functions, by multiplying them with a windowing function, such as a Hamming window or a sinc function. The performance of different filtering methods in presence of noise can be further investigated. This can be tested by adding noise to the simulated images, and specific window functions during reconstruction.
In conclusion, the ADD filtered backprojection is a fast and accurate reconstruction method that improves upon the current two most commonly used filtered backprojection methods.
Chapter 3  Sinogram Correlation Function by Phase Difference Analysis

3.1 Introduction

Seminal work by Radon has provided the theorems describing the recovery of an $n$-dimensional object from its $(n-1)$-dimensional projections (Radon, 1917). These theorems, though strictly applicable only to the reconstruction of continuous object functions from a projection set containing an infinite number of projections, have provided the basis for methods that recover an approximation of the object from discretely sampled projections.

Reconstruction of the 3D electron density map of biological macromolecules from their 2D projections obtained by electron microscopy is becoming a more and more important technique in solving molecular structures. This reconstruction requires knowledge of relative projection orientations, and has promoted the development of the angular reconstitution technique to reconstruct randomly orientated molecules (van Heel, 1987, Farrow and Ottensmeyer, 1992, 1993). Angular reconstitution, in principle, is based on the Central Section Theorem, which states that a 2D Fourier transform of a 3D object is a central section of the 3D Fourier transform of that object, and therefore any two projections of the same object will share a common line in 3D Fourier space. The search for this common line forms the core of this angular reconstitution. Since
a central line in Fourier space corresponds to a line projection in real space, the search is normally implemented in real space by calculating the so-called sinogram correlation function (SCF). The sinogram is a collection of line projections of an image from different orientations, and is in fact a discrete implementation of 2D Radon transform. The sinogram correlation function is a 2D function that determines the similarity, or correlation, between any possible two line projections, one each from two respective sinograms. The peak in the SCF corresponds to the common line direction in the local coordinate system of the two projections.

Usually the cross correlation coefficient (CCC) is adopted as the measure of similarity between two line projections. Since line projections of an object, unless the object is very non-symmetrical, generally have similar overall shapes and only differ in small details, the SCFs calculated by CCC tends to give high correlation values for most pairs of line projections. This results in broad peaks in the SCFs and makes it difficult to determine the locations of the peaks. Moreover, electron micrographs are generally noisy and the molecules imaged by microscopes suffer from radiation damage, which may cause the peaks at the correct locations to have correlation values lower than the neighbouring peaks in the SCFs. This error in the determination of common lines will result in an error in the determination of projection directions for molecular images.

To minimize the effect of errors in the determination of common line directions, quaternion mathematics has been introduced to assist angular reconstitution in processing the common line data, after they are determined from the SCFs. It has been shown to be a fast and elegant approach
to find the global optimal solution to the image orientations when considering more than three images simultaneously (Horn, 1987, Harauz, 1990, Farrow and Ottensmeyer, 1992, 1993). To reduce errors in the determination of image orientations further, efforts have been made in selecting multiple peaks in the SCFs (Ottensmeyer et al., 1997). Instead of using only the highest peak in each SCF as the common line determinant, several highest peaks are chosen simultaneously as candidates. In the currently used implementation all combinations of candidate peaks among seven SCFs are evaluated simultaneously. The combination that gives the smallest consistency error among images is chosen as the solutions for the common lines.

These methods have provided greater capacity of handling common line data, and therefore better accuracy in angular reconstitution. This accuracy, however, can not be determined quantitatively for images of protein particles with unknown structure, and it can only be estimated by simulation (Farrow and Ottensmeyer, 1992). When the quaternion assisted method was first applied in angular reconstitution it was able to determine the projection orientations at an error of less than 10 degrees for simulated projections with a signal-to-noise ratio of 5. This signal-to-noise ratio resulted in an average error of 32 degrees in the calculation of common lines. Clearly, more accurate common line data are desired for further improvement in the orientation calculation.

In this chapter, we present a new measure of similarity between two line projections based on the phases in the Fourier transform of the line projections. Applied in the SCF, this method, called
phase difference analysis, or PDA, significantly enhances peaks of high similarity in comparison to the use of the cross correlation coefficient.

3.2 The SCF using the cross correlation coefficient

For two discrete line projections $f(k)$ and $g(k), k = 1, 2, \ldots, N$, the cross correlation coefficient is defined as:

$$r(f, g) = \frac{\sum_{k=1}^{N} (f(k) - \bar{f}) \sum_{k=1}^{N} (g(k) - \bar{g})}{\sqrt{\sum_{k=1}^{N} (f(k) - \bar{f})^2 \sum_{k=1}^{N} (g(k) - \bar{g})^2}} \quad (3-1-1)$$

So the SCF as a function of the CCC can be written as:

$$SCF_{ccc}(\theta_1, \theta_2) = r(f_{\theta_1}, g_{\theta_2}) \quad (3-1-2)$$

To show the behavior of this form of SCF, two projections from different orientations were produced from the three-dimensional model of the letter “F” as shown in Figure 2-2b (page 26). The sinograms of these two projections were calculated and the SCF for these sinograms was determined using Equation (3-1-1) and (3-1-2) (Figure 3-1).
Figure 3-1  Sinogram correlation function calculated using the cross correlation coefficient. On the left bottom and right top corners are two projections of the letter model "F" in Figure 2-2b. Adjacent to the two projections are their sinograms. The square region in the middle is the sinogram correlation function, calculated from the two projections. The vertical direction in the sinogram correlation function represents the rotation of the first projection (bottom left) from 0 (top) to 360 degrees (bottom). The horizontal direction represents the rotation of the second projection (top right) from 0 (left) to 360 degrees (right). Bright spots represent good correlations.
A direct observation of Figure 3-1 suggests that most pairs of \((\theta_1, \theta_2)\) give high correlation values and the two global maxima (one symmetrically related to the other by 180 degrees in \(\theta_1\) and \(\theta_2\)) are quite broad and flat-topped. The breadth of these peaks can result in rather imprecise determination of the common line angles, especially in the situation where noise in the images can enhance or suppress the correlation values at any point in the SCF.

### 3.3 SCF using phase difference analysis

To make the peaks in the SCFs more distinguishable, we present a different measure of similarity, or correlation, between line projections, based on the phase differences in the Fourier transforms of the line projections. This so-called phase difference analysis (PDA) will be shown capable of defining peaks of greatest similarity more sharply in the SCF.

Assume the Fourier transform of \(f(k)\) and \(g(k)\) are \(F(h)\) and \(G(h)\), where

\[ h = -N/2 + 1, \ldots, N/2, \]

respectively (assuming \(N\) is a power of 2 for convenience). Expanded in terms of amplitude functions \(A(h)\) and phase function \(\phi(h)\), \(F(h)\) and \(G(h)\) can be written as:

\[
F(h) = A_F(h)e^{i\phi_F(h)} \quad (3-2-1)
\]

\[
G(h) = A_G(h)e^{i\phi_G(h)} \quad (3-2-2)
\]
Let

\[ \Delta \phi(h) = \frac{\phi_F(h) - \phi_G(h)}{h} \]  \hspace{1cm} (3-2-3)

Then \( \Delta \phi(h) \) represents the “normalized” phase difference between the two signals at each frequency. The negative frequencies are omitted due to the hermitianality of real signals. When \( f(k) \) and \( g(k) \) are two identical functions except for a translation, i.e.,

\[ f(k) = g(k + \Delta k) \]  \hspace{1cm} (3-2-4)

then,

\[ F(h) = G(h) e^{i 2 \pi h \Delta k} \]  \hspace{1cm} (3-2-5)

and,

\[ \Delta \phi(h) = 2\pi \Delta k / N \]  \hspace{1cm} (3-2-6)

Thus \( \Delta \phi(h) \) is a constant, independent of frequency \( h \), when \( f(k) \) and \( g(k) \) are identical; in particular, when there is no shift between the two functions, \( \Delta \phi(h) \) will be zero. In general, for any two functions, the standard deviation of all the \( \Delta \phi(h) \), will be small for a high correlation and
zero for a perfect correlation. We can therefore define this standard deviation as a measure of correlation by PDA:

\[ \Psi(f, g) = SD\{\Delta \Phi_{fg}(h) | h = 0, 1, \ldots, M \} \]  \hspace{1cm} (3-2-7)

where \( M \) is a cutoff frequency index.

The corresponding SCF is therefore defined as:

\[ SCF_{PDA}(\theta_1, \theta_2) = \Psi(f_{\theta_1}, g_{\theta_2}) \]  \hspace{1cm} (3-2-8)

It is clear in this correlation function that, instead of searching for the maxima, the minima determine the best correlated pair(s) of line projections.

In Figure 3-2, the same two sinograms (those in Figure 3-1) have been used to calculate the PDA-based SCF and the SCF via the cross correlation coefficient. For a direct comparison between the two SCFs, the contrast in the PDA-SCF map is reversed and scaled linearly to the same range as the CCC-SCF map (Figure 3-2a). The two SCFs are thresholded by increasing cutoffs to provide a visual representation of the search for peaks (Figure 3-2b, c, d). This comparison shows that in the PDA-based SCF fewer pairs of line projections result in high correlations and that the peaks containing the maxima are more compact, defined by steeper gradients around them.
Figure 3-2 a. The CCC-based SCF (left) and the PDA-based SCF (right, contrast reversed) of the two "F" projections in Figure 3-1.

Figure 3-2 b. The CCC-based SCF (left) and the PDA-based SCF (right, contrast reversed) thresholded at 50% of their maxima.
Figure 3–2 c. The CCC–based SCF (left) and the PDA–based SCF (right, contrast reversed) thresholded at 75% of their maxima.

Figure 3–2 d. The CCC–based SCF (left) and the PDA–based SCF (right, contrast reversed) thresholded at 90% of their maxima.
3.4 Discussion

Advantages of PDA are its property in SCF peak enhancement, its flexibility in phase manipulation in different frequency ranges, and its robustness to shift errors. These are discussed below, as are considerations of phase wraparound and computational speed.

3.4.1 Peak enhancement

As shown above, clearly the SCF calculated by phase difference analysis gives more distinctive peaks. However the different patterns in the SCF maps in Figure 3-2 suggest that phase difference analysis is a different measure of similarity, not simply a type of transform of the cross correlation coefficient.

The difference in pattern can be explained as follows. In Fourier space, both amplitude function and phase function are required to define a function. However for most line projections of an object the amplitude functions are very similar to each other, big at low frequencies and small at high frequencies. In particular, this similarity is more pronounced in the low frequency range. When comparing two line projections in real space using the cross correlation coefficient, the similarity in the low frequency components results in high correlation values for most pairs of line projections. This causes the broad and flat-topped peaks in the SCFs. Given the similarities at low frequencies, the phase function at higher frequencies is then more characteristic for each line
projections. So when using phase difference analysis, the low frequency similarity can be suppressed in relation to differences in high frequency components. This causes the suppression of the less characterizing peaks.

Since phase information is not a sufficient but only a necessary condition to define a function, there is a chance of creating false positive peaks in the SCF calculated using PDA. However, our experiments on model data have shown that the global maxima in the SCF determined by PDA are always at the same positions as those determined by CCC, but are more distinctively defined. Nevertheless, a consideration of multiple peaks in the SCF as possible candidates for common line angles has been implemented in the determination of best correlated pair of line projections (Ottensmeyer et al., 1997). The concern of false positivity can be virtually laid to rest.

3.4.2 Flexibility in the manipulating phases at different frequencies

A particular advantage of this phase method is the flexibility in the choice of the range of Fourier coefficients that can be selected for use in the analysis. For example, phases for the Fourier indices at high frequencies can be omitted to reduce the influence of noise on the correlation. In addition the phases at low frequencies can be omitted or less weighted when the molecule is known to have indistinguishable overall shapes from different orientations.
3.4.3 Robustness to shift error

As a routine procedure prior to the calculation of cross correlation based SCF, shifting the center of mass of the molecule to the image center is required. Only when this alignment is carried out can the peak(s) in a SCF indicate the best correlated pair. However since the molecular boundary is always hard to determine accurately, the determination of center of mass is susceptible to error, and this shift error will decrease the cross correlation coefficient. In the SCF calculated by PDA, the correlation is invariant to real space translation, and the shift error can be determined by Equation (3-2-6).

3.4.4 Phase wraparound

Since phases can only take values in one period of $2\pi$, the phase differences calculated in PDA can only take values within a range of $4\pi$ (or +/-2$\pi$). However when the shift between two line projections is big enough, the corresponding phase difference will be beyond the range of $4\pi$. This problem of phase wraparound has to be carefully corrected. There can be a variety of approaches to deal with this problem. One way is to adjust each phase difference by an integer number times of $2\pi$ so that the plot of phase difference versus frequency correlates best to a line passing through the origin. To minimize the computation, we simplify this procedure as follows: to determine the phase difference at frequency $h$, adjust phase difference $\Delta\phi(h)$ by an integer times of $2\pi$, so that $\Delta\phi(h)$ and the previously determined phase difference $\Delta\phi(h-1)$ have the smallest
difference. This implementation is based on two assumptions: one is that the shift error between line projections is smaller than half the total length of the line projections, the other is that the low frequency components are always less noisy than the high frequency components. This approach solves the phase wraparound.

3.4.5 Computational speed

The SCF by PDA can be combined with a fast sinogram algorithm (Lanzavecchia et al., 1996) to speed up the process of SCF calculation.

The fast sinogram algorithm (see the flow chart in Figure 3-3a) takes the 2D Fourier transform of an image, selects a central section of the transform, and then takes an inverse 1D Fourier transform of that section to obtain a line projection in real space. The sinogram is formed by calculating the line projections at all orientations. Usually obtaining the line projections is the starting point for the calculation of SCF. Normally the fast sinogram calculation and the calculation of SCF are performed in a sequence shown in Figure 3-3b. When the phase difference

\[ \text{Based on this assumption, the phase difference at the fundamental frequency is always considered to be less than } \pi. \]
Figure 3–3 a. A fast sinogram algorithm
Figure 3–3 b. Calculation of a point in the SCF by the fast sinogram combined with a real space correlation method.
analysis is applied, the inverse 1D Fourier transform in the fast sinogram can be omitted and the central section in the 2D Fourier transform of one projection can be used directly with another central section in the 2D Fourier transform of another projection to calculate a point in the sinogram correlation function of the two projections (see Figure 3-3c). This will increase the speed of calculating the SCFs, and the time used in calculating one point in the SCFs is estimated to be half of the time spent using the previous method.

3.5 Conclusion

We have shown that phase difference analysis can be used to measure similarity between functions. Compared to the cross correlation coefficient in defining the sinogram correlation function for electron microscopic image reconstruction, it enhances peaks in the sinogram correlation functions. It is robust to shift errors, and it speeds up the process of angular reconstitution when combined with a fast sinogram algorithm. It also has the advantages of flexibility in selecting ranges of frequencies for effects such as noise reduction or suppression of non-informative gross structural similarities. Given these advantages of the phase difference analysis method, further investigation is still required for a quantitative assessment of this approach in handling noisy data compared in comparison to the previous cross correlation method.
Figure 3–3 c. Calculation of a point in the SCF by the fast sinogram combined with PDA.
Chapter 4  Summary and Future Directions

4.1 Summary

In the determination of protein three-dimensional structures by electron microscopy, the calculation of projection orientations and the reconstruction of 3D structures from the two-dimensional projections are critical procedures in computation. The errors in these procedures limit the quality, or achievable resolution, of determined structures from a given set of images. To form a 3D reconstruction from 2D density projections, filtered backprojection is the most commonly used method, which includes a filtering function applied to each projection and a backprojection operation. To calculate the orientations of projections, the sinogram correlation function has to be calculated to determine the common lines between pairs of projections. In this project, an angular distribution dependent (ADD) filtering algorithm is developed for the filtered backprojection method to improve the accuracy of three-dimensional reconstruction; and a phase difference analysis approach (PDA) is utilized in the calculation of sinogram correlation functions.

Compared with the previously established filtering methods, i.e., ramp filtering and exact filtering, ADD filtering provides more accurate reconstructions. Analysis with the Fourier
two filtering methods in the whole frequency range for both uniformly and nonuniformly distributed orientations. Although, like the “exact” filtering method, ADD filtering individualizes the filtering functions for projections according to their orientation distribution, the computational complexity of ADD filtering is significantly lower than that of exact filtering (by an order of $N$, where $N$ is the number of projections). This makes ADD filtering not only more accurate but also faster than exact filtering.

The introduction of the PDA approach as a measure of correlation for the sinogram correlation functions results in more distinguishable peaks compared with the original cross correlation approach. Eliminating a procedure of transformation from Fourier space to real space, it speeds up the orientation determination process when combined with a fast sinogram algorithm (Lanzavecchia, et al., 1996). It is robust to shift errors and has the flexibility in selecting ranges of frequency components for noise reduction or suppressing non-informative gross structural similarity.

Taking advantage of the improvements in the two aspects above, protein structure determination by electron microscopy is expected to be more accurate and at higher resolution.
4.2 Future directions

4.2.1 Introduction

Application of electron microscopy to protein structure determination is a new area in structural biology. There are great potentials in further improving the image analysis and image reconstruction techniques in this field. The following sections cover two future directions investigated during this project.

In 4.2.2 a method of automated molecular segmentation in electron micrographs using a "shape factor" calculation is described. Some results of preliminary investigation are presented. In 4.2.3 a novel 3D reconstruction method, the "finer voxel" method, is proposed for STEM images. A theoretical discussion is presented to show protein structure at atomic resolution is potentially obtainable using this method.
4.2.2 Automated molecular segmentation in electron micrographs using a shape factor calculation

4.2.2.1 Introduction

To determine the molecular orientations in electron micrographs by angular reconstitution techniques, it is necessary to process the images before the orientations are calculated. This includes filtering images for noise reduction, segmenting molecular density from the surrounding carbon background, normalizing the molecular density in relation to the carbon background, and subtracting the average carbon background density level from the molecular density distribution.

Accurate molecular segmentation is one of the most important steps in the preprocessing before angular reconstitution because the determination of orientations requires that all the images be of the same molecular structure except for the difference in projection directions. Errors in the segmentation, such as the inclusion of carbon background surrounding the molecular image or the exclusion of part of the density of molecules, will alter the line projections of images and therefore result in errors in calculating particle orientations.

This type of segmentation problem is, in fact, the same as edge detection. In principle many methods in edge detection and image segmentation can be used for this purpose, such as density thresholding, the gradient operator, or the Laplacian operator (Jain, 1989, Marr, 1982). However,
since electron micrographs of molecular images are generally noisy and of low contrast, the molecular boundaries are not well defined by those methods.

Currently a “grow” algorithm is used to segment a molecule from the background in the image (Jain, 1989). This algorithm starts from an image pixel that is known to be within the molecule. It looks for a connecting pixel with the maximum density and includes it in the boundary. Then it looks for the next brightest pixel that connects to the area currently determined to be within the molecule. After gradually including more and more pixels, the molecule “grows”. The most critical parameter for this algorithm is the constraint for which the “growing” stops. A so called etching factor is defined for this purpose:

\[
Etching \text{ factor} = \frac{\text{Total density of the grown area}}{\text{Total density of the whole image}}
\]

A global etching factor is chosen heuristically from a small subset of the images. This etching factor is then applied as the constraint parameter to each image to stop the “growth” process.

The heuristic single choice of the etching factor raises the problem of inaccuracy in molecular segmentation, since small local variations in carbon support thickness or non-uniform mass loss during electron irradiation can then alter the boundary of individual molecular images. To compensate for such effects, it is necessary to individualize the etching factor for each image so that the boundary in each image can be more accurately defined.
For the derivation of individualized etching factor criteria, a “shape factor” was defined and employed as a shape descriptor to investigate a new method for molecular segmentation. Combined with the grow algorithm, this method potentially takes advantage of the compact nature of protein structure and will not be affected by the low contrast and blurry boundaries in the micrographs. Preliminary studies are presented below.

4.2.2.2 Shape factor and compactness-etching plot

The shape factor is defined as:

\[ \text{Shape factor} = \frac{\text{perimeter of a region}^2}{\text{area of the region}} \]

It is a non-dimensional quantity, reflecting the compactness of a region, or the complexity of the boundary of a region. Figure 4-1 shows the shape factor of several geometries. A circle has the smallest shape factor, equal to \(4\pi\). The more complex or the less compact a shape is, the bigger is its shape factor.

By combining the grow algorithm and the shape factor, a compactness-etching plot can be obtained to describe the relationship between the shape factor of the “grown” molecule and the change of the etching factor during the “growth” process.
<table>
<thead>
<tr>
<th>Geometry</th>
<th>Shape Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image.png" alt="Circle" /></td>
<td>12.56</td>
</tr>
<tr>
<td><img src="image.png" alt="Square" /></td>
<td>16</td>
</tr>
<tr>
<td><img src="image.png" alt="Triangle" /></td>
<td>20.79</td>
</tr>
<tr>
<td><img src="image.png" alt="Complex Shape" /></td>
<td>35</td>
</tr>
<tr>
<td><img src="image.png" alt="Complex Shape" /></td>
<td>70</td>
</tr>
</tbody>
</table>

Figure 4-1 The shape factors for several geometries
Theoretically we can predict the general trend of such a compactness-etching plot. When the etching factor is 1, the shape of the grown region will be the entire square image, with a shape factor of 16. As the etching factor decreases gradually, the boundary of the grown region will shrink into the carbon film. Due to the grainy variation in the density of carbon film, the boundary becomes more complicated, or the grown shape is less compact. So the shape factor gradually increases. There will be a maximum value for the shape factor as the boundary changes on the carbon film. When the etching factor is such that the grown region just covers the molecule, the shape factor should drop to a small value. This is because proteins are in general compact and presumably pack themselves to a lowest energy state where the hydrophilic surface is used most efficiently. As the etching factor continues decreasing, the grown boundary starts cutting into the molecule. Due to the internal density variation in the molecule, the grown boundary becomes complex again, with a consequent shape factor increase. The factor will eventually decrease again, when only a very small area is enclosed in the grown region. Based on this analysis, we can estimate the hypothetical compactness-etching plot (Figure 4-2).

Experiments were carried out to test this hypothesis, and 128 images of protein Enzyme IIA were extracted from electron micrographs for this purpose. A series of grown contours enclosing the molecule were made for each image with the etching factor changing in steps of 0.01 (see Figure 4-3 for an example). The etching factor that corresponds to the molecular boundary in each series of grown images was selected by visual inspection. Then a compactness-etching plot was
Figure 4-2 Hypothetical compactness-etching plot
Figure 4–3 A series of "grown" images for an Enzyme IIA image. The etching factor varies from 0.99 (left top corner) to 0.01 (right bottom corner), with changing step of 0.01. The contrast is reversed. The circled image corresponds to the visually determined molecular boundary.
generated for each image (see Figure 4-4 for an example). Finally the hypothesis was tested by comparing the intermediate major minima of the plot with the visually determined boundary.

As the results of the experiments, for 109 images (85%) the visually determined boundary corresponds to a major intermediate minimum or "valley" in the compactness-etching plot. For the other 19 (15%) images, there is no direct correspondence between the visually determined boundaries and the compactness-etching plot. Among the compactness-etching plots of the 109 images, 39 (36%) have only one valley, 37 (34%) have two valleys, and 33 (30%) have three or four.

4.2.2.3 Conclusion

The preliminary study presented above suggests that, it is promising to use the shape factor in molecular segmentation in conjunction with a changing etching factor. Searching the compactness-etching plot for major minima will indicate the candidates for the correct etching factor. Since many compactness-etching plots currently do not supply a unique solution, a secondary criterion must be sought to select the correct minima among a small number of choices. The advantage of employing the shape factor is the individualization that is possible to automate in the search for molecular boundaries in electron micrographs.
Figure 4–4 An example of the compactness–etching plot. The cirled valley in the plot agrees with the image grown to the visually determined molecular boundary (shown in Figure 4–3).
4.2.3 Determination of protein structure at atomic resolution with STEM: Three-dimensional reconstruction by a finer voxel method

4.2.3.1 Introduction

In the practice of applying electron microscopy to structural biology, researchers have been challenged with the difficulties of solving protein structures at atomic resolution. The difficulties are due to the spherical aberration and chromatic aberration of the imaging system, as well as the fragility of protein specimens. In particular, this fragility prevents protein particles in the specimen from being exposed to high dose electron beams that would otherwise provide a sufficient signal-to-noise ratio for recognizing high resolution details. When a 2D crystalline protein array can be prepared, the high dose of electrons can be spread across those regularly arranged protein particles giving both images and diffraction patterns at a low dose per unit cell in the crystal. It is then possible to perform electron crystallography to solve structures at higher resolution, but atomic resolution is still a dream of electron microscopists (Kühlbrandt et al., 1994, Savage et al., 1996, Baker and Johnson, 1996, Kimura, 1997, Chiu and Schmid, 1997).

A scanning transmission electron microscope (STEM), operating in a dark field mode, is particularly efficient at utilizing electrons that have interacted with a specimen. It thus provides images of protein specimens with a reasonable signal-to-noise ratio on individual single protein molecules even for very low doses. However, even these doses are still too high for a protein to
preserve its details at atomic resolution, necessitating special approaches to reduce the dose further while still recovering the high resolution structural information. These approaches may be either instrumental or computational. On the instrumental side, detection of a larger portion of interacting electrons should result in a higher signal-to-noise ratio in images for the same dose of electrons. One of the most recent examples is to utilize both elastic and inelastic signals in the imaging system (Ottensmeyer et al., 1997). On the computational aspect, three-dimensional reconstruction techniques and the concomitant 3D signal averaging should allow high resolution structures to be retrieved from images acquired at low doses of electrons. Here a novel 3D reconstruction approach, designated as “finer voxel” method, is proposed. The theoretical discussion in this section will show that using this approach higher resolution structures can be reconstructed from low resolution images, i.e., images obtained at low magnification, or, for an STEM, obtained with big scanning intervals. Since a lower electron dose is applied to a specimen imaged at a lower magnification, this approach will realize the idea of irradiating protein specimens with a lower dose of electrons and achieving the same or higher resolution in the reconstruction.

Usually for any 3D reconstruction approach, the resolution of a reconstruction is thought to be no better than the imaging resolution, which is determined by the sampling (scanning) interval of the imaging system (van Heel and Harauz, 1986). However it is easy to show that using this “finer voxel” method higher resolution can be obtained in three-dimensional reconstruction than that in the original images and that even atomic resolution is possible to achieve.
The method is first described in 4.2.3.2, then the underlying principles of the method will be elucidated in details in 4.2.3.3. Following that, in 4.2.3.4 discussions will be given on some requirements for this method to achieve a reconstruction at atomic resolution, such as the size of molecules, the number of images, and the tolerable orientation errors.

4.2.3.2 Finer voxel method

Assume $M$ images of a molecule along different known orientations are obtained from a STEM and that the beam profile function of the STEM is known. The dimension of the images is $N \times N$, and the sampling interval is $d$. From the Nyquist's sampling theorem (Bracewell, 1965), the best resolution present in each individual image is $2d$. Now we will use the finer voxel method to determine the 3D structure of the molecule at an $m$ times better resolution, i.e., a resolution of $2d/m$, where $m$ is an integer greater than 1.

Step 1. Subtract the carbon background intensity for each image, determine the boundaries of molecules, and set the region outside the boundary to be zero in each image. The determination of molecular boundary should be conservative, i.e., the boundary has to enclose the whole molecule.

Step 2. Build a cubic box of $Nd \times Nd \times Nd$ to hold the reconstructed volume and divide this box into $(m \times N) \times (m \times N) \times (m \times N)$ cubic voxels. Consequently the dimension of each voxel is $(d/m) \times (d/m) \times (d/m)$ (Figure 4-5). The integrated density in voxel $(h, k, l)$ is denoted as $y_{hkl}$, where
Figure 4–5 Subdivision of voxels in the finer voxel method. Images are acquired at a scanning (sampling) interval of $d$. The reconstruction volume is divided into voxels of dimension $d/m$, to obtain an $m$ times higher resolution than that in the images.
\[ l \leq h, k, l \leq m \times N. \] Since the molecule can be put into this box, we have changed the reconstruction problem into the determination of the total density in each voxel, i.e., determining \( y_{hkl}. \) Because the dimension of voxels in the reconstruction is smaller than the dimension of pixels in the images, we call this reconstruction method “finer voxel” method.

Step 3. Calculate the center of mass for each image and shift the image so that its center of mass is aligned to the center of the box.

Step 4. Form a system of linear equations and solve the equations to determine the density in each voxel, \( y_{hkl}. \) This will be explained in details as follows.

Let \( q \) index the number of images. When image \( q \) is obtained, the molecule in the box has been rotated by an Euler angle of \( \Omega_q \) about the center of box, where \( l \leq q \leq M \) and \( M \) is the total number of images. Now consider the beam profile function of the STEM is a 2D point spread function \( p(x, y) \), where the \( x-y \) plane maps to each image plane, or alternatively, the projection direction is always the same as the normal of the image plane. The beam profile function convolves with the 3D density distribution function of the molecule in the \( x-y \) plane, integrated over the \( z \) direction, or projection direction, giving rise to a density of a pixel in an image. The density of pixel \((i, j)\) in image \( q \) is denoted as \( I^{(q)}_{ij} \), where \( l \leq i, j \leq N. \)

Before the rotation of \( \Omega_q \), the center of voxel \((h,k,l)\) is at \((hd/m, kd/m, ld/m)\). In image \( q \) the center
of pixel \((i,j)\) is at \((x^{(a)}_i, y^{(a)}_j)\). We can define a 2D square interval \(W_{hkl}\) for voxel \((h, k, l)\), where

\[
W_{hkl} = \{(x, y, z) \mid \text{condition}\}
\]

where

\[
W_{hkl} = \{(x, y, z) \mid \text{condition}\} = \{(x, y, z) \mid \text{condition}\}
\]

(4-1)

i.e. \(W_{hkl}\) is a square interval bounded by voxel \((h, k, l)\), passes through its center, and is parallel to the \(x\)-\(y\) plane.

After the box is rotated by \(\Omega_q\), the center of voxel \((h, k, l)\) and the four corners of the \(W_{hkl}\) will be rotated to new positions. The projection of this rotated square region into the \(x\)-\(y\) plane is a parallelogram and the projection of the center of voxel \((h, k, l)\) into the \(x\)-\(y\) plane is in the geometric center of this parallelogram. If we denote this parallelogram interval as \(W'_{hkl}\), we can define the operation from \(W_{hkl}\) to \(W'_{hkl}\) by a symbol \(\xi\):

\[
W'_{hkl} = \xi_q \left\{ W_{hkl} \right\}
\]

(4-2)

where \(\xi_q\) refers to rotating a square interval parallel to the \(x\)-\(y\) plane by the Euler angle \(\Omega_q\) and taking its projection region in the \(x\)-\(y\) plane (Figure 4-6).
Figure 4–6 The operation \( \zeta \). \( W \) is a square interval bounded by a voxel, passing through the center of the voxel, and parallel to the \( x-y \) plane. The \( \zeta \) operation rotates \( W \) by an Euler angle and takes its projection in the \( x-y \) plane. "x" marks the center of the voxel.
Then for image \( q \), the intensity of its pixel \((i,j)\), \( I_{ij}^{(q)} \) can be considered as the sum of the integration of the beam profile function over each \( W_{hkl} \) weighted by the total density in voxel \((h,k,l)\), or,

\[
I_{ij}^{(q)} = \sum_{hkl} y_{hkl} \left( \int_{\Omega_{hkl}} p(x - x_{ij}^{(q)}, y - y_{ij}^{(q)})\,dx\,dy \right)
\]  

(4-3)

For each pixel in each image we can write such a linear equation, given the beam profile function \( p(x,y) \) is known. The total density in each voxel can be calculated by solving these equations. There are \((m \times N)^3\) variables \((y_{hkl})\) in this system of linear equations. Thus a minimum of \((m \times N)^3\) equations are required. In every image we can derive \( N^2 \) equations, so the minimum number of images \( M \) is \((m^3 \times N)\). When we have more images than required, solving this system of equations evolves to a minimization problem.

4.2.3.3 Underlying principles

To illustrate the underlying principles of the finer voxel method, two questions have to be answered:

(1) where does the “extra” information, i.e., higher resolution, come from?

(2) will the frequency band limit of the beam profile function become a limit of resolution in the reconstruction?
These two questions will be answered in 4.2.3.3.1 and 4.2.3.3.2 respectively using one-dimensional or two-dimensional analogues. Then the principles of the finer voxel method will be further elucidated in 4.2.3.3.3.

4.2.3.3.1 Superposition of multiple measurements

Superposition of multiple measurements can result in higher resolution than that in each individual measurement. This principle will be explained using the following two-dimensional analogues.

Let us start by looking at low dose TEM (transmission electron microscope) images of molecules deposited in a specimen with the same orientation. These images are sparsely represented with randomly located quanta (electrons) (Figure 4-7). The resolution in the images is given by the average distance between the sampling points and the Nyquist resolution limit, or,

\[ R = 2d \]

where \( R \) stands for the resolution and \( d \) stands for the averaged distance between sampling points.

In this case, the resolution can be increased by increasing the dose in each individual image to reduce the average distance between sampling points, or alternatively by superposing sufficient individual images with sparsely irradiated samples (low dose) after appropriate alignment of their center of mass.
Low dose TEM images are sparsely represented with randomly located electrons, as indicated in the four molecular images in panel (a). Each black dot represents the location of an electron. The superposition of the four images produces the image in (b). This is equivalent to imaging the molecule with a higher dose of electrons. The resolution of an image can be estimated by twice the averaged distance between electrons. An image produced by superposition, like that in (b), has a higher resolution than the original images.
In the STEM case, a similar approach can be applied. When the beam scans across the specimen at regular intervals, the positions of the center of beam form a 2D grid (Figure 4-8). At each grid point, the beam profile function convolves with the molecular density distribution function, giving rise to a pixel in the detected image. In general, two molecules will be located at different positions with respect to the grid interval (beam sampling interval), and their different offsets can be characterized by the difference in their centers of mass relative to the sampling interval.

If one molecular image can be superimposed upon the other with their centers of mass aligned, it will be effectively the same as sampling the molecule by twice as many samples. Then it is not surprising that the resultant image will have a higher resolution than the individual images. The more images to superimpose, the higher resolution will be obtained in the resultant image. For example, if we have four molecular images, with center of mass offsets \((0, 0)\), \((0, d/2)\), \((d/2, 0)\), and \((d/2, d/2)\) respectively, where \(d\) is the sampling or scanning interval, after superimposing the four images, a resultant image with twice the resolution can be obtained.

So consider the center of mass offsets are distributed uniformly (in Appendix 1 this condition is shown to be not necessary). It requires four times as many images to obtain twice the resolution than that determined from the sampling interval of each image; similarly, for a one-dimensional signal, it requires twice as many measurements to obtain twice the resolution. It can be deduced that eight times as many images are required to obtain a 3D reconstruction with a two-fold better resolution, provided a proper superposition of images can be performed, or, as the number of
Figure 4-8 Hypothetical molecular particles sitting at the same orientations are sampled under an STEM. "x" marks the center of each molecule. Grid points stand for the positions of the beam when it scans across the specimen.
images increases by a factor of $N$, the resolution of reconstruction will increase (become smaller) by a factor of $\sqrt[3]{N}$. Without being limited by the Nyquist's sampling resolution. This is consistent with the minimum requirement of the number of images discussed in 4.2.3.2.

Appearing to be in conflict with Nyquist's sampling theorem (Bracewell, 1965), this is actually an extension to the sampling theorem under the condition that multiple measurements are obtained. It can be explained with Fourier analysis (see Appendix 1).

Thus as the answer to the first question above, the "extra" information can come from proper superposition of multiple measurements. So the extra resolution achieved by the finer voxel method is effectively derived from a sort of superposition. This will be further discussed in 4.2.3.3.3.

However, rigorously speaking, the signal we are trying to recover in the example above is the original signal convolved with the beam profile function. When the effective width of beam profile function is sufficiently small, the effect of the beam profile can be neglected. While in the situations where the width of the beam profile function is comparable to the sampling interval, conceptually a deconvolution is required to obtain the original signal accurately.
4.2.3.3.2 Deconvolution of beam profile function

The beam profile function of a STEM probe is normally considered to be a point spread function whose Fourier transform is band limited, and this cutoff is usually regarded as the resolution limit of the device. So the convolution of the point spread function and the signal seems to be band-limited too. Then it raises the question, equivalent to the second question asked above, whether the point spread function can be deconvolved from the convolved signal to recover the Fourier components in the original signal at the frequency higher than the cutoff frequency in the beam profile function.

Actually this deconvolution will be possible when the signal to be recovered, the molecular density distribution function, is space-limited. Consider only a one-dimensional situation (Figure 4-9). Assume the signal is all zero outside a region of size $D$ (this $D$ can be estimated based on the molecular weight of the protein). Then when the point spread function is convolved with the specimen from one end of the signal to the other end, only the central $2D$ portion of the point spread function "interacts" with the signal. Or alternatively, the point spread function at the two tails is effectively zero in the process of convolution. So, instead of considering the beam profile as a space-unlimited point spread function, we can safely truncate the point spread function in the center by a rectangular function of width $2D$. This "blunt tailed" point spread function has a Fourier transform extending to infinity. The convolution of this "blunt tailed" point spread function
Figure 4-9 One-dimensional illustration of the fact that only the central $2D$ portion of beam profile function convolves with the molecular density function as the beam scans across the molecule from one end to the other. $D$ is the maximum dimension of the molecule.
with the signal has frequency components up to infinity. So the beam profile function can be possibly deconvolved to recover the original signal. This answers the second question asked above, i.e., the frequency band limit of the beam profile function will not limit the achievable resolution in the reconstruction.

Since this deconvolution is based on the knowledge that the signal is space-limited, to perform this deconvolution, one has to identify a region outside which the original signal can be considered all zero. It is this knowledge that gives the extra constraint to make the deconvolution possible. After superimposing the several measurements, the deconvolution can be carried out in the Fourier domain simply by dividing by the Fourier transform of the “blunt tailed” point spread function.

In Appendix 2, we provide a real space method to implement a one-dimensional superposition and deconvolution simultaneously by solving a system of linear equations. The extension of the method in Appendix 2 into three dimensions becomes the finer voxel method for reconstruction.
4.2.3.3.3 Principles of finer voxel method

The finer voxel method implements the principles of both superposition of multiple measurements and deconvolution of beam profile function. This method is a direct extension of the method in Appendix 2 from one dimension to three dimensions.

In the situation where molecular images are obtained in STEM along different orientations, an image can be understood as a measure of the 3D density distribution of the molecule. In this measurement, the resolution in the image plane is twice the sampling interval, and the resolution along the projection direction is effectively the dimension of the molecule along that direction, provided that the dimension of the molecule is known. When a second image projected along a different direction is obtained, it will contribute to improving the resolution in the reconstruction. As long as the two projection directions are not perpendicular to each other, the second image will always contribute to improving the resolution in the plane where the first image lies. Or alternatively, it can be understood as a "partial superposition" of the second image onto the first image. When enough measurements from different directions are obtained, this superposition will enhance the reconstruction resolution in the first image plane, making it possibly higher than the resolution determined by the sampling interval. The reconstruction resolution in all the other planes will also be improved simultaneously by this superposition. The deconvolution in this method, just like that in Appendix 2, is implemented by solving linear equations.
When the targeted detail in the reconstruction is much bigger than the effective width of the beam profile function, the deconvolution of the beam is in fact not necessary. The finer voxel method can be modified just for the purpose of superposition of image information. However, our current STEM has a beam size of 3 angstroms, to achieve atomic resolution the beam profile function cannot be neglected.

4.2.3.4 Discussion

There will be a few factors that are related to the achievable resolution for the finer voxel method. They are the beam sampling interval, the size of the protein, the number of images, and the tolerable orientation error. These factors will be discussed in this section.

In Step 3 of the finer voxel method, images are aligned to the cubic box according to the calculated center of mass for each image. However the center of mass can never be determined accurately after a continuous function is discretely sampled. Statistically we can estimate the error in the calculation of center of mass, in relation to the sampling interval and the number of pixels that form the protein image, and we can show the targeted resolution can still be achieved with good confidence, as long as the error is sufficiently small (see Appendix 3). Appendix 3 suggests that we will have 95% confidence of reconstructing a protein at a targeted resolution as along as the number of pixels that form each protein image is greater than $1.28d^2 / x^2$, where $d$ is the sampling interval, and $x$ is the targeted resolution. Therefore we can calculate, for different beam
sampling intervals. the minimal sizes of protein that can be reconstructed at atomic resolution (around 1.5 Å) using the finer voxel method (Table 1).

We can also calculate the number of images required for the reconstruction (Table 2) and the tolerable orientation errors for different sizes of protein (Table 3). The formulae used for these calculations are derived in Appendix 4.

Table 1 suggests, for a 3 Å sampling interval, all sizes of protein can be reconstructed at atomic resolution (1.5 Å) by the finer voxel method if one disregards radiation damage; when the sampling interval increases to 6 Å, almost all the proteins can still be reconstructed, with the dose decreased 4-fold (square of 2); when the sampling interval increases to 10 Å, proteins of median or big size can still be reconstructed and the dose is decreased about 12-fold (square of 10/3).

The advantage of this reconstruction approach is that it provides the potential of reconstructing the molecular structure at atomic resolution, even from images obtained with low magnification. This permits the molecule to be sparsely sampled with an electron beam without degrading the reconstruction resolution. It is effectively spreading electrons to more molecules and therefore decreasing radiation damage to each molecule.

The trade off for this approach, as shown in Table 2, is that a large number of protein images are required for a reconstruction, especially at a large sampling interval (low magnification) and for a
high targeted resolution. This will result in a great amount of work and computer time in selecting protein particles, determining their orientations, and solving linear equations for the reconstruction.

Table 3 indicates another requirement for a reconstruction to achieve a certain resolution, a sufficient accuracy in the determination of protein orientations. The determination of image orientations, or angular reconstitution, is based on the common axis principle (van Heel, 1987). In the search for common lines, comparisons between line projections from different images are performed. The reduction of information from two dimensional images to one-dimensional line projections makes the search difficult for noisy data, and the determination of orientations with a sufficiently small error for atomic-resolution reconstruction is more challenging. Again for a big number of images, big computer power is required for angular reconstitution. An iterative quaternion assisted method for determining image orientations is an elegant and fast approach to align a giant number of images (Farrow and Ottensmeyer, 1992, 1993). The accuracy of orientation determination is related to many issues, such as imaging signal-to-noise ratio and radiation damage to the protein specimen. Currently, for protein structure reconstruction at atomic resolution, the required angular accuracy shown in
Table 3 has not been achieved. It may still take some time for angular reconstruction to achieve the desired accuracy. Nevertheless, as techniques advance, eventually protein atomic structures will be possible by electron microscopy.

Table 1 Minimal size requirements for a protein to be reconstructed at atomic resolution with different sampling intervals

<table>
<thead>
<tr>
<th>Sampling interval (Å)</th>
<th>Minimum of protein size (kD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0.18</td>
</tr>
<tr>
<td>6</td>
<td>12</td>
</tr>
<tr>
<td>10</td>
<td>250</td>
</tr>
<tr>
<td>15</td>
<td>2,800</td>
</tr>
</tbody>
</table>

Table 2 Number of images required for different sizes of protein to be reconstructed at atomic resolution with different sampling intervals

<table>
<thead>
<tr>
<th>Sampling interval (Å)</th>
<th>50 kD</th>
<th>100kD</th>
<th>200kD</th>
<th>400kD</th>
<th>1,200kD</th>
<th>2,500kD</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>710</td>
<td>890</td>
<td>1,100</td>
<td>1,400</td>
<td>2,000</td>
<td>2,600</td>
</tr>
<tr>
<td>6</td>
<td>2,800</td>
<td>3,600</td>
<td>4,500</td>
<td>5,700</td>
<td>8,200</td>
<td>10,000</td>
</tr>
<tr>
<td>10</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>16,000</td>
<td>23,000</td>
<td>29,000</td>
</tr>
<tr>
<td>15</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>65,000</td>
<td>/</td>
</tr>
</tbody>
</table>

3 The estimated error in the determination of orientations is around 10 degrees for images with a signal-to-noise ratio of 5, according to Farrow's investigation using simulated data (Farrow and Ottensmeyer, 1992).
Table 3 Acceptable orientation error for different sizes of protein to be reconstructed at atomic resolution

<table>
<thead>
<tr>
<th>Size of protein (kD)</th>
<th>Acceptable orientation error (degree)</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>1.7</td>
</tr>
<tr>
<td>100</td>
<td>1.4</td>
</tr>
<tr>
<td>200</td>
<td>1.1</td>
</tr>
<tr>
<td>400</td>
<td>0.9</td>
</tr>
<tr>
<td>1,200</td>
<td>0.6</td>
</tr>
<tr>
<td>2,500</td>
<td>0.4</td>
</tr>
</tbody>
</table>
4.3 Conclusion

For current instrumental capabilities of electron microscopy, a more accurate protein structure determination at a higher resolution mainly depends on computational improvements. In this project, some image analysis and image reconstruction techniques are developed or investigated for this purpose.

An angular distribution dependent (ADD) filtering algorithm was developed to replace the filtering methods in the filtered backprojection reconstruction. Reconstructions obtained by this approach are shown to be more accurate and contain higher resolution information. The phase difference analysis (PDA) is employed in the calculation of the sinogram correlation functions (SCFs). The peaks in the SCFs are enhanced, to facilitate a more accurate determination of image orientations. The shape factor is defined for an investigation of automated molecular segmentation for electron micrographs. Some preliminary results are obtained. The finer voxel method is proposed as a new 3D reconstruction method. Determination of protein structure at atomic resolution by electron microscopy is theoretically possible using this approach.

With future progress, electron microscopy will become more powerful in helping biologists with the understanding of the structure of proteins which have not been crystallized or are too big for NMR spectroscopy.
Appendix 1 Fourier analysis on superposition of multiple measurements

For the sake of simplicity, we will use a one-dimensional analogue to explain, by Fourier analysis, that higher resolution can be obtained when multiple discrete measurements are superimposed upon each other.

For a space-limited real continuous signal, its Fourier transform is a hermitian continuous function, with its frequency spectrum extending to infinity. When the real space signal is discretely sampled at regular interval \( d \), its Fourier transform is convolved with a Comb function whose spacing is \( 1/d \) (Figure A1-1). This convolution leads to a periodicity in the Fourier domain with a period of \( 1/d \). Since the frequency of the object extends to infinity, the periodicity results in the so-called aliasing, which means that the frequency spectrum in each period overlaps with each other. Then from one discrete measurement of a signal, a frequency component higher than \( 1/2d \), or the Nyquist’s frequency \( f_c \), will never be recovered. This Nyquist’s frequency determines the resolution limit for a discrete measurement. Aliasing always exists when the sampling frequency is lower than twice the highest frequency in the measured signal convolved with the sampling function, or, for STEM, the beam profile function. So aliasing occurs in STEM imaging, not only because the imaged molecules are space-limited, but also because molecules contain atomic resolution structure while the sampling interval is normally much bigger. In the presence of aliasing, the Fourier transform of the measured signal even within the range of \( [-f_c, f_c] \) is not the real frequency spectrum of the
Figure A1–1  A space-limited continuous one-dimensional signal (a) has a Fourier transform (b) extending to infinity. A Comb function spaced by intervals of $d$ (c) can be Fourier-transformed to another Comb function spaced by $1/d$ (d). The signal (a) is sampled by the Comb function (c), giving rise to a discrete signal (e), whose Fourier transform is function (f), or, a convolution of function (b) and function (d).
original signal at that range, because the aliased frequency spectrums extended from other periods add to the spectrum in the $[f_c, f_e]$ range. When no further measurements are carried out, only the Fourier transform of the measured signal can be approximated as the real frequency spectrum in $[f_c, f_e]$. When multiple discrete measurements of the same signal are performed with different sampling offsets, extra information not only permits one to determine the frequency components in $[f_c, f_e]$ with more accuracy, but also allows the frequency components outside $[f_c, f_e]$ to be determined. In this situation, consider the Fourier transform of the measured signal at each frequency within $[f_c, f_e]$ as the sum of all the frequency components that extend to that frequency point. For example, in Figure 1-2a, to determine the frequency components up to $2f_c$, we consider the Fourier transform of the original signal extends to $2f_c$ and $-2f_c$. Then for each frequency $f (0 \leq f \leq f_e)$, the measured Fourier component $F'(f)$ is the sum of two overlapping Fourier components of the original signal: one is $F(f)$ from the period centered at 0, the other is $F(f-2f_c)$ from the period centered at $2f_c$.

Then for the first measurement, we have

$$F'_1(f) = F(f) + F(f - 2f_c) \quad (A1-1)$$

The second measurement is performed by a shift of a fraction of the sampling interval, $\Delta d$, with respect to the first measurement. This shift in real space will add a phase delay to each Fourier
Fourier transform of measured signal

Figure A1–2 a. To recover the frequency components up to twice the Nyquist's frequency, consider that each period of the Fourier transform of the original signal is overlapped with the neighbouring periods. The overlapping region is only considered to be half of the period. The Fourier transform of the measured signal at each point then will be the sum of the overlapping frequency components. Solid curve, the Fourier transform of the measured signal, $F'$; dash curve, the Fourier transform of the original signal, $F$. Then we have $F'(f) = F(f) + F(f - 2fc)$ for a nonnegative $f$ smaller than $2fc$. 
Figure A1-2 b. To recover the frequency components up to three times the Nyquist's frequency, consider each period of the Fourier transform of the original signal is overlapped with the neighboring periods. The overlapping region is considered to be one period. The Fourier transform of the measured signal at each point then will be the sum of three overlapping frequency components. Solid curve, the Fourier transform of measured signal; dash curve, the Fourier transform of the original signal.
component of the original signal, by an amount proportional to the frequency. Then we have, for the second measurement,

\[ F'_2(f) = F(f)e^{i2\pi f_A N} + F(f - 2f_c)e^{i2\pi f_{2f_c} N} \]  

where \( N \) is the number of samples in each measurement, the same as the number of samples in the discrete Fourier transform.

Combining Equation (A1-1) and Equation (A1-2), \( F(f) \) and \( F(f-2f_c) \) can be solved. So the Fourier components up to frequency \( 2f_c \) can be determined. It also agrees with the discussion in 4.2.3.3.1 in that two measures are required for the equations to be solvable.

Similarly, to resolve the signal at three times better resolution than the Nyquist’s resolution we need to consider more overlap in the Fourier domain, and to perform more measurements (Figure A1-2b). It can be proved that \( m \) measurements are required to recover a one-dimensional signal at \( m \) times better sampling resolution. It should be pointed out that in the argument above we did not assume the shift \( \Delta d \), or offset difference, to be half the sampling interval \( d \), which suggests the sampling offsets of measurements are not required to be uniformly distributed.
Appendix 2 Superposition of multiple measurements and deconvolution of beam profile function by solving linear equations

Here we give a real space method to implement both superposition of multiple measurements and deconvolution of a beam profile function in a one-dimensional situation.

Suppose multiple discrete measures on a continuous space-limited one-dimensional signal are performed, and each measurement is composed of \( N \) samples at sampling interval \( d \) (Figure A2-la, b). The Nyquist’s resolution limit in each individual measurement is \( 2d \). To recover the signal at \( m \) times better the resolution, \( m \) measures at different sampling offsets are required from the argument above.

Assume we can divide the original signal by regular intervals of \( d/m \) (Figure A2-1c), and the integrated density of the signal in each interval will give rise to a discrete signal sampled at \( d/m \) interval (Figure A2-1d). If this series of samples, composed of \( m \times N \) samples, can be determined, the signal is recovered at resolution \( 2d/m \).

Assume each sample in this series is at \( x_k \), and the magnitude of each sample is \( y_k \), where \( 1 \leq k \leq m \times N \). In measurement \( q \), we denote the \( j \)th sample as \( I_j^{(q)} \), and the beam giving rise to the
Figure A2-1 (a) and (b), two measurement of a continuous signal with different sampling offsets; (c), divide the signal by intervals smaller than the sampling interval; (d), integrated density in each small interval; (e), beam positions for measurement (a).
sample is centered at $C_j^{(')}$, where $1 \leq q \leq m$, $1 \leq j \leq N$ (Figure A2-1e). The beam function is $p(x)$. For the sake of simplicity, we also denote each $d/m$ interval centered at $x_k$ as $W_k$, i.e.,

$$W_k = \{x | x_k - d/2m \leq x \leq x_k + d/2m\} \quad (A2-1)$$

So for sample $j$ in measurement $q$, we have

$$I_j^{('q')} = \sum_k y_k \left( \int_{W_k} p(x - c_j^{('q')})dx \right) \quad (A2-2)$$

Then we can form a system of linear equations with $m \times N$ variables ($y_k$). Each measurement will provide $N$ different equations in form of Equation (A2-2). It is apparent that $m$ different measures are required to solve this set of equations. This further proves that, in order to recover a signal at $m$ times better resolution than the Nyquist's limit, at least $m$ measures of the signal sampled at different offsets are required. This is consistent with the Fourier analysis in Appendix 1. If we have more measurements, the extra equations will be linear combinations of others under ideal situation. In the presence of noise and other source of error, solving the set of linear equations becomes a minimization problem; more equations than necessary will provide us with more information which can be used to obtain an optimized solution with increased signal-to-noise ratio.

Below is an example to demonstrate this method.
Assume we have a one-dimensional density distribution and a beam profile function as those in Figure A2-2a. In Figure A2-2b and Figure A2-2c, the beam profile function is used to scan over the density function twice at a sampling interval of 2, with each time a different offset. Each measurement is a convolution of the beam profile function with the density distribution function. Since the density distribution has two points spaced by 1 and the scanning interval for each measurement is 2, it is impossible to resolve the two points from only one of the measurements. Moreover the frequency response of the beam profile function would not allow the two points to be resolved even if the scanning interval were 1.

Assume the relative offsets of the measurements are known, the beam profile function is known, and more importantly, the original density distribution is known to be space-limited (for example, the support of the density distribution function is not bigger than 5), we will be able to use the above-proposed method to retrieve the original density distribution from the two measurements.

Now we are to determine the values of a discrete function at five contiguous positions. The values of the function at other positions are known to be zero. Assume the values of the five positions are \(x_1, x_2, x_3, x_4,\) and \(x_5\). To solve five variables, five linear equations are required. We can choose any five data from the two measurements to form equations, as along as the two measurements are translationally aligned with respect to each other. For example we use five data shown in Figure A2-2d, we will get a system of linear equations as follow.
Figure A2-2 a. Top, a one-dimensional density distribution function, which includes only two points separated by 1; the numbers are the integrated density of the two points. Bottom, a beam profile function; dash line, the continuous function; solid line, the discrete approximation to the function; the numbers are the integrated intensity of the function at each interval (or the area of each rectangle).
The beam profile function

Beam scanning

At interval of 2

The density distribution function

Figure A2-2 b. The beam profile function (top) scans over an object, the density distribution function (middle) at a scanning interval of 2 with a certain offset (indicated by the relative positions of the beam and the object shown in this figure), giving rise to the first measurement (bottom). When the relative positions of the beam and the object are those shown in the figure, the measured intensity of the object is the number associated with the shaded rectangle. As the beam scans to the right, more data of the object are obtained.
Figure A2–2 c. The beam profile function (top) scans over the object (middle) at a scanning interval of 2 with a different offset. When the relative positions of the beam and the object are those shown in the figure, the measured intensity of the object is the number associated with the shaded rectangle. As the beam scans to the right, more data are obtained.
Figure A2–2d Assume the object, or the density distribution function, only has a length not bigger than 5. Then five variables are to be determined. From the two measurements, any five data can be used to formed a system of linear equations for the five variables. Assume the relative offset between the two measurements are known (they are aligned as shown in the figure). Then five data are chosen as those pointed by the arrows in the figure. The absolute position of the object function with respect to the beam function may be unknown, but the relative positions of the object function in two measures are known. The absolute position of object with respect to the beam can be chosen arbitrarily. When five linear equations are formed, the solution to the equations recovers the original object. The two points in the density distribution function can be resolved.
From first measurement,

\[
0.01x_5 + 0.02x_4 + 0.05x_3 + 0.1x_2 + 0.3x_1 = 0.4 \\
0.05x_5 + 0.1x_4 + 0.3x_3 + 0.3x_2 + 0.1x_1 = 0.7 \\
0.3x_5 + 0.3x_4 + 0.1x_3 + 0.05x_2 + 0.02x_1 = 0.22
\]

From the second measurement,

\[
0.02x_5 + 0.05x_4 + 0.1x_3 + 0.3x_2 + 0.3x_1 = 0.5 \\
0.1x_5 + 0.3x_4 + 0.3x_3 + 0.1x_2 + 0.05x_1 = 0.65
\]

Solving the five equations we get

\[
x_1 = 1 \\
x_2 = 0 \\
x_3 = 2 \\
x_4 = 0 \\
x_5 = 0
\]

This demonstrates that the original density function can be recovered.
Appendix 3  Estimation the error in the calculation of center of mass

When the center of mass of an image is calculated, an implicit assumption is made that the center of mass of each pixel is at the center of the pixel. This approximate assumption has an error, and it is related to the error of calculating center of mass of a molecule from an image. The error in the calculation of center of mass can be estimated statistically by the Central Limit Theorem.

Since the calculation of a center of mass in 2D coordinates always breaks down into its x and y components. A discussion in a one-dimensional context is sufficient here. Assume \( f(x) \) is a continuous density distribution function defined in a closed region \( R \) and \( R \) is divided into \( N \) intervals \( R_1, R_2, \ldots, R_N \). Each interval has the same size \( d \). Let \( i \) index the number of interval.

The measured mass in the \( i \)th interval can be determined as:

\[
M_i = \int_{R_i} f(x) dx \quad (A3-1)
\]

The real center of mass in the \( i \)th interval is:

\[
C_i = \int_{R_i} f(x) x dx \quad (A3-2)
\]
So the real center of mass of \( f(x) \) is:

\[
C = \frac{\sum_i M_i C_i}{\sum_i M_i} \tag{A3-3}
\]

Assume the center of interval \( R_i \) is \( x_i \). And it is used to approximate the real center of mass \( C_i \) to calculate the center of mass of \( f(x) \). Then the calculated center of mass is:

\[
C' = \frac{\sum_i M_i x_i}{\sum_i M_i} \tag{A3-4}
\]

The error in calculating center of mass

\[
E = C - C'
\]

or,

\[
E = \frac{\sum_i M_i (C_i - X_i)}{\sum_i M_i} \tag{A3-5}
\]

where the error in measuring \( M_i \) is neglected. Let
Then $e_i$ represents the error term in approximating the center of mass of an interval as the center of the interval. Equation (A3-5) becomes:

$$E = \sum_i e_i \left( \sum_i \frac{M_i}{M} \right)$$  \hspace{1cm} (A3-7)

Since the center of mass in an interval can be anywhere in the interval, we assume the center of mass in an interval has a uniform probability distribution in the interval, i.e., for all $e_i$, the probability density function of $e_i$, $p(e)$, is uniform distribution in $[-d/2, d/2]$.

$$p(e) = \begin{cases} 
1/d, & -d/2 \leq e \leq d/2 \\
0, & \text{else}
\end{cases}$$  \hspace{1cm} (A3-8)

The mean of $e$, $\mu(e)$, is:

$$\mu(e) = \int_{-\infty}^{\infty} p(e)de = 0$$  \hspace{1cm} (A3-9)

The standard deviation of $e$, $\sigma(e)$, can be determined by:
\[ \sigma^2(e) = \int_{-\infty}^{\infty} p(e)(e - \mu(e))^2 \, de = d^2 / 12 \]

or,

\[ \sigma(e) = d / \sqrt{12} \quad \text{(A3-10)} \]

From Equation (A3-7), \( E \) is a weighted average of \( e \) among \( N \) samples. Approximate the weight for each weighting factor \( M_i / (\sum M_j) \) as \( 1/N \). Then \( E \) is the mean of \( N \) samples of \( e \). From the Central Limit Theorem, the standard deviation of \( E \) can be estimated as:

\[ \sigma(E) = \sigma(e) / \sqrt{N} \quad \text{(A3-11)} \]

Then we have

\[ \sigma(E) = \sigma(e) / \sqrt{12N} \quad \text{(A3-12)} \]

When \( N \) is big enough, \( E \) approaches a normal distribution. To achieve a resolution of \( x \), \( E \) has to satisfy:

\[ |E| \leq x / 2 \quad \text{(A3-13)} \]
Although this can not be satisfied deterministically, we will satisfy this condition with 95% confidence by considering $E$ to be within the range of $[-1.96 \sigma(E), 1.96 \sigma(E)]$. Then we have

$$1.96\sigma(E) \leq x / 2$$  \hspace{1cm} (A3-14)

Combining with Equation (A3-12), we have

$$N \geq 1.28d^2 / x^2$$  \hspace{1cm} (A3-15)

This indicates that to reconstruct a protein structure to a resolution of $x$ with images obtained by a sampling interval of $d$, the number of pixels that form the protein image has to be at least $1.28d^2/x^2$. 
Appendix 4 Some formulae used in the calculation of the tables

This section derives some formulae in the calculation of Table 1, Table 2, and Table 3.

The average density of a protein is 1.3g/cm³. For a protein of molecular weight M kD, its volume V is:

\[ V = \frac{M \times 10^{24}}{6.02 \times 10^{23} \times 1.3} (\text{Å}^3) \]

or,

\[ V = 1280M (\text{Å}^3) \]  \hspace{1cm} (A4-1)

Assume proteins are spherical, the radius of a protein is R, then we have:

\[ 4\pi R^3 / 3 = V \]

or,

\[ R = 6.73\sqrt[3]{M} (\text{Å}) \]  \hspace{1cm} (A4-2)
Assume the pixel size is \( d \), then the number of pixels that form the protein in an image is:

\[
N = \frac{\pi R^2}{d^2}
\]

or,

\[
N = \frac{142^\frac{3}{2}M^2}{d^2}
\]

(A4-3)

From Equation (A3-13), to have 95% confidence of reconstructing the protein at resolution \( x \) using the finer voxel method,

\[
\frac{142^\frac{3}{2}M^2}{d^2} \geq 1.28d^2/x^2
\]

or,

\[
M \geq d^6/1170x^3 \text{ (kD)}
\]

(A4-4)

has to be satisfied. (A4-4) is used to calculate Table 1.
For the \( x \) A targeted resolution, the number of voxels in reconstructed volume is:

\[
Q = \frac{V}{(\frac{x}{2})^3}
\]

thus \( Q \) variables in the linear equations are to be solved. Every image can provide \( N \) equations, then the number of images required is:

\[
I = \frac{Q}{N}
\]

or,

\[
I = 72d^{2\frac{3}{2}}\sqrt{M} / x^3
\]  \hspace{1cm} (A4-5)

(A4-5) is used to calculate Table 2.

An error of \( \phi \) radian in the determination of its orientation will cause a tangential error at the protein surface of \( R\phi \), degrading the resolution to not better than \( 2R\phi \). Thus for the targeted resolution \( x \), the orientation error has to satisfy

\[2R\phi \leq x\]
or,

\[ \phi \leq \frac{x}{13.5\sqrt{M}} \text{ (radian)} \]  

(A4-6)

(A4-6) is used to calculate Table 3.
References


Harauz, G. and van Heel, M., “Exact filters for general geometry three-dimensional reconstruction,” *Optik* 73, 146-156 (1986).


