

University of Alberta

Transmission Electron Tomography: Imaging Nanostructures in 3D

by

Xiongyao Wang

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Physics

©Xiongyao Wang

Fall 2012

Edmonton, Alberta

Permission is hereby granted to the University of Alberta Libraries to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only. Where the thesis is converted to, or otherwise made available in digital form, the University of Alberta will advise potential users of the thesis of these terms.

The author reserves all other publication and other rights in association with the copyright in the thesis and, except as herein before provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatsoever without the author's prior written permission.

Abstract

In studies of nanoscale materials, electron tomography (ET) can be used to obtain three dimensional information on the morphology and spatial distribution of nanoparticles. Electron tomography records a tilt series of projected images of an object and then mathematically reconstructs a 3D map of the object from the recorded images. Filtered back-projection (FBP), which is based on spatial Fourier transforms, and simultaneous iterative reconstruction techniques (SIRT), which are purely algebraic, are the two common reconstruction methods used to generate a tomographic “3D matrix” from the recorded images. Three aspects of ET were investigated in this thesis: the reconstruction and visualization, experimental considerations, and practical applications.

First, to quantify the quality of the reconstruction and the selection of visualization threshold was discussed. The quality of the reconstruction by FBP and SIRT methods was evaluated by root mean square (RMS) difference frequency analysis, a *quantitative* description of similarity between the original test and its reconstruction. A quality index (QI) method was proposed and successfully applied to set the visualization threshold for volume rendering of tomographic reconstructions. Setting the threshold according to *a priori* known space-filling volume fraction of nanoparticles was found not to be a suitable parameter for visualization. The effect of the filter used in FBP was examined.

On the experimental front of electron tomography, a new ET sample preparation method was developed. The new method combines standard thin film deposition techniques and focused ion beam (FIB) milling. The proposed method minimized the effect of the projected thickness and missing wedge by controlling the thickness of the thin film and the width of the bar. Furthermore, the new method reduces gallium implantation problems and is suitable for tomographic sample preparation of samples in solution.

Finally, the tomographic results of latex nanoparticles, Au nanocrystal multilayer, and Si nanocrystals embedded in silica glass were examined. A new method was developed by

combining high angle annular dark-field (HAADF) and energy-filtered STEM techniques simultaneously to obtain parallel recording of 3D tomographic data from two different types of nanoparticles. This method was successfully applied to investigate the Er doped Si nanocrystals system.

Acknowledgment

The most rewarding time of my life is probably the last five and half years. During my Ph.D studies in the Department of Physics at University of Alberta, I have experienced a high quality of education and great research experience. I believe the accomplishment that I am going to achieve would not be possible without the help of people around me. First and foremost, I would like to thank my supervisors, Dr. Marek Malac and Prof. Al Meldrum, for mentoring me all the way in electron tomography and nanoscale material research. Their patience, encouragement and support is the basis of the completeness of this thesis research.

I am grateful to Ross Lockwood for his help in simulation, sample fabrication, and result discussions; to Peng Li for ET sample preparation using FIB technique.

I also would like to thank other people: Hui Qian, Dr. Douglas Vick, Robert McLeod, Dr. Jian Chen, Dr. Kai Cui, Dr. Huairuo Zhang, Martin Kupsta, Florian Lenz, Yanyan Zhi, Shalon McFarlane, Prof. Ray Egerton, Dr. Hiromitsu Furukawa and Dr. Carol Wang. To learn from them and share the thoughts has always been enjoyable experience.

In addition, I thank other professors and staff members in the Department of Physics at University of Alberta for striving to help me achieve my educational goal in such a prime institution.

Last, but not least my special thanks go to my family for their constant love and support.

Contents

1	Introduction	1
1.1	The Principle of Tomography	3
1.1.1	Radon Transform	3
1.1.2	Inverse Radon Transform	5
1.2	The Filtered Back-Projection (FBP)	7
1.2.1	The simple back-projection algorithm	7
1.2.2	The Algorithm of the Filtered Back Projection (FBP)	11
1.3	The Iterative Algorithm	15
1.3.1	The Algebraic Representation of the Cross Section of the Object and its Projection	16
1.3.2	Algebraic Reconstruction Techniques (ART) and Simultaneous Iterative Reconstruction Technique (SIRT)	21
1.4	Introduction of Two TEM Techniques	23
1.4.1	High-Angle Annular Dark-Field (HAADF) Imaging Mode	24
1.4.2	Energy Filtered Transmission Electron Microscopy (EFTEM) Imaging Mode	27
1.5	Summary	29
	Bibliography	30
2	Reconstruction and visualization of nanoparticle composites by transmission electron tomography	33
2.1	Simulations and Experiment	36
2.2	Results and Discussion	37
2.2.1	Visualization and rendering	37
2.3	Summary	57

Bibliography	58
3 Experimental Considerations of Electron Tomography	60
3.1 Sample Preparation Methods for Electron Tomography	60
3.1.1 Grid Support Method	61
3.1.2 Physical thinning Methods	61
3.1.3 The FIB Micro-Sampling Technique	62
3.1.4 A New Method of 3D Electron Tomography Sample Fabrication by Combining FIB Technique and Thin film deposition methods	65
3.2 Electron Tomography Imaging Modes	69
3.3 Summary	70
Bibliography	71
4 Experimental Results and Discussion	74
4.1 Latex Nanoparticles	74
4.2 Au nanocrystal multilayer	77
4.3 Si nanocrystals embedded in glass	82
4.4 Er-doped Si nanocrystals	87
4.5 Summary	93
Bibliography	95
5 Summary and Future Directions	98
Bibliography	104
List of Publications	105
Journal Papers	105
Conference Abstracts	106
Appendix A Reading and Writing the Reconstruction File Generated by TEMography Software	107
Appendix B Calculation of RMS Difference	110

List of Figures

1.1	An example showing information loss along the incident beam direction. The distance between the largest sphere and the smallest sphere cannot be obtained directly from the projected image, due to the 2D projection nature of the image.	3
1.2	(a) Illustration of the angle relation (φ) between the line L and x -axis ; (b) the illustration of the coordinate system xoy and sot	4
1.3	An illustration of the simple back-projection algorithm. (a): the object consists two dot-shape disks. The line P in (b) represents the projection of the object, which only contains two projected points of the object for this example. (b)(c)(d): the back-projected body (B) is illustrated along three projected angle (Θ_1 , Θ_2 , and Θ_3). The back-projected body is a line perpendicular to the projected line P. The intersection of B and P is the projected points. The width of the back-projected body is the diameter of the projected dot-shape object. (e): the summation is to sum/overlap the back-projected bodies at different angles. The result of the summation shows two dot-shape disks marked by two red circles. Revised from Ref [1].	8
1.4	An illustration of the relationship between then different coordinate systems. (x, y) is the fixed coordinate; (x_r, y_r) is the rotating coordinate; (r, θ) is the polar coordinate.	9
1.5	An illustration of the Fourier slice theorem. The 2D object in the x - y plane is projected along the y_r direction to the x_r line. $p_\phi(x_r)$ is the intensity of the object projected onto the line x_r . The coordinate system for the 2D Fourier transform of the object is $\omega_1 o \omega_2$. The line B in Fourier space is the 1D Fourier transform of $p_\phi(x_r)$. Revised from Ref [31].	12
1.6	The illustration of the relationship of coordinate systems used in FBP method. Revised from Ref [31].	14

1.7	In algebraic methods a square grid is superimposed over the volume that is being reconstructed. The value (intensity) within the volume is assumed to be constant within each cell (voxel) of the grid denoted by x_i . The projection process is decomposed into a series of projected ray y_i with constant width τ . The line integral of the projection of the object is called a <i>ray-sum</i> . The weighting factor a_{ij} is equal to the area of the grid superimposed by the ray. Revised from Ref [31].	17
1.8	The illustration of the concept of solving algebraic equations with two equations. One starts with an arbitrary initial guess (point H noted using x^0) and then projects onto the line y_1 corresponding to the first equation to obtain projected point x^1 . The projected point (x^1) on the line y_1 is then projected onto the line of the second equation y_2 to obtain the projected point x^2 . The obtained projected point x^2 is the result of two equations (y_1 and y_2) corresponding to one iterative loop. If necessary, the operation is repeated as illustrated by treating the obtained solution (or point) as the initial guess in the next loop until the solution is satisfactory. Revised from Ref [31]. . .	19
1.9	Simulated angular dependence of the differential cross sections for elastic and inelastic scattering of 100 keV electrons from a carbon atom, using the Lenz model. Modified from Ref [42].	24
1.10	Schematic of the common techniques used in STEM mode with the range of electron scattering angles gathered by each detector. Modified from Ref [1].	25
1.11	Illustration of HAADF technique. The electrons scattered at a high angle are collected using the annular detector. The collecting angle range is equal to $\theta_2 - \theta_1$. If the transmitted electrons are collected, the STEM bright-filed image is obtained. I is the intensity of the incident electrons. Revised from ref [43].	26
1.12	Comparison of energy-loss data acquisition using (a) an energy-filtered TEM, and (b) a STEM equipped with a parallel detection spectrometer. Revised from [47].	28
2.1	Illustration of the sampling of data in Fourier space. The relatively larger number of data points at low frequencies results in a blurred reconstruction. The angular increment between projections is θ and the maximum tilt angle α . Cited from Ref [6].	34
2.2	Illustration showing how a dual-axis tilt series collapses a missing wedge into a missing pyramid of information. Cited from Ref [6].	35

2.3	Projected images for 100-particle phantom data. (a) shows the original phantom data, while (b,c,d) show a FBP reconstruction using different thresholding for visualization.(a) has a quality index $QI = 1.00$, (b) has a quality index $QI = 0.37$, (c) has a quality index $QI = 0.56$ and d) has a quality index $Q = 0.21$. Figure (c) appears the most similar to the original phantom, as determined by QI	38
2.4	Visualization of the reconstructed results comparing: (a) volume rendering; (b) isosurface rendering with an threshold from 250 to 254; and (c) isosurface rendering with an threshold range from 245 to 254.	40
2.5	Zoom-in visualization of the FBP reconstruction for phantom data with different numbers of particles. (a) Reconstruction of 10 particles, the apparent density of the spheres is uniform in 10-particle case. (b) Reconstruction of 300 particle phantom data. The density of the spheres is not uniform. The two spheres marked by the red rectangle cannot be easily separated in the reconstruction although they can appear separated in some viewing angles. The red circle shows the distortion of one particle. (c) Visualization of the 500-particle reconstruction. The spheres in the red rectangle cannot be separated, from any viewing angle. “Fog” and streaking between spheres are obvious.	41
2.6	The profile of R-L filter in (a) frequency space and (b) real space. In the frequency space, the filter emphasizes the high frequency and de-emphasizes the low frequency. The filter acts as a high-pass filter to decrease the effect of low frequencies. The effect of the filter in real space is to behave like a window with the shape in (b) in the convolution in Eq. 1.20.	44
2.7	The profile of the R-L filter (a) discrete situation $h_{R-L}(nd)$ (dark circle) (b) using linear interpolation $\bar{h}_{R-L}(x_r)$ (blue line) and corresponding discrete value (dark-circle). The discrete form of the R-L filter is widely adopted in practical applications.	46
2.8	The profile of the S-L filter in (a) frequency space and (b) real space. In frequency space, the filter emphasizes the high frequency and de-emphasizes the low frequency. The filter acts as a high-pass filter to decrease the effect of low frequencies. Compared to the R-L filter, the S-L filter change is more smooth at high frequency range. Correspondingly, in real space, the S-L filter change is sharper around the zero point.	47

2.9	The profile of S-L filter and S-L filter with linear interpolation (a) discrete situation $h_{S-L}(nd)$ (b) using linear interpolation. $\bar{h}_{S-L}(x_r)$ (blue line) and corresponding discrete value (dark-circle). Compared to the discrete form of the R-L filter, the S-L filter becomes closer to x -axis when $x > d$. However, the R-L filter intersects with x -axis when the ratio of the interval over d is even.	48
2.10	Effect of various back projection filters on reconstruction of a two-dimensional circle by FBP, from 1D projections. (a) computer-generated circle 100 pixels in diameter in a 360×360 pixel plane. (b) the reconstruction without a filter in the back projection. (c) the reconstruction with the S-L filter. (d) shows a reconstruction with same parameters as (c) but with a logarithm applied to the intensity of projected data. The reconstruction in (d) shows clear enhancement of boundaries. The projections were generated using 1° tilt increments and a projection tilt angle from -90° to 90°	50
2.11	Zoom-in visualization of the reconstructed results by SIRT for various numbers of particles in the original phantom data. (a) Reconstruction of 10 particle phantom data. The apparent density (defined as pixel intensity) of spheres is uniform. (b) Reconstruction of 300 particle phantom data. Two spheres in the red rectangle can be separated more easily than that in FBP in Fig 2.5b). The particle marked by red circle again shows distortion from spherical shape. (c) Reconstruction of 500 particle volume. The two spheres marked by the red rectangle can be easily separated. "Fog" and streaking between spheres do not appear in 500 particle volume.	51
2.12	Comparison of the frequency distribution of the normalized RMS difference of the reconstructed results by FBP (blue circles) and SIRT (green crosses) respectively. The RMS frequency is higher for the FBP for almost all RMS differences. Both FBP and SIRT show significant number of pixels with high difference from the original phantom data. The increase in the RMS difference between the phantom data and the reconstructed volume is primarily due to increased frequency of pixels with small RMS difference.	52
2.13	Comparison of the RMS differences for reconstructions by FBP (a) and SIRT (b) using either 1° or 2° tilt increments, for 200 particles in the object volume. The differences, especially for the SIRT method, appear minor.	53
2.14	Zoom-in visualization showing pixels with an RMS difference equal to 0.5, from a) SIRT; and b) FBP. The object was a 300-particle phantom volume.	54

2.15	An illustration of sampling mechanism and its influence on the frequency of the RMS=0.5 voxels. (a) a circular particle with fine sampling; (b) the grid is coarsened such that each pixel has double the edge length; (c) the RMS distribution of a real 3D phantom with three different pixel sizes, corresponding to image of 600^3 , 300^3 , and 100^3 pixel size. The RMS=0.5 spike becomes less noticeable as the grid is coarsened. The maximum RMS value is normalized for each case, for better visual comparison of the curve shapes.	55
2.16	Zoom-in visualization of the 300-particle reconstruction and imaging results for: (a) FBP, volume fraction set to known fraction in the phantom (4.7%); (b) FBP for the best QI, giving a volume fraction of 1.5%; (c) SIRT, volume fraction set to known fraction in the phantom; (d) SIRT for the best QI, giving a volume fraction of 0.8%.	56
3.1	The illustration of a physical thinning preparation method. 1. Two pieces of the sample are adhered together using glue. 2. A 2mm×3mm cut bar is obtained by cutting the glued sample. 3. The cut bar is attached to a Cu grid using wax. 4. The Cu grid with the cut bar is then attached to a stub to be mechanically polished. 5. After mechanically polishing, the sample is further thinned using an Argon ion beam. Revised from [2]	62
3.2	An illustration of thickness effect caused by the rotation and projection. The sample is viewed along the x-axis. The parallel dashed lines represents the sample after rotating by θ degrees. The projected thickness (d_2) along z-axis is equal to d_1 at zero degrees. d_2 changes with the tilt angle θ	63
3.3	Schematic illustration of the FIB micro-sampling technique. Revised from [4].	64
3.4	Steps involved in the bridge-type sample preparation. Step 1 is the deposition of a thin film that is either the sample itself or a support film used for a suspended sample. Step 2 shows the collection of the thin film prepared in Step1 onto a TEM grid. Step 3 is the fabrication of a bridge-shaped sample using a focused ion beam instrument. L is the length of the side of the rectangular hole perpendicular to the bridge, and h is the thickness of the thin film.	66

3.5	The maximum achievable tilt angle as a function of the ratio L/h . For a given value of L/h , the maximum tilt corresponds to the greatest angle that would not lead to a projected overlap of the bridge and the film beyond the cut out rectangle. If L/h is larger than 3.73, a $\pm 75^\circ$ tilt range can be achieved. Ultimately the bars of the supporting Cu grid will limit the tilt angle to about $\pm 80^\circ$. The inset shows the values of the bridge width and film thickness for which the maximum projected sample thickness does not exceed 100 nm from any viewing direction (blue region in the inset).	67
3.6	An SEM image of the finished bar-shaped sample for TEM tomography. Area 1 is the original film, while area 2 is vacuum after the film was cut out. The inset shows a higher magnification of the bridge region of the sample. In this case, the bar slightly detached in the center, but this does not cause a problem for collecting a tilt series in the TEM.	68
3.7	The illustration of parallel (TEM) and convergent (STEM) illumination modes respectively in EM. Revised from Ref [15].	70
4.1	An example HAADF STEM image from the tilt series shows the 2D distribution of latex particles. A series of tilted images are collected for the 3D reconstruction. The tilt range covers from -73° to $+73^\circ$ with 2° increments.	75
4.2	Tomographic 3D reconstruction results. The 3D reconstruction result in (b) corresponds to the area marked by the yellow rectangle in (a). The latex nanoparticle marked in the red box shown in (b) shows that the elongation effect is not obvious in the reconstruction result. x-axis (red): 1216 nm; y-axis (green): 185 nm; z-axis (blue): 626 nm.	76
4.3	(a) Bright field STEM image of a needle-like sample consisting of Au particles embedded in SiO_2 . The Au particles appear dark. The other images show the SIRT reconstruction results with a quality index of (b): 0.32, (c): 0.28 and (d): 0.38.	79
4.4	3D visualization of Au particles by SIRT viewed along (a) the x-axis, (b) the y-axis, and (c) the z-axis.	80

4.5	Visualization of the tomographic reconstructions for the Au nanoparticles. Panels (a) and (b) show the FBP method, and (c) and (d) used SIRT. (a) volume rendering with the 16% estimated nanoparticle filling fraction. (b) image of the same region, thresholding for the best quality index. (c) and (d) are, again, thresholding according to the estimated “real” filling fraction and the best- <i>QI</i> filling fraction, respectively, but this time for SIRT. Scale: <i>x</i> -axis (red): 16 nm; <i>y</i> -axis (green): 58 nm; <i>z</i> -axis (blue): 54 nm. The <i>QI</i> and volume fractions are 0.35 and 13%, for (b), while for (d) they are 0.38 and 12%, respectively.	81
4.6	The EFTEM image using a 4 eV window centered at 17 eV shows Si nanoparticles in the SiO _{0.6} sample. This confirms that the Si nanoparticles are formed and can be visualized in EFTEM.	83
4.7	Three dimensional electron tomography reconstruction of the silicon distribution (volume rendering) in the SiO _{0.6} film by the FBP method. Volume rendering was used for this figure. The point-like contrast that appears in the red rectangle may originate from the low contrast of Si-NCs. The lengths of the axes are 53.90 nm (red, bottom vector), 51.15 nm (green, top vector), and 66.00 nm (blue, middle vector) [27].	84
4.8	HRTEM micrography of the SiO _{0.6} thin film. The HRTEM image indicates the presence of complex morphologies of silicon nanocrystals. The diffraction pattern in the inset is consistent with diamond-structure silicon [27].	85
4.9	HRTEM micrography of the SiO _{0.6} thin film. The HRTEM image shows particles touching together but with different lattice fringe orientation, which indicates the presence of interconnected silicon nanocrystals.	86
4.10	Illustration of the sensitization process. In (a), an incident photon is absorbed by a silicon nanocrystal, exciting an electron-hole pair. This electron-hole pair recombines non-radiatively and transfers its energy to the Er ³⁺ ion as shown in (b) by a Förster-type mechanism. Finally, the Er ³⁺ ion emits light with a wavelength near 1.5 μm in (c). (After Ref [34]).	87
4.11	Calculation of total elastic cross section for Si (lower, green curve) and Er (upper, red curve), respectively, based on the Lenz model. Most contrast contribution in HAADF images will arise from Er. The total cross section is the cross section integrated from zero degree to <i>x</i> ₀ degrees.	89

4.12	A 3D visualization of the reconstruction from HAADF visualized with volume rendering (a) and (b) and isosurface rendering (c). The shape of the Er cluster in (c) is based on the shape of Er visualized in volume rendering (b). The figure shows that the Er clusters are not uniform and two particles appear to be connected. The apparent connection between particles may be caused by the elongation effect due to the missing wedge, combined with small separation distances.	90
4.13	A 3D visualization of the reconstruction from EFTEM visualized with volume rendering (a) and isosurface rendering (b). Similar to HAADF reconstruction, the shape of Si-NCs visualized with isosurface rendering is determined based on the shape of Si-NCs with volume rendering. It shows that Si-NCs are not spherical in shape.	91
4.14	A 3D visualization of the combined reconstruction visualized by isosurface rendering. The images show the distribution of silicon (green) and erbium (red). One erbium cluster is located at the boundary of a Si-NC.	92
5.1	Flowchart of the DART algorithm. Revised from [1].	100

Chapter 1

Introduction

Transmission electron microscopy (TEM) and scanning TEM ((STEM) are well established methods for characterization of objects [1]. The high popularity of (S)TEM in (nano) materials characterization arises from the wide range of structural and chemical properties (S)TEM can measure. In favorable circumstances, (S)TEM measurement can be performed at nearly atomic resolution, and, often, structural, chemical, and dielectric information can be obtained from a *single* nanoscale object.

In order to investigate the three-dimensional (3D) structure of objects in conventional TEM, diffraction pattern images can be used. The method records a set of 2D electron diffraction patterns of the sample along several directions (angles) by tilting the sample along one direction. Owing to the property of symmetry in crystalline materials, three diffraction pattern images are usually enough to obtain 3D information. Then the reciprocal information is obtained directly from the diffraction images, since the diffraction pattern images tilt along one line or direction with known angles. Therefore, crystallographic information on the object can be obtained from the reconstructed reciprocal lattice. This tomographic method is quite simple and easy to apply when crystallographic information is only required [2, 3, 4, 5].

However, diffraction-based methods in the conventional TEM tomography face a significant disadvantage that limits its extensive application in practice. The method is only suitable in the case of crystalline samples since diffraction-based methods take advantage of crystallographic symmetry. If the sample does not exhibit crystallographic symmetry, the reciprocal information is not easy to obtain directly from a set of diffraction images. This means there is no way to resolve 3D information of the sample using conventional TEM tomography.

In nanoscience and nanotechnology, obtaining 3D information on nanoparticles becomes crucial for revealing the physical properties of nanoscale materials. One reason for the interest in 3D information of nanoscale materials arises from quantum size effects in semiconductor quantum dots which depend on the particle size and 3D shape [6]. In metal nanoparticles, size and 3D morphology are important because they may have more effect on electronic and chemical behavior of materials than their structure and composition [7]. For example, gold in its bulk state exhibits no catalytic activity, while nanoparticle gold is a good catalyst for selective oxidation of hydrocarbons and the complete combustion of carbon monoxide in air [8, 9, 10, 11, 12]. Also, the spatial distribution of nanoparticles and spacing between nanoparticles can control collective ensemble behavior due to various cooperative effects. Therefore obtaining 3D information on nanoparticles can be desirable.

So far, the projection illustrated in Fig 1.1 is the main factor that limits the application of conventional TEM techniques for revealing spatial information on a sample. The projection causes spatial information loss along the projected direction in the recorded image. As shown in Fig 1.1, the distance between the largest sphere and the smallest sphere cannot be obtained directly from the projected image. As for the spatial distribution of three spheres, conventional TEM techniques provide little information. As a result, conventional 2D TEM techniques are not effective in obtaining 3D information on the nanoscale sample.

These problems can be resolved by electron tomography (ET). ET is a technique which was developed to reconstruct the 3D structure/morphology of a sample from a series of (S)TEM 2D projected images collected as a tilt series [13]. ET was first successfully employed by DeRosier and Klug in 1968 to reconstruct the helical structure of the T4 phage tail from its projections [14]. Since then, the fast development of computation speed and increasing computer memory capacity led to great progress in tomographic practices in biology and medicine [15, 16, 17, 18, 19, 20]. Recently, ET has been applied in nano-materials science [21, 22, 23, 24, 25, 26, 27]. For example, Friedrich *et al.* reconstructed the location, size, distribution of NiO catalyst particles in individual Santa Barbara Amorphous type material (SBA-15 pores) [25]. Arslan *et al.* [24] used ET based on Z-contrast in the STEM mode to obtain the complete 3D size and shape of embedded structures, with a resolution of approximately 1 cubic nanometer. For more details on the development of ET, the reader can refer to the review papers given in Refs [28, 29].

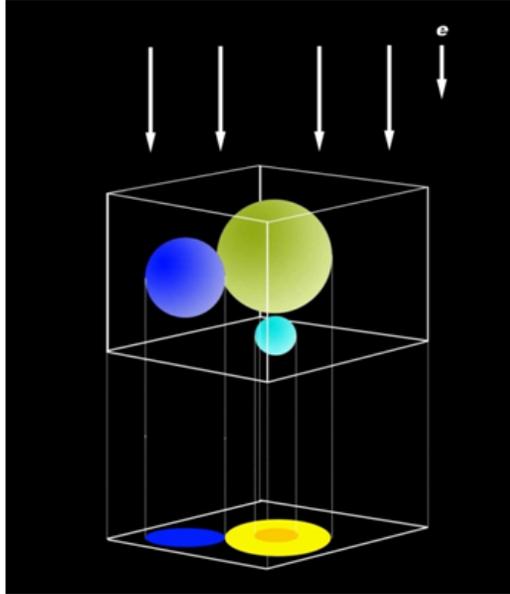


Figure 1.1: An example showing information loss along the incident beam direction. The distance between the largest sphere and the smallest sphere cannot be obtained directly from the projected image, due to the 2D projection nature of the image.

1.1 The Principle of Tomography

Whether it is possible to reconstruct an object from its tilt series projections was initially investigated by an Austrian mathematician Johann Radon in 1917 [30]. He rigorously proved the feasibility of reconstructing an object from a set of projections. His theory for obtaining the 3D reconstructed volume is based on two steps involved in the projection and reconstruction of an object: the “forward Radon transform” that applies to the 2D projection of a 3D object and the inverse Radon transform that can reconstruct the 3D volume from the set of its 2D projections. The Radon transform is at the core of all ET reconstruction methods and deserves some discussion of its basic theory.

1.1.1 Radon Transform

The general mathematical principle of the Radon transform, as proposed by J. Radon [30], is used to describe the relationship between an object in space and its projection along a certain direction. It is defined as the mapping of a function $f(x, y)$ describing a real space object by the projection or line integral through $f(x, y)$ along all possible lines \mathbf{M} with unit length dl :

$$Rf : \quad p_f(s, \varphi) = \int_L f(x, y) dl \quad (1.1)$$

The geometry of the Radon transform is illustrated in Figure 1.2. For the projection step, L is a line parallel to the projected direction. The parameters s , the length of a projection line from the origin of a suitable coordinate system, and φ , the angle between L 's normal line and the x axis, are known and comply with the following relationship:

$$L : \quad s = x \cos \varphi + y \sin \varphi \quad (1.2)$$

A function $p_f(s, \varphi)$ describing the projection of the object $f(x, y)$ along a particular direction is referred to as Radon transform of the function $f(x, y)$. Here, as described before, L is a line within the xy plane which is parallel to the projected direction of the object $f(x, y)$. The defined parameter s and φ describes the projected line L whose projection corresponds to a point on the projection of the object.

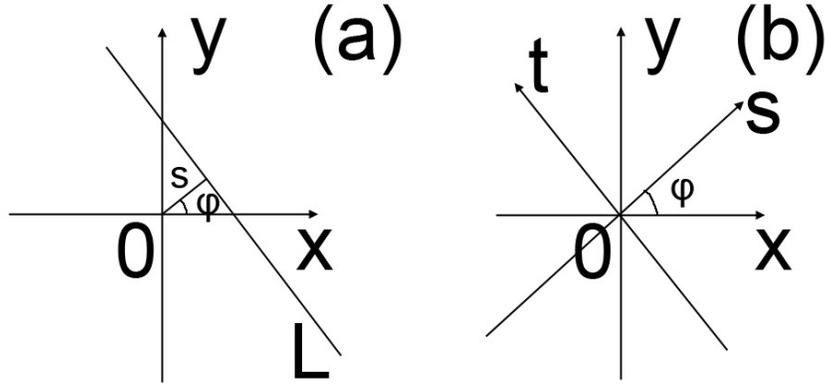


Figure 1.2: (a) Illustration of the angle relation (φ) between the line L and x -axis ; (b) the illustration of the coordinate system xoy and sot .

To facilitate the calculation of the Radon projection of an object, the coordinate system is counterclockwise rotated around the origin, o , by φ degrees to obtain the sot coordinate system shown in Fig. 1.2b. Therefore, the equation of a *projection* line, L , can be expressed as follows:

$$\begin{cases} x = s \cos \varphi - t \sin \varphi \\ y = s \sin \varphi + t \cos \varphi \end{cases} \quad (1.3)$$

Then the Radon transform can be rewritten by combining equation 1.1 and 1.3:

$$p_f(s, \varphi) = \int_{-\infty}^{\infty} f(s \cos \varphi - t \sin \varphi, s \sin \varphi + t \cos \varphi) dt \quad (1.4)$$

The relationship between the Radon transform and the ET reconstruction process can now be examined. Mathematically, the Radon transform represents a general relationship between an object and its projections, and is not specifically limited to three dimensions. In practice, the Radon transform corresponds to the recording process of a 3D object in TEM, where $p_f(s, \varphi)$ is the projection data (*i.e.* standard TEM or STEM images), φ is the projected angle or direction, and s is related to the position of a pixel on a detector such as charge coupled device or camera film.

1.1.2 Inverse Radon Transform

In electron tomography, retrieval of the three-dimensional object morphology (*i.e.* its three dimensional density distribution) from a set of projected images is of primary interest. Mathematically, the ET reconstruction is equivalent to obtaining the integrand from the line integrals of a function along different projection directions. This process is often referred to as the inverse Radon transform. The precise formula describing inverse Radon transform was also first derived by Johann Radon [30].

If a generalized function $f(x, y)$ satisfies the following conditions:

- (1) $f(x, y)$ is continuous;
- (2) $\frac{f(x, y)}{\sqrt{x^2 + y^2}}$ is integrable;
- (3) for an arbitrary point $P = (x, y)$ and each $r \geq 0$.

Let

$$\overline{f(x, y; r)} = \frac{1}{2\pi} \int_0^{2\pi} f(x + r \cos \varphi, y + r \sin \varphi) d\varphi \quad (1.5)$$

so that for any arbitrary point $P(x, y)$, the following equation is valid:

$$\lim_{r \rightarrow \infty} \overline{f_p(r)} = 0$$

where $\overline{f(x, y; r)}$ stands for the average value of the integral of $f(x, y)$ along the circumference of a circle with radius r (the origin of the circle is located at $P(x, y)$).

The value of a generalized function $f(x, y)$ is uniquely determined by its Radon transform and can be calculated using the following formula:

$$f(x, y) = -\frac{1}{\pi} \int_0^\infty \frac{d(\overline{Rf(x, y; r)})}{r} \quad (1.6)$$

Here, $\overline{Rf(x, y; r)}$ is the average value of the integral of $f(x, y)$ along the tangents $p = x \cos \varphi + y \sin \varphi + r$ of the circle with center $P(x, y)$ and radius r . The expression for $\overline{Rf(x, y; r)}$ is:

$$\overline{Rf(x, y; r)} = \frac{1}{2\pi} \int_0^{2\pi} [Rf](x \cos \varphi + y \sin \varphi + r, \varphi) d\varphi \quad (1.7)$$

Combining equation 1.7 into equation 1.6 and adopting polar coordinates, the analytical expression for $f(x, y)$ can be obtained as follows:

$$\overline{f}(r, \theta) = \frac{1}{2\pi^2} \int_0^\pi \int_{-\infty}^\infty \frac{1}{r \cos(\theta - \varphi) - s} p'_f(s, \varphi) ds d\varphi \quad (1.8)$$

here, $p'_f(s, \varphi)$ is the s partial derivative of $p_f(s, \varphi)$ and $\overline{f}(r, \theta)$ is $f(x, y)$ in polar coordinates. Therefore the density of the object ($\overline{f}(r, \theta)$) can be attained through the inverse Radon transform ($p_f(s, \varphi)$) of the actual projected data in Radon space using equation 1.8.

As a summary of the Radon transform and the inverse Radon transform, the inverse Radon transform theoretically proves that the reconstruction of an object from its projections is practical if the projection process satisfies the Radon transform conditions. The

Radon transform and inverse Radon transform provide the theoretical basis of electron tomography. Radon theory guarantees that the reconstruction result is unique, which implies that reconstructing an object from its projections is perfectly feasible.

However, because the projected angle is not continuous (*i.e.*, in practice one does not have an infinite number of images in tilt series), it is impossible to directly operate the inverse Radon transform on the experimental projected data. In fact, Radon theory cannot be directly applied toward object reconstruction in practical electron tomography.

In order to deal with the lack of continuity in the projection angle, different methods have been proposed to realize the inverse Radon transform from a non-continuous tilt series. These methods include filtered back-projection (FBP), real-space iterative techniques and techniques based on Fourier interpolation. Since the FBP method and the simultaneous iterative reconstruction technique (SIRT) (classified in the category of the real-space iterative techniques) method are common and employed in our research, we focus on discussing these two methods.

For both the FBP method and SIRT method, the prerequisite to obtain the reconstruction is that the projection process satisfies the Radon transform and the projected function satisfies the conditions required by the inverse Radon transform (as described in Section 1.1.2). In practice, this means that the intensity of the object's projected image intensity must be linearly related to the imaged properties of the object.

1.2 The Filtered Back-Projection (FBP)

1.2.1 The simple back-projection algorithm

Before discussing the details of the FBP method, an intuitive reconstruction example is introduced which forms the basis of FBP. This method is referred to as a back-projection or summation technique. The technique can be illustrated using the example of a simple 2D binary object consisting of two dot-shaped regions with intensity value 1 inside and 0 outside. This object will be reconstructed from its 1D projection as shown in Fig 1.3.

The object appears in the projections shown here at three different angles Θ_1 , Θ_2 and Θ_3 . The first step in the reconstruction process is to extend the projections along the corresponding direction to obtain the so-called back-projection bodies (B) that have the same intensity of the projected point on the line (P), as shown in Fig 1.3b, 1.3c, and 1.3d. Then, the reconstruction can be obtained by summing all the back-projection bodies, as shown in Fig 1.3e. In Fig 1.3e, the shape of the reconstructed object can be obtained which is marked by two red circles. Usually, the result is normalized (divided) by the number of

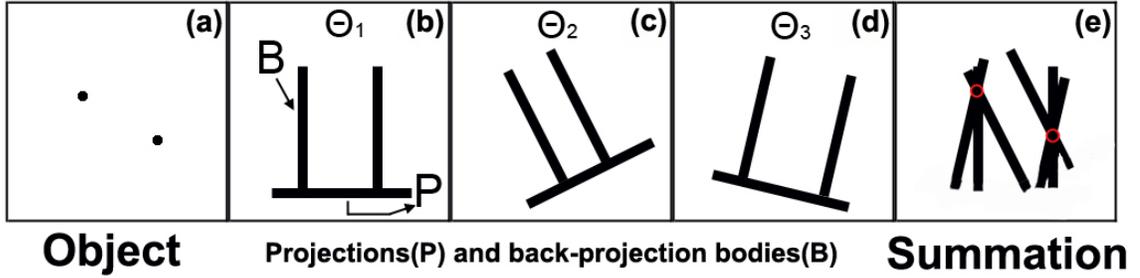


Figure 1.3: An illustration of the simple back-projection algorithm. (a): the object consists two dot-shape disks. The line P in (b) represents the projection of the object, which only contains two projected points of the object for this example. (b)(c)(d): the back-projected body (B) is illustrated along three projected angle (Θ_1 , Θ_2 , and Θ_3). The back-projected body is a line perpendicular to the projected line P. The intersection of B and P is the projected points. The width of the back-projected body is the diameter of the projected dot-shape object. (e): the summation is to sum/overlap the back-projected bodies at different angles. The result of the summation shows two dot-shape disks marked by two red circles. Revised from Ref [1].

the projections in order to keep the intensity of the reconstructed result close to that of the original object.

As shown in Fig 1.3e, the simple back-projection method results in an approximation of the original object. Two dot-shaped regions can be identified as marked by two red circles in the back-projection because of their higher intensity compared to the surrounding background. However, some points which have originally no intensity will have intensity after the simple back-projection (Fig 1.3e) in the reconstruction. This situation results in an asterisk-like artifact as shown in the reconstruction in Fig 1.3e (part outside of the red circles). The reason for this artifact is that, in the simple back-projection method, the intensity of the back-projection body along the back-projected direction is assumed to be the same. In reality, this assumption is not correct.

Next, the mathematical origin of the asterisk-like artifact is investigated. For simplicity, one can assume that there is only a point $\delta(x_0, y_0)$ at the origin in an x - y plane as shown in Fig 1.4. This means that the density ($f(x, y)$) of the object is equal to $\delta(0, 0)$. We take two steps to simulate the back-projection process:

- (1) The first step is to obtain the projection of a point (P). To obtain the projections of the point is equivalent to rotating the coordinate system around its origin and keeping the projection direction along y_r -axis of the rotating coordinate system. The

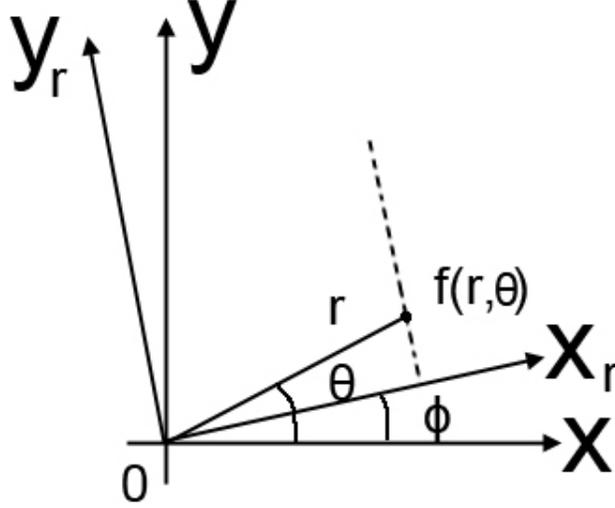


Figure 1.4: An illustration of the relationship between then different coordinate systems. (x, y) is the fixed coordinate; (x_r, y_r) is the rotating coordinate; (r, θ) is the polar coordinate.

rotated coordinate system is defined as $x_r o y_r$. Its origin coincides with xoy 's, which ensures the perfect alignment of the coordinate system for the reconstruction process. Therefore the projected direction/angle ϕ can be determined. The position of the projected line/back-projected line can be completely located using (x_r, ϕ) . Fig 1.4 illustrates the geometrical relation between the original coordinate system (xoy and (r, θ) polar coordinate system) and rotated coordinate system. Any point in space can be represented using (x, y) , (x_r, y_r) and (r, θ) in the polar coordinate system.

In practice, ϕ is a discrete quantity. If we assume the projection angle is ϕ_1 , the corresponding projection is:

$$\begin{aligned}
 p_{\phi_1}(x_r) &= p(x_r, \phi_1) = \int_{-\infty}^{\infty} f_r(x_r, y_r) dy_r = \int_{-\infty}^{\infty} \delta(x_r, y_r) dy_r \\
 &= \int_{-\infty}^{\infty} \delta(x_r) \delta(y_r) dy_r = \delta(x_r) |_{\phi=\phi_1} = \delta[r \cos(\theta - \phi_1)]
 \end{aligned} \tag{1.9}$$

Thus, for an arbitrary angle ϕ_n , there exists $p_{\phi_n} = \delta[r \cos(\theta - \phi_n)]$ where $n = 1, 2, \dots, N$ and N is the number of the total projected images.

- (2) The second step is to reconstruct the point P from 1D back to 2D. According to the simple back-projection algorithm, the reconstructed point (x_r, y_r) can be expressed as follows in the polar coordinate system:

$$\begin{aligned}
 f(r, \theta) = f_r(x_r, y_r) &= \frac{1}{N} \sum_{i=1}^N p_{\phi_i}(x_r) = \frac{1}{N} \sum_{i=1}^N p_{\phi_i} [r \cos(\theta - \phi_i)] \\
 &= \frac{1}{\pi} \sum_{i=1}^N p_{\phi_i} [r \cos(\theta - \phi_i)] \Delta\phi \quad (1.10)
 \end{aligned}$$

The physical meaning of Eq. 1.10 is that the intensity of a point in the reconstructed result is equal to the summation of the intensity of all projected lines containing the point. If the projected lines are infinite, *i.e.*, the projected angle is continuous, the expression of the back-projection can be obtained from equation 1.10:

$$f(r, \theta) = \frac{1}{\pi} \int_0^\pi p_\phi [r \cos(\theta - \phi)] d\phi \quad (1.11)$$

where the integral range is $[0, \pi]$.

For only a single point object, as assumed here, the reconstructed image ($h(r, \theta)$) of a point is:

$$h(r, \theta) = \frac{1}{\pi} \int_0^\pi \delta[r \cos(\theta - \phi)] d\phi \quad (1.12)$$

As we know, for a δ -function, $\delta[\alpha(\phi)] = \delta(\phi - \phi_0)/\alpha'(\phi_0)$ where ϕ_0 is the solution of $\alpha(\phi) = 0$. Let $\alpha(\phi) = r \cos(\theta - \phi)$, then:

$$\alpha(\phi_0) = r \cos(\theta - \phi_0) = 0 \quad \text{or} \quad \sin(\theta - \phi_0) = 1 \quad (1.13)$$

Therefore:

$$\begin{aligned}
h(r, \theta) &= \frac{1}{\pi} \int_0^\pi \delta[r \cos(\theta - \phi)] d\phi = \frac{1}{\pi} \int_0^\pi \frac{\delta(\phi - \phi_0)}{|r \sin(\theta - \phi_0)|} d\phi \\
&= \frac{1}{\pi} \frac{1}{|r \sin(\theta - \phi_0)|} = \frac{1}{\pi} \frac{1}{|r|} = \frac{1}{\pi} \frac{1}{\sqrt{x^2 + y^2}}
\end{aligned} \tag{1.14}$$

Equation 1.14 shows that the reconstructed result ($h(r, \theta)$) of the point $P \delta(x, y)$ is not a δ -function. At $r = 0$ where the original point is located, the intensity of the reconstruction result is infinite. The intensity $h(r, \theta)$ is supposed to decrease with r ($r \neq 0$). The intensity change results in the appearance of the asterisk-like artifact in practice. The Equation 1.14 quantitatively describes the origin of the asterisk-like artifact for the simple back-projection algorithm. This shows the system/algorithm of reconstruction corresponding to the simple back-projection method is not perfect and thus results in artifacts in the reconstructed volume.

1.2.2 The Algorithm of the Filtered Back Projection (FBP)

This section discusses modifications of the simple back-projection method that partially corrects the asterisk-like artifacts in the reconstructed volume. The correction of these artifacts is achieved by suitably filtering the projected data (images). In order to obtain a reliable reconstruction, it is practical to correct the projected data before the back-projecting step, then carry on the back-projection from the corrected (filtered) projected data. The process of correcting the projected data is typically done in Fourier space, using a weighting filter on the various spatial frequencies present in the image. This method is referred to as “filtered back-projection” (FBP).

Before discussing the algorithm of FBP in detail, the Fourier slice theorem needs to be introduced since it provides the mathematical basis of the FBP. The Fourier slice theorem states that the one-dimensional Fourier transform $F(\rho, \phi)$ of a parallel projection $p_\phi(x_r)$ of the object $f(x, y)$ at angle ϕ is equal to a slice of the 2D Fourier transform $F(\omega_1, \omega_2) = \hat{F}(\rho)$ of the object. In other words, the Fourier transform of $p_\phi(x_r)$ gives the values of $F(\omega_1, \omega_2)$ subtending an angle ϕ with the ω_1 axis in Fig 1.5.

The Fourier slice theorem can be expressed mathematically as:

$$\mathbf{F}_1[p_\phi(x_r)] = \hat{F}(\rho, \phi)|_{\phi=\arctan(\omega_2/\omega_1)} = F(\omega_1, \omega_2, \phi)|_{\phi=\arctan(\omega_2/\omega_1)} \tag{1.15}$$

here $\mathbf{F}_1[\bullet]$ stands for a one-dimensional Fourier transform.

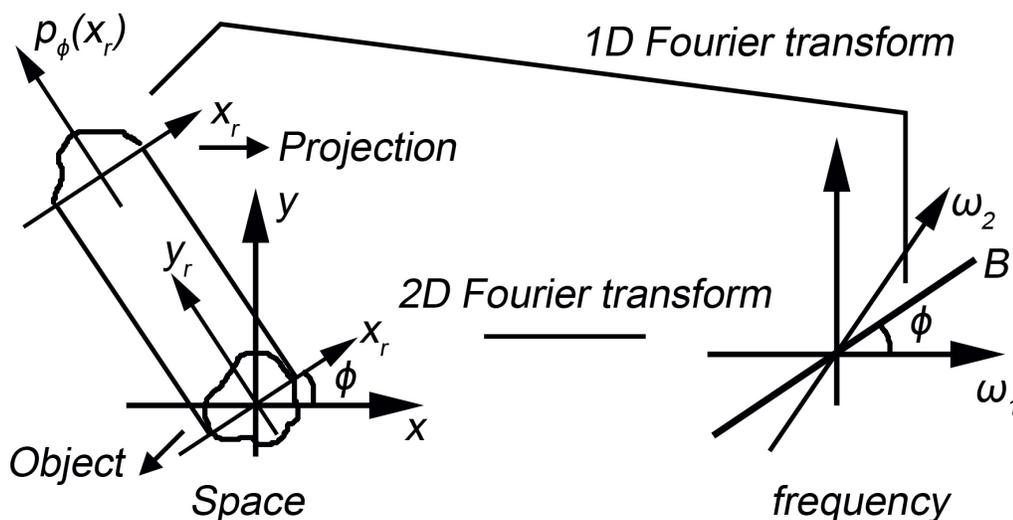


Figure 1.5: An illustration of the Fourier slice theorem. The 2D object in the x - y plane is projected along the y_r direction to the x_r line. $p_\phi(x_r)$ is the intensity of the object projected onto the line x_r . The coordinate system for the 2D Fourier transform of the object is $\omega_1\omega_2$. The line B in Fourier space is the 1D Fourier transform of $p_\phi(x_r)$. Revised from Ref [31].

The details of the mathematical proof of the Fourier slice theorem can be found, for example, in Ref [31]. Here, just the main ideas are briefly summarized. In Fig 1.5, the projected angle is ϕ and the system coordinates, (x_r, y_r) and (x, y) , have the following relation:

$$\begin{cases} x_r = x \cos \phi + y \sin \phi \\ y_r = -x \sin \phi + y \cos \phi \end{cases} \quad (1.16)$$

ω_1 and ω_2 are not independent satisfying the formula 1.17

$$\begin{cases} \omega_1 = 2\pi\rho \cos \phi \\ \omega_2 = 2\pi\rho \sin \phi \end{cases} \quad (1.17)$$

Therefore, according to the Fourier slice theorem, the reconstruction of the projected

images can be realized by the following steps (for simplicity, we take a 2D object projected to 1D situation as an example):

- collect the projected data at different angles (theoretically the range of angle should be continuous from 0° to 180°);
- perform the 1D Fourier transform of all projected data at different angles;
- obtain the 2D Fourier transform of the object by aligning the 1D Fourier transform data from 0° to 180° in a 2D Fourier plane;
- reconstruct the object by applying the 2D inverse Fourier transform.

As described above, the Fourier projection theorem gives a method for obtaining tomographic reconstructions from projected data sets. We now move on to discuss the principle and implementation of *filtered* back projection (FBP) by applying the Fourier slice theorem. As the name implies, the FBP algorithm can be considered as two independent steps: the filtering step, which can be visualized as a simple weighting of each projection in the frequency domain (Fourier space), and the back-projection step, which is equivalent to finding the elemental reconstructions. The Fig 1.6 is used to illustrate the process of FBP and the relationship of coordinate systems used in FBP method.

Let's consider an object $a(x, y)$ to be reconstructed and its 2D Fourier transform $A(\omega_1, \omega_2) = \hat{A}(\rho, \theta)$. For simplicity, we again consider the case of 1D projections of a 2D object. The generalization for a 2D to 3D case is straightforward. According to the Fourier slice theorem, $\hat{A}(\rho, \theta)$ can be obtained by a 1D Fourier transform of the projections $p_\phi(x_r)$ of the object $a(x, y)$ at different angles ϕ , which is:

$$A(\omega_1, \omega_2) = \hat{A}(\rho, \theta) = \mathbf{F}_1[p_\phi(x_r)] = P_\phi(\rho) = P(\rho, \phi) \quad (1.18)$$

The image to be reconstructed can be expressed as:

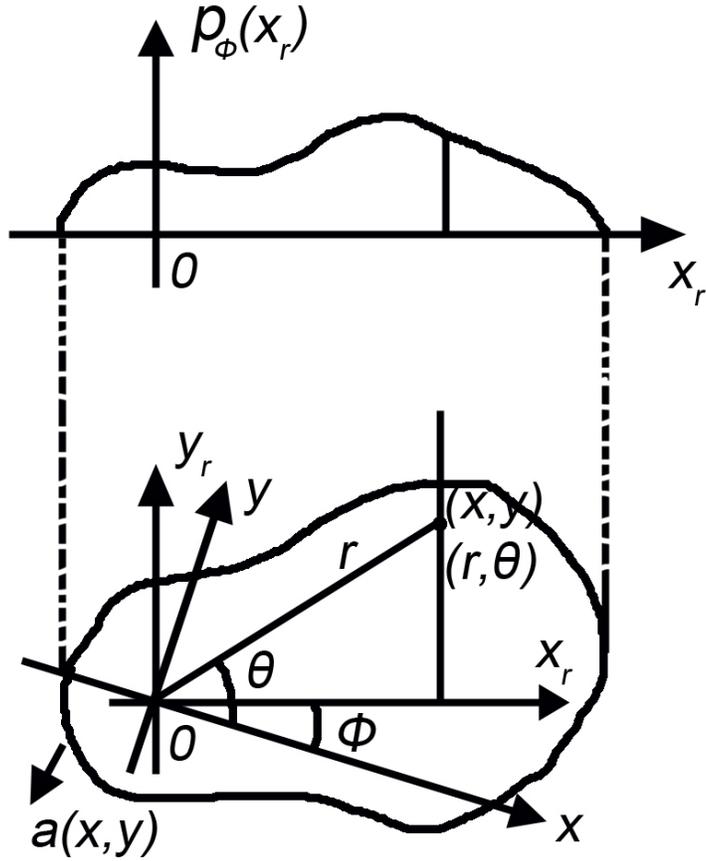


Figure 1.6: The illustration of the relationship of coordinate systems used in FBP method. Revised from Ref [31].

$$\begin{aligned}
\hat{a}(r, \theta) = a(x, y) &= \mathbf{F}_2^{-1}[A(\omega_1, \omega_2)] = \frac{1}{4\pi^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} A(\omega_1, \omega_2) e^{j(\omega_1 x + \omega_2 y)} d\omega_1 d\omega_2 \\
&= \int_0^\pi \int_{-\infty}^{\infty} |\rho| \hat{A}(\rho, \theta) e^{j2\pi \rho r \cos(\theta - \phi)} d\rho d\phi \\
&= \int_0^\pi \int_{-\infty}^{\infty} |\rho| P(\rho, \phi) e^{j2\pi \rho r \cos(\theta - \phi)} d\rho d\phi \\
&= \int_0^\pi \int_{-\infty}^{\infty} |\rho| P(\rho, \phi) e^{j2\pi \rho r \cos(\theta - \phi)} d\rho d\phi \\
&= \int_0^\pi d\phi \int_{-\infty}^{\infty} |\rho| P(\rho, \phi) e^{j2\pi \rho r \cos(\theta - \phi)} d\rho
\end{aligned} \tag{1.19}$$

The second integral in 1.19 $\int_0^\pi \int_{-\infty}^\infty |\rho| p(\rho, \phi) e^{j2\pi\rho r \cos(\theta-\phi)} d\rho d\phi$ can be rewritten as the inverse spatial Fourier transform of variable x_r , based on the definition of the convolution:

$$\begin{aligned}
\int_{-\infty}^\infty |\rho| P(\rho, \phi) e^{j2\pi\rho r \cos(\theta-\phi)} d\rho &= \int_{-\infty}^\infty |\rho| P(\rho, \phi) e^{j2\pi\rho x_r} \Big|_{x_r=r \cos(\theta-\phi)} d\rho \\
&= h(x_r) * p(x_r, \phi) \Big|_{x_r=r \cos(\theta-\phi)} \\
&= g(x_r, \phi) \Big|_{x_r=r \cos(\theta-\phi)} \\
&= g[r \cos(\theta - \phi), \phi]
\end{aligned} \tag{1.20}$$

here $g(x_r, \phi) = p(x_r, \phi) * h(x_r)$; $h(x_r) = \mathbf{F}_1^{-1}[|\rho|]$ and $p(x_r, \phi) = \mathbf{F}_1^{-1}[P(\rho, \phi)]$. The physical meaning of equation 1.19 and 1.20 is that the projection $p(x_r)$ is filtered by the transfer function $|\rho| = \mathbf{F}[h(x_r)]$ to obtain the value of the corrected/filtered projection $g(x_r, \phi)$ corresponding to the projected line/angle $x_r = r \cos(\theta - \phi)$ passing through the point (r, θ) .

Put equation 1.20 into equation 1.19:

$$\hat{a}(r, \theta) = \int_0^\pi g[r \cos(\theta - \phi), \phi] d\phi \tag{1.21}$$

Formula 1.20 describes the filtering process, which is the step 1 in FBP. The above formula 1.21 means that the value of the summation/integral of all projections of the point (r, θ) after filtering gives the value $(\hat{a}(r, \theta))$ of a point after reconstruction. This is step 2 in a FBP.

1.3 The Iterative Algorithm

An approach different from Fourier transform-based methods makes use of the iterative algorithms. The main concept underlying the iterative algorithms is to assume that the cross section of the reconstructed object is composed of an array of unknown matrices. Algebraic equations are then set up for the unknown matrices (the object) in terms of the recorded projection data (the experimental images). The reconstruction result is obtained by resolving the algebraic equations. The process of resolving the algebraic equations is usually an iterative operation. Therefore the method of resolving algebraic equations is referred to as an iterative method.

Although conceptually this method is much simpler than the FBP method discussed previously, iterative methods have some disadvantages. The algorithm focuses on solving a

series of equations so that the solving process of the equation group lacks a corresponding physical process. Therefore, artifacts appearing in the reconstruction can be hard to interpret. In addition, iterative methods usually require a long computation time to obtain the satisfactory reconstructed results compared to FBP method.

Of course, iterative methods have advantages as well. For situations in which it is impossible to obtain a large number of projected images, or the projected angles are not uniformly distributed over 180° , the FBP method cannot provide reconstructed results with high accuracy since both these conditions strongly affect the quality of the reconstruction of FBP method. Fortunately iterative techniques can be employed under these conditions. A further advantage of iterative techniques is the ability to apply prior knowledge to the original solutions in order to attain the best solutions, such as the boundary confinement of the object [32].

This section is divided into two parts: in the first part, we demonstrate how to set up the linear equations describing the relationship between the density of the object and its projected images. We describe how to construct a set of linear equations with variables representing density cross sectional density of the object. The second part presents the Kaczmarz method [33] for solving these equations (referred as the reconstruction process) and introduces the algebraic reconstruction techniques (ART).

1.3.1 The Algebraic Representation of the Cross Section of the Object and its Projection

Presenting the projection process of an object using a matrix relation or equation group is the basis of the iterative methods (or the first step). The way of presenting the projection process determines parameters of the equation group affecting the reconstructed result. Therefore, the method used to quantify the projection process is critical. Here, a graphic procedure is adopted to illustrate the quantification of the projection process, which is a good way to connect the projection process with the equation group.

The operation of projection quantification is shown in Fig 1.7. First, the spatial discretization is applied on the tomographic volume $f(x, y)$. This means that we superimpose a square grid on the image $f(x, y)$ and assume that the pixel value of $f(x, y)$ is constant in each square. Let x_j denote this constant value in the j -th square and N be the total number of pixels. Then, for the iterative method, the ray running through the (x, y) plane has a certain width τ along a certain angle to obtain one projection of the image. All projections can be obtained by the ray running through the recording tilt angles. The i -th ray is shaded in Fig 1.7 in order to illustrate this line well. The width of each ray is

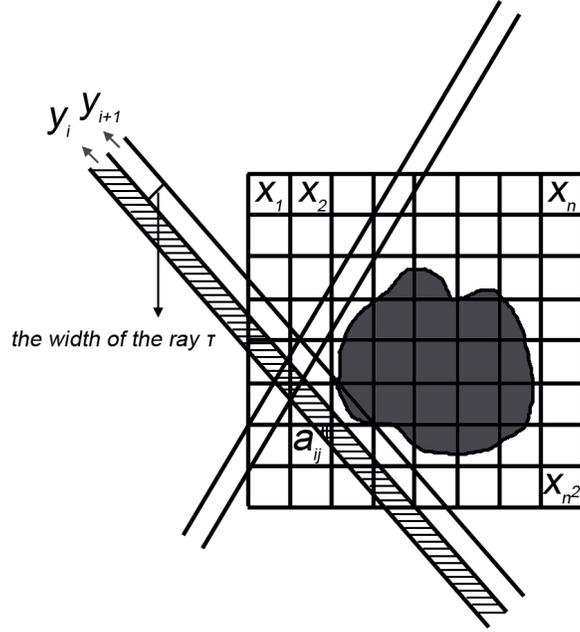


Figure 1.7: In algebraic methods a square grid is superimposed over the volume that is being reconstructed. The value (intensity) within the volume is assumed to be constant within each cell (voxel) of the grid denoted by x_i . The projection process is decomposed into a series of projected ray y_i with constant width τ . The line integral of the projection of the object is called a *ray-sum*. The weighting factor a_{ij} is equal to the area of the grid superimposed by the ray. Revised from Ref [31].

approximately equal to the image square width. In this case, the line integral is referred to as *ray-sum* corresponding to the intensity of a point on the projected line for 2D situation (or the projected plane for 3D).

For convenience, it is customary to adopt a one-index representation for the projection data. For example, y_i is the ray-sum of the i -th ray as shown in Fig 1.7. Then the relationship of x_j and y_i can be expressed as:

$$\sum_{j=1}^N a_{ij} x_j = y_i, \quad i = 1, 2, \dots, M \quad (1.22)$$

where M is the total number of rays and a_{ij} is the weighted factor that represents the

contribution of the j th square to i th ray-sum. The value of the weighting factor is equal to the fractional area of the i th ray running through the j th pixel as marked in Fig 1.7. For a given ray-sum in practice, most of a_{ij} are often equal to zero since only a small number of squares have a contribution from the sample while many are passing through vacuum.

After describing the projection process, the next step in the iterative method is how to resolve the equations of the projection to reconstruct the object. If M and N were small, the equations 1.22 can be solved using conventional matrix theory. However, in practice, N is very large and direct matrix inversion is not possible. For the case of large M and N , an alternative method based on the method of projections was first proposed by Kaczmarz [33] and later further improved by Tanabe [34]. In order to illustrate the computational steps of these alternative methods, equation 1.22 is rewritten in an expanded form:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1N}x_N = y_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2N}x_N = y_2 \\ \vdots \\ a_{M1}x_1 + a_{M2}x_2 + \cdots + a_{MN}x_N = y_M \end{cases} \quad (1.23)$$

Equations can be understood as follows: Let \mathbf{x} be represented as $x = (x_1, x_2, \dots, x_N)^T$. Mathematically, the image represented by \mathbf{x} can be considered as a generalized point in an N -dimensional space. In this N -dimensional space, each of the equations in 1.23 represents a hyperplane. If a unique solution of these equations exists, the intersection of all these hyperplanes is a single generalized point and furthermore the intersected generalized point coincide with the solution of these equations.

This concept is further illustrated in Fig 1.8. For the purpose of the convenience of formulation, one only considers the situation of two projections y_1 and y_2 that satisfy the following equations:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 = y_1 \\ a_{21}x_1 + a_{22}x_2 = y_2 \end{cases} \quad (1.24)$$

here a_{11}, a_{12}, a_{21} and a_{22} are the weighting factors as defined by the equation 1.22.

The search procedure for the solution of algebraic equations in Fig 1.8 starts with an initial guess x^0 , projecting x^0 on the first line y_1 , re-projecting the resulting point x^1 onto the second line y_2 to obtain the solution of two equations corresponding to one iterative

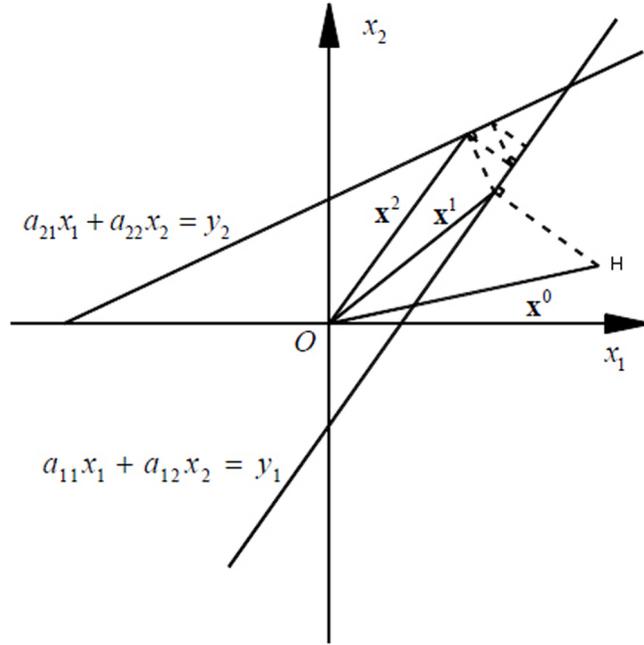


Figure 1.8: The illustration of the concept of solving algebraic equations with two equations. One starts with an arbitrary initial guess (point H noted using x^0) and then projects onto the line y_1 corresponding to the first equation to obtain projected point x^1 . The projected point (x^1) on the line y_1 is then projected onto the line of the second equation y_2 to obtain the projected point x^2 . The obtained projected point x^2 is the result of two equations (y_1 and y_2) corresponding to one iterative loop. If necessary, the operation is repeated as illustrated by treating the obtained solution (or point) as the initial guess in the next loop until the solution is satisfactory. Revised from Ref [31].

loop. If necessary, the operation is repeated by treating the obtained solution (or point) as the initial guess in the next loop until the solution is satisfactory. If the solution exists, the result of the iterations will eventually converge to a point for two equations situation, which is the desired solution.

We apply this concept in solving equations 1.23 with a computer. To realize this procedure using a computer, one first starts with an initial guess $x^0 = (x_1^0, x_2^0, \dots, x_N^0)$. In most cases, x^0 is set to be zero and is represented as $x^0 = (0, 0, \dots, 0)$. x^0 is then projected on the first equation in 1.23 generating x^1 . x^1 is projected on the hyperplane represented by the second equation in 1.23 to give x^2 and so forth. When x^{i-1} is projected on the hyperplane represented by the i -th equation to yield x^i , this process can be mathematically

expressed as:

$$x_j^i = x_j^{i-1} - \frac{A_i x_j^{i-1} - y_i}{A_i \cdot A_i} A_i \quad (1.25)$$

here $A_i = (a_{i1}, a_{i2}, \dots, a_{iN})$ is the weighting factor of the i th ray and $A_i \cdot A_i$ is the dot product of A_i with itself.

As mentioned above, the computational process for the iterative algorithm is as follows: assuming an initial solution, taking continuous projection on the hyperplanes represented by the equations in 1.23 to eventually obtain x^M by formula 1.25. This is the first iteration. In the next iteration, x^M is treated as the initial solution to repeat the first iteration process, obtaining the second iterative loop result x^{2M} , and so on. A satisfactory solution can be obtained after sufficient iteration. Tanabe [34] shows that, if the equations in 1.23 have an unique solution \hat{x} , then

$$\lim_{k \rightarrow \infty} x^{kM} = \hat{x} \quad (1.26)$$

The time required to solve the equations in 1.23 is related to the angle between the hyperplanes. If the hyperplanes representing the equations in 1.23 are perpendicular to one another and there is a unique solution, only one iteration is needed. However, if the hyperplanes have only a small angle between them, the convergence speed of the algorithm is slow, since more projections are needed to reach the unique point/solution. For instance, in Fig 1.8, a smaller angle between line y_1 and line y_2 would need more projections for the point H to reach the intercept point of line y_1 and line y_2 , which is the solution of the equation group (y_1 and y_2). Thus a longer computation time is required to reach the correct resolution for smaller angles.

Many alternative iterative techniques were developed on the basis of equation 1.25 [13]. Most of them are the approximate expressions of equation 1.25. For the purpose of discussing other iterative techniques, the equation 1.25 can be rewritten as:

$$x_j^i = x_j^{i-1} + \frac{y_i - z_i}{\sum_{k=1}^N a_{ik}^2} a_{ij} \quad (1.27)$$

where

$$z_i = A_i x^{i-1} = \sum_{k=1}^N a_{ik} x_k^{i-1} \quad (1.28)$$

Eq. 1.27 shows that when the solution is projected on the i -th hyperplane in the $(i-1)$ -th iteration, the intensity value of the j -th square (at this step, the intensity is x_j^{i-1}) can be obtained from the difference Δx_j^i :

$$\Delta x_j^i = x_j^i - x_j^{i-1} = \frac{y_i - z_i}{\sum_{k=1}^N a_{ik}^2} a_{ij} \quad (1.29)$$

1.3.2 Algebraic Reconstruction Techniques (ART) and Simultaneous Iterative Reconstruction Technique (SIRT)

With the theory discussed above, the next is to discuss two computer implementations of iterative techniques. They are algebraic reconstruction technique (ART) and simultaneous iterative reconstruction technique (SIRT) respectively.

1.3.2.1 Algebraic Reconstruction Techniques (ART)

As discussed before, the value of weighting factor is crucial for the reconstruction in iterative methods. In many ART methods, the value of the weighting factor a_{ik} in equation 1.29 is simply set to be equal to 1 or 0, depending on whether the center of the k -th square is within the i -th ray. If they are, the weighting factor is equal to 1. Otherwise, it is equal to 0. This saves the computation time of determining the value of a_{ik} . Under this condition, the denominator in equation 1.29 is expressed as $\sum_{k=1}^N a_{ik}^2 = H_i$, equal to the number of the squares whose centers are within the i -th ray. The value of the j -th square corrected by the i -th hyperplane/equation in equation 1.23 could be written as:

$$\Delta x_j^i \approx \frac{y_i - z_i}{H_i} \quad (1.30)$$

for all the squares whose centers are within the i -th ray.

Although the above idea is easily implemented in a computer, the disadvantage is that it leads to artifacts in the reconstructed images, especially in a situation where H_i has a large difference from the actual situation. For example, considering the situation that the

ray just runs through the edge of the square with actual contribution to the projection, the weighting factor is treated as 0 resulting in a reduced contribution of the square to the projection. In order to avoid this situation, the following formula could be used to replace equation 1.30:

$$\Delta x_j^i = \frac{y_i}{L_i} - \frac{z_i}{H_i} \quad (1.31)$$

here L_i is the length of the i th ray through the reconstruction region. However, the approximation mentioned above induces yet another kind of error. This error normally causes the appearance of *salt and pepper noise* in the ART reconstruction [31].

There are other ways to supplement ART. We should clarify that what is commonly referred to as ART is actually the original *Kaczmarz* algorithm. The procedure follows the steps mentioned in last section and the iterative formula employs equation 1.25. Eggermont [35] and Censor [36] proposed a possible way to reduce the noise in ART reconstruction by introducing relaxation coefficient α which is less than 1. In their methods, equation 1.25 is replaced as:

$$x^i = x^{i-1} - \alpha_i \frac{A_i x^{i-1} - y_i}{A_i \cdot A_i} A_i \quad (1.32)$$

where α_i is the relaxation coefficient that usually is a function of iterations which decreases with increasing number of iterations i .

1.3.2.2 Simultaneous Iterative Reconstruction Technique (SIRT)

An alternative method to solve the equations in 1.23 is referred to as simultaneous iterative reconstruction technique (SIRT). The method adopts the Jacobi method of numerical linear algebra. Different from the ART method, the Jacobi method is a parallel iteration; that is, the value of the j th square is not changed until the end of an iteration going through all equations. More specifically, equation 1.29 is adopted to calculate Δx_j^i , but the value of x_j^i is corrected only after all equations are calculated. Therefore this method corrects the value of each square by averaging all the computed changes for that square.

In the SIRT method, based on Jacobi's idea, the iterative steps are: (1) with an initial guess solution $x^0 \subset R^N$ and typically taking $x^0 = (0, 0, \dots, 0)^T$; (2) this guess is then iterated according to:

$$x^i = x^{i-1} + \frac{1}{M} \sum_{i=1}^M \frac{y_i - A_i x^{i-1}}{A_i \cdot A_i} A_i \quad (1.33)$$

The outcome of this iteration is taken as an input for the next cycle of the iterative loop until a suitable convergence criterion is met.

1.4 Introduction of Two TEM Techniques

The transmission electron microscope (TEM) is a remarkably versatile instrument. The interaction of the electron beam with the sample results in a number of possible imaging modes. Several conventional TEM techniques can be used, in principle, for 3D ET. Early examples of the application of electron tomography in materials science adopted bright-field TEM. They were based on studies in the biological sciences of stained polymers and the internal network structure of block copolymers [37]. For example, bright-field TEM was used to investigate the 3D properties and porosity of zeolitic materials [38]. Recently, other imaging modes were also applied for electron tomography, such as EFTEM [39] and HAADF [40] imaging modes.

Although bright field TEM (BFTEM) is a mode frequently used to acquire projected images, an important question is whether the result of 3D ET reconstruction using the bright-field imaging mode is reliable. For both biological and non-crystalline inorganic systems, the application of bright-field TEM for 3D ET is acceptable because mass-thickness contrast generally satisfies the ‘projection requirement’ of the Radon transform that the collected signal should be a monotonic function of a physical property of the sample such as mass thickness [41]. However, in general, BFTEM is not suitable for the study of crystalline materials. The problem is that diffraction contrast does not satisfy the projection requirement and can lead to serious artifacts in the tomographic reconstruction.

As a better alternative, high-angle annular dark-field (HAADF) imaging mode and energy filtered imaging mode (EF-TEM or EF-STEM) are widely employed in obtaining reliable projection data. They strictly satisfy the projection requirement constrained by the Radon transform. Since those two imaging modes are employed in my research, the HAADF and EFTEM methods are discussed in detail below.

1.4.1 High-Angle Annular Dark-Field (HAADF) Imaging Mode

Annular dark-field (ADF) imaging is a method of mapping samples in a scanning transmission electron microscope (STEM). The image is formed by collecting scattered electrons with an annular dark-field detector. If this detector collects high-angle incoherently scattered electrons, and is sensitive to variations in the atomic number of the elements in the sample, this technique is referred to as high-angle annular dark-field imaging (HAADF).

When an electron beam passes through a sample, the scattering angle of most of the elastically scattered electrons is larger than that of the inelastically scattered electrons. Fig 1.9 shows the theoretical simulation of the angular dependence of the differential cross sections for elastic and inelastic scattering of 100 keV electrons from a carbon atom, using the Lenz model [42]. Below an angle of ~ 20 mrad, most of the scattering arises from inelastic interactions. Most of the scattering arises from the elastic scattering when the angle is above ~ 20 mrad.

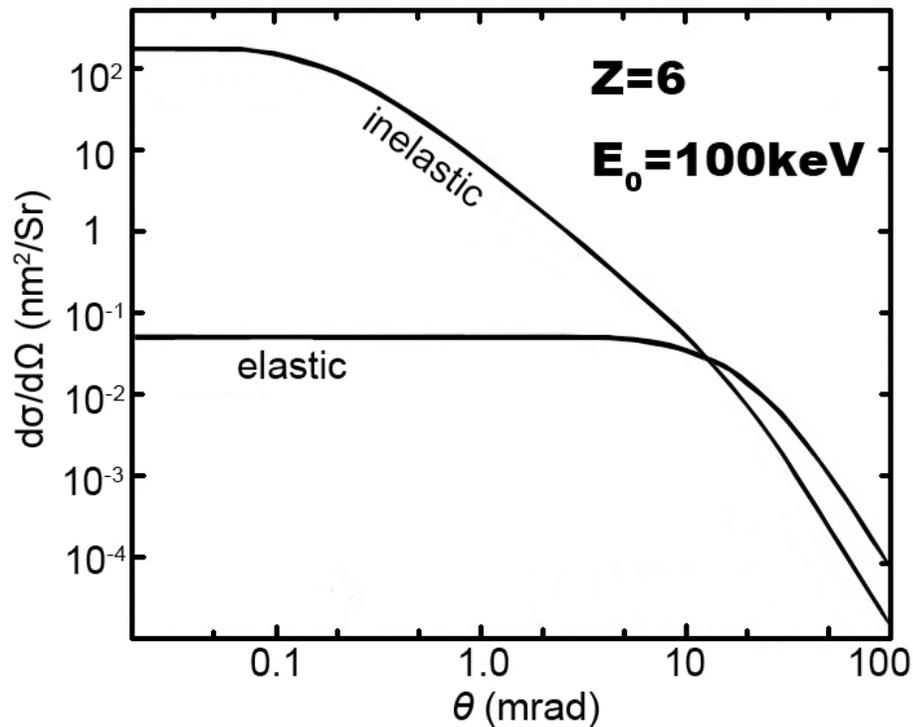


Figure 1.9: Simulated angular dependence of the differential cross sections for elastic and inelastic scattering of 100 keV electrons from a carbon atom, using the Lenz model. Modified from Ref [42].

Using a suitable detector to collect the electrons scattered at a high angle in STEM mode, the image will be formed mainly using elastically scattered electrons. The transmitted electrons and most of the inelastically scattered electrons are not used in this mode. They pass through the center opening in an annular dark field detector in STEM. An annular detector with a large opening is used to collect the high-angle scattered electrons. This is the high-angle annular-dark field (HAADF) imaging mode (Fig. 1.10 and 1.11).

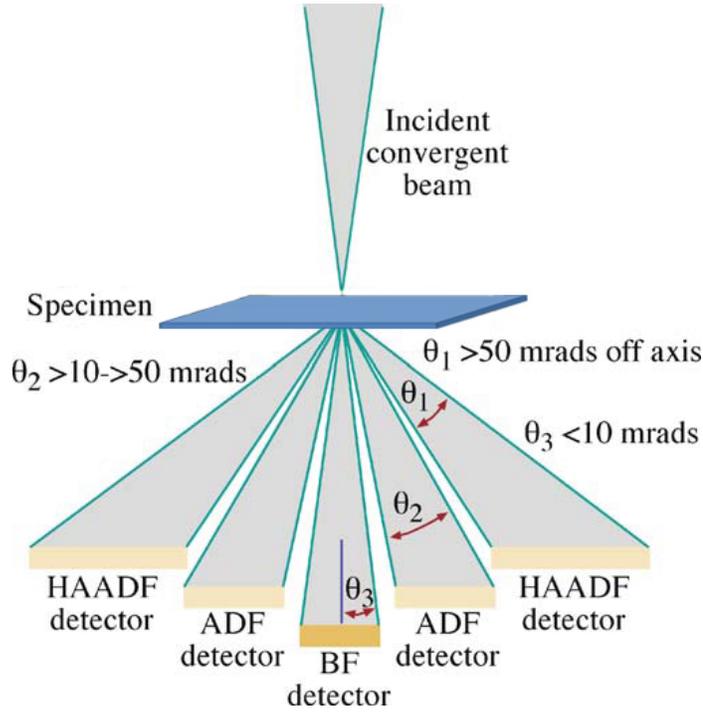


Figure 1.10: Schematic of the common techniques used in STEM mode with the range of electron scattering angles gathered by each detector. Modified from Ref [1].

According to the theory derived by Pennycook *et al.* [44], the scattering cross-section $\sigma_{\theta_1, \theta_2}$ of the annular area between θ_1 and θ_2 can be presented using the integral of the Rutherford scattering intensity from θ_1 to θ_2 :

$$\sigma_{\theta_1, \theta_2} = \left(\frac{m}{m_0}\right) \frac{Z^2 \lambda^4}{4\pi^3 a_0^2} \left(\frac{1}{\theta_1^2 + \theta_0^2} - \frac{1}{\theta_2^2 + \theta_0^2}\right) \quad (1.34)$$

here, m — the mass of the incident electron;

m_0 — the rest mass of an electron;

Z — the atomic number;

λ — the wavelength of the electron;
 a_0 — the Bohr radius;
 θ_0 — the characteristic angle of elastic scattering.

Therefore, the scattering intensity I_s of a sample or the recorded intensity of the sample is presented as:

$$I_s = \sigma_{\theta_1, \theta_2} \cdot NtI \quad (1.35)$$

here, N is the number of atoms per volume unit of the sample, t is the thickness of the sample and I is the intensity of the incident electrons.

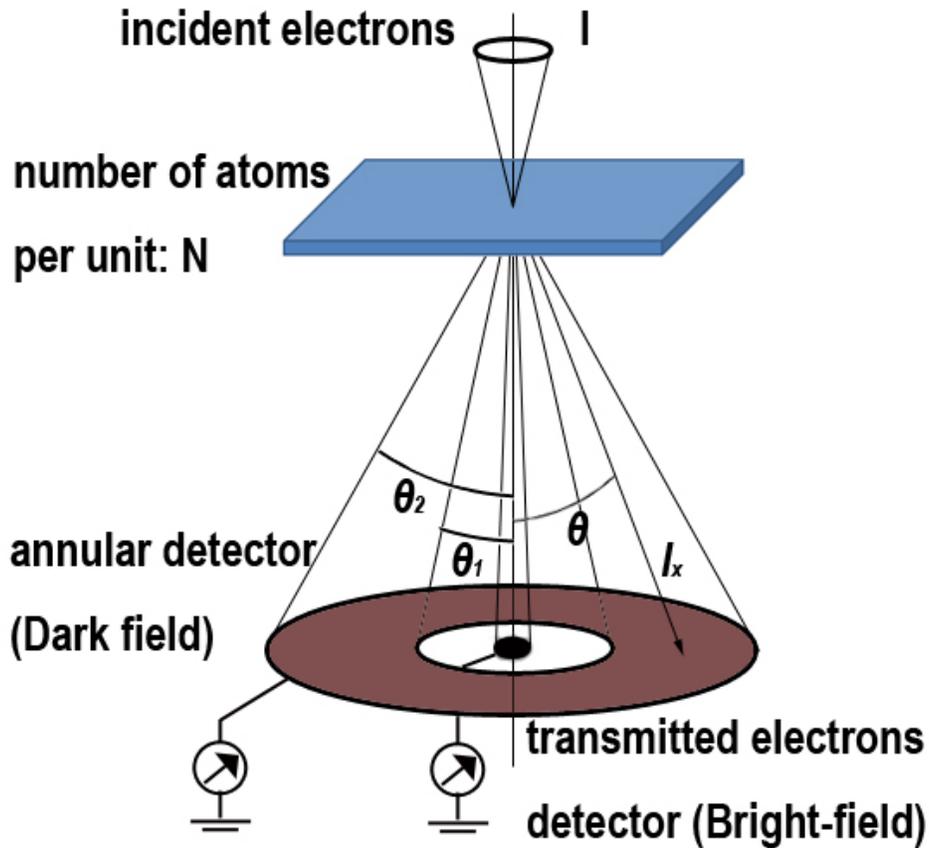


Figure 1.11: Illustration of HAADF technique. The electrons scattered at a high angle are collected using the annular detector. The collecting angle range is equal to $\theta_2 - \theta_1$. If the transmitted electrons are collected, the STEM bright-field image is obtained. I is the intensity of the incident electrons. Revised from ref [43].

According to Eq. 1.34 and Eq. 1.35, the intensity of HAADF is proportional to square

of the atomic number (Z^2). HAADF images are therefore sensitive to the atomic number and chemical composition. Therefore, the HAADF image is also named as Z-contrast or Z^2 -contrast imaging. If the sample thickness is constant, the intensity in HAADF images increases with increasing atomic number. The contrast in HAADF images is approximately monotonic with the thickness of the sample when the sample composition is homogeneous. Furthermore, as described in reference [45], the signal contributed by electrons collected by STEM HAADF detector can be considered to be incoherent. Diffraction and phase contrast that cause concern with the “projection theorem” in BFTEM are virtually absent in HAADF imaging. These properties make the STEM HAADF technique ideal for the application of electron tomography [3].

The earliest application of STEM HAADF tomography was reported in reference [40]. The technique focused on the study of heterogeneous catalysts based on metallic nanoparticles distributed within highly porous siliceous and carbonaceous support structures. In the paper, the STEM HAADF imaging was able to discriminate nanoparticles with approximately 3 nm in diameter from the background support. For comparison, the contrast from the particles imaged using the bright-field TEM technique was quite weak [46].

1.4.2 Energy Filtered Transmission Electron Microscopy (EFTEM) Imaging Mode

As discussed in the preceding section, scattering can be classified according to whether it is elastic or inelastic. Inelastic scattering occurs as a result of Coulomb interaction between a fast incident electron and the atomic electrons that surround each nucleus. The following inelastic scattering processes are of relevance in the TEM: (1) phonon scattering (typically less than 0.1 eV loss); (2) valence electron scattering or plasmon scattering (typically less than 50 eV loss); (3) interband transmission (typically less than 10 eV loss); (4) excitation of inner-shell electrons (more than 13 eV loss); (5) Bremsstrahlung radiation.

By collecting scattered electrons from processes (2) and (4) to form an image, chemically specific information can be obtained. This method, making use of inelastically scattered electrons, is referred to as energy filtered TEM (EFTEM) or energy filtered STEM (EFSTEM), depending on whether the signal was collected in the parallel or scanning mode of the TEM. In EFTEM analysis, energy-selected images are normally recorded using fixed beam illumination, with spectral information integrated over a particular energy-loss range defined by an “energy-selecting slit”. In EFSTEM mode, spatial information is recorded by scanning a small probe over the sample, while spectral (and hence chemical) information

is probed serially by acquisition of a number of images (or pixels) corresponding to different energy losses. Figure 1.12 shows the energy filtered recording mechanism of TEM and STEM modes [47].

Energy filtered STEM recording mode has certain advantages compared to that in TEM mode. An important one is that the scanning time on each point in scanning mode is considerably shorter than that in TEM mode. Accordingly, the total electron dose for the scanning techniques is considerably less than in EFTEM mapping, reducing possible radiation damage to the sample.

The EFTEM mode has advantages in some situations. The spatially parallel nature of EFTEM allows sample drift between individual acquisitions to be corrected with considerably greater ease than for scanning methods. In addition, although the total electron dose on the sample is larger in EFTEM, the beam current density is considerably smaller. Hence dose-rate-dependent damage experienced in scanning probe studies, such as hole-drilling [48], can be less severe in EFTEM than that in EFSTEM.

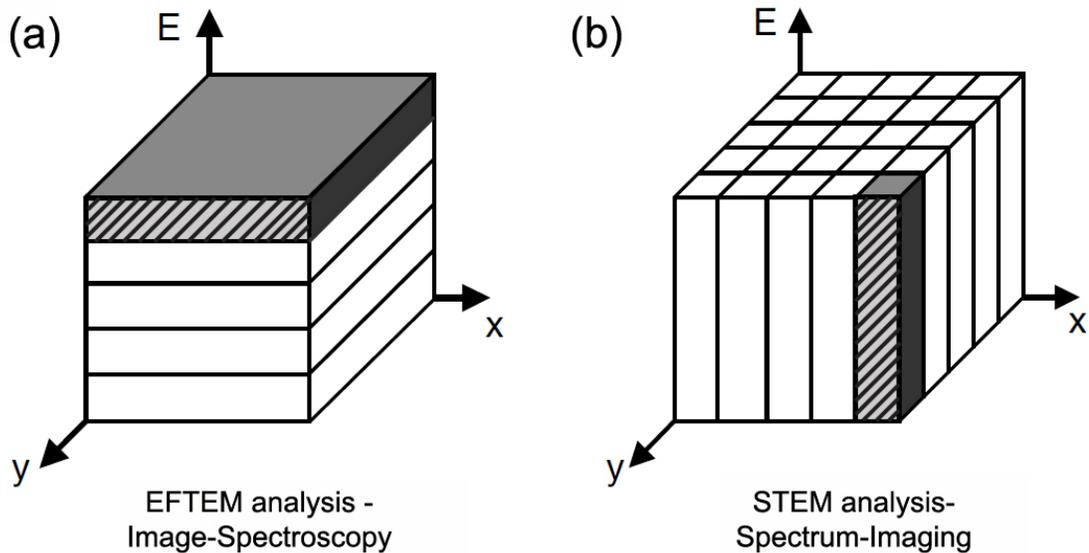


Figure 1.12: Comparison of energy-loss data acquisition using (a) an energy-filtered TEM, and (b) a STEM equipped with a parallel detection spectrometer. Revised from [47].

Both (2) (plasmon loss) and (4) (core loss) scattering electrons can provide chemical information. The former (plasmon energy loss) method will be used here to investigate silicon nanoparticles in an SiO_x matrix. The plasmon signature in the low-energy-loss part

of the spectrum is well suited for distinguishing Si nanoparticles from an amorphous SiO₂ matrix. The Si plasmon is sharply peaked at 16.7 eV with a width at half-maximum of a few eV, while the SiO₂ plasmon is broad and centered around 23 eV. Therefore, the energies of the plasmon peaks of Si and SiO₂ are well separated (about 6 eV) compared to the energy resolution (~ 1 eV) of JEOL 2200FS used in this work.

1.5 Summary

This chapter described the background of ET and two TEM imaging modes used in my research. First of all, the mathematical basis of electron tomography (Radon theory) was introduced. Then, the physical principles of two 3D ET reconstruction methods (FBP and SIRT) used in practice were introduced. FBP is a method based on the Fourier transform. The method is widely employed in practice since the computation time using FBP is short. SIRT is an algebraic iterative method. The essence of the SIRT method is to resolve a set of linear equations. Usually SIRT takes more computing time compared to FBP but SIRT has advantage in the situations in which it is impossible to obtain a large number of projected images, or the projected angles are not uniformly distributed over 180°. Finally, the principles of the HAADF and EFTEM imaging techniques were introduced. Both imaging modes satisfy the projection requirement and are widely applied in material science.

Bibliography

- [1] David B. Williams and C. Barry Carter, *Transmission Electron Microscopy: A Textbook for Materials Science*, 2nd edition, Springer (2009).
- [2] Sakamoto, Y., Kaneda, M., and Terasaki O.. *et al.* *Nature*, 408(6811):449-453 (2000).
- [3] Gao, C. B., Sakamoto, Y., Sakamoto, K., Terasaki, O., and Che, S. N.. *Angew. Chem., Int. Ed.*, 45:4295-4298 (2006).
- [4] Garcia-Bennett, A. E., Miyasaka, K., Terasaki, O., and Che, S. N.. *Chem. Mat.*, 16(19):3597-3605 (2004).
- [5] Han, Y., Zhang, D. L., Chng, L. L., Sun, J. L., *et al.* *Nat. Chem.*, 1(2):123-127 (2009).
- [6] Reed, M.A., Randall J. N., Aggarwal, R. J., Matyi, R. J., Moore, T. M., and Wetsel A. E.. *Phys. Rev. Lett.*, 60(6):535-537 (1988)
- [7] Roduner, E.. *Chem. Soc. Rev.* 35:583-592 (2006).
- [8] Reddy, B. V., Khanna, S. N., and Dunlap, B. I.. *Phys. Rev. Lett.*, 70:3323-3326 (1993).
- [9] Moseler, M., Hakkinen, H., Barnett, R. N., and Landman, U.. *Phys. Rev. Lett.*, 86:2545-2548 (2001).
- [10] Haruta, A.. *Chem. Rev.*, 3:75-87 (2003)
- [11] Chen, M. S., and Goodman, D. W.. *Acc. Chem. Res.*, 39:739-746 (2006).
- [12] Hashmi, A. S. K., and Hutchings, G. J.. *Angew. Chem., Int. Ed.*, 45:7896-7936 (2006).
- [13] Frank J.. *Electron Tomography: Three-dimensional imaging with the Transmission Electron Microscope*. New York: Plenum Press (1992)
- [14] De Rosier, D.J., and Klug, A.. *Nature*, 217:130-134 (1968).

- [15] Sui, H. X., and Downing, K. H. *Nature*, 442:475-478 (2006).
- [16] Nicastro, D., Schwartz, C., Pierson, J., *et al.* *Science*, 313:944-948 (2006).
- [17] Beck, M., Forster, F., Ecke, M., *et al.* *Science*, 306:1387-1390 (2004).
- [18] Beck, M., Lucic, V., Forster, F., Baumeister, W., Medalia, O.. *Nature*, 449:611-615 (2007).
- [19] Kurner, J., Frangakis, A. S., and Baumeister, W.. *Science*, 307:6-438 (2005).
- [20] Liu, J., Taylor, D. W., Krementsova, E. B., Trybus, K. M., and Taylor, K. A.. *Nature*, 442:208-211 (2006).
- [21] Midgley, P. A., Ward, E. P. W., Hungria, A. B., and Thomas, J. M.. *Chem. Soc. Rev.*, 36:1477-1494 (2007).
- [22] Kim, H. S., Hwang, S. O., Myung, Y. *et al.* *Nano Lett*, 8:551-557 (2008).
- [23] Kaneko, K., Inoke, K., Freitag, B., *et al.* *Nano Lett*, 7:421-425 (2007).
- [24] Arslan, I., Yates, T. J. V., Browning, N. D., and Midgley, P. A.. *Science*, 309:2195-2198 (2005).
- [25] Friedrich, H., Sietsma, J. R. A., de Jongh, P. E., *et al.* *J. Am. Chem. Soc.*, 129:10249-10254 (2007).
- [26] Friedrich, H., de Jongh, P. E., Verkleij, A. J., and de Jong, K. P.. *Chem. Rev.*, 109:1613-1629 (2009).
- [27] Ersen, O., Parmentier, J., Solovyov, L. A., *et al.* *J. Am. Chem. Soc.*, 130:16800-16806 (2008).
- [28] Midgley, P. A., and Weyland, M.. *Ultramicroscopy*, 96:413 (2003).
- [29] Midgley, P. A., Weyland, M., and Stegmann, H. Applications of Electron Tomography. In *Advanced Tomographic Methods in Materials Research and Engineering*; Banhard, J., Ed.; Oxford University Press Oxford, U.K., p 335 (2008)
- [30] Radon, J., Ber. Verh. K. Sachs. Ges. Wiss. Leipzig. *Math. Phys. Kl.*, 69:262 (1917).
- [31] Avinash, C., and Kak, Malcolm Slaney. *Principles of Computerized Tomographic Imaging*. IEEE Press, New York (1987)

- [32] Batenburg, K. J., and Sijbers, J. *Proc. IEEE Conf. Image Processing*, 4:133-136 (2007)
- [33] Kaczmarz, S.. *Bull. Acad. Pol. Sci. Lett. A*, 6-8A:355-357 (1937)
- [34] Tanabe, K.. *Numer. Math.*, 17:203-214 (1971)
- [35] Eggermont, P. P. B., Herman, G. T., and Lent, A.. *Linear Algebra and its Applications*, 40:37-67 (1981)
- [36] Censor, Y.. *Mathematical Programming*, 42:307-325 (1988)
- [37] Spontak, R. J., Williams, M. C. and Agard, D. A.. *Polymer*, 29:387-395 (1988)
- [38] Koster, A. J., Ziese, U., Verkleij, A. J., Janssen, A. H. and de Jong, K. P.. *J. Phys. Chem. B*, 104:9368-9370 (2000)
- [39] Yurtsever, A., Weyland, M., and Muller, D. A.. *Appl. Phys. Lett.*, 89, 151920 (2006)
- [40] Midgley, P. A. Weyland, M., Thomas, J. M. and Johnson, B. F. G.. *Chem. Commun.*, 10:907-908 (2001)
- [41] Hawkes, P. W.. *The Electron Microscope as a Structure Projector in Electron Tomography: Three-Dimensional Imaging with the Transmission Electron Microscope* (ed. Frank, J.), 17-39 Plenum (1992)
- [42] Egerton, R. F.. *Electron Energy-Loss Spectroscopy in the Electron Microscope*, 3rd edition. Springer Science+Business Media (2011)
- [43] Shindo, D., and Oikawa, T.. *Analytical Electron Microscopy for Materials Science*, Kyoritsu Shuppan Co. Ltd (1999)
- [44] Pennycook, S. J., Berger, S. D., and Culberston, R. J.. *J. Microscopy*, 144:229 (1986)
- [45] Muller, D. A.. *Nature Mater.*, 8:263-270 (2009)
- [46] Thomas, J. M., *et. al.* *Angew. Chem. Int. Ed.*, 43:6745-6747 (2004)
- [47] Thomas, P. J., and Midgley, P. A.. *Topics in Catalysis*, 21:109 (2002)
- [48] Ozkaya, D., Yuan, J., Brown, L. M., and Flewitt, P. E. J.. *J. Microsc.*, 180:300 (1995)

Chapter 2

Reconstruction and visualization of nanoparticle composites by transmission electron tomography

The first example of tomographic imaging in an electron microscope was reported four decades ago [1, 2]. However, the processing capacity of affordable computers has become sufficient for electron tomography (ET) much more recently. Consequently ET in (scanning) transmission electron microscopy ((S)TEM) has become a more viable option for investigating the morphologies, spatial distributions, and chemical compositions of nanostructures [1]. In order to obtain information needed to quantitatively characterize nanomaterials, such as particle volume filling fraction and the nanoparticle shape and size distribution, high quality 3D reconstructions without artifacts, or at least with well-understood artifacts, are required.

There are many difficulties associated with transmission electron tomography at the nanoscale. First, the individual objects (*i.e.*, nanocrystals) have sampling limitations, *e.g.*, due to instrument imaging resolution. For single-axis tilting from -90° to 90° , the resolution parallel to the tilt axis (assumed as x -axis) d_x , is equal to the original resolution of the projections in a perfect tilt series alignment. In general, the resolution in the other perpendicular directions, d_y and d_z , of a reconstructed tomogram is determined by the tilt range and increment, and can be approximated with the following formula:

$$d_y = d_z = \pi * \frac{D}{N} \quad (2.1)$$

where D is the diameter of the object and N is the number of tilt projections [4, 5].

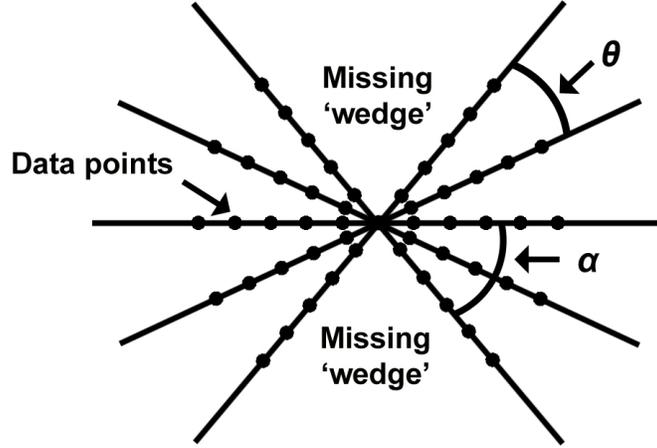


Figure 2.1: Illustration of the sampling of data in Fourier space. The relatively larger number of data points at low frequencies results in a blurred reconstruction. The angular increment between projections is θ and the maximum tilt angle α . Cited from Ref [6].

An additional factor that adversely affects the resolution in single-axis tilt tomography is the presence of a missing wedge due to the limited tilt range available in a TEM, as shown in Fig 2.1. This arises either due to blocking of the electron beam by the TEM holder, TEM grid or other parts of the sample, or because the projected sample thickness to be penetrated by the beam becomes too great at high tilt angles [7]. The missing wedge problem can lead to artifacts and anisotropy in the spatial resolution [8]. The result can be an image elongation or distortion parallel to the optic axis. The “elongation factor” e_{yz} is simply given by:

$$e_{yz} = \frac{d_z}{d_y} \quad (2.2)$$

The elongation factor is a function of the maximum tilt angle, α [8]:

$$e_{yz} = \sqrt{\frac{\alpha + \sin \alpha \cos \alpha}{\alpha - \sin \alpha \cos \alpha}} \quad (2.3)$$

In practice, the elongation factor e_{yz} is also affected by other factors such as the experimental recording conditions. For example, the e_{yz} in an ultra-thin film is considerably larger than the prediction by the theoretical equation due to the poor image quality especially at high tilt angles [9]. Further limitation on the reconstruction quality comes from nonlinear relationships between the transmission intensity and sample properties, such as local mass thickness [9, 10].

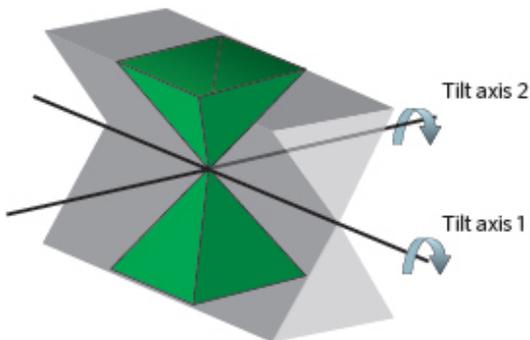


Figure 2.2: Illustration showing how a dual-axis tilt series collapses a missing wedge into a missing pyramid of information. Cited from Ref [6].

The quality of reconstruction results in transmission electron tomography has been steadily improving in recent years, for several reasons. First, new computational algorithms have been developed that can, for example, employ maximum entropy methods [11, 12] which seek the image with the least information content among those that could have given rise to the measured data, given what is known about the noise statistics. Second, the development of new TEM holders permit a full 90° tilt range reducing or eliminating the effect of the missing wedge [13]. Taking the advantage of focused ion beam sample preparation techniques, one can now fabricate needle-shaped samples that can take advantage of the full tilt range. Another way to obtain projections over as wide a tilt range as possible is to employ "dual-axis" recording geometry [14, 15, 16], as illustrated in Fig 2.2. More recently widespread imaging methods in electron microscopy, such as Z-contrast and energy filtered TEM (EFTEM) tomography minimize the effect of multiple electron scattering and

reduce nonlinearities between image intensity and projected thickness, as compared to more standard bright-field (BF) TEM tomography [8, 17].

In this chapter, different reconstruction and visualization methods in the transmission electron tomography of nanocomposite materials are evaluated and compared. Reconstruction and visualization using computer-generated phantom data consisting of random distributions of spherical nanoparticles are first performed. In this case, one can perform quantitative 3-D volume correlations in order to quantify the reconstruction and visualization artifacts, because the original object (the phantom) is perfectly known. Subsequently (in Chapter 4), a real sample consisting of a multilayer of gold nanocrystals embedded in an SiO₂ matrix is examined. The work aims to address the following questions:

- (1) How do the widely-used reconstruction methods FBP and SIRT compare in terms of the quality of the reconstruction of nanocomposite samples, when the missing wedge problem is absent?
- (2) How does one deal effectively with the absence of fiducial markers in nanoscale transmission electron tomography which are used for alignment as a marker by implanting a kind of nanoparticles (such as gold nanoparticles)?
- (3) What visualization method is suitable for quantitative interpretation in transmission electron tomography of nanocomposite samples?
- (4) Can the nanoparticle volume filling fraction, which can be important in many practical examples, be reliably ascertained by tomographic methods?
- (5) What are the common errors and artifacts, and what are the suitable metrics to reveal such artifacts?

2.1 Simulations and Experiment

For testing the tomographic reconstruction methods, a set of codes was written in the Matlab environment (much of the work in this chapter was published in reference [18]). These allow one to populate a desired volume with an arbitrary number of nanoparticles to produce a phantom data set. The code then creates simulated projection images using a selected range of tilt angles about a specified rotation axis. The conditions for domain population and image projection are that (i) particles are separate; (ii) the nanoparticles have no internal structure (they have uniform internal density); and iii) the transmitted

intensity observed in the projected image is linear with particle thickness as would correspond to mass thickness in an ideal experiment. The pixel intensity in the projection image "slice" was therefore simply the sum of the number of voxels inside the nanoparticles along the projection direction. Once a single image slice was produced, a rotation transform was applied along the axis traversing the center of the simulated geometry, essentially mimicking the rotation of a needle-like sample inside a TEM imaging system. Finally, the 600×600 pixel projection images were saved as a normalized 8-bit bitmaps for tomographic reconstruction. The 8-bit dynamic range provides 256 levels for the projected mass thickness ensuring sufficient projected image intensity sampling.

A set of simulated images was generated using the above technique to investigate the limitations of the reconstruction methods. The data set was based on six simulated phantom volumes having 10, 100, 200, 300, 400 and 500 spherical particles, respectively. The particles were 20 pixels in diameter and were randomly located within a 600-pixel-diameter sphere. The tilt range of projection of this data set was from -90° to $+90^\circ$ with a 2° increment. Reconstruction and analysis were performed using the TEMographyTM software package [19], utilizing either the FBP or SIRT reconstruction method. Twenty iterations were used in the SIRT reconstructions since no visible improvements were observed beyond this number. The reconstructed results were analyzed using Matlab software (Appendix A and B)

2.2 Results and Discussion

2.2.1 Visualization and rendering

The display of tomographic data requires the setting of an intensity threshold value that defines whether a voxel in the reconstructed volume is within or outside a nanoparticle. This selection is important, as it ultimately determines the apparent particle size and shape, and the volume fraction occupied by the particles (Fig 2.3). Determination of an appropriate threshold value is a non-trivial process, because there is no a-priori way to know the appropriate threshold value. Simply "estimating by eye the threshold is not always sufficiently well defined for quantitative work. Here, threshold values are estimated using the following steps:

First, a projection of the original simulated phantom dataset along a specific axis was obtained. Using the tomographic reconstruction, the projections of the reconstructed volume were taken along the same axis as the original "phantom" images, using several

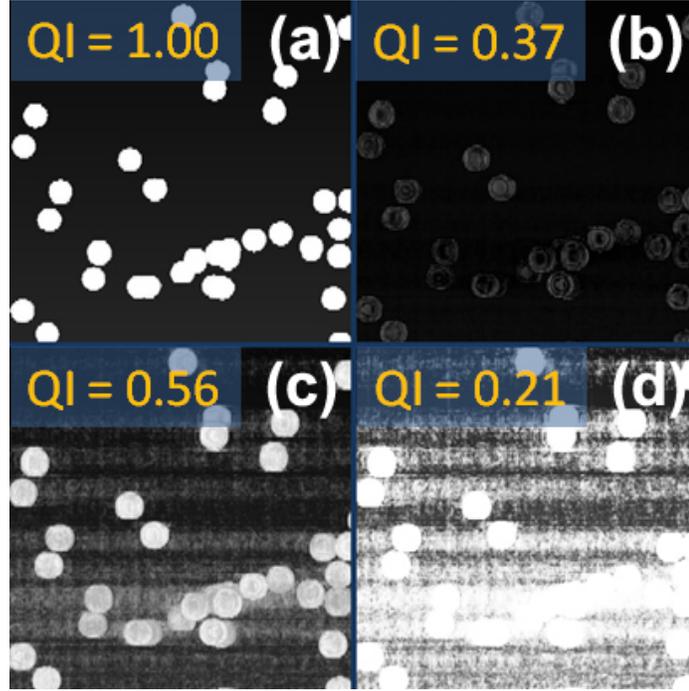


Figure 2.3: Projected images for 100-particle phantom data. (a) shows the original phantom data, while (b,c,d) show a FBP reconstruction using different thresholding for visualization. (a) has a quality index $QI = 1.00$, (b) has a quality index $QI = 0.37$, (c) has a quality index $QI = 0.56$ and d) has a quality index $QI = 0.21$. Figure (c) appears the most similar to the original phantom, as determined by QI.

different threshold values for a good “visual” similarity of the projected phantom and projected reconstruction images. Then the original projected phantom image was compared with the projected reconstructed images (with several threshold values) by calculating the 2D quality index for each one according to the procedure developed in Ref [20].

Before describing the next step, the 2D quality index (QI) is introduced. This is a mathematically universal objective image quality index. It is easy to calculate and applicable to various image processing applications. One considers a 2D image where $x = \{x_i, i = 1, 2, \dots, N\}$ and $y = \{y_i, i = 1, 2, \dots, N\}$ is the image signal or intensity respectively, in which N is the total number of pixels in the image. The definition of QI is:

$$QI = \frac{4\sigma_{xy}\bar{x} \cdot \bar{y}}{(\sigma_x^2 + \sigma_y^2)[(\bar{x})^2 + (\bar{y})^2]} \quad (2.4)$$

where

$$\begin{aligned}\bar{x} &= \frac{1}{N} \sum_{i=1}^N x_i, \bar{y} = \frac{1}{N} \sum_{i=1}^N y_i \\ \sigma_x^2 &= \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2, \sigma_y^2 = \frac{1}{N-1} \sum_{i=1}^N (y_i - \bar{y})^2 \\ \sigma_{xy} &= \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y}).\end{aligned}$$

In order to clarify the definition of the QI, we rewrite Equation 2.4 as a product of three components:

$$QI = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \cdot \frac{2\bar{x} \cdot \bar{y}}{(\bar{x})^2 + (\bar{y})^2} \cdot \frac{2\sigma_x \sigma_y}{(\sigma_x^2 + \sigma_y^2)} \quad (2.5)$$

In Equation 2.5, the first component is the correlation coefficient for image x and image y . Even if x and y have a linear relationship, relative distortions might still exist, which can be compensated by the second and third components of Equation 2.5. The second component evaluates how close the mean luminance/brightness is between image x and image y . The value of the second component is equal to 1 if and only if $\bar{x} = \bar{y}$. σ_x and σ_y can be viewed as the estimate of the contrast of image x and image y , so the third component measures how similar the contrasts of the images are. Therefore QI describes an image distortion by a combination of three factors: loss of correlation, luminance distortion, and contrast distortion. The range of the QI is from -1 to 1. A high QI value indicates close similarity between image x and image y .

In practice, the similarity of images is often related to the space variant, although usually a single overall quality value is used to quantify the similarity of images. Therefore the local QI can be calculated statistically and the average value of all local QI values can be obtained. A sliding window with $B \times B$ size moves B pixels horizontally and vertically through all the rows and columns of the image from the top-left corner of the image to the bottom-right corner. Given there are a total of S steps, the average QI can be rewritten as:

$$QI_{avg} = \frac{1}{S} \sum_{j=1}^S QI_j \quad (2.6)$$

where QI_j is the quality index at the j -th step.

The threshold value corresponding to the best QI was then chosen for the subsequent visualization of the tomographic results. Essentially, the quality index compares the pixel intensities in the projection of the phantom data with those in the projection of reconstructed data for various threshold values. While this method is somewhat arbitrary since it is based on an initial visual comparison of the original data with the reconstruction, it was combined with a quantifiable selection of the best threshold value.

Commonly, tomographic data from the TEM is visualized using either a volume rendering or an isosurface rendering method. In volume rendering, only those voxels with intensities higher than the threshold value (or within a range of lower and upper threshold values) will be considered to belong to nanoparticles within the visualization region of the reconstructed data. The “haze” visible in Fig. 2.3b is due to “reconstruction noise” *i.e.* pixels that are above the threshold but are not inside any defined particle. Since their greyscale value is above threshold they will incorrectly contribute to the volume filling fraction.

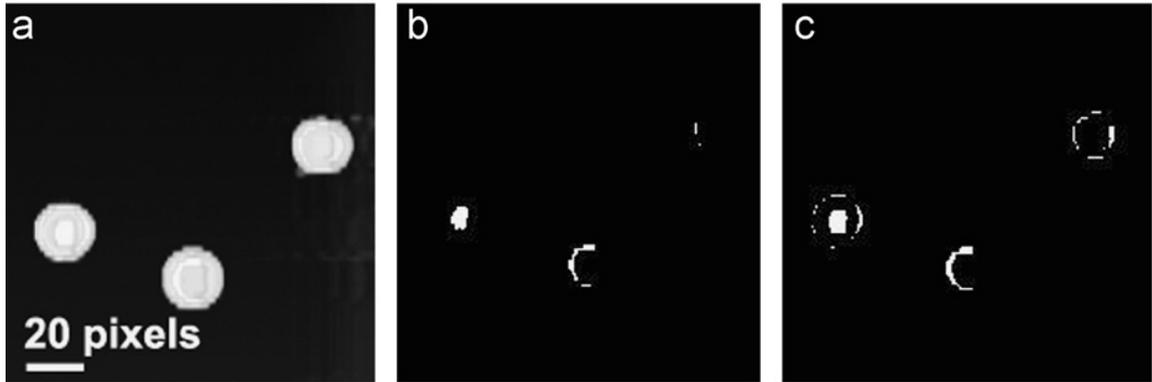


Figure 2.4: Visualization of the reconstructed results comparing: (a) volume rendering; (b) isosurface rendering with an threshold from 250 to 254; and (c) isosurface rendering with an threshold range from 245 to 254.

In visualization using isosurface rendering, a certain *range* of pixel intensities is chosen (as shown in the image Fig. 2.4). Any pixel within this range is then considered to be a part of a “surface” of a particle. Only those particles with closed surfaces will contribute to the volume filling fraction. Since some of the isosurfaces may not completely close (depending on the selected range of threshold values), the volume fraction obtained from isosurface reconstruction can be lower, sometimes considerably lower, than the true volume fraction in the original object. In Fig. 2.4b, one can see an image slice (*i.e.*, not a projection but a

slice through the 3D data cube) showing an example of one such surface in the tomographic image reconstruction. In this case, the pixel intensity range was selected from 250 to 254, and only those pixels are shown that indicate the particle surfaces. The same data is shown in Fig. 2.4c, except with an intensity range from 245 to 254, demonstrating that even with this small change the resulting structure looks quite different. Thus isosurface rendering with the surface determined by thresholding makes a reliable quantitative measure of volume filling fraction difficult to obtain. Therefore, volume rendering was employed to visualize the reconstruction results in the following work.

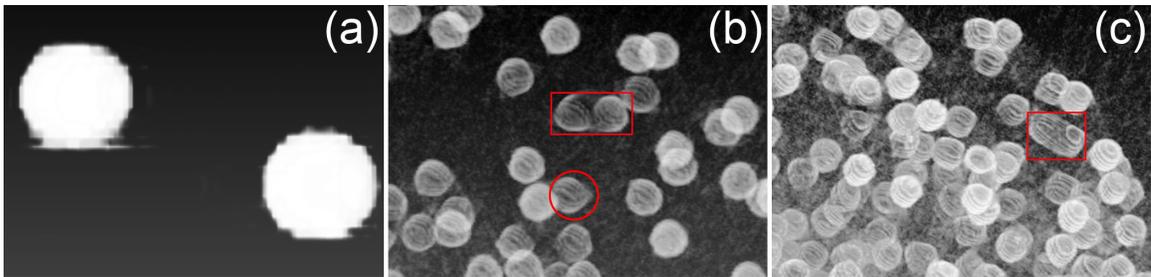


Figure 2.5: Zoom-in visualization of the FBP reconstruction for phantom data with different numbers of particles. (a) Reconstruction of 10 particles, the apparent density of the spheres is uniform in 10-particle case. (b) Reconstruction of 300 particle phantom data. The density of the spheres is not uniform. The two spheres marked by the red rectangle cannot be easily separated in the reconstruction although they can appear separated in some viewing angles. The red circle shows the distortion of one particle. (c) Visualization of the 500-particle reconstruction. The spheres in the red rectangle cannot be separated, from any viewing angle. “Fog” and streaking between spheres are obvious.

Next the quality of the tomographic reconstructions themselves is examined, using the volume rendering method with the image quality (QI)-based thresholding as described above. First, the reconstruction results for simulated data set are examined qualitatively, using the FBP reconstruction method (Fig. 2.5). The reconstructions by FBP appear visually satisfactory except that the image intensity at the particle boundaries tends to be higher than in the interior of the spheres. This gives many nanoparticles a shell-like appearance (especially apparent in Fig. 2.5c).

In order to clearly understand how this phenomenon appears, the effect of filtering on the reconstruction is examined. As introduced in the preceding chapter in the FBP method section, the summation of all projections of the point, after image filtering, gives the value of the point after reconstruction. The function of the filter is to avoid the artifacts (*i.e.*, the asterisk-like artifact) caused by the simple back-projection algorithm. As shown in

Eq.1.20, the transfer function of the filter is $H(\rho) = |\rho| = \mathbf{F}[h(x_r)]$. The problem is that the bandwidth of this filter is infinite, so that this ideal filter cannot be implemented in practice. Essentially $\int_{-\infty}^{\infty} |H(\rho)|^2 d\rho = \int_{-\infty}^{\infty} |\rho|^2 d\rho \rightarrow \infty$ does not satisfy the Paley-Wiener criterion [21] for a practical filter. In order to resolve this problem, the filter is usually selected based on the following two criteria: 1) the filter is applicable; 2) the filter can help improve the reconstruction accuracy. Thus, a filter can be obtained by applying a window function $W(\rho)$ which is used to confine the frequency range. The filter in frequency (Fourier) space can be expressed as:

$$H(\rho) = |\rho| \cdot W(\rho) \quad (2.7)$$

here $|\rho| < B$ is confined by the window function $W(\rho)$. Therefore, different filters can be obtained by selecting different window functions.

Assuming B is higher than the highest frequency component in each projection, then according to the Shannon sampling theorem [22] the projections can be sampled at intervals d of

$$d = \frac{1}{2B} \quad (2.8)$$

without inducing an additional error. Thus if d is sufficiently small, it is reasonable to ignore the high frequency contributions. The bandwidth is defined as $|\rho| < B = 1/(2d)$. Therefore, the projected data can be represented as:

$$P_\phi(nd), \quad n = -N/2, \dots, -1, 0, 1, \dots, N/2. \quad (2.9)$$

$P_\phi(nd)$ is set to zero for the extended points outside the original image. This process is named zero-padding; it can avoid the inter-period interference artifacts in the calculation of a convolution as shown in Ref. [23]. Then a fast Fourier transform (FFT) algorithm can be applied to calculate the Fourier transform $P(\rho, \phi)$ of a projection at angle ϕ :

$$P(\rho, \phi) = P_\phi(\rho) \approx P\left(n\frac{2B}{N}\right) = \frac{1}{2W} \sum_{n=-N/2}^{N/2} p\left(\frac{n}{2B}, \phi\right) e^{-j2\pi n(2B/N)} \quad (2.10)$$

As discussed above, the Fourier transform $P(\rho, \phi)$ is confined in frequency space by the window function. The filtered version of Eq 1.20 is:

$$\begin{aligned} g(x, \phi) &= \sum_{-B}^B P_{\phi}(\rho) |\rho| e^{j2\pi\rho x d\rho} \\ &\approx \frac{2B}{N} \sum_{n=-N/2}^{N/2} P_{\phi}\left(n\frac{2B}{N}\right) \left|n\frac{2B}{N}\right| W\left(n\frac{2B}{N}\right) e^{j2\pi n(2B/N)x} \end{aligned} \quad (2.11)$$

here N is sufficiently large and $W(n\frac{2B}{N})$ represents the window function used. One comment here is that the bandlimited $|\rho|$ is usually called the *transfer function* $H(\rho) = |\rho| \cdot W(\rho)$ of the system, which is written as:

$$H\left(n\frac{2B}{N}\right) = \left|n\frac{2B}{N}\right| \cdot W\left(n\frac{2B}{N}\right) \quad (2.12)$$

As derived in Eq 1.20, $g(x, \phi)$ can also be calculated by the convolution. The discrete expression of $g(x, \phi)$ sampled by the window function is:

$$g(x, \phi) = g_{\phi}\left(\frac{k}{2B}\right) \approx \frac{2B}{N} p_{\phi}\left(\frac{k}{2B}\right) * h\left(\frac{k}{2B}\right) \quad (2.13)$$

$$k = -N/2, \dots, -1, 0, 1, \dots, N/2$$

where $*$ denotes circular convolution and where $h(\frac{k}{2B})$ is the inverse discrete Fourier transform (DFT) of the discrete function $|n(2B/N)| H(2B/N)$. In addition, $h(k/2B)$ is usually named the *impulse response* of the filter in space.

In practice, the selection of the filter should be based on the structure of the object to be reconstructed. Here, two widely employed filters are introduced.

1 R-L (Ramachandran-Lakshminarayana) filter

(a) the transfer function of the system $H_{R-L}(\rho)$

$$H_{R-L}(\rho) = |\rho| W(\rho) = |\rho| \text{rect}\left(\frac{\rho}{2B}\right) \quad (2.14)$$

with

$$\text{rect}\left(\frac{\rho}{2B}\right) = \begin{cases} 1, & |\rho| < B \\ 0, & |\rho| \geq B \end{cases}$$

(b) the corresponding impulse response $h_{R-L}(x_r)$

There exists $h_{R-L}(x_r) = \mathbf{F}^{-1}(H_{R-L}(\rho))$, then:

$$\begin{aligned}
 h_{R-L} &= \int_{-B}^B |\rho| e^{j2\pi\rho x_r} d\rho \\
 &= 2B^2 \frac{\sin 2\pi x_r B}{2\pi x_r B} - B^2 \left(\frac{\sin \pi x_r B}{\pi x_r B} \right)^2 \\
 &= 2B^2 \text{sinc}(2x_r B) - B^2 \text{sinc}^2(x_r B)
 \end{aligned} \tag{2.15}$$

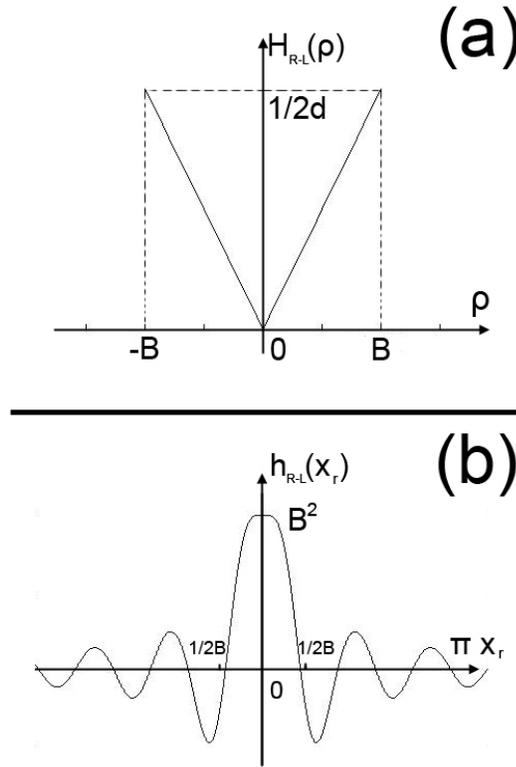


Figure 2.6: The profile of R-L filter in (a) frequency space and (b) real space. In the frequency space, the filter emphasizes the high frequency and de-emphasizes the low frequency. The filter acts as a high-pass filter to decrease the effect of low frequencies. The effect of the filter in real space is to behave like a window with the shape in (b) in the convolution in Eq. 1.20.

The illustration of $H_{R-L}(\rho)$ and $h_{R-L}(x_r)$ is shown in Fig. 2.6. Fig. 2.6a shows

that the R-L filter decreases the affect of low frequency signal on Eq. 1.20. The effect of the filter in real space is to behave like a window with the shape in Fig. 2.6b in the convolution in Eq. 1.20.

- (c) the discrete expression of $h_{R-L}(nd)$ Inserting the $x_r = nd$ into Eq. 2.15 to obtain the discrete expression of $h_{R-L}(nd)$, one obtains:

$$h_{R-L}(nd) = \begin{cases} \frac{1}{4d^2}, & n = 0 \\ 0, & n : \text{even} \\ -\frac{1}{n^2\pi^2d^2}, & n : \text{odd} \end{cases} \quad (2.16)$$

The illustration of discrete $h_{R-L}(nd)$ is shown in Fig. 2.7a. The continuous $\bar{h}_{R-L}(x_r)$ based on $h_{R-L}(x_r)$ is attained (in Fig 2.7b) by the linear interpolation. Eq. 2.14 was introduced by Bracewell [5]. Its discrete expression (Eq. 2.16) was proposed by Ramachandran and Lakshminarayanan [24]. Eq. 2.16 is the one of the most commonly used filters since it is simple and easy to implement. However the disadvantage of R-L filter is “ringing”. This is due to Gibbs phenomenon, wherein a sharp transition in the frequency domain is caused by the ideal rectangular filter windows.

2 S-L (Shepp-Logan) filter

In order to reduce the Gibbs phenomenon (oscillatory response), a suitable window function $W(\rho)$ is selected. For example, a *sinc* function is applied instead of the ideal rectangular window so that:

- (a) the transfer function of the system $H_{S-L}(\rho)$

$$H_{S-L}(\rho) = |\rho| \operatorname{sinc}\left(\frac{\rho}{2B}\right) \operatorname{rect}\left(\frac{\rho}{2B}\right) = \left| \frac{2B}{\pi} \sin \frac{\pi\rho}{2B} \right| \operatorname{rect}\left(\frac{\rho}{2B}\right) \quad (2.17)$$

as shown in Fig. 2.8a.

- (b) the corresponding impulse response $h_{S-L}(x_r)$

$$h_{S-L}(x_r) = \int_{-B}^B \left| \frac{2B}{\pi} \sin \frac{\pi\rho}{2B} \right| e^{2\pi\rho x_r} d\rho = \frac{1}{2} \left(\frac{4B}{\pi} \right)^2 \frac{1 - 4Bx_r \sin\left(\frac{\pi}{2 \times 4Bx_r}\right)}{1 - (4Bx_r)^2} \quad (2.18)$$

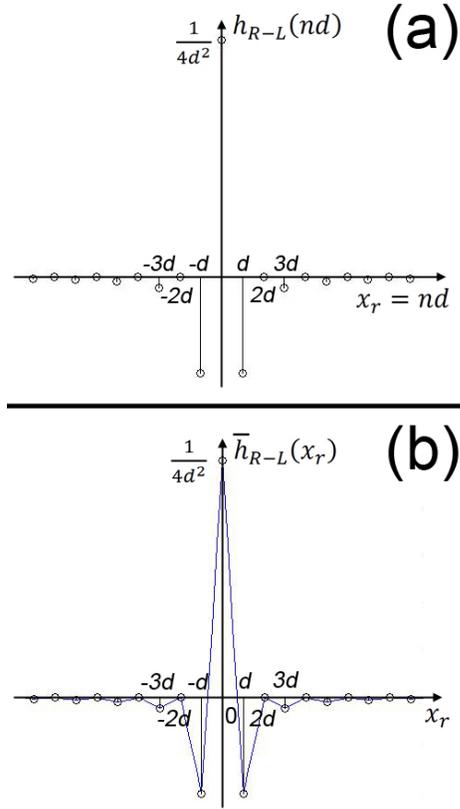


Figure 2.7: The profile of the R-L filter (a) discrete situation $h_{R-L}(nd)$ (dark circle) (b) using linear interpolation $\bar{h}_{R-L}(x_r)$ (blue line) and corresponding discrete value (dark-circle). The discrete form of the R-L filter is widely adopted in practical applications.

Let $y = 4Bx_r$, then:

$$h_{S_L}(y) = \frac{1}{2} \left(\frac{4B}{\pi} \right)^2 \frac{1 - y \sin \frac{\pi}{2} y}{1 - y^2} \quad (2.19)$$

the profile of $H_{S-L}(y)$ is illustrated in Fig 2.8b.

(c) the discrete expression of $h_{S-L}(nd)$

Similar to the R-L filter, the sampling interval still is represented as $d = 1/(2B)$. The uniform sampling is employed, which is $x_r = nd = n/(2B)$. Hence:

$$h_{S-L}(nd) = \frac{-2}{\pi^2 d^2 (4n^2 - 1)}, \quad n = 0, \pm 1, \pm 2, \dots \quad (2.20)$$

the profile of $h_{S_L}(nd)$ is shown in Fig 2.9. Compared to the R-L filter, the S-L filter change is more smooth at high frequency range. Correspondingly, in real

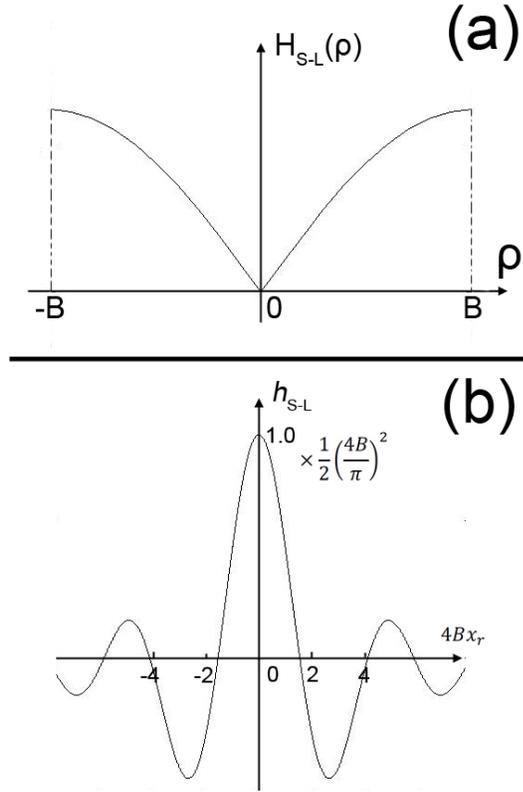


Figure 2.8: The profile of the S-L filter in (a) frequency space and (b) real space. In frequency space, the filter emphasizes the high frequency and de-emphasizes the low frequency. The filter acts as a high-pass filter to decrease the effect of low frequencies. Compared to the R-L filter, the S-L filter change is more smooth at high frequency range. Correspondingly, in real space, the S-L filter change is sharper around the zero point.

space, the S-L filter change is sharper around the zero point.

- (d) $\bar{h}_{S_L}(x_r)$ (the continuous $h_{S-L}(nd)$ by the linear interpolation) and its Fourier transform expression. By applying the linear interpolation method, $\bar{h}_{S_L}(x_r)$ can be obtained (Fig 2.9b). Its analytical expression is:

$$\bar{h}_{S_L}(x_r) = [h_{S-L}(x_r) \sum_{n=-\infty}^{\infty} \delta(x_r - nd)] * \Psi(x_r) \quad (2.21)$$

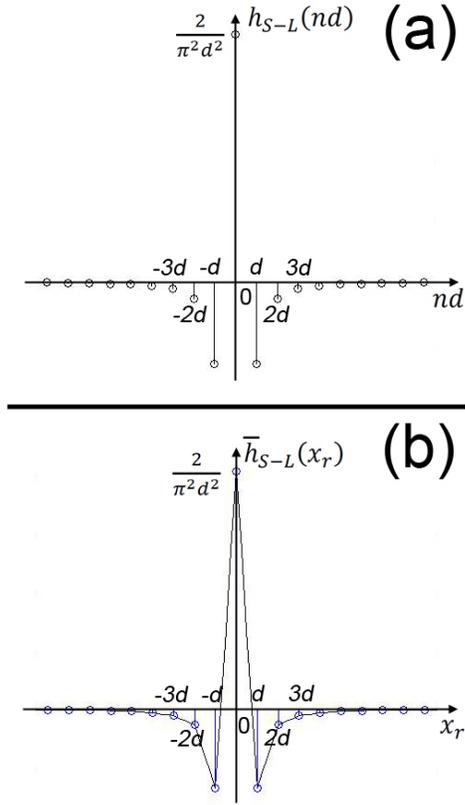


Figure 2.9: The profile of S-L filter and S-L filter with linear interpolation (a) discrete situation $h_{S-L}(nd)$ (b) using linear interpolation. $\bar{h}_{S-L}(x_r)$ (blue line) and corresponding discrete value (dark-circle). Compared to the discrete form of the R-L filter, the S-L filter becomes closer to x -axis when $x > d$. However, the R-L filter intersects with x -axis when the ratio of the interval over d is even.

here the linear interpolation function [26] is:

$$\Psi(x_r) = \begin{cases} 1 - |x_r|/d, & |x_r| < d \\ 0, & |x_r| \geq d \end{cases}$$

The linear interpolation can be rewritten as:

$$\Psi(x_r) = \frac{1}{d} \text{rect}\left(\frac{x_r}{d}\right) * \text{rect}\left(\frac{x_r}{d}\right) \quad (2.22)$$

Hence:

$$\bar{h}_{S-L}(x_r) = [h_{S-L}(x_r) \sum_{n=-\infty}^{\infty} \delta(x_r - nd)] * \frac{1}{d} \text{rect}\left(\frac{x_r}{d}\right) * \text{rect}\left(\frac{x_r}{d}\right) \quad (2.23)$$

If let $\omega = 2\pi\rho$, the Fourier transform of $\bar{h}_{S-L}(x_r)$ or the transfer function of the system is represented as:

$$\mathbf{F}[\bar{h}_{S-L}(x_r)] = \left| \frac{1}{\pi d} \sin \frac{\omega d}{2} \right| \left(\frac{\sin(\omega d/2)}{(\omega d/2)} \right)^2 \quad (2.24)$$

Eq. 2.24 is given by Shepp-Logan [27]. Usually all equations related to Eq. 2.24 are referred to as the S-L filter such as Eq. 2.17, Eq. 2.18, Eq. 2.20 and Eq. 2.23. In general, the most common S-L filter formula used in the FBP method is Eq. 2.20 since it is simple and easy to calculate.

One now turns to the problem of why the reconstructions by FBP appear tend to have a region of high image intensity at the particle boundaries. Fig 2.10b shows that the intensity (greyscale level) increases from the boundary towards the particle center in reconstructions using a simple back-projection without a filter. In Fig 2.10c the intensity of the sphere appears uniform in a reconstruction with the S-L filter, which weights the result of Fourier transformation of the intensity of the projected images using Eq. 2.24. In practice, a logarithm of the image intensity is sometimes taken, in order to reduce the effect of the exponential relation between image intensity and the local mass thickness in some TEM imaging modes [28]. Fig 2.10d shows the reconstruction result in which a logarithm was applied to the projected 1D images before reconstruction. This acts like a low-pass filter on the images. The reconstruction was then performed with S-L filter with the same parameters as in Fig 2.10c. The image in Fig 3d has a stronger intensity at the boundary than in the center. Thus, the increased intensity at the particle boundaries is related to the selected filter used in the back projection method. Similar results were found for the SIRT reconstructions: If a logarithm filter was employed on the images, unrealistically bright nanoparticle boundaries developed in the reconstruction.

Next a tomographic reconstruction of sample volumes with many nanoparticles will be examined. For a 10-particle object reconstructed using the FBP method, there is no noticeable particle elongation in any of the particles (Fig. 2.5a). This is in agreement with the standard elongation formula (2.3). However, reconstructions containing 100, 200, 300, 400 and 500 particles in the phantom volume do show particle shape irregularities (*e.g.*, the

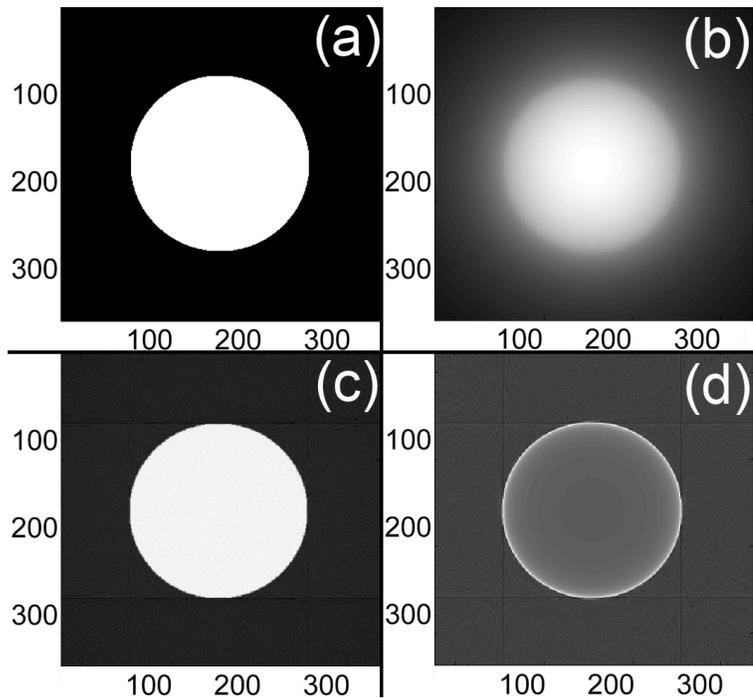


Figure 2.10: Effect of various back projection filters on reconstruction of a two-dimensional circle by FBP, from 1D projections. (a) computer-generated circle 100 pixels in diameter in a 360×360 pixel plane. (b) the reconstruction without a filter in the back projection. (c) the reconstruction with the S-L filter. (d) shows a reconstruction with same parameters as (c) but with a logarithm applied to the intensity of projected data. The reconstruction in (d) shows clear enhancement of boundaries. The projections were generated using 1° tilt increments and a projection tilt angle from -90° to 90° .

particle outlined with the red circle in Fig. 2.5b is elongated along the vertical axis). This distortion is not caused by the missing wedge since the simulated tilt range was a full $\pm 90^\circ$. The elongation direction is non-uniform throughout the reconstructed volume and appears to be more pronounced when there are several neighboring particles in close proximity.

At least two additional qualitative effects are visible in the FBP reconstructions, as illustrated in Fig. 2.5. First, there is an apparent “streaking” between adjacent particles and a “fog” around the individual particles, with both effects becoming stronger with an increasing particle number (*i.e.*, with increasing volume fraction occupied by particles) within the phantom volume. Particles that are closely-spaced were not always separated in the reconstruction. In numerous reconstructions by FBP, the center-to-center distance between two spheres needed to be larger than about 2.2 times the radius in order to be

imaged as clearly separate in the reconstructed volume. Often, particles closer together than about twice their radius become linked by “tendrils”, leading to highly non-spherical, elongated images of what were originally isolated, spherical objects. The practical impact of the above on morphology evaluation of an unknown material is obvious.

Next, the SIRT reconstruction results were examined and compared to the FBP images (Fig. 2.11). For 10 particles, the SIRT reconstruction appears visually satisfactory. Elongation was not observed for the 10-particle case. However, the shape distortion became a little worse with increasing particle number for the 100, 200, 300, 400 and 500 sphere cases. Once again, particles closer than about twice their radius became linked by filaments extending between adjacent particles.

In addition to these overall similarities between the two reconstruction methods, some clear differences exist as well. Most obviously, the streaking between particles, the surrounding “fog”, and particle overlap do not obviously appear in SIRT reconstructions compared to FBP reconstructions. The mottled, fog-like background in Fig. 2.5 is not present in Fig. 2.11. The streaking between particles, especially obvious in the two lower particles in Fig. 2.5b and 2.5c, is not present in Fig. 2.11b and 2.11c. Thus, a visual inspection suggests that the SIRT reconstruction is closer to the original object than for the FBP case.

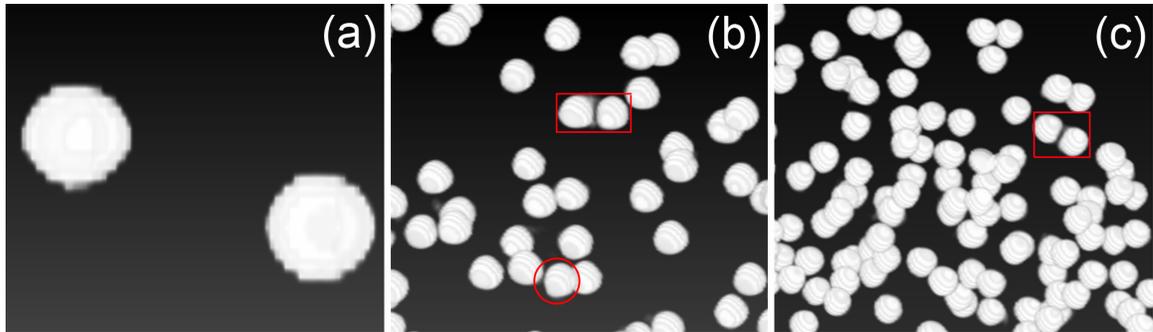


Figure 2.11: Zoom-in visualization of the reconstructed results by SIRT for various numbers of particles in the original phantom data. (a) Reconstruction of 10 particle phantom data. The apparent density (defined as pixel intensity) of spheres is uniform. (b) Reconstruction of 300 particle phantom data. Two spheres in the red rectangle can be separated more easily than that in FBP in Fig 2.5b). The particle marked by red circle again shows distortion from spherical shape. (c) Reconstruction of 500 particle volume. The two spheres marked by the red rectangle can be easily separated. “Fog” and streaking between spheres do not appear in 500 particle volume.

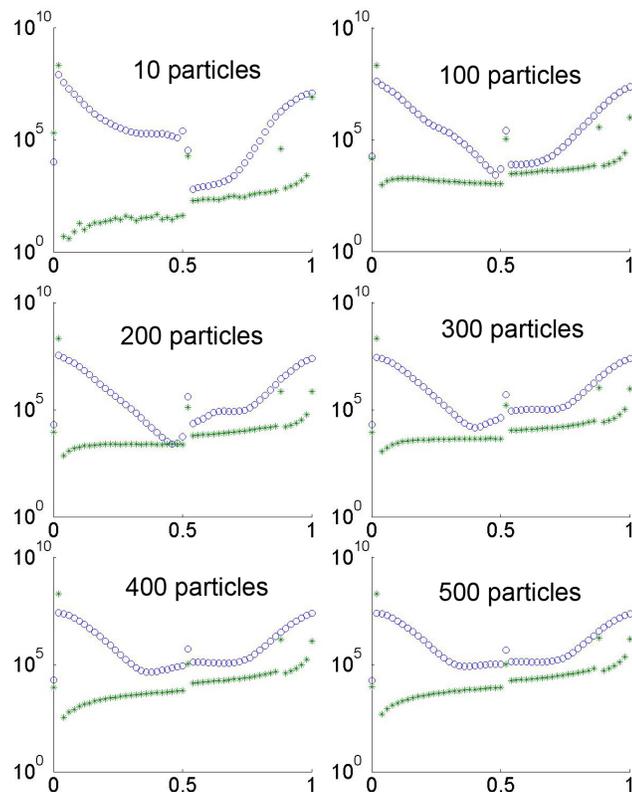


Figure 2.12: Comparison of the frequency distribution of the normalized RMS difference of the reconstructed results by FBP (blue circles) and SIRT (green crosses) respectively. The RMS frequency is higher for the FBP for almost all RMS differences. Both FBP and SIRT show significant number of pixels with high difference from the original phantom data. The increase in the RMS difference between the phantom data and the reconstructed volume is primarily due to increased frequency of pixels with small RMS difference.

Figs. 2.5 and 2.11 allow a visual comparison of the FBP and SIRT tomographic reconstruction; however, a more quantifiable way of determining the quality of the 3D reconstruction can be achieved by comparing the nanoparticle volume fraction in the reconstructed domain with the original known object volume fractions. In the case of simulated objects, this is easy enough to accomplish in real samples, as discussed further below, it requires independent knowledge of the volume filling fraction. The volume fractions could also in principle be used to set a threshold value (so that the object and image volume fractions match); however, we found that this did not produce high quality-index images, because of the tendency to count “noise pixels” above the threshold as part of the filling fraction.

A second method to compare reconstruction results is to calculate the voxel-by-voxel root-mean-square (RMS) difference between the reconstruction and the original phantom object.

In this work, a frequency analysis of the RMS differences between the reconstructions and the simulated objects was employed. The RMS value of each voxel was defined as $\sqrt{|I_0^2 - I_i^2|}$ where I_0 is the normalized intensity value in the object and I_i is the normalized intensity in the corresponding voxel of the reconstructed result. Given the large number of voxels in a $600 \times 600 \times 600$ pixel data set, it is impractical to examine the RMS value of individual voxels. Instead, we analyze the frequency of occurrence of a given RMS value. Thus, a high frequency for an RMS difference near zero indicates a close correlation between the ghost image and the reconstruction; whereas an RMS difference value of 1 indicates the maximum possible difference between the original and the reconstructed data (*i.e.*, zero intensity in the reconstruction where there is a maximum value in original phantom, or vice-versa). Finally, we also examine difference images in order to spatially locate the source of errors in the reconstruction. The RMS differences have nothing to do with the thresholding, since they are based on the evaluation of the reconstructed volume itself and not on the visualization method.

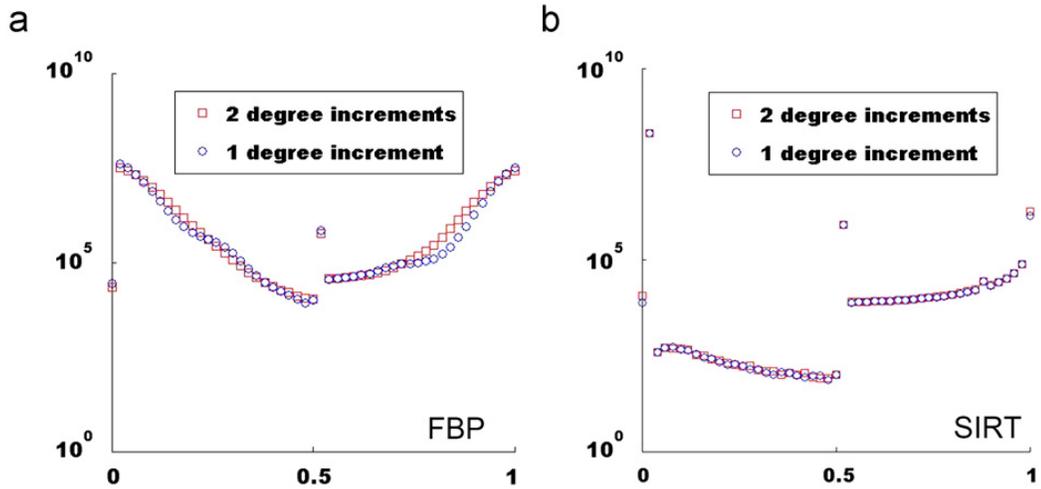


Figure 2.13: Comparison of the RMS differences for reconstructions by FBP (a) and SIRT (b) using either 1° or 2° tilt increments, for 200 particles in the object volume. The differences, especially for the SIRT method, appear minor.

The RMS differences were found to depend on the reconstruction method (FBP *vs.* SIRT) and on the particle density in the object volume (Fig. 2.12). However, perhaps somewhat surprisingly, the effect of the tilt increment (either 1° or 2° increments) was

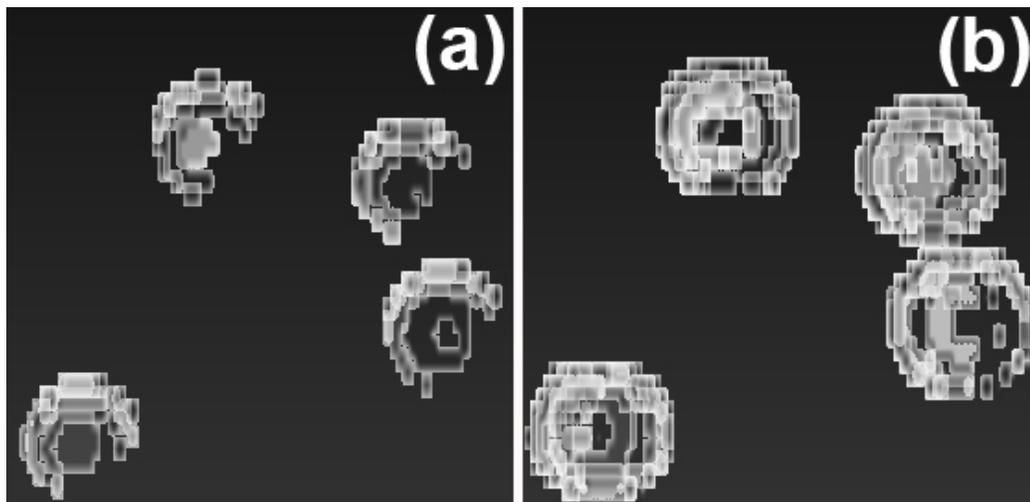


Figure 2.14: Zoom-in visualization showing pixels with an RMS difference equal to 0.5, from a) SIRT; and b) FBP. The object was a 300-particle phantom volume.

found to be relatively minor over this range (Fig. 2.13), suggesting that fairly little is to be gained on going from a tilt series of 180 to 360 images (in contrast, the RMS differences were noticeably worse for a 4° tilt increments in the case of SIRT reconstructions). Generally, the reconstruction results using the SIRT algorithm showed lower RMS differences for all particle numbers, except for RMS=0; in the latter case, the SIRT reconstruction results showed a spike, indicating a relatively large number of voxels for which the reconstruction closely or exactly matches the original data. Thus, a quantitatively better reconstruction was always obtained with SIRT than with FBP. The discontinuities at each end of the RMS difference spectrum for both reconstruction methods were found to be due to binning; a finer bin size resulted in noisier-looking curves but without discontinuities, except for the one at RMS=0.5.

For all reconstructions a pronounced discontinuity near an RMS difference of 0.5 was observed. In order to understand where, spatially, the source of this RMS difference “spike” originates, those voxels with an RMS difference equal to 0.5 are visualized in Fig. 2.14. Once again a boundary effect was observed: those voxels with an RMS difference equal to 0.5 almost always are located at the surface of the spheres. This leads one to suspect that the RMS=0.5 anomaly originates from problems associated with the sampling mechanism in the computer-generated data (this is separate from the bright boundary effect which did not occur in SIRT and was minimized in the FBP by avoiding the use of a log filter). Instead, the grid boxes shown in Fig. 2.15 describe the pixelation in the projected image. If the

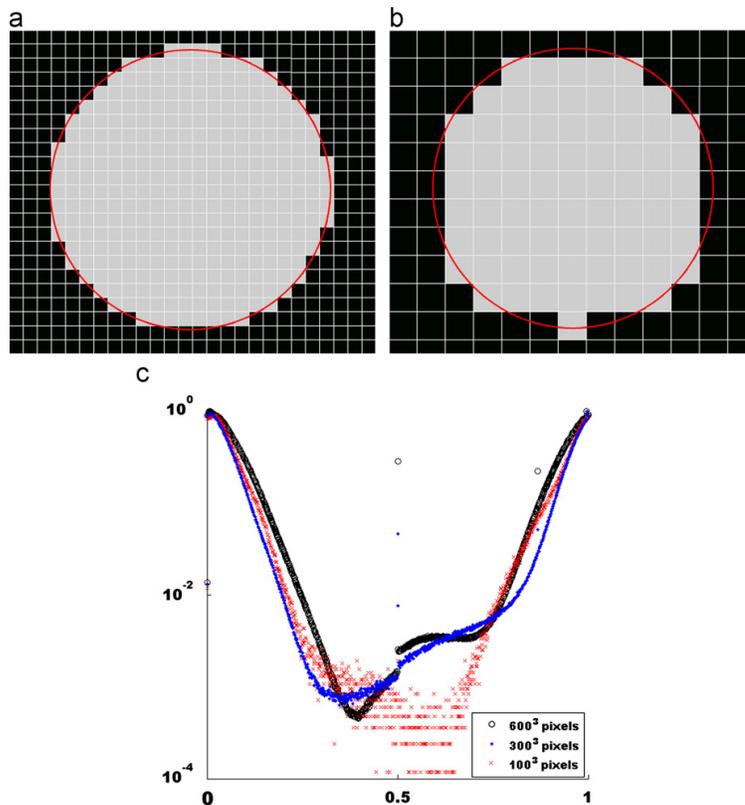


Figure 2.15: An illustration of sampling mechanism and its influence on the frequency of the RMS=0.5 voxels. (a) a circular particle with fine sampling; (b) the grid is coarsened such that each pixel has double the edge length; (c) the RMS distribution of a real 3D phantom with three different pixel sizes, corresponding to image of 600^3 , 300^3 , and 100^3 pixel size. The RMS=0.5 spike becomes less noticeable as the grid is coarsened. The maximum RMS value is normalized for each case, for better visual comparison of the curve shapes.

pixel center distance falls inside the radius of the circle, the value of the grid box is equal to 1; otherwise, it is zero. Therefore, considering many projections at different tilt angles, sometimes a given boundary pixel will be assigned a value of 1, and at other times a value of 0. The average value of that pixel from all projections will be close to or equal to 0.5; whereas in the original object it was *either* zero or one, and the RMS difference between 0.5 and either 0 or 1 is, obviously, equal to 0.5. One also sees in Fig. 2.15 that there will be proportionally fewer of these ambiguous surface pixels (voxels in 3D) when the gridding is coarser. By reconstructing the original data with three different voxel sizes (Fig. 2.15), one does indeed observe a decrease in the value of the RMS=0.5 spike, consistent with this

spatial quantization hypothesis. In an experiment, the object is also sampled by a finite number of pixels, for example pixels of a CCD camera. However, in an experiment, the object is often heavily oversampled (many pixels in projected images correspond to single pixel in reconstruction) thus reducing the effect of spatial quantization.

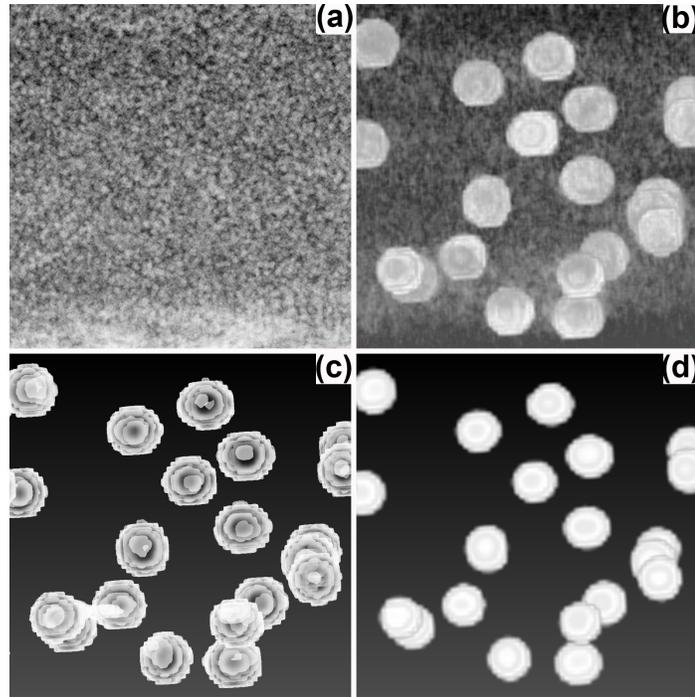


Figure 2.16: Zoom-in visualization of the 300-particle reconstruction and imaging results for: (a) FBP, volume fraction set to known fraction in the phantom (4.7%); (b) FBP for the best QI, giving a volume fraction of 1.5%; (c) SIRT, volume fraction set to known fraction in the phantom; (d) SIRT for the best QI, giving a volume fraction of 0.8%.

A suitable method to set threshold values for data visualization might be to match the particle volume fractions in the reconstruction with the known values in the simulated object (or the approximately-known values in the case of a real sample as discussed below). However, as shown in Fig 2.16, volume fraction matching is not a reliable way to set the threshold. The results showed that the best QI will always have a lower particle volume filling fraction than that in the original object. This is mainly account of the tendency to otherwise count above-threshold noise (*i.e.*, streaks and fog) as contributing to the calculated volume fraction. This is especially apparent in Fig. 2.16a, where the FBP reconstruction with a threshold set to match the volume fraction produced an especially

poor image. For the SIRT reconstruction, the reconstruction most closely matching the volume fractions also produced a much lower-than-optimum quality index (see Figs. 2.16c and 2.16d). Therefore the volume fraction is not a reliable parameter to determine the visualization threshold.

2.3 Summary

Although transmission electron tomography is becoming an increasingly important and more widely used tool for characterizing the shape and spatial distribution of nanocomposite materials, various artifacts need to be considered when attempting quantitative measurements. In this chapter, some of the important considerations for generating and visualizing tomographic data from the TEM have been highlighted. First, the thresholding is a key aspect that can have a profound effect on how the reconstruction looks. Changing the thresholding can change the size, shape, and apparent connectivity of the particles. One cannot set a threshold to match a known volume fraction; this produces poor results due primarily to the unavoidable counting of high-intensity-value “noise pixels” as belonging to the nanoparticle volume. Both FBP and SIRT reconstruction methods were tested for simulated data, and in general, superior results were obtained using SIRT. Several additional issues must be considered in the visualization: first, certain common types of filtering and image processing before reconstruction can lead to artifacts such as a high intensity at the interfaces. This can give the nanoparticles a shell-like appearance for both FBP and SIRT reconstruction methods. Turning these filters “on” or “off” is necessary in the imaging software. Finally, for a tilt interval of 2° and a full $\pm 90^\circ$ tilt range, particles spaced closer than about 2.2 times their radius will erroneously appear to be a single, elongated particle even under optimal reconstruction and visualization methods for both FBP and SIRT methods.

Bibliography

- [1] De Rosier, D. J., and Klug, A.. *Nature*, 217:130-134 (1968)
- [2] Hart, R. G. *Science*, 159:1464-1467 (1968)
- [3] Frank, J.. *Three-Dimensional Electron Microscopy of Macromolecular Assemblies*, Academic Press, San Diego (1996).
- [4] Crowther, R. A., de Rosier, D. J., and Klug, A.. *Proc. R. Soc. London, Ser. A*, 317:319 (1970)
- [5] Bracewell, R. N. and Riddle, A. C.. *Astrophys. J.*, 150:427-437 (1967)
- [6] Midgley, P. A., and Dunin-Borkowski, R. E.. *Nature Mat.*, 8:271-280 (2009)
- [7] Kawase, N., Kato, M., Nishioka, H., Jinnai, H... *Ultramicroscopy*, 10:8-15 (2007)
- [8] Midgley, P. A., and Weyland, M.. *Ultramicroscopy*, 96:413-431 (2003)
- [9] Kaneko, T., Nishioka, H., Nishi, T., and Jinnai, H.. *J. Electron Microsc.*, 54:437-444 (2005)
- [10] Cao, M., Zhang, H. B., Li, C., and Nishi, R.. *Rev. Sci. Instrum.*, 80:026104 (2009)
- [11] Barth, M., Bryan, R. K. *et al.* *Scanning Microsc.*, Suppl. 2:277-284 (1988).
- [12] Skoglund, U., Ofverstedt, L. G. *et al.* *J. Struct. Biol.*, 117:173-188 (1996).
- [13] Jinnai, H., and Spontak, R. J.. *Polymer*, 50:1067-1087 (2009)
- [14] Penczek, P., Marko, M., Buttle, K., and Frank, J.. *Ultramicroscopy*, 60:393-410 (1995)
- [15] Mastronarde, D. N.. *J. Struct. Biol.*, 120:343-352 (1997)
- [16] Sugimori, H., Nishi, T., and Jinnai, H.. *Macromolecules*, 38:10226-10233 (2005)

- [17] Li, P., Wang, X., Malac, M., *et al.* *Microscopy and Microanalysis*, Vol 15, Suppl. 2:1256 (2009)
- [18] Wang, X. Y., Lockwood, R., Malac, M., Furukawa, H., Li, P., and Meldrum, A.. *Ultramicroscopy*, 113:96-105 (2012)
- [19] Tomography.com
- [20] Wang, Z., and Bovik, A. C.. *IEEE Signal Processing Letters*, Vol 9, No.3 81 (2002)
- [21] Oppenheim, A. V., Willsky, A. S. and Hamid, S.. *Signals and Systems* (2nd Edition), Prentice Hall (1996)
- [22] Shannon, C. E.. *Proc. Institute of Radio Engineers*, vol.37, no.1, p10-21 (1949)
- [23] Jakowatz, Jr. C. V., and Kak, A. C.. "Computerized tomography using *x*-rays and ultrasound," Research Rep. TR-EE 76-26, School of Electrical Engineering, Purdue Univ., Lafayette, IN (1976)
- [24] Ramachandran, G. N., and Lakshminarayanan, A. V.. *Proc. Nat. Acad. Sci.*, 68:2236-2240 (1971)
- [25] Shepp, L. A., and Logan, B. F.. *IEEE Trans. Nucl. Sci. NS*, 21:237-245 (1981)
- [26] Parker, J. A, Kenyon, R. V., and Troxel D. E. *IEEE Transaction on Medical Imaging*, 2(1):31-39 (1983)
- [27] Shepp, L. A. and Logan, B. F. *IEEE Trans. Nuci. Sci.*, NS-21:21-43 (1974)
- [28] Reimer, L.. *Transmission Electron Microscopy: Physics of Image Formation and Microanalysis*. Second Edition. Springer-Verlag Berlin Heidelberg New York (1989)

Chapter 3

Experimental Considerations of Electron Tomography

The sample preparation methods and image recording processes are crucial aspects of electron tomography. Therefore the methods for sample preparation, including a newly developed method, will be discussed in this chapter. All the samples used in this thesis will also be discussed here. The results obtained will be discussed in the next chapter. Finally, the common image recording modes will briefly be discussed.

3.1 Sample Preparation Methods for Electron Tomography

As discussed in Chapter 2, reconstruction methods and algorithms used in electron tomography affect the quality of the reconstruction. Sample preparation methods also play an important role in the quality of the final reconstruction. In conventional transmission electron microscopy (TEM) observation, the image quality of the sample is affected by the preparation methods. Important factors including the thickness of the sample and contamination can affect the results. Furthermore, sample preparation methods for ET applications should guarantee the quality of all images in the whole tilt series, although obtaining high quality images is especially difficult at high tilt angles.

Sample preparation methods have been developed in two main fields. In biological research, many methods, such as negative staining, glucose embedding, use of tannic acid, ice-embedded specimens, hybrid techniques (cryo-negative staining) and labeling with gold clusters to enhance contrast of biological samples, were developed. An example of a detailed discussion on biological sample 3D ET preparation in bio-TEM appears in Ref [1]. In

materials science, sample preparation methods can be classified into three categories: grid support methods, physical thinning methods and focused ion beam (FIB) micro-sampling techniques. All three methods have been used in this work.

3.1.1 Grid Support Method

The support method makes use of a suitable amorphous thin film, often amorphous carbon, to support the sample. In general, carbon films with holes $1\sim 2\ \mu\text{m}$ in diameter are prepared on 200 mesh copper or molybdenum grids to provide support and stability to the sample. The sample can be directly deposited onto the holey carbon film for the experiment.

The advantage of the grid support method is its simplicity and relative experimental ease. This preparation method causes little physical damage or contamination on the sample. Therefore, this method is widely applied in the preparation of 3D tomography samples, especially for nanoparticles suspended in a liquid buffer.

The disadvantage is that the grid support method does not allow tilting up to $\pm 90^\circ$. This limitation arises from the fact that the projected sample image is blocked by the grid bars of the TEM grid at high tilts. The resulting missing wedge can lead to elongation artifacts in the reconstructed objects [1].

3.1.2 Physical thinning Methods

Plan-view and cross-sectional view are two basic approaches in TEM sample preparation methods. Both methods typically utilize cutting, gluing, polishing and ion milling processes (Fig 3.1). In both cases, the most important problem is the fact that the projected thickness increases with increasing tilt angle. This phenomenon is usually referred to as the thickness effect (Fig 3.2). The projected thickness increases with increasing the tilt angles according to $d_2 = d_1/|\cos\theta|$, where θ is the tilt angle, d_1 is the projected thickness at a zero degrees of tilt and d_2 is the projected thickness at tilt angle θ . The projected thickness d_2 of the sample along the beam direction affects the quality of the recorded image in TEM. The range of tilt angles can be further limited by blocking of the beam by the grid bar.

A simple way to decrease the thickness effect is to prepare a very thin sample. However, this is not easy to implement by physical thinning because of the limitation of the physical thinning methods, such as the polishing process and the ion milling process. Therefore, the tilt for cross section samples ranges usually from -60° to $+60^\circ$ in most cases.

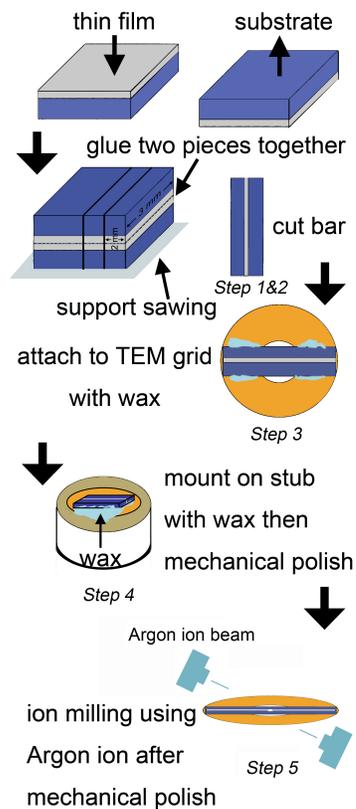


Figure 3.1: The illustration of a physical thinning preparation method. 1. Two pieces of the sample are adhered together using glue. 2. A 2mm×3mm cut bar is obtained by cutting the glued sample. 3. The cut bar is attached to a Cu grid using wax. 4. The Cu grid with the cut bar is then attached to a stub to be mechanically polished. 5. After mechanically polishing, the sample is further thinned using an Argon ion beam. Revised from [2]

3.1.3 The FIB Micro-Sampling Technique

FIB is a technique applied in the semiconductor industry, materials science, and increasingly in biology for site-specific analysis, deposition, and ablation of materials [3]. In a FIB, a beam of Ga ions is used to sputter materials away. The most basic form of a FIB microscope consists of a liquid metal ion source producing a Ga^+ ion beam with a range of energies typically from 2 to 50 keV. The Ga^+ ion beam, which is converged to a finely focused probe using a method similar to that of the scanning electron microscope, prepares the sample by localized milling. The current in the Ga^+ ion beam can be controlled from approximately 1 pA to 30 nA or more by varying (electrostatic) lenses and selecting various beam limiting apertures. The beam diameter can range from about 5 nm for fine milling

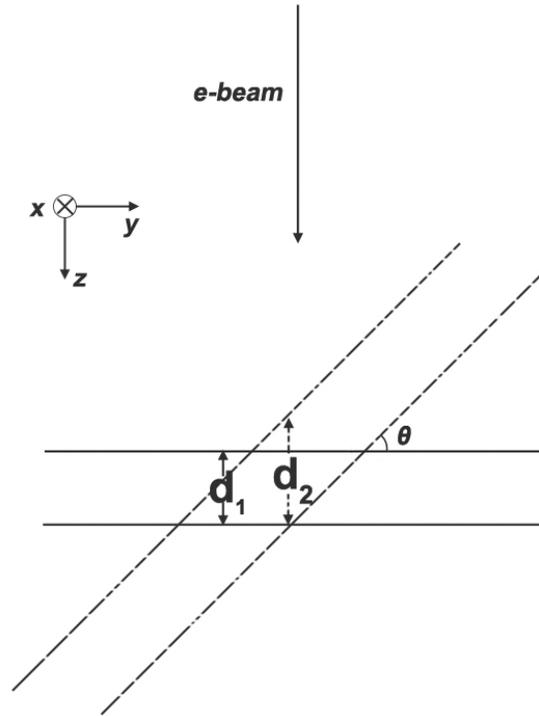


Figure 3.2: An illustration of thickness effect caused by the rotation and projection. The sample is viewed along the x-axis. The parallel dashed lines represents the sample after rotating by θ degrees. The projected thickness (d_2) along z-axis is equal to d_1 at zero degrees. d_2 changes with the tilt angle θ .

to several micrometers for rough sample modifications. With modern instruments capable of delivering 30 nA in sub-micrometer space, current densities in excess of 10 A/cm² can be achieved. Thus FIB is a “nano-scale milling machine”.

A FIB micro-sampling technique has been developed for electron tomography sample preparation. The method includes the following steps: milling or cutting the sample, lifting out the micro-sample, mounting the micro-sample on a special carrier, and thinning the micro-sample by FIB milling. All operation steps are performed in the FIB system. The procedure for the FIB micro-sampling is schematically illustrated in Figure 3.3.

The region of interest is first covered with a protective layer such as W, Pt or carbon. Next, the sample is deep trenched by FIB milling to cut out a portion of the sample (referred to as the micro-sample) from the region of interest (Fig 3.3a). Then, the micro-sample is separated from the bulk sample and lifted out with the assistance of a micromanipulator placed inside the FIB system (Fig 3.3b and 3.3c). The micro-sample is attached to an edge

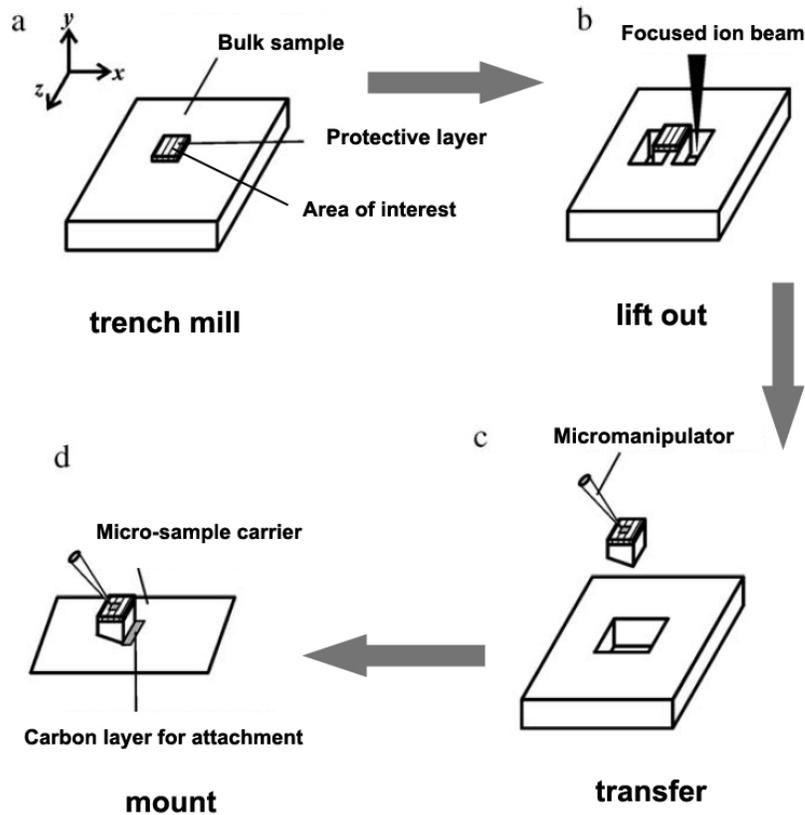


Figure 3.3: Schematic illustration of the FIB micro-sampling technique. Revised from [4].

of a micro-sample carrier (or a TEM grid designed for electron tomography) usually by carbon deposition (Fig 3.3d) in the FIB and then the micro-sample is cut off the micromanipulator. Finally, the micro-sample carrier is mounted onto an FIB/(S)TEM compatible sample holder.

The FIB micro-sampling technique has several advantages. The most important is that it eliminates the missing wedge problem if a special TEM holder with $\pm 90^\circ$ tilt range is employed in the electron tomography experiment [5]. Secondly, the FIB micro-sampling technique can investigate the area of interest of the sample in structurally complicated materials.

There are two main disadvantages of the FIB micro-sample technique. One is that it is time consuming. The other one is that it may cause Ga ion radiation damage. This makes the method difficult for fabricating samples used for high resolution tomographic

reconstruction, which is sensitive to Ga irradiation.

3.1.4 A New Method of 3D Electron Tomography Sample Fabrication by Combining FIB Technique and Thin film deposition methods

If the sample is prepared with a thickness limitation along one dimension such as thin film samples, the micro-sample preparation time will be greatly minimized. However, decreasing the thickness of the sample is not always simple [6, 7, 8, 9]. As discussed in Section 3.1.1, one of advantages of the grid support technique is that the sample is naturally thin along one dimension. This permits a bridge-like sample preparation method which is described below. Some of this work was published in [10].

Here a new method is reported for sample preparation suitable for high-tilt-angle ET on materials that can either be prepared in thin film form or deposited onto a thin film substrate either from liquid or using thin film deposition processes. One experimentally demonstrates the suitability of this method by applying it to an Er-doped $\text{SiO}_{1.5}$ thin film containing silicon nanocrystals. The method reduces the problems associated with increasing projected sample thickness and allows one to obtain data over sample tilt range up $\pm 75^\circ$, while eliminating the delicate “pluck-out” that is needed for rod-shaped sample preparation [9, 11]. The effect of missing wedge is very small in data collected over a tilt range greater than $\pm 75^\circ$ [9, 11]. Additionally, using the method reported here, only a small part of the sample is consumed in order to accomplish the electron tomography. Other TEM characterization methods can be used to investigate adjacent areas of the same sample during the same microscope session.

The method involves the deposition of the sample material onto a thin film substrate, such as an amorphous carbon or SiN membrane, followed by a simple FIB fabrication step. Figure 3.4 shows the steps involved in the sample preparation. First, the sample is prepared by suitable deposition methods such that the thin film can be “floated” from the substrate and “picked up” onto a TEM grid. FIB milling is then used to cut out two large rectangles, leaving a thin bar or “bridge” of the original film. It is preferable to align the bridge parallel to the grid bars for easy alignment with respect to the tilt axis when the sample is mounted in the TEM holder. The details of the fabrication procedure are listed as follows:

- (1) The sample is deposited as a thin film on mica or rock salt. The film is then floated on water and picked up onto a standard copper grid with 75 mesh or less. Alternatively, for a sample which is prepared in a solution or as a suspension in a liquid, the sample can be dropped onto an existing thin film support, such as sub-10-nm-thick amorphous carbon on a copper support grid and then dried in air;

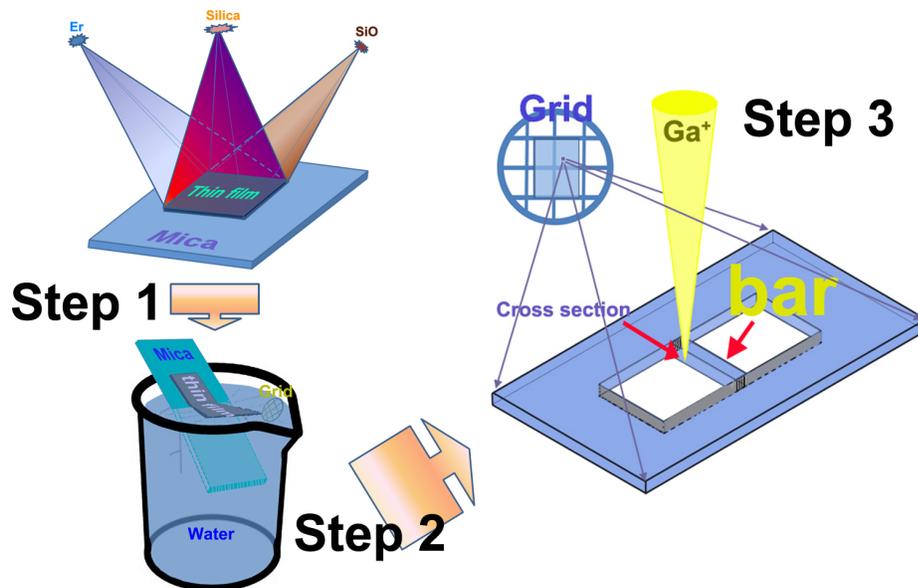


Figure 3.4: Steps involved in the bridge-type sample preparation. Step 1 is the deposition of a thin film that is either the sample itself or a support film used for a suspended sample. Step 2 shows the collection of the thin film prepared in Step 1 onto a TEM grid. Step 3 is the fabrication of a bridge-shaped sample using a focused ion beam instrument. L is the length of the side of the rectangular hole perpendicular to the bridge, and h is the thickness of the thin film.

- (2) A carbon layer with a thickness of approximately 5 nm can be deposited on the surface of the film in order to reduce sample charging during FIB milling and TEM analysis;
- (3) Two rectangles are cut out by the FIB, leaving a narrow bridge with the desired width. The bridge is preferably aligned parallel to the grid bars. When the length of the side of the rectangle perpendicular to the bridge is more than 3.73 times the thickness of the thin film, a $\pm 75^\circ$ tilt range can be achieved as shown in Figure 3.5 (this can be more than six times, in practice, for easier alignment and to account for the possibility that the bridge axis is not perfectly aligned along the tilt axis).

It is desirable to limit the maximum projected thickness below 1 inelastic mean free path for the incident electron energy in the TEM. This translates to about 100 nm for a microscope operated at a few hundred kV mainly composed of light-element samples [13]. Keeping the projected thickness below about 100 nm (inset in Figure 3.5) not only allows good image contrast, but also reduces overlap of individual objects (such as nanocrystals)

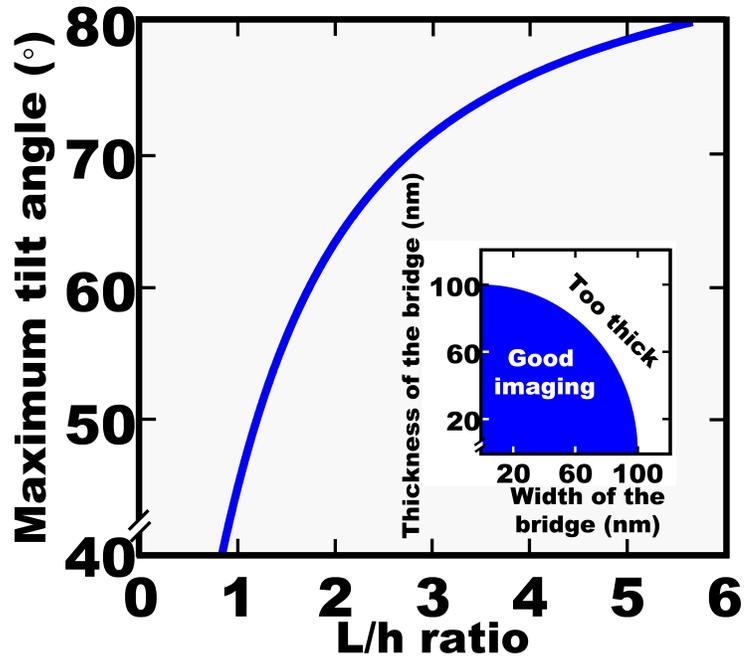


Figure 3.5: The maximum achievable tilt angle as a function of the ratio L/h . For a given value of L/h , the maximum tilt corresponds to the greatest angle that would not lead to a projected overlap of the bridge and the film beyond the cut out rectangle. If L/h is larger than 3.73, a $\pm 75^\circ$ tilt range can be achieved. Ultimately the bars of the supporting Cu grid will limit the tilt angle to about $\pm 80^\circ$. The inset shows the values of the bridge width and film thickness for which the maximum projected sample thickness does not exceed 100 nm from any viewing direction (blue region in the inset).

in the projected images.

Here an example of the application of the above method is given. An Erbium-doped $\text{SiO}_{1.5}$ thin film was used to demonstrate the bridge sample suitability for electron tomography. The thin film sample was deposited onto a mica substrate by simultaneous thermal evaporation of silicon monoxide (SiO), silica (SiO_2), and metallic erbium (Er). The film composition was targeted to be $\text{SiO}_{1.5}$ (10 at.% excess Si) with an erbium doping level of about $6 \times 10^{20} \text{ cm}^{-3}$ (1 at.%). The film thickness was 20 nm. The thin film was floated off the mica substrate in distilled water and collected onto a molybdenum TEM grid. Subsequent annealing in forming gas (5% H_2 + 95% N_2) at 1000°C resulted in phase segregation that resulted in the growth of silicon nanocrystals in an oxide matrix [12].

The next step was to prepare the bar-shaped sample using Ga ion milling in a Hitachi

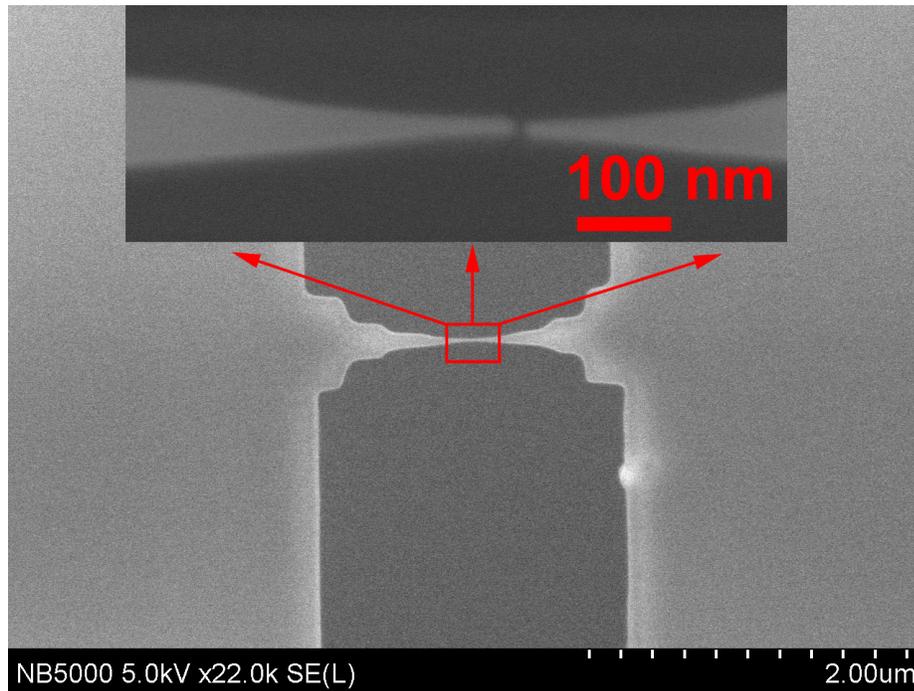


Figure 3.6: An SEM image of the finished bar-shaped sample for TEM tomography. Area 1 is the original film, while area 2 is vacuum after the film was cut out. The inset shows a higher magnification of the bridge region of the sample. In this case, the bar slightly detached in the center, but this does not cause a problem for collecting a tilt series in the TEM.

NB 5000 dual beam instrument (FIB/SEM). As shown in the SEM image (Figure 3.6), two rectangular slabs are removed from the film to produce a bar-shaped sample. Although in this particular case the bar cracked in the center (as observable in the high-magnification inset in Figure 3.6), it did not affect the observation of the sample in TEM since the width of the bridge is sufficiently small for good imaging at high tilt angles. The FIB step is fast; it takes about one hour in FIB compared to three hours or more for rod-shaped sample preparation. Furthermore it avoids the difficult lift-out step in the FIB. As mentioned at the beginning of this chapter, the experimental results will be discussed in the next chapter.

The sample preparation is not particularly difficult. Although the method cannot cover the full $\pm 90^\circ$ tilt range, it can easily reach $\pm 75^\circ$ tilt range which is sufficient in many practical applications [9]. One requirement is that the side of the cut-out rectangle perpendicular to the bar is six times the width of the bar, in order to obtain sufficient tilt range. The process also reduces Ga^+ contamination compared to needle shape sample

method, since only two sides of a thin film are FIB milled as opposed to extensive milling of entire needle. Conceivably, the method can be also used for samples deposited onto thin substrate from a solution. Furthermore, the substrate can be milled before the sample is deposited onto it. The fact that there is large area in addition the narrow bridge for tomography provides ample opportunity for additional analysis in a TEM.

3.2 Electron Tomography Imaging Modes

A transmission electron microscope provides many different modes to probe the sample. However, not all TEM imaging modes can provide an image intensity that is linearly related to the thickness and density of the sample. As a result, there are a considerable number of possible ET modes which could be applied for electron tomography. (Detailed information about 2D TEM/STEM techniques mentioned in the following can be found in Ref [14].) These modes mainly are classified into two distinct modes to form the image: parallel conventional (TEM) and convergent (STEM) illumination as illustrated in Fig 3.7. For parallel illumination, ET imaging modes include bright-field (BF) ET [16, 17, 18, 19, 20, 21], annular dark-field (ADF) ET [22], energy-filtered (EF) ET [23, 24, 25, 26, 27, 28] and diffraction ET imaging [29, 30]. For convergent illumination, high-angle annular dark-field (HAADF) ET and energy-dispersive X-ray (EDX) ET are also applied [31, 32].

ET can be classified into three groups depending on the structural nature of the objects [33, 34]. The details of the three imaging modes for ET are described as follows:

- (1) Diffraction contrast imaging. These modes are used to obtain the shape of the object. Thus they include all modes that reliably can attain the surface contour of an object. The most common of these modes is the BF-ET (after contrast inversion). BF-ET imaging mode can be directly applied to the amorphous materials such as biological and polymeric samples. However, for the crystalline materials, this method will induce artifacts, since the contrast does not satisfy the linear intensity dependence requirement. Conventional dark-field ET and STEM mode can also be included in diffraction contrast imaging modes;
- (2) Atomic number contrast imaging. The local sample density is related to the atomic-number rather than mass density. HAADF-STEM is a common example;
- (3) Elemental-mapping. These methods involve spectroscopic imaging using a suitable spectrometer to achieve spectroscopic electron tomography. These methods have been applied in both EFTEM-ET and EDX mapping. The main advantage compared to

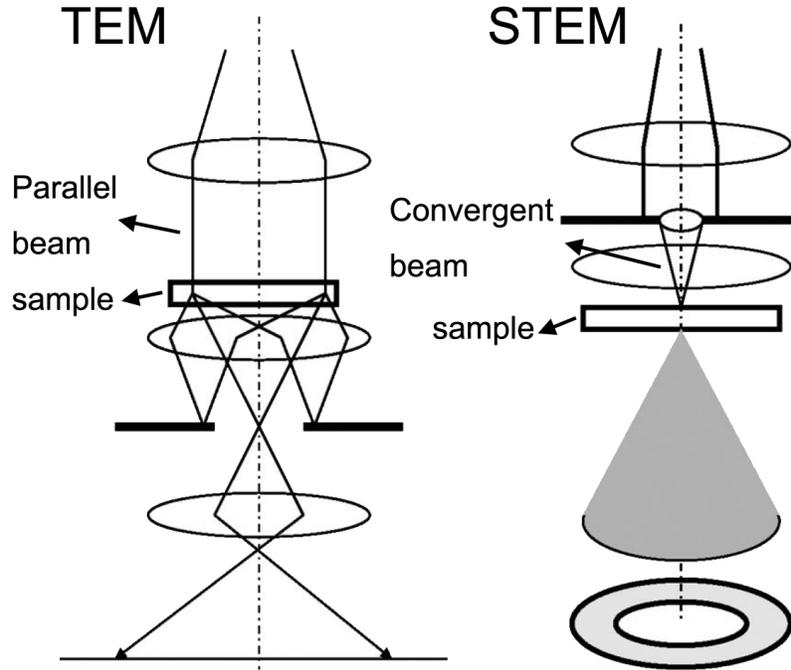


Figure 3.7: The illustration of parallel (TEM) and convergent (STEM) illumination modes respectively in EM. Revised from Ref [15].

other imaging methods is that elemental mapping allows the highest thickness limit and to obtain a certain elemental information from a sample.

3.3 Summary

In this chapter, the main 3D electron tomography sample preparation methods were reviewed. A new method to prepare the tomography sample by combining the advantage of both the thin film deposition method and FIB micro-sample technique was reported. The sample preparation is not particularly difficult. This method can easily reach $\pm 75^\circ$ tilt range which is sufficient in many practical applications [9]. The method alleviates the irradiation damage and Ga^+ ion contamination by reducing the milling time of the sample in the FIB system. Finally, the common electron tomography imaging modes used in practical application have been summarized and classified.

Bibliography

- [1] Frank, J.. Three-Dimensional Electron Microscopy of Macromolecular Assemblies, Academic Press, San Diego (1996)
- [2] Ayache, J., Beaunier, L., Boumendil, J., Ehret, G., and Laub, D.. Sample Preparation Handbook for Transmission Electron Microscopy. Springer Science (2010)
- [3] Giannuzzi, L. A, and Stevie, F. A.. Introduction to Focused Ion Beams: Instrumentation, Theory, Techniques and Practice. Springer Science+Business Media, Inc. Boston (2005)
- [4] Hayashida, M., Terauchi, S., and Fujimoto, T.. *Micron*, 42:307-310 (2011)
- [5] Kawase, N., Kato, M., Nishioka, H., and Jinnai, H.. *Ultramicroscopy*, 107:8-15 (2007)
- [6] Kamino, T., Yaguchi, T., Konno, M., Ohnishi, T., and Ishitani, T.. *J. of Electron Microscopy*, 53(6):583-588 (2004)
- [7] Kawase, N., Kato, M., Nishioka, H., and Jinnai, H.. *Ultramicroscopy*, 10:8-15 (2007)
- [8] Hayashida, M., Terauchi, S., and Fujimoto, T.. *Micron*, 41:540-545 (2010)
- [9] Midgley, P. A., and Dunin-Borkowski, R. E.. *Nature Mater.*, 8:271-280 (2009)
- [10] Wang, X., Lockwood, R., Vick, D., Li, P., Meldrum, A., Malac, M.. *Micro. Res. Tech* DOI: 10.1002/jemt.22044
- [11] Kawas, N., Kato, M., Nishioka, H., Jinnai, H.. *Ultramicroscopy*, 107:8-15 (2007)
- [12] Wang, J., Wang, X. F., Li, Q., Hryciw, A., and Meldrum, A.. *Philos. Mag.*, 87:11 (2007)
- [13] Egerton, R. F.. Electron energy-loss spectroscopy in the electron microscope. 2nd. Plenum Press. New York (1996)

- [14] David B. Williams and C. Barry Carter, *Transmission Electron Microscopy: A Textbook for Materials Science*, 2nd edition, Springer (2009).
- [15] Friedrich, H., de Jongh, P. E., Verkleij, A. J., and de Jong, K. P.. *Chem. Rev.*, 109:1613-1629 (2009).
- [16] Friedrich, H., Sietsma, J. R. A., de Jongh, P. E., *et al.*, *J. Am. Chem. Soc.*, 129:10249-10254 (2007).
- [17] Ersen, O., Parmentier, J., Solovyov, L. A., *et al.* *J. Am. Chem. Soc.*, 130:16800-16806 (2008).
- [18] Koster, A. J., Ziese, U., Verkleij, A. J., Janssen, A. H., and de Jong, K. P.. *J. Phys. Chem. B*, 104:9368 (2000)
- [19] de Jong, K. P., and Koster, A. J.. *Chem. Phys. Chem.*, 3:27 (2002)
- [20] Janssen, A.H., Koster, A. J., and de Jong, K. P.. *Angew. Chem. Int. Ed.*, 40:1102 (2001)
- [21] Ersen, O., Werckmann, J., Houille, M., Ledoux, M. J., and Pham-Huu, C.. *Nano Lett.*, 7:1898 (2007)
- [22] Bals, S., Van Tendeloo, G., and Kisielowski, C.. *Adv. Mater.*, 18:892 (2006)
- [23] Midgley, P. A., and Weyland, M.. in *International workshop on Strategies and advances in Atomic-Level Spectroscopy and Analysis (SALSA)*, Elsevier Science Bv, Guadeloupe, Guadeloupe, p 413 (2002)
- [24] Mobus, G., and Inkson, B. J.. *Appl. Phys. Lett.*, 79:1369 (2001)
- [25] Mobus, G. *et al.* in *International workshop on Strategies and advances in Atomic-Level Spectroscopy and Analysis (SALSA)*, Elsevier Science Bv, Guadeloupe, Guadeloupe, p 433 (2002)
- [26] Weyland, M., Yates T. J. V., *et al.* *Sr. Mater.*, 55:26 (2006)
- [27] Gass, M. H., Kozilo, K. K., Windle, A. H., and Midgley, P. A.. *Nano Lett.*, 6:376 (2006)
- [28] Yurtsever, A., Weyland, M., and Muller, D. A.. *Appl. Phys. Lett.*, 89:151920 (2006)
- [29] Kolb, U., Gorelik, T., *et al.* *Ultramicroscopy*, 107:507 (2007)

- [30] Kolb, U., Gorelik, T., and Otten, M. T.. *Ultramicroscopy*, 108:763 (2008)
- [31] Midgley, P.A., Weyland, M., Thomas, J. M., and Johnson, B. F. G.. *Chem. Commun.*, 10:907-908 (2001)
- [32] Weyland, M., Midgley, P. A., and Thomas, J. M.. *J. Phys. Chem. B*, 108:7882 (2001)
- [33] Xu, X. J., Saghi, Z., Gay, R., and Mobus, G.. *Nanotechnology*, 8:18 (2007)
- [34] Friedrich, H., McMartney, M. R., and Buseck, P. R.. *Ultramicroscopy*, 106:18 (2005)

Chapter 4

Experimental Results and Discussion

This chapter focuses on the experimental results. The organization of this chapter follows the steps of increasing experimental difficulty and scientific interest. All 3D reconstruction results in this chapter are obtained and visualized using TEMography.comTM software package (www.temography.com).

4.1 Latex Nanoparticles

The well-separated nanoscale latex spheres are optimal for testing tomographic reconstruction methods. The latex particles purchased from *Purest Colloids, Inc.* were spherical, with a nominal diameter of 90 nm. The concentration and amount of spatial overlap of the nanoparticles can be controlled by adjusting the dilution ratio of water and the latex nanoparticle solution.

The electron tomography sample was prepared using the grid support method introduced in Section 3.1.1. The nanoparticles were diluted using distilled water. Then the suspension with nanoparticles was agitated in an ultrasonic bath for twenty minutes in order to control the dispersion of the nanoparticles. Finally, a drop of the suspension with nanoparticles was deposited onto the TEM grid with the holey carbon. The sample was then left to dry in laboratory air before it was examined in the transmission electron microscopy (TEM).

High angle annular dark field (HAADF) STEM was used to image the latex nanoparticles. The HAADF image in Fig 4.1 shows that the nanoscale latex spheres are well

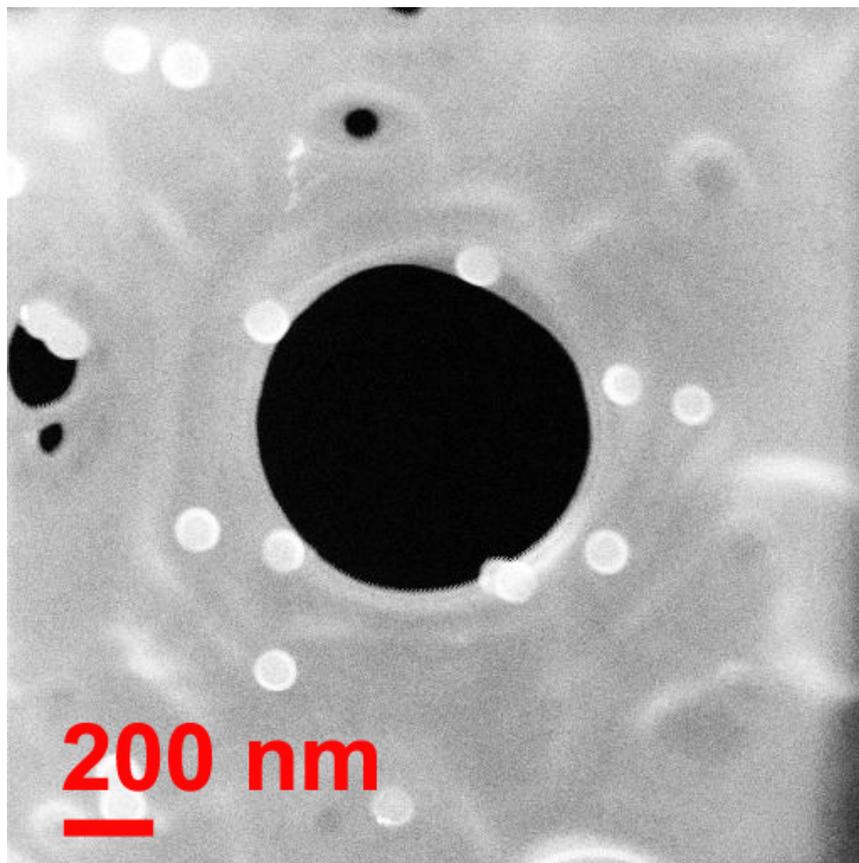


Figure 4.1: An example HAADF STEM image from the tilt series shows the 2D distribution of latex particles. A series of tilted images are collected for the 3D reconstruction. The tilt range covers from -73° to $+73^\circ$ with 2° increments.

separated. An image series was then collected over a tilt range $\pm 73^\circ$ at 2° intervals using JEOL 2200FS TEM. This tilt series formed the basis of the 3D reconstruction.

The tomographic reconstruction was obtained by the FBP method. The FBP method was chosen because the computation time required by the FBP method is much less than for the SIRT method. The origin of the artifacts is easier to interpret for the FBP reconstruction as compared with SIRT (Section 1.3).

The final visualization of the reconstructed result was obtained by comparing projection images of reconstruction to the experimental image at zero tilt. The best threshold was estimated by a visual comparison. Although this method is not quantitative compared to QI-factor method adopted in Chapter 2, it is commonly used in practice. These initial tests

were primarily used to investigate the visualization of the morphology of the reconstructed object. Volume rendering was used to present the visualization of the reconstructed results according to the discussion in Chapter 2, as shown in Fig 4.2.

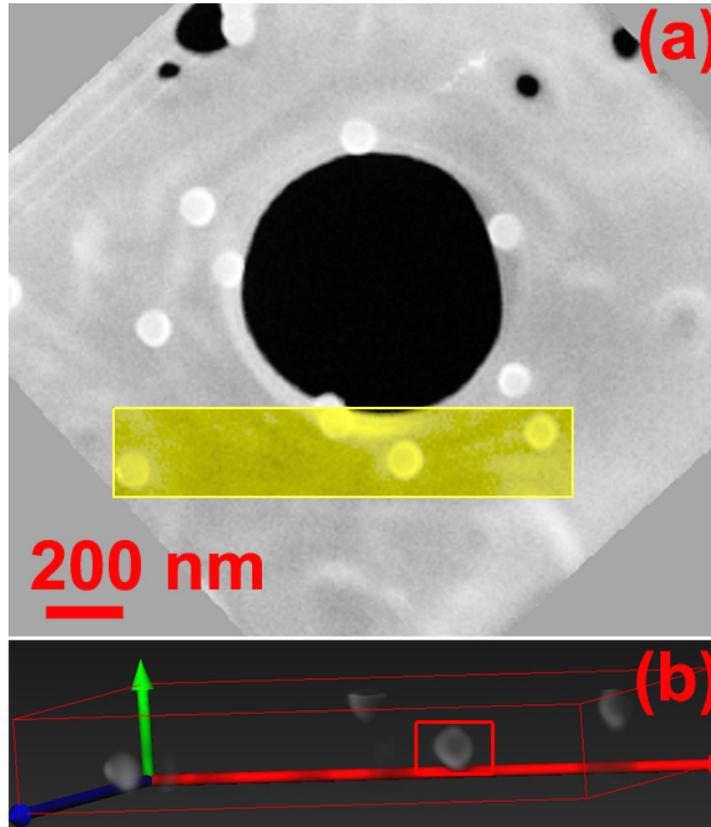


Figure 4.2: Tomographic 3D reconstruction results. The 3D reconstruction result in (b) corresponds to the area marked by the yellow rectangle in (a). The latex nanoparticle marked in the red box shown in (b) shows that the elongation effect is not obvious in the reconstruction result. x-axis (red): 1216 nm; y-axis (green): 185 nm; z-axis (blue): 626 nm.

The reconstructed area shown in Fig 4.2b corresponds to the yellow rectangle in Fig 4.2a. The reconstruction shows that the latex nanoparticles have a spherical shape, which is consistent with the expectation of a latex sphere. The latex nanoparticle shown in the red box in Fig 4.2b does not have any obvious elongation caused by the limited tilt range. The result is consistent with the conclusions in Refs [1, 2, 3, 4, 5, 6], which suggested that a tilt range of $\pm 75\text{--}80^\circ$ should minimize elongation artifacts.

However, the faint contrast near the red box and near the left sphere in Fig 4.2b shows

that some issues do exist in the reconstruction. As discussed in the Chapter 2, this kind of blurring is not caused by the reconstruction method (FBP), nor is it due to the missing wedge effect. Considering that the image contrast for the latex particles on a carbon film is only a little higher than that for the support carbon film, the “artifact” may be caused by the intensity arising from the carbon support. Although the carbon film is thin, it still contributes image contrast (or intensity) to the 2D projected image [7].

In summary, the reconstruction of non-overlapping nanoscale latex particles on a carbon thin film confirms that the visualization of the reconstruction obtained by a visual comparison method is acceptable since the morphology of the latex nanoparticles is correct. The grid support method is an efficient way to prepare a 3D tomography sample, especially when the nanoparticles can be supported directly on the carbon film. The missing wedge problem is apparently insignificant for the tilt range of $\pm 73^\circ$ [1, 2, 3, 4, 5, 6]. However, the support film may contribute to the final reconstruction, resulting in the slight blurriness in the reconstruction [7].

4.2 Au nanocrystal multilayer

In order to connect the simulation results (discussed in Chapter 2) to “real world” TEM tomography of nanocomposites, the tomographic reconstruction of a sample made of Au nanoparticles arranged in a multilayered structure within a SiO_2 matrix was examined. Embedded Au nanoparticles are fairly resistant to electron beam irradiation [8]. The image contrast of Au nanoparticles is high compared to the SiO_2 matrix in the recorded images. This should help minimize blurring appearing in the reconstruction.

A multilayered film consisting of alternating layers of silicon dioxide (thickness = 25 nm) and gold (thickness = 2.5 nm) was prepared by sequential electron beam and thermal evaporation of SiO_2 and Au, respectively, on a standard silicon wafer (p-type, 20 Ω -cm, [100] oriented). Five periods of the SiO_2 -Au bilayer were capped by a final 50-nm thick protective SiO_2 film. After deposition, the sample was annealed at 900°C for 2 minutes in an N_2 atmosphere and rapidly cooled back to room temperature after trial and error. The annealing step was intended to induce formation and growth of well-defined Au particles. A needle-shaped electron tomography sample was prepared using the focused ion beam (FIB) micro-sampling technique presented in Section 3.1.3. The tilt range of the sample can reach $\pm 90^\circ$. Therefore, the elongation effect due to the missing wedge does not affect the tomographic reconstruction.

A tomographic series of Au nanoparticles within SiO_2 was acquired using Hitachi

HF3300 microscope at an accelerating voltage of 300 kV in bright field (BF) STEM mode. A BF STEM image is shown in Figure 4.3a. The sample was tilted about a single axis (tilt range -90° to $+90^\circ$ with a 2° increments). Image alignment and reconstruction of the tilt series were carried out using the FBP and the SIRT reconstruction methods. Visualization of the reconstructed results was done by both volume and isosurface rendering of the tomographic data cube.

Obviously, the experimental data brings an additional level of uncertainty to the evaluation of the reconstruction result, compared to computer-generated phantom data in Chapter 2. For example, the projected images contain electron shot noise due to electron arrival statistics, and the image pixel intensity can be nonlinear with the local mass thickness (or geometrical thickness) of the sample. Unlike the case for the simulated data, the shape and volume filling fraction of the particles in an experimental sample may not be exactly known, although the latter can sometimes be estimated from sample preparation parameters. The objective of this section is to determine whether the results on the simulated samples apply to a “real” reconstruction of an unknown object.

The visualization of the reconstruction of the Au/SiO₂ multilayer sample was obtained using QI-factor method to set the threshold value and was visualized using volume rendering. The calculation of QI was introduced in Chapter 2. Here, the process of calculating QI to attain the visualization is shown in Fig 4.3. Fig 4.3d has the highest QI value. Hence, the threshold value for Fig 4.3d was adopted to visualize the reconstruction. The volume-rendered reconstruction for different angles is shown in Fig 4.4.

The morphology and spatial distribution of Au nanoparticles are revealed by the reconstruction (Fig 4.4). The Au nanoparticles are well separated and the size of Au nanoparticles is not uniform. No obvious artifacts are visible. Therefore, the reconstruction of Au nanoparticles sample is suitable for the further investigation.

In order to investigate whether electron tomography can be used for *quantitative* measurements of volume filling fraction, the volume fraction of Au nanoparticles contained in the reconstructed volume (Fig 4.5) was estimated. The volume filling fraction obtained from the tomographic reconstruction was then compared to the filling fraction estimated from sample preparation parameters. Each deposited gold layer was nominally 2.5 nm thick, while the SiO₂ layers were 25 nm thick. The volume fraction of Au within the multilayered part of the sample was estimated by direct conversion from the molar volume, assuming that all the deposited Au is present in the particles rather than dissolved in the matrix. The reconstructed volume contains some areas of the sample away from the Au nanoparticle multilayer, which must be factored into the fraction of particles in the total reconstructed

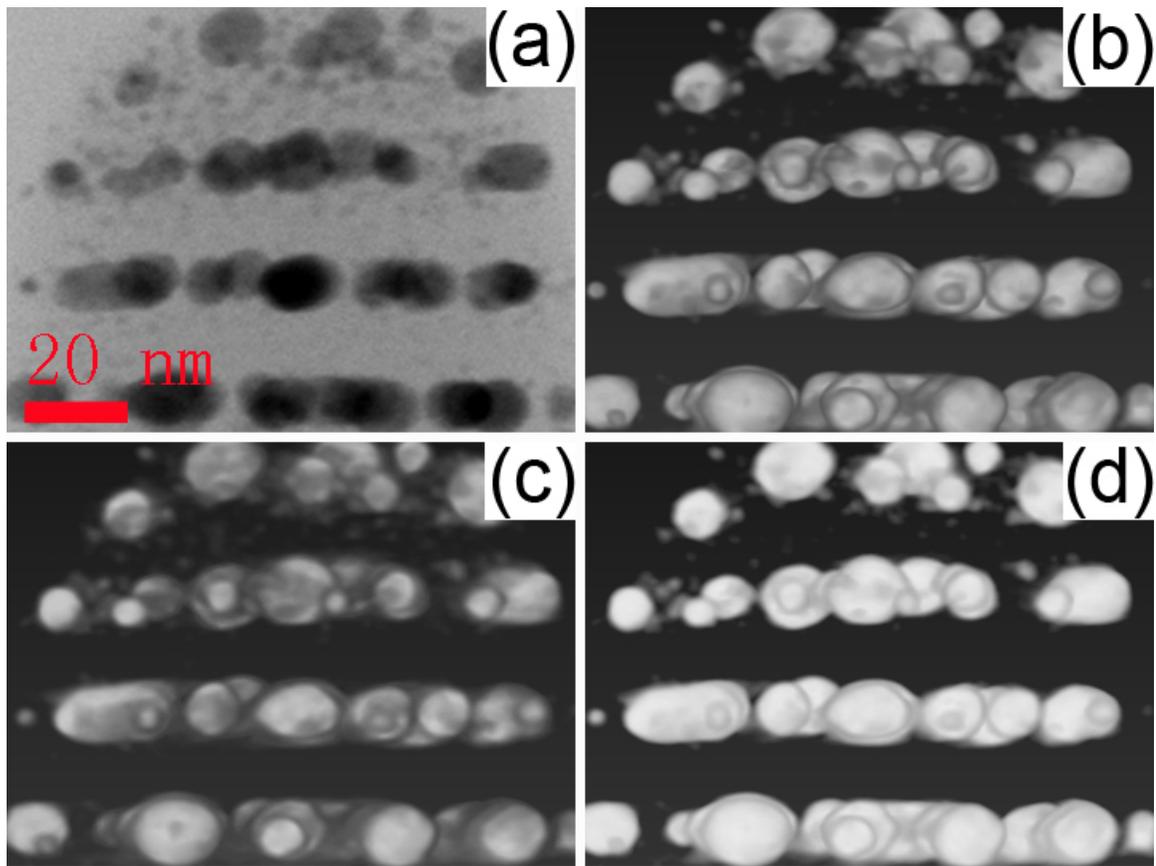


Figure 4.3: (a) Bright field STEM image of a needle-like sample consisting of Au particles embedded in SiO_2 . The Au particles appear dark. The other images show the SIRT reconstruction results with a quality index of (b): 0.32, (c): 0.28 and (d): 0.38.

volume. The volume fraction of Au particles within the reconstructed region in Fig 4.5 was thus estimated to be around 16%.

The results in Fig 4.5 are similar to the observations previously obtained on the simulated data in Chapter 2. Fig 4.5 zooms in on one of the Au layers and shows a comparison of the FBP and SIRT reconstruction results for a few Au particles. First, both FBP and SIRT reconstructions have apparent distortions when thresholding by setting the Au nanoparticle volume filling fraction to the 16% expected from the deposition parameters. Similar to the simulated phantom object results, although less severe, the FBP result (Fig 4.5a) with the “correct” volume-filling fraction bears no resemblance to how one would expect the Au/ SiO_2 sample should look. As with the simulated data, the reconstruction with a

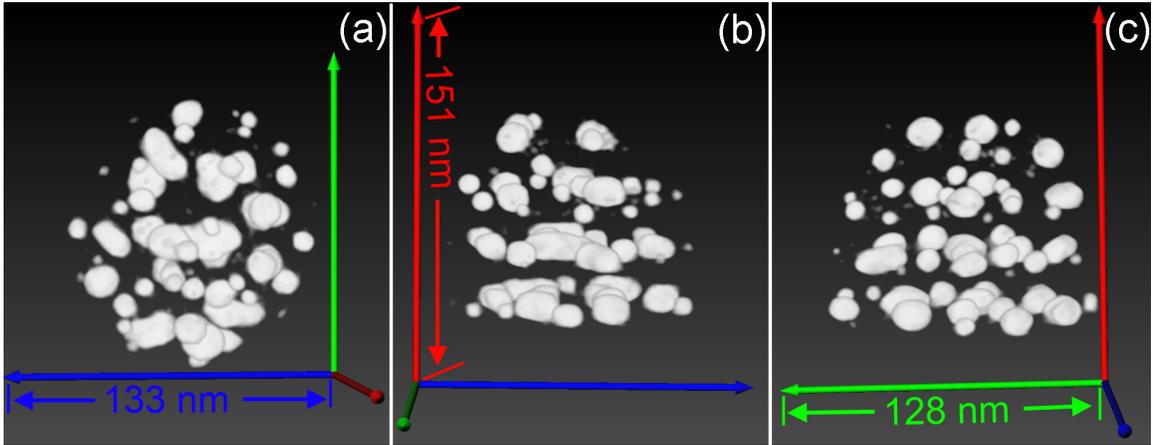


Figure 4.4: 3D visualization of Au particles by SIRT viewed along (a) the x-axis, (b) the y-axis, and (c) the z-axis.

much better QI (Fig 4.5b and 4.5d) gives lower filling fractions but much better quality reconstructed images. The visualization obtained via the QI method appears to resemble the original Au/SiO₂ sample more closely than that obtained by setting the volume fraction equal to the estimated volume fraction.

Two conclusions on the visualization of the reconstruction for *both* FBP and SIRT can be obtained based on the above discussion. First, the proposed QI method to determine the visualization of the reconstruction appears reasonable for both simulated results and experimental results. Second, the volume fraction is not a suitable parameter used to determine the threshold for the visualization of the reconstruction in *both* simulated and experimental tomographic reconstructions.

Two methods were adopted to determine the final visualization of the reconstruction. One is to use the QI value, whereas the second method (referred as the visual method) is to visually compare the reconstruction to the projected images, as in the preceding Section 4.1. In the literature, the second method is adopted more commonly than the QI method. The reason that the visual method is commonly used is based on the two following aspects:

- (1) The reconstruction obtained by the visual method usually does not have significant difference in shape and morphology compared to the reconstruction obtained by the QI factor method.
- (2) Commercial electron tomography reconstruction software usually includes manual threshold adjustment to visualize the reconstruction, and no quantitative method

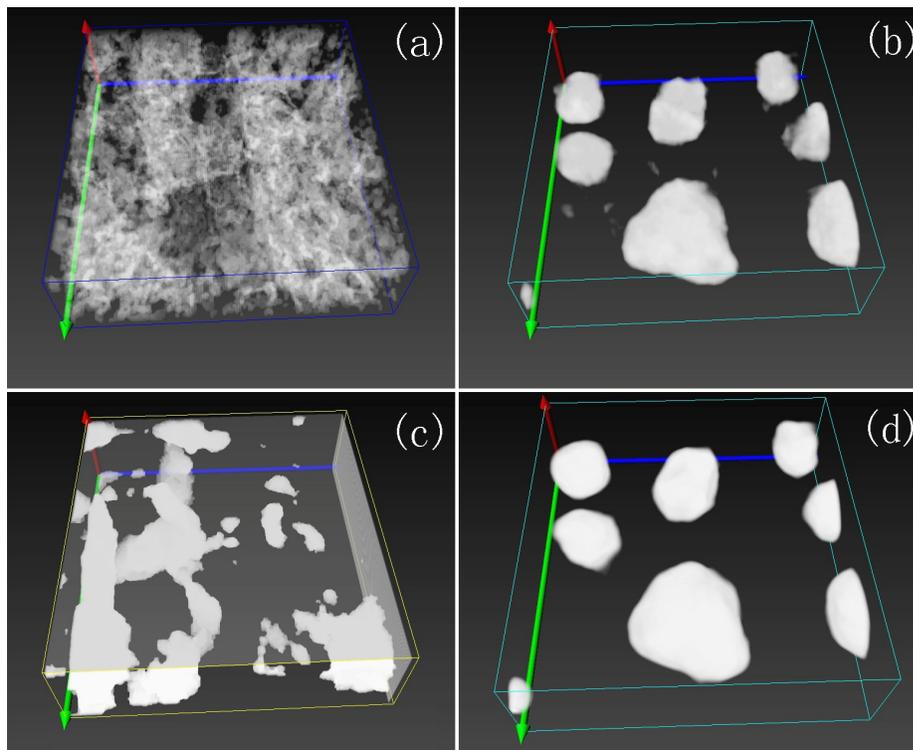


Figure 4.5: Visualization of the tomographic reconstructions for the Au nanoparticles. Panels (a) and (b) show the FBP method, and (c) and (d) used SIRT. (a) volume rendering with the 16% estimated nanoparticle filling fraction. (b) image of the same region, thresholding for the best quality index. (c) and (d) are, again, thresholding according to the estimated “real” filling fraction and the best- QI filling fraction, respectively, but this time for SIRT. Scale: x -axis (red): 16 nm; y -axis (green): 58 nm; z -axis (blue): 54 nm. The QI and volume fractions are 0.35 and 13%, for (b), while for (d) they are 0.38 and 12%, respectively.

is usually available.

Finally, the appearance of the filamentary texture in the reconstructed results in Fig 4.5a was investigated. The extrapolation between Fourier planes during reconstruction is a possible origin of the filamentary structures. The intensity of voxels located between two adjacent Fourier planes is not directly attained from the experimental data due to the fact that the tilt angle is not continuous (the data was obtained in 2° steps). The information between the individual experimental images, corresponding to slices through Fourier space, are instead obtained by interpolation, resulting in the appearance of the filamentary texture in Fig 4.5a [9].

4.3 Si nanocrystals embedded in glass

Silicon nanoparticles embedded in an amorphous silicon oxide matrix are a promising material for integrated photonics [10, 11, 12, 13, 14, 15], memory devices [13, 16] and for multiple-exciton generation in solar cells [17, 18]. This system shows efficient room temperature photoluminescence in the range of 700-900 nm [19]. Net optical gain, which has been reported [10] to be comparable to that of direct-band-gap quantum dots, demonstrates a potential application for a silicon laser. The size and 3D morphology of the Si nanocrystal composite can strongly affect the macroscopic properties of the material [20, 21].

A $\text{SiO}_{0.6}$ thin film sample was prepared using the cross-sectional method introduced in Section 3.1.2. The $\text{SiO}_{0.6}$ was fabricated by simultaneous electron beam and thermal evaporation of Si, SiO, and SiO_2 onto fused quartz substrates. The SiO_x film thickness was 200 nm. The $\text{SiO}_{0.6}$ film was annealed in a 95% N_2 + 5% H_2 environment at 1.2 atm at $T_a = 1100$ °C. Thermal annealing at temperature ≥ 900 °C results in phase segregation and the formation of crystalline silicon nanocrystals embedded in SiO_2 [22]. A electron tomography sample was prepared using the physical thinning method described in Section 3.1.2.

Imaging Si nanocrystals (NCs) in an SiO_x matrix by conventional TEM techniques is somewhat difficult due to the low mass contrast between Si-NCs and the matrix. A more optimal method is to take advantage of the different plasmon loss energies for Si and SiO_x [23], as discussed in Section 1.4.2. Therefore, the sample was imaged using the energy-filtered TEM mode of the JEOL 2200 FS TEM/STEM equipped with an Omega filter and operated at an accelerating voltage of 200 kV. The Si nanoparticles were visualized in energy-filtered bright field STEM mode by placing a 4 eV-wide energy-selecting slit around a 17 eV energy loss window. The Si-NCs appear bright when imaged using the bulk plasmon loss of Si at 17 eV, as shown in Fig 4.6.

Radiation damage in SiO_x thin films poses a limitation on the acquisition of images for ET. Radiolysis, knock-on damage and surface sputtering of oxygen atoms [24, 25] can cause the Si/O to increase and result in the growth of silicon nanocrystals. Irradiation of a 50-nm thick amorphous SiO_2 thin film with 200 keV electrons to an accumulated dose of 1×10^4 C/cm² results in the formation of Si nanocrystals as oxygen is removed from the sample. Oxygen can be almost entirely removed from a 15-nm thick amorphous SiO_2 layer for an irradiation dose exceeding 3×10^5 C/cm² at 100 keV [26], precluding the investigation of oxide-embedded Si nanocrystals at high doses. Therefore, the experimental tilt range was limited to $\pm 50^\circ$ with 2° intervals in order to minimize irradiation damage.

After collecting a series of tilted EFTEM images, the tomographic result was obtained

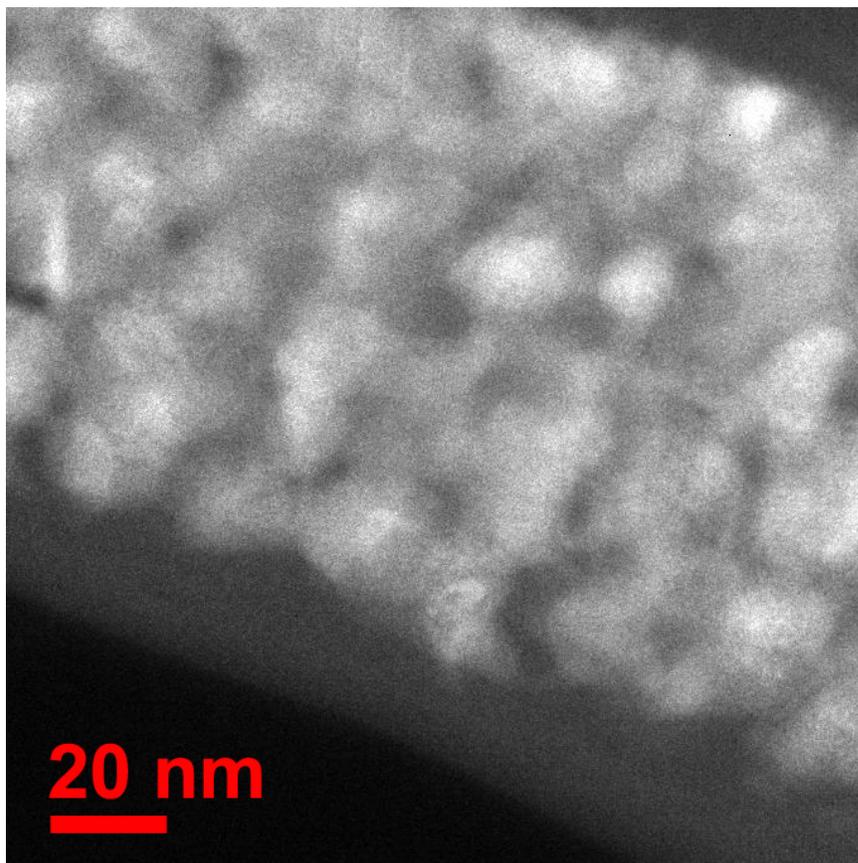


Figure 4.6: The EFTEM image using a 4 eV window centered at 17 eV shows Si nanoparticles in the $\text{SiO}_{0.6}$ sample. This confirms that the Si nanoparticles are formed and can be visualized in EFTEM.

using the FBP reconstruction method. The reasons FBP was used were (1) FBP is less time-consuming than SIRT; and (2) the tomographic contrast in the reconstructed result obtained by FBP is easier to interpret. As discussed in Section 4.2, the visualization of the reconstruction used the visual method provided by the software. The visualization was presented using volume rendering, as discussed in Chapter 2.

The Si-NCs do not appear to have a simple spherical symmetry (Fig 4.7). The shape of Si-NCs in the reconstruction is affected by elongation due to the missing wedge. However, a single spherical zirconia nanoparticle affected by the missing wedge has an elliptical shape [6] under the same tilt range in the 3D reconstruction. The HRTEM (Fig 4.8) image also shows the irregular shape of Si-NCs. Therefore, complex rather than spherical morphologies of Si-NCs were observed.

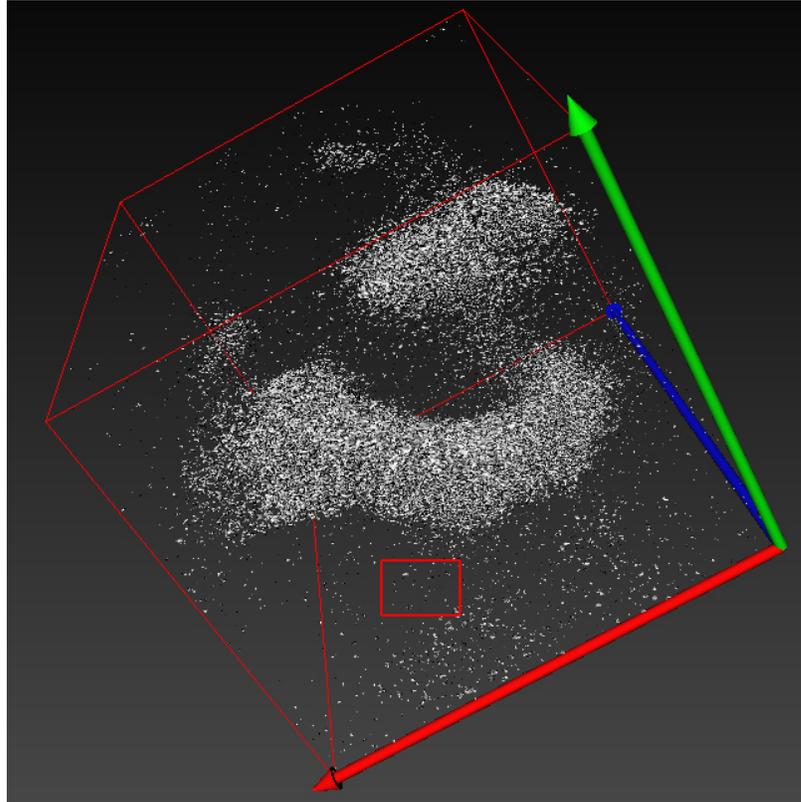


Figure 4.7: Three dimensional electron tomography reconstruction of the silicon distribution (volume rendering) in the $\text{SiO}_{0.6}$ film by the FBP method. Volume rendering was used for this figure. The point-like contrast that appears in the red rectangle may originate from the low contrast of Si-NCs. The lengths of the axes are 53.90 nm (red, bottom vector), 51.15 nm (green, top vector), and 66.00 nm (blue, middle vector) [27].

The tomographic result (Fig 4.7) shows the presence of interconnected silicon-rich regions, as presented in [27]. However, the Si-NCs may have up to a factor of ~ 1.9 apparent elongation [28] along the z -axis (blue axis in Fig 4.7) due to the fact that the maximum tilt angle was 50° . Therefore, the interconnectivity is somewhat uncertain from the ET reconstruction result, due to the missing wedge effect. However, in this work, the interconnectivity was confirmed from the corresponding HRTEM result (Fig 4.9) which shows interconnected particles.

Secondly, the tomographic result shows point-like contrast, as depicted in the red rectangle in Fig 4.7. This type of artifact may originate from the low contrast in Fig 4.6, similar to the artifact or blurring in the reconstruction of the latex particles, as shown in Fig 4.2.

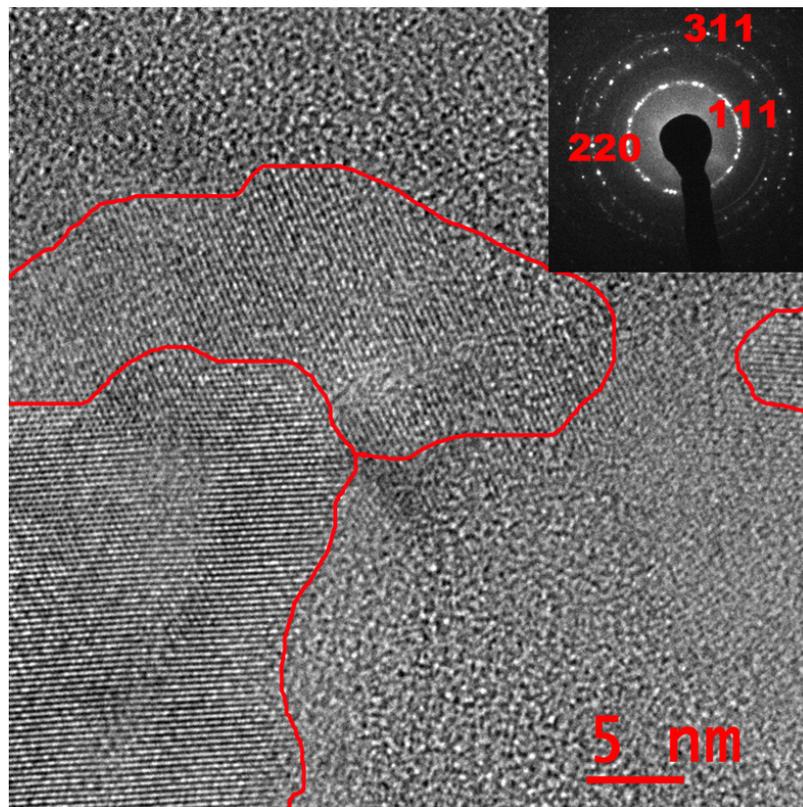


Figure 4.8: HRTEM micrograph of the SiO_{0.6} thin film. The HRTEM image indicates the presence of complex morphologies of silicon nanocrystals. The diffraction pattern in the inset is consistent with diamond-structure silicon [27].

Although the contrast of SiO₂ is low compared to the contrast of Si-NCs, it still contributes image contrast to the 2D projected image. Thus, the point-like contrast may be caused by the image contrast from amorphous SiO₂ [7].

In this experiment, electron beam irradiation limits the resolution of the reconstruction. Beam damage limits the number of tilt images that can be collected at an acceptable signal-to-noise ratio, resulting in the missing wedge effect or extensive extrapolation in Fourier space. One way to reduce the effect of electron beam irradiation damage is to image Si-NCs in EF-STEM mode instead of EFTEM, since the total electron dose for the EF-STEM modes is considerably less than in EFTEM mapping [29].

In addition, the plasmon energy loss imaging method affects the resolution of the reconstruction. The delocalization of the plasmon scattering process [30] limits the spatial resolution at 17eV to $\sim 1\text{--}3$ nm [31] in the projected images. Compared to the effect of the

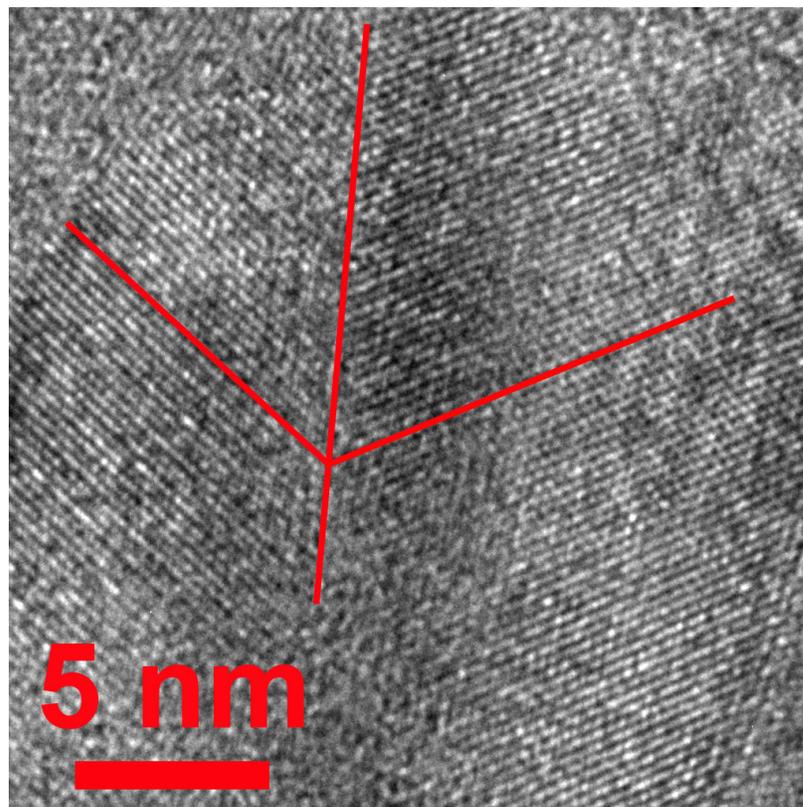


Figure 4.9: HRTEM micrograph of the $\text{SiO}_{0.6}$ thin film. The HRTEM image shows particles touching together but with different lattice fringe orientation, which indicates the presence of interconnected silicon nanocrystals.

missing wedge, delocalization is less severe in this experiment. For example, assuming that a Si-NC is a sphere with a 10-nm diameter, the reconstruction using the plasmon energy loss imaging method without the missing wedge results in a particle with a diameter less than 13 nm. The apparent diameter of the Si-NC with a 50° maximum tilt angle is 19 nm [28].

The experimental results show that energy-filtered electron tomography can be used to characterize the 3D morphology and distribution of Si-NCs. The Si-NCs present in $\text{SiO}_{0.6}$ thin film do not have a simple spherical symmetry. Instead, more complex morphologies were observed.

4.4 Er-doped Si nanocrystals

In this section, a scientifically interesting but especially challenging nanocrystal composite was chosen for electron tomography: Er-doped Si nanocrystals in an SiO₂ matrix. This material system has potential applications in erbium-doped waveguide amplifiers [32] and microphotonic light emitters, and may possibly lead to a silicon-compatible laser [33]. This system is challenging for tomography because two kinds of nanoclusters, Si nanocrystals (NCs) and Er-rich clusters, need to be located spatially.

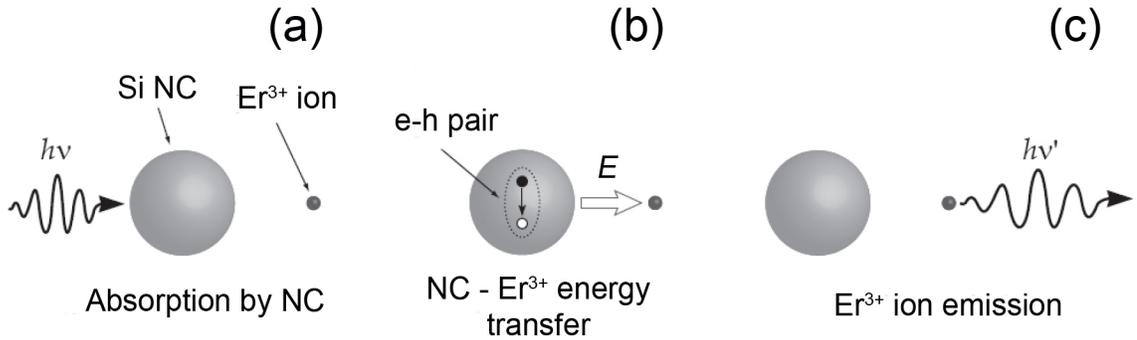


Figure 4.10: Illustration of the sensitization process. In (a), an incident photon is absorbed by a silicon nanocrystal, exciting an electron-hole pair. This electron-hole pair recombines non-radiatively and transfers its energy to the Er³⁺ ion as shown in (b) by a Förster-type mechanism. Finally, the Er³⁺ ion emits light with a wavelength near 1.5 μm in (c). (After Ref [34]).

The Er-doped Si-NC system has been of interest for many years [35, 36, 37, 38, 39, 40, 41], because of its possible application in waveguide amplifiers. Essentially, Si-NCs can sensitize the fluorescence arising from the Er³⁺ $^4\text{I}_{13/2} - ^4\text{I}_{15/2}$ $4f$ -shell transition (Figure 4.10). By using nanocrystals as sensitizers for the optically active erbium, the absorption cross-section of the Er³⁺ ions (\sim on the order of 10^{-20} cm²) effectively becomes similar to that of the Si-NCs ($\sim 10^{-16}$ cm²) [42]; a difference of about four orders of magnitude. A strong decrease in transfer efficiency with increasing distance was reported in Refs [43, 44, 45, 46], indicating an interaction characteristic length approximately 0.5 nm [43] for amorphous Si-NCs and approximately 2.1 nm for crystalline Si nanocrystals [44]. Recently, the characteristic transfer distance between crystalline Si-NC and Er³⁺ ions updated in Ref [47] shows the distance can be 6 to 7 nm.

Several physical and chemical vapor deposition methods can be used to grow SiO_{*x*} films lightly doped with Er³⁺ such as radio frequency co-sputtering [36] and thermal evaporation

[48]. Essentially, the result is a distribution of Si-NCs and Er^{3+} ions embedded in an oxide host. However, little is known about the distribution of Er^{3+} ions. As discussed above, this is important since the performance of the Si/Er system will depend on the physical distribution of NCs and Er^{3+} ions. Spectral evidence suggests that the Er^{3+} is located within the SiO_2 matrix and not inside the Si-NCs [49], although there has been no direct experimental confirmation of this.

Conventional ET techniques can obtain the spatial information and morphology of Si-NCs and erbium by recording Si nanoparticles and erbium 3D information sequentially from the same area. However, this method is confronted by the following difficulties:

- (1) Reliable spatial alignment at high spatial resolution of two reconstructions is difficult to achieve. Although the Si and Er information could come from the same area by imaging the same area sequentially, the tilt angle is normally not exactly the same due to the tilt value uncertainties and stage positioning errors. The differences in the experimental recording conditions results in alignment errors between the Si reconstruction and Er reconstruction. This can be significant compared to the desired resolution.
- (2) Radiation damage in the SiO_x thin film poses a limitation on the time for image acquisition, as discussed in Section 4.3. Recording Si and Er data sequentially from the same area increases the total electron irradiation on the sample, making it difficult to obtain satisfactory images with high signal–noise–ratio.

Here, a new electron tomography technique combined the HAADF and EF-STEM methods to simultaneously measure the 3D morphology of Si nanocrystals and Er-rich clusters in an Er-doped SiO_x matrix. This method takes advantage of the STEM mode and the fact that different detectors can acquire signals simultaneously. The two methods used for parallel STEM image acquisition were HAADF and EFSTEM. The HAADF is especially sensitive to the high atomic number of Er, as illustrated in Fig 4.11, while the EF-STEM at the Si plasmon loss energy can be used to image the Si-NCs. The fact that the two signals are obtained simultaneously eliminates the misalignment of the Si and Er reconstruction and reduces the irradiation damage.

The TEM sample for analysis was prepared using the method described in the preceding Chapter 3 to minimize the Ga^+ ion irradiation damage. To reduce artifacts arising from electron-beam damage and the electron-beam induced growth of Si nanocrystals and amorphous Si nanoparticles, as discussed in the preceding section, the tilt range in this experiment was limited to $\theta = -60^\circ$ to $+60^\circ$ with 4° increments. The $\theta = \pm 60^\circ$ maximum tilt

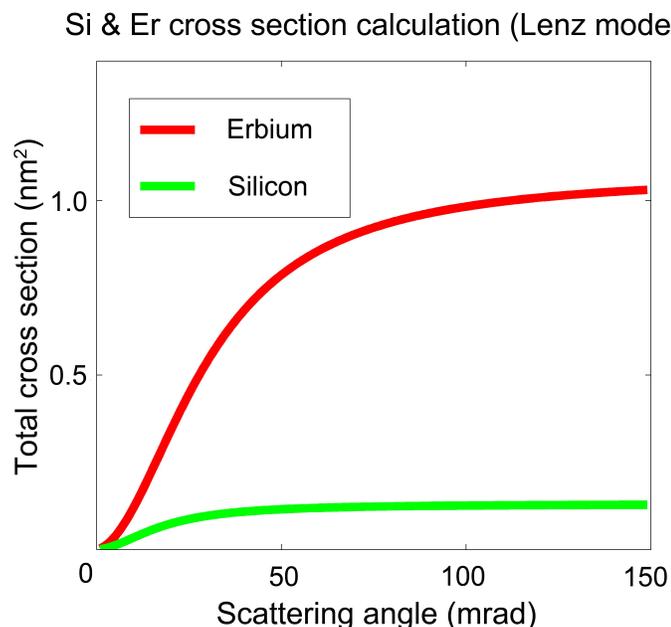


Figure 4.11: Calculation of total elastic cross section for Si (lower, green curve) and Er (upper, red curve), respectively, based on the Lenz model. Most contrast contribution in HAADF images will arise from Er. The total cross section is the cross section integrated from zero degree to x_0 degrees.

angle with 4° increments represented the largest number of images that could be collected under STEM conditions with an acceptable signal-to-noise ratio, before observable beam-induced changes in the sample morphology occurred. This procedure limited the electron irradiation dose to 4.4×10^3 C/cm².

After simultaneously collecting two sets of tilt series data from the same area, the tomographic result was obtained by FBP method. Although the quality of the reconstruction by SIRT is better than that by FBP for a $\pm 90^\circ$ tilt range, the FBP was chosen in this experiment for the same reason as discussed in Section 4.1. After obtaining the reconstruction, the reconstruction was initially visualized by volume rendering by the visual comparison method for the same reason as discussed in Section 4.2. In order to display the 3D combination reconstruction in the software, isosurface rendering had to be used to visualize the reconstruction, since the color of the combination reconstruction can be adjusted only in isosurface rendering. The shape of the reconstruction result visualized with isosurface rendering was based on the shape with volume rendering, as shown in Fig 4.12(b) and Fig 4.12(c).

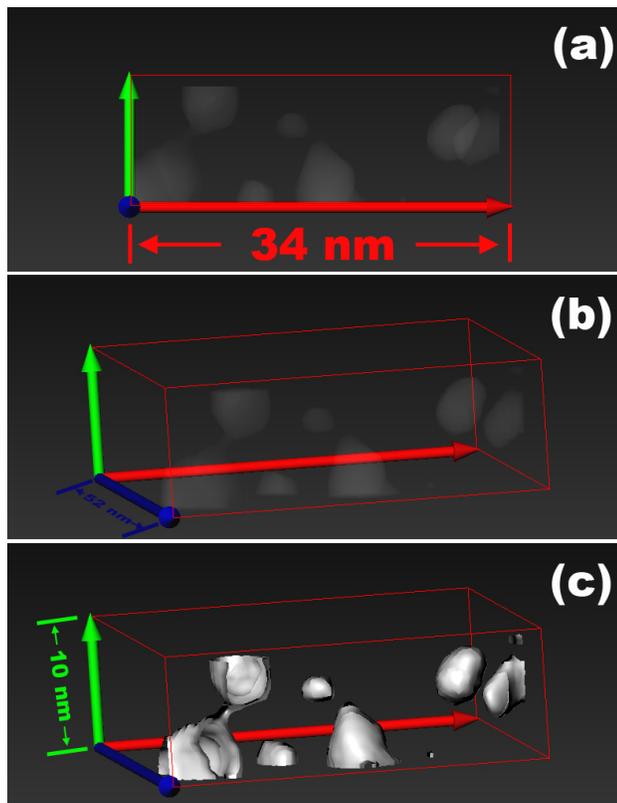


Figure 4.12: A 3D visualization of the reconstruction from HAADF visualized with volume rendering (a) and (b) and isosurface rendering (c). The shape of the Er cluster in (c) is based on the shape of Er visualized in volume rendering (b). The figure shows that the Er clusters are not uniform and two particles appear to be connected. The apparent connection between particles may be caused by the elongation effect due to the missing wedge, combined with small separation distances.

The reconstructed 3D images of the Er clusters (Fig 4.12) show the size, shape and spatial distribution of Er clusters. The size of Er clusters is not uniform. Two Er clusters that are connected to each other are shown in the reconstruction (Fig 4.12(b)). Similar to the discussion in Section 4.3, this phenomenon may be caused by the elongation effect due to the missing wedge, although the elongation is less severe here than in the Section 4.3, because the elongation can be theoretically as high as a factor of 1.55 [28] for $\pm 60^\circ$ compared to ~ 1.9 for $\pm 50^\circ$. Er cluster interconnection may be also caused due to the fact that the separation distance of Er clusters are small because particles spaced closer than 2.2 times their radius will erroneously appear to be a single particle as discussed in Chapter 2.

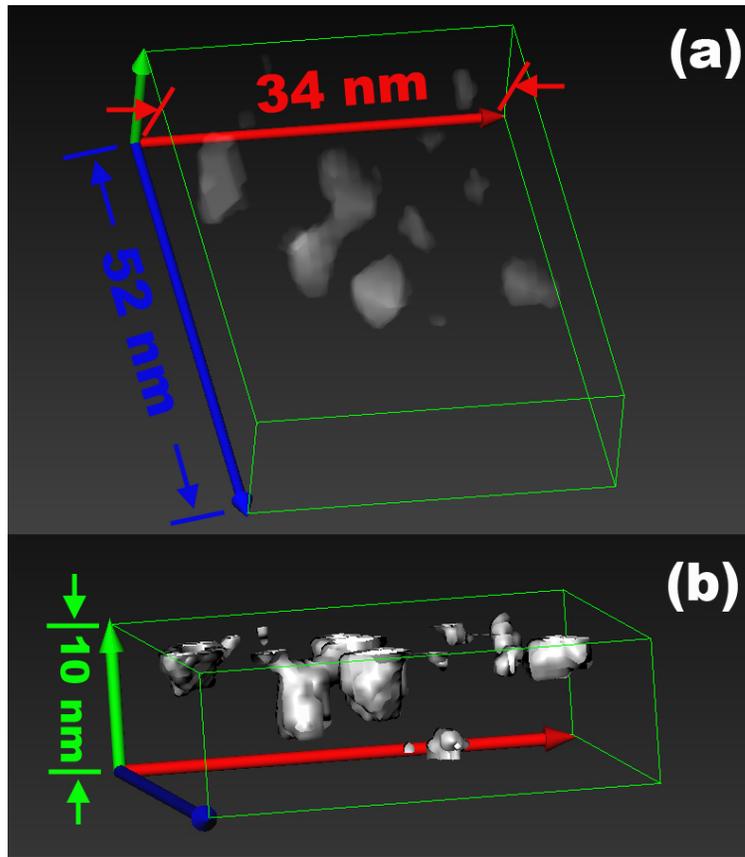


Figure 4.13: A 3D visualization of the reconstruction from EFTEM visualized with volume rendering (a) and isosurface rendering (b). Similar to HAADF reconstruction, the shape of Si-NCs visualized with isosurface rendering is determined based on the shape of Si-NCs with volume rendering. It shows that Si-NCs are not spherical in shape.

Thus, whether two Er clusters actually connect to each other in the reconstruction cannot be determined simply from the reconstruction.

The reconstructed 3D images of the Si-NCs (Fig 4.13) show the size, shape and spatial distribution of the Si-NCs. The Si-NCs have an irregular, non-spherical shape in general agreement with Ref [50] and the experimental result in the preceding section. Similar to the HAADF reconstruction situation, the Si-rich filamentary regions may be due to the missing wedge effect or close particle spacing.

The combined visualization of the reconstruction of the silicon and Er clusters (Fig 4.14) illustrates their spatial relationships. Little evidence of overlap between the Si-NCs and Er clusters is apparent in Fig 4.14(b). One erbium cluster located at the boundaries of

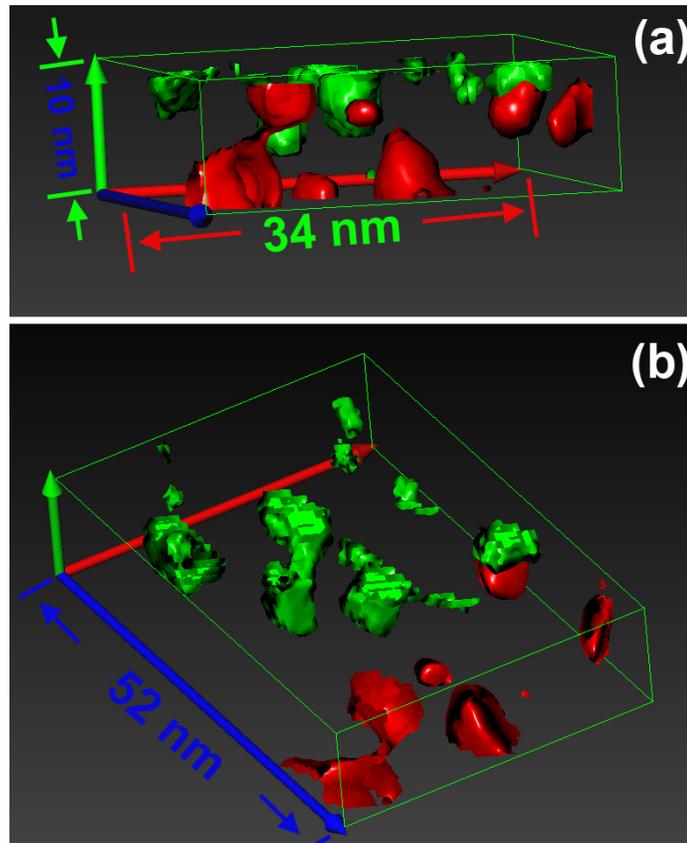


Figure 4.14: A 3D visualization of the combined reconstruction visualized by isosurface rendering. The images show the distribution of silicon (green) and erbium (red). One erbium cluster is located at the boundary of a Si-NC.

a Si-NC is shown in Fig 4.14(b). Other erbium clusters are separated from Si-NCs. Thus, the experimental result show that Si clusters are distributed outside of the Si-NCs.

Artifacts in the tomographic reconstruction imaged in Fig 4.14 can arise from:

- (1) For EF-STEM reconstruction, delocalization of inelastic electron scattering (plasmon energy loss), as discussed in the preceding section, the missing cone of tilt data, and the fairly large tilt increment of 4° can lead to an elongation effect and decreased resolution.
- (2) For the HAADF reconstruction, artifacts can arise from contamination by gallium that would also appear bright in HAADF images, the missing cone of tilt data, and the large tilt increment. The Ga contamination should be limited to the surface of

the sample only, however.

The proposed method of combining the HAADF and EF-STEM in STEM mode simultaneously provides a way to explore the spatial relationships between silicon nanocrystals and Er clusters in an amorphous SiO₂ matrix. The 3D tomographic result of Si-NCs shows that Si-NCs do not have simple spherical symmetry and the size of Si-NCs is not uniform. Instead, Si-NCs with irregular morphology were observed and the connection of Si-NCs by Si-rich areas was revealed. Clustering of Er was also observed in the tomographic result. The Er clusters do not appear inside Si-NCs, but rather appear to “decorate” the boundaries. The experimental result shows that the proposed method has the potential to solve the full 3D nanostructure of solid nanocomposites in which the nanocluster distribution, shape, and spacing determines the collective ensemble behavior.

4.5 Summary

This chapter presented the experimental results and discussion of electron tomography on latex nanoparticles, Au nanoparticle multilayers in amorphous SiO₂, Si nanoparticles in SiO₂, and Er doped Si nanoparticles in SiO₂ matrix. Methods to determine the threshold for the visualization of the reconstruction were examined. A method to obtain 3D information of two kinds of materials in a sample was reported. To summarize the results:

- (1) The shape and spatial morphology of latex nanoparticles prepared by the grid support method was obtained by HAADF electron beam tomography. The results showed that the grid support method is an efficient way to prepare a sample for 3D tomography, especially for cases in which the material is available in solution to begin with. The missing wedge problem can be minimized or ignored in most of cases in which the tilt range reaches $\pm 75^\circ$. The image contrast arising from the support film may result in blurring [7].
- (2) Tomographic imaging of Au nanoparticle multilayers confirms that one cannot set a threshold to match a known volume fraction. The visualization obtained via the QI method is preferable when quantitative interpretation is intended [51].
- (3) The shape and spatial morphology of Si nanoparticles in the SiO_{0.6} thin film sample was determined by energy-filtered electron tomography. The Si-NCs do not have a simple spherical symmetry. Instead, irregular morphologies were observed.
- (4) A new electron tomography method combining HAADF and EFTEM techniques simultaneously in STEM mode was applied to investigate a sample consisting of an

Er-doped silicon nanocrystal composite. The results revealed the 3D distribution between the Si-NCs and the Er clusters, and shape of each. The size of the Si-NCs and Er clusters is not uniform. The Si-NCs do not have a spherical symmetry. Er clusters do not appear inside Er-NCs, but rather decorate the Si-NCs.

Bibliography

- [1] Midgley, P. A., and Dunin-Borkowski, R. E.. *Nature Mater.*, 8:271-280 (2009)
- [2] Kawase, N., Kato, M., Nishioka, H., and Jinnai, H.. *Ultramicroscopy*, 107:8-15 (2007)
- [3] Midgley, P. A., and M. Weyland. *Ultramicroscopy*, 96:413 (2003)
- [4] Baumeister, W., Grimm, R., and Walz, J.. *Trends Cell Biol.*, 9:81 (1999)
- [5] Koster, A. J., Grimm, R., Typke, D., Hegerl, R., Stoschek, A., Walz, J., and Baumeister, W.. *J. Struct. Biol.*, 120:276 (1997)
- [6] Kawas, N., Kato, M., Nishioka, H., Jinnai, H.. *Ultramicroscopy*, 107:8-15 (2007)
- [7] Comolli, L. R., and Downing, K. H.. *J. Struct. Biol.*, 152:149 (2005)
- [8] Rizza, G., Cheverry, H., Gacoin, T., Lamasson, A., and Henry, S.. *J. Appl. Phys.*, 101:014321 (2007)
- [9] Cao, M., Zhang, H. B., Lu, Y., Nishi, R., and Takaoka, A.. *J. Microsc.*, 239:66 (2010)
- [10] Pavesi, L., Negro, L. D., Mazzoleni, C., Fronzo, G., and Priolo, F.. *Nature*, 408:441 (2000)
- [11] Canham, L.. *Nature*, 408:441 (2000)
- [12] Polman, A.. *Nat. Mater.*, 1:10 (2002)
- [13] Fiori, A. T., and Ravindra, N. M.. *J. Electron. Mater.*, 32:1043 (2003)
- [14] Rong, H., Liu, A., Jones, R., Cohen, O., Hak, D., Nicolaescu, R., Fang, A., and Paniccia, M.. *Nature*, 433:209 (2005)
- [15] Han, H. S., Seo, S. Y., and Shin, J. H.. *Appl. Phys. Lett.*, 79:4568 (2001)

- [16] Tiwari, S., Rana, F., Hanafi, H., Hartstein, A., and Crabbe, E. F.. *Appl. Phys. Lett.*, 68:1377 (1996)
- [17] Beard, M. C., Knutsen, K. P., Yu, P., Luther, J. M., Song, Q., Metzger, W. K., Ellingson, R. J., and Nozik, A. J.. *Nano Letters*, 7:2506 (2007)
- [18] Timmerman, D., Izeddin, I., Stallinga, P., Yassievich, I. N., and Gregorkiewicz, T. *Nature Photonics*, 2:105 (2008)
- [19] Iacona, F., Franzo, G., and Spinella, C.. *J. Appl. Phys.*, 87:1295 (2000)
- [20] Alivisatos, A. P.. *Science* 271:933 (1996)
- [21] Pavesi, L., Gaponenko, S., and Dal Negro, L. *Towards the First Silicon Laser* Springer (2003)
- [22] Wang, J., Wang, X. F., Li, Q., Hryciw, A., and Meldrum, A.. *Philos. Mag.*, 87:11 (2007)
- [23] S. Schamm, C. Bonafos, H. Coffin, N. Cherkashin, M. Carrada, G. Ben Assayag, A. Claverie, M. Tence, C. Colliex. *Ultramicroscopy*, 108:346-357 (2008)
- [24] X. W. Du, M. Takeguchi, T. Tanaka, and K. Furuya. *Appl. Phys. Lett.*, 82:1108 (2003)
- [25] D. A. Muller and J. Silcox. *Philos. Mag. A*, 71:1375 (1995)
- [26] G. S. Chen, C. B. Boothroyd and C. J. Humphreys. *Philos. Mag. A*, 78:491 (1998)
- [27] Titova, L. V., Cocker, T. L., Cooke, D. G., Wang, X., Meldrum, A., and Hegmann, F.. *Phys. Rev. B* 83: 085403 (2011)
- [28] Frank J.. *Electron Tomography: Three-dimensional imaging with the Transmission Electron Microscope*. New York: Plenum Press (1992)
- [29] Egerton, R. F.. *Electron Energy-Loss Spectroscopy in the Electron Microscope*, 3rd edition. Springer Science+Business Media (2011)
- [30] Egerton, R. F. *Micron*, 34:127-139 (2003)
- [31] Egerton, R. F. *Ultramicroscopy*, 107:575-586 (2007)
- [32] Lee, H., Shin, J. H., and Park, N.. *Optics Express*, 13:9881 (2005)
- [33] Pavesi, L. *J. Phys.: Cond. Matter.*, 15:1169 (2003)

- [34] Hryciw, A. C.. “Optical properties of rare-earth doped silicon nanocomposites”, Ph.D dissertation, University of Alberta, Edmonton, Alberta, Canada, 2007
- [35] Desurvire, E., Simpson, J. R., and Becker, P. C.. *Optics Letters* 12:888 (1987)
- [36] Fujii, M., Yoshida, M., Kanzawa, Y., Hayashi, S., and Yamamoto, K.. *Appl. Phys. Lett.* 71:1198 (1997)
- [37] Kik, P. G., Brongersma, M. L., and Polman, A.. *Appl. Phys. Lett.* 76:2325 (2000)
- [38] Kenyon, A. J., Chryssou, C. E., Pitt, C. W., Shimizu-Iwayama, T., Hole, D. E., Sharma, N., and Humphreys, C. J.. *J. Appl. Phys.* 91:367-374 (2002)
- [39] Walters, R. J., Bounanoff, G. I., and Ateater, H. A.. *Nat. Mater.* 4:143 (2005)
- [40] Liang, D., and Bowers J. E.. *Nat. Photonics* 4:511-517 (2010)
- [41] Pavesi, L., and Turan, R.. Silicon Nanocrystals. Wiley-VCH (2010)
- [42] Chryssou, C. E., Kenyon, A. J. and Pitt, C. W.. *Mater. Sci. Eng. B* 81:16 (2001)
- [43] Jhe, J. H., Shin, J. H., Kim, K. J., and Moon, D. W.. *Appl. Phys. Lett* 82:4489 (2003)
- [44] Kimura, T., Isshiki, H., Ide, S., Shimizu, T., Ishida, T., and Saito, R.. *J. Appl. Phys* 93:2595 (2003)
- [45] Gourbilleau, F., Dufour, C., Madelon, R., and Rizk, R.. *J. Lumines.*
- [46] Garrido, B., Garcia, C., Pellegrino, P., Navarro-Urrios, D., Daldosso, N., Pavesi, L., Gourbilleau, F., and Rizk, R.. *Appl. Phys. Lett* 89:163103 (2006)
- [47] Choy, K., Lenz, F., Liang, X. X., Marsiglio, F., Meldrum, A. *Appl. Phys. Lett.* 26:261109 (2009)
- [48] Lenz, F., DeCorby, R., Meldrum, A.. *Phys. Status Solidi A*, 206:989 (2009)
- [49] Pellegrino, P., Garrido, B., Arbiol, J., Garcia, C., Lebour, Y., and Morante, J. R.. *Appl. Phys. Lett.* 88:121915 (2006)
- [50] Yurtsever, A., Weyland, M and Muller, D. A. *Appl. Phys. Lett.*, 89:151920 (2006)
- [51] Wang, X. Y., Lockwood, R., Malac, M., Furukawa, H., Li, P., and Meldrum, A.. *Ultramicroscopy*, 113:96-105 (2012)

Chapter 5

Summary and Future Directions

This thesis focused on electron tomography in transmission electron microscopy (TEM). Electron tomography reconstruction algorithms, sample preparation techniques, and the practical applications of electron tomography in materials research were investigated. This chapter will summarize the main results and discuss possible future directions.

Over the past few years, electron tomography has experienced rapid development, as a result of the increased computational speed of modern computers and the ability to control and automate data acquisition in modern electron microscopes. The algorithms for electron tomography reconstruction are most commonly based on either filtered back-projection (FBP) or the simultaneous iterative reconstruction techniques (SIRT).

The two main reconstruction methods were discussed in detail and compared in Chapter 2. Some of the important considerations for generating and visualizing tomographic data were highlighted. Imaging of the reconstruction is an important aspect that can strongly influence our interpretation of the data. Changing the thresholding can change the size, shape, and apparent connectivity of the particles. One cannot set a threshold to match a known volume fraction; this produces poor results due primarily to the unavoidable counting of high-intensity-value “noise pixels” as belonging to the nanoparticle volume. Instead, a quantitative comparison with projected images produces a much more realistic 3D picture of the sample. Both volume rendering and isosurface rendering can be used in the visualization, but volume rendering avoids issues associated with non-closing surfaces and is therefore the preferred method. Both FBP and SIRT reconstruction methods were tested, both for simulated and experimental data, and in general, superior results were obtained using SIRT.

Several additional issues must be considered in the visualization. First, certain common types of filtering and image processing before reconstruction can lead to artifacts such as a high intensity at the interfaces, which can give the nanoparticles a shell-like appearance for

both FBP and SIRT methods. Second, for a tilt interval of 2° and a full $\pm 90^\circ$ tilt range, particles spaced closer than about 2.2 times their radius will erroneously appear to be a single, elongated particle, even under optimal reconstruction and visualization methods for FBP reconstructions.

Due to the data acquisition methods, there are some extra requirements for electron tomography sample preparation as compared to that for the conventional TEM. In Chapter 3, a new method for the fabrication of samples for electron tomography was reported. The sample preparation is less time-consuming when compared to the “standard” rod-shaped sample preparation method. The new method involves making FIB-cut rectangular holes that can be fabricated without the need for precise alignment. Although this method cannot be used to cover the full $\pm 90^\circ$ tilt range, it can reach a $\pm 75^\circ$ tilt range, which is sufficient to minimize missing wedge problems. The requirement for a $\pm 75^\circ$ tilt range is that the side of the cut rectangle perpendicular to the bar is at least six times the length of the thickness of the thin film—a requirement which is easily satisfied. The reported method simplifies the fabrication in FIB by avoiding the lift-out process. The method can also reduce Ga^+ contamination, as compared to the rod-shaped sample preparation method, since only two sides of a thin film are FIB-milled as opposed to extensive milling of the entire needle length in the standard rod-shaped sample preparation method. The method can be also used for samples dissolved in solution dropped onto thin substrates. The fact that there is a large area in addition to the narrow bridge for tomography or 2D projection imaging provides ample area for additional analysis in a TEM.

An electron tomography method was introduced in Chapter 4, in which high angle annular dark-field (HAADF) and energy-filtered STEM (EF-STEM) imaging was combined simultaneously in the STEM mode. This provides information on two different materials without spatial shift error, while minimizing radiation damage issues. Although the reconstruction using this parallel method for an Er-doped Si nanocomposite was successfully presented, reaching the desired resolution was difficult due to the inherent resolution of EF-STEM at low energy loss, the sample preparation process, radiation damage, and the missing wedge effect. In future work, two experimental aspects should improve the resolution of the combined reconstruction. One is to obtain a thinner sample to further limit the thickness along the projected direction and decrease the Ga^+ ion contamination in FIB system. The second is to achieve a $\pm 75^\circ$ tilt range, if radiation damage can be minimized.

In addition to the improvement of the data acquisition experiment, the resolution of the reconstruction may be perhaps further improved by adopting recently proposed reconstruction algorithms such as DART (Discrete Algebraic Reconstruction Technique) [1].

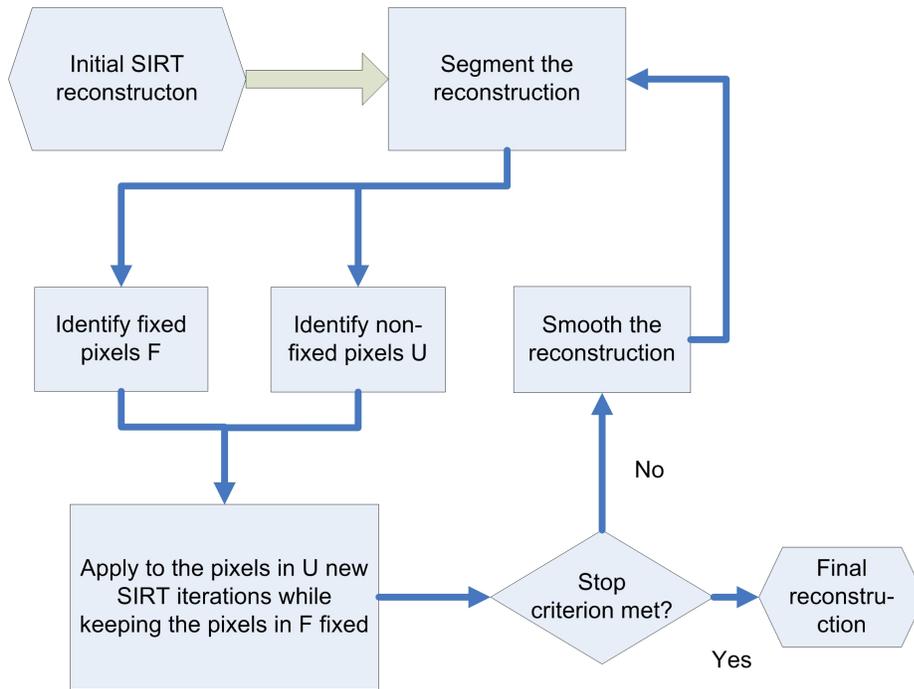


Figure 5.1: Flowchart of the DART algorithm. Revised from [1].

DART is a reconstruction method based on prior knowledge. Prior knowledge in this case refers to an assumption that the unknown 3D object consists of a small number (*i.e.*, two to five) of different materials, each with a constant grey level (intensity) in the tomographic reconstruction. Such prior knowledge is available in some special cases. For example, for the Si-NCs embedded in silica glass sample, the intensity of Si-NCs should be constant in the reconstruction while the intensity of the region outside of Si-NCs should be a constant and different from Si-NCs. For the Er-doped in Si-NCs system, the intensity of Si-NCs, Er clusters, and background in the reconstruction should be constant and different from one another.

A full description of the algorithmic details of the DART method can be found in Ref [1]. Figure 5.1 shows a flowchart for the method. The initial reconstruction is calculated using SIRT or ART. Then DART is introduced in the iterative loop. In each iteration, five steps are executed as follows: (1) Segmentation: the intensity of the reconstruction is “discretized” based on the prior knowledge. For example, for the Si-NCs embedded in glass sample, the intensity of Si-NCs in the reconstruction is set to 1 while the region outside of Si-NCs is set to 0. (2) Selection of non-fixed pixels: the boundary of objects

after the segmentation is commonly referred to as non-fixed pixels and they need to be further evaluated; others pixels are fixed. (3) A new reconstruction is calculated using SIRT method. During the process of computation, the intensity of the fixed pixels is not changed, only the non-fixed pixels are adjusted. (4) Smoothing operation: a Gaussian smoothing filter with 1 pixel radius is commonly applied to the non-fixed boundary pixels of the new reconstruction to obtain a smoothed boundary. (5) The iteration is terminated based on a suitable termination criterion. For example, after the difference value between the object and the reconstruction is sufficiently small or after a fixed number of iterations is performed.

From the above description, one can see that DART is a heuristic algorithm [1]. This means the convergence of the algorithm is not guaranteed. Obviously the convergence of the solution depends on the selection of parameters in the iteration, such as the parameter used for the segmentation step and determination of the set of non-fixed pixels. By optimizing parameters, a satisfactory reconstruction for the Er-doped Si-NCs system should be obtainable with the DART method. Therefore, application of the DART method to the Er-doped in Si-NCs system is one of future research directions.

The term “resolution” is often used to ascertain the quality of the tomographic result, but the definition of the resolution is not always clear and deserves a brief discussion [2]. According to the definition in Ref [3], resolution is a boundary in reciprocal space defining the 3D domain within which Fourier components contribute significantly to the density map. Therefore resolution is a quantity in Fourier space and hence has dimension 1/length. Some authors also refer to the real-space quantity 1/resolution as “resolution” [3, 4]. To avoid this confusion, 1/resolution is termed a “resolution distance”, following the terminology in Ref [3].

The theoretical resolution distance of a reconstruction is given by the Equation 2.1 for a single-tile axis geometry [4]. However, there are many effects that prevent reaching the theoretical resolution. Therefore, methods of practical resolution estimation were developed based on Fourier shell correlation (FSC) [5, 6] or 3D spectral signal-to-noise ratio (SSNR) [7, 8, 9, 10, 11]. There is uncertainty because the value of the estimated resolution strongly depends on the criterion adopted in these methods [12]. Therefore, the corresponding resolution distance in electron tomography can not be treated in the same way as, for example, the optical resolution in light optics (*e. g.*, the Rayleigh criterion) or information limit in 2D electron microscopy imaging.

Although the numerical estimates of the resolution can be obtained as discussed above,

the ultimate assessment of the claimed resolution is always done by examining the appearance of the reconstructed volume [12]. For example, resolution can be assessed by examining the resolvability of known features of the reconstructed object, such as the sub-structure information [13] and the crystal lattice structure [14]. Although this approach might not be adequate to evaluate the “resolution” of the tomographic reconstruction with a desirable degree of accuracy, the general validity of the results within the claimed resolution limit can be appraised.

The development of improved tomographic techniques is ongoing. To improve the resolution of the reconstruction in all three dimensions in the reconstructed volume would be desirable. Therefore, the factors limiting the resolution of the reconstruction of electron tomography were reviewed by considering the entire electron tomography process from sample preparation to data visualization.

- (1) The sample preparation method. As discussed in Chapter 3, the projected thickness is the main limiting factor that affects the contrast of the recorded image. If the FIB technique is used to prepare the sample, Ga ion damage and contamination may be the limiting factor, especially for the HAADF imaging mode.
- (2) The recording of the projected images. As discussed earlier, the missing wedge is an important limiting factor. In addition, uncertainties in the tilt value measurements and stage positioning errors may affect the quality of the reconstruction, since they affect the accuracy of the alignment. Electron radiation damage is also an important limiting factor for some materials. In biological electron tomography, a low dose cryo-ET is usually used to minimize the radiation damage. The pixel size of the detector also needs to be considered, because if it is too large, it will cause under-sampling as discussed in Chapter 2.
- (3) The reconstruction of the tomographic volume. As discussed in Chapter 2, current reconstruction methods need to be developed further for samples containing many different phases. The alignment of the projected images plays a critical role in determining the outcome of the reconstruction process.
- (4) The visualization of the reconstructed volume. Strictly speaking, the visualization process does not improve the resolution of the reconstruction. However, as discussed in Chapter 2, determining the threshold for the volume visualization is critical for interpreting the result.

Many future developments in electron tomography to improve the resolution of the reconstruction will involve automation. Automated electron tomography is a potential

avenue towards increasing the reconstruction resolution at high magnification. Future work is ongoing to develop the following methods related to automated electron tomography:

- (1) Automate image recording. The purpose of automated imaging recording is to minimize the recording time. Therefore, the accuracy of the predictive strategy on the image movement during the tilt recording process is important. A potential method may make use of *a priori* knowledge about the image movement to increase the accuracy.
- (2) Automate alignment for the reconstruction. There are at least two possible ways to improve the accuracy of automatic alignment for the high magnification situation. One is to further develop the algorithm of the cross-correlation function that is used to align images. The second is to apply the predictive strategy to increase the accuracy by obtaining the image movement information from the automatic image recording process.
- (3) Automate quality index (QI) thresholding. Obtaining the best QI value within an acceptable time is important. Using *a priori* knowledge of the sample to restrict the range of thresholds could speed the method considerably.

Bibliography

- [1] Batenburg, K. J., and Sijbers, J.. *IEEE Trans. Image Process.* 20(9):2542 (2011)
- [2] Banhart, J.. “Advanced Tomographic Methods in Materials Research and Engineering. Oxford University Press, USA (2006)
- [3] Banhart, J.. “Three-Dimensional Electron Microscopy of Macromolecular Assemblies. Oxford University Press, USA (2006)
- [4] Crowther, R. A., DeRosier, D. J., and Klug, A.. *Proc. R. Soc. Lond.* 317:319-340 (1970)
- [5] Grigorieff, N.. *Acta Crystallogr. D Biol. Crystallogr* 56:1270-1277 (2000)
- [6] Stewart, A., and Grigorieff, N. *Ultramicroscopy* 102:67-84 (2004)
- [7] Penczek, P. A.. *J. Struct. Biol.* 138:34-46 (2002)
- [8] Penczek, P. A., Renka, R., and Schomberg, H.. *J. Opt. Soc. Am. A* 21:499-509 (2004)
- [9] Unser, M., Sorzano, C. O., Thevenaz, P., Jonic, S., El-Bez, C., De carlo, S., Conway, J. F., and Trus, B. L.. *J. Struct. Biol.* 149:243-255 (2005)
- [10] Cardone, G., Grunewald, K., and Steven, A. C.. *J. Struct. Biol.* 151:117-129 (2005)
- [11] Banhart, J.. “Electron Tomography Methods for Three Dimensional Visualization of Structures in the Cell. 2nd Edition. Springer Science+Business Media, LLC (2006)
- [12] Penczek, P. A.. Chapter 3 in *Methods in Enzymology* Volume 481. Cryo-EM, Part B 3D Reconstruction (2010)
- [13] Baker, M. L., Ju, T., and Chiu, W.. *Structure* 15:7-19 (2007)
- [14] Scott, M. C., Chen, C. C., Mecklenburg, M, Zhu, C, Xu, R., Ercius, P., Dahmen, U., Regan, B. C., and Miao, J.. *Nature* 483:444-447

List of Publications

Journal Papers

1. Wang, X., Lockwood, R., Vick, D., Meldrum, A., and Malac, M.. “A convenient method for electron tomography sample preparation using a focused ion beam”. *Microw. Res. Tech.* DOI: 10.1002/jemt.22044 (2012)
2. Wang, X., Lockwood, R., Malac, M., Li, P., and Meldrum, A.. “Reconstruction and visualization of nanoparticle composites by transmission electron tomography”. *Ultramicroscopy* 113: 96 (2012)
3. Wang, X., Li, P., Malac, M., Lockwood, R., and Meldrum, A.. “The spatial distribution of silicon NCs and erbium ion clusters by simultaneous high-resolution energy filtered and Z-contrast STEM and transmission electron tomography”. *Phys. Status Solidi C* 208: 1038 (2011)
4. Bianucci, P., Wang, X., Veinot, J., and Meldrum, A.. “Silicon nanocrystals on bottle resonators: Mode structure, loss mechanisms and emission dynamics”. *Opt. Express* 18: 8466 (2010)
5. Titova, L. V., Cocker, T. L., Cooke, D. G., Wang, X., Meldrum, A., and Hegmann, F.. “Ultrafast percolative transport dynamics in silicon nanocrystal films”. *Phys. Rev. B* 83: 085403 (2011)
6. Lockwood, R., McFarlane, S., Rodriguez-Nunez, J. R., Wang, X., Veinot, J. G. C., and Meldrum, A.. “Photoactivation of silicon quantum dots”. *J. Lumin* 131: 1530 (2011)

Conference Abstracts

1. Wang, X., Lockwood, R., Vick, D., Meldrum, A., and Malac, M.. “A new method to fabricate 3D electron tomography sample using FIB technique”. *Microsc Microanal* 17 (S2) 690 (2011)
2. Wang, X., Malac, M., Meldrum, A., Lockwood, R., Li, P., and Furukawa, H.. “Limits of tomographic reconstruction of discrete nanoparticles”. *Microsc Microanal* 16(S2) 1868 (2010)
3. Li, P., Wang, X., Malac, M., Egerton, R., Meldrum, A., Liang, X., Lenz, F., and Wang, J.. “3D imaging of Si and Er nanoclusters in Er doped SiO_{1.5} films by STEM tomography”. *Microsc Microanal* 15(S2) 1868 (2009)

Appendix A

Reading and Writing the Reconstruction File Generated by TEMography Software

The following matlab code is used to read the binary file generated by the TEMography software.

```
1 % Reading code: Read binary file generated by the TEMography software
2 fid=fopen('ReconstructionFile.bin','r');
3 C=fread(fid);
4
5 size_x=C(1)+C(2)*256;
6 size_y=C(3)+C(4)*256;
7 size_z=C(5)+C(6)*256;
8
9 % Create a Matrix A with the same size as x\ast y\ast z
10 a=size_x*size_y*size_z;
11 A=zeros(a,1);
12 k=1:a;
13 A(k,1) = C(2*k+5) + C(2*k+7)*256;
14
15 % Store the data cubic using matrix M
16 l = 1:size_z;
17 n = 1:size_y;
18 m = 1:size_x;
19 M(m,n,l) = A;
```

```

20
21
22 % Normalization by itself
23 M=abs (max (max (max (M) ) )-M) /max (max (max (M) ) ) ;
24
25
26 save 'Readfile.mat' M;

```

The following matlab code is used to write a data-cube into the binary file recognized by Visualizer which is a part of TEMography software package.

```

1 % The safe way to visualize the imported data is to substitute the data ...
   part of the binary file using the imported data to avoid the coding ...
   problem of generating the binary file for the software.
2 fid=fopen('Original.bin');
3 C=fread(fid);
4 fy=fopen('GeneratedFile.bin','w');
5
6
7 size_x=C(1)+C(2)*256;
8 size_y=C(3)+C(4)*256;
9 size_z=C(5)+C(6)*256;
10
11 D1=load('ImportedData.mat');
12 D1=double(D1);
13 D_size=size(D1);
14 zhuan=zeros(D_size(1),D_size(2),D_size(3));
15
16 for z1=1:D_size(1)
17     for z2=1:D_size(2)
18         for z3=1:D_size(3)
19             z4=D_size(1)+1-z3;
20             z5=D_size(1)+1-z2;
21             zhuan(z1,z2,z3)= D1(z5,z4,z1);
22         end
23     end
24 end
25
26 D=zeros(D_size(1)*D_size(2)*D_size(3),1);
27 b=1;
28 for x1=1:D_size(1)
29     for x2=1:D_size(2)
30         for x3=1:D_size(3)

```

```

31         D(b,1)=zhuan(x3,x2,x1);
32         b=b+1;
33     end
34 end
35 end
36
37 clear zhuan
38
39
40 D=D*65535;
41
42 a=size(D);
43
44 maxD=max(max(max(D)));
45 KK=maxD/256;
46
47 if KK ≤ 256
48
49     for i=1:a
50         a1=2*i+5;
51         D(i)=round(D(i));
52         m1=fix(D(i)/256);
53         m2=mod(D(i),256);
54         C(a1)=m2;
55         C(a1+1)=m1;
56     end
57 elseif KK>256
58     disp('WRONG');
59 end
60
61 fwrite(fy,C);

```

Appendix B

Calculation of RMS Difference

The following matlab code is used to calculate the RMS difference between the reconstruction and the original object with considering the position shift between them.

```
1 % Calculation of RMS difference
2 x_shift=a;
3 y_shift=b;
4 z_shift=c; % a,b,c are the shift pixels between the reconstruction and ...
              the original object. They could be measured from the Visualizer by ...
              visualizing two data-cubes at the same time.
5
6 size_RMS=(600-x_shift)*(600-y_shift)*(600-z_shift);
7 RMS=zeros(size_RMS,1);
8
9 Rec=load('Reconstructeddata.mat');
10 RecI=Rec(1:(600-x_shift),1:(600-y_shift),1:(600-z_shift));
11
12 zhuan=load('OriginalData.mat');
13 GenI=zhuan((1+x_shift):600,(1+y_shift):600,(1+z_shift):600);
14 clear Rec;
15 clear zhuan;
16
17 M1=(abs(GenI.^2-RecI.^2)).^(1/2);
18 b=1;
19 for k=1:(600-z_shift)
20     for l=1:(600-y_shift)
21         for m=1:(600-x_shift)
22             RMS(b)=M1(m,l,k);
23             b=b+1;
```

```
24         end
25     end
26 end
27
28
29     save 'RMSvalue.mat' RMS;
30 end
31
32 % NOTE: this code strongly depends on the storage way of the generated ...
    date. This code is valid for the generated way cited in this thesis.
```