TUNGSTEN TIP AND ATOMIC SITE TOMOGRAPHY

by

Ho Yee Timothy Poon

A thesis submitted to the University of Birmingham for the degree of

DOCTOR OF PHILOSOPHY

Nanoscale Physics Research Laboratory

School of Physics and Astronomy

College of Engineering and Physical Sciences

University of Birmingham

May 2019

Abstract

Atomic electron tomography aims to precisely locate individual atoms of a nanoparticle in three-dimensional space. In this work, a tomography method based on tungsten tips is developed to allow images to be taken over a full angular range by placing a nanoparticle on the apex of an etched tungsten tip. There is no interference of signal from supporting materials with the suspended nanoparticle. A new reconstruction algorithm, atomic site tomography, is developed using the principle of regularisation in multiple linear regression. This algorithm is specifically designed for identifying the precise locations of individual atoms in three-dimensional space, and the algorithm is validated by an experimental dataset. A gold nanoparticle dataset is successfully obtained by tungsten tip tomography, and the dataset is processed to remove scanning artefacts. Selected region of the gold nanoparticle dataset is used to demonstrate the new reconstruction algorithm and the whole gold nanoparticle is then reconstructed. A tuning fork atomic force microscope is developed to provide a more flexible method to prepare samples for tungsten tip tomography and its progress is reported. This work contributes to the field of atomic electron tomography by improving the experimental techniques for acquiring high-quality tomography dataset and proposing a new reconstruction algorithm which aims at locating individual atoms of nanoparticles precisely.

To my family and my lovely wife Vianni

for their love, patience and support

Acknowledgements

I would like to thank my supervisor, Dr Wolfgang Theis, for giving me the opportunity to undertake this project. His enthusiasm for science and tireless work ethic are always an inspiration. His insights and guidance have had a huge impact on the project and it is difficult not to feel scientifically illiterate after taking his advice. I have learnt so much from him and I am grateful to him for always standing patiently behind my desk and seeing me slowly typing 'ax.plot(…)'.

I would also like to thank everyone at NPRL for all the help during my study and their friendships: particularly, Ale and Alex for keeping me awake during some endless nights in weekends; Shane, for taking me in during my difficult time and introducing the UHV-STM world to me. I thank NCEM for giving access to the TEAM 0.5 microscope, and in particular Peter and Colin for making my short stay in Berkeley both fruitful and wonderful. I also thank the university for giving me the U21 PhD Scholarships for a one-month academic visit to Ohio State University.

Immense thanks go to my family for their support and understanding in pursuing my PhD. Thanks Vianni, for putting up with me during my study and still hasn't abandoned me. Finally, without God I could do nothing, and with confidence and hope I trust God will guide me along the best pathway for my life.

Contents

List of figures

List of tables

List of abbreviations

ADC analog-to-digital converter

AET atomic electron tomography

AFM atomic force microscope

ART algebraic reconstruction technique

AST atomic site tomography

bcc body-centred cubic

CBED convergent-beam electron diffraction

CT computed tomography

DAC digital-to-analog converter

DSP digital signal processor

EST equally sloped tomography

ET electron tomography

fcc face-centred cubic

FFT fast Fourier transform

FIB focused ion beam

GENFIRE generalised Fourier iterative reconstruction

GXSM Gnome X Scanning Microscopy

HAADF high-angle annular dark-field

HOPG highly oriented pyrolytic graphite

HV high vacuum

lasso least absolute shrinkage and selection operator

PPFFT pseudo-polar fast Fourier transform

SART simultaneous algebraic reconstruction technique

SIRT simultaneous iterative reconstructive technique

SPM scanning probe microscope

STEM scanning transmission electron microscopy

STM scanning tunnelling microscopy

TEAM Transmission Electron Aberration-corrected Microscope

TEM transmission electron microscopy

TIA TEM Imaging & Analysis

UHV ultra-high vacuum

1

Introduction

Tomos means 'slice, section' in Greek, and *tomography* is 'a technique for displaying a representation of a cross section through a human body or other solid object using X-rays or ultrasound' [1]. The motivation of tomography is to provide three-dimensional structural information, instead of two-dimensional projections from a three-dimensional object. This is vital in medical imaging so that the structure of the object of interest is not shadowed by surrounding tissues, and a more accurate diagnostics can be made. The concept of tomography was first suggested by the Polish radiologist Karol Mayer in 1914 [2], not long after the discovery of X-ray. Afterwards, numerous scientists and radiologists have developed and patented techniques about tomography, but it was not until 1972 that the first commercial X-ray computed tomography (CT) scanner was available [3]. In 1979, Allan M. Cormack and Sir Godfrey N. Hounsfield were awarded the Nobel Prize in Physiology or Medicine 'for the development of computer assisted tomography' [2, 4].

The same motivation of exploring materials in three dimensions existed in the physical science community. Although early tomography experiments were largely performed using X-ray, any penetrating waves or particles could be

used in a tomography experiment. The invention of electron microscopy in 1931 by Ernst Ruska and Max Knoll greatly enhanced our understanding of materials and the resolution first surpassed the wavelength of light in 1933 [5]. In 1968, the three-dimensional structure of the tail of bacteriophage *T*4 was resolved using transmission electron microscopy (TEM) [6] and since then, the field of electron tomography (ET) has been flourishing. High-resolution electron microscopy also enabled the retrieval of three-dimensional structural information such as interfaces [7], stacking sequence [8], stacking faults [9] and antiphase boundaries [10] by using intersecting atomic columns from a few projections — the idea of discrete tomography (see Section 1.3.1). With the development of aberration-corrected scanning transmission electron microscopy (STEM), atomic-scale resolution in ET becomes feasible and this is known as atomic electron tomography (AET). In 2012, the precise locations of some individual atoms in a gold nanoparticle were determined [11] and in 2015, thousands of individual atoms in a tungsten needle were identified without any prior knowledge of the structure [12].

This thesis aims to advance the field of AET. A new platform for AET, which forms a nanoparticle on the apex of an etched tungsten tip (*tungsten tip tomography*), is discussed in Chapter 2. Scanning artefacts such as random drift and scanning-coil distortion need to be minimised, and the projections need to be accurately aligned so that a reliable reconstruction is possible. These steps are discussed in Chapter 3. A new reconstruction algorithm, atomic site tomography (AST), is developed in this work. The principle of AST is demonstrated in Chapter 4 and it is validated by an experimental dataset in Chapter 5. A gold nanoparticle dataset obtained by tungsten tip tomography is also reconstructed by AST. To allow a more flexible way of placing a nanoparticle on the apex of an etched tungsten tip, an ambient table-top atomic force microscope (AFM) based on tuning forks is constructed and the progress is reported in Chapter 6.

In this chapter, the principles of electron microscopy are introduced in Section 1.1 as this is central to ET. The principles, techniques and reconstruction algorithms of conventional ET and AET are reviewed in Section 1.2 and Section 1.3 respectively. The major principle of AST is based on regularisation in multiple linear regression and it is introduced in Section 1.4. Lastly, the fundamental principle of scanning probe microscope (SPM) is introduced in Section 1.5.

1.1 Electron microscopy

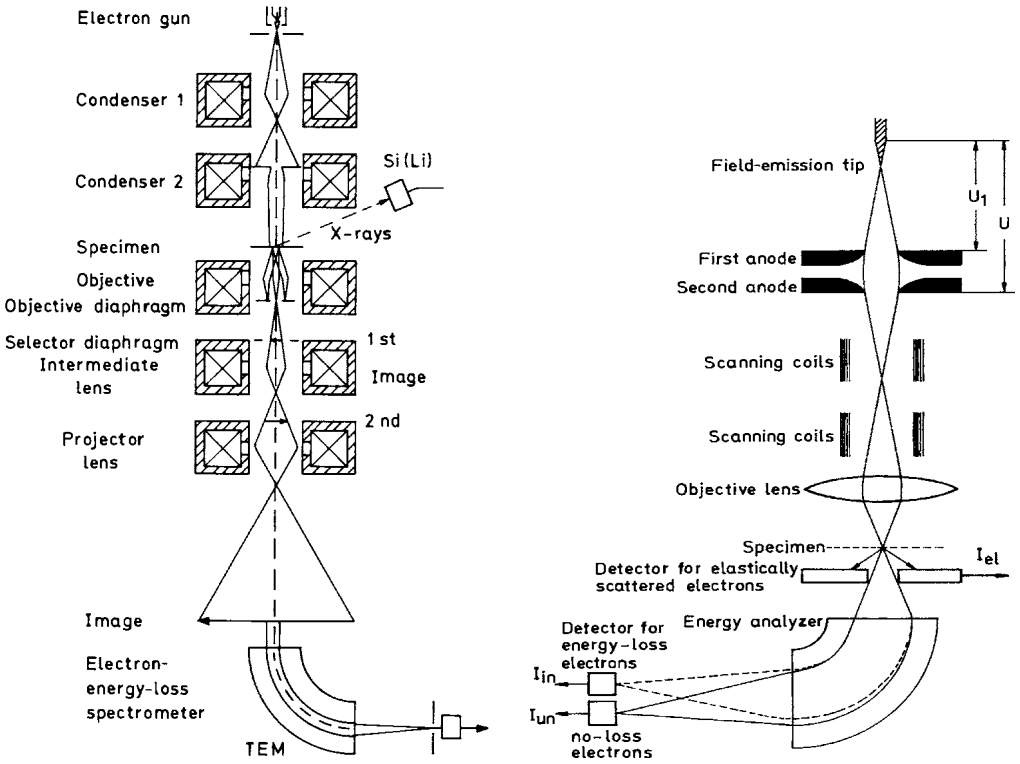
The French physicist Louis de Broglie proposed that a wavelength is associated with any particle (wave-particle duality), inspired by the work of Max Planck and Albert Einstein on light [13]. This was verified experimentally by the famous Davisson-Germer experiment [14]. The de Broglie wavelength is given by:

$$\lambda = \frac{h}{p} \tag{1.1}$$

h is the Planck constant (6.626×10^{-34} J s) and $p = mv$ is the momentum of the particle with m the rest mass of the particle and v its velocity. If the relativistic effect of electrons is considered, Equation (1.1) is modified as

$$\lambda = \frac{h}{mv}\sqrt{1 - \frac{v^2}{c^2}} \tag{1.2}$$

From Equation (1.2) the wavelengths of the electrons accelerated at 300 kV, 200 kV and 100 kV are 1.96 pm, 2.51 pm and 3.7 pm respectively. In electron microscopy, the accelerating voltages of electrons are typically 80 kV–500 kV and the wavelengths of electrons are much smaller than lattice spacings. In transmission electron microscopy (TEM), a thin specimen, typically of the order of 5 nm–100 nm in thickness, is irradiated by a parallel electron beam [15]. The



(a) TEM (b) STEM

Figure 1.1: The configurations of TEM and STEM with their major components are shown in (a) and (b) respectively. The schematic diagrams are adapted from [15].

electrons are scattered both elastically and inelastically, and they are collected behind the specimen to form an image (see Figure 1.1a).

In scanning transmission electron microscopy (STEM), the emitted electrons are focused by the electron lenses to form an electron probe in the sample plane. A raster scan is performed on the specimen by the action of scanning coils (see Figure 1.1b). The typical diameter of the electron probe is 0.2 nm–0.5 nm without aberration correction [15]. Different modes of imaging are possible in STEM and in electron tomography (ET), and it is important to satisfy the 'projection requirement' (see Section 1.2.3). High-angle annular dark-field (HAADF) imaging by STEM is preferable in ET as the intensity is approximately linear with the number of atoms, which relates to Z^{γ} (Z is the atomic number and γ is 1–2, and this is also known as Z-contrast imaging) [16–20].

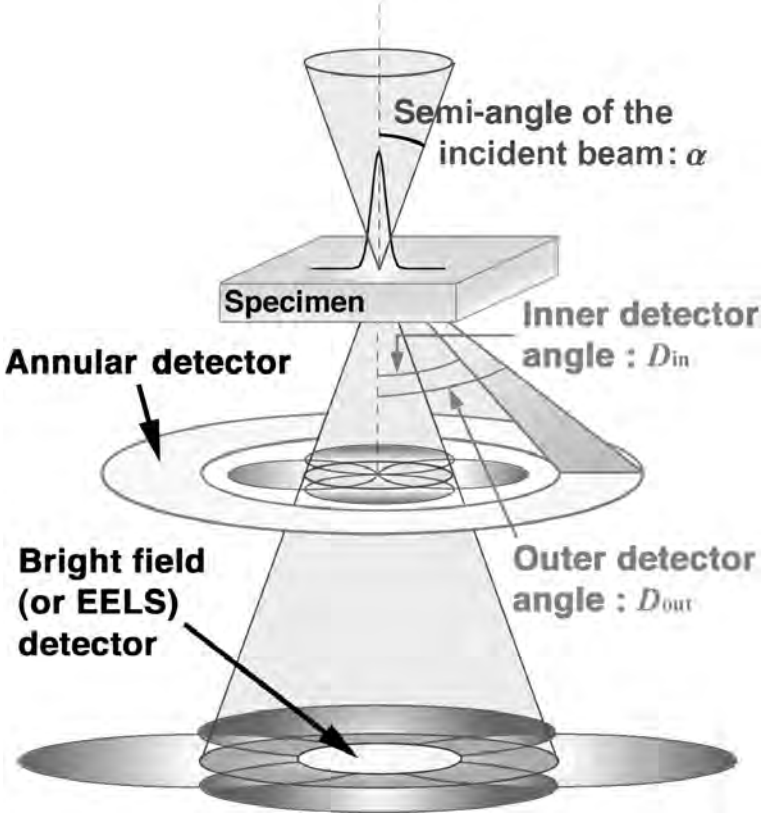


Figure 1.2: In HAADF-STEM, an electron probe is formed by electron lenses and a raster scan is performed on a specimen. An annular detector is used to collect electrons scattered to high angles, which are incoherent and quasi-elastic. The intensity is related to the atomic number Z when the inner detector angle is large enough that the intensity contribution from diffraction can be neglected. The schematic diagram is adapted from [24].

HAADF-STEM imaging collects mostly the incoherent electrons which scatter quasi-elastically to high angles (thermal diffuse scattering), where diffraction contrast can be neglected [21–23]. The configuration of HAADF-STEM imaging is shown in Figure 1.2. The typical collection angle for HAADF imaging in the annular detector is >80 mrad and the actual annular range depends on the microscope [25–28].

Although the wavelengths of electrons in electron microscopy are of the order of pm, the true resolution is limited by the size of electron probe. The major reason that limits the size of electron probe is the aberrations introduced by electron lenses [15]; this is the major limiting factor for achieving the theoret-

ical resolution. For high-resolution electron microscopy, the aberrations from electron lenses need to be corrected. Aberrations can be divided into two major types: geometric and chromatic aberrations [15]. Geometric aberrations include spherical aberration, astigmatism and coma. The optimum resolution is limited by spherical aberration, and it has the effect of reducing the focal length when the electrons are travelling at larger distances from the optic axis. This is equivalent to introducing an additional phase to the electrons. Chromatic aberration is the variation of focal length because of the variation of electron energy, which can be the result of fluctuation in the accelerating voltage or energy spread in the electron beam [15]. In STEM, these aberrations limit the minimum size of an electron probe in the sample plane as electrons with different energies cannot be focused on the same point; hence, the electron probe appears as a disk.

In 1936, the German physicist Otto Scherzer showed that it is theoretically impossible for the positive spherical aberration coefficient C_s to be zero in any system of rotationally symmetric electron lenses that have a static magnetic field [29]. In 1949, he outlined the principle to correct spherical aberration by reducing C_s [30]. The limit of resolution δ_{min} is related to spherical aberration by

$$\delta_{min} = 0.43 \left(C_s \lambda^3\right)^{\frac{1}{4}} \tag{1.3}$$

Scherzer proposed that by using multipole lenses where their magnetic field is not rotationally symmetric, the positive C_s from the condenser and objective lenses can be compensated by the negative C_s from the multipole lenses system [30]. The multipole system can be a quadrupole-octupole corrector or a hexapole corrector [31] and this was confirmed experimentally in 1977 [32]. In 1998, the first image from an aberration-corrected electron microscope was obtained with a resolution of 1.4 Å [33, 34]. As the value of C_s in Equation (1.3)

is reduced by aberration-corrected optics, atomic details of many materials were resolved and a comprehensive review can be found in [35]. Aberration-corrected electron microscopy must be used to acquire dataset for AET to ensure high-resolution images are taken.

1.2 Conventional electron tomography

Principles and techniques in conventional ET are relevant and applicable to atomic electron tomography (AET) and they will be discussed in this section. 'Conventional' here concerns mostly with biological samples with a thickness greater than 20 nm. The resolution can be about 2 nm but individual atoms cannot be precisely located. The primary mode of conventional ET is TEM instead of STEM in AET. Common techniques for specimen preparation in conventional ET are described Section 1.2.1. The techniques of data acquisition in conventional ET are described in Section 1.2.2 and traditional reconstruction algorithms are described in Section 1.2.3.

1.2.1 Specimen preparation

There are different considerations for preparing biological samples as it is their native states that are of interest, which are hydrated. Unlike metallic nanoparticles, biological samples are mostly proteins or organic molecules and they are weak phase objects. The way of preparing the specimen dictates the structural information that can be retrieved and how the reconstruction is to be interpreted. A detailed description of different specimen preparation techniques can be found in [36].

Biological samples are commonly stained by heavy metal ions to increase their contrast under TEM. The metal ions react with the background in negative staining, whereas they react with the actual biological sample in positive stain-

ing [37]. Although the contrast is improved, the projections are signals from metal ions instead of the biological sample itself and it impacts on the interpretation of the reconstruction. To preserve the structure, the biological sample is usually fixed chemically and dehydrated; however, chemical fixation is relatively slow and structural change is inevitable.

Currently the most popular way to preserve the structure is by rapid freezing, known as cryofixation. The sample is plunged into liquid coolants such as nitrogen, propane or ethane [38]. The freezing happens in less than a millisecond and the near-native state of the biological sample can be preserved. The vitrified water is usually replaced by an organic solvent with chemical fixatives to prevent the growth of non-vitrified ice crystal [39]. For conventional ET, the sample is then embedded in resin and it is ready to be put in an electron microscope.

1.2.2 Data acquisition

A tilt series consists of images taken at different tilt angles. This is achieved by supporting the sample on a TEM grid and the grid is rotated by a single- or dual-axis holder [40–43]. The tilt angular range is limited by the small gap between the two pole pieces and the mechanics of the holder, and this is often ±70°. The images acquired in high tilt angles often have a lower signal-to-noise ratio. This is caused by the increase of the effective thickness of the sample (see Figure 1.3a). It is also possible that the holder will shield the electron beam and this makes imaging at high tilt angles impossible.

The limitation of tilt angle in a tilt series results in a region where there is no information of the sample available, and in tomography this is known as the 'missing wedge' (see Figure 1.3b) [44]. The missing wedge is a region where the information of the sample is lost permanently. When the sample is rotating

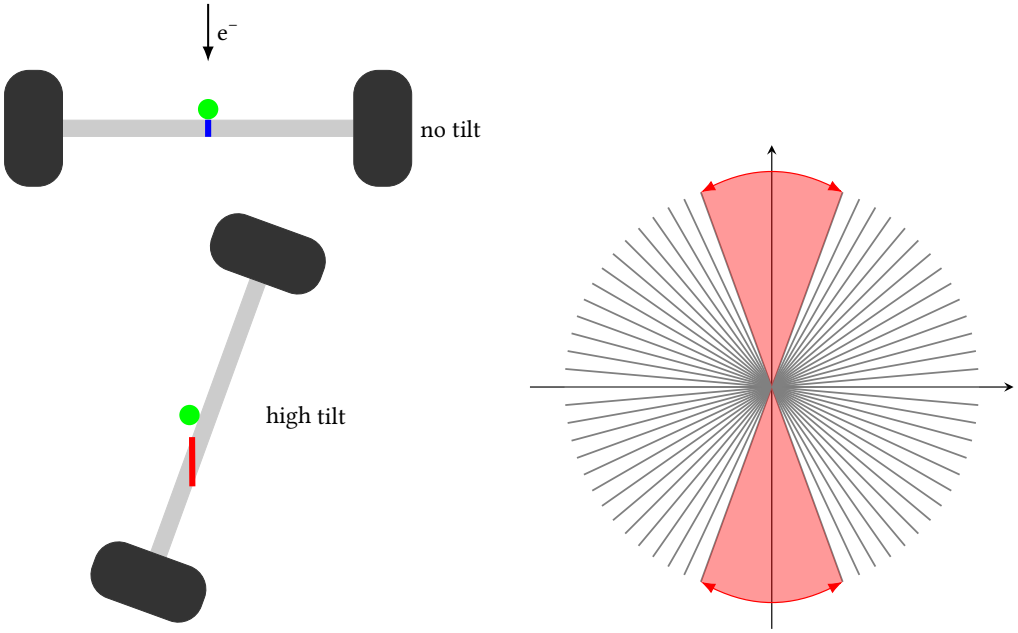


Figure 1.3: The limitations of acquiring images at high tilt angles are illustrated here. (a) when the sample (green) is not tilted or at a low tilt angle, the effective thickness (blue) is small; when it is at a high tilt angle, the effective thickness of the support (red) increases, giving a lower signal-to-noise ratio. (b) a 'missing wedge' in tomography (highlighted in red) is a region where the information is permanently lost due to the limitation of the tilt angular range, which could be caused by the limit of mechanics in sample rotation or shadowing of the sample by the grid or the holder.

around the tomography axis, it is important to ensure the images in the tilt series have a common point of rotation. In conventional ET, fiducial markers such as colloidal gold beads are used to provide a point of reference for the alignment of different images in a tilt series [44].

1.2.3 Reconstruction

To have a faithful reconstruction, the measured signals of every point in the projection need to be line integrals through the sample; in other words, the signals need to be linear with the mass-thickness of the sample. This is known as the projection requirement [45]. All reconstruction algorithms implicitly assume that the projection requirement is satisfied; however, this is usually not

the case experimentally. In physical science, most of the samples are crystalline and the atoms are of high atomic number, in contrast to biological samples. Dynamical scattering and diffraction contrast contribute to the projection in a non-monotonic way [15].

In HAADF-STEM, electrons reaching the annular detector are incoherently scattered and this eliminates contributions due to phase contrast. The signal is proportional to Z^γ and is monotonic with the mass-thickness; hence, HAADF-STEM is preferred for ET in physical science [46]. However, in some types of sample the image intensities have a non-linear dependence with thickness both in TEM and STEM [19, 47, 48]. In HAADF-STEM, the reconstruction will show a downward arc for a homogeneous nanoparticle and this is known as the 'cupping artefact' [45, 49, 50]. This is the consequence of different amount of scattered electrons in different layers. The artefact can be corrected and minimised in optimised experimental conditions [49, 51].

The reconstruction could still be a faithful representation of the object — the conclusion from the reconstruction is valid — even the data does not strictly satisfy the projection requirement. In the following sections, traditional reconstruction algorithms and related mathematics will be reviewed.

Radon transform. In 1917, the Austrian mathematician Johann Radon laid the mathematical foundation of tomography and showed that a two-dimensional object can be fully represented by an inverse operation on an infinite number of one-dimensional projections [52]. This is known as the *Radon transform*. To understand the Radon transform, it is convenient to parametrise a line in its point-normal form $l_{\theta,t}$ (see Figure 1.4). Every point $\vec{p}$ on $l_{\theta,t}$ can be expressed as

$$\vec{p}(s) = t\vec{\omega} + s\vec{\omega}^{\perp} \qquad (t, s \in \mathbb{R}) \tag{1.4}$$

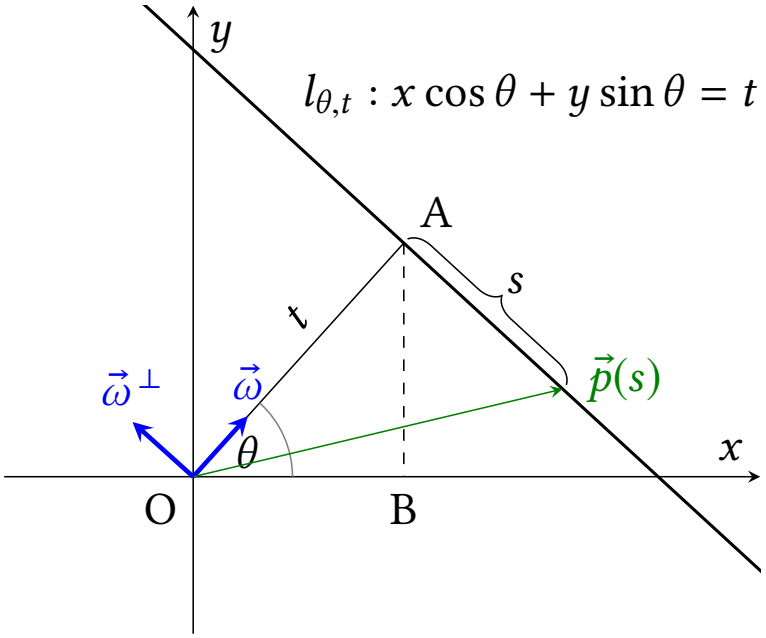


Figure 1.4: Parametrisation of a line by the perpendicular distance t to the line from the origin and its angle θ from the x-axis. Any point $\vec{p}(s)$ on the line $l_{\theta,t}$ can be expressed as $t\vec{\omega} + s\vec{\omega}^{\perp}$.

where $\vec{\omega}$ and $\vec{\omega}^{\perp}$ are defined as

$$\vec{\omega} = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} \tag{1.5}$$

$$\vec{\omega}^{\perp} = \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix} \tag{1.6}$$

Every point on $l_{\theta,t}$ can be written explicitly as

$$\begin{pmatrix} x(s) \\ y(s) \end{pmatrix} = \begin{pmatrix} t\cos\theta - s\sin\theta \\ t\sin\theta + s\cos\theta \end{pmatrix} \tag{1.7}$$

The Radon transform $\mathcal{R}$ of a function $f(x,y) \in \mathbb{R}^2$ is the mapping:

$$\mathcal{R}\left[f(x,y)\right] = \int_{l_{\theta,t}} f(x,y)\ \mathrm{d}l \tag{1.8}$$

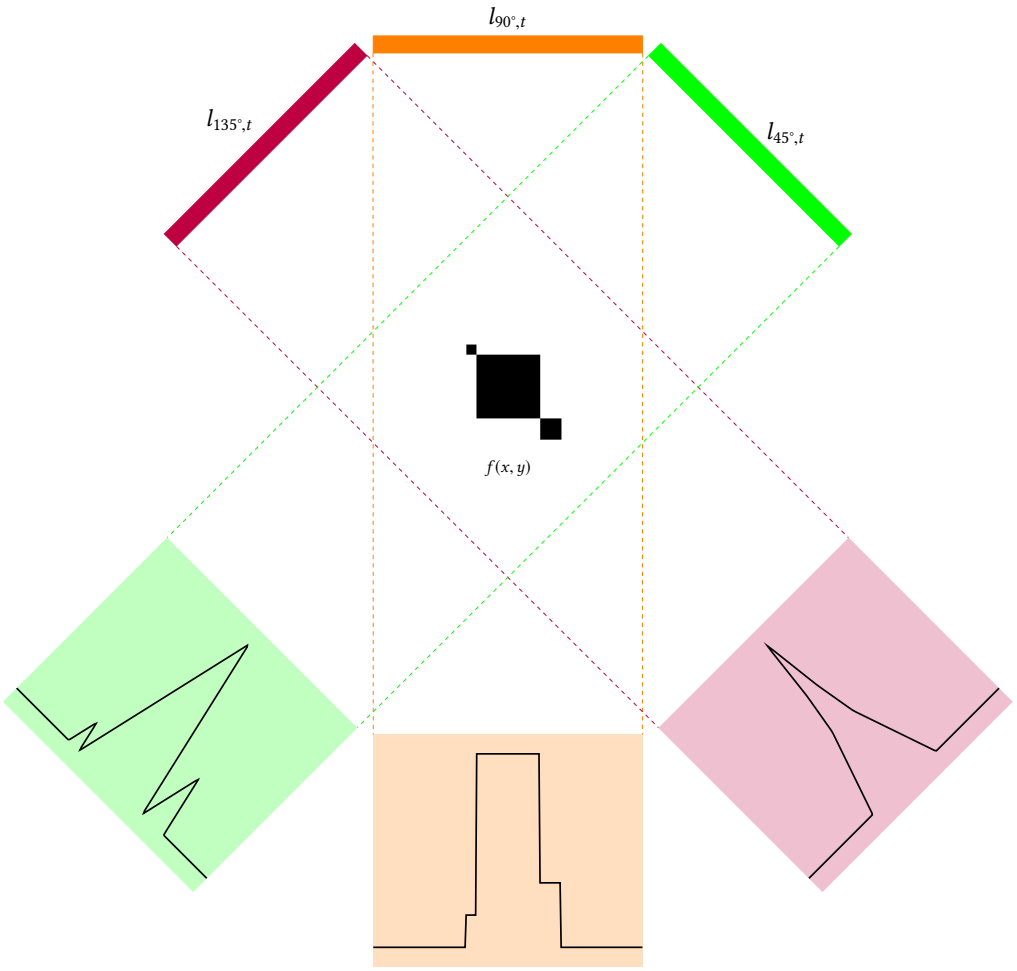


Figure 1.5: A projection measured in a tomography experiment is the Radon transform of an image function $f(x, y)$ along a set of parallel lines denoted by $l_{\theta,t}$. Three projections with $\theta = 45°, 90°, 135°$ are shown with their corresponding colours. A tomography dataset is a collection of Radon transforms at different sets of parallel lines at different θ.

The Radon transform is the line integral of a function $f(x, y)$ along a particular line $l_{\theta,t}$. This gives the total density of $f(x, y)$ along $l_{\theta,t}$. If $f(x, y)$ is an image function, a projection in a tomography experiment is the Radon transform of $f(x, y)$ from a set of parallel lines denoted by $l_{\theta,t}$ (see Figure 1.5). We can immediately appreciate the importance of the projection requirement. The reconstruction reduces to an inverse Radon transform which can give back f from $\mathcal{R}f$.

Using the parametrisation in Equation (1.7), the Radon transform can be expressed in a more explicit form:

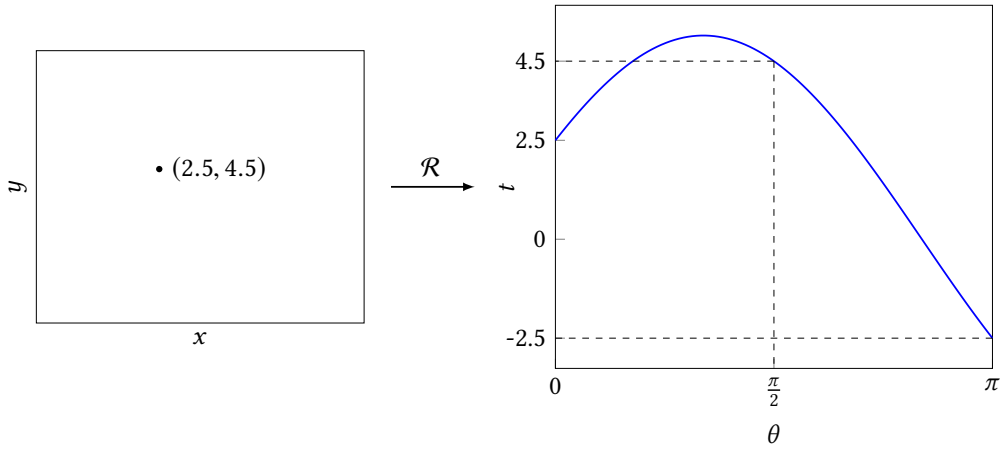
$$\begin{aligned}[\mathcal{R}f](\theta, t) &= \int_{-\infty}^{\infty} f(x(s), y(s))\, \mathrm{d}s \\ &= \int_{-\infty}^{\infty} f(t\cos\theta - s\sin\theta, t\sin\theta + s\cos\theta)\, \mathrm{d}s \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\delta(x\cos\theta + y\sin\theta - t)\, \mathrm{d}x\mathrm{d}y \end{aligned} \tag{1.9}$$

Equation (1.9) limits the integration for $(x, y) \in \mathbb{R}^2$ satisfying $x\cos\theta + y\sin\theta = t$, which is the parametrised line $l_{\theta,t}$ (see Figure 1.4). The delta function δ restricts the integration only along $l_{\theta,t}$.

If the object is a point source at (x_0, y_0), $f(x, y)$ can be written as $\delta(x-x_0)\delta(y-y_0)$. Using Equation (1.9), the Radon transform of a point source at (x_0, y_0) is

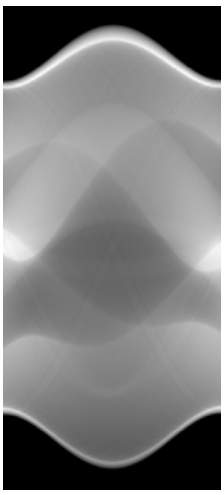
$$\begin{aligned}[\mathcal{R}f](\theta, t) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \delta(x - x_0)\delta(y - y_0)\delta(x\cos\theta + y\sin\theta - t)\, \mathrm{d}x\mathrm{d}y \\ &= \delta(x_0\cos\theta + y_0\sin\theta - t) \end{aligned} \tag{1.10}$$

In an experiment, projections from different orientations are taken and assembled into a two-dimensional matrix. As the Radon transform is a sinusoidal delta function in the (θ, t) space (see Figure 1.6a), this matrix is commonly known as the *sinogram*. The (θ, t) space is known as the *Radon domain*. Figure 1.6b shows the famous *Shepp-Logan* phantom, originated in 1974 for testing image reconstruction algorithms [53], and Figure 1.6c shows its sinogram. Some of the reconstructions in this section will be demonstrated by the Shepp-Logan phantom.



(a) point source

(b) real space



(c) Radon domain

Figure 1.6: (a) The Radon transform of a point source at $(x_0, y_0) = (2.5, 4.5)$ is a sinusoidal delta function in the Radon domain spanned by θ and t. In the Radon domain, it follows the equation $t = x_0 \cos\theta + y_0 \sin\theta$. For $\theta = \pi/2$, $t = y_0$; for $\theta = 0$ or π, $t = \pm x_0$. (b) and (c) are the Shepp-Logan phantom in real space and Radon domain respectively.

Fourier transform and convolution. A one-dimensional Fourier transform $\mathcal{F}_{1D}$ of a function $f(x)$ is:

$$\begin{aligned} \hat{f}(\omega) &= (\mathcal{F}_{1D} f)(\omega) \\ &= \int_{-\infty}^{\infty} f(x) e^{-i\omega x} \, \mathrm{d}x \end{aligned} \tag{1.11}$$

The inverse one-dimensional Fourier transform $\mathcal{F}_{1D}^{-1}$ is:

$$\mathcal{F}_{1D}^{-1}\left[\hat{f}(x)\right] = \frac{1}{2\pi}\int_{-\infty}^{\infty}\hat{f}(\omega)e^{i\omega x}\,\mathrm{d}\omega \tag{1.12}$$

This is straightforward to extend to higher dimensions. For the two-dimensional Fourier transform $\mathcal{F}_{2D}$:

$$\begin{aligned}\hat{f}(u,v) &= (\mathcal{F}_{2D}f)(u,v)\\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(x,y)e^{-i(ux+vy)}\,\mathrm{d}x\mathrm{d}y\end{aligned} \tag{1.13}$$

The Fourier transform represents the function in the frequency domain. The Fourier transform at a particular spatial frequency ω is equal to the sum of the function multiplying with the complex exponential $e^{-i\omega x}$. This is equivalent to sampling the function at every point with a particular spatial frequency ω. The higher the magnitude of the Fourier transform at ω, the more prevalent the frequency ω is in the function.

The convolution of two functions f and g can be denoted by $f * g$, and it is defined as:

$$(f * g)(x) = \int_{-\infty}^{\infty}f(\tau)g(x-\tau)\,\mathrm{d}\tau \tag{1.14}$$

The convolution theorem is

$$f * g = \mathcal{F}^{-1}\left\{\mathcal{F}\left\{f\right\}\cdot\mathcal{F}\left\{g\right\}\right\} \tag{1.15}$$

Convolution of a signal f by a transfer function g can be regarded as averaging f by g. If the functions are in spatial domain, f is the object and g can be regarded as the filter. The filtered object is modified by the filter, often blurred or smeared. One convenient way to evaluate the convolution of two functions is to multiply their Fourier transforms and perform an inverse Fourier transform

of the product, as stated in Equation (1.15). The proof of convolution theorem can be found in standard mathematical textbooks such as [54].

Central slice theorem. Many reconstruction algorithms rely heavily on the central slice theorem, as it provides a simple connection between one-dimensional Radon transform and two-dimensional Fourier transform of the image function [55]. The central slice theorem states that the one-dimensional Fourier transform of the Radon transform of $f \in \mathbb{R}^2$ in the direction θ is equal to the central slice along θ in the two-dimensional Fourier transform of f. Mathematically, it can be written as

$$\mathcal{F}_{1D}\left[(\mathcal{R}f)(\theta, t)\right](\omega) = \mathcal{F}_{2D}\left[f(\omega\cos\theta, \omega\sin\theta)\right] \tag{1.16}$$

To prove Equation (1.16), the Radon transform written in the form of Equation (1.9) and polar coordinates in the two-dimensional Fourier space are used. The left hand side of Equation (1.16) is

$$\begin{aligned}
\mathcal{F}_{1D}\left[\mathcal{R}f\right](\omega) &= \int_{-\infty}^{\infty} \mathcal{R}f(\theta, t)e^{-i\omega t}\,\mathrm{d}t \\
&= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\delta(x\cos\theta + y\sin\theta - t)e^{-i\omega t}\,\mathrm{d}x\mathrm{d}y\right)\mathrm{d}t \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\left(\int_{-\infty}^{\infty}\delta(x\cos\theta + y\sin\theta - t)e^{-i\omega t}\,\mathrm{d}t\right)\mathrm{d}x\mathrm{d}y \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)e^{-i\omega(x\cos\theta + y\sin\theta)}\,\mathrm{d}x\mathrm{d}y \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)e^{-i(x\omega\cos\theta + y\omega\sin\theta)}\,\mathrm{d}x\mathrm{d}y \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)e^{-i(xu+yv)}\,\mathrm{d}x\mathrm{d}y \\
&= \hat{f}(u,v) \\
&= \hat{f}(\omega\cos\theta, \omega\sin\theta) \\
&= \mathcal{F}_{2D}f(\omega\cos\theta, \omega\sin\theta)
\end{aligned}$$

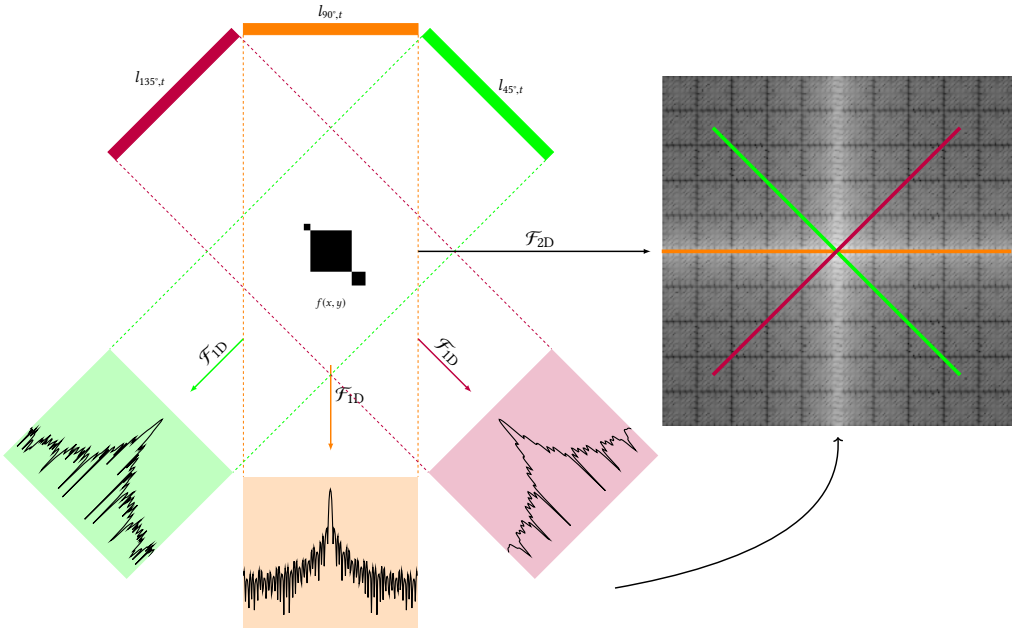


Figure 1.7: Application of one-dimensional Fourier transform on the projections in different orientations is equal to the two-dimensional Fourier transform of the sample along the respective directions. The Fourier-transformed projections can be assembled in the Fourier space and this is known as a layergram.

The significance of the central slice theorem is that the Fourier-transformed projections in different orientations can be put together in the Fourier space (see Figure 1.7). This is sometimes regarded as a *layergram* [56].

One might think that by doing a direct two-dimensional inverse Fourier transform on the layergram, the object can be recovered; however in an actual experiment, there are only discrete values of θ and t. Since the layergram is constructed by polar coordinates, an interpolation to Cartesian grid is required and it is difficult to interpolate accurately (see Figure 1.8) . This is known as *regridding*. As a result, direct inverse two-dimensional Fourier transform of a layergram is seldom used to recover the object.

Back-projection methods. All reconstruction problems in tomography reduce to finding inverse operations on the Radon transform; however, there is

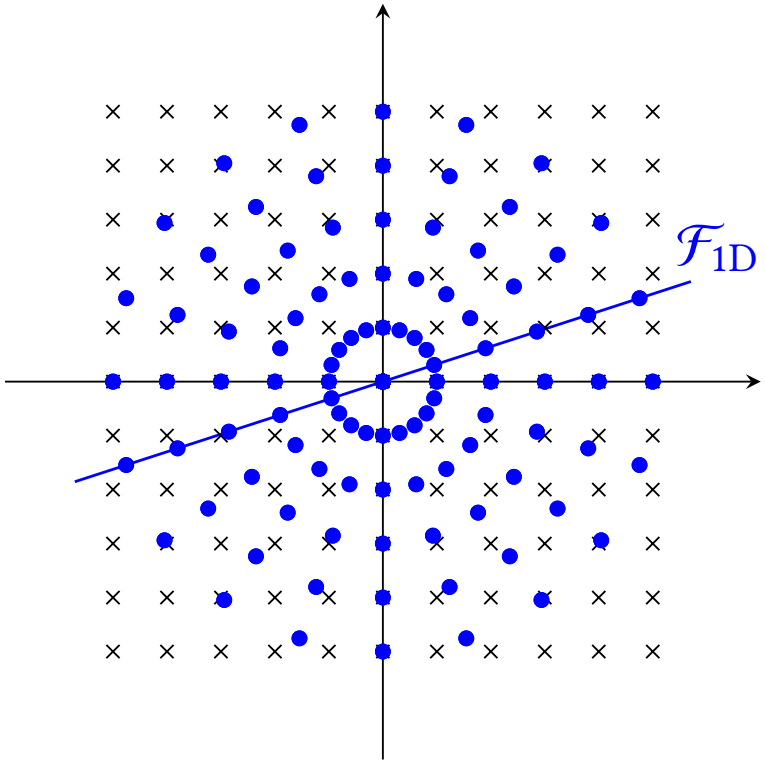


Figure 1.8: The blue dots are polar coordinates and the black crosses are Cartesian coordinates. For an inverse two-dimensional Fourier transform in polar coordinates, it needs to interpolate to the Cartesian coordinates as they do not overlap. This is known as regridding.

no simple and straightforward method available. One of the most common methods is to back-project from the Radon domain. A more detailed discussion of back-projection can be found in [57]. Back-projection $\mathcal{B}$ yielding the intensity at a point $(x, y) \in \mathbb{R}^2$ from the Radon transform is

$$\{\mathcal{B}[\mathcal{R}f]\}(x, y) = \int_0^\pi \mathcal{R}f(x\cos\theta + y\sin\theta)\,\mathrm{d}\theta \tag{1.17}$$

For every line $l_{\theta,t}$ that passes through the point (x, y), the total intensity of all points along $l_{\theta,t}$ is captured by the intensity at θ in the sinogram. Back-projection is equivalent to summing the 'smeared sinogram' and is the accumulation of line integrals passing through the point (x, y) (see Figure 1.9). *Simple back-projection* is an attempt to recover $f(x, y)$ directly from the back-projection.

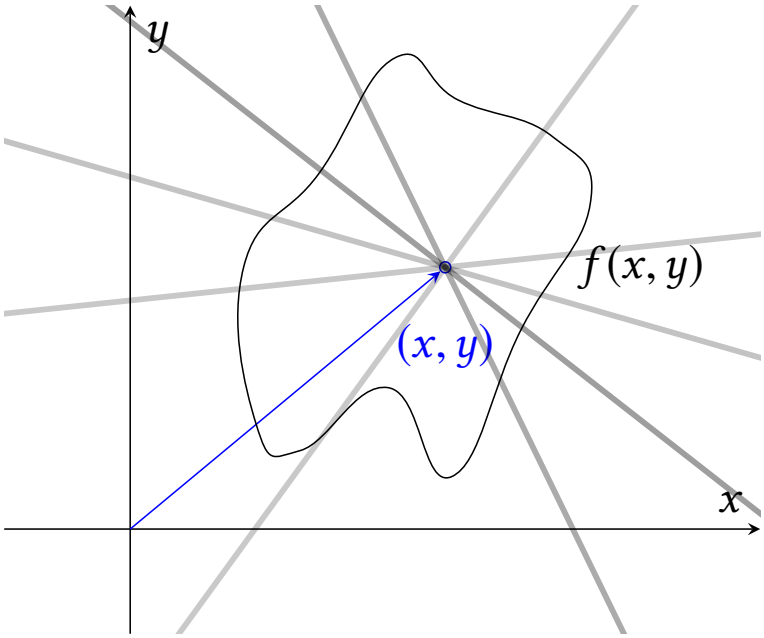


Figure 1.9: The back-projection of Radon transform of a point (x, y) inside the image function $f(x, y)$ involves two steps. First, the intensities in the sinogram at different θ are smeared back into the point. The darker the line, the higher the intensity. Finally, all intensities are summed.

The integrand of back-projection can be expressed as

$$\begin{aligned}
\mathcal{R}\left[f(x\cos\theta + y\sin\theta)\right] &= \mathcal{F}_{1\mathrm{D}}^{-1}\left\{\mathcal{F}_{1\mathrm{D}}\mathcal{R}f(\omega)\right\} \\
&= \mathcal{F}_{1\mathrm{D}}^{-1}\left[\hat{f}(\omega\cos\theta, \omega\sin\theta)\right] \quad \text{(central slice theorem)} \\
&= \int_{-\infty}^{\infty} \hat{f}(\omega\cos\theta, \omega\sin\theta) e^{i\omega t}\, \mathrm{d}\omega \\
&= \int_{-\infty}^{\infty} \hat{f}(\omega\cos\theta, \omega\sin\theta) e^{i\omega(x\cos\theta + y\sin\theta)}\, \mathrm{d}\omega \qquad (1.18)
\end{aligned}$$

Substitute Equation (1.18) into Equation (1.17), the back-projection can be written as

$$\mathcal{B}\left\{\mathcal{R}f\right\}(x, y) = \int_0^{\pi}\int_{-\infty}^{\infty} \hat{f}(\omega\cos\theta, \omega\sin\theta) e^{i\omega(x\cos\theta + y\sin\theta)}\, \mathrm{d}\omega\mathrm{d}\theta \qquad (1.19)$$

Equation (1.19) is a polar integral, but the integration bound is not correct. To address this issue, note that the line $l_{\theta,-t}$ is the same line as $l_{\theta+\pi,t}$.

$$\begin{aligned} l_{\theta+\pi,t} : x\cos(\theta+\pi) + y\sin(\theta+\pi) &= t \\ -x\cos\theta - y\sin\theta &= t \\ x\cos\theta + y\sin\theta &= -t \end{aligned}$$

As $l_{\theta,-t} = l_{\theta+\pi,t}$, their line integrals are the same; therefore, $\mathcal{R}f(\theta,-t) = \mathcal{R}f(\theta+\pi,t)$. By splitting Equation (1.19) into two parts:

$$\begin{aligned} &\int_0^{\pi}\int_{-\infty}^{0} \hat{f}(\omega\cos\theta,\omega\sin\theta)e^{i\omega(x\cos\theta+y\sin\theta)}\,\mathrm{d}\omega\mathrm{d}\theta + \dots \\ &\int_0^{\pi}\int_0^{\infty} \hat{f}(\omega\cos\theta,\omega\sin\theta)e^{i\omega(x\cos\theta+y\sin\theta)}\,\mathrm{d}\omega\mathrm{d}\theta \end{aligned} \tag{1.20}$$

Recognising the property $\mathcal{R}f(\theta,-t) = \mathcal{R}f(\theta+\pi,t)$, the first part of Equation (1.20) can be written as

$$\int_{\pi}^{2\pi}\int_0^{\infty} \hat{f}(\omega\cos\theta,\omega\sin\theta)e^{i\omega(x\cos\theta+y\sin\theta)}\,\mathrm{d}\omega\mathrm{d}\theta \tag{1.21}$$

Combining the two integrals in Equation (1.20) by using Equation (1.21), the back-projection can be written as

$$\mathcal{B}\left\{\mathcal{R}f\right\}(x,y) = \int_0^{2\pi}\int_0^{\infty} \hat{f}(\omega\cos\theta,\omega\sin\theta)e^{i\omega(x\cos\theta+y\sin\theta)}\,\mathrm{d}\omega\mathrm{d}\theta \tag{1.22}$$

To change from polar coordinate to Cartesian coordinate, let $u = \omega\cos\theta$ and $v = \omega\sin\theta$ in the two-dimensional Fourier space. The Jacobian is

$$\mathbf{J} = \begin{vmatrix} \frac{\partial u}{\partial\omega} & \frac{\partial u}{\partial\theta} \\ \frac{\partial v}{\partial\omega} & \frac{\partial v}{\partial\theta} \end{vmatrix} = \begin{vmatrix} \cos\theta & -\omega\sin\theta \\ \sin\theta & \omega\cos\theta \end{vmatrix} = \omega \tag{1.23}$$

Rewrite Equation (1.22) as Cartesian coordinate:

$$\begin{aligned}\mathcal{B}\left\{\mathcal{R}f\right\}(x,y) &= \int_0^{2\pi}\int_0^{\infty} \hat{f}(\omega\cos\theta,\omega\sin\theta)e^{i\omega(x\cos\theta+y\sin\theta)}\,\mathrm{d}\omega\mathrm{d}\theta \\ &= \int_0^{2\pi}\int_0^{\infty} \frac{1}{\omega}\cdot\hat{f}(\omega\cos\theta,\omega\sin\theta)e^{i\omega(x\cos\theta+y\sin\theta)}\,\omega\mathrm{d}\omega\mathrm{d}\theta \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{\sqrt{u^2+v^2}}\cdot\hat{f}(u,v)e^{i(xu+yv)}\,\mathrm{d}u\mathrm{d}v \end{aligned} \tag{1.24}$$

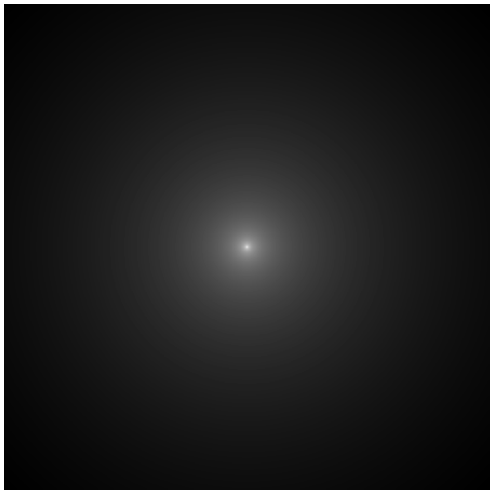
The convolution theorem in Equation (1.15) can be used to simplify Equation (1.24). It follows from the convolution theorem that

$$\mathcal{F}^{-1}(f) * \mathcal{F}^{-1}(g) = \mathcal{F}^{-1}\left\{f\cdot g\right\} \tag{1.25}$$

The back-projection is now

$$\begin{aligned}\mathcal{B}\left\{\mathcal{R}f\right\}(x,y) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{\sqrt{u^2+v^2}}\cdot\hat{f}(u,v)e^{i(xu+yv)}\,\mathrm{d}u\mathrm{d}v \\ &= \mathcal{F}_{\mathrm{2D}}^{-1}\left\{\frac{1}{\sqrt{u^2+v^2}}\right\} * \mathcal{F}_{\mathrm{2D}}^{-1}\left\{\hat{f}(u,v)\right\} \\ &= \frac{1}{\sqrt{x^2+y^2}} * f(x,y) \end{aligned} \tag{1.26}$$

Simple back-projection does not give back the original image function $f(x,y)$, instead it is convoluted with the function $\frac{1}{\sqrt{x^2+y^2}}$ (see Figure 1.10a). It gives a blurred version of $f(x,y)$, as shown in Figure 1.10b. This is known as the *star-like artefact* (Figure 1.11a). Even with infinite projections, the result is always blurred. The origin of the blurring is the $\frac{1}{\omega}$ factor from the Jacobian when polar coordinate is used — the intensity of the reconstruction is proportional to $\frac{1}{r}$.



(a) $\frac{1}{\sqrt{x^2+y^2}}$

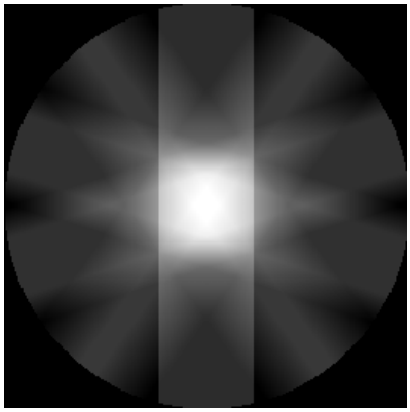
(b) Simple back-projection result

Figure 1.10: Simple back-projection always gives the original object convoluted with the function $1/\sqrt{x^2+y^2}$ (see (a)) and results in a blurred version of the object as shown in (b).

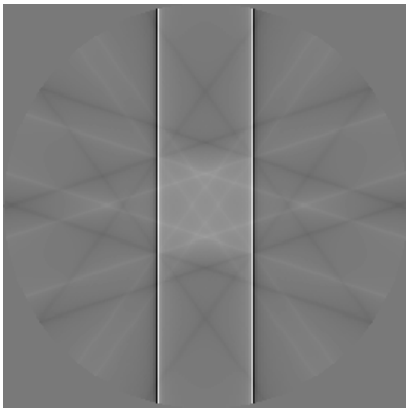
If a certain kind of 'filter' $\mathcal{W}$ is used, the factor $\frac{1}{\omega}$ can be removed. For $\mathcal{W} = \omega$, the back-projection with the filter is

$$\begin{aligned}
\mathcal{B}\{\mathcal{W}\mathcal{R}f\}(x,y) &= \int_0^{2\pi}\int_0^{\infty} \mathcal{W}\hat{f}(\omega\cos\theta, \omega\sin\theta)e^{i\omega(x\cos\theta+y\sin\theta)}\,\mathrm{d}\omega\mathrm{d}\theta \\
&= \int_0^{2\pi}\int_0^{\infty} \omega\hat{f}(\omega\cos\theta, \omega\sin\theta)e^{i\omega(x\cos\theta+y\sin\theta)}\,\mathrm{d}\omega\mathrm{d}\theta \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \hat{f}(u,v)e^{i(xu+yv)}\,\mathrm{d}u\mathrm{d}v \\
&= f(x,y) \qquad (1.27)
\end{aligned}$$

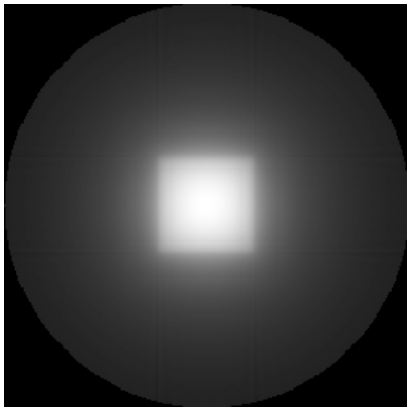
The object $f(x,y)$ is completely recovered for an infinite number of projections. This is known as *filtered back-projection* [58]. The star-like artefact is suppressed as shown in Figure 1.11. The filter $\mathcal{W} = \omega$ is called a ramp filter, and it corrects the measurement in Fourier space as the low-frequency region is over-sampled (see Figure 1.8). It puts different weights on different points in the projections and it is also known as *weighted back-projection*.



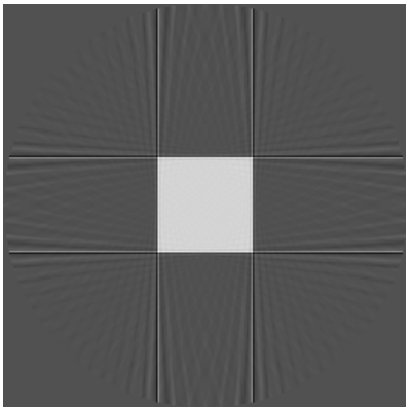
(a) 5 projections



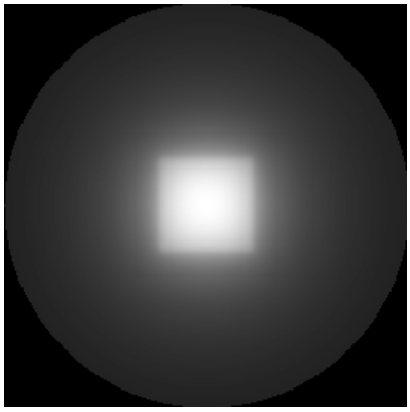
(b) 5 projections



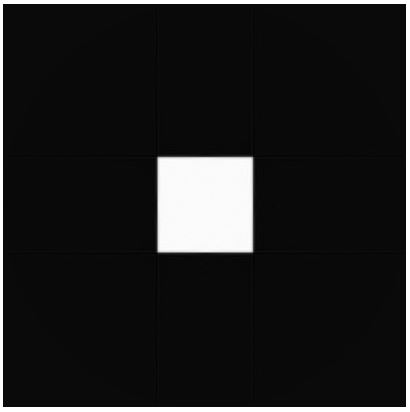
(c) 50 projections



(d) 50 projections



(e) 500 projections



(f) 500 projections

Figure 1.11: Reconstructions of a square object: (a), (c) and (e) are from simple back-projection and (b), (d) and (f) are from filtered back-projection. In (a) the star-like artefact is very prominent with 5 projections, while in (b) the reconstruction is poor due to insufficient projections. When 50 projections are used, filtered back-projection gives a very good result and with 500 projections, it completely recovers the square; no matter how many projections are used, the results are always blurred in simple back-projection.

For experimental datasets, there are only a finite number of projections and noise is always present, so in practice additional low-pass filters (for reduction of high-frequency noise) are applied to improve the quality of the reconstructions [59–61].

Algebraic reconstruction methods. Back-projection is a promising way to reconstruct tomographic data as it provides an analytical solution to the inversion of Radon transform. It assumes an infinite number of ideal projections distributed across the whole angular range; however, this is never achieved in practice. Tomographic data often contains a missing wedge (see Section 1.2.2), suffers from noise and has a non-uniform distribution of tilt angles. While back-projection methods work very well on complete tomographic data, algebraic methods have a larger tolerance for imperfect datasets.

All algebraic methods translate the problems into a system of linear equations and the reconstruction reduces to solving the system with an unique solution, which is the object of interest. In algebraic methods, the image $f(x, y)$ is divided into N pixels such that the value in each pixel is constant (see Figure 1.12). The value in pixel j is denoted by f_j. In a pixellated image, the line integral of a beam b_i can be written as

$$\sum_{j=1}^{N} a_{ij} f_j = b_i \tag{1.28}$$

a_{ij} is the contribution of pixel j to b_i for the beam i and is proportional to the overlap of the beam i on pixel j. For a particular beam i, most of the a_{ij} are zero because it only intersects a very small number of pixels. If there are M beams,

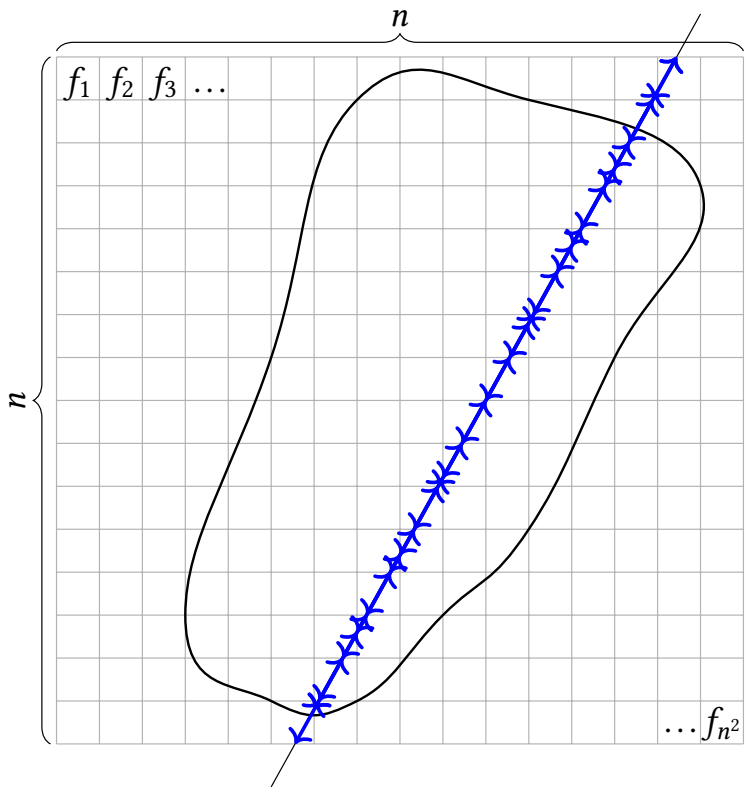


Figure 1.12: The image is divided into $n \times n$ pixels with each pixel having an intensity value $f_1, f_2, \ldots, f_{n^2}$. The line integral of a beam in a pixellated image is the summation of the contribution from each pixel and this contribution is proportional to the length of the beam passing through that pixel. If the beam does not pass through that pixel, the contribution is zero.

a system of linear equations can be set up:

$$
\begin{aligned}
a_{11}f_1 + a_{12}f_2 + \ldots + a_{1N}f_N &= b_1 \\
a_{21}f_1 + a_{22}f_2 + \ldots + a_{2N}f_N &= b_2 \\
\vdots \qquad\quad & \vdots \qquad\quad \vdots \\
a_{i1}f_1 + a_{i2}f_2 + \ldots + a_{iN}f_N &= b_i \\
\vdots \qquad\quad & \vdots \qquad\quad \vdots \\
a_{M1}f_1 + a_{M2}f_2 + \ldots + a_{MN}f_N &= b_M
\end{aligned}
$$

The image consisting of pixels with values $f_1, f_2, \ldots, f_N$ is a point in $\mathbb{R}^N$ and can be denoted by $\vec{f}$. The system of linear equations can be represented in

vector form:

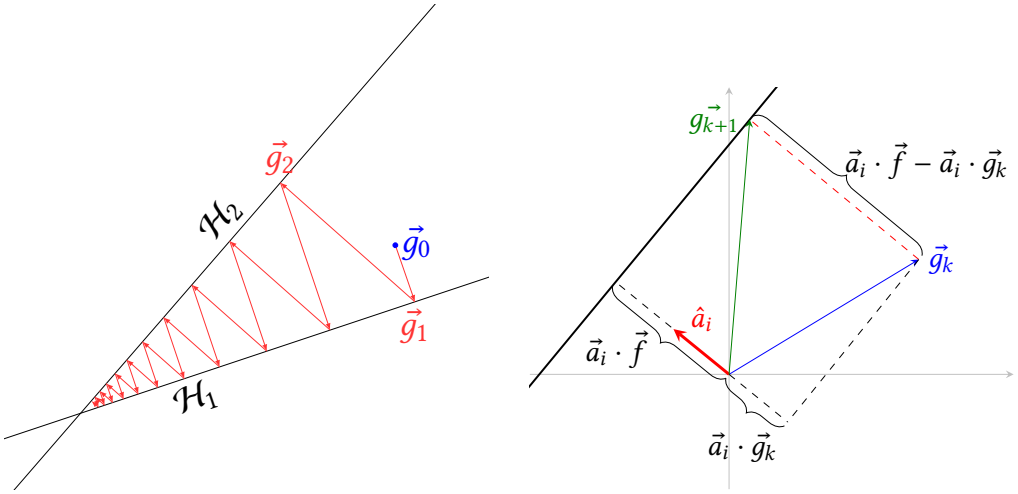
$$\begin{aligned}\vec{a_1} \cdot \vec{f} &= b_1 \\ \vec{a_2} \cdot \vec{f} &= b_2 \\ \vdots &= \vdots \\ \vec{a_M} \cdot \vec{f} &= b_M\end{aligned}$$

$\vec{a_i}$ is a $(1 \times N)$-dimensional vector containing the contribution of each pixel j for the beam i. Each linear equation is an affine hyperplane in $\mathbb{R}^N$ for $b_i \neq 0$. The intersection of all these hyperplanes is a single point in $\mathbb{R}^N$ giving the image $\vec{f}$, assuming there is a unique solution. Instead of solving the system analytically, algebraic methods employ an iterative approach.

In 1937, the Polish mathematician Stefan Kaczmarz laid an iterative approach to solve a system of linear equations and this is known as the *Kaczmarz method* [62]. With the developments of tomography, scientists independently suggested the same algorithm under the name algebraic reconstruction technique (ART) in 1970 [63]. For the first iteration, it starts at a random point $\vec{g_0} \in \mathbb{R}^N$ and evaluates the orthogonal projection of $\vec{g_0}$ on a hyperplane which is denoted by $\vec{g_1}$. $\vec{g_1}$ is then projected onto the next hyperplane in the second iteration, giving $\vec{g_2}$. The iteration stops when the projection no longer changes within a threshold [62, 63]. This is illustrated graphically in Figure 1.13 by two intersected lines.

To find the orthogonal projection of $\vec{g_k}$ on the hyperplane $\vec{a_i} \cdot \vec{f}$, note that $\vec{a_i}$ is perpendicular to the hyperplane $\vec{a_i} \cdot \vec{f}$. The unit vector orthogonal to the hyperplane $\hat{a_i}$ is

$$\hat{a_i} = \frac{\vec{a}}{\|\vec{a}\|_2^2} \tag{1.29}$$



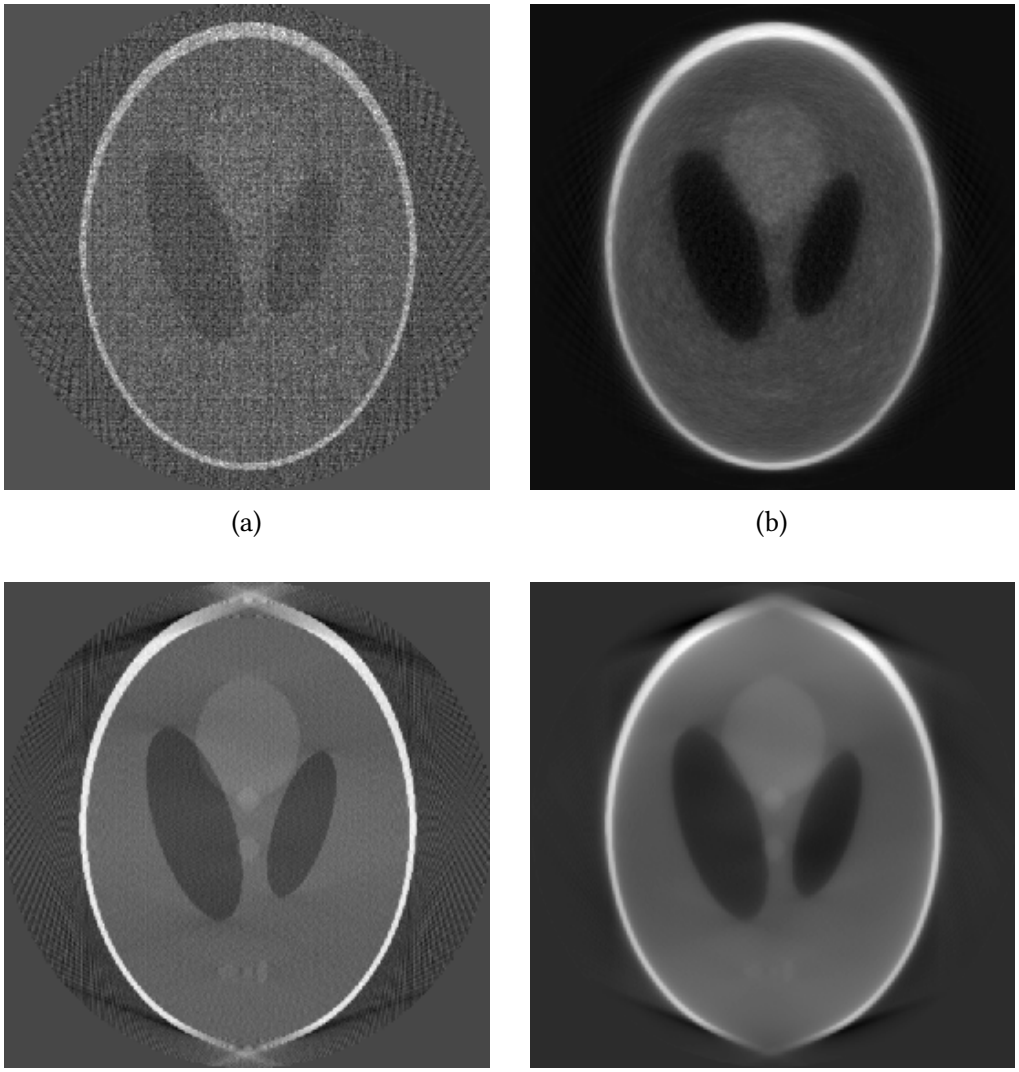
(a) Kaczmarz method (b) Orthogonal projection

Figure 1.13: (a) suppose there are two intersected hyperplanes $\mathcal{H}_1$ and $\mathcal{H}_2$ in a two-dimensional space (two lines), a random point $\vec{g}_0$ is selected and orthogonally projected onto one line. The orthogonal projection is then projected onto the other line, and this repeats until convergence. (b) the orthogonal projection of $\vec{g}_k$ is $\vec{g_{k+1}}$, which is $\vec{g}_k + c\hat{a}_i$ for some constant c. c has a value of $\vec{a}_i \cdot \vec{f} - \vec{a}_i \cdot \vec{g}_k$.

The orthogonal projection is $\vec{g}_k + c\hat{a}_i$ for some constant c. In Figure 1.13b, c is equal to $\vec{a}_i \cdot \vec{f} - \vec{a}_i \cdot \vec{g}_k = b_i - \vec{a}_i \cdot \vec{g}_k$; therefore, the orthogonal projection $\vec{g_{k+1}}$ is

$$\vec{g_{k+1}} = \vec{g}_k + \frac{b_i - \vec{a}_i \cdot \vec{g}_k}{\|\vec{a}\|_2^2}\vec{a}_i \tag{1.30}$$

After M iterations, all the hyperplanes have been considered and the rate of convergence depends on the angles between hyperplanes. It is shown that if a single intersection point $\vec{f}$ is present, the algorithm will be converged for any system of linear equations [64]. Different approaches are used to increase the convergence rate: orthogonalisation of hyperplanes [65] or considering the hyperplanes in an optimised order [66]. Constraints can be enforced in ART, and the most common constraints are non-negativity and the 'box constraint' — the box constraint limits the pixel values from 0 to 1. Figure 1.14 compares the reconstructions from ART and filtered back-projection with incomplete dataset and it is clear that ART is superior.



(a) (b)

(c) (d)

Figure 1.14: (a) and (c) are reconstructions from filter back-projection with Poisson noise and missing wedge respectively; (b) and (d) are reconstructions from ART with noise and missing wedge respectively. (a) and (b) are reconstructions with 50 projections from 0° to 180°; (c) and (d) are reconstructions from ±70° with 2°-interval.

ART usually suffers from salt-and-pepper noise as in each iteration, it only changes the values of the pixels along a particular beam. This artefact can be reduced by introducing a relaxation parameter λ to control the size of the update [67]. Alternatively, the number of beams considered can be increased so that the update is averaged among them [68].

Equation (1.30) is the foundation of all algebraic methods. Instead of performing the iteration in a sequential way, simultaneous iterative reconstructive technique (SIRT) obtains the update by considering all the beams in a single iteration [69], and this is more robust to noise. ART generally produces very noisy reconstructions, but the convergence is quick; SIRT produces a smoother reconstruction at the expense of slower convergence. Simultaneous algebraic reconstruction technique (SART) takes the advantages of ART and SIRT to produce a smoother reconstruction in a single iteration. It considers a subset of beams in a particular direction and updates simultaneously in that direction [70]. The convergence of SART was proven for any sequence of the subset chosen [71, 72]. Numerous extensions of ART were proposed, and to name a few: component averaging (CAV) [73], diagonally relaxed orthogonal projections (DROP) [74], discrete algebraic reconstruction technique (DART) [75], ordered-subset simultaneous algebraic reconstruction technique (OS-SART) [76], frequency adapted simultaneous algebraic reconstruction technique (FA-SART) [77].

Compressed sensing. The quality of reconstruction improves with the number of projections in back-projection and algebraic reconstruction methods; however, it is generally not desirable to obtain a very large number of projections due to radiation damage, which is particularly serious in biological samples, or it is not practical and physically impossible because of the missing wedge and other constraints (see Section 1.2.2). There is a great interest in the ET community to develop methods which can retain a high quality of reconstruction by using fewer projections, and compressed sensing has been gaining a lot of interests in recent years in the field of ET [78].

Compressed sensing is a sampling technique that allows a large reduction in sampling rate (lower than the classical Nyquist rate) while still reconstructs the original signal at high accuracy. It was proposed by Emmanuel Candés,

Terence Tao, Justin Romberg and David Donoho in 2006 [79, 80]. The idea of compressed sensing relies on two important concepts: *sparsity* and *incoherence*. A signal x is sparse if most of its entries are zero; it is *k-sparse* if there are at most k non-zero entries. Most signals are not sparse in their natural representation, but they are or can be well approximated as sparse when they are described in some suitable representation bases. For an image of a periodic structure, the pixel intensities are mostly non-zero but this image can be regarded as sparse when it is represented by spatial frequencies; in this case, the representation basis Ψ is the discrete Fourier transform. A sparse signal can be written as Ψx when it is sparse in the representation basis Ψ. A sensing basis Φ samples the sparse signal and it can be written as $\Phi\Psi x$. The relationship between Φ and Ψ needs to be incoherent in compressed sensing, and the coherence between two bases can be expressed as

$$\mu(\Phi, \Psi) = \sqrt{n} \cdot \max_{1 \leq k,j \leq n} |\langle \phi_k, \psi_j \rangle| \tag{1.31}$$

μ is between 1 and $\sqrt{n}$ (when Φ and Ψ are normalised), where n is the number of pixel if the object is an image. If an image is obtained by a pixel detector, the sensing basis is the spike basis (Dirac delta function). Maximum incoherence is achieved, for example, by the spike basis and the discrete Fourier transform [81]. The recovered signal $\hat{x}$ can be found by solving the optimisation problem

$$\hat{x} = \text{argmin}_z \|z\|_1 \qquad z \in \mathcal{S}(y) \tag{1.32}$$

with

$$\mathcal{S}(y) = \{z : \Phi\Psi z = y\} \tag{1.33}$$

$\hat{x}$ has minimal ℓ_1-norm consistent with the measurement y and the original signal can be obtained by $\Psi^{-1}\hat{x}$. The theoretical foundation of compressed sensing can be found in [81, 82].

Compressed sensing in ET can be considered as a reconstruction technique using prior information of the sample, and the key prior information is that the signal is sparse in a specific representation basis. The first reconstruction in ET by compressed sensing was reported in 2011 for determining the three-dimensional morphology of iron oxide nanoparticles [83]. Since then, different sparse domains have been explored such as total variation (finite differences) [84, 85] and redundant dictionary [86]. More recent methods utilised an adaptive (learning) approach to seek the best sparse representation basis instead of using a fixed representation basis to reduce artefacts [87–90].

Compressed sensing in ET allows reconstruction to be performed from a very small number of projections and it assumes all the important information is sampled in those projections; therefore, careful experimental design is required to include the important structural information in those projections. The choice of the representation basis for one sample cannot be directly translated to another sample — for example, total variation is effective for samples that can be described by piece-wise constant functions (in other words, with sharp edges), and this may not be applicable to other samples without sharp edges. An adaptive approach in compressed sensing is important to reduce the bias on the choice of the representation basis. If there are undesirable defects present in the sample, the training dataset for any data-driven approach should account for those non-ideal structures; otherwise, strong justification is needed to assume the structure is free from any defects.

1.3 Atomic electron tomography

Conventional ET is well established and used widely in commercial scanners for medical purposes and for imaging biological samples, but the resolution of reconstruction does not go below 1 nm, limiting the application in physical sciences. With the developments in modern electron microscopy, the resolution of electron microscopes can reach below 50 pm because of improvements in aberration correction [91, 92].

Atomic electron tomography (AET) targets resolving the individual atomic positions in nanoparticles. To achieve atomic resolution in AET, all projections need to be aligned to a common axis with atomic precision. This can be achieved by using the centre-of-mass method [11]. However, under conventional tilting geometries, the nanoparticles cannot usually be tilted beyond ±70° and the reconstruction still suffers from a missing wedge. This can be avoided by supporting the sample on a needle-shaped platform [12].

ET was used to identify dislocations separated by a few nm in Mg-doped α-GaN (using filtered back-projection) [93] and pores with sub-nm resolution in porous SiCOH (using SIRT) [94]. In 2011, AET successfully identified individual atoms of silver nanoparticles of 2 nm–3 nm in diameter by discrete tomography [95], although discrete tomography requires atoms to be on a single lattice. Since then, information such as the locations of individual atoms, lattice strain and elemental information in three dimensions have been obtained [11, 12, 96–100]. In this section, important techniques in AET will be described: discrete tomography in Section 1.3.1, equally sloped tomography (EST) in Section 1.3.2, generalised Fourier iterative reconstruction (GENFIRE) in Section 1.3.3.

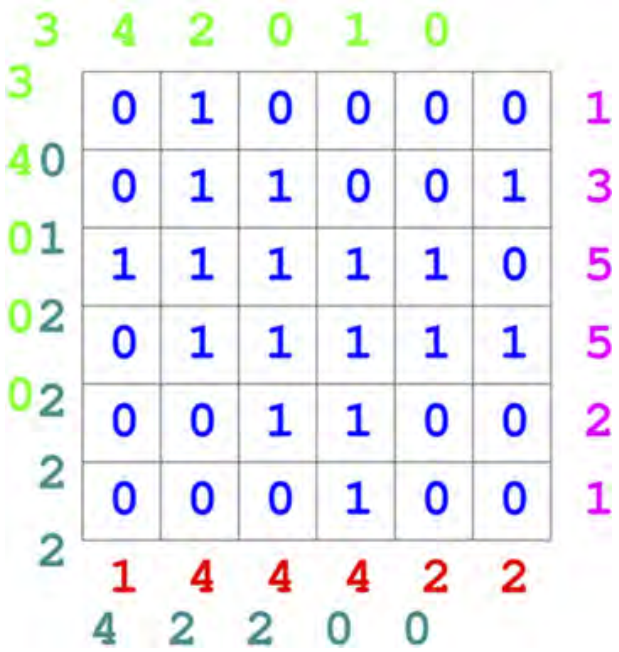


Figure 1.15: The basic mathematical model of discrete tomography: '1' indicates occupied grid point and '0' indicates vacant site. Atoms are counted in four projections in this grid: horizontal (purple), vertical (red), diagonal (cyan) and anti-diagonal (green). The figure is adapted from [102].

1.3.1 Discrete tomography

The first experimental AET to identify individual atoms was achieved by discrete tomography [95]. Discrete tomography utilises prior knowledge of the sample: atoms in a crystalline material are in a regular grid. All the atoms are also 'discrete' in nature, in which each grid point is either occupied or vacant. The discrete nature of atoms allows us to formulate the reconstruction as a binary problem [101].

The mathematical model of discrete tomography is shown in Figure 1.15 [102]. The occupied and vacant grid points are represented as '1' and '0' respectively. Projections from different directions can be considered as different linear equations where the unknowns are either 1 or 0. This assumes all the atoms are anchored to a fixed grid point without deviation and respond to the electron probe identically. As the intensities in the projections are directly related to

the number of occupied atoms, it is important to count the number of atoms in each projection accurately and this relies on accurate quantification of HAADF-STEM images [16].

To interpret the intensity values from HAADF-STEM image accurately, dynamical scattering needs to be considered as all projections are taken along zone axes in discrete tomography. When imaging along low-index zone axes, the positive electrostatic potential of the atoms acts as a channel for the incident electron probe, causing the intensity to focus tightly along the atomic columns. This is called the *channelling effect* [103]. When the electrons propagate along the column, some intensities can appear on neighbouring atomic columns. This is called *cross-talk* [104]. These dynamical effects can impact the interpretation of pixel intensities and the accurate counting of atoms. HAADF-STEM imaging contrast can be described by using the Bloch wave method with tightly bound *s*-type Bloch states of the electrons in the sample; when imaging along low-index zone axes, the contribution from neighbouring atomic columns can be neglected if their *s*-states do not overlap with each other [22, 23]. The contributions due to channelling effect and cross-talk while imaging along low-index zone axes can be quantitatively accounted for by using the Bloch wave method; nonetheless, if the specimen is thin (<10 nm), these dynamical effects are negligible [105–108].

By using only two images along the $[1\,0\,\bar{1}]$ and $[1\,0\,0]$ directions, individual atoms of a silver nanoparticle can be traced using discrete tomography [95]. Further restrictions are imposed on the discrete model: it is a face-centred cubic (fcc) lattice and there are no vacancies. Multi-slice simulations confirmed the cross-talk between atomic columns in $[1\,0\,\bar{1}]$ and $[1\,0\,0]$ zone-axis orientations were negligible; this is not true if the inter-column spacing decreases as the electronic states start to overlap, for example, in the $[4\,\bar{1}\,\bar{1}]$ direction [95].

Discrete tomography exploits prior knowledge and quantitative HAADF-STEM imaging; however deviation from the lattice model or number of atoms counted along a particular direction can result in errors in the reconstruction — for a 10 % atom counting error, at most 72 % of reconstructions can be considered perfect [102]. This limits the size of nanoparticles that is applicable to discrete tomography, which is around 2 nm–3 nm in diameter. In some cases, there could be one or more solutions available from the experimental projections [102]. These are the main concerns of discrete tomography.

1.3.2 Equally sloped tomography

Equally sloped tomography (EST) was proposed in 2005 and is an iterative algorithm between real and Fourier space [109]. The iterative algorithm uses pseudo-polar fast Fourier transform (PPFFT) and oversampling method to reconstruct the object.

Projections can be assembled in Fourier space in polar coordinates; however the object is in Cartesian coordinates and interpolation is needed to recover the object (see Figure 1.8). PPFFT is the polar-coordinate analogue of fast Fourier transform (FFT) in Cartesian coordinate. It allows efficient computation of discrete Fourier transform in polar coordinates, comparable to FFT. No interpolation to Cartesian coordinates is required. In PPFFT, it constructs a pseudo-polar grid and the Fourier transform at every grid point is evaluated. Pseudo-polar grid consists of concentric-squares grid (see Figure 1.16) and fractional Fourier transform can be used to obtain such grid [110]. For the measured projections to be superimposed on the pseudo-polar grid, the projections should be spaced equally in slope, rather than in angle [111]. This is the reason this algorithm is called ‘equally sloped’.

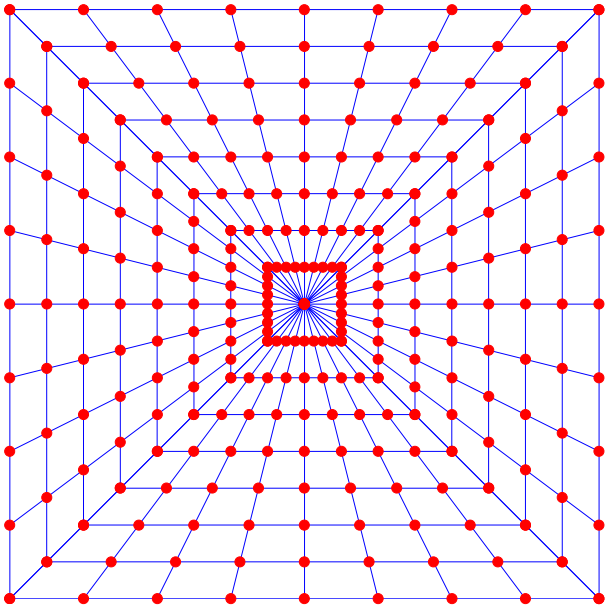


Figure 1.16: The pseudo-polar grid consists of concentric-squares grid. The grid points are equally spaced in the radial direction for each ray, and the rays are equally sloped in the angular direction. The figure is adapted from [111].

There is still the missing wedge problem in EST and some information in the Fourier space is then permanently lost. By oversampling in Fourier space, the missing information can be partially recovered as each grid point in Fourier space is correlated [109]. This is originated from the oversampling phase method which is used to retrieve phase information from diffraction patterns [112–115]. Oversampling a diffraction pattern is equivalent to padding zeros around the actual object, and this creates a *support* of the object.

The iterative approach in EST can be traced back to the Gerchberg-Saxton algorithm in 1972 for retrieving the phase of image wave functions from known diffraction patterns [116]. This is also known as the *error-reduction* algorithm [117]. At the kth iteration, an estimation of the object, $g_k(x)$, is Fourier transformed; the Fourier transform G' is made to satisfy the constraints from measured data; inverse Fourier transform is applied to give an image $g'_k(x)$; the new estimation of the object $g_{k+1}(x)$ is made to satisfy the constraints in real

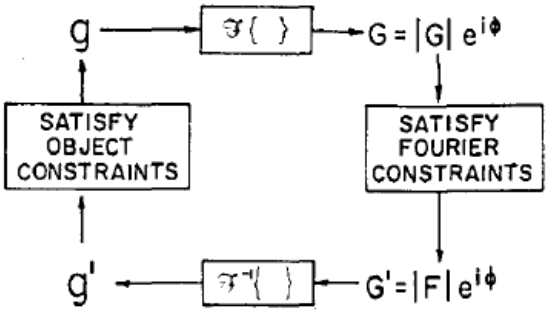


Figure 1.17: The block diagram of the error-reduction algorithm, showing the iteration between real and Fourier space and enforcement of real- and Fourier-space constraints. The figure is adapted from [117].

space:

$$g_{k+1}(x) = \begin{cases} g'_k(x) & x \notin \gamma \\ 0 & x \in \gamma \end{cases} \tag{1.34}$$

γ is the region where the point x violates the constraints: the intensity of the image at x is not positive, or x is outside the support of the object. Figure 1.17 shows a block diagram of the algorithm. A modified algorithm was also suggested and this is known as the *input-output* algorithm [117]. This has a faster convergence and EST employs this approach. The only difference of the input-output algorithm is the real-space constraint:

$$g_{k+1}(x) = \begin{cases} g'_k(x) & x \notin \gamma \\ g_k(x) - \beta g'_k(x) & x \in \gamma \end{cases} \tag{1.35}$$

β is a constant and fast convergence is achieved with β between 0.5 and 1. It controls the rate of pushing the intensities of invalid points to zero. In EST, the Fourier-space constraints are always the Fourier transform of the measured projections with equally-sloped separation.

In 2012, EST was used to identify individual atoms of a 10 nm gold nanoparticle without any prior knowledge [11], unlike discrete tomography. In 2013, EST

was used to study a multiply-twinned platinum nanoparticle and its dislocations with atomic resolution [96]. In 2015, the top section of a tungsten tip was reconstructed by EST without any missing wedge with atomic resolution [12]. Besides the application in AET, EST was used to reduce the radiation damage for biological samples and optimise the process of conventional ET [118–121].

1.3.3 Generalised Fourier iterative reconstruction

Generalised Fourier iterative reconstruction (GENFIRE) can be regarded as the 'generalised' version of EST [122]. It matches the benefits of EST, but it does not require equally-sloped projections. However, without using the pseudo-polar grid, interpolation is required during iterations between real and Fourier space and this process is known as regridding (see Central slice theorem). Discrete values of the Fourier transform in polar coordinates are interpolated to Cartesian coordinates, and there are numerous methods available [123]. They can be roughly divided into two types: global and local. Global methods make use of all the data points for interpolation, while local methods consider only a subset of data points. GENFIRE uses the local method which sets a threshold distance to limit the data points considered. The method is a modification of the inverse-distance weighted interpolation method, proposed in 1968 [124].

In Figure 1.18, interpolation is required to determine the Fourier transform of the grid point $\vec{P}$ from two projections p_1 and p_2. The perpendicular distances between the projections and $\vec{P}$ are denoted by D_1 and D_2, with their coordinates (u_1, v_1) and (u_2, v_2) respectively. Since (u_1, v_1) and (u_2, v_2) are not necessarily integers, discrete Fourier transform with oversampling is used to compute their Fourier transforms [122], and their values are denoted by f_1 and f_2 respectively. Suppose D_1 and D_2 are within the set threshold, in other words, close enough,

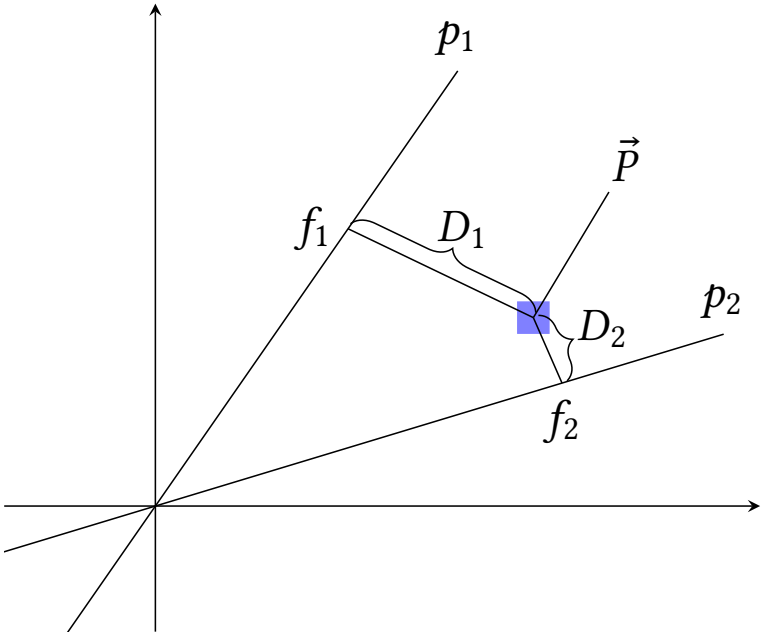


Figure 1.18: To find the Fourier transform at grid point $\vec{P}$ from two projections p_1 and p_2, interpolation is required. D_1 and D_2 are the perpendicular distances from $\vec{P}$ to the projections, with coordinates (u_1, v_1) and (u_2, v_2). The discrete Fourier transforms at (u_1, v_1) and (u_2, v_2) are f_1 and f_2 respectively.

the Fourier transform at $\vec{P}$ is

$$F(\vec{P}) = \begin{cases} \frac{\frac{f_1}{D_1} + \frac{f_2}{D_2}}{\frac{1}{D_1} + \frac{1}{D_2}} & D_1, D_2 \neq 0 \\ f_1 & D_1 = 0 \\ f_2 & D_2 = 0 \end{cases} \tag{1.36}$$

The recommended values of the threshold distance and oversampling ratio are 0.5 and 3, and the general formula of this interpolation can be found in [122]. Once the Fourier grid is obtained, the iteration steps are similar to those of EST; the only difference being that the error-reduction algorithm in Equation (1.34) is used instead of the input-output algorithm in EST.

GENFIRE introduces two optional optimisations: resolution extension/suppression and angular refinement. Resolution extension/suppression gradually changes the way the Fourier-space constraint is enforced to cope with higher

noise levels in higher spatial frequencies, and the reconstruction is smoother (see the supplementary information in [122]). Angular refinement tries to correct the deviation of experimental projection angles due to experimental errors by comparing the cross-correlation of re-projections with slight variations of the experimental angles [122].

GENFIRE is a relatively new algorithm which was reported in 2017 in the context of determining the structure and the elemental composition of an FePt nanoparticle with atomic precision [100]. Since then, the dynamics of nucleation of an FePt nanoparticle was followed by AET with 26 pm precision using GENFIRE [125]. Very recently, GENFIRE was used to locate three-dimensional crystal defects in a Re-doped MoS_2 monolayer with a precision of 4 pm [126]. GENFIRE was also used in conventional ET as reported in [127, 128].

1.4 Regularisation in regression

Multiple linear regression is frequently used to describe a target or an observation, $\boldsymbol{y} \in \mathbb{R}^n$, from a weighted sum of features, $\boldsymbol{x} \in \mathbb{R}^p$. An instance in the observation, y, can be predicted by $\hat{y}$ using the weight $\boldsymbol{\beta} \in \mathbb{R}^p$, which is written as:

$$\hat{y} = \beta_0 + x_1\beta_1 + x_2\beta_2 + \ldots + x_p\beta_p \tag{1.37}$$

β_0 is called the intercept and it is not associated with a feature. β_i is the weight for the feature x_i. The ordinary least squares method minimises the residual sum between the actual observation y and the prediction $\hat{y}$ to determine the optimal weights:

$$\min \|y - \hat{y}\|_2^2 = \min_{\boldsymbol{\beta}\in\mathbb{R}^p} \|y - (\beta_0 + \boldsymbol{x}^\mathsf{T}\boldsymbol{\beta})\|_2^2 \tag{1.38}$$

Ordinary least squares provides a reliable estimation of the weights as long as the observation truly comes from the provided set of features. If the features are correlated with each other (multicollinearity), the estimation of the weights are not reliable as they become indeterminable [129]. Ordinary least squares assumes the features are independent from each other and does not select features if correlations are present among them — it does not set some weights to zero. Feature selection reduces the complexity of the model and provides an easier interpretation. Feature selection is central to the proposed new reconstruction algorithm in this work, atomic site tomography (AST), and this is discussed in the following section.

1.4.1 Lasso

Feature selection can be achieved by penalising the residual sum by the ℓ_1-norm of the weights, and this is called ℓ_1-*regularisation*:

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p}\{\|y-(\beta_0+\boldsymbol{x}^\mathsf{T}\boldsymbol{\beta})\|_2^2+\alpha\,\|\boldsymbol{\beta}\|_1\} \tag{1.39}$$

α controls the strength of the regularisation. The larger the value of α is, the more weights become zero. If the weight is zero, it is equivalent to the corresponding feature having been eliminated from the model. The optimal regularisation strength is usually found by cross-validation. Cross-validation selects the best α from a set of α by comparing the mean square errors between the observation and their corresponding predictions.

ℓ_1-regularisation is also called least absolute shrinkage and selection operator (lasso) since it has been proposed in 1996 [130]. *Shrinkage* indicates the weights are smaller compared with the estimates by ordinary least squares, while *selection* sets some weights to zero (both controlled by α). There are several algorithms to find the solution of lasso, such as least angle regression (LARS) [131]

and homotopy [132]. The algorithm adapted in our work is the pathwise coordinate descent [133] and it is discussed below.

For an observation $\boldsymbol{y}$ having N instances, the lasso problem can be written as

$$\min R = \min_{\boldsymbol{\beta} \in \mathbb{R}^p} \left[\frac{1}{2N} \sum_{i=1}^{N} \left\| y_i - \left(\beta_0 + \boldsymbol{x_i}^\mathsf{T} \boldsymbol{\beta} \right) \right\|_2^2 + \alpha \left\| \boldsymbol{\beta} \right\|_1 \right] \tag{1.40}$$

The pathwise coordinate descent method partially optimises an individual weight, β_j, and in each iteration it cycles through all the weights. The gradient of Equation (1.40) with respect to β_j is

$$\begin{aligned} \frac{\partial R}{\partial \beta_j} &= \frac{\partial}{\partial \beta_j} \left[\frac{1}{2N} \sum_{i=1}^{N} \left\| y_i - \left(\beta_0 + \boldsymbol{x_i}^\mathsf{T} \boldsymbol{\beta} \right) \right\|_2^2 + \alpha \left\| \boldsymbol{\beta} \right\|_1 \right] \\ &= \frac{1}{N} \sum_{i=1}^{N} \left[y_i - \left(\beta_0 + \boldsymbol{x_i}^\mathsf{T} \boldsymbol{\beta} \right) \right] \frac{\partial}{\partial \beta_j} \left[y_i - \left(\beta_0 + \boldsymbol{x_i}^\mathsf{T} \boldsymbol{\beta} \right) \right] + \alpha \\ &= -\frac{1}{N} \sum_{i=1}^{N} x_{ij} \left[y_i - \left(\beta_0 + \boldsymbol{x_i}^\mathsf{T} \boldsymbol{\beta} \right) \right] + \alpha \end{aligned}$$

Suppose $\widetilde{\boldsymbol{\beta}}$ is the set of weights without the contribution of the j-th feature ($\beta_j = 0$), the gradient can be written as

$$\begin{aligned} \left. \frac{\partial R}{\partial \beta_j} \right|_{\boldsymbol{\beta} = \widetilde{\boldsymbol{\beta}}} &= -\frac{1}{N} \sum_{i=1}^{N} x_{ij} \left[y_i - \left(\widetilde{\beta_0} + \boldsymbol{x_i}^\mathsf{T} \widetilde{\boldsymbol{\beta}} \right) \right] + \alpha \\ &= -\frac{1}{N} \sum_{i=1}^{N} x_{ij} \left(y_i - \widetilde{y}_i^{(j)} \right) + \alpha \end{aligned} \tag{1.41}$$

$\widetilde{y}_i^{(j)}$ is the partial residual of β_j; in other words, it is the residual without the contribution from the j-th feature, and can be written explicitly as

$$\widetilde{y}_i^{(j)} = \widetilde{\beta_0} + \sum_{l \neq j}^{N} x_{il} \widetilde{\beta_l} \tag{1.42}$$

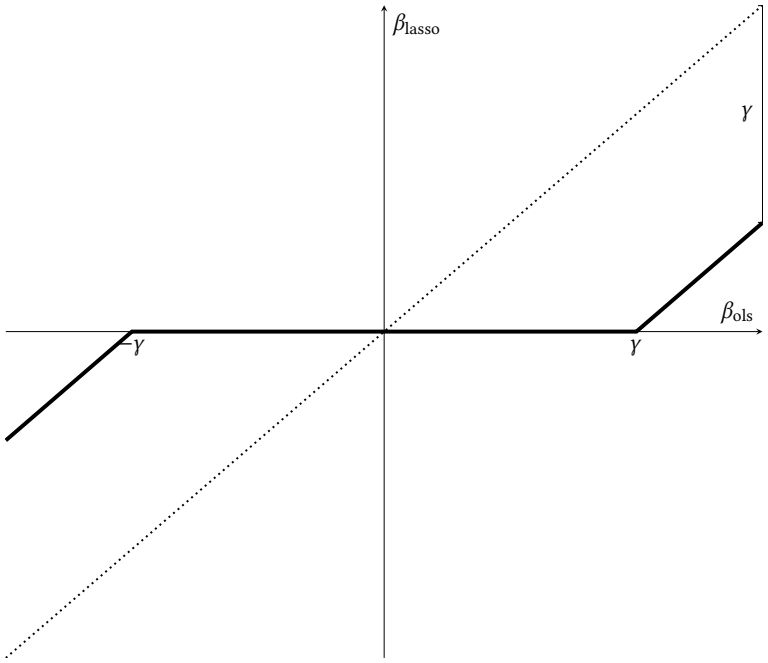


Figure 1.19: The soft-thresholding operator for the weight update of lasso is shown here in solid line. The ordinary least squares weight is shown in dotted line for comparison. The lasso weight is shrunken, by γ, and some of them are set to zero, as controlled by γ.

The update to the weight β_j is based on the soft-thresholding operator $S(z, \gamma)$ and its gradient [134]. The soft-thresholding operator is

$$S(z,\gamma) = \text{sign}(z)(|z| - \gamma) = \begin{cases} z - \gamma & z > 0 \text{ and } |z| > \gamma \\ z + \gamma & z < 0 \text{ and } |z| > \gamma \\ 0 & |z| \leq \gamma \end{cases} \tag{1.43}$$

The update to β_j after applying the soft-thresholding operator is

$$\beta_j \leftarrow S\left(\frac{1}{N}\sum_{i=1}^{N} x_{ij}\left(y_i - \widetilde{y}_i^{(j)}\right), \alpha\right) \tag{1.44}$$

From Equations (1.43) and (1.44), it can be clearly seen that the weights are shrunken and some of them are set to zero, depending on α. A comparison of the weights between ordinary least squares and lasso is shown in Figure 1.19. The uniqueness of lasso solution is discussed in [135]. The package *scikit-*

learn [136] in Python is used in this work to implement the lasso fitting and its implementation details can be found in [137].

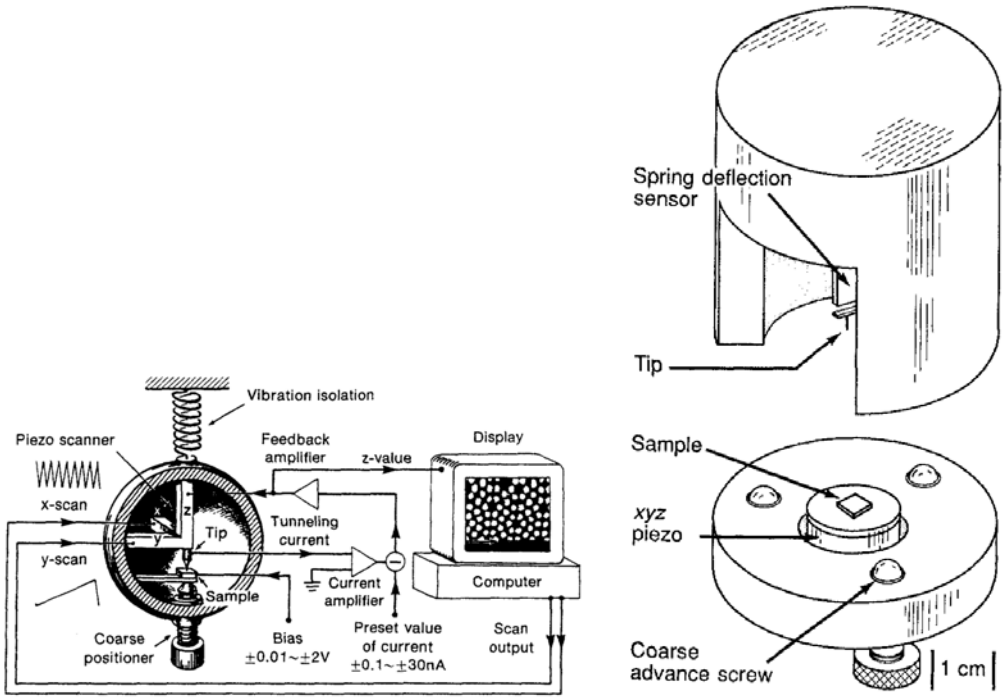
1.5 Scanning probe microscopy

Scanning probe microscope (SPM) allows probing features on surfaces with atomic resolution. The main application of SPM in tungsten tip tomography is for its sample preparation — by using techniques in SPM, a nanoparticle is intended to be picked up by an etched tungsten tip. Two major types of SPM are scanning tunnelling microscopy (STM) and atomic force microscopy (AFM), and they are discussed in the following sections.

1.5.1 STM

The first atomic resolution of surface topography by scanning tunnelling microscopy (STM) was reported by Binnig, Rohrer, Gerber and Weibel in 1982 using gold and $CaIrSn_4$ [138]. In 1986, Binnig and Rohrer were awarded the Nobel Prize in Physics 'for their design of the scanning tunnelling microscope'. STM consists of a conductive tip attached to a piezoelectric scanner. Upon applying voltages to the piezoelectric scanner, it can expand or contract vertically (the z direction) and perform a raster scan horizontally (the x and y direction). A schematic diagram of an STM is shown in Figure 1.20a.

The conductive tip is usually an electrochemically etched tungsten tip, or a mechanically cut Pt–Ir tip for ordinary application, although other metal tips can be used. The tip is brought very closely to a conductive substrate, of the order of a few Å. Quantum tunnelling occurs when a bias voltage is applied across the tip and the substrate. The electrons are able to tunnel across the tip-substrate gap because of the overlap of their electron wave functions when



(a) STM (b) AFM

Figure 1.20: Schematic diagrams of STM and AFM are shown in (a) and (b) respectively. The schematic diagrams are adapted from [139].

they are in close proximity. A theoretical treatment of the tunnelling current can be found in [140].

The tunnelling current is very small, of the order of nA; therefore, a transimpedance amplifier (current-to-voltage converter) is used to amplify the current to a voltage signal. This voltage is compared to a reference value (the setpoint) and a negative feedback acts on the z piezoelectric component — if the voltage from the amplifier is larger than the setpoint, the z piezoelectric component contracts to increase the distance between the tip and the substrate, which reduces the tunnelling current, and vice versa. The tunnelling current is maintained at the setpoint by enabling the feedback control during a raster scan in x and y direction, and the variation of the z piezoelectric voltage produces a surface topographical image. A comprehensive treatment of STM can be found in [139].

1.5.2 AFM

STM provides a simple but a direct method to probe any conductive surface in real space. However, it is limited to conductive substrates. AFM was first reported by Binnig, Quate and Gerber in 1986 [141], 4 years after the invention of STM. It is a very similar technique to STM: they both have a tip attached to a piezoelectric scanner for raster scan and z positioning, and feedback control.

The main difference of AFM from STM is the feedback signal. AFM uses a pyramidal tip attached on a cantilever, and the tip is brought very close to the surface. A schematic diagram of an AFM is shown in Figure 1.20b. There are different operating modes of AFM and they depend on the force regime between the tip and the substrate. The tip is at the attractive regime in non-contact mode, and at the repulsive regime in contact mode. The tapping mode is in the middle between non-contact and contact mode: the tip touches the surface and maintains in the attractive regime.

The cantilever deflects because of the variation of the force between the tip and the surface. The deflection of the cantilever is detected by a photo-diode detector and uses as the feedback signal to the system, instead of the tunnelling current in STM. By maintaining a constant deflection, a topographical image can be obtained at the constant force mode. There are many variants of AFM such as electric force microscopy, magnetic force microscopy and piezoresponse force microscopy. A comprehensive review of AFM can be found in [142].

The tuning fork AFM in Chapter 6 is in non-contact mode as the interaction between an etched tungsten tip and the substrate should be minimised. This can ensure potential damage on the tungsten tip is minimised. The frequency

change of the tuning fork is the feedback signal, which is similar to the change of deflection of the cantilever in traditional AFM.

2

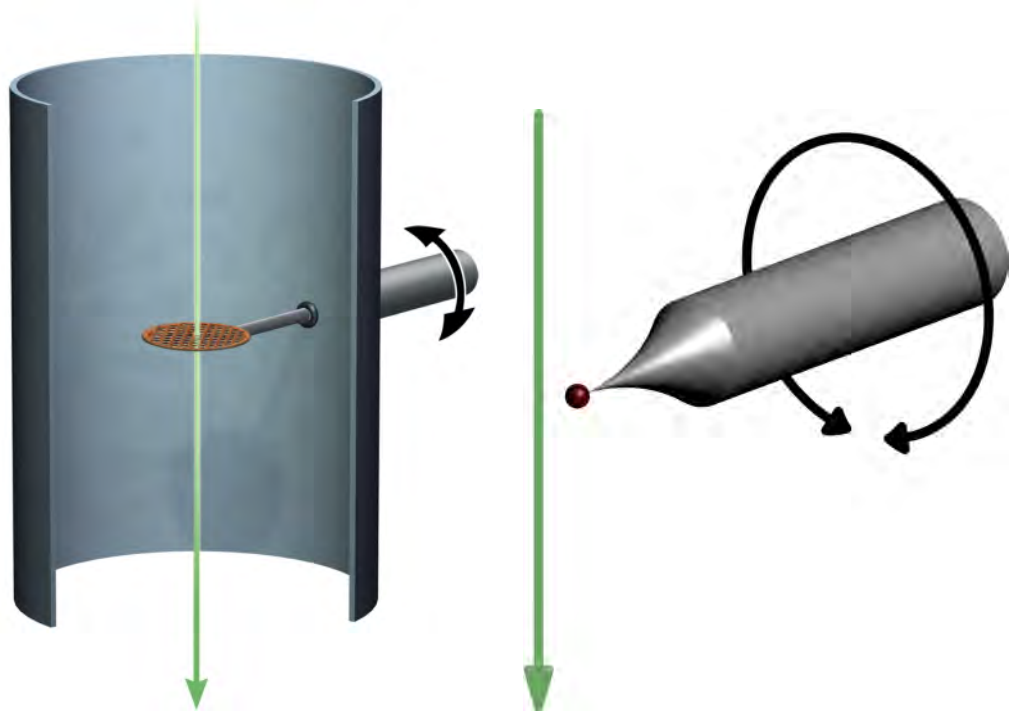
Tungsten tip tomography

Samples for electron tomography (ET) in physical science are often nanoparticles. In a typical tomography experiment, a tiny drop of colloidal solution is drop-cast on a transmission electron microscopy (TEM) grid and the grid is normally placed in a single-tilt sample holder to obtain projections from different tilt angles (see Figure 2.1a); alternatively, other slab-like substrates such as Quantifoil or carbon replica can also be used [143, 144]. This is a simple, convenient and flexible method; however, there are several drawbacks. There is a limit of the tilt angles due to mechanical limit of the stage and shadowing, so the maximum tilt angular range is usually around ±70°. This results in a missing wedge and some information from the sample is permanently lost. Although some reconstruction algorithms like equally sloped tomography (EST) and generalised Fourier iterative reconstruction (GENFIRE) attempt to recover those information, the recovered information is only partial and it depends on how the oversampling ratio is set (see Section 1.3.2). The supporting material of the TEM grid, although it is a very thin layer (3 nm–5 nm) of amorphous carbon film, will impact on the signal-to-noise ratio of the image. The impact becomes significant when the size of the nanoparticle is below 10 nm, which is often the case in atomic electron tomography (AET).

Instead of putting the nanoparticles on a flat surface, a needle-shaped platform is desired for full-range tomography. This resolves the long-standing problem of the missing wedge in tomography experiments. Needle-shaped samples can be prepared by focused ion beam (FIB) [145–148]; however, nanoparticles cannot be shaped by FIB and for some materials, it is challenging to prevent induced damage from the ion beam. Recently, powder specimens were attached to carbon nanofibres that extend beyond the end of a tungsten needle for full-range tomography [149]. This is a simple and convenient sample preparation method, but it is difficult to control the orientation of carbon nanofibres with respect to the longitudinal axis of the tungsten needle; the supporting carbon nanofibres also impact on the signal-to-noise ratio, which is similar to using a TEM grid. The method involves dipping the tungsten needle into a colloidal solution in which the nanoparticles tend to aggregate, so it is challenging to perform analysis of an individual nanoparticle.

Tungsten tip tomography uses a processed tungsten tip as a platform for ET (see Figure 2.1b). An individual nanoparticle is located on the apex of the tungsten tip and imaging can be performed over the full angular range (±90°) by rotating the tungsten tip around its axis. The nanoparticle on the apex extends out from the tungsten tip along the tomography axis, so the signal does not mix with the amorphous carbon film on the TEM grid or carbon nanofibres, which cannot be reconstructed. Different crystallographic planes of tungsten are used to accurately determine the tilt angles, reducing the errors from sample-stage mechanics and software.

The idea of tungsten tip tomography was first proposed in 2013 [151] with some initial tests [152]. In 2015, ET on a bare tungsten tip was performed and positions of individual tungsten atoms could be identified with atomic precision [12]. By growing the nanoparticle on the apex of a bare tungsten



(a) (b)

Figure 2.1: Two configurations in electron tomography are shown here. (a) when the sample is on a slab-like substrate such as a conventional TEM grid, there will be a missing wedge due to the mechanical limits of the stage rotation and shadowing, and the amorphous carbon film will impact on the signal-to-noise ratio. The figure is adapted from [150]. (b) when the sample is located on the apex of a tungsten tip, ±90° tilt angular range can be realised and there will be no degradation of signal as the sample is free-standing.

tip, the structure of the nanoparticle can also be determined. This chapter outlines the general preparation and experimental procedures for tungsten tip tomography.

2.1 Tip processing

To perform ET on the apex of a tip, the curvature of the apex needs to be very sharp. Tungsten is chosen as the material for tip tomography because procedures to produce sharp tungsten tips are well established in the field of scanning tunnelling microscopy (STM). The tip should also be stable under the conditions for high-resolution electron microscopy; otherwise diffused atoms

will contaminate the sample of interest. Atoms at the surface are more vulnerable to the incident electron beam and are easier to be displaced. If the energy transferred elastically from the electron beam to the nuclei exceeds the displacement energy, electron-beam sputtering or knock-on damage occurs; however, surface atoms could still be sputtered by incident electrons with an energy lower than the knock-on threshold under prolonged exposure [153]. For tungsten, the knock-on threshold for the electron beam is above 500 keV [154, 155]. The accelerating voltage in high-resolution electron microscopy is typically not higher than 300 kV; therefore, a tungsten tip can be considered as a robust platform for ET. Although higher accelerating voltages could attain higher resolutions, they would induce severe beam damage on the samples, so they are not routinely used.

To prepare tungsten tips suitable for ET, the tungsten wire requires some processing to obtain large crystalline domains and etching to give the correct geometry. They will be outlined below.

2.1.1 Recrystallisation

Commercial tungsten wire with a diameter of 250 µm (99.95 % purity, Thermo Shield) is used as the starting material to obtain the tungsten tip for tomography. Wire drawing at room temperature (cold drawing) is commonly used to manufacture thin metal wires and strain energy created by deformation is stored in various types of defects during the process. Dislocations and grain boundaries are produced during wire drawing and the average grain size in the as-received tungsten wire is usually of the order of µm [156]. In tungsten tip tomography, the crystallographic planes of tungsten are used to calibrate the stage for the accurate determination of tilt angles and it is essential to maintain

a homogeneous grain for this purpose. This can be achieved by increasing the grain size by recrystallisation.

Since the internal energy of as-drawn tungsten wire is higher due to the presence of defects, there is a tendency to reduce the defect density so that the system is in a lower energy state. Recrystallisation is the process of reducing the defect density by annealing, which promotes the formation of large defect-free grains. During annealing, the dislocations will rearrange and the area of grain boundaries will be reduced. These contribute to the growth of large defect-free grains. The same material with a larger grain size has less strength, and this inverse relationship between grain size and strength is known as the Hall–Petch relation [157, 158]. The dislocations need more energy to propagate through grain boundaries and if the areas of grain boundaries are reduced, less energy is needed for the movements of dislocations and hence less energy is required to deform the material. This results in a much more easily broken wire after recrystallisation and requires careful handling.

To obtain tungsten wire with large grains, commercial tungsten wire is cut to a length of 5 cm and annealed under tension. The tungsten wire is held tightly by screwed metal plates and compressed springs provide the tension. The wire is resistively heated by a ramping current starting from 8.5 A with a ramp rate of $0.4\,\mathrm{A\,min^{-1}}$ in vacuum ($< 10^{-6}$ mbar). The recrystallisation temperature of pure tungsten is around 1200 °C [159]. While the current is increasing, the tungsten wire becomes softer due to the formation of large grains. The force from the compressed springs will eventually pull the wire apart, where the wire is fractured in a ductile manner [160, 161]. Typically, the final current before breaking the wire is approximately 10 A in the current system. A schematic illustration of this process is shown in Figure 2.2. Two pieces of recrystallised tungsten wire are obtained and the ends can be used for subsequent etching.

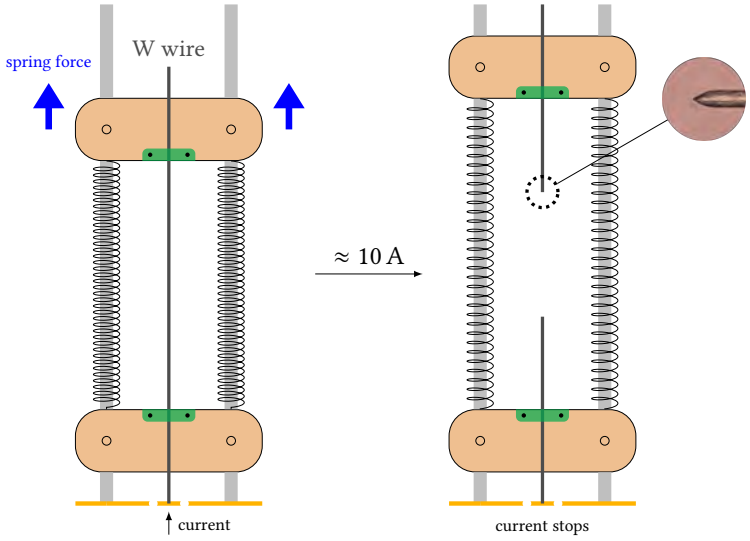
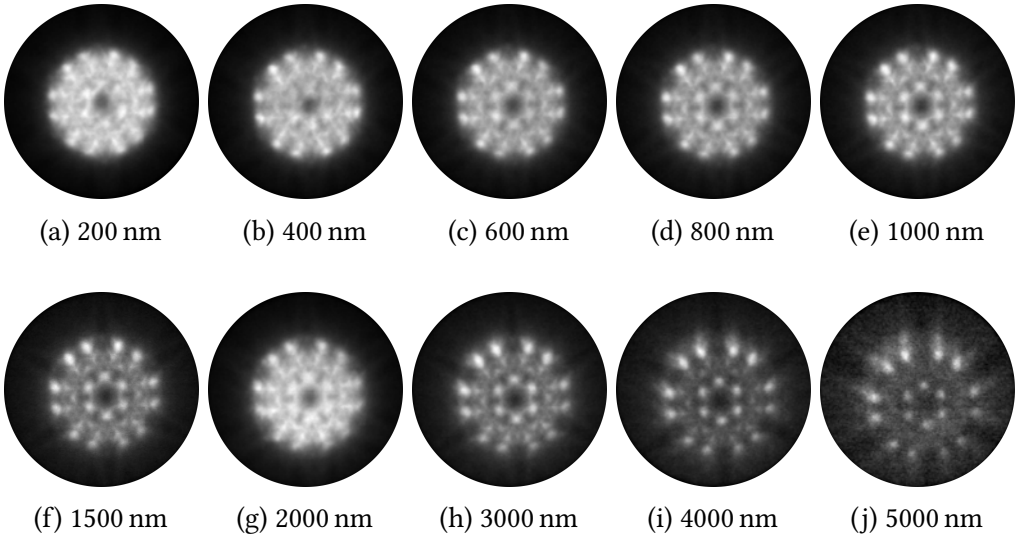


Figure 2.2: A schematic diagram of the recrystallisation of tungsten wire is shown here. The upward spring force will keep pulling the tungsten wire apart until it breaks as a result of the formation of large grains, which will make the tungsten wire softer. Ceramic split bushes are used to prevent current flowing through the rods. The inset shows an optical microscopic image of the end of the recrystallised tungsten wire.

The colour of the recrystallised tungsten wire is shinier due to the removal of tungsten oxides.

The resulting grain size could be as large as 4 mm [156], which is 5–6 orders of magnitude greater than the scale of the apex's curvature. Figure 2.3 shows convergent-beam electron diffraction (CBED) patterns along a recrystallised tungsten wire at the (1 1 1) zone axis, and it shows the crystallographic structure of the grain is homogeneous. This shows recrystallisation successfully increased the grain size.

The deformation of tungsten wire during cold drawing orients the [1 1 0] axis to the wire axis, a typical characteristics for a body-centred cubic (bcc) material. This is considered to be the result of double slip with respect to the [1 1 0] axis [162–164]. This texture is retained after recrystallisation of pure tungsten wire [156, 159, 165]. This is important as the [1 1 0] direction of the tungsten tip is the tilt axis in tomography experiments.



(a) 200 nm (b) 400 nm (c) 600 nm (d) 800 nm (e) 1000 nm

(f) 1500 nm (g) 2000 nm (h) 3000 nm (i) 4000 nm (j) 5000 nm

Figure 2.3: CBED patterns of (1 1 1) are taken at different distances below the apex of a recrystallised tungsten tip, from 200 nm–5000 nm in (a)–(j). The identical nature of the CBED patterns indicates the top 5 µm of the recrystallised tip is a homogeneous single grain. The gradual shift of the centre of CBED patterns is due to larger beam tilt when it is moving down the tip. The convergence angle is 24.96 mrad and the radius of the cut-off area is 45.1 mrad.

2.1.2 Electrochemical etching

After recrystallisation, electrochemical etching is used to produce a sharp tip with a suitable geometry. Recrystallised tungsten wire has a homogeneous crystal orientation and upon etching, it gives a smooth surface; as-drawn tungsten wire is polycrystalline with many defects, so the rate of etching is different in different crystal planes, giving a rough surface with multiple protrusions [156]. Recrystallisation ensures the apex of etched tungsten tip is smooth without tiny protrusions interfering the sample on the apex.

To produce sharp tungsten tips, electrochemical etching with an alkaline solution, usually NaOH or KOH, is used. The tungsten tip acts as the anode, while an inert metal acts as the cathode. A voltage is applied and the following redox

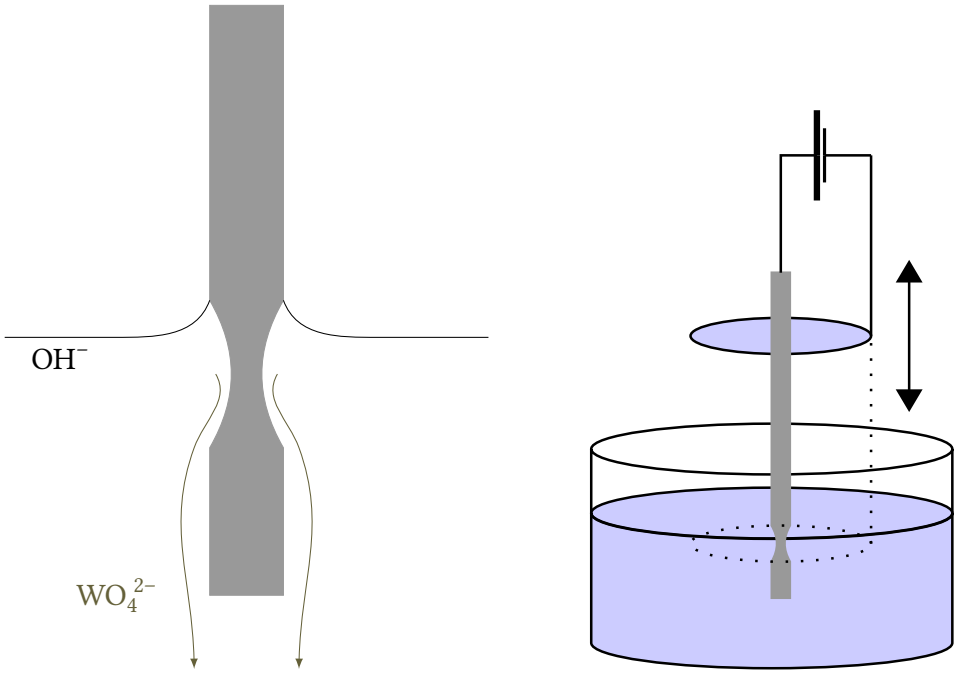
reaction takes place [166]:

$$\begin{aligned}
&\text{anode:} && \mathrm{W(s) + 8\,OH^-(aq) \longrightarrow WO_4^{2-}(aq) + 4\,H_2O(l) + 6\,e^-} \\
&\text{cathode:} && \mathrm{6\,H_2O(l) + 6\,e^- \longrightarrow 3\,H_2(g) + 6\,OH^-(aq)} \\
&\text{---} && \text{---} \\
& && \mathrm{W(s) + 2\,OH^-(aq) + 2\,H_2O(l) \longrightarrow WO_4^{2-}(aq) + 3\,H_2(g)}
\end{aligned}$$

The upper part of the meniscus has a lower concentration of OH^- as the interfacial area between the tip and the solution is smaller, and this results in a lower etching rate. The lower end of the tungsten wire is protected by a thin layer of WO_4^{2-} sinking down from above, and this also lowers the etching rate (see Figure 2.4a) [167]. Therefore, the etching rate is the highest between the meniscus and the lower end of the wire, and a 'notch' is formed until the lower end drops off.

The etched tungsten tip requires the correct geometry to fit into a 'puck' (see Figure 2.5), so the etching is not simply to produce a sharp tip like in STM. A 5 M NaOH solution is used to create a notch with ≈100 µm in diameter, while also thinning the area around the notch to ≈230 µm in diameter. The thinning is required for the ease of fitting the etched tungsten tip into the hole of the puck which is 250 µm in diameter. The cathode loop is lifted 1.6 cm above from this notch, which will become the final apex, to create another notch (see Figure 2.4b), allowing gentle separation manually. This avoids the use of wire cutters which introduces mechanical stress on the recrystallised wire that may propagate to the area near the apex.

After gentle washing with running tap water, the wire is further etched with 1 M NaOH solution until drop-off to obtain the final etched tip. The geometry of the etched tip is shown in Figure 2.6. To remove the tungsten oxide layer,



(a) at meniscus (b) loop etching

Figure 2.4: The electrochemical etching of tungsten wire is shown here. (a) the upper part of the meniscus contains a lower concentration of OH^-, while the lower end of the wire is protected by a thin layer of WO_4^{2-} which is flowing down from above. The concentration of OH^- is thus highest between these two regions, resulting in a notch. (b) the cathode is an insert metal loop wire, and it is lowered into the solution to form the notch in (a); it is lifted to create a tiny notch on the tungsten wire so that it can break easily.

the tip is flashed at 900 °C (measured by a thermocouple) by resistive heating in vacuum ($< 10^{-6}$ mbar). Suitable tips for tomography require an apex diameter of 10 nm–20 nm.

2.1.3 Sample preparation

For metal samples, the nanoparticles can be formed by growing a thin film on the etched tungsten tip followed by annealing. This can be achieved by melting a small piece of metal wire on a rounded V-shaped tungsten wire by resistive heating. This small reservoir of metal acts as the source of evaporation when the wire is heated. Two concentric metal tubes with inner diameters 0.9 mm and 0.5 mm in a barrel connector are used to hold the etched tungsten tip,

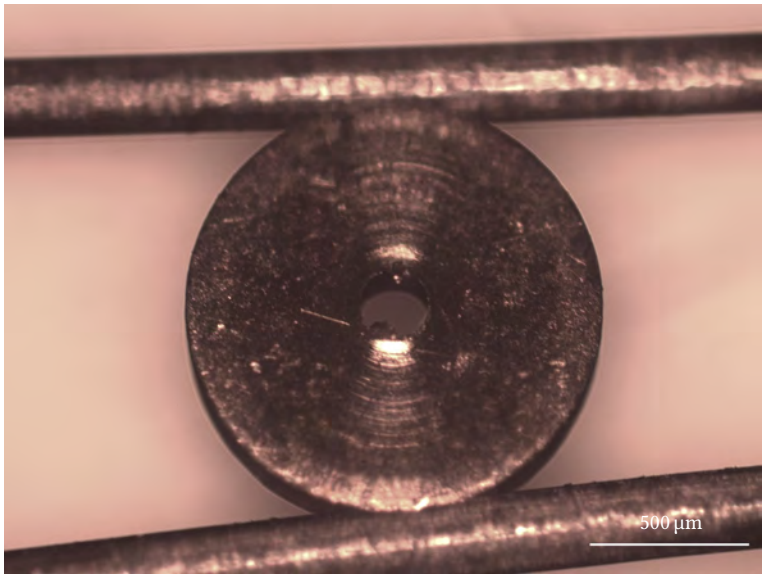
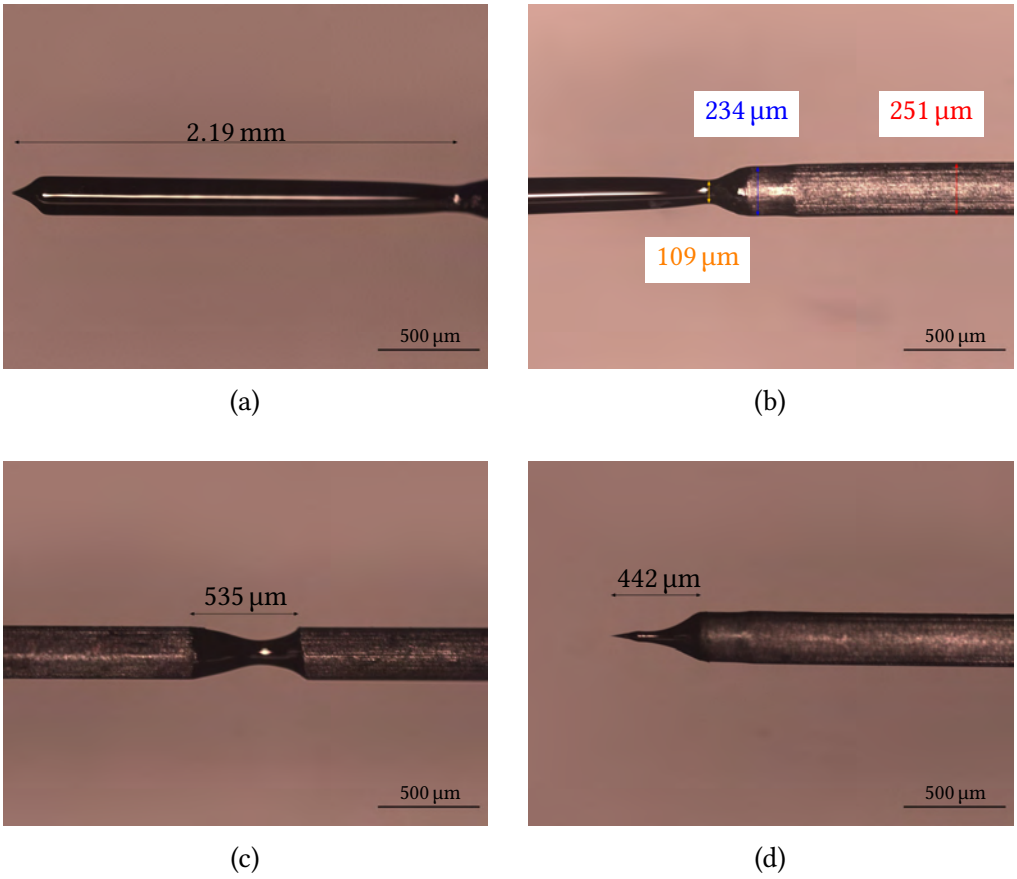


Figure 2.5: A puck is a metal disc with a hole in the middle and a groove along its perimeter. It is custom-made for the TEAM stage to hold tip-shaped samples. The puck in the image is held by two metal rods along its groove.

with the apex pointing to the metal source. The evaporation flange is shown in Figure 2.7. For gold, the rounded V-shaped tungsten wire is resistively heated by a minimum current of 4 A in vacuum ($<10^{-6}$ mbar) for deposition of a thin gold layer on the tip. The amount of current depends on the metal and several trial runs are needed to find the appropriate parameters for different metals. The etched tungsten tip is then annealed at 800 °C–900 °C for 5 s–30 s, and nanoparticles are grown on the etched tungsten tip. A quick inspection by TEM is needed to check if a nanoparticle is formed on the apex of the tip, and no other nanoparticles are too close to the nanoparticle on the apex.

After confirming the nanoparticle is on the apex, the etched tungsten tip is ready to be inserted into a puck. The insertion is done meticulously to avoid bending or breaking the tip. Once the etched tungsten tip is inserted into the puck, the back end is electrochemically etched away while a small plastic cover is used to protect the puck from the etching solution. The front (the side of apex) and the back side cannot extend further than 500 µm and 700 µm out of the puck respectively; otherwise the tip cannot be put into the TEAM 0.5



(a) (b)

(c) (d)

Figure 2.6: The geometry of etched tungsten tip at different stages observed under an optical microscope: (a) the end of tungsten wire after thinning, (b) the thinning of the tungsten wire with the notch which later becomes the apex, (c) the notch formed by lifting the loop as in Figure 2.4b for easy manual breaking and (d) the final apex of the tungsten tip.

microscope, which the gap between the pole pieces is very small. Figures 2.8a and 2.8b show two photos during and after the tip insertion.

2.2 Experimental set-up

After the etched tungsten tip is mounted on the puck, it can be loaded into the TEAM microscope for tomography. Transmission Electron Aberration-corrected Microscope (TEAM) is a collaborative effort to construct a new generation of high-resolution electron microscopes, focusing on advanced aberration-corrected optics. Two microscopes were built in the National Center for Electron

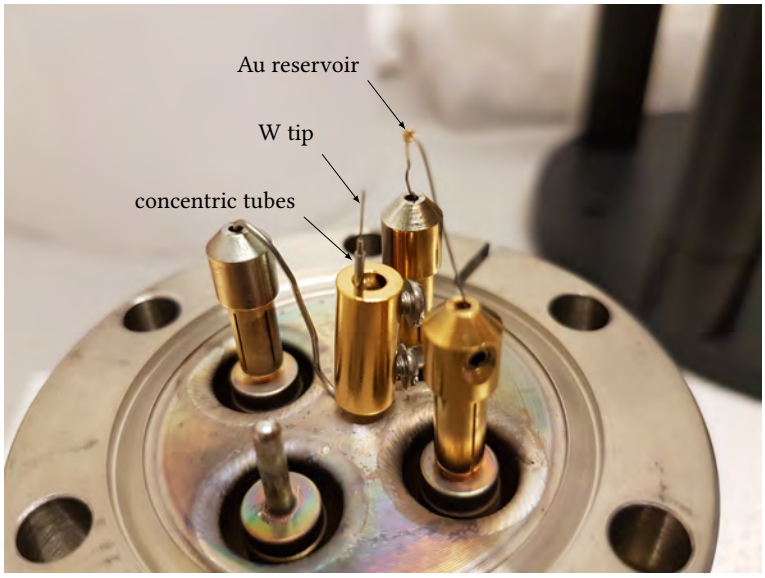


Figure 2.7: The evaporation flange for preparing gold nanoparticles on etched tungsten tip is shown here. The gold reservoir is formed by resistive heating of a small amount of gold wire. The etched tungsten tip is inserted into two concentric metal tubes which are fixed in a barrel connector. The etched tungsten tip points towards the gold reservoir, and evaporation of gold is performed by resistive heating under vacuum.

(a) (b)

Figure 2.8: Images of pucks during the insertion of an etched tungsten tip are shown here. (a) the etched tungsten tip is being inserted into the puck and (b) the etched tungsten tip is mounted on the puck.

Microscopy (NCEM) in Berkeley, California: TEAM 0.5 and TEAM I, and their specifications can be found in [168, 169]. TEAM 0.5 operated in high-angle annular dark-field (HAADF)-scanning transmission electron microscopy (STEM) mode was used for all the tomography experiments in this work. The electron probe size can be 50 pm at 300 kV, which is ideal for AET.

Interaction with TEAM 0.5 is through the TEM Imaging & Analysis (TIA) software. For tungsten tip tomography, a custom GUI[1] is used to interact with the TIA software and its settings for automated data acquisition. A server is launched on the microscope computer and the client is connected to the server remotely. For successful automation, some set-up is needed and they will be outlined below (see Figure 2.9).

2.2.1 TEAM stage

The TEAM stage is a tailor-made stage for the TEAM microscopes [170–172]. There are five axes of movement: x, y, z for translational movement, and α, γ for rotational movement (see Figure 2.10a). The movements are entirely controlled by piezoelectric motors, which allow a precision down to 14 pm. The displacements are measured by sets of capacitor plates, whose their overlaps change during movement [173]. The puck is held along its groove by a pair of Ti wires, the 'chopsticks', whose translational movement rotates the puck around its axis with a range of ±180°, the γ rotation. The puck can also be tilted ±180° around the x axis and this is the α tilting. The design of TEAM stage allows tilting of ±180° and completely eliminates the missing wedge problem for a tip sample. The only limiting factor is the shadowing of the tip by the puck and this occurs when the angle between the tip and the puck is deviated from 90° by more than 15°; in other words, the tungsten tip is not straight enough.

After loading the puck in the TEAM stage, α is tilted by ±90° such that the electron beam is approximately perpendicular to the transverse axis of the tungsten tip (see Figure 2.10b). Two photos of the TEAM stage inside TEAM 0.5 with $\alpha = 0°$ and $\alpha = 90°$ are shown in Figure 2.11. The apex of the tip is located by moving the x and y axis of the stage, and it is usually ≈400 µm from the puck. The distance moved depends on how the etched tip is put into the

[1] It is developed and maintained by Dr Wolfgang Theis.

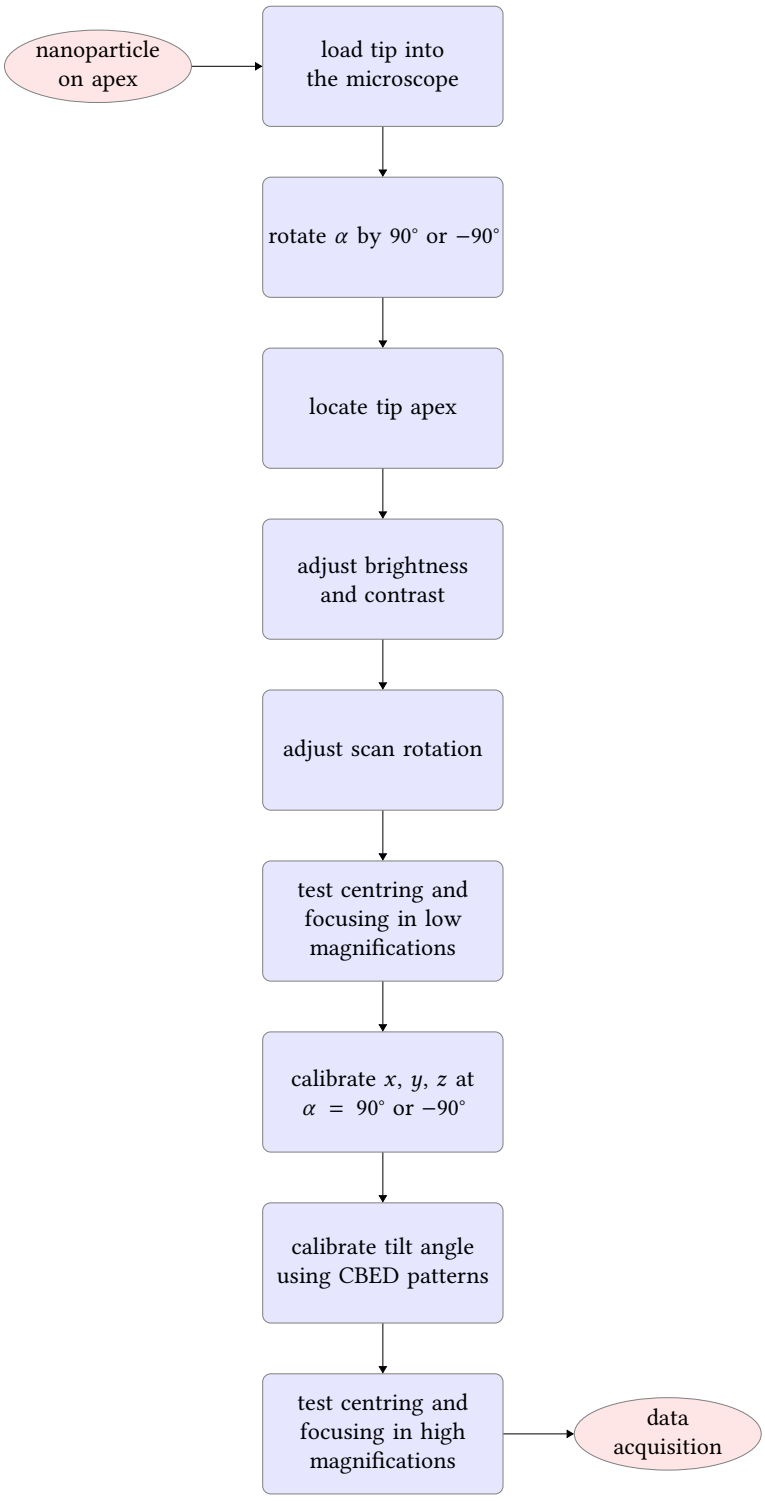
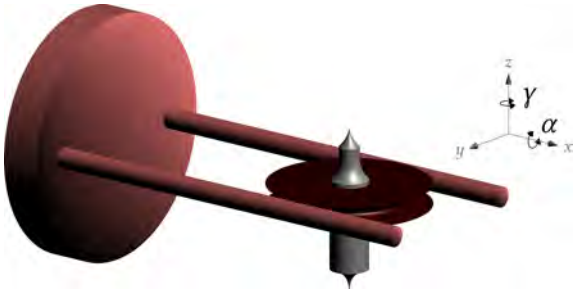
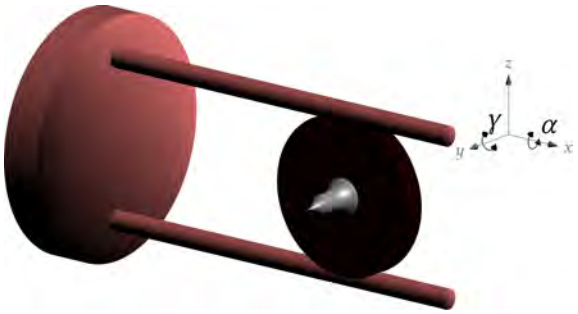


Figure 2.9: A flowchart describing the required steps before automated data acquisition is shown here. Low magnifications refer to magnifications ≤160 k× and they are usually 5 k×, 20 k×, 80 k× and 160 k×; high magnifications refer to magnifications >160 k× and they are usually 640 k×, 1.3 M×, 2.55 M×, 3.6 M× and 5.1 M×, or up to the magnification at which the projections are being recorded.

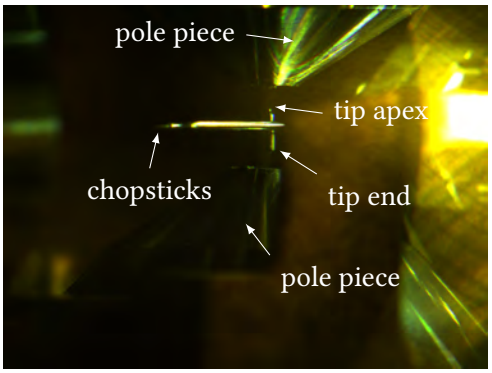


(a) $\alpha = 0°$

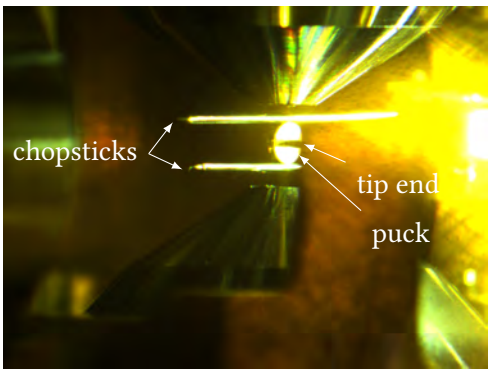


(b) $\alpha = 90°$

Figure 2.10: The schematic diagrams show the coordinate system of the TEAM stage. The puck with inserted etched tungsten tip is held by a pairs of 'chopsticks'. (a) shows the stage when $\alpha = 0°$, and this is the transfer position of the TEAM stage. (b) shows the stage when $\alpha = 90°$, and this is how tungsten tip tomography's experiments are done.



(a) $\alpha = 0°$



(b) $\alpha = 90°$

Figure 2.11: Photos of TEAM stage inside TEAM 0.5 with a puck holding an etched tungsten tip at (a) $\alpha = 0°$ and (b) $\alpha = 90°$.

puck. To confirm the stage has moved correctly, the requested movement in the client interface is compared with the measured movement from the stage and appropriate scaling factors are applied to the client interface accordingly.

Table 2.1: An example of contrast factors at different magnifications, with absolute contrast as 0.8 and brightness as 0.43 with a dwell time of 12 µs

Magnifications	Contrast factors
5 k×	0.43
80 k×	0.44
160 k×	0.46
640 k×	0.49
2.55 M×	0.50

2.2.2 Tip focusing, finding and centring

A lot of automated functionalities depend on the pixel intensities and it is desirable not to saturate the detector or have a too low intensity value. This can be adjusted by setting the brightness and contrast dependent on the magnification. The recorded images are 16-bit images, and the dark count should be ≈5000 and the brightest feature should be ≈50 000. There is a contrast factor associated with each magnification (see Table 2.1 for typical values); when the imaging area is large, the intensity count for the tip is very high due to its thickness so contrast should be smaller. Between these magnifications, the contrast factors are linearly interpolated and remain constant if the magnification is set beyond the highest setting.

Once the brightness and contrast are set, the scan rotation can be adjusted so that the tip is pointing vertically in the acquired image. This depends on the inclination between the transverse axis of the tip and the horizontal axis of the image. The inclination can be recorded using a 5 k× magnification and the scan direction is changed accordingly. It is essential to automatically focus the apex of the tip and locate the tip once it is lost in the field of view during movement. These procedures are described below.

Tip focusing. A good focus is essential for obtaining high-resolution images for AET. In order to avoid excess electron dose on the sample at the apex, the

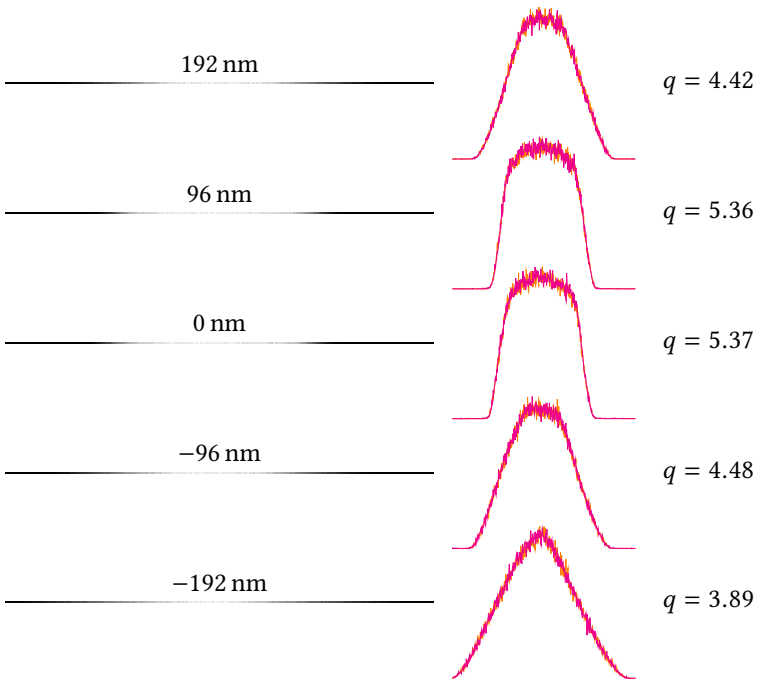
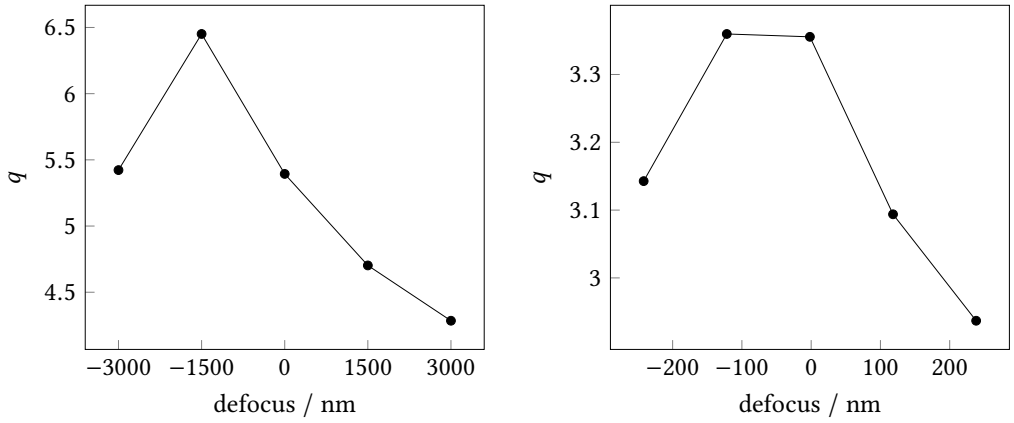


Figure 2.12: Line scans are used to determine the quality factors q with different defocus values. The left side shows the line scan obtained below the apex of the tungsten tip with different defocus values given above, and the corresponding line plots on the right. These line scans are taken at 80 k×.

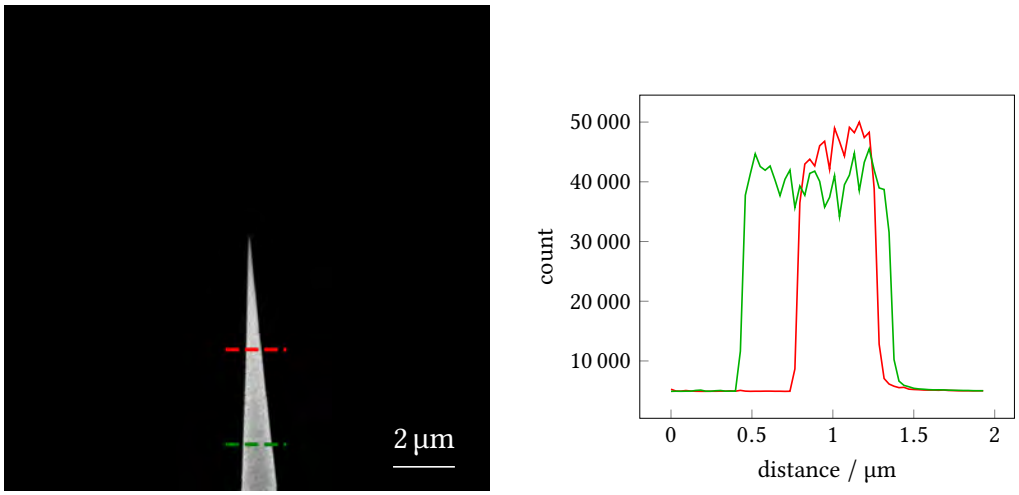
focusing is done by a line scan below the apex with a defocus series (see Figure 2.12). Different strategies are used to determine the quality of the focus: in low magnifications (≤160 k×), the line scan is fitted by a model function — a function representing the shape of the tungsten tip — convoluted with a Gaussian function with the defocus step as its width; in high magnifications (>160 k×), the maximum intensity of the peaks after fast Fourier transform (FFT) is found. These are used to calculate a quality factor q at each defocus step and the optimal focus is chosen by comparing their q.

Two modes of obtaining defocus series are implemented, the coarse and the fine mode, with larger and smaller defocusing steps respectively (see Figure 2.13). After the optimal defocus step is determined, the z axis of the stage is moved (up or down with respect to the transverse axis of the tip) to reset the defocus to 0 nm.



(a) coarse mode (b) fine mode

Figure 2.13: The quality factors q of defocus series are shown for the coarse and fine defocus series mode in (a) and (b) respectively. The optimal defocus value is found by using a defocus value close to the peak and another defocus series is done until the defocus value gives the highest q. The coarse mode here are taken at 20 k× and fine mode, 640 k×.



(a) tip at 5 k× (b) line scans

Figure 2.14: The line scans of tungsten tip at different locations at 5 k× are shown here. In (a), the red line shows a line scan 3.55 µm below the apex, and the green line shows a line scan 6.61 µm below the apex. The intensity of the corresponding line scan is shown in (b), which a drop of intensity in the middle of the tip is visible.

The intensity decreases with the thickness and this is especially apparent for 5 k× (see Figure 2.14). When the line scan is fitted, the drop in intensity is accounted for by including an exponential decay in the centre. The constant

of the exponential decay depends on the actual thickness of the tip and this is determined empirically.

Tip finding and centring. The tip is often out of the field of view after rotating, and it is important to have a routine to find the tip. The widths of the tip at different magnifications are used for this purpose. The width of tip is defined as the horizontal distance of the tip at the bottom of the field of view for a particular magnification when the apex is at the centre. During the search for the tip, x and y axes of the stage are stepped until some features (pixels with intensities higher than the dark count) are detected. The stage movement will then follow the edge of the tip — the edge of the tip is defined by the slope of the tip width at different magnifications — until the horizontal distance of the tip matches the measured tip width.

After every rotation, the tip is found by the above steps and progressively magnified from 5 k×, 20 k×, 80 k× and finally 160 k×. At 160 k×, a defocus series is performed to obtain an optimal value of defocus. The apex of the tip is at the centre with a good focus, and it is ready for higher magnifications.

2.2.3 Stage calibration

The movement of TEAM stage determines the position of the tip's apex at different rotations, and calibration is needed in order to find and acquire images at the apex of the tip at different angles. The calibration ensures high-quality images are taken at the correct tomography tilt angles ϕ, which depends on α and γ of the stage. These are discussed below.

Calibrating x, y and z with γ. When the tip is rotated by γ, the two chopsticks move in an opposite direction by the same distance. The asymmetric

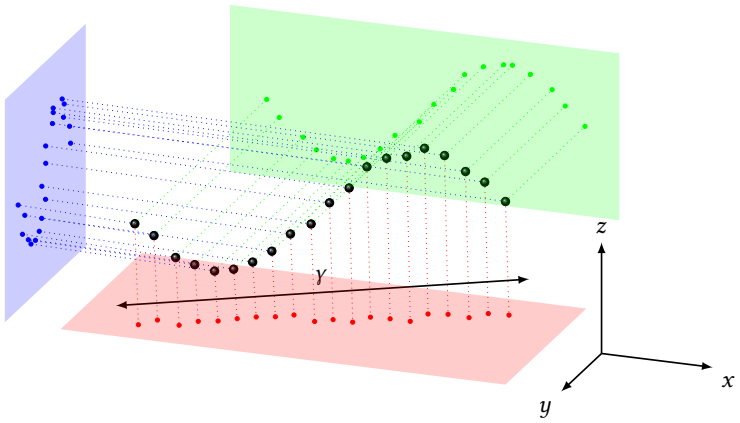


Figure 2.15: A schematic diagram shows the precession of the apex of the tip due to its asymmetric geometry and the movement of the puck. The black dots show the trace of the apex of the tip while the chopsticks are moving for γ rotation. The x, y, z coordinates of the stage are recorded and their variations with γ depend on the actual tip. Figure 2.16 shows variations of x, y and z coordinates with γ for an actual etched tungsten tip.

geometry of the tip and the movement of the puck cause the apex of the tip to precess when the puck is rotated by the two chopsticks (see Figure 2.15).

It is important to have an idea of how big the precession is for a particular tip, so the x, y and z range of the apex are known. This is done by fixing α at 90° and recording the x, y and z coordinates of the stage with γ from −180° to 180°, with the apex of the tip at the centre. The γ step is usually 20° and starts from $\gamma = 0°$ to $\gamma = 180°$ first and after that, from $\gamma = 0°$ to $\gamma = -180°$. After moving to a new γ, the apex of the tip is found and focused with procedures described in Section 2.2.2. At each γ a CBED pattern is also recorded. The x, y and z coordinates are approximately sinusoidal with γ as the trace of apex is a sinusoidal curve in three-dimensional space (see Figure 2.16). Every tip precesses in a different way so the variations could be very different than that of Figure 2.16. With this information, the tip can go to a different γ with a prior knowledge of its apex's coordinates, reducing the chance of losing the tip.

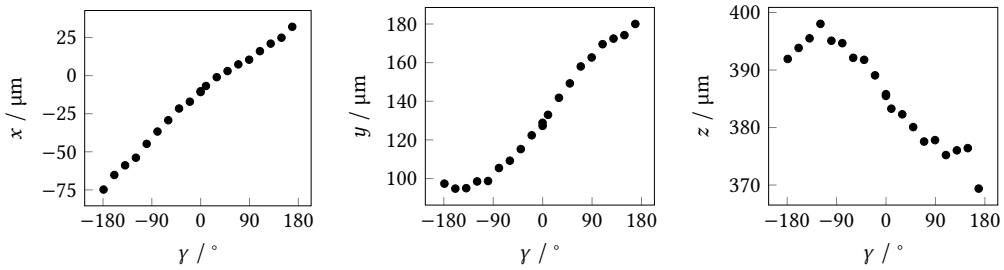
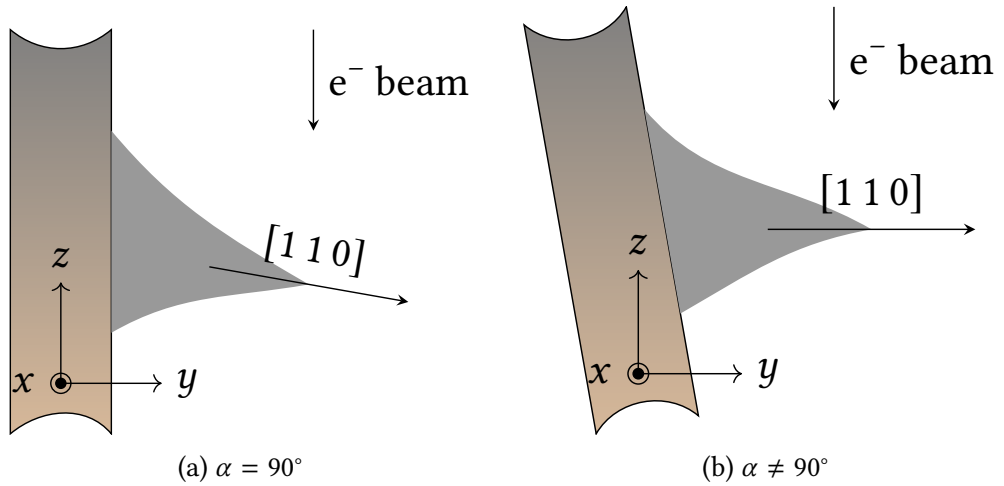


Figure 2.16: The variations of x, y and z coordinates of the TEAM stage with γ are shown with α fixed at 90°. The calibration of the stage is usually done with a γ step of 20°.



(a) $\alpha = 90°$ (b) $\alpha \neq 90°$

Figure 2.17: α can be altered to orient the [1 1 0] axis perpendicular to the electron beam. In (a), α is at 90° but the [1 1 0] axis is not perpendicular to the electron beam. By altering α as in (b), it can be made perpendicular to the electron beam.

Calibrating ϕ with α and γ. In tungsten tip tomography, the [1 1 0] axis is the tilt axis of the tomography experiment. All tilt angles ϕ are defined by rotation around the [1 1 0] axis of the tungsten tip. The [1 1 0] axis needs to be perpendicular to the direction of the electron beam. If the transverse axis of the tungsten tip is exactly perpendicular to the puck at all γ, then α can stay at 90°. This is very unlikely, so α needs to be altered in order to orient the [1 1 0] axis perpendicular to the electron beam (see Figure 2.17).

Rotation around the [1 1 0] axis of the tungsten tip is equivalent to going along the primitive great circle of the stereographic projection of (1 1 0) at the centre

(see Figure 2.18). There are low-index zone axes which can be easily identified by taking a high-resolution image or a CBED pattern. To find those low-index zone axes, Kikuchi lines from the CBED patterns obtained in the calibration of x, y and z can be used as a starting point. Some guesses of α and γ are needed: go to the guess α and γ, obtain a CBED pattern, and try to see if this is a low-index zone axis. This can be confirmed by taking a high-resolution image. Once a low-index zone axis on the primitive great circle is identified, the remaining zone axes can be found by their angular relationship. Kikuchi lines are moving when the tip is tilted, and they converge on zone axes. The (1 1 0) Kikuchi line is usually the line deviating not too much from $\alpha = 90°$ if the precession of the tungsten tip is small.

Every movement along the primitive great circle is achieved by rotating α and γ with translational movement, and this defines the tilt angle ϕ. After identifying several low-index zone axes on the primitive great circle, their α and γ values are used for calibration to map tilt angles ϕ to a set of (α, γ); this is possible because all ϕ lie on the primitive great circle. The relationship of α and γ when it is moving along the (1 1 0) Kikuchi line is a sinusoidal curve if the tip apex is precessed (see Figure 2.15). An experimental example of the relationship between α and γ is shown in Figure 2.19. Different tip geometries have different relationships of ϕ to α and γ, depending on the precession. Some common CBED patterns along the primitive great circle of (1 1 0) are shown in Figure 2.20 as they are invaluable for searching low-index zone axes.

2.2.4 Imaging parameters

Images for AET are acquired at a much higher magnification than those used in tip finding and centring. This magnification is reached by progressively increasing the magnification once the tip is found after rotation; therefore, con-

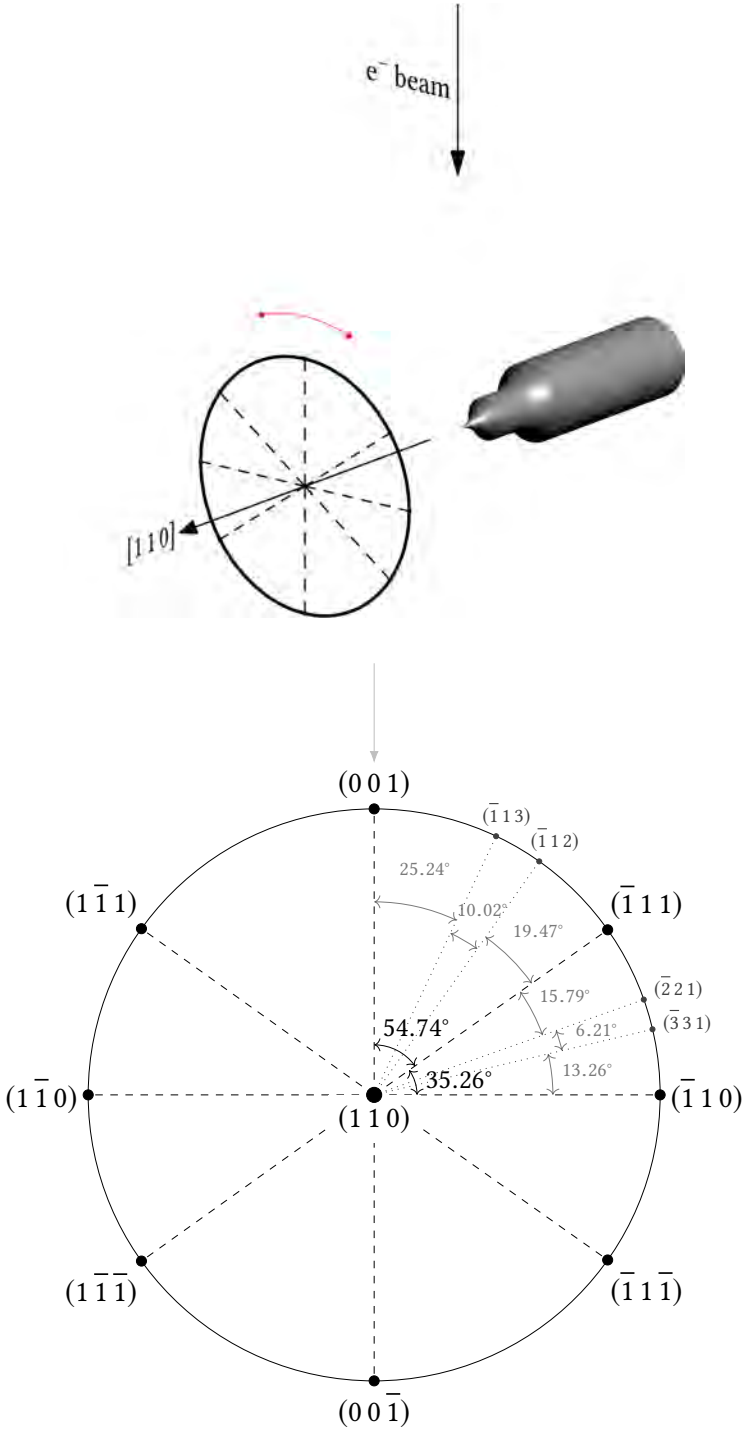


Figure 2.18: The [1 1 0] axis of the tungsten tip is the tilt axis for tungsten tip tomography and all tilt angles are defined around the [1 1 0] axis. Rotation around [1 1 0] axis is equivalent to going through the primitive great circle of the (1 1 0) pole in the centre. The angles between different planes on the (1 1 0) primitive great circle are shown with high-index planes in grey as they are usually more difficult to identify in images.

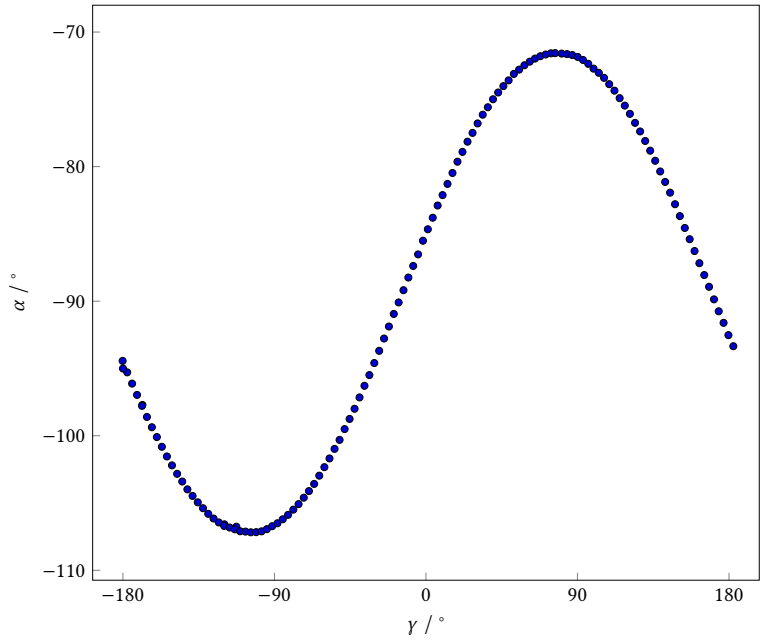


Figure 2.19: The experimental relationship between α and γ along the primitive great circle of (1 1 0) of an etched tungsten tip is shown here. The relationship is a sinusoidal curve as the apex of the etched tungsten tip is precessed.

trast factors are needed for magnifications — this set of magnifications can be changed — at 640 k×, 1.3 M×, 2.55 M×, 3.6 M× and 5.1 M×. This can be done similarly as in Section 2.2.2. The magnification used for the actual tomography experiment is usually >3 M× for identifying individual atoms. Once the imaging magnification is selected, a position inside the field of view with a threshold intensity is selected for vertical alignment; this position is usually at the top of the nanoparticle on the apex. The vertical alignment attempts to place the sample roughly at the same location in the field of view of the same magnification at different tilt angles.

A line scan is used to focus the sample (see Section 2.2.2). The position of the line scan can be set to avoid intersecting the region of interest of tomography. A defocus offset can also be set to account for the height difference between the sample and the position of line scan. A region inside the field of view is used to calculate a quality factor of the focused nanoparticle, and this region can be adjusted with an appropriate size and location. A threshold quality factor can

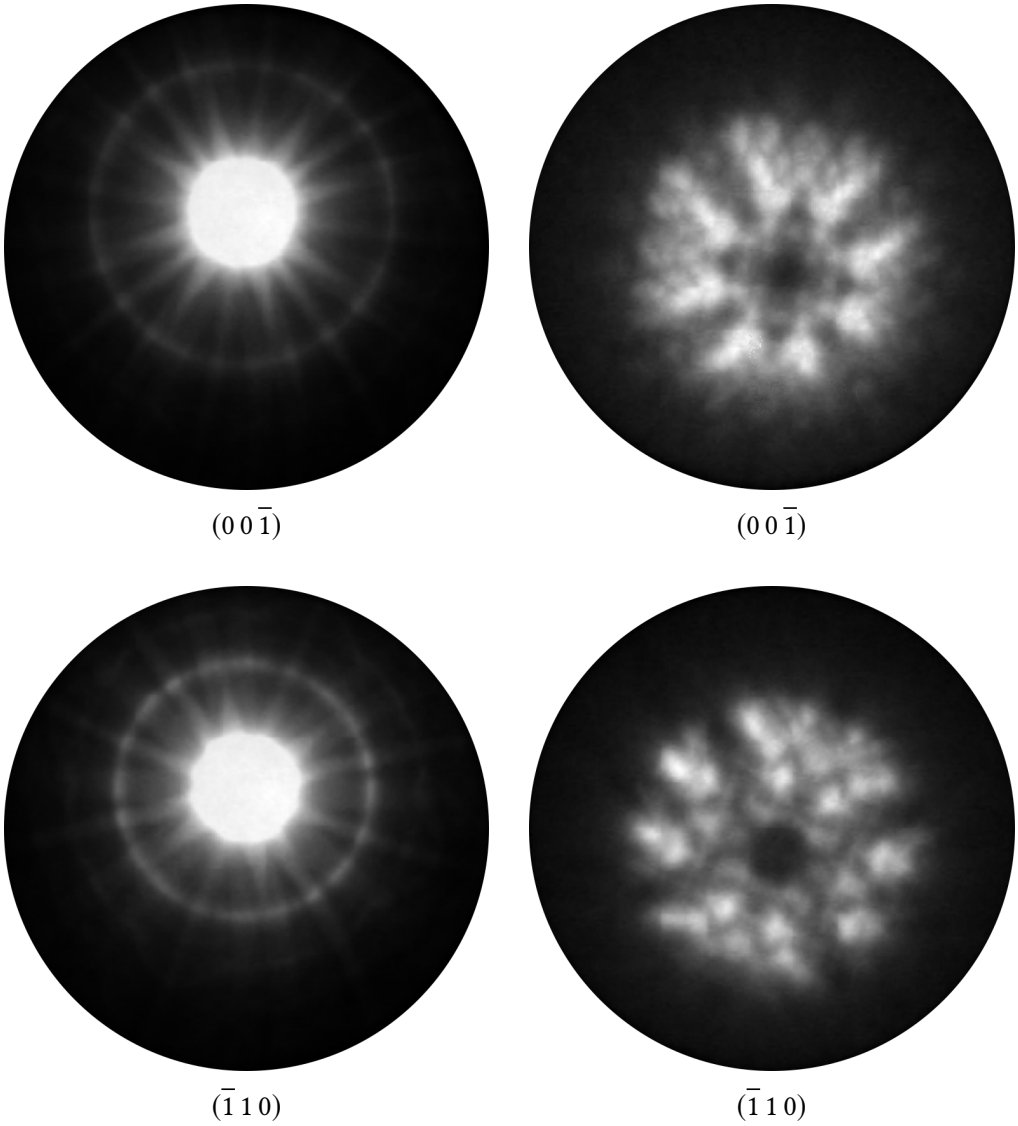


Figure 2.20: CBED patterns of (0 0 $\bar{1}$) and ($\bar{1}$ 1 0) — images on the left are taken 200 nm below the apex with camera length 34 mm, and images on the right, 1500 nm below the apex with camera length 135 mm. The convergence angle is 24.96 mrad. The radius of the cut-off area on the left and right side is 140.6 mrad and 45.1 mrad respectively.

be set to ensure the quality of the focused sample; if the quality factor is lower than the threshold, the image will be automatically retaken.

The scan direction can be adjusted to avoid scanning along the {1 1 0} planes. This is usually set at 45°. Dwell time and the camera length for HAADF imaging are also set, usually 3 µs and 135 mm at 3.6 M×, respectively.

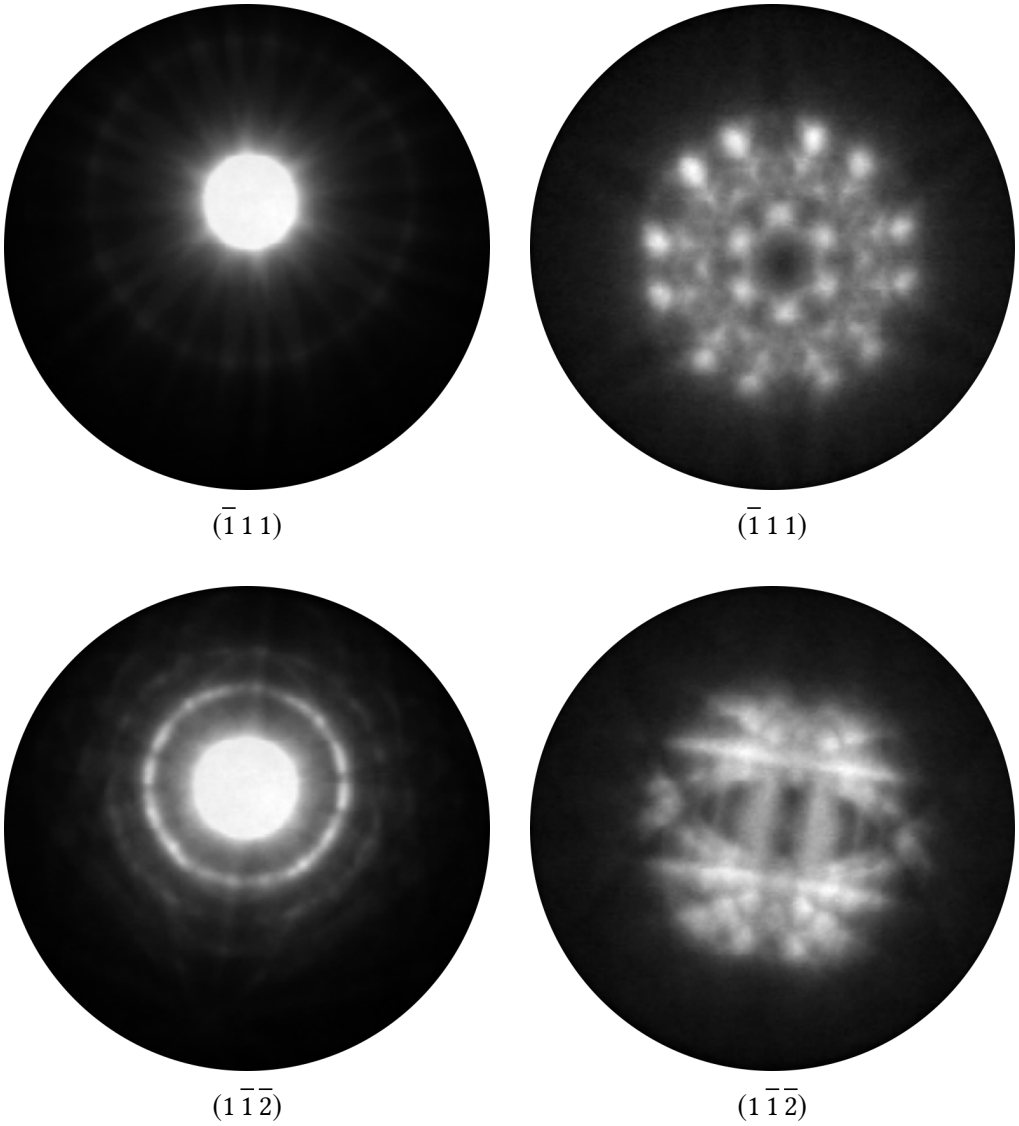


Figure 2.20 *(continued)*: CBED patterns of $(\bar{1}\,1\,1)$ and $(1\,\bar{1}\,\bar{2})$ — images on the left are taken 200 nm below the apex with camera length 34 mm, and images on the right, 1500 nm below the apex with camera length 135 mm. The convergence angle is 24.96 mrad. The radius of the cut-off area on the left and right side is 140.6 mrad and 45.1 mrad respectively.

2.3 Data acquisition

After going through all the procedures in Section 2.2, a high-resolution tomography dataset can be taken. A tilt angle series is defined by the angular range and the number of tilt angles. Although theoretically a 360°-angular range does not provide additional information due to the property of Radon transform

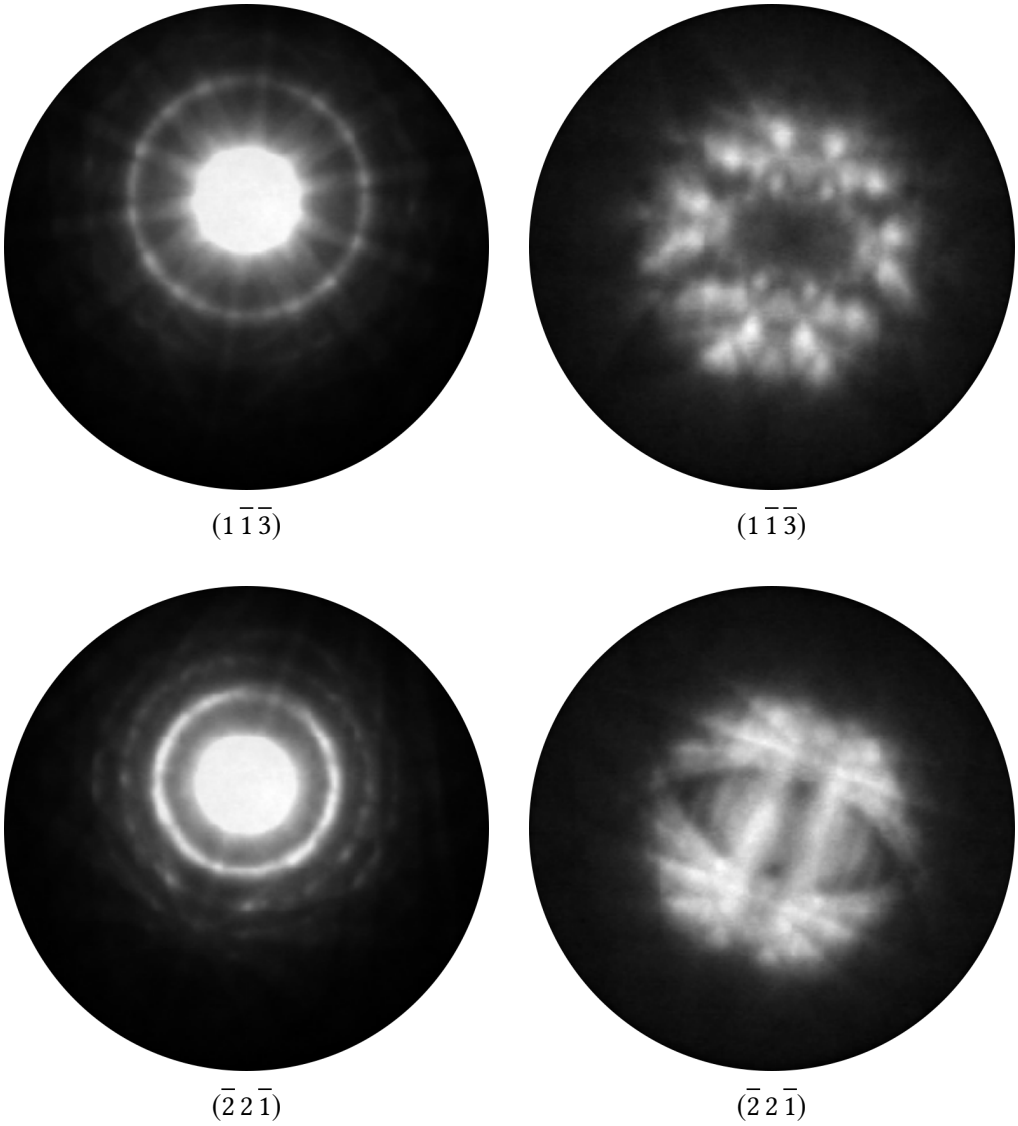
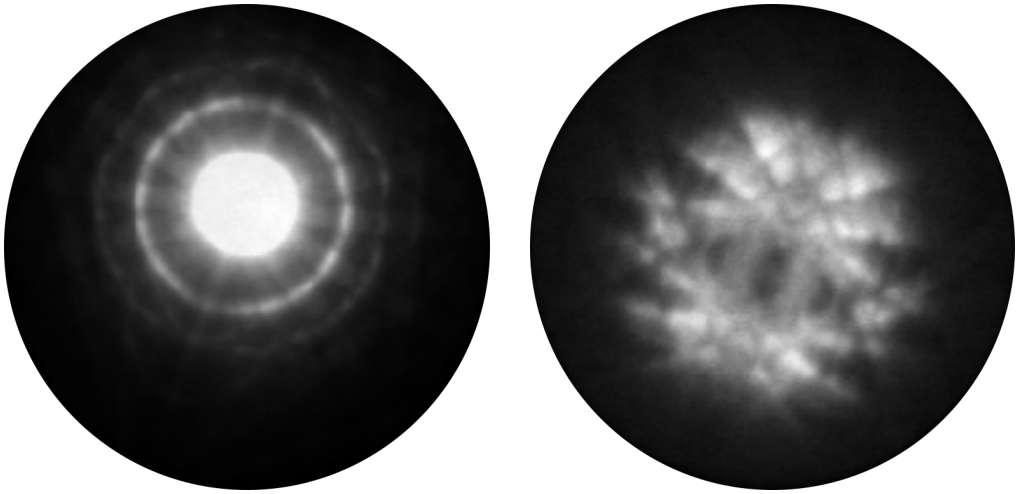


Figure 2.20 *(continued)*: CBED patterns of $(1\,\bar{1}\,\bar{3})$ and $(\bar{2}\,2\,\bar{1})$ — images on the left are taken 200 nm below the apex with camera length 34 mm, and images on the right, 1500 nm below the apex with camera length 135 mm. The convergence angle is 24.96 mrad. The radius of the cut-off area on the left and right side is 140.6 mrad and 45.1 mrad respectively.

(see Back-projection methods), images acquired by an angular difference of 180° are different in actual experiments due to different depths of focus, channelling effect and other artefacts. The number of tilt angles should be divisible by the angular range for easy analysis, but this is not required; for example, 61 tilt angles are usually used for a range of 180°. The tilt angles $\phi = 0°$ and



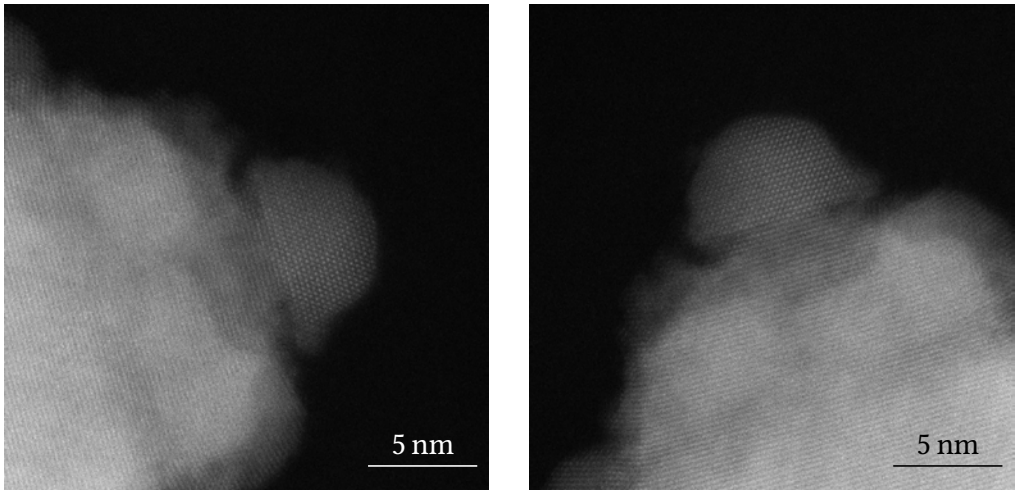
$(\bar{3}\,3\,\bar{1})$ $(\bar{3}\,3\,\bar{1})$

Figure 2.20 *(continued)*: CBED patterns of $(\bar{3}\,3\,\bar{1})$ — images on the left are taken 200 nm below the apex with camera length 34 mm, and images on the right, 1500 nm below the apex with camera length 135 mm. The convergence angle is 24.96 mrad. The radius of the cut-off area on the left and right side is 140.6 mrad and 45.1 mrad respectively.

$\phi = 90°$ are defined by the values of α and γ at the $(0\,0\,1)$ and $(1\,\bar{1}\,0)$ plane respectively.

A list of tilt angles is calculated by the requested angular range and the number of tilt angles. If γ is too close to $\pm 180°$, an offset can be applied to shift all the tilt angles as the stage is not calibrated outside $\gamma = \pm 180°$. Before a full dataset is taken, a CBED pattern on a low-index zone axis can be taken to confirm its γ does not deviate too much from the value in the calibration as the puck may 'slip' after a period of time. If there is a systematic difference for different low-index zone axes, a γ offset can be applied to the list of calculated tilt angles.

The process of acquiring a full tungsten tip tomography dataset is fully automated. It goes to the respective tilt angle ϕ by changing x, y, z, α and γ of the stage; the tip is found, centred and focused at the target magnification; two images are taken at each ϕ, with scan directions 90° apart from each other (say



(a) (b)

Figure 2.21: Two acquired images in tungsten tip tomography for a gold nanoparticle are shown with a scan direction of (a) 45° and (b) 135°. The images were acquired at 3.6 M×, with a convergence angle of 30 mrad and a dwell time of 3 µs. The images were taken at a tilt angle close to the Au {1 1 0} plane and lattice structure can be seen clearly.

at 60° and another at 150°) (see Figure 2.21). Optionally, the defocus used in taking these two images can be set slightly different, usually a few nm. The image with a higher quality factor as determined by the setting in Section 2.2.4 is used for the final data analysis.

Minimal human intervention is required. Occasionally the tip is lost and the stage will keep looking for the tip, and sometimes the nanoparticle is too close to the edge of the image; this is usually due to carbon contamination around the nanoparticle, and the threshold in vertical alignment (see Section 2.2.4) is no longer suitable. This can be avoided by resetting the threshold or the position for vertical alignment.

After a full dataset is taken, all images are reviewed to check if there are any badly focused images or the nanoparticle is too close to the edge of the image. The bad images are retaken manually. It is useful to have more images around

low-index zone axes as more information on the lattice can increase the quality of reconstruction in AET.

3

Data processing

After obtaining a full tungsten tip tomography dataset, the images are processed and aligned before reconstruction. The processing is required to remove scanning artefacts and the alignment ensures all projections have the same centre of rotation. This is essential for obtaining a high-quality atomic-resolution reconstruction.

During scanning, the image suffers distortions due to environmental, electronic and mechanical instabilities [174]. The main concerns in electron microscopy are usually sample drift, temperature variation, electromagnetic interference, acoustic vibration, power and pressure fluctuation. The design of modern electron microscopes can minimise the image distortions from these effects, but they cannot be completely eliminated [175]. The drift happens in an arbitrary direction so the actual scan area is a rhombus instead of a square and different algorithms were developed to recover the distortion-free image. The distortion of a HAADF-STEM image can be removed by referencing the Fourier transform of a high-resolution TEM image of the same area [176, 177]. For a typical STEM image with horizontal fast-scan direction, there are vertical streaks in the Fourier domain due to random displacements of rows along the vertical direction.

The distortion can be corrected by the phase analysis of the vertical streaks in the Fourier domain [178], and if atomic features are also assumed, both linear and non-linear distortions can be corrected [179]. Later algorithms attempted to correct the distortion by using a series of images, either using multiple images in the same scan direction [180] or different scan directions [181, 182]. In an actual crystal, the atoms are oscillating around their equilibrium positions but the time scale is of the order of ps [183–185]. As these oscillations happen in a much shorter time scale than the dwell time, it is assumed the average positions of atoms are taken in the STEM image.

In this chapter, the approach described in [182] and reference lattice structures are used to process the images before alignment. These processing steps are described and illustrated by a dataset of a gold nanoparticle on the apex of tungsten tip, although the principles are not limited to datasets of tungsten tip tomography.

3.1 Image distortion corrections

If the drift is not corrected, it makes images at different tilt angles inconsistent with each other — the same atom appears to be moved in different projections due to the drift. The inconsistency degrades the quality of the reconstruction. For the TEAM 0.5 microscope, the thermal drift is very small as the microscope is heavily isolated (the temperature only changes by 0.25 °C over a duration of 16 h [172]). However, in a tomography experiment, the sample is frequently rotated and moved; if there is insufficient time for the stage to settle, the drift cannot be ignored. With a dwell time of 3 µs, it takes ≈3 s to acquire a 1024×1024 image and two images are acquired at each tilt angle; however, the stage remains stationary at a tilt angle for a few minutes as it needs time to focus and identify the tip apex (see Section 2.2.2). This is not sufficient for the stage to stabilise as

this usually takes hours. For practical reasons, there is no designated settling time for each tilt angle so this drifting should be dealt with.

Besides the drift from the sample, the distortion from scanning coils can also be significant. Scanning coils deflect the electron beam into different positions on the sample so that a raster scan on the sample can be performed by the beam. Two scanning coils are used separately for x and y deflection. The mechanical difference of the two scanning coils causes different magnetic fields, so the trajectories in the x and y direction are different.

3.1.1 Drift correction

An orthogonal pair of images is used for drift correction [182]. This method does not require any assumption with regard to the image features. This is the reason for taking two images during data acquisition of tungsten tip tomography (see Section 2.3). The algorithm and implementation of drift correction are detailed in [182]. Figure 3.1 shows an example before and after the drift correction. The streaks along the slow scan direction in Fourier space are due to random noise. After the image is drift-corrected, those streaks disappear. The image with a higher quality focus factor as determined in Section 2.2.4 is selected for further processing and reconstruction.

3.1.2 Scanning-coil correction

To determine the distortion from scanning coils, some reference structures are needed. For this purpose, high-resolution images (usually at 5.1 M×) of the tungsten (1 1 1) lattice are used (see Figure 3.2). The images of (1 1 1) tungsten lattice are taken with the same scan directions as other tomography projections. After the (1 1 1) lattice images are drift-corrected, they are Fourier-transformed; in the Fourier space, as expected at the (1 1 1) zone axis for a body-centred

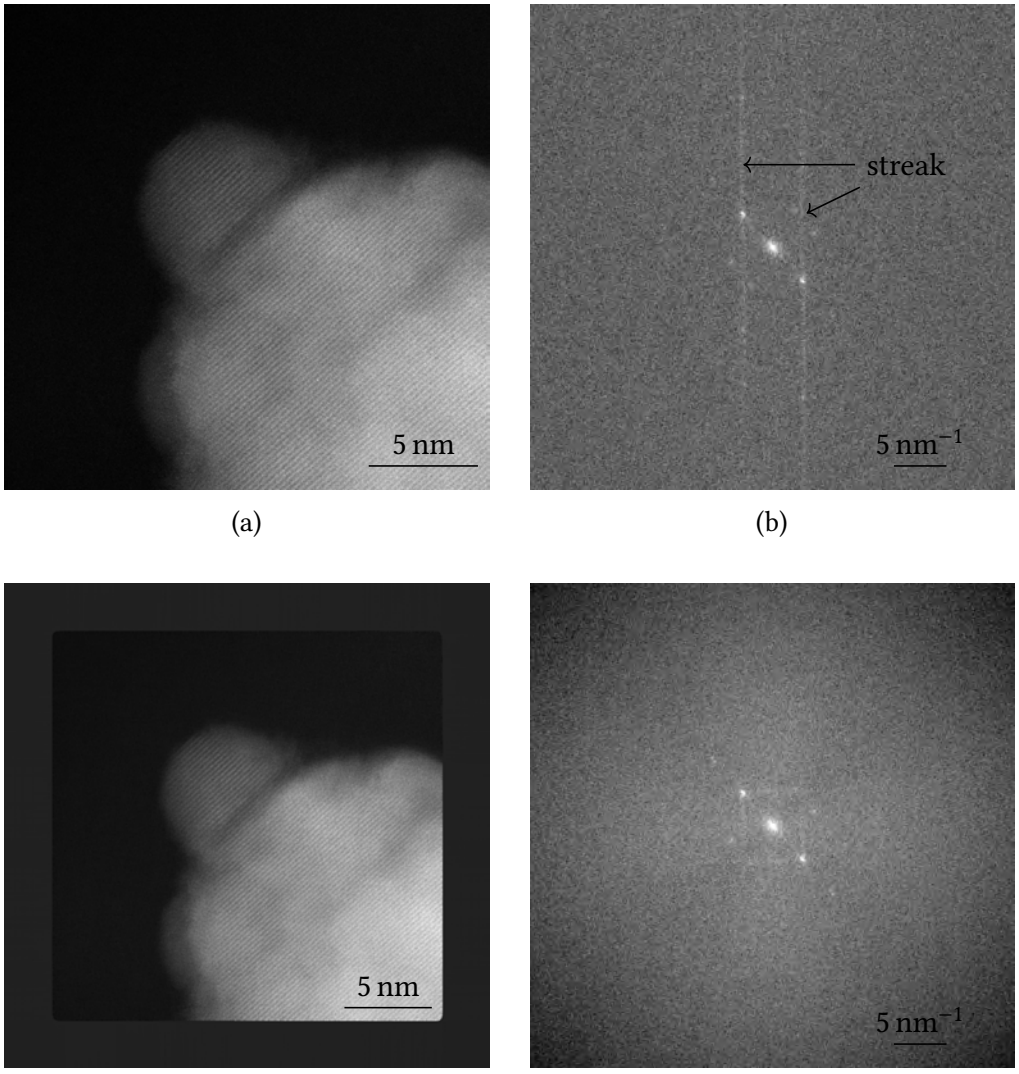
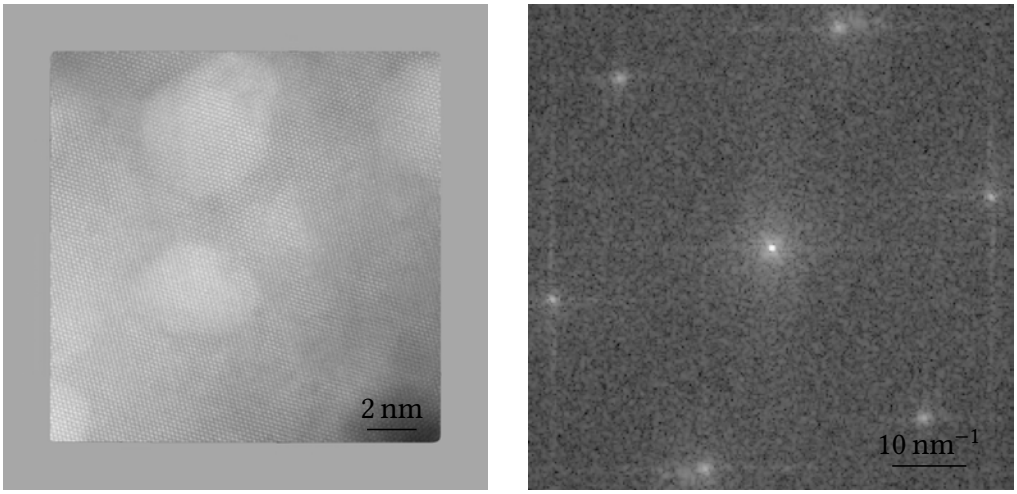


Figure 3.1: Microscopic images and their Fourier transforms before, (a) and (b), and after, (c) and (d), drift correction are shown here. Streaks are visible along the slow scan direction because of random drift. The drift-corrected image is padded by 128 pixels in every direction. The Fourier transform of drift-corrected image does not show streaks along the slow scan direction. The magnitudes of all Fourier transforms are in logarithmic scale and a Hanning filter is applied before Fourier transform.

cubic (bcc) material, a central spot, in addition to six spots, is seen. If there is no distortion from scanning coils, the vectors from the centre of the Fourier transform to each spot should be of the same length and exactly 60° from each other.



(a) (1 1 1) lattice (b) Fourier transform

Figure 3.2: The lattice image for scanning-coil distortion correction and its FFT image are shown here. (a) is the drift-corrected image at 5.1 M× taken along the (1 1 1) plane and a hexagonal lattice structure can clearly be seen. (b) is a zoom-in of its Fourier transform and six spots, characteristics of a hexagonal structure, are visible. The magnitudes of the Fourier transform are in logarithmic scale and a Hanning filter is applied before Fourier transform.

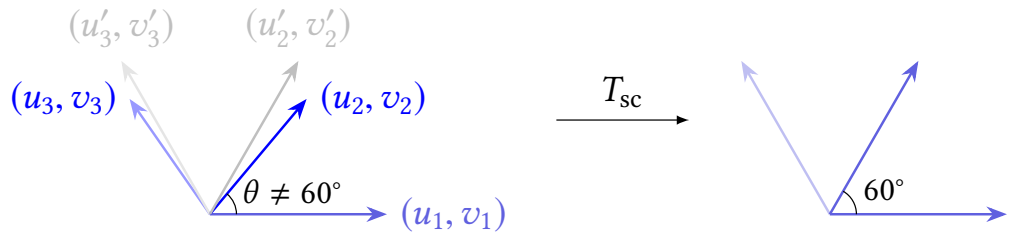


Figure 3.3: (u_1, v_1), (u_2, v_2) and (u_3, v_3) are the spots in Fourier transform from the experimental high-resolution image taken at the (1 1 1) lattice plane (see Figure 3.2b). (u_2', v_2') is exactly 60° from (u_1, v_1) and of the same length as (u_1, v_1). The required affine transformation matrix T_{sc} transforms (u_2, v_2) to (u_2', v_2'), while (u_1, v_1) remains unchanged. The two vectors (u_1, v_1) and (u_2, v_2) fully define a two-dimensional lattice.

For the spot (u_1, v_1) in Fourier space, the nearest spot in anticlockwise direction is labelled as (u_2, v_2). An anticlockwise rotation of 60° of (u_1, v_1) gives the point (u_2', v_2') (see Figure 3.3). The aim is to find the transformation matrix T_{sc} such that

$$\begin{pmatrix} u_1 & u_2' \\ v_1 & v_2' \end{pmatrix} = T_{sc} \begin{pmatrix} u_1 & u_2 \\ v_1 & v_2 \end{pmatrix} \tag{3.1}$$

Table 3.1: Deviations of FFT spots from the theoretical 60° before and after scanning-coil distortion correction

	% deviation from 60°			
	$\measuredangle(\vec{v}_1, \vec{v}_2)$	$\measuredangle(\vec{v}_2, \vec{v}_3)$	$\measuredangle(\vec{v}_3, \vec{v}_1)$	$\sum_{i,j=1,2,3} \measuredangle(\vec{v}_i, \vec{v}_j)$
no correction	−0.911	−1.214	2.125	0
with correction	−0.029	0.106	−0.077	0

Table 3.2: Typical values for scaling and shearing factors

Factor	Value
s_u	1.000
s_v	1.026
k_u	−0.003
k_v	−0.001

This transformation is a combination of scaling and shearing in both u and v direction, so T_{sc} can be decomposed into

$$T_{\text{sc}} = \begin{pmatrix} s_u & 0 \\ 0 & s_v \end{pmatrix} \begin{pmatrix} 1 & k_u \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ k_v & 1 \end{pmatrix} \tag{3.2}$$

s_u, s_v, k_u and k_v are the scaling and shearing factors in u and v direction. According to Equation (3.1), T_{sc} can be found by

$$T_{\text{sc}} = \begin{pmatrix} u_1 & u_2' \\ v_1 & v_2' \end{pmatrix} \begin{pmatrix} u_1 & u_2 \\ v_1 & v_2 \end{pmatrix}^{-1} \tag{3.3}$$

All projections undergo affine transformations by T_{sc}. Table 3.1 shows the typical deviations of FFT spots (see Figure 3.3) before and after scanning-coil distortion correction, and Table 3.2 shows some typical values in the T_{sc} matrix.

3.2 Alignment

The purpose of the alignment is to ensure the same atomic planes from different projections are at the same locations and they are rotated around the same axis. This ensures all projections are consistent with each other. In tungsten tip tomography, the tilt axis is the [1 1 0] axis of the tungsten tip. As the high-angle annular dark-field (HAADF) intensity depends on the number of atoms, the integrated intensities of every atomic plane along the [1 1 0] direction remain unchanged when the atomic planes are rotated around [1 1 0], if all atoms are within the field of view in all projections (see Figure 3.4). This is a valid assumption as the lattice structure of tungsten is very stable under electron beam (see Section 2.1). The atomic planes can be aligned by comparing their integrated intensities and this is known as the 'common lines alignment'; the common line is the projection of intensity onto the tomography axis — this is the [1 1 0] axis in tungsten tip tomography. To obtain the common lines in different projections, all images are rotated such that the vertical axis of the images are aligned with the [1 1 0] tilt axis (see Figure 3.5). A common line at a particular tilt angle is chosen as a reference and the remaining common lines are cross-correlated with the reference. Figure 3.6 shows the aligned common lines of the tomography dataset of a gold nanoparticle on an etched tungsten tip.

After aligning the projections vertically, the projections are aligned horizontally and this is achieved by the centre-of-mass alignment [11]. The images are projected on the horizontal axis and the centre of mass x_{CM} is defined as

$$x_{\mathrm{CM}} = \sum_{x=-\frac{N}{2}}^{\frac{N}{2}-1} \frac{xI(x)}{I(x)} \tag{3.4}$$

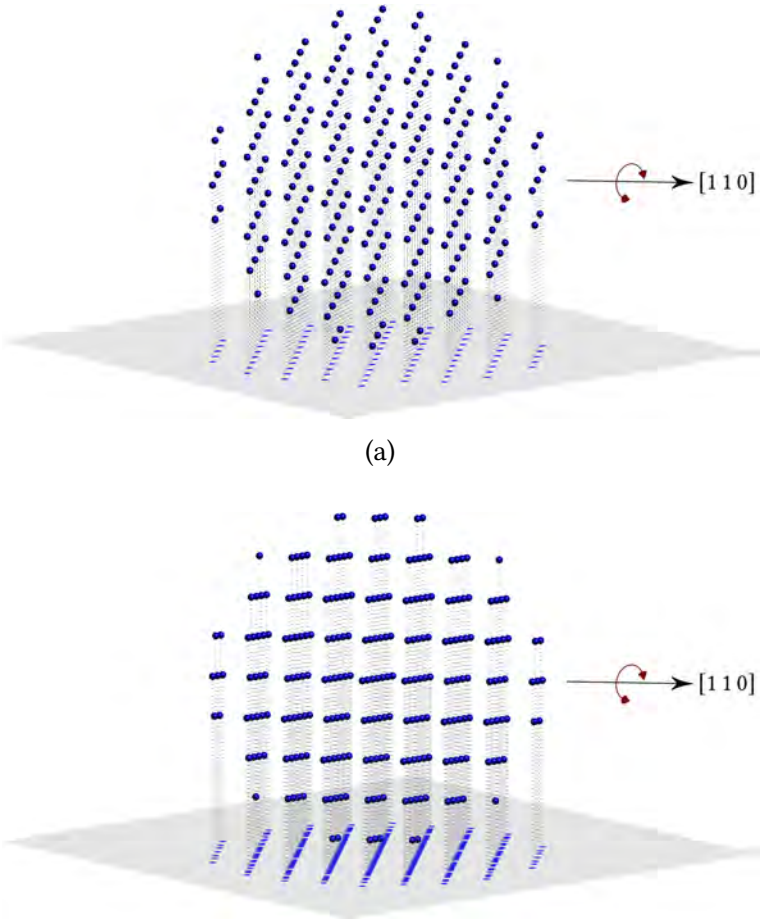
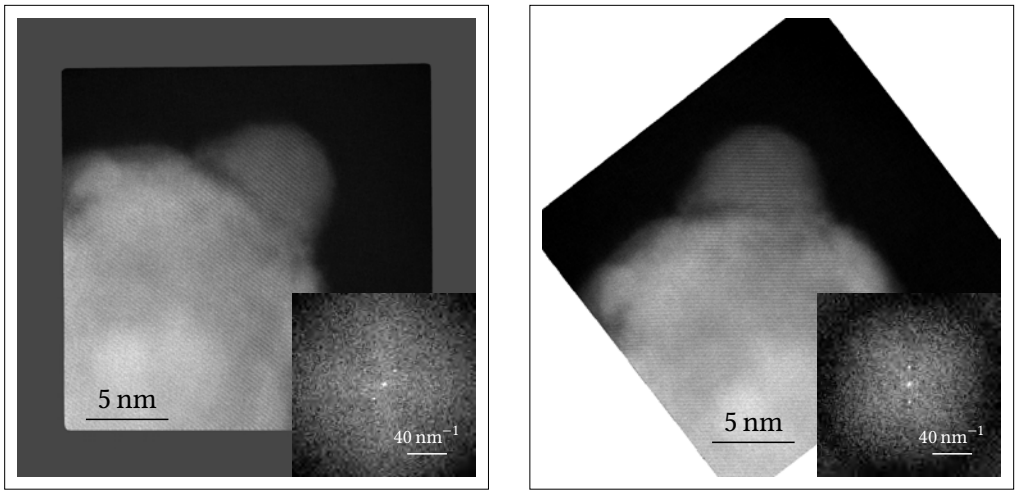


Figure 3.4: An atomic model of a bcc crystal, the same crystal structure of tungsten, is shown here. When the bcc crystal is rotated around the [1 1 0] axis, the number of atoms projected on the [1 1 0] axis is unchanged, this is the common line projection; therefore, the HAADF intensity is constant. By comparing the HAADF intensity of different atomic planes along the [1 1 0] axis, the atomic planes can be precisely aligned with the assumption that the atoms do not move around. (a) and (b) have an angular difference of 10° to each other around the [1 1 0] direction.

N is the total number of pixels along the horizontal axis and $I(x)$ is the pixel intensity at pixel x which ranges from $-\frac{N}{2}$ to $\frac{N}{2} - 1$. All images are shifted horizontally by $-x_{\mathrm{CM}}$ so that the new centre of mass x'_{CM} is at the centre of the image, $x'_{\mathrm{CM}} = 0$. In practice, only the region of sample and several layers of tungsten tip is used for the centre-of-mass alignment. Those layers must also be fully within the field of view. An example of centre-of-mass alignment is shown in Figure 3.7.



(a) (b)

Figure 3.5: To obtain the common line of [1 1 0], the corrected images are rotated such that the tip is pointing vertically. The common line is then the summation of intensities along the horizontal axis. The tip in (a) is not pointing vertically and it is rotated in (b) such that the [1 1 0] axis is pointing vertically upwards. The insets show the Fourier transforms of a region of the tungsten tip.

All the vertical and horizontal alignments are done in the Fourier space so that sub-pixel accuracy can be attained. This is based on the Fourier shift theorem, and it states that multiplying a linear phase factor of the shift s with the Fourier transform of a function $f(x)$ is equivalent to the Fourier transform of $f(x-s)$:

$$\begin{aligned} e^{ius}\mathcal{F}[f(x)] &= e^{ius}\int_{-\infty}^{\infty} f(x)e^{-iux}\,\mathrm{d}x \\ &= \int_{-\infty}^{\infty} f(x)e^{-iu(x-s)}\,\mathrm{d}x \\ &= \int_{-\infty}^{\infty} f(x-s)e^{-iu(x-s)}\,\mathrm{d}(x-s) \\ &= \mathcal{F}[f(x-s)] \end{aligned} \tag{3.5}$$

To shift an image vertically or horizontally, the integral part of the required shift is first performed by rolling the pixels and then the remaining decimal part is achieved by the Fourier shift theorem. This is because rolling pixels

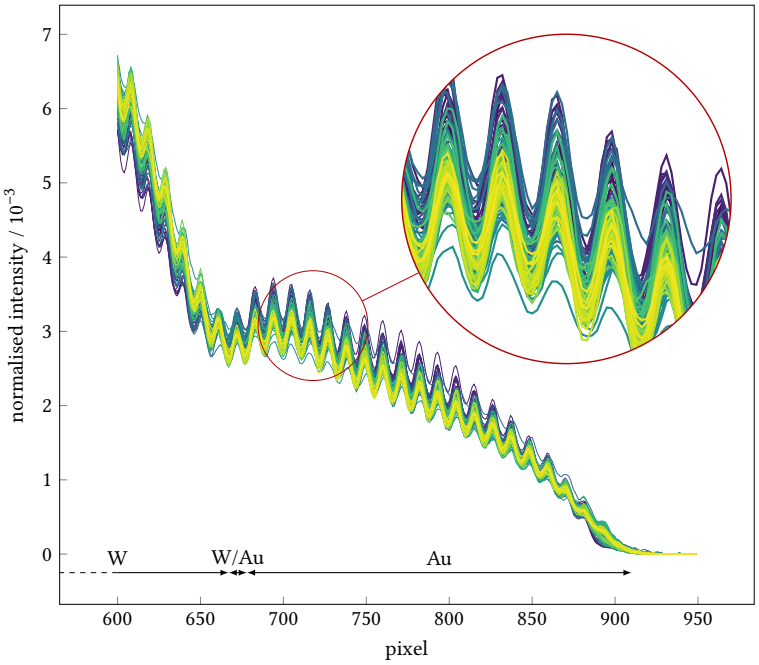


Figure 3.6: The aligned common line of 110 projections of a gold nanoparticle on an etched tungsten tip is shown here. Every peak is an atomic plane and in principle, those peaks should overlap with each other exactly if they represent the same atomic planes. However, every projection is taken with a slightly different focus so the widths and amplitudes of the peaks are different. The arrows at the bottom identify the regions corresponding to the gold nanoparticle and the tungsten tip.

does not cause a loss of information and this can minimise potential artefacts during alignment.

3.3 Refinement

After the discussed alignment steps, the dataset is almost ready for reconstruction. To achieve a reconstruction with an atomic precision, the lattice structure of the sample needs to be accurately aligned. The information of lattice structure is present in the images taken at the major low-index zone axes, (1 0 0), (1 1 0) and (1 1 1). The alignment of those images requires a very high accuracy so the lattice information is reliable, and this makes tomography with atomic precision possible. This is done by shifting the images according to their phases in the Fourier domain.

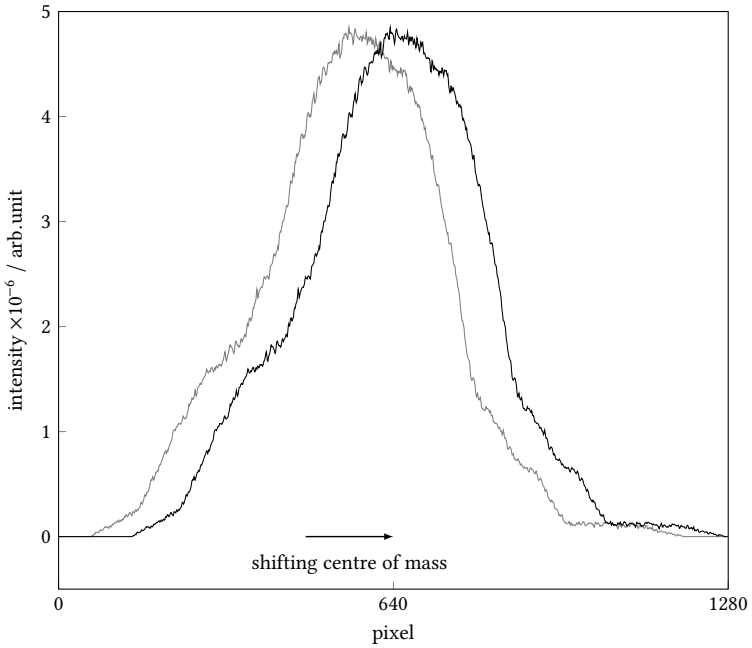


Figure 3.7: Projections of a certain region of the sample and tungsten tip on the horizontal axis of the image are used for the centre-of-mass alignment. The centre of mass of the projection is calculated and it is moved to the centre so that every projection has the centre of mass at the centre.

For simplicity a single layer along the [1 1 0] direction, either of the sample or of the tungsten, is used for this purpose, so the operations are in two-dimensional space. This is possible because if one layer is aligned, other layers must also be aligned. In Figure 3.8a, the atom is projected exactly to the centre for the projections along zone axes. In this case, the projections are symmetric above the centre of the projections and their Fourier phase angles are zero. In Figure 3.8b, the atom is shifted such that the projections are not symmetric around the centre for two projections. The Fourier phase angles in those two projections are non-zero. They are opposite in sign as Δs_2 and Δs_3 are the same in magnitude but opposite in direction. The generalised case is shown in Figure 3.8c, where the following relationship is always valid:

$$\Delta s_1 + \Delta s_2 + \Delta s_3 = 0 \tag{3.6}$$

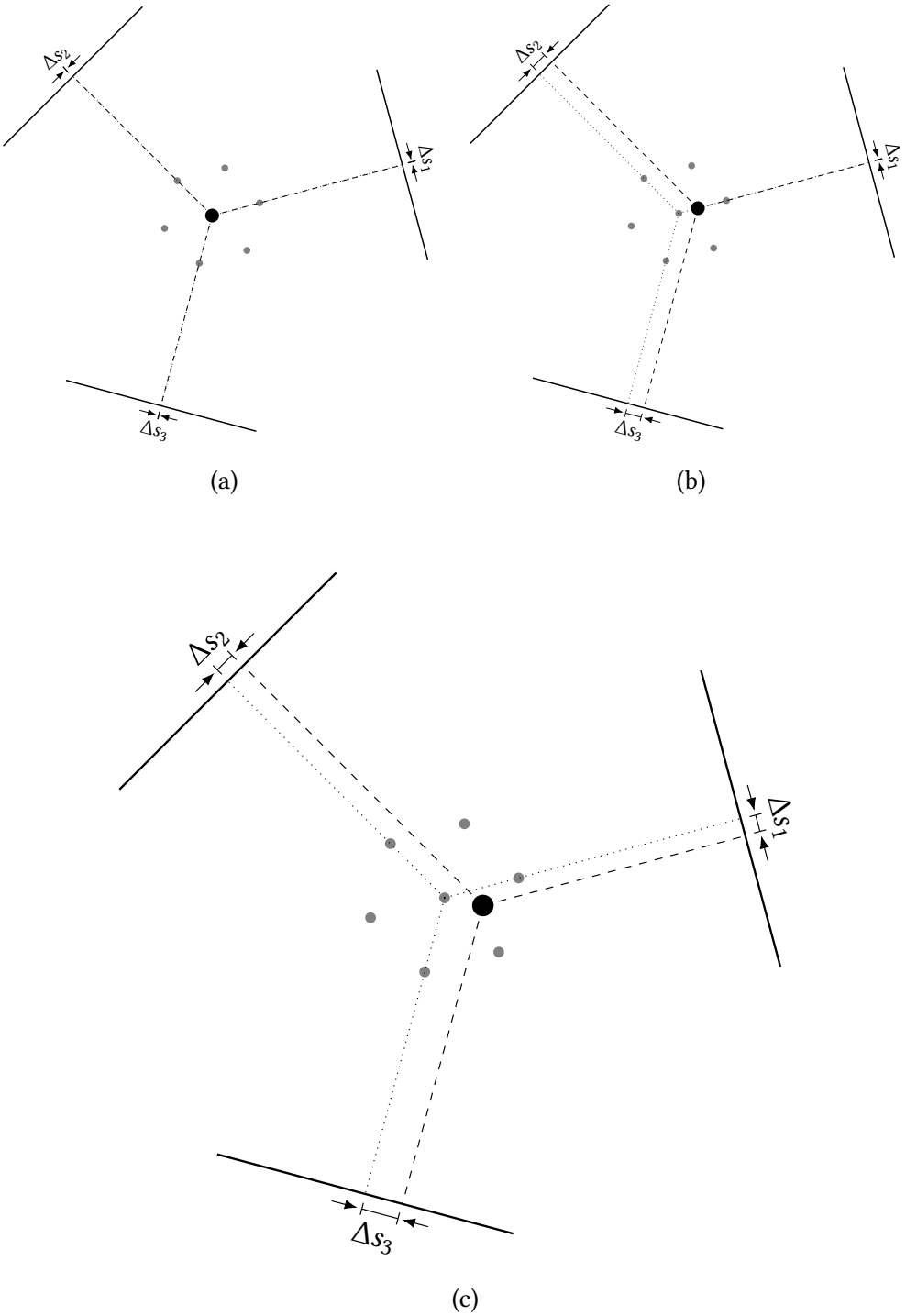


Figure 3.8: Fourier phase angles for a particular atomic lattice are zero when the lattice is symmetric above the centre of the projection; otherwise, they are non-zero. The sum of all Fourier phase angles for a particular atomic lattice is always zero. (a) shows the case where all Fourier phase angles are zero, (b) shows two Fourier phase angles are non-zero and opposite in sign, and (c) shows a generalised case where all Fourier phase angles are non-zero.

Figure 3.8 shows the case for a particular atomic lattice and this can be generalised: the sum of all Fourier phase angles of the same atomic lattice is always zero. Using this property, the projections with the same atomic lattice are aligned.

The low-index zone axes which display clear column contrast are aligned using the property in Equation (3.6) and fixed afterwards. All other projections are re-aligned by cross-correlation with the projections from an initial reconstruction. The re-alignment stops until all projections are consistent with the reconstruction.

4

Atomic site tomography

A variety of reconstruction algorithms were presented in Section 1.2 and Section 1.3 to recover the three-dimensional structure of the materials. Traditional algorithms such as filtered back-projection and variants of algebraic reconstruction methods (see Section 1.2.3) work well when the tomography dataset is well defined and they are more suitable in conventional electron tomography (ET), where atomic precision is not required. Reconstruction methods that utilise the concept of compressed sensing and employ a data-driven approach are gaining traction (see Section 1.2.3); however, they often require some prior knowledge of the sample and those assumptions cannot be easily translated between different types of samples. In the field of atomic electron tomography (AET), discrete tomography (see Section 1.3.1) also exploits prior knowledge of the sample; it assumes the atoms are in a regular grid and projections must be taken along zone axes. State-of-the-art techniques such as equally sloped tomography (EST) (see Section 1.3.2) and generalised Fourier iterative reconstruction (GENFIRE) (see Section 1.3.3) are iterative algorithms between real and Fourier space. EST uses a pseudo-polar grid during reconstruction and equally-sloped projections must be taken; GENFIRE does not have this limitation, but regridding is needed during iterations between real and Fourier space.

In this work, a new reconstruction algorithm, atomic site tomography (AST), is developed to provide a promising way to reconstruct tomography dataset with atomic precision. AST works in the real space only; therefore, there is no loss of information when a layergram (see Figure 1.7) is constructed, such as the case in GENFIRE. It minimises the required prior knowledge of the sample during reconstruction — it only assumes projections are coming from individual atoms and no limitations are imposed on the locations of individual atoms; this is a very general assumption and is applicable to all nanoparticles whether they have crystalline or amorphous structures. There is also no requirement on the tilt series, unlike discrete tomography or EST where projections along zone axes or equally-sloped tilt angles must be acquired, respectively. This chapter outlines the principle of AST and presents validation of its application to simulated data based on atomic models. AST is then compared to algorithms that require iterations between real and Fourier space and its performance is evaluated.

4.1 Simulated data

The aim of all the tomography reconstruction algorithms is to provide a model representing the structure we are interested in. AST is specifically designed for AET to identify the precise locations of atoms (*atomic site*), or more generally identical entities that are located in a finite region of three-dimensional space. To validate the algorithm, routines to generate different crystal structures and obtain mathematical projections under simulated experimental conditions were written.

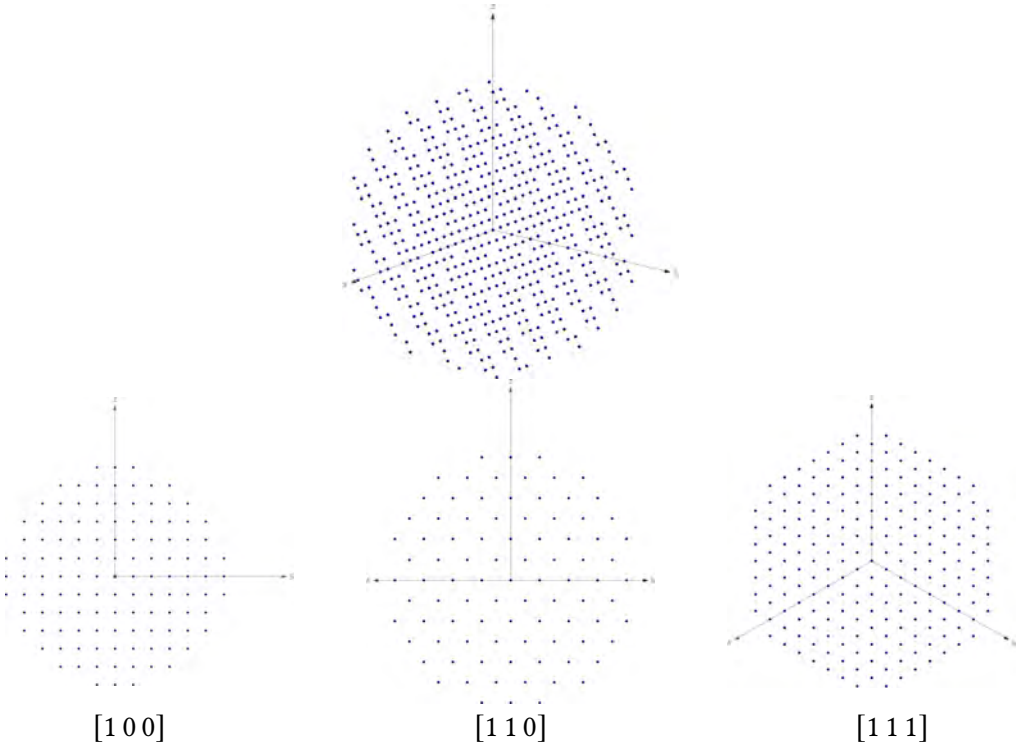
AET usually involves recovering the individual atomic positions of crystalline nanoparticles; therefore, spherical models of crystalline nanoparticles are generated. Figure 4.1 shows face-centred cubic (fcc) and body-centred cubic (bcc)

spherical crystal models viewed in different directions. To validate the AST algorithm, a tungsten spherical crystal model with a diameter of 2 nm is constructed. There is one vacancy site in this model and the atoms are displaced from their lattice positions with a Gaussian probability distribution having a width of 0.1 Å. The use of a relatively large atomic displacement (compared with those observed in real systems which are normally of the order of 10^{-2} Å [186–188]) is to demonstrate the robustness of AST if the lattice is not perfectly crystalline. The crystal model shown in Figure 4.2 will be used to demonstrate the algorithm of AST.

With the three-dimensional coordinates of the atoms known, mathematical projections can be calculated at different sample tilts. Figure 4.3 shows the simulated dataset from the crystal model in Figure 4.2 taken with 15 tilt angles evenly distributed between $-90°$ and $90°$. The tomography axis is along the [1 1 0] direction of the crystal model. The projected intensity of a single atom is modelled by the following function:

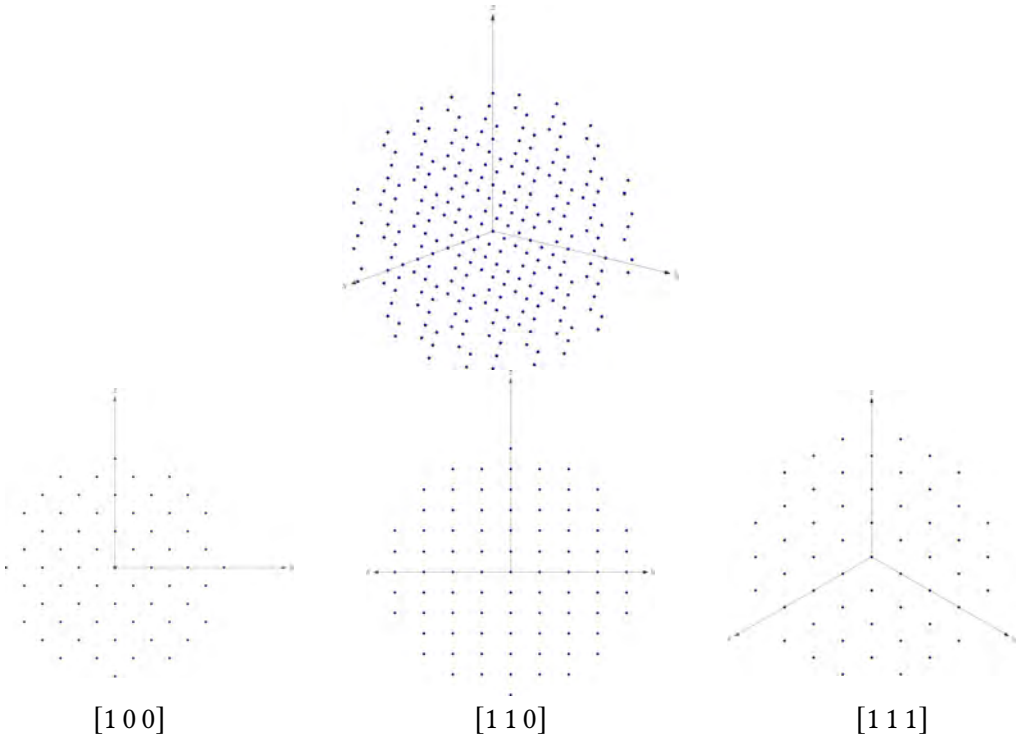
$$\frac{A}{2\pi\,|\Sigma|^{\frac{1}{2}}} e^{-\frac{M_{d_1}^2}{2}} + \frac{B}{2\pi\,|\Sigma|^{\frac{1}{2}}} e^{-\frac{M_{d_2}^2}{2}} \qquad \left(\frac{\sigma_{Y1}}{\sigma_{X1}} = \frac{\sigma_{Y2}}{\sigma_{X2}}\right) \tag{4.1}$$

Equation (4.1) is the addition of two bivariate Gaussian functions having different widths along the x and y directions. The parameters used in the simulated dataset are $A = 80$, $\sigma_{X1} = 0.11$, $\sigma_{Y1} = 0.081$, $\rho = 0$, $B = 20$ and $\sigma_{X2} = 0.061$; $M_d(\vec{x}, \vec{\mu}, \Sigma)$ is the Mahalanobis distance and $\Sigma(\sigma_X, \sigma_Y, \rho)$ is the covariance matrix (see Table 5.1 for their definitions). These parameters are representative of a real-world experiment. Poisson noise is added to the dataset. There is no misalignment of the dataset because all reconstruction algorithms assume the dataset is already aligned (see Section 3.2).



[1 0 0] [1 1 0] [1 1 1]

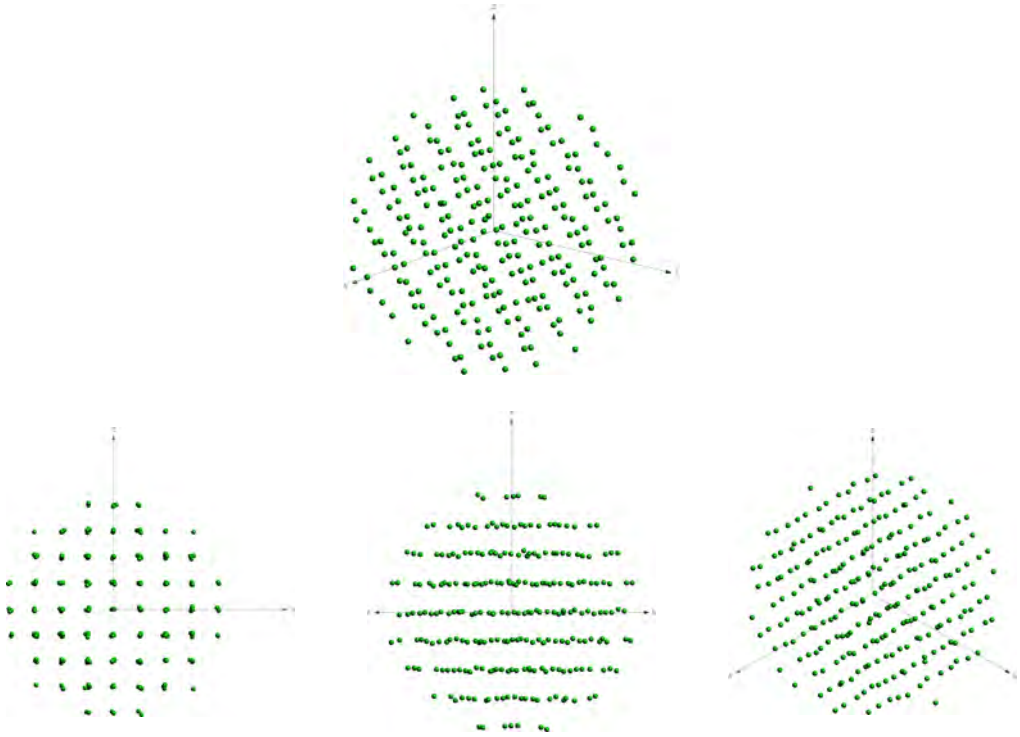
(a) fcc



[1 0 0] [1 1 0] [1 1 1]

(b) bcc

Figure 4.1: The crystal models, generated by the written routines, of (a) fcc and (b) bcc are shown and viewed in the direction of [1 0 0], [1 1 0] and [1 1 1].



(1 0 0) zone (1 1 0) zone (1 1 1) zone

Figure 4.2: A 2 nm tungsten spherical crystal model (lattice constant is 316.52 pm) is generated and is shown here. There are 258 atoms in this crystal model. The crystal contains one vacancy site (not visible here) and random fluctuations from their equilibrium positions (the displacement is drawn from a Gaussian distribution having a width of 0.1 Å). This is the crystal model that is used to demonstrate the principle of AST.

These mathematical projections do not consider the effects of dynamical scattering and aberrations from electron optics. This is particularly valid in the current model as the crystal is very small (2 nm in diameter) while dynamical scattering usually becomes significant with a specimen thickness greater than 10 nm [15]; lens aberrations such as astigmatism are accounted for by using an elliptical function to represent its atomic shape. Therefore, the tilt series is sufficiently representative of an actual tomography experiment that it can be used to verify the AST algorithm.

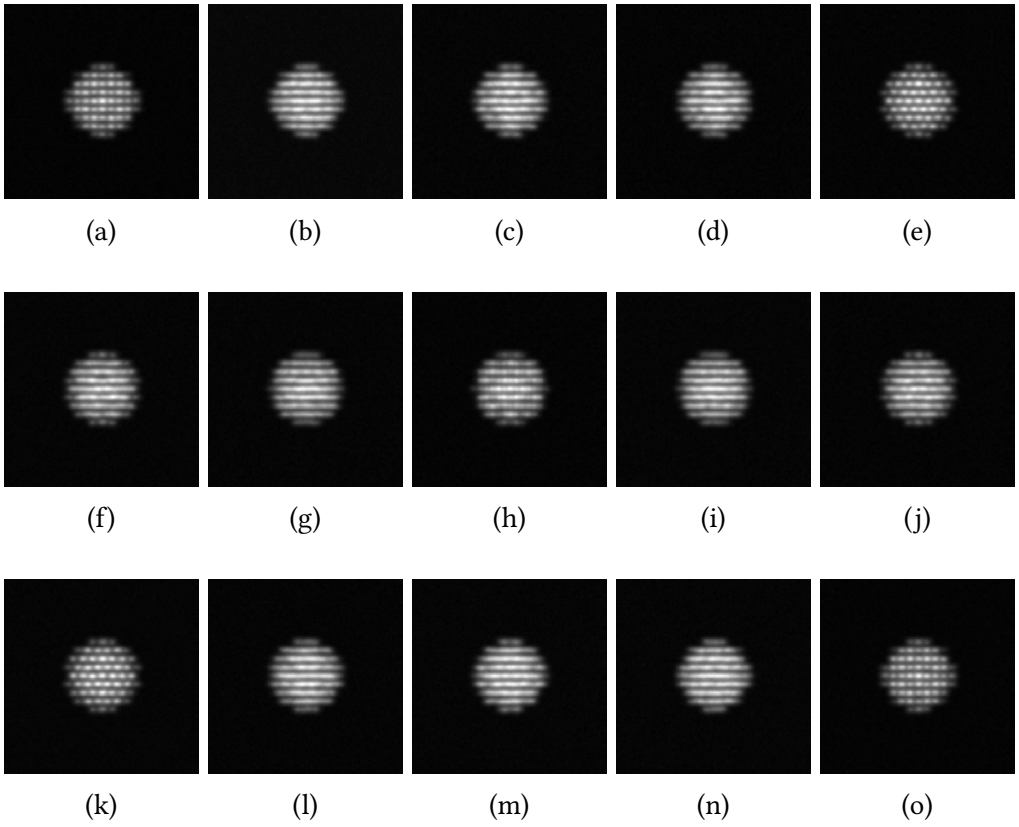


Figure 4.3: Projections of the crystal model in Figure 4.2 are simulated at different tilt angles. The tilt axis is along the [1 1 0] direction of the crystal. The 15 tilt angles are evenly distributed between -90° and 90° and the respective images are arranged in order from (a) to (o). The size of the projections is 128×128 pixel with a pixel size of 0.0405 nm.

4.2 Algorithm

AST is a tomography reconstruction algorithm that recovers the coordinates of identical entities by multiple linear regression. For a nanoparticle, the entities are the atoms.

A 'search grid' is constructed to cover the full range of the nanoparticle (see Section 4.2.1) — this is called the *support* of the nanoparticle. Initial positions of the atoms are found by least absolute shrinkage and selection operator (lasso) (see Section 4.2.2). A constraint can be enforced on the remaining grid points by using the knowledge of the minimal nearest neighbour distance of atoms of

a given element in a solid. The locations of these points are further refined by splitting them into 'cross-hairs ' (see Section 4.2.3).

This section will go through the steps in AST using the simulated dataset in Figure 4.3 generated from a spherical tungsten (bcc) crystal model (see Figure 4.2).

4.2.1 Search grid

The search grid is a set of coordinates in three-dimensional space and the associated weights correspond to the relative likelihood of the presence of an atom. The search grid is a close-packed grid with the nearest neighbour distance three times smaller than that of the crystal to be reconstructed — if the nearest neighbour distance of the nanoparticle is d, the lattice constant of the fcc search grid is then $\sqrt{2}d/3$, giving a nearest neighbour distance of $d/3$. Figure 4.4 shows the construction of a search grid in AST.

The nearest neighbour distance of a material is well known, and the size of the search grid should be slightly larger than the expected size of the crystal. By inspecting the acquired images, it is straightforward to gain an estimate of the support of the nanoparticle. The only knowledge to construct the search grid is the nearest neighbour distance and the estimated size of the crystal to be reconstructed, and these are readily available.

4.2.2 First iteration

AST is essentially a multiple linear regression problem (see Section 1.4):

$$\boldsymbol{y} = \beta_0 + \boldsymbol{x}\boldsymbol{\beta} \tag{4.2}$$

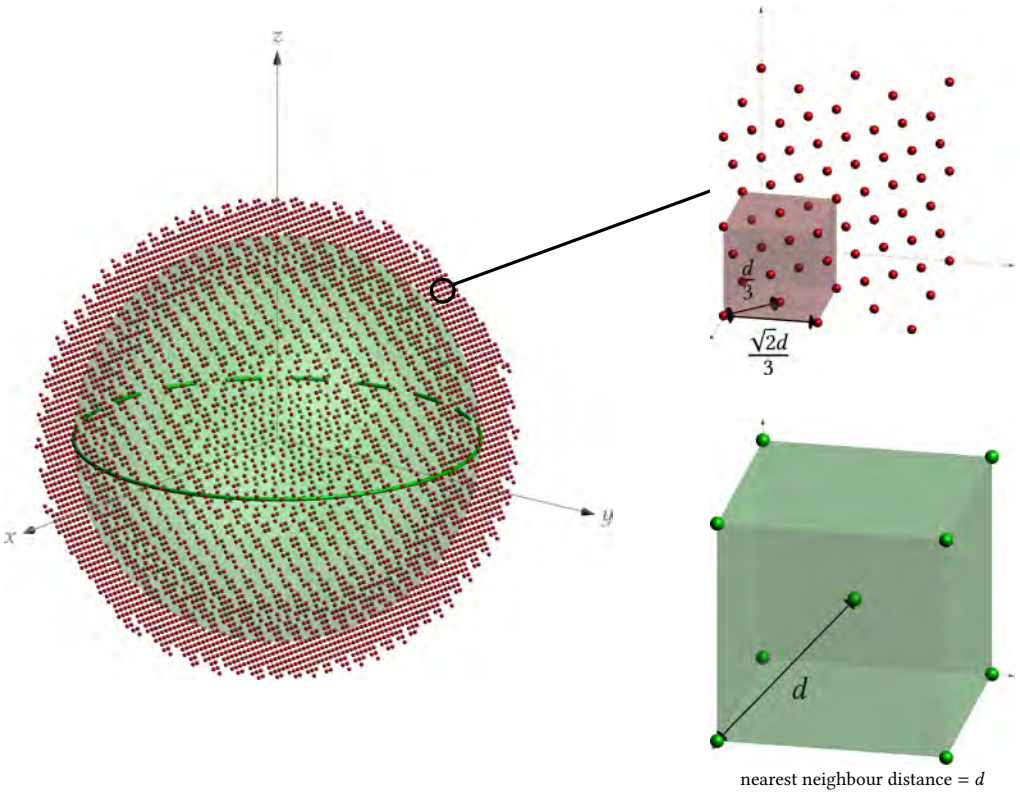


Figure 4.4: The search grid is chosen to be slightly larger than the size of the expected crystal. The region containing the expected crystal is shaded in green, while the points of the search grid outside the region of the expected crystal are shown in red. The grid points of the search grid inside the green region are not shown for clarity. The search grid is an fcc close-packed structure with the nearest neighbour distance three times smaller than that of the expected crystal; if the expected crystal has a nearest neighbour distance of d, that in the search grid is $d/3$. The unit cells of the search grid and the expected crystal are shaded by red and green respectively.

$\boldsymbol{y}$ is the observation and this is the collection of all pixel intensities in the set of experimental images taken (the sinogram in Figure 4.3). The experimental sinogram is reshaped into a $N \times 1$ column matrix, where $N = n_x \times n_y \times n_{\text{projections}}$ (n_x and n_y are the numbers of pixels along the horizontal and vertical dimension of the projections respectively, and $n_{\text{projections}}$ is the number of projections in the sinogram).

$\boldsymbol{x}$ is the predictor vector and each single predictor is the intensity factor of an atom at its corresponding point in the search grid constructed in Section 4.2.1. A sinogram for every point in the search grid is obtained under the same tilt angle

series as the experimental dataset. The sinograms from the positions in the search grid are generated from mathematical functions representing the atomic shapes. These are not known in real-world experiments (see Section 5.1.1 for methods to find them), but they are precisely known in simulation so they are used directly.

All sinograms generated from the search grid are put into a $N \times p$ matrix, where p is the number of predictors/features; in other words, p is the number of grid points in the search grid. This is the *design matrix* in multiple linear regression. The column k ($k = 1, 2, 3, \ldots, p-1, p$) in the design matrix contains the sinogram of the point k in the search grid.

$\boldsymbol{\beta}$ is the coefficient for the predictors/features and it is a $p \times 1$ column matrix. β_0 is the background of the experimental sinogram. The multiple linear regression in matrix form can be written as:

$$\underbrace{\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}}_{\text{sinogram}} = \underbrace{\begin{pmatrix} 1 & x_{11} & x_{12} & x_{13} & \ldots & x_{1p} \\ 1 & x_{21} & x_{22} & x_{23} & \ldots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & x_{n3} & \ldots & x_{np} \end{pmatrix}}_{\text{search grid design matrix}} \underbrace{\begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \vdots \\ \beta_{p-1} \\ \beta_p \end{pmatrix}}_{\text{coefficients}} \tag{4.3}$$

Each entry in $\boldsymbol{y}$ is a pixel intensity of an image taken at a particular tilt angle and it can be written as:

$$y_i = \beta_0 + x_{i1}\beta_1 + x_{i2}\beta_2 + \ldots + x_{ip}\beta_p \tag{4.4}$$

Multiple linear regression is used to solve the following minimisation problem (see Equation (1.38)):

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \|\boldsymbol{y} - \beta_0 - \boldsymbol{x}\boldsymbol{\beta}\|^2 \tag{4.5}$$

The solution of $\boldsymbol{\beta}$ must be sparse as this is how the search grid is constructed — the nearest neighbour distance of the search grid is three times smaller than that of the crystal. ℓ_1-regularisation is particularly suitable for solving a multiple linear regression problem with sparse coefficients [130]. The minimisation problem of ℓ_1-regularisation is:

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \|\boldsymbol{y} - \beta_0 - \boldsymbol{x}\boldsymbol{\beta}\|^2 + \alpha \|\boldsymbol{\beta}\|_{\ell_1} \tag{4.6}$$

AST uses lasso (pathwise coordinate descent) to solve the minimisation problem in Equation (4.6). Lasso promotes a sparse solution and the sparsity is controlled by the regularisation constant α (see Section 1.4.1).

The first lasso iteration uses cross-validation to determine which α gives the best fit; in other words, a range of α is tried and the α which gives the lowest least squares value is selected. Each point in the search grid is associated with a coefficient representing the relative likelihood of a presence of an atom. As the solution to the search grid is sparse, most of the entries in $\boldsymbol{\beta}$ are zero (see Figures 4.5a and 4.5c).

There is only one atom within a spherical shell having a diameter of the minimum nearest neighbour distance of the crystal (see Figure 4.5e). All local maxima within this spherical shell are identified; the local maxima are currently the most likely coordinates of atoms. To refine the locations of the atoms and reduce the number of grid points, centres of masses are calculated around all local maxima using all the grid points within this spherical cell. The coefficients

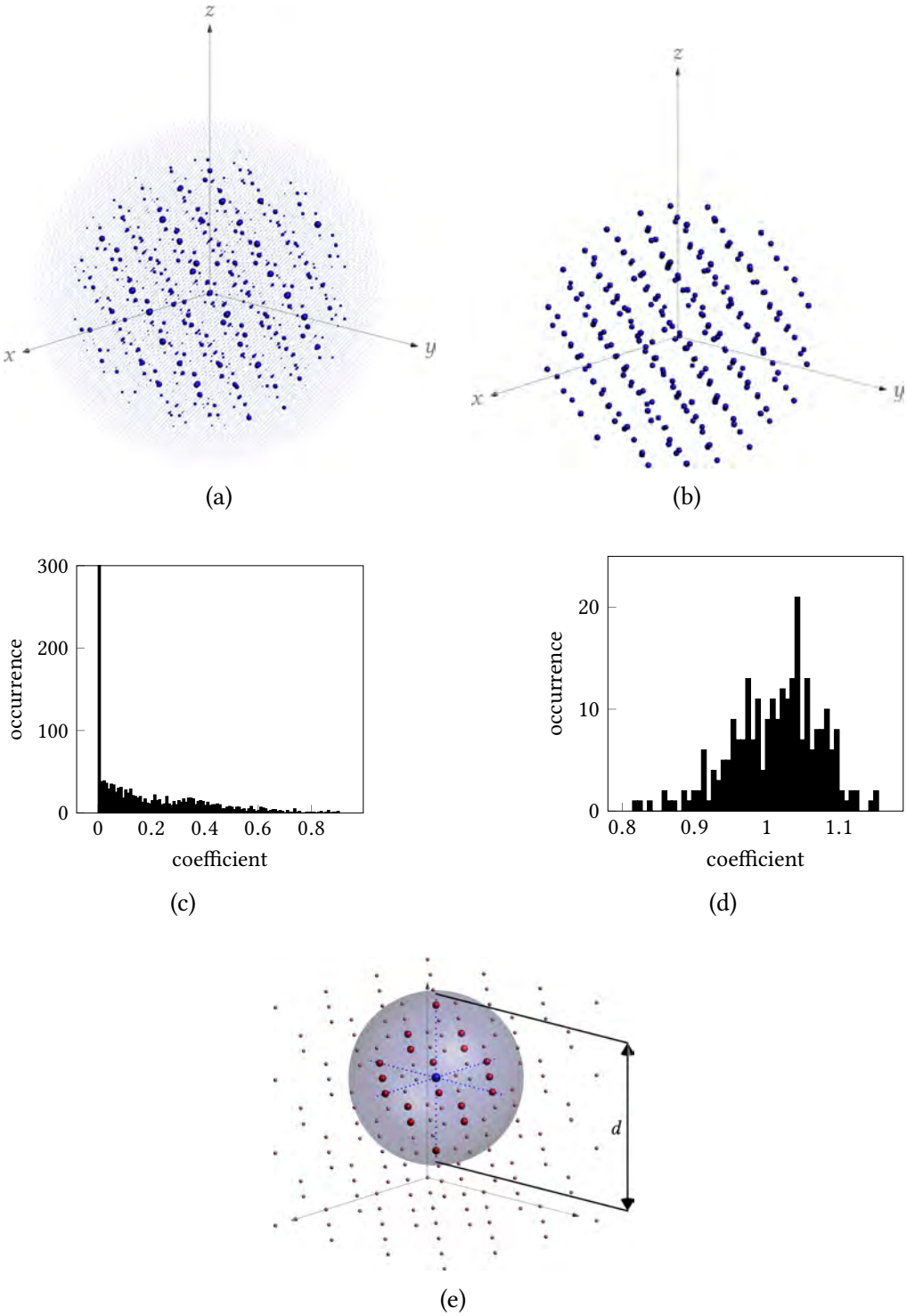


Figure 4.5: The initial positions of the atoms are found by a lasso fit on the search grid. (a) and (b) show the search grid after the first lasso iteration and their centres of masses respectively. Their distributions of coefficients are shown in (c) and (d) respectively. (e) shows a spherical shell (shaded in blue) with a diameter of the nearest neighbour distance d from the expected crystal. The local maximum is shown in blue and the centres of masses are calculated using all the emphasised points within the shell.

of the local maxima are summed and Figures 4.5b and 4.5d show the identified centres of masses and the distribution of coefficients.

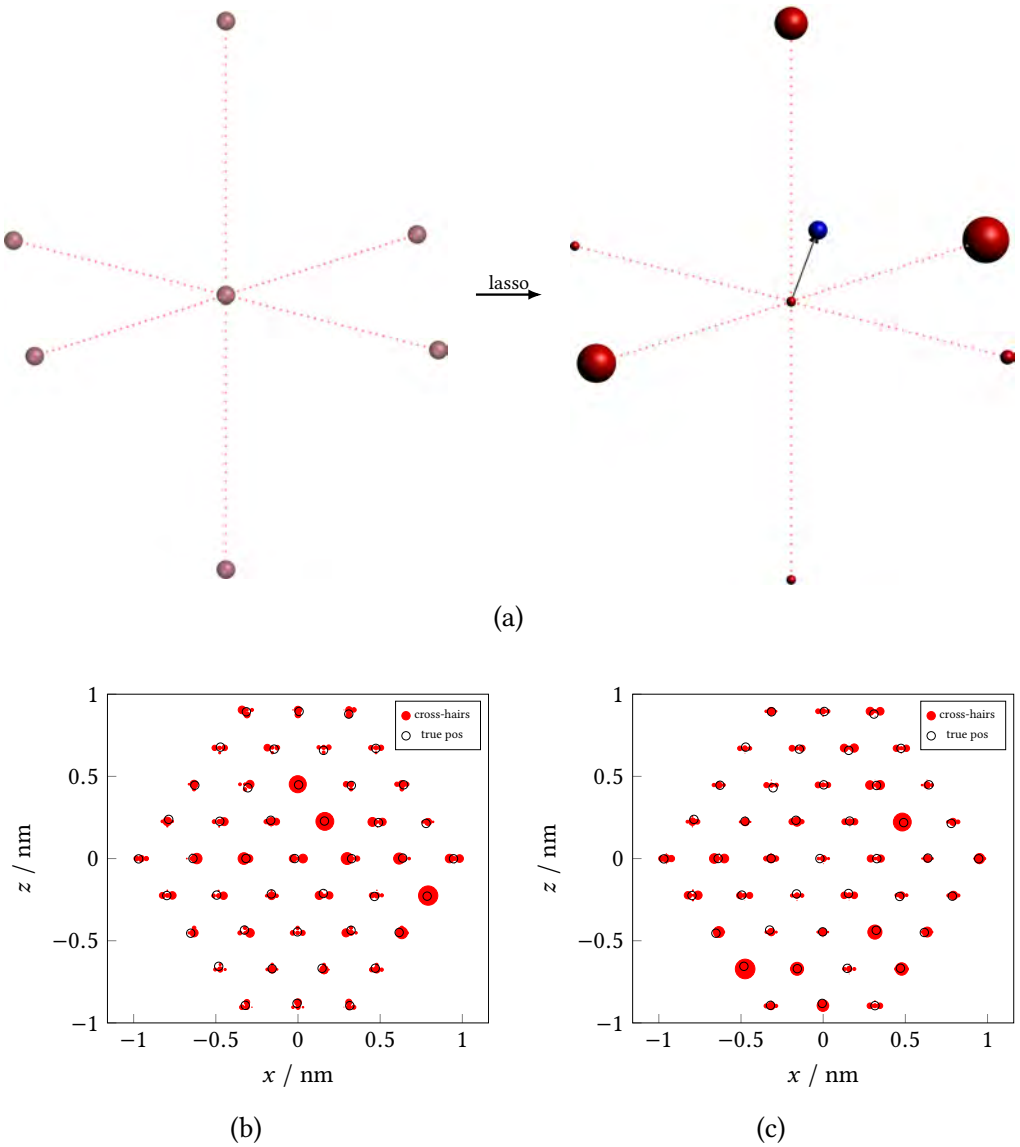
4.2.3 Refinement

The centres of masses obtained in Section 4.2.2 are, at this stage, the most probable locations of atoms. Those coordinates are split into 'cross-hairs' (see Figure 4.6). The cross-hairs can be regarded as a new 'search grid' and another lasso regression is performed on the following system:

$$\underbrace{\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}}_{\text{sinogram}} = \underbrace{\begin{pmatrix} 1 & x_{11} & x_{12} & x_{13} & \dots & x_{1q} \\ 1 & x_{21} & x_{22} & x_{23} & \dots & x_{2q} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & x_{n3} & \dots & x_{nq} \end{pmatrix}}_{\text{cross-hairs design matrix}} \underbrace{\begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \vdots \\ \beta_{q-1} \\ \beta_q \end{pmatrix}}_{\text{coefficients}} \tag{4.7}$$

The regularisation parameter of lasso in the cross-hairs fit is set at a very small value (of the order of 10^{-6}) as no feature selection is needed in this stage. After the fitting, every point in all cross-hairs is associated with a coefficient (see Figure 4.6). Centres of masses are calculated within each cross-hairs using the newly obtained coefficients. Another lasso is performed on this set of centres of masses as in Equation (4.7) and the above procedures are repeated. The centres of masses are 'moving' closer and closer to the positions of the atoms (see Figures 4.6b and 4.6c).

Figure 4.7 shows the distance of each cross-hairs moved during each iteration. As they are approaching the true positions of the atoms, the distances moved



(a)

(b)

(c)

Figure 4.6: The locations of atoms are refined by 'cross-hairs '. (a) shows the construction of cross-hairs. Centres of masses from the previous lasso iteration are split into cross-hairs and these cross-hairs are the new 'search grid'. After lasso on all cross-hairs, every point on cross-hairs is associated with a coefficient. Centres of masses of all cross-hairs are determined and the above steps are done iteratively. The new centre of mass in this cross-hairs is shown in blue. (b) and (c) show the movements of the cross-hairs in two-dimensional projections on the central slice of the crystal model of two consecutive iterations. The cross-hairs can be seen moving towards the expected locations of the atoms.

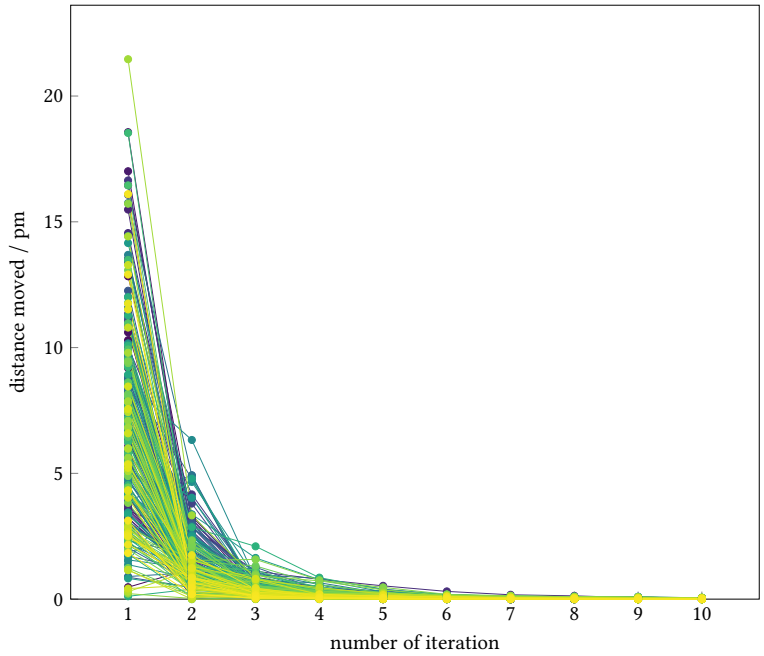


Figure 4.7: The movements of centres of masses of all cross-hairs are shown for 10 iterations. After every iteration, new centres of masses are determined from cross-hairs with associated coefficients from lasso. The movements of centres of masses are less and less as they are approaching the true positions of atoms.

become less and less. During the iterations, if two centres of masses are having a separation smaller than the nearest neighbour distance of the crystal, the centre of mass with a smaller coefficient is removed.

When the positions of centres of masses are being adjusted during the iterations, their coefficients are also changing as the likelihood of being an atom is also changing. If a coefficient is below a certain threshold value, that centre of mass is removed as the likelihood representing the presence of an atom is too low. The variations of the coefficients during the iterations are shown in Figure 4.8 for the simulated bcc crystal.

The final set of centres of masses is regarded as the reconstruction; they represent an atomic model of the crystal. The reconstruction only uses 15 tilt angles (see Figure 4.3). The reconstruction of the simulated bcc crystal is shown in Figure 4.9 and the distances between atoms of the reconstruction and the actual

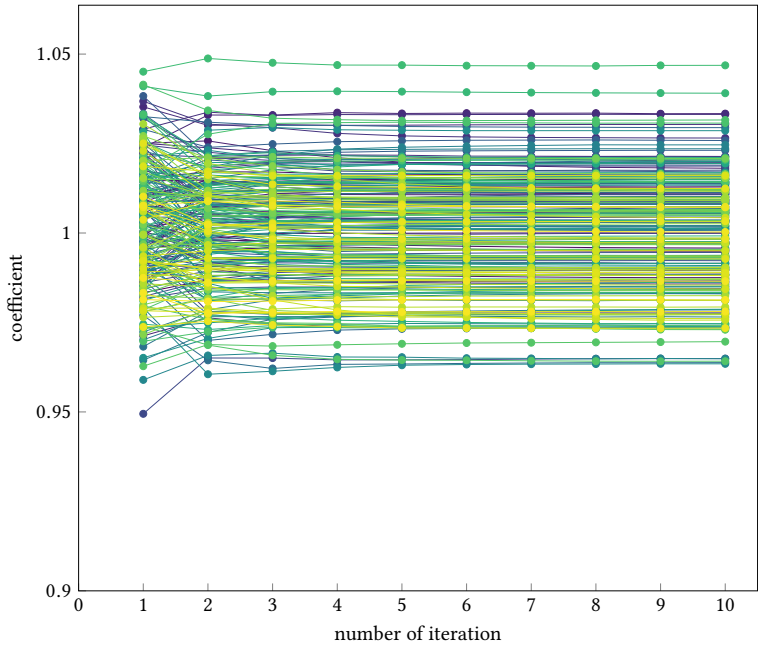
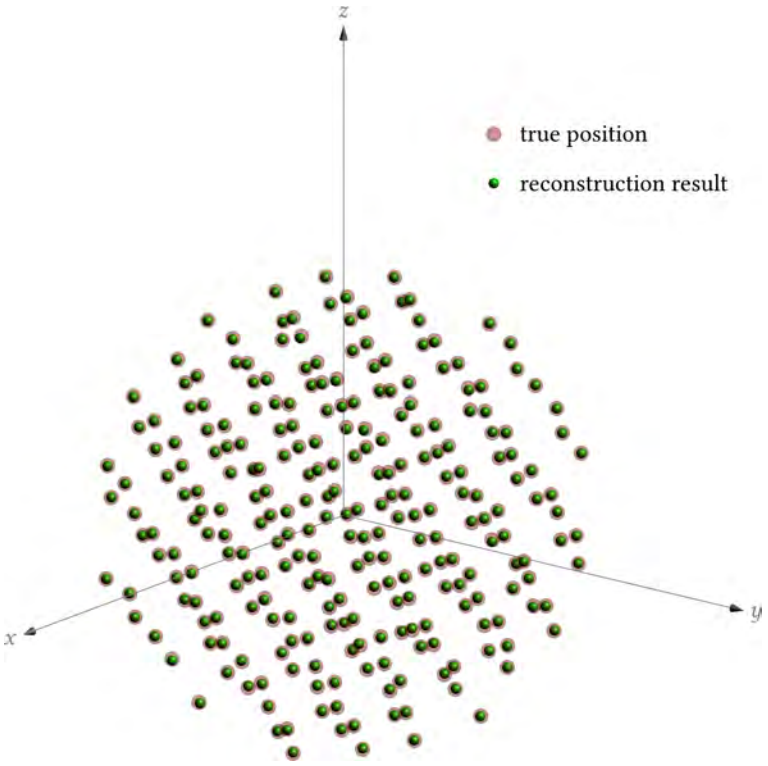


Figure 4.8: The variations of coefficients of all centres of masses of cross-hairs are shown for 10 iterations. Most of the coefficients are almost unchanged after 5 iterations as the locations of the centres of masses are closer and closer to the true positions of atoms. The more the movement of a centre of mass, the more drastic change of its coefficient.

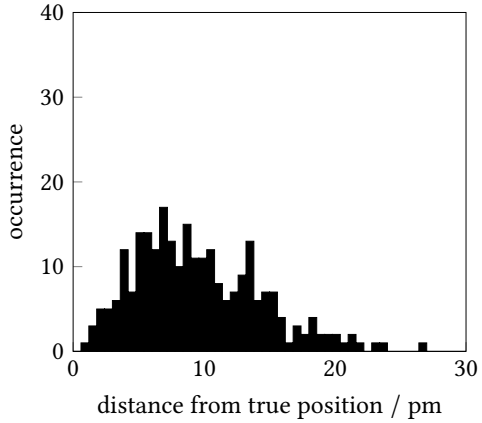
model are plotted. The average distance from the reconstruction to the actual model is ≈4 pm and this is below the resolution of high-resolution electron microscopes. There is no intrinsic limit on the precision as the centres of masses move in any direction in a real three-dimensional space.

4.3 Performance

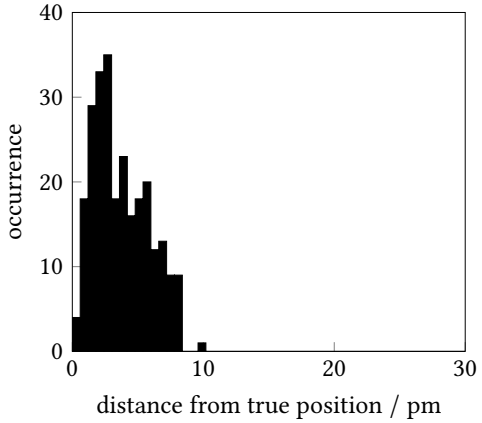
The algorithm of AST is demonstrated in Section 4.2 by using simulated data showing that AST is able to recover the crystal model accurately. In contrast to many state-of-the-art tomography reconstruction techniques such as EST and GENFIRE, AST does not require iterations between real and Fourier space; AST uses multiple linear regression with ℓ_1-regularisation in real space only. To compare the performance of AST to existing techniques, a script was written to implement the commonality of EST and GENFIRE — the iterations between



(a)



(b) 1st iteration



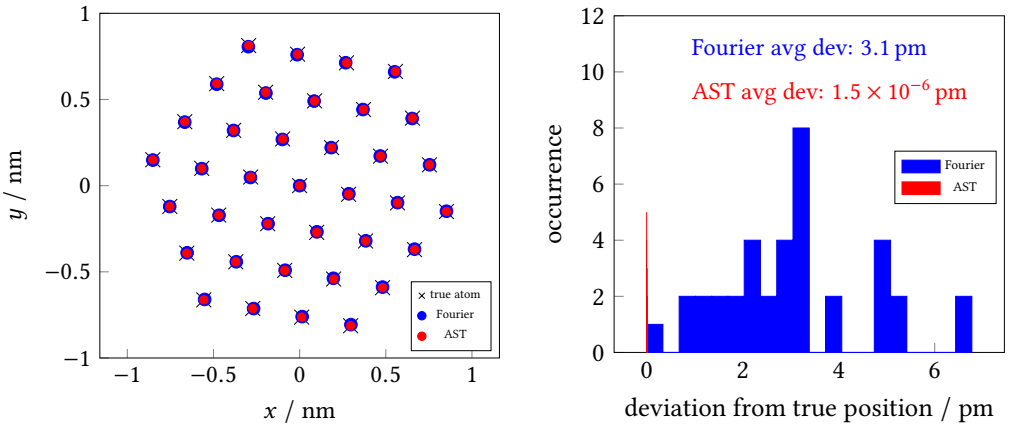
(c) last iteration

Figure 4.9: The reconstruction of the simulated dataset from the bcc crystal in Figure 4.2 by AST is shown in (a). The reconstruction almost overlaps the true positions of the crystal model, and the differences from the true positions in the first and last iteration are shown in (b) and (c) respectively.

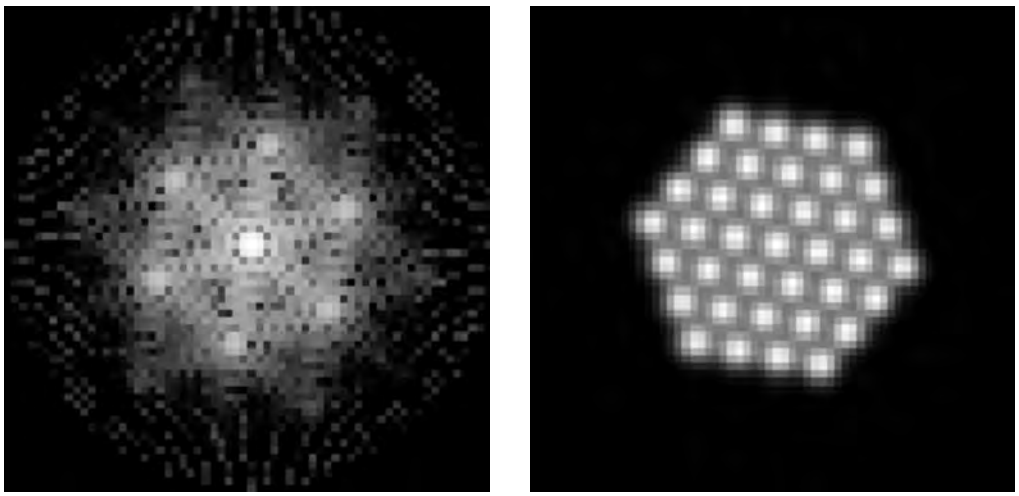
the real and Fourier space and the application of constraints in real and Fourier space (see Section 1.3.2 and Section 1.3.3). This is termed the 'Fourier-based algorithm' in this section.

For the sake of clarity, the comparison was done by using a single (1 1 1) layer of an fcc crystal with a lattice constant of 407.82 pm (the crystal structure of gold). The layer is tilted by 20° from the horizontal axis. The diameter of the crystal model is 2 nm, without any vacancy sites and the atoms are not deviated from their equilibrium positions (see Figure 4.10a). The sinogram is generated by using a single bivariate Gaussian function with $\sigma_X = \sigma_Y = 0.085$ nm, with a pixel size of 0.052 nm and without any Poisson noise. To demonstrate how the Fourier-based algorithm works, 61 tilt angles regularly spaced in ±90° are used and they are shifted by 20° to ensure projections of low-index zone axes are included. The reconstructions from both the Fourier-based algorithm and AST are shown in Figure 4.10a together with the true atomic positions. As the tilt series includes projections of low-index zone axes and there are sufficient number of projections, the deviations from the true atomic positions of the reconstructions from Fourier-based algorithm and AST are tiny (see Figure 4.10b). The larger average deviation of the Fourier-based algorithm could be attributed to some missing information due to discrete intervals of tilt angles; nonetheless, the periodicity of the crystal, the most important information, is sufficient to give an accurate crystal model. The final Fourier-space intensity map and its density map from the Fourier-based algorithm are shown in Figure 4.10c and Figure 4.10d respectively. As the projected intensity of an atom is accurately known, the atoms can be identified in the final density map (see Figure 4.10a).

In this section, AST and the Fourier-based algorithm are compared by using different tilt series with the crystal model in Figure 4.10a. The number of iterations between real and Fourier space is 100 for all the Fourier-based reconstruction and the number of refinement steps (see Section 4.2.3) is 10 for all the AST reconstruction. For AST, the maximum number of iterations and the convergence tolerance in lasso are set at 10^6 and 10^{-10} respectively. Two types



(a) (b)



(c) (d)

Figure 4.10: The reconstructions of a single layer of atoms are shown here. (a) the true positions of the atoms are shown together with the results from the Fourier-based algorithm and AST. Both reconstructions overlap with each other. (b) the deviations from the true atomic positions of the reconstructions from the Fourier-based algorithm and AST are shown here. The deviations of AST result are very small and the histogram appears as a single line near 0 pm. (c) the final Fourier-space intensity map (after 100 iterations) in the Fourier-based algorithm is shown here. The size of the Fourier-space intensity map is $9.31\,\mathrm{nm}^{-1} \times 9.31\,\mathrm{nm}^{-1}$ and it is shown in logarithmic scale. (d) the final density map is obtained from the inverse Fourier transform of (c) with real-space constraints. The positions of atoms are determined from the density map and shown in blue in (a). The size of the density map is $3.33\,\mathrm{nm} \times 3.33\,\mathrm{nm}$ with 64×64 pixels (the pixel size is 0.052 nm).

of tilt series will be considered: tilt series with very few projections and tilt series with missing wedges. The reconstruction from both AST and the Fourier-based algorithm assume the accurate knowledge of the projected intensities of atoms.

4.3.1 With very few projections

It is desirable to take as few projections as possible in a tomography experiment to minimise the damage from the electron beam. 'Very few' projections here is in respect of ordinary AET experiments in which more than 60 projections are normally taken [11, 12, 100]. The performance of AST and the Fourier-based algorithm are compared by using only 13 projections distributed evenly over the full ±90° angular range. Figure 4.11a shows the reconstructions from AST and the Fourier-based algorithm using 13 projections which hit major low-index zone axes and Figure 4.11b shows their deviations from the true atomic positions. The corresponding Fourier-space intensity map and density map of the last iteration are shown in Figures 4.12a and 4.12b. Both techniques can successfully recover the crystal model, with the crystal model from AST being more accurate.

Figures 4.11c and 4.11d show the reconstructions and their deviations from the true atomic positions by using the above 13 tilt angles shifted by 2°. The corresponding Fourier-space intensity map and density map of the last iteration are shown in Figures 4.12c and 4.12d. This tilt series misses the major low-index zone axes by 2° and the result from the Fourier-based algorithm deviates from the true positions much more than that of AST. The deviations from true positions of the Fourier-based algorithm result become significant when the 13 tilt angles are shifted by 5° (see Figures 4.11e, 4.11f, 4.12e and 4.12f).

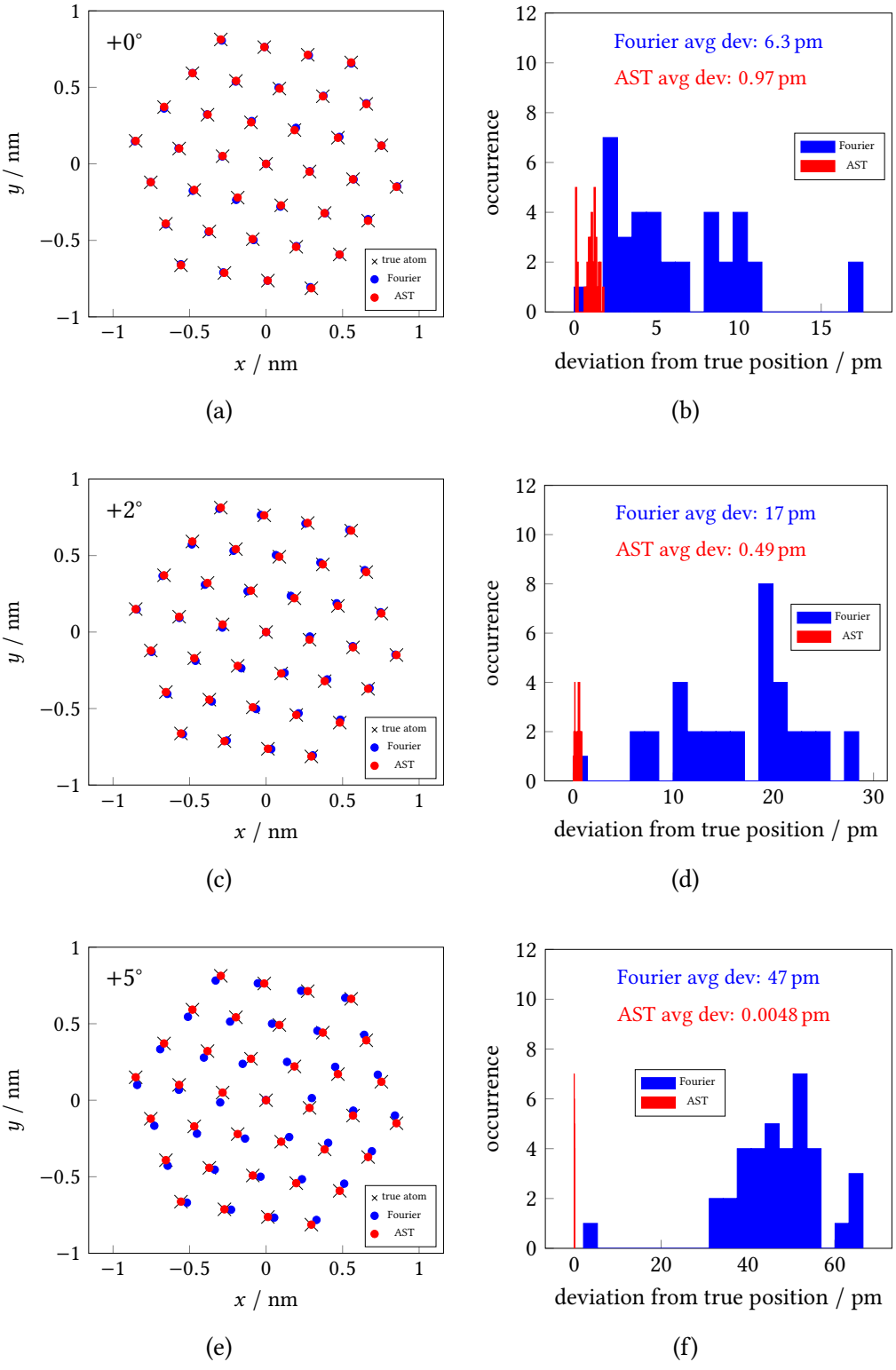
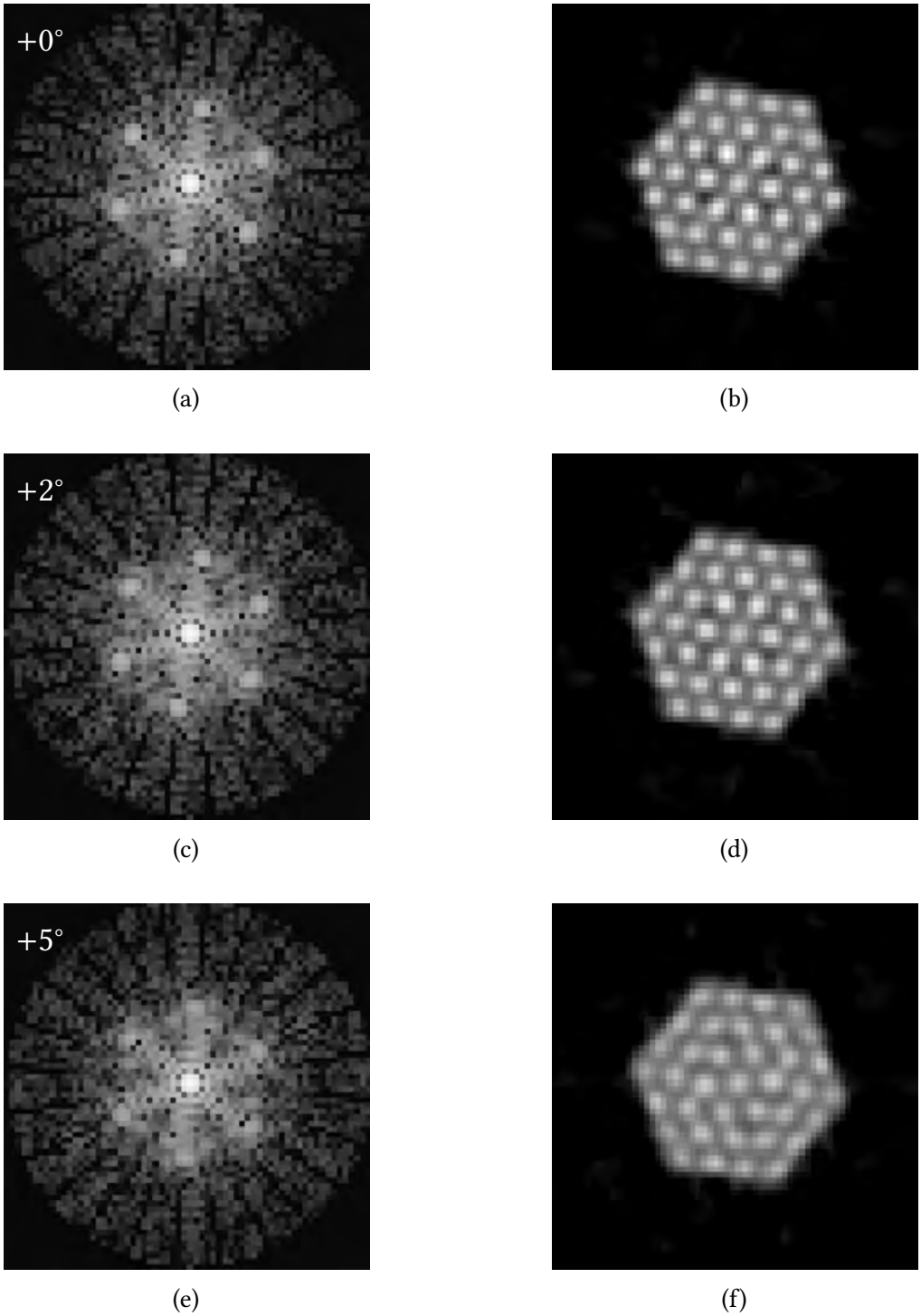


Figure 4.11: The reconstructions by AST and the Fourier-based algorithm with different angular shifts from zone axes are shown here. (a), (c) and (e) are the reconstructions, and (b), (d) and (f) are their deviations from true atomic positions using 13 tilt angles which hit the major zone axes, shifted by 2° and 5° respectively.



(a) (b)

(c) (d)

(e) (f)

Figure 4.12: The final iteration of Fourier-space intensity maps and density maps from the Fourier-based algorithm using 13 tilt angles are shown here. The positions of atoms are determined from the density maps in (b), (d) and (f) for 13 tilt angles hitting major zone axes, shifted by 2° and 5° respectively, and they are shown in Figure 4.11. The size of the Fourier-space intensity maps is $9.31\,\mathrm{nm}^{-1} \times 9.31\,\mathrm{nm}^{-1}$ and they are shown in logarithmic scale. The size of the density maps is $3.33\,\mathrm{nm} \times 3.33\,\mathrm{nm}$ with 64×64 pixels (the pixel size is $0.052\,\mathrm{nm}$).

AST can still successfully locate the atoms when the tilt angles are shifted away from zone axes, while the Fourier-based algorithm shows a systematic 'spiral-like' deviation (see Figures 4.11e and 4.12f). This 'spiral-like' deviation is a consequence of systematic shifts of tilt angles away from the zone axes and AST is robust to this deviation.

4.3.2 With missing wedge

The presence of a missing wedge in a tomography dataset can severely degrade the quality of the reconstruction. It results in an elongation of features along the direction of the missing wedge in the reconstruction (see Figures 1.14c and 1.14d). For a single-axis tilting geometry and samples supported on grids, the maximum tilt angular range is around ±70° (see Section 1.2.2); therefore, the reconstructions from AST and the Fourier-based algorithm are compared for 61 tilt angles regularly spaced between ±70°, ±60° and ±50°.

The Fourier-based algorithm performs satisfactorily for angular ranges ±70° and ±60° (see Figures 4.13a to 4.13d and Figures 4.14a to 4.14d). For ±50°, the deviations from true atomic positions are significant (see Figures 4.13e, 4.13f, 4.14e and 4.14f); AST is rather insensitive to the increasing size of the missing wedge and it is able to recover the crystal model consistently. The discrepancies of the true atomic positions and the reconstructions from the Fourier-based algorithm are the consequence of missing the information on the zone axes and the deviations are along the direction of the missing wedges — this is most apparent in Figures 4.13e and 4.14f. AST is robust to missing wedges and does not necessarily require the full angular range ±90° for a faithful reconstruction.

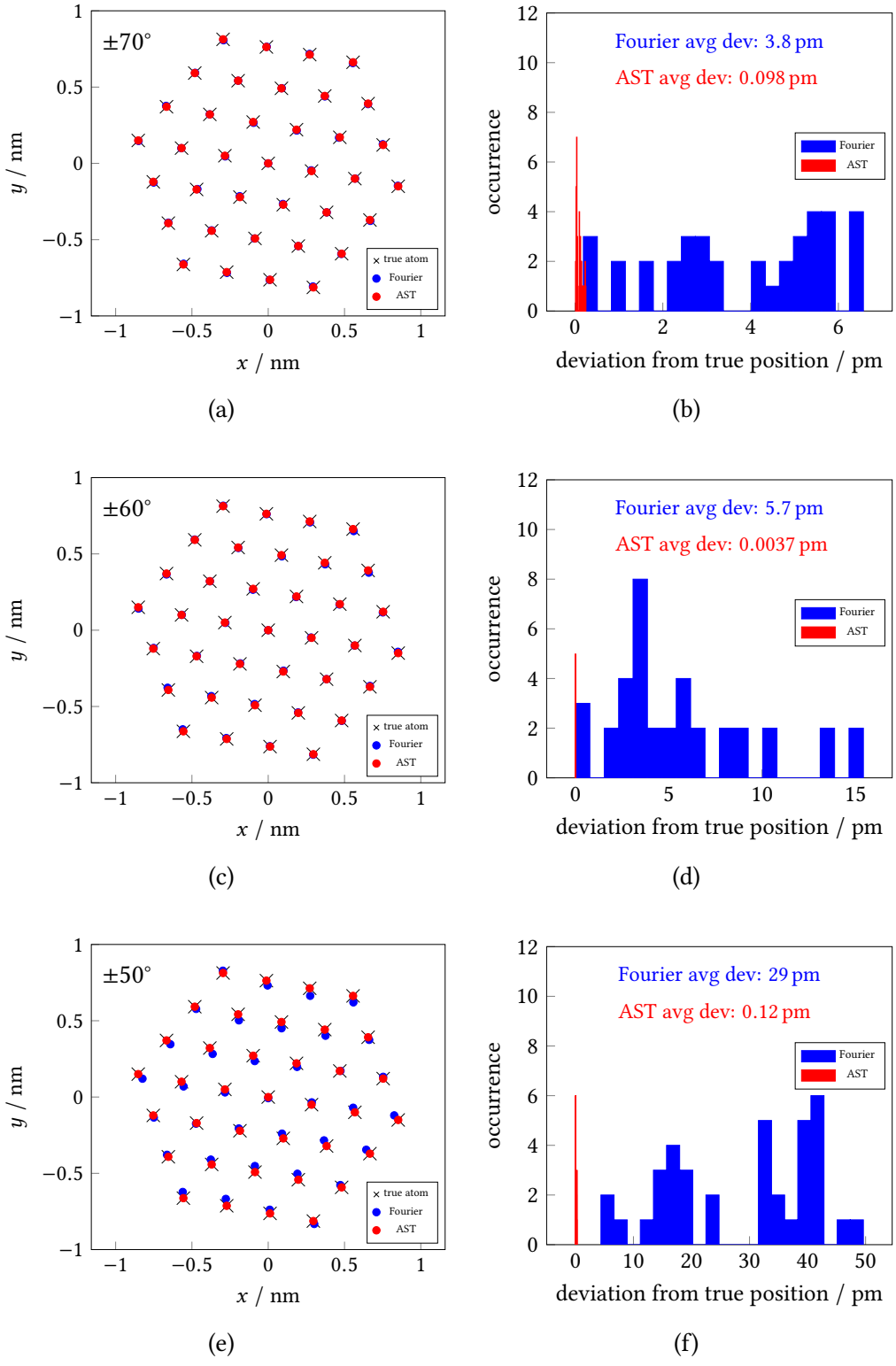


Figure 4.13: The reconstructions by AST and the Fourier-based algorithm with different missing wedges are shown here. (a), (c) and (e) are the reconstructions, and (b), (d) and (f) are their deviations from true atomic positions using 61 tilt angles evenly distributed in the angular range of ±70°, ±60° and ±50° respectively.

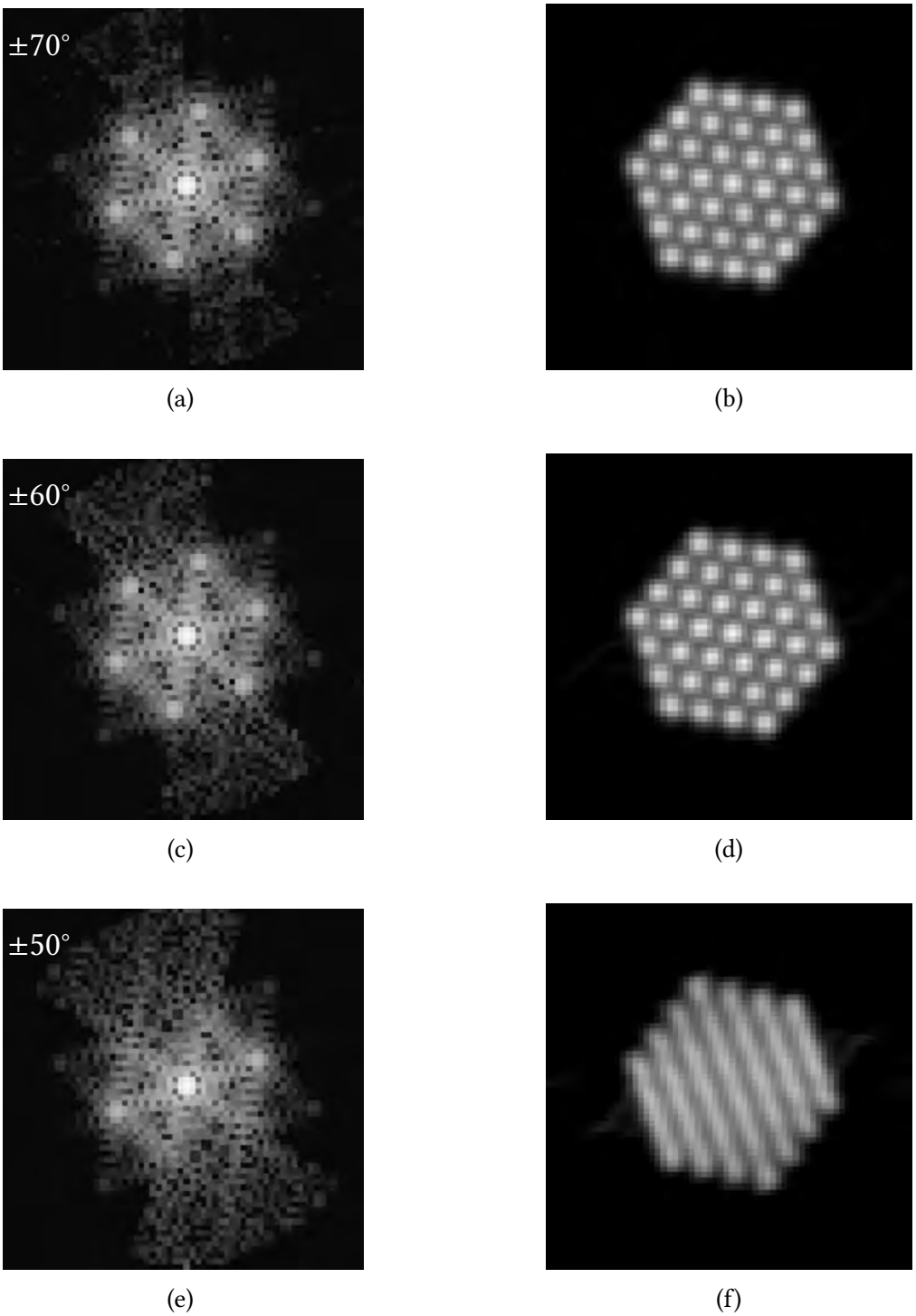


Figure 4.14: The final iteration of Fourier-space intensity maps and density maps from the Fourier-based algorithm with different missing wedges are shown here. The positions of atoms are determined from the density maps in (b), (d) and (f) for 61 tilt angles evenly distributed in ±70°, ±60° and ±50° respectively, and they are shown in Figure 4.13. The size of the Fourier-space intensity maps is $9.31\,\mathrm{nm}^{-1} \times 9.31\,\mathrm{nm}^{-1}$ and they are shown in logarithmic scale. The size of the density maps is $3.33\,\mathrm{nm} \times 3.33\,\mathrm{nm}$ with 64×64 pixels (the pixel size is 0.052 nm).

4.4 Discussions

AST, the application of ℓ_1-regularisation in multiple linear regression in tomography reconstruction, is shown to be superior to state-of-the-art techniques, which involve iterations between real and Fourier space and application of respective constraints. The Fourier-based techniques such as EST and GENFIRE attempt to retrieve missing information by iterating between real and Fourier space; however, the performance of these methods in AET suffer from artefacts if there are only limited number of projections or a large missing wedge, and it is not trivial to optimise the parameters during reconstruction (see Section 1.3.2 and Section 1.3.3). AST performs ℓ_1-regularisation directly on the final crystal model with the experimental projections, while the Fourier-based techniques recover the final density map of the crystal model first and then the atomic positions are determined from the density map. The ℓ_1-regularisation is solely performed in the real space, and there is no interpolation from the Fourier space to the real space (see Figure 1.8). In AET, AST seeks the best crystal model that produces all the experimental projections by performing multiple linear regression with ℓ_1-regularisation. It is much more robust to datasets that have very few projections or a missing wedge, as it never tries to 'retrieve' those missing information. The pathwise coordinate descent algorithm in ℓ_1-regularisation (see Section 1.4.1) minimises the differences between projections generated from the crystal model and the experimental projections.

The application of compressed sensing and data-driven approaches to tomography reconstruction provide the possibility of using only a few projections. However, it relies primarily on the prior knowledge of the sample and it is not easy to account for structural deviations from expected lattices such as vacancy sites (see Section 1.2.3). The only assumption in AST is that the projections are coming from atoms (*atomic site*); hence, AST is particularly suitable for

AET. AST provides the best crystal model using all the available projections, and it does not impose limits on where these projections are taken; it does not necessarily require projections from zone axes, unlike discrete tomography (see Section 1.3.1).

However, there are two limitations of AST. The number of points in the initial search grid (Section 4.2.1) could be very large and it grows cubically with the size of the expected crystal. The current script that implements the pathwise coordinate descent algorithm in *scikit-learn* (0.19.2) [136] in Python limits the non-zero entries in the design matrix to $(2^{31} - 1)$, the maximum positive value for a 32-bit signed binary integer. This limits the size of the search grid, so reconstruction of large crystalline nanoparticles (>5 nm in diameter) are done by splitting the sinogram slab by slab — the reconstruction is done only on a section of the sinogram. The number of entries can also be decreased by binning the sinogram. The memory to hold the design matrix of the search grid typically exceeds the capacity of ordinary desktop computers as a design matrix in a sparse representation is ≈25 GB for $(2^{31} - 1)$ non-zero 64-bit floating-point numbers. In view of this, most of the reconstruction by AST in this work were carried out on computer clusters (BlueBEAR, University of Birmingham). Even when the memory issue in AST is addressed by splitting or binning the sinogram, AST might not be suitable for reconstructing large nanoparticles (>20 nm) as it assumes the projections are from linear combinations of individual atoms and this is not valid when dynamical scattering becomes significant.

AST works best if the mathematical functions representing the projected shapes of atoms are accurately known; however, these information are not available in real-world experimental data. Approximations of the projected atomic shapes can be found by comparing simulated images from a model with the experi-

mental sinogram (see Section 5.1.1). To conclude, AST provides a promising approach to reconstruct tomography datasets and it is especially suitable for AET.

5

Reconstruction of experimental data

The principle of atomic site tomography (AST) is outlined and discussed in Chapter 4 by using a simulated dataset. This chapter validates the method of AST by using real-world experimental data. A tungsten tip dataset published in [12, 189] is selected for validation, and a gold nanoparticle dataset acquired by tungsten tip tomography during the course of this work is reconstructed by AST.

5.1 Tungsten tip reconstruction

There are five high-quality experimental datasets available in [189] for researchers to validate their reconstruction algorithms for atomic electron tomography (AET). A dataset of a tungsten tip acquired over a full 180° range using tilt angles defined by equally sloped tomography (EST) (see Section 1.3.2) is chosen, as the stability of tungsten (see Section 2.1) ensures the projections are consistent with each other. The dataset was acquired by procedures as outlined in Chapter 2 using tilt angles defined by EST, and a full projection from this dataset is shown in Figure 5.1. The results from EST can be found in [189]. In this section, a

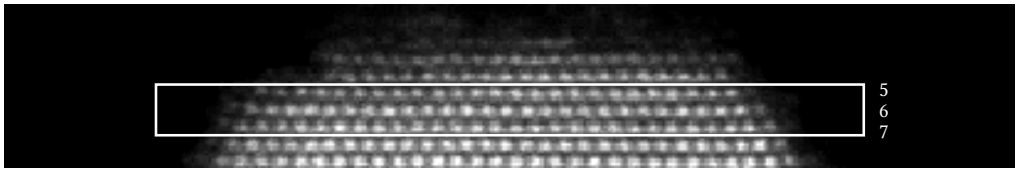


Figure 5.1: A zone-axis projection (at tilt angle 125.2°) from the tungsten tip dataset in [189] is shown here. There are 62 projections in this tilt series. The pixel size is 0.0405 nm and the size of the image is 2.19 nm × 13.41 nm. Layer 5, 6 and 7 of the dataset (highlighted by the box) is cut out for the validation of AST.

three-layer slab of the tungsten tip in the dataset is chosen to illustrate and validate the procedures of AST.

5.1.1 Atomic shapes

The simulation in Section 4.1 provides a set of functions and parameters to generate mathematical projections from atomic structures in the sinogram, and the same set of functions and parameters are used in the AST reconstruction. This ensures that features in the design matrix (see Section 4.2.2) are true representations of the sinogram. The correct choices of functions and parameters representing the atomic shapes are essential for an accurate AST reconstruction. A wrong mathematical representation of an atom can result in extra or missing atoms in the reconstruction as illustrated in Figure 5.2. This is because the least absolute shrinkage and selection operator (lasso) is trying to minimise the regularised residuals (see Equation (1.39)) using the provided features.

However, there is no prior knowledge of the projected atomic shapes for experimental datasets. Therefore, an atomic model inside the support of the nanoparticle is constructed to estimate the atomic shapes, and a reconstruction is done to refine the atomic model using the newly determined atomic shapes. This is done iteratively until the atomic model and shapes are self-consistent. As the traced positions of individual tungsten atoms from EST are available

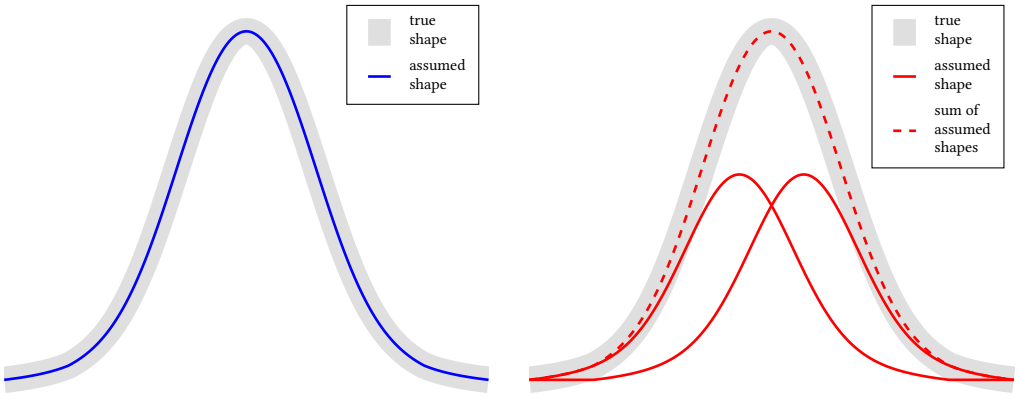


Figure 5.2: The consequence of not having a correct mathematical representation of a projected atom is illustrated here. In (a), the assumed atomic shape matches the true shape of the atom, so one atom is correctly identified; in (b), the assumed atomic shape does not match the true atom and it requires two assumed atoms to fully represent the true atom. This results in two atoms being identified while there is only one true atom. The atom appears to be 'split'.

in [189], the above steps are skipped and the traced positions are used to generate projections using the same tilt series as in the experiment.

The parameters in the mathematical functions are adjusted to minimise the differences between the calculated projections from traced positions and the corresponding experimental projections. The same atom appears differently in different projections due to different foci and possible changes in the aberrations of the electron optics over the course of data acquisition. Table 5.1 lists the functions that are used for this purpose. Each projection is assigned a function and a set of parameter that have the smallest difference between the calculated projection and the corresponding experimental projection.

The projected intensities of a single atom from the mathematical functions smoothly decay to zero from the centre within a certain range. The region where the intensities of a single tungsten atom can be non-zero is chosen to be within a radius of 0.35 nm from its central position; the radius chosen should be at least larger than the nearest neighbour distance in the lattice, which

Table 5.1: Mathematical functions used to represent the intensities of projected atoms

(1) bivariate Gaussian	$(A, \sigma_X, \sigma_Y, \rho)$
$$\frac{A}{2\pi \left\lvert\Sigma\right\rvert^{\frac{1}{2}}} e^{-\frac{M_d^2}{2}}$$	
(2) two bivariate Gaussian	$(A, \sigma_{X1}, \sigma_{Y1}, \rho, B, \sigma_{X2})$
$$\frac{A}{2\pi \left\lvert\Sigma\right\rvert^{\frac{1}{2}}} e^{-\frac{M_{d_1}^2}{2}} + \frac{B}{2\pi \left\lvert\Sigma\right\rvert^{\frac{1}{2}}} e^{-\frac{M_{d_2}^2}{2}} \qquad \left(\frac{\sigma_{Y1}}{\sigma_{X1}} = \frac{\sigma_{Y2}}{\sigma_{X2}}\right)$$	
(3) bivariate Gaussian and Lorentzian	$(A, \sigma_X, \sigma_Y, \rho, B, \gamma_X)$
$$\frac{A}{2\pi \left\lvert\Sigma\right\rvert^{\frac{1}{2}}} e^{-\frac{M_{d_\sigma}^2}{2}} + \frac{B}{2\pi \left\lvert\Sigma\right\rvert^{\frac{1}{2}}} \left[1 + M_{d_\gamma}^2\right]^{-\frac{3}{2}} \qquad \left(\frac{\gamma_Y}{\gamma_X} = \frac{\sigma_Y}{\sigma_X}\right)$$	
(4) bivariate Gaussian with exponential decay	$(A, \sigma_X, \sigma_Y, \rho, \lambda)$
$$\frac{A}{2\pi \left\lvert\Sigma\right\rvert^{\frac{1}{2}}} e^{-\lambda \frac{M_d^2}{2\lambda + M_d}}$$	
(5) bivariate Gaussian with exponential decay and Lorentzian	$(A, \sigma_X, \sigma_Y, \rho, B, \gamma_X, \lambda)$
$$\frac{A}{2\pi \left\lvert\Sigma\right\rvert^{\frac{1}{2}}} e^{-\lambda \frac{M_{d_\sigma}^2}{2\lambda + M_{d_\sigma}}} + \frac{B}{2\pi \left\lvert\Sigma\right\rvert^{\frac{1}{2}}} \left[1 + M_{d_\gamma}^2\right]^{-\frac{3}{2}}$$	
M_d, Mahalanobis distance	$(\vec{x}, \vec{\mu}, \Sigma)$
$$\sqrt{(\vec{x} - \vec{\mu})^\intercal \Sigma^{-1} (\vec{x} - \vec{\mu})}$$	
Σ, covariance matrix	$(\sigma_X, \sigma_Y, \rho)$
$$\begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix}$$	

in a tungsten lattice is 0.274 nm. This is to ensure the projected atoms are sufficiently represented by the functions, while keeping the number of non-zero entries in the design matrix smaller than $(2^{31} - 1)$ so that a lasso fitting is feasible using the current version of *scikit-learn* (0.19.2) [136] (see Section 4.4). The projections of an atom generated by different mathematical functions of two tilt angles and a comparison between the generated and experimental sinograms are shown in Figure 5.3. The projections of individual atoms at different tilt

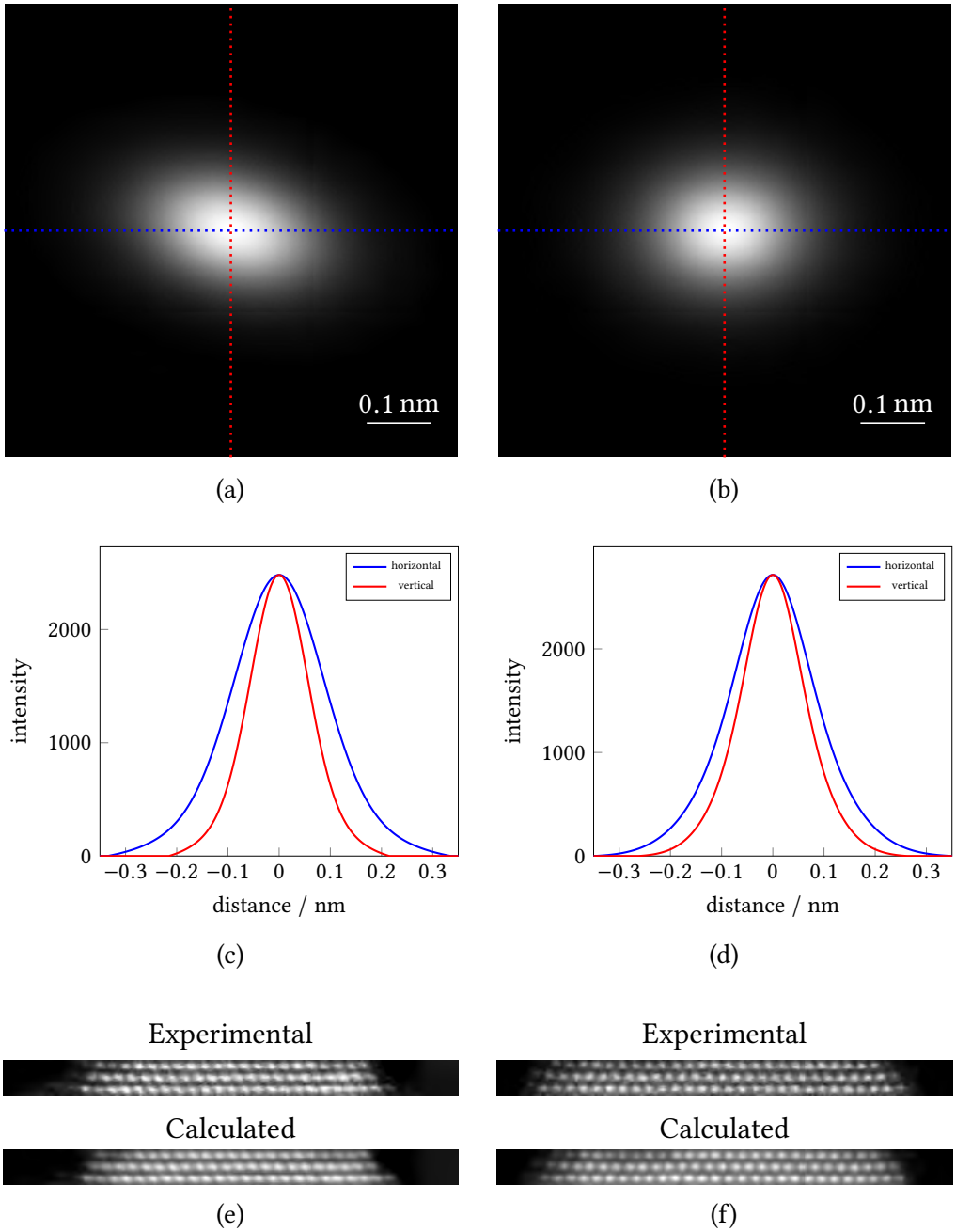


Figure 5.3: Two examples of the projected atomic shapes as represented by mathematical functions are shown here. (a), (c) and (e) show the two-dimensional shape, their horizontal and vertical line plots, and the calculated projection at tilt angle 55.6°; (b), (d) and (f) show the same information at tilt angle 125.2°. The function for the tilt angle 55.6° and 125.2° are function (3) and (2) in Table 5.1 respectively. The size of the projections is 0.69 nm × 8.95 nm.

angles could be elliptical ($\sigma_X \neq \sigma_Y$) and slanted ($\rho \neq 0$) to account for different foci and lens aberrations such as astigmatism (see Section 4.1).

The search of atomic shapes assumes all atoms contribute equally to the intensities in the projections regardless of their locations; this ignores, for example, any contribution from dynamical scattering (see Section 1.2.3). As the traced positions from EST are used to find the mathematical functions, the minimisation will be over-fitted — the mathematical functions will fit the noise in the projections. This is corrected by broadening the atomic shapes. In the current case, a broadening of 10 % is chosen. This makes the positions of the atoms less sensitive to the noise in the projections.

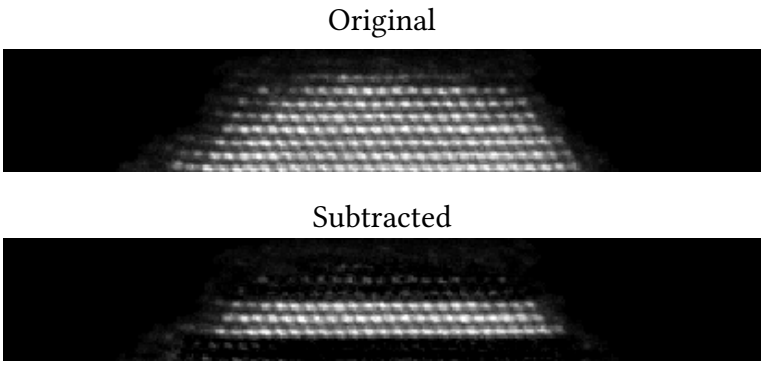
5.1.2 Input sinogram

There are nine atomic layers along the tomography axis (as determined by EST) in the tungsten tip dataset and the details of data acquisition can be found in [12]. The available dataset was aligned and the background was subtracted — the intensities outside the region of interest are zero. The dataset is also corrected for random drift and scanning-coil distortion.

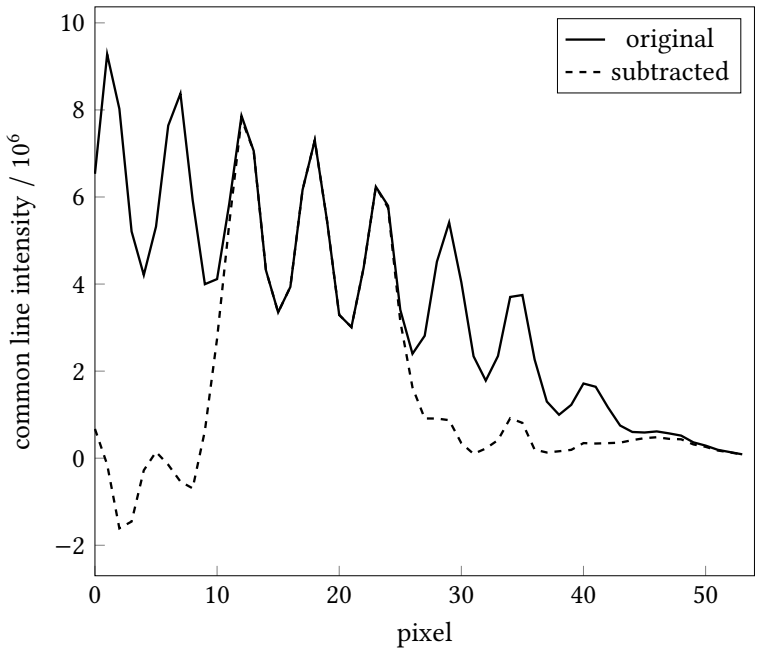
Layer 5–7 of the dataset are used for the validation of AST. The top section of the tungsten tip is not considered because surface atoms are always less stable [154, 155]; the bottom layer is not considered because the contributions from the layers further down are not known. All contributions from atoms other than the layer 5–7 are subtracted from the full experimental sinogram, using the traced positions and mathematical functions determined in Section 5.1.1. A projection before and after the subtraction are shown in Figure 5.4.

5.1.3 Initial fit

A search grid is constructed as described in Section 4.2.1. The bounds of the search grid can be approximated by the region of interest in the sinogram and it must cover the full volume which could have contributions to the sinogram.



(a)



(b)

Figure 5.4: Layer 5, 6 and 7 of the tungsten tip dataset are selected for AST reconstruction, and the contributions from neighbouring layers are subtracted as they are not reconstructed. The original and subtracted sinograms and their common line projections are compared in (a) and (b) respectively. The projections in (a) are taken from the tilt angle 55.6° with a dimension of 2.19 nm × 13.41 nm.

In experimental datasets, the atomic positions may be slightly deviated from their lattice points due to intercalated atoms and lattice strains so the minimum nearest neighbour distance could be smaller than the value in a perfect lattice [12].

In this analysis, the expected smallest nearest neighbour distance is taken to be 80 % of its theoretical value, $0.8 \times 0.274 = 0.219$ nm; 80 % is chosen because the expansion and compression as determined in [12] do not cause a deviation greater than 80 % from its equilibrium value. Using the estimated bounds from the sinogram and an expected smallest nearest neighbour distance of 0.219 nm, the search grid for reconstructing layer 5–7 contains 175 993 points and its design matrix in a sparse representation is ≈16 GB with 1 380 714 961 non-zero entries.

The relative importance of different pixels in the sinogram can be adjusted by applying different weights on the corresponding features in the design matrix; this changes the contribution to the overall residual sum when a multiple linear regression is performed. In the initial lasso fit of the search grid to the sinogram, the weights of pixels having zero intensities are increased to restrict the likelihood of atoms appearing on the background. The projections from major low-index zone axes are essential to provide information on the atomic lattice structure; therefore, the weights of pixels in those projections, 55.6°, 125.2° and 180.6°, are also increased. In the initial search grid lasso fit, the weights on pixels with zero intensities and low-index zone-axis projections in this analysis are 50 and 10 respectively. The weights for all other pixels in the sinogram are 1.

The optimal regularisation parameter α_o (see Equation (1.39)) is determined by cross-validation on the centres of masses after lasso fits of the search grid to the experimental sinogram. Different sets of centres of masses are determined (as described in Section 4.2.2) by using a range of α. Simple least squares fits (with positivity constraint) are performed on the experimental sinogram by using different sets of centres of masses determined from different values of α. The optimal α_o is chosen by selecting one which gives the highest adjusted R^2

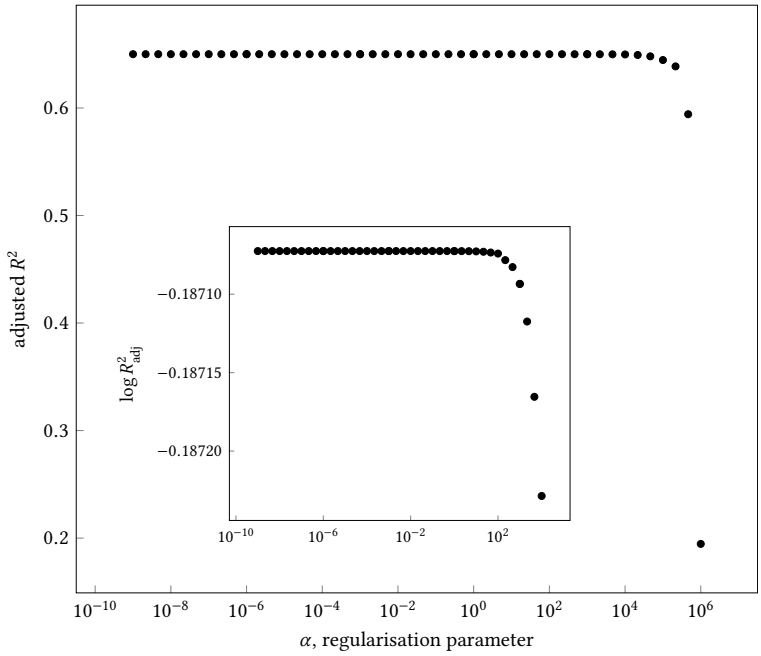
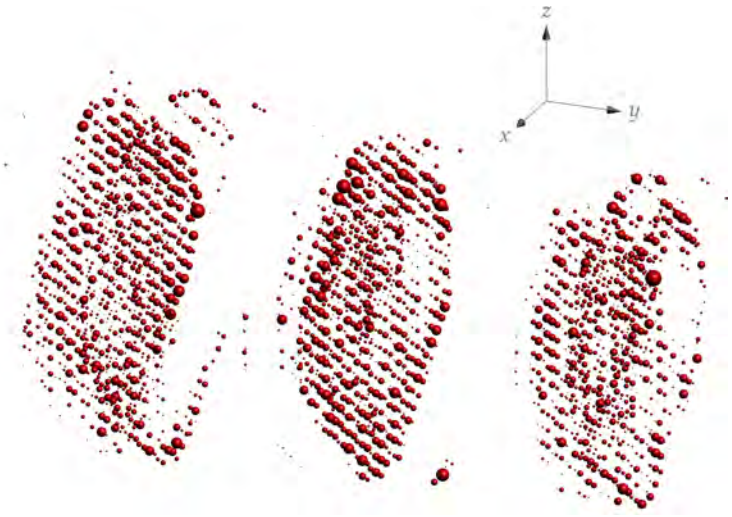


Figure 5.5: Different sets of centres of masses are determined from lasso fits of the search grid to the experimental sinogram using a range of regularisation parameters, and their centres of masses are fitted to the experimental sinogram by simple least squares. The variation of adjusted R^2 values from simple least squares fits with different sets of centres of masses is shown here. For the tungsten tip dataset, the adjusted R^2 values remain stable at 0.65 for $\alpha < 10^4$ (the change is less than 10^{-5} as shown in the inset).

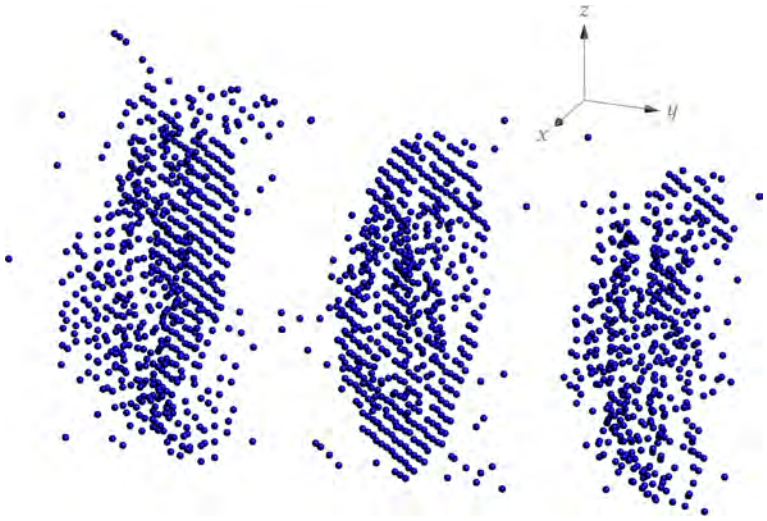
value. The adjusted R^2 value is defined as

$$R^2_{\text{adj}} = 1 - (1 - R^2)\frac{n - k}{n - k - 1} \tag{5.1}$$

R^2 is the coefficient of determination from the simple least squares fit, n is the total number of pixels of the experimental sinogram and k is the number of centres of masses. The adjusted R^2 value is used because it accounts for the number of centres of masses. Figure 5.5 shows the variation of adjusted R^2 values with a range of α. For the tungsten tip dataset, the adjusted R^2 values remain stable for $\alpha < 10^4$ so α_o is chosen to be 10^4 to speed up the initial lasso fit of the search grid (convergence is slower for a lower α value).



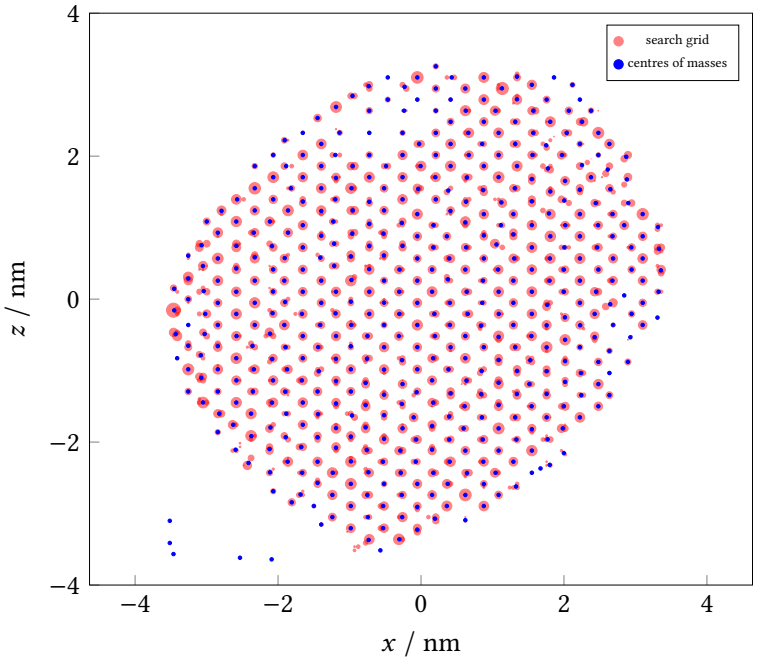
(a) search grid



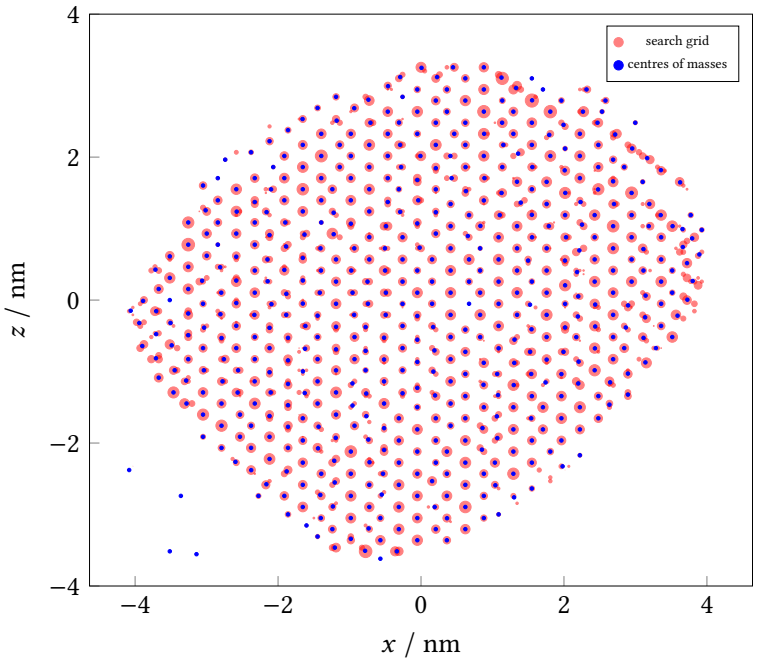
(b) centres of masses

Figure 5.6: The results of lasso fit of search grid to the experimental sinogram and after identifying centres of masses in three-dimensional space are shown in (a) and (b) respectively. For clarity, the y dimension is stretched by a factor of 20.

The results of the initial lasso fit of the search grid and after identifying their centres of masses are shown in three-dimension and two-dimensional projections of a single atomic layer in Figure 5.6 and Figure 5.7 respectively.

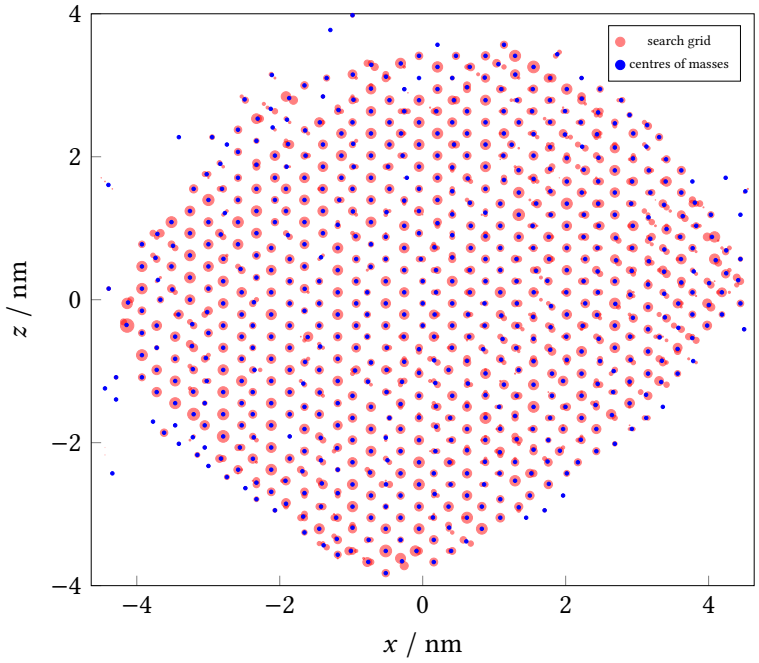


(a) layer 5



(b) layer 6

Figure 5.7: The results of initial lasso fit of the search grid and their centres of masses are shown in (a) and (b) for layer 5 and layer 6 respectively. The radii of the points for the search grid are proportional to the square root of the weights for clarity.



(c) layer 7

Figure 5.7 *(continued)*: The results of initial lasso fit of the search grid and their centres of masses are shown in (c) for layer 7. The radii of the points for the search grid are proportional to the square root of the weights for clarity.

5.1.4 Cross-hairs refinement

The centres of masses are currently the most probable locations of the atoms, and the positions can be refined by splitting them into cross-hairs as described in Section 4.2.3. During the movements of cross-hairs, their centres of masses may get too close to each other. If this happens, they are merged by calculating their common centres of masses. The merging happens when two or more points are within a certain distance.

The local maximum shell is progressively increased from 50 % to 60 % of the theoretical nearest neighbour distance; this ensures the final positions do not get closer than 60 % of the theoretical nearest neighbour distance. The movements of cross-hairs of a region in layer 6 are shown in Figure 5.8. After 10 iterations, the cross-hairs become stable.

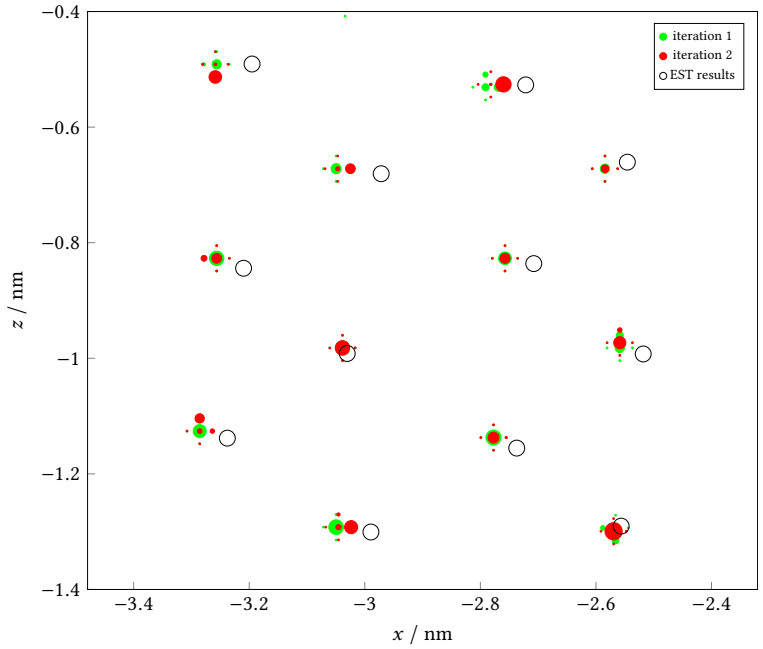
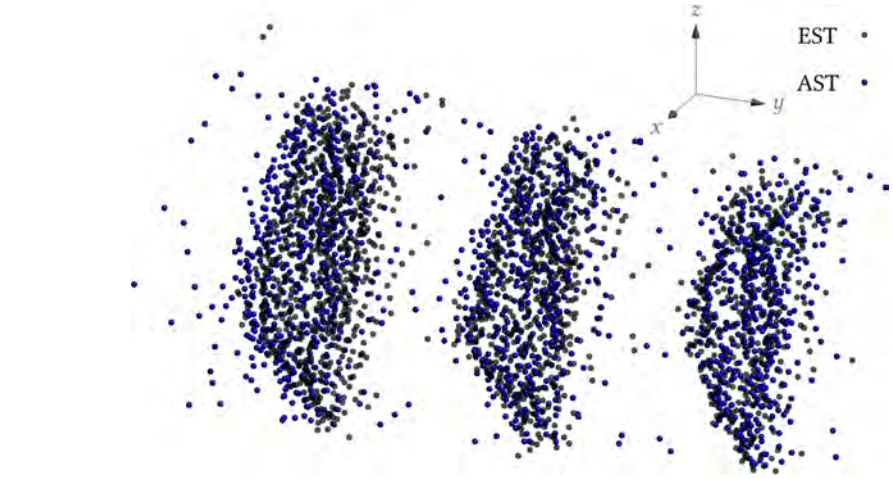


Figure 5.8: Two iterations of the refinement process are shown here to illustrate the movements of cross-hairs. The green dots are the cross-hairs one iteration before the red dots. The EST results are shown as hollow circles.

5.1.5 Results of reconstruction

The final positions of the centres of masses of cross-hairs are taken as the reconstruction. AST gives the total number of tungsten atoms as 1618 and EST, 1593. The coordinates of the reconstructions from AST and EST in three-dimensional space and their two-dimensional projections along the [1 1 0] direction of each layer are shown in Figure 5.9 and Figure 5.10 respectively. The reconstruction from AST agrees well with the EST reconstruction.

The differences between the results of AST and EST are mainly atoms at the surface, which are always challenging in AET. The surface atoms always carry some degree of uncertainty in reconstructions. The uncertainty is caused by the movements of the surface atoms induced by the electron beam during data acquisition and non-predictable intensity contributions from carbon contamination. This causes the projections to be inconsistent with each other and degrade the quality of the reconstruction.



(a)

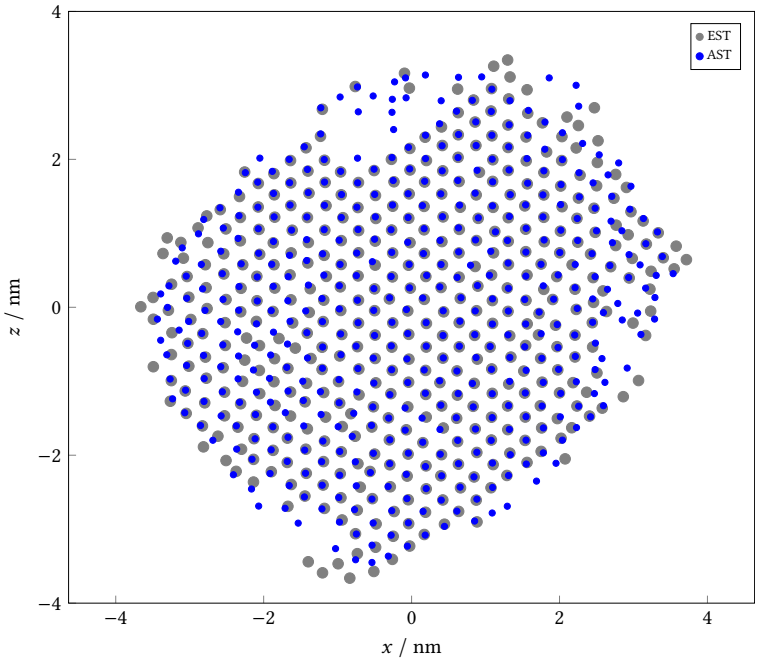
(b) along ⟨1 0 0⟩

(c) along ⟨1 1 0⟩

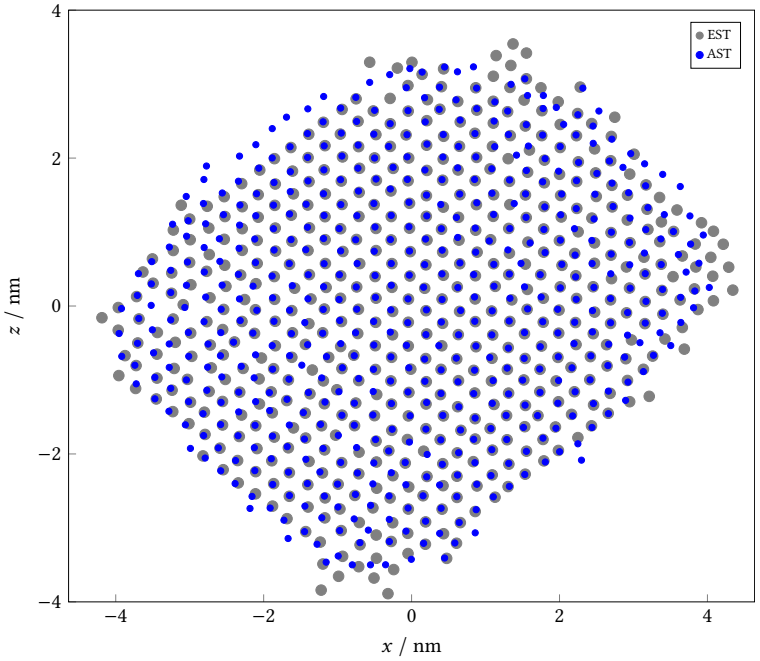
(d) along ⟨1 1 1⟩

Figure 5.9: The reconstructions from AST and EST are compared here. (a) shows their reconstructions in three-dimensional space. For clarity, the y dimension is stretched by a factor of 20. (b), (c) and (d) show their views along the ⟨1 0 0⟩, ⟨1 1 0⟩ and ⟨1 1 1⟩ direction respectively.

All reconstruction algorithms aim to provide a valid model for the sample of interest and a 'unique' solution to the sample does not exist. It is a solution if it does not violate some physical constraints (such as overlapping or isolated atoms) or prior knowledge of the sample. The model should also be consistent with the experimental projections, and this can be used to verify if the reconstruction is one of the valid models. Projections generated from the AST reconstruction using mathematical functions as found in Section 5.1.1 are com-



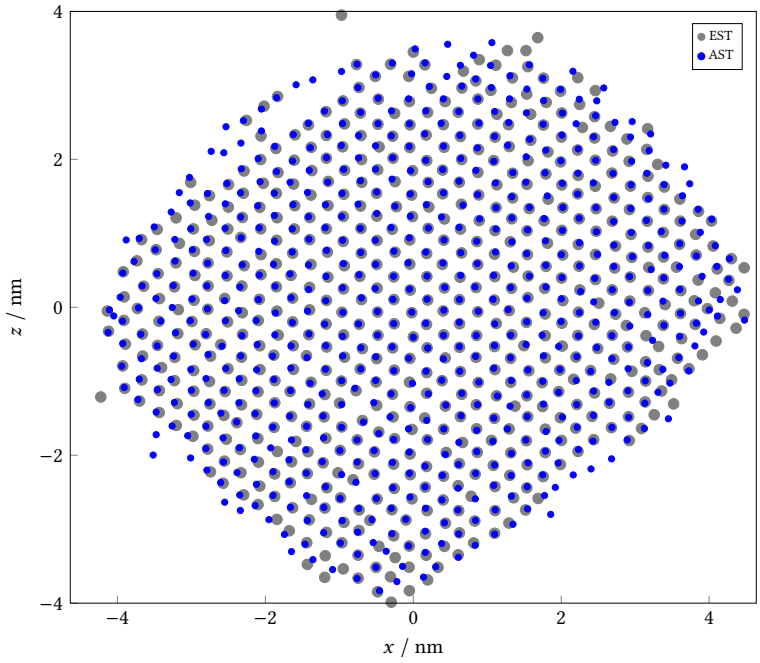
(a) layer 5



(b) layer 6

Figure 5.10: Comparison of projections along the [1 1 0] direction between the AST and EST results are shown here for layer 5 and layer 6, in (a) and (b) respectively.

pared with the corresponding experimental projections and they are shown in Figure 5.11.



(c) layer 7

Figure 5.10 *(continued)*: Comparison of projections along the [1 1 0] direction between the AST and EST results are shown here for layer 7 in (c).

A simple least squares fit (with positivity constraint) of the AST reconstruction to the experimental sinogram is performed and the weight distribution is shown in Figure 5.12. The adjusted R^2 value of the final AST result is 0.74, which is greater than the value from the positions of centres of masses from the search grid (0.65, see Figure 5.5). This proves that the cross-hairs refinement improves the accuracy of the reconstruction. The weight distribution of Figure 5.12 is centred around 1 (the mean and standard deviation of the distribution are 0.956 and 0.228 respectively) as each atom should contribute to the experimental projections equally; the deviation could be caused by dynamical effects during imaging, surrounding carbon contamination which contributes to the intensities in projections randomly, and noise in the experimental data.

The nearest neighbour distance distributions from the results of AST and EST are shown in Figure 5.13. Interstitial carbon atoms from electron beam contam-

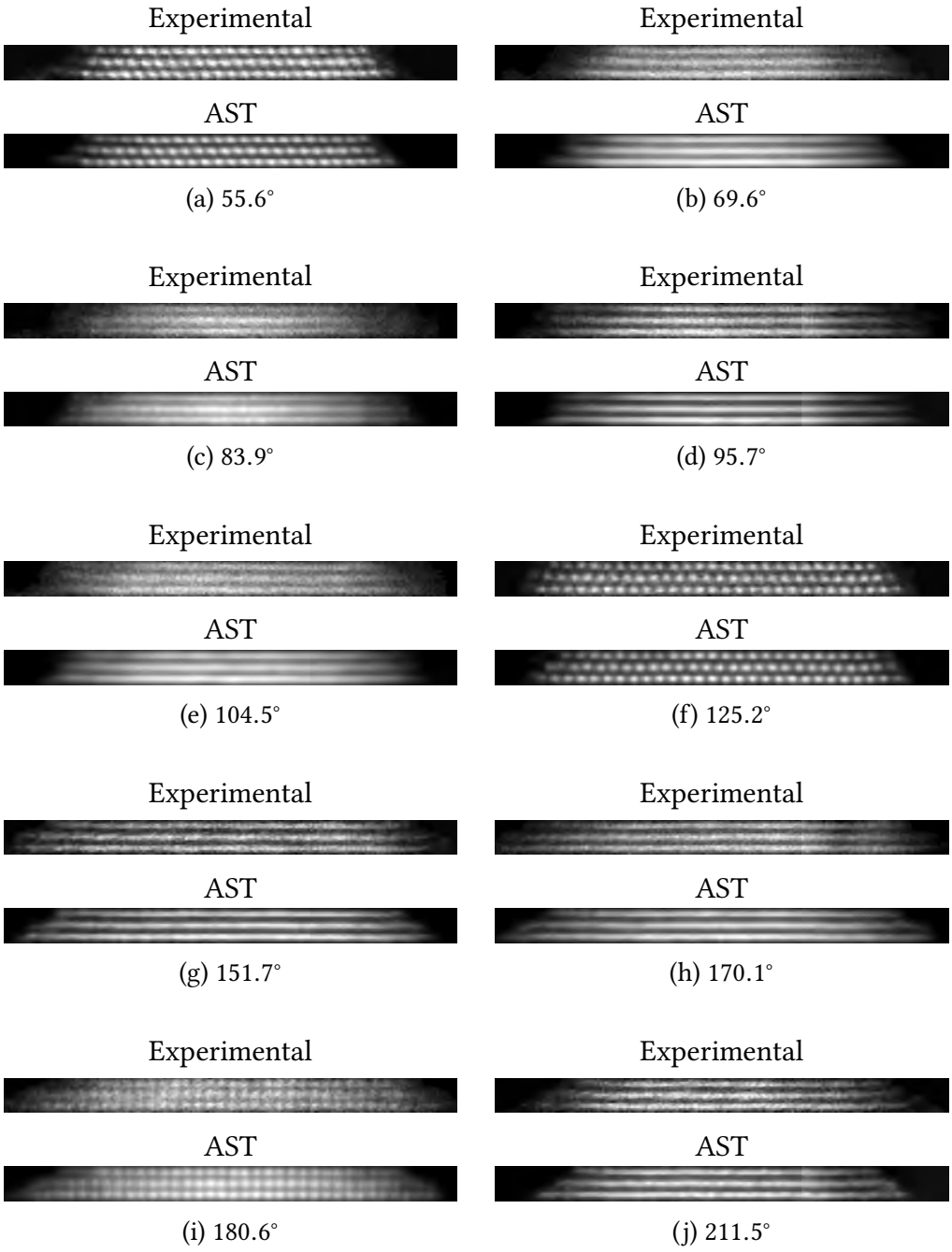


Figure 5.11: Selected projections from the experimental sinogram and the projections calculated from the results of AST are shown here. The consistency between these projections indicates the model from AST is valid. The size of the projections is 0.69 nm × 8.95 nm.

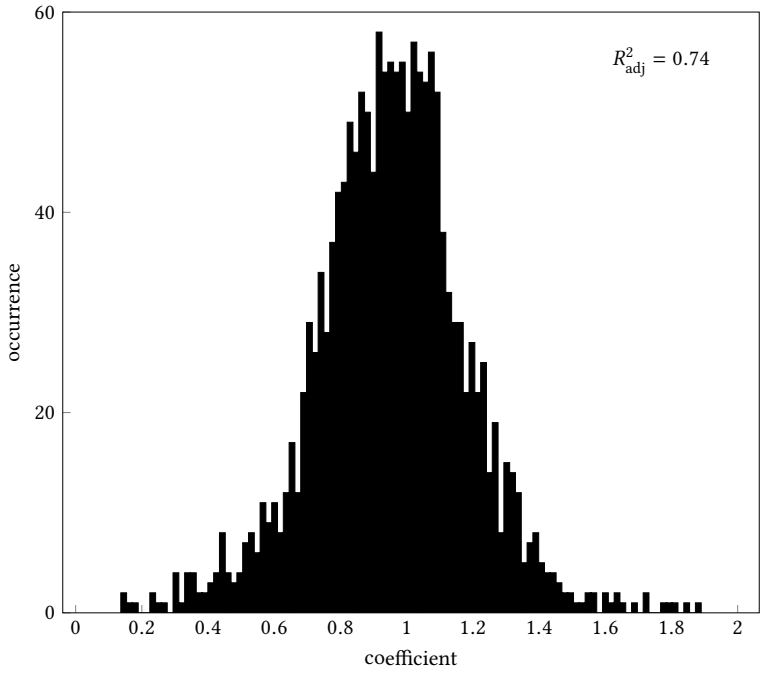
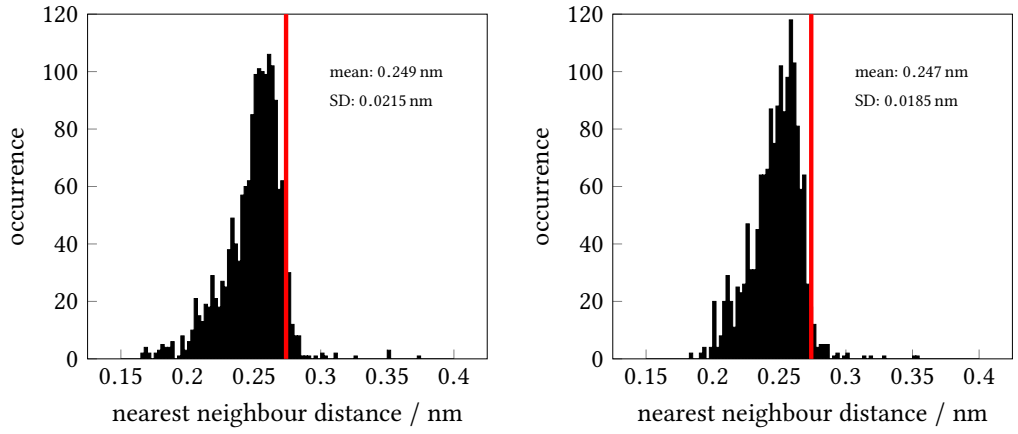


Figure 5.12: A simple least squares fit (with positivity constraint) of the AST reconstruction to the experimental sinogram is performed and the weight distribution is shown here. The adjusted R^2 value of the fit is 0.74.



(a) AST (b) EST

Figure 5.13: The nearest neighbour distance distributions of the results from AST and EST are shown in (a) and (b) respectively. The red line indicates the theoretical nearest neighbour distance in a perfect tungsten lattice, 0.274 nm.

ination and random noise distort the tungsten atoms from their equilibrium positions and both reconstructions give the same conclusion.

To conclude, the AST reconstruction successfully identified the locations of individual tungsten atoms. The positions of the tungsten atoms are consistent

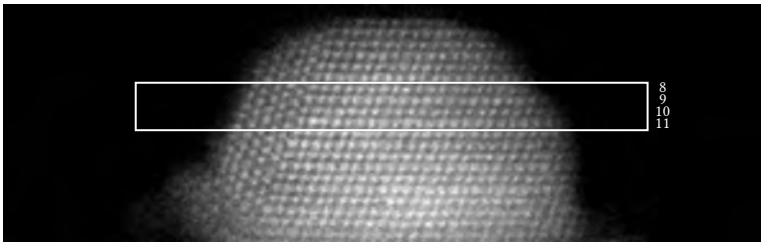


Figure 5.14: A zone-axis projection (at tilt angle −122.29° and binned by a factor of 2) of the gold nanoparticle suspended on the apex of the tungsten tip is shown here. There are 110 projections in this tilt series. The pixel size is 0.0436 nm. The selected section (highlighted by the box) to demonstrate the reconstruction by AST is labelled as 8–11. The size of the whole image is 5.12 nm × 16.39 nm.

with the experimental data as shown in Figure 5.11. This provides a validation of the AST algorithm.

5.2 Gold nanoparticle reconstruction

AST is validated in Section 5.1 by using a tungsten tip dataset published in [12, 189]. In this section, a gold nanoparticle dataset acquired by tungsten tip tomography is reconstructed by AST.

5.2.1 Input sinogram

The dataset was obtained by tungsten tip tomography over the full ±90° tilt range as described in Chapter 2. The tilt angles are separated by 3° and additional images were obtained ±1.5° and ±0.75° around the zone axes. More zone-axis projections in the sinogram has the same effect of increasing the weights of zone-axis projections in lasso fits (see Section 5.1.3). The total number of images in the tilt series is 110. Scanning artefacts were removed and the tilt series was aligned as described in Chapter 3. A zone-axis projection of the whole gold nanoparticle is shown in Figure 5.14.

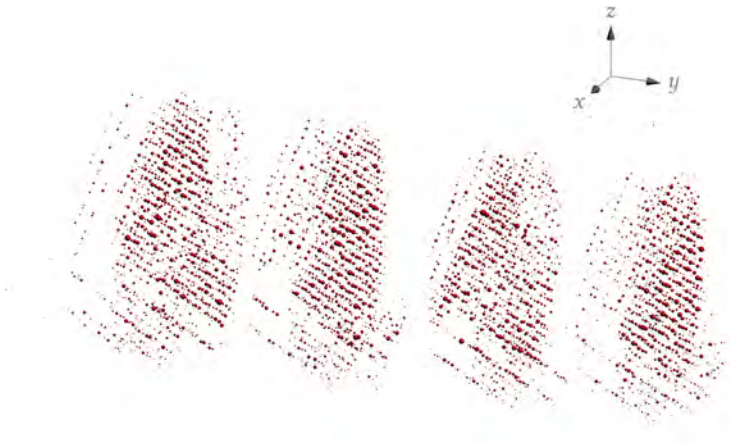
A section of the gold nanoparticle is selected to demonstrate the reconstruction by AST. The top section of the gold nanoparticle is avoided as the surface atoms

are less stable; gold is less stable than tungsten under the electron beam, so the atoms at the top are more susceptible to being displaced by the electron beam. As visible in Figure 5.14, different crystal grains are present in the gold nanoparticle. This poses a challenge as the contributions from neighbouring layers cannot be estimated by a simple exponential decay at the edge; therefore, the reconstruction is focused on the central two sections, indicated as 9 and 10 in Figure 5.14. For practical purposes, the selected section is also binned by a factor of 2 to reduce the total number of pixels (see Section 4.4).

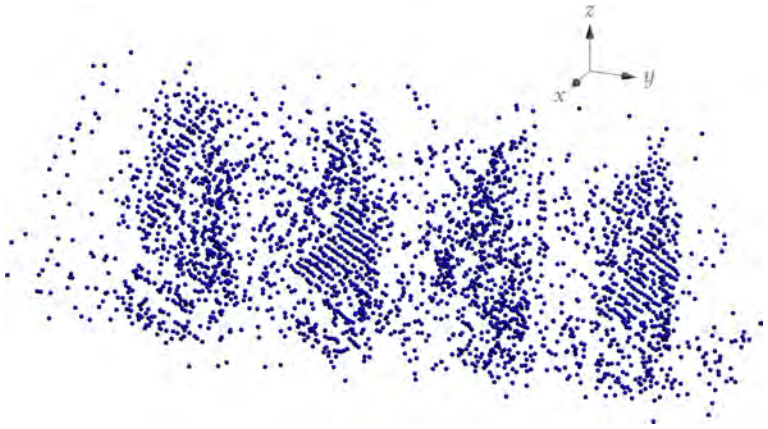
A preliminary face-centred cubic (fcc) crystal model is constructed by estimating the support of the gold nanoparticle from the sinogram. Atomic shapes are found by the steps as described in Section 5.1.1. A search grid is constructed and a lasso fit is performed. The support of the gold nanoparticle is refined by using the result from the search grid; an fcc crystal is constructed by using the refined support and new atomic shapes are found by using the refined crystal model. This is done iteratively until the atomic shapes become consistent with the crystal model.

5.2.2 Reconstruction of selected section

Initial lasso fit of the search grid to the selected section (see Figure 5.14) is performed and the result is shown in Figure 5.15 and Figure 5.16 in three-dimensional space and two-dimensional projections respectively. A clear close-packed crystal structure can be seen, which agrees with the fcc lattice structure of gold. Arrangement of some gold atoms is different and this indicates the presence of different crystal grains. This is especially clear in Figure 5.15a with non-zero weights showing up between the atomic layers, and one of the different crystal grains is highlighted in Figure 5.16b.



(a) search grid

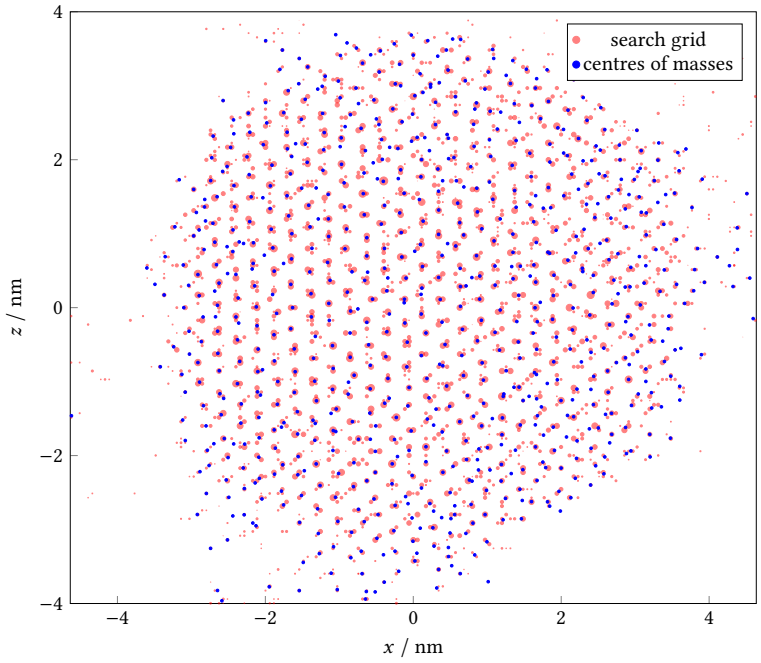


(b) centres of masses

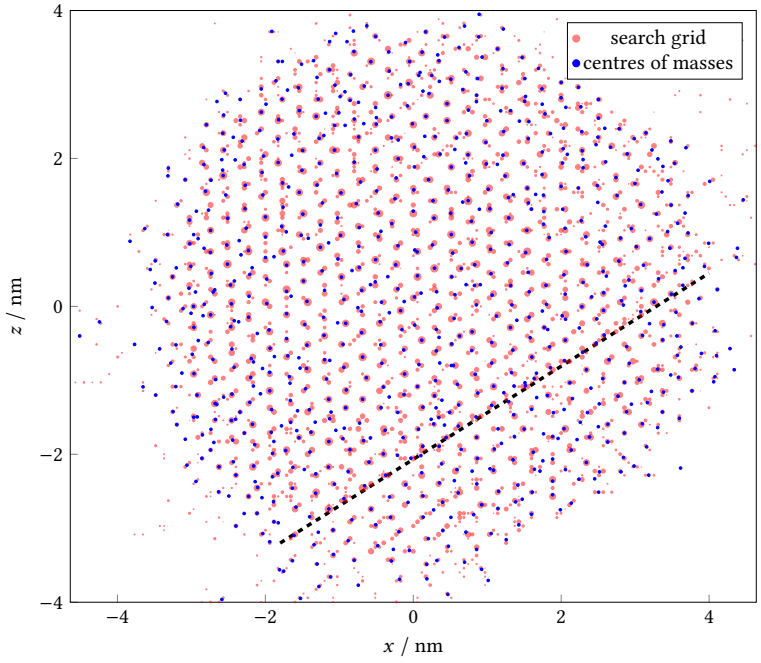
Figure 5.15: The result after lasso fit of the search grid to the sinogram of gold nanoparticle and its centres of masses are shown in (a) and (b) respectively. For clarity, the y dimension is stretched by a factor of 20.

Cross-hairs refinement is performed as described in Section 4.2.3 and the reconstruction of section 9 and 10 in three-dimensional space and two-dimensional projections are shown in Figure 5.17 and Figure 5.18 respectively.

Selected experimental projections and calculated projections from the AST reconstruction are shown in Figure 5.19. The calculated projections are consistent with the corresponding experimental projections, and it successfully shows the presence of different crystal grains. The calculated projections are generated by assuming all weights of atoms are 1 — they are all identical. The distribution

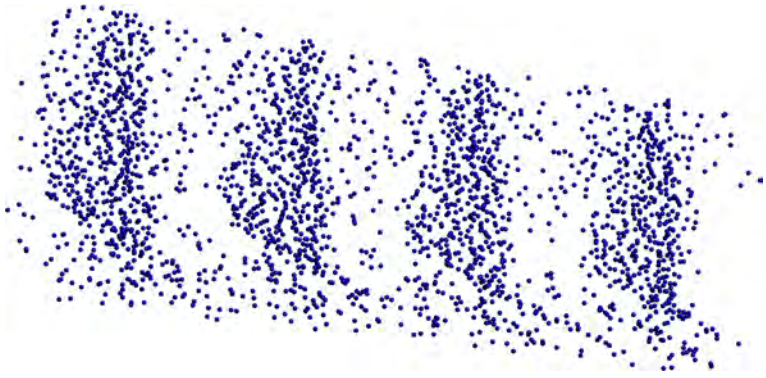


(a) section 9

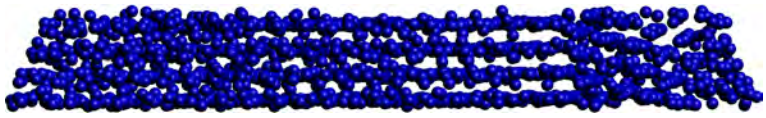


(b) section 10

Figure 5.16: The result of initial lasso fit of the search grid and their centres of masses of the projections along the [1 1 1] direction of the gold nanoparticle are shown in (a) and (b) for section 9 and 10 respectively. The radii of the points for the search grid are proportional to the square root of the weights for clarity. The dashed line in (b) indicates a region of different crystal grains.



(a)



(b)

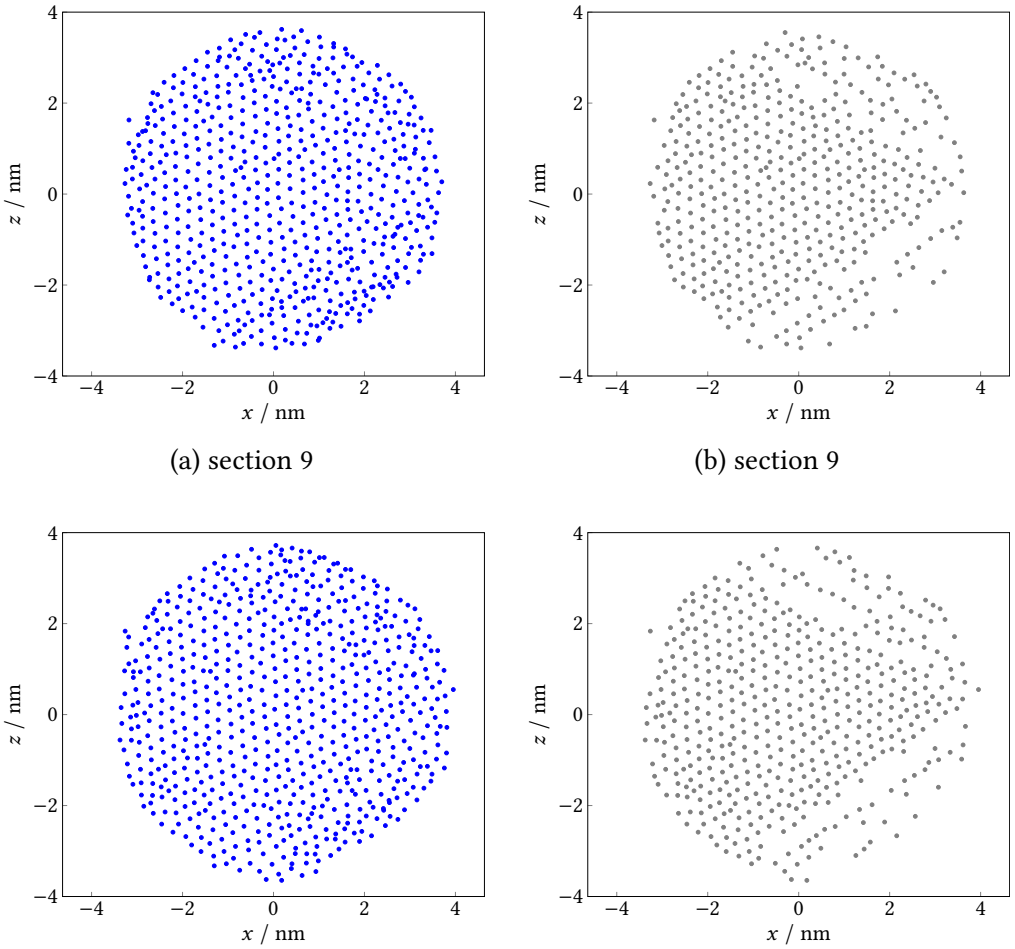
Figure 5.17: The AST reconstruction in three-dimensional space is shown here. (a) is at the same perspective as shown in Figure 5.15 with the y dimension stretched by a factor of 20 for clarity. (b) shows a cross-sectional view of the gold nanoparticle with a clear different grain on the right side. As the contributions from neighbouring layers of the top and bottom section are not known, their reconstructions should be ignored and only the central sections are reliable.

of nearest neighbour distances and the weights after simple least squares fit of the reconstruction to the sinogram are shown in Figure 5.20.

The selected section of a gold nanoparticle dataset is successfully reconstructed by AST. Unlike the tungsten tip dataset in Section 5.1, where the traced positions determined from EST are used to approximate the atomic shapes (see Section 5.1.1), only the smallest nearest neighbour distance of the expected crystal is utilised in the AST reconstruction of the gold nanoparticle. This shows AST is applicable to datasets with different crystalline structures.

5.2.3 Full reconstruction

The reconstruction of a region of the gold nanoparticle is consistent with the fcc structure of a gold crystal lattice. From the experimental images and the reconstruction, it is noticed that different grains are present in this gold nanoparticle.



(a) section 9

(b) section 9

(c) section 10

(d) section 10

Figure 5.18: The two-dimensional projections along the [1 1 1] direction of the gold nanoparticle of section 9 and 10 are shown here. The projected coordinates in (a) and (c) are taken within the range of $\pm 0.625 d_{111}$ from the centre of the slab, where d_{111} is the interplanar distance of (1 1 1) planes; The projected coordinates in (b) and (d) are taken within the range of $\pm 0.25 d_{111}$ from the centre of the slab. When the range is reduced, fewer atoms are projected, indicating they are not on the same plane as with other atoms. This indicates the existence of different crystal grains.

To have a better understanding of the relationship between these grains, it is desirable to obtain the reconstruction of the whole gold nanoparticle instead of using only the selected section in Section 5.2.1.

The whole gold nanoparticle is reconstructed by AST with the procedures described in Chapter 4 and Sections 5.2.1 and 5.2.2, except it is done by combining different sections of the gold nanoparticle due to limitation of memory (see Sec-

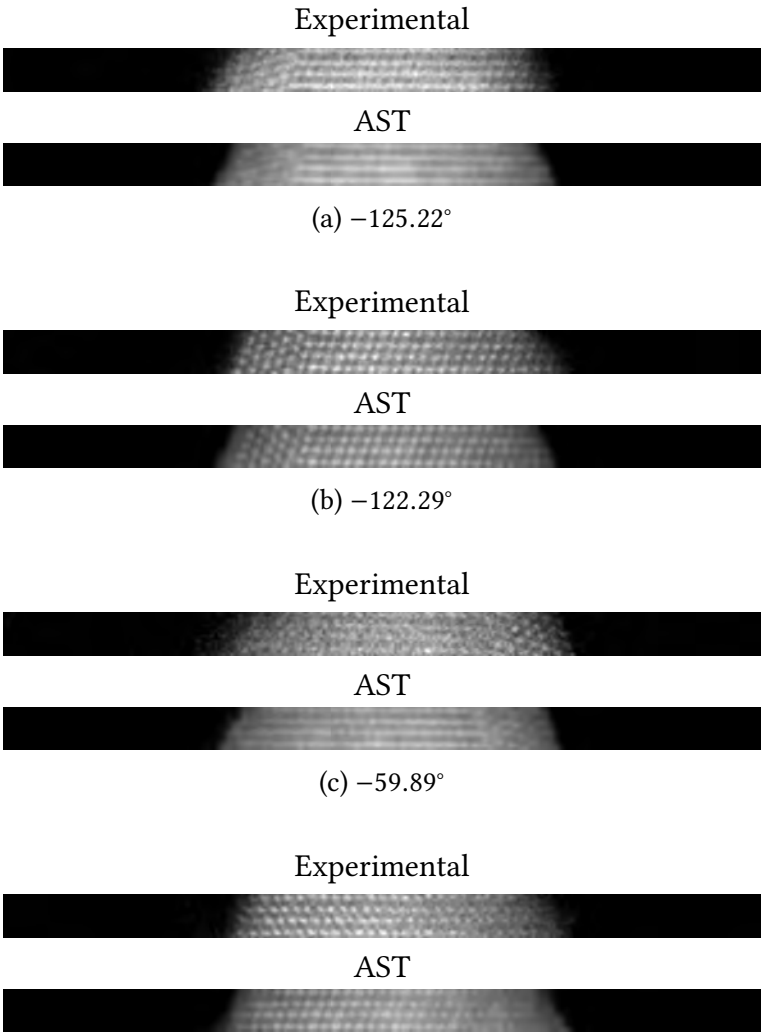
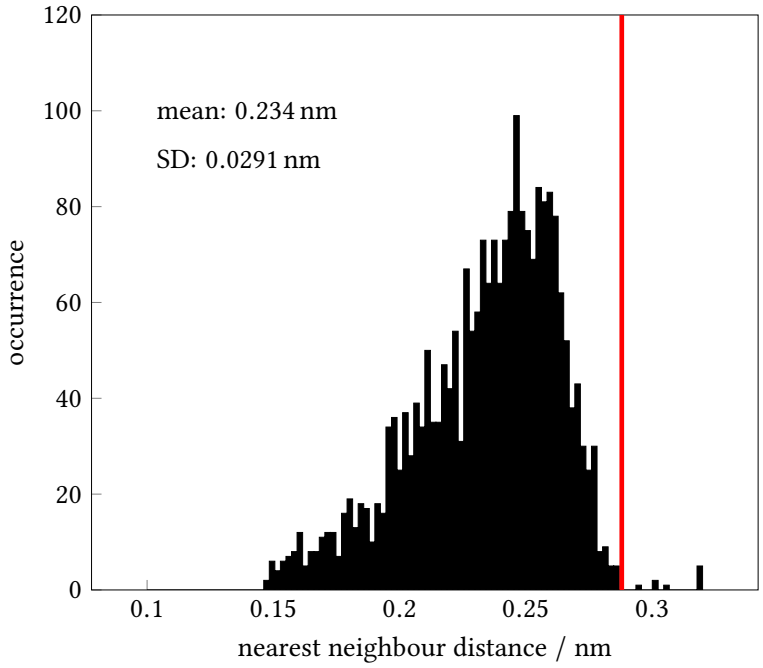
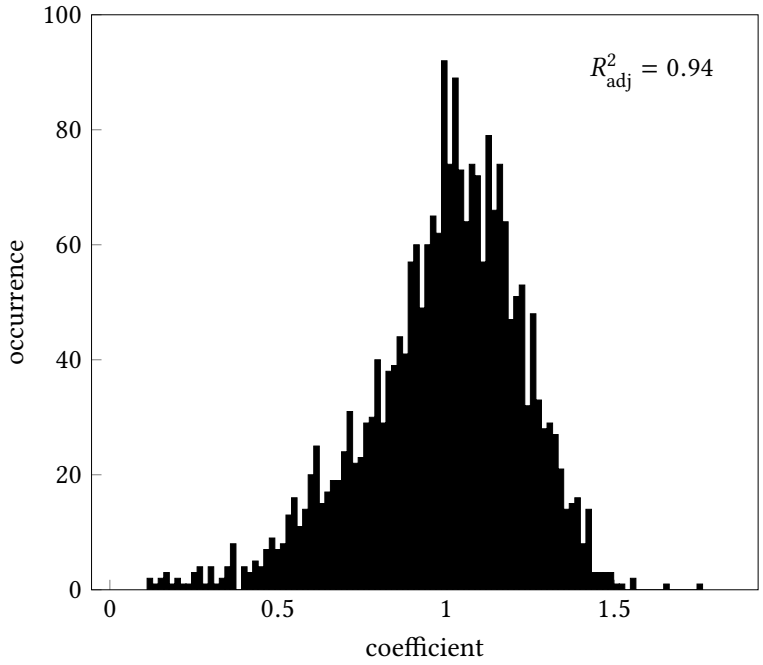


Figure 5.19: Selected projections from the experimental sinogram and the projections calculated from the results of AST are shown here. The consistency of these projections indicates the model from AST is valid. Each pair of the images is shown in the same colour range. The size of the projections is 0.95 nm × 16.39 nm.



(a) nearest neighbour



(b) least squares weights

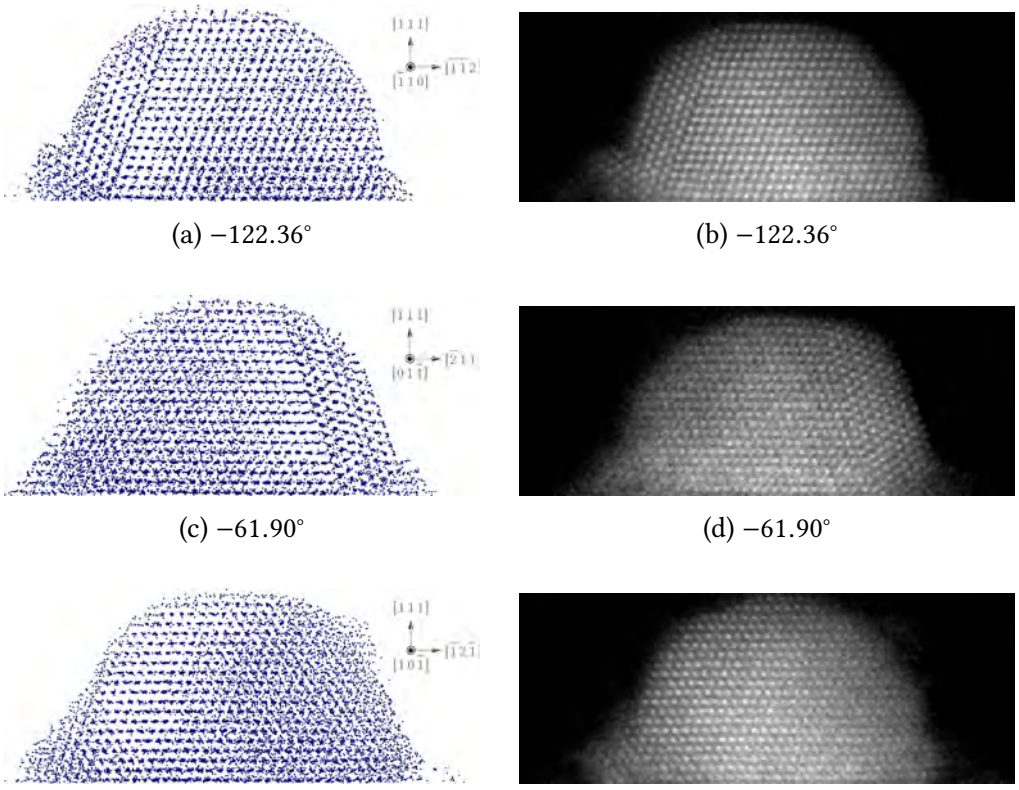
Figure 5.20: The distribution of nearest neighbour distances and the weights from simple least squares fit of the AST reconstruction are shown in (a) and (b) respectively. The red line in (a) indicates the theoretical nearest neighbour distance in a perfect gold lattice, 0.288 nm.

tion 4.4). The region of the gold nanoparticle close to the tungsten tip is not included in the reconstruction as the tungsten tip and its interfacial region with the gold nanoparticle cannot be accurately reconstructed due to the presence of additional nanoparticles and tungsten oxides surrounding the tungsten tip (see Figure 2.21).

The three-dimensional reconstruction of the gold nanoparticle and experimental projections at different tilt angles are shown in Figure 5.21. There are 12 100 gold atoms in the reconstruction and the tomography cross section of the gold nanoparticle is approximately 8 nm in diameter and 4 nm in height. The reconstruction clearly shows that the gold nanoparticle consists of three different grains and this is consistent with the experimental projections (see Figure 5.21). The grain in the central region is denoted by *main grain*; the grain on the left side of Figure 5.21a is denoted by *grain 1*; the grain on the right side of Figure 5.21c is denoted by *grain 2*. The fcc structure of the gold nanoparticle is most evident when it is viewed from the top (along [1 1 1]), and the well-known ABCABC… close-packed arrangement can be seen in Figure 5.22 when 4 layers of gold atoms are overlapped with each other.

5.3 Discussions

AST is used to reconstruct an experimental dataset of tungsten tip in Section 5.1 and the results are consistent with the published results by EST [12, 189]. This validates the algorithm of AST. With the knowledge of three-dimensional coordinates of thousands of individual tungsten atoms, the crystal structure can be accurately identified and crystal defects such as point defects can be visualised in a single-atom level [12]. This even allows the direct measurement of atomic displacement field in three-dimensional space [12]. The understanding of material properties is advanced greatly by the accurate knowledge of three-



(a) −122.36°

(b) −122.36°

(c) −61.90°

(d) −61.90°

(e) −1.62°

(f) −1.62°

Figure 5.21: The three-dimensional coordinates of the atoms of the whole gold nanoparticle at different tilt angles and their corresponding experimental projections are shown here. There are 12 100 atoms identified for the gold nanoparticle, with an approximate diameter of 8 nm and height of 4 nm. (a) is at tilt angle −122.36° and the grain on the left side is denoted by *grain 1*. (c) is at tilt angle −61.90° and the grain on the right side is denoted by *grain 2*. (b), (d) and (f) show the respective experimental projections taken at the tilt angles as in (a), (c) and (e). It can be seen that the grains at different parts of the gold nanoparticle agree with the three-dimensional coordinates. The size of the projections is 4.90 nm × 12.75 nm.

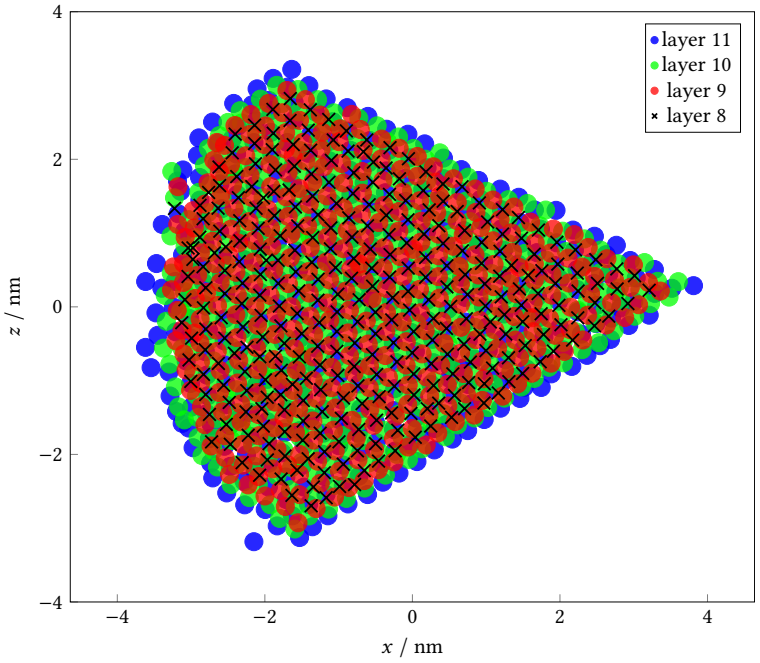
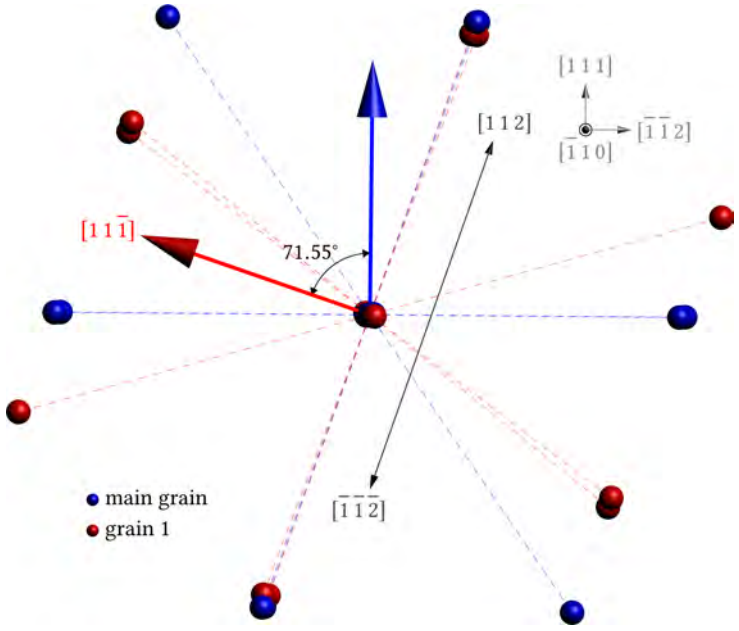


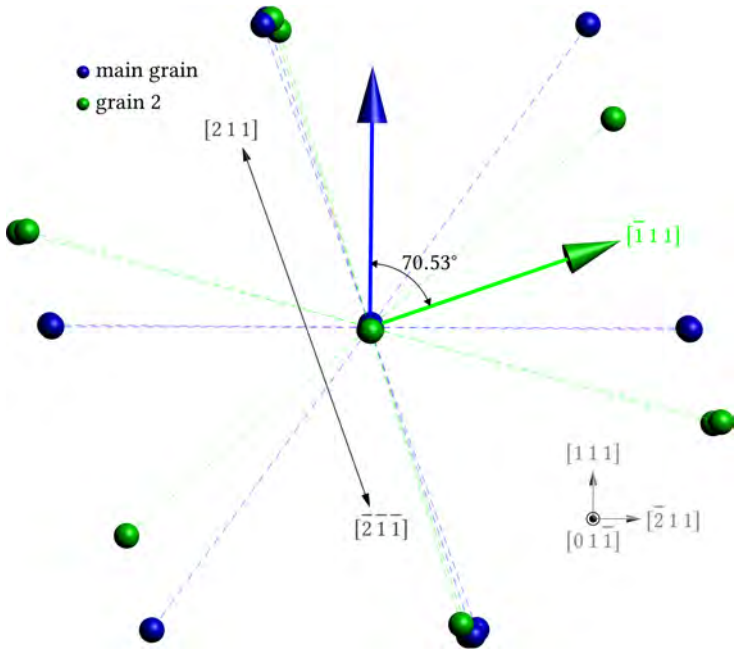
Figure 5.22: 4 layers of the gold nanoparticle (denoted by 8–11 in Figure 5.14) are viewed along the [1 1 1] direction. Only the main grain is included here for clarity. The bottom and top layer are shown in blue circles and black crosses respectively. The well-known ABCABC… stacking of fcc crystal can be clearly seen.

dimensional coordinates of individual atoms and AET is essential to achieve this.

AST is also used to reconstruct a gold nanoparticle dataset acquired by tungsten tip tomography (see Section 5.2). The knowledge of three-dimensional coordinates of the individual gold atoms allows a thorough analysis of the orientations of different grains in this gold nanoparticle. Different grains are first separated by inspecting their boundaries; the average local nearest neighbour configurations for each grain are then obtained by identifying the 12 nearest neighbours of every atom. The average local nearest neighbour configuration of the main grain, grain 1 and grain 2 are shown in Figure 5.23 by blue, red and green spheres, respectively. They are projected along ⟨1 1 0⟩ as in Figure 5.21.



(a)



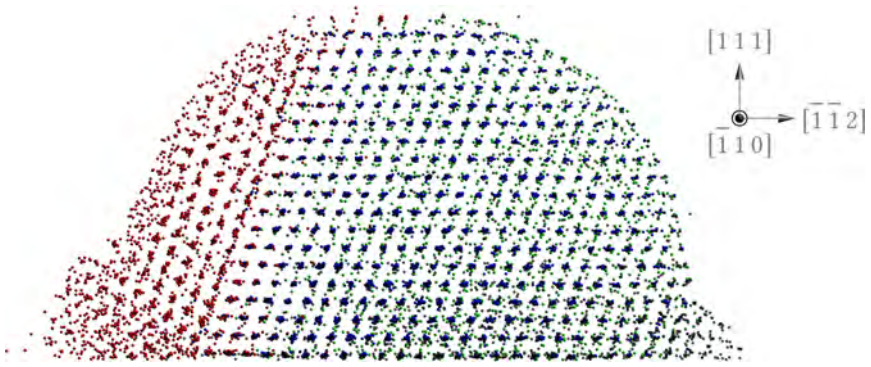
(b)

Figure 5.23: The local nearest neighbour configurations of different grains in the gold nanoparticle are shown here. (a) the blue and red spheres are the local nearest neighbours of the main grain and grain 1 respectively. (b) the blue and green spheres are the local nearest neighbours of the main grain and grain 2 respectively. The big arrows with the corresponding colours denote the close-packed stacking direction of the corresponding grains and their relative angles are indicated. The twinning directions between different grains are also shown by black double-headed arrows.

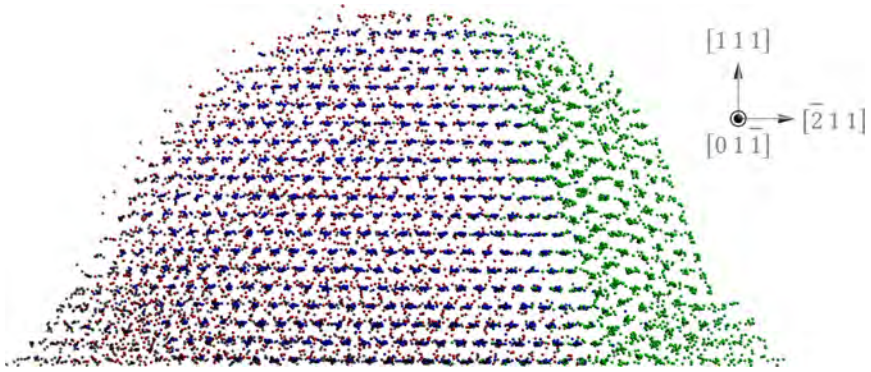
The grain boundaries are ⟨1 1 0⟩ tilt boundaries as the axis of rotation is ⟨1 1 0⟩. The orientations of these grains are determined by identifying the close-packed stacking directions of each grain: it is 71.55° between the main grain and grain 1 (see Figure 5.23a) and 70.53° between the main grain and grain 2 (see Figure 5.23b). The twin plane of the main grain and grain 1 is $(1\,1\,\bar{1})$ and the main grain and grain 2, $(\bar{1}\,1\,1)$. These grain boundaries are the well-known (1 1 1) twin boundaries [190]. The theoretical misorientation of grains having a (1 1 1) twin boundary is 70.53° [190, 191], which agrees well with the determined angles. The twinning direction is along ⟨1 1 2⟩ (see Figure 5.23) and this further confirms the two grain boundaries are (1 1 1) twin boundaries.

The local nearest neighbours of all atoms are now compared with the average nearest neighbour configurations of the three grains. 'Scores' are given to every atom in the gold nanoparticle by comparing the vectors to their 12 nearest neighbours with those three nearest neighbour configurations; a higher score is given if the nearest neighbour is closer to the average nearest neighbours. The atoms are then classified by the spectral clustering algorithm [136], by using the radial basis function with a kernel coefficient of 1 [192]. Atoms are classified according to their three-dimensional coordinates and the given scores. The classification separates atoms into four categories: belonging to the main grain, grain 1, grain 2 or unclassified. The four categories are coloured respectively and are shown in Figure 5.24. The two-dimensional projection of section 9 (see Figure 5.14) of the gold nanoparticle is shown in Figure 5.25 with atoms being coloured according to the grain they belong to.

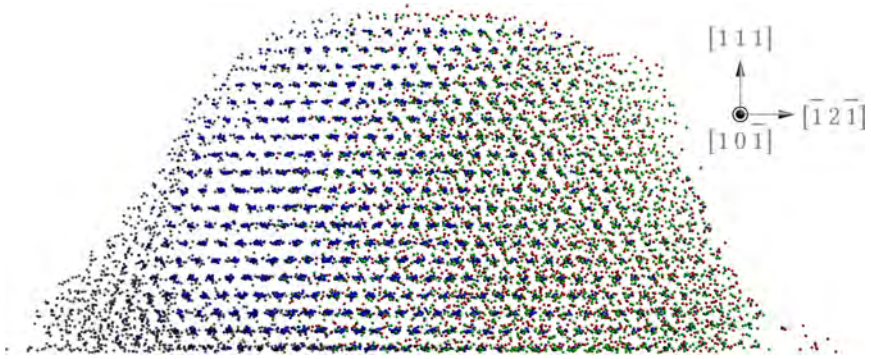
There are some uncertainties in the classification; some atoms are labelled as 'unclassified' (atoms that do not belong to any of the present grain) and these are shown as black crosses in Figure 5.25. The atoms near the boundary planes and surface atoms have different nearest neighbour configurations than



(a) −122.36°



(b) −61.90°



(c) −1.62°

Figure 5.24: The three-dimensional coordinates of the whole gold nanoparticle are shown with atoms classified into different grains. The atoms belonging to the main grain, grain 1 and grain 2 are coloured as blue, red and green respectively. The atoms that do not belong to any of the present grain are shown in black. The viewing angles of (a), (b) and (c) are −122.36°, −61.90° and −1.62° respectively, the same as in Figure 5.21.

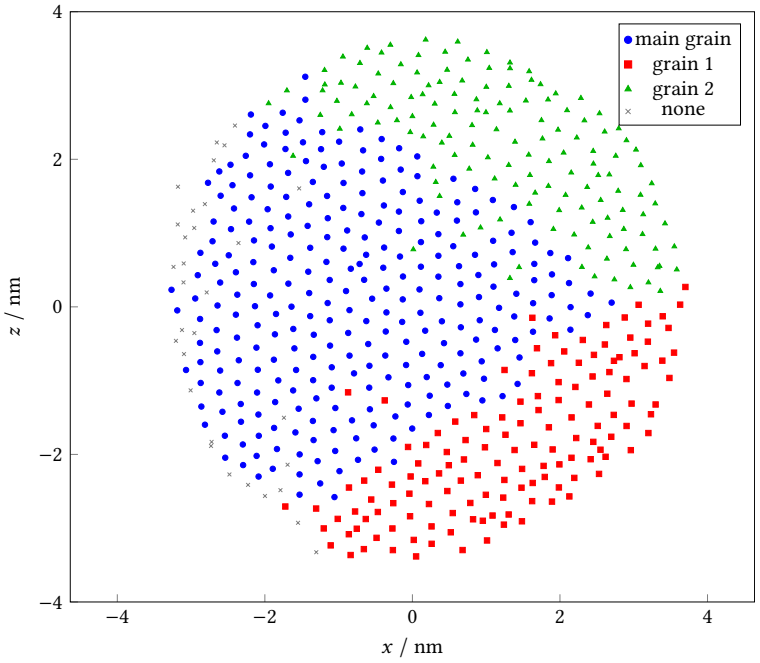


Figure 5.25: A single layer of the gold nanoparticle is viewed along ⟨1 1 1⟩, with atoms coloured according to the grain they belong to. The main grain, grain 1 and grain 2 are shown in blue circles, red squares and green triangles respectively. Unclassified atoms are depicted as black crosses. It is evident that the spectral clustering algorithm provides a highly accurate classification of the atoms regarding the grains they belong to.

those in the bulk; only the atoms in the bulk have the full nearest neighbour configurations as the grains. It is expected that the classification, which depends on the three-dimensional coordinates and nearest neighbours of the atoms, will give some uncertainties. Some of the misclassified atoms could be due to the noise in the sinogram. Those atoms are not removed as long as they satisfy the physical constraint on nearest neighbour distance; if they are removed based on their nearest neighbour configurations, this would enforce a particular crystal structure, which is not desired in AST.

To conclude, AST not only resolves a single crystal structure (as in the tungsten tip) but also resolves multiple grains with atomic precision. This can greatly advance our understanding of crystal structures and defects. AST is applicable to a variety of nanoparticles whether they have crystalline or amorphous structures.

The improved understanding of crystal structures and defects of nanoparticles can have a significant impact on their applications such as catalysis [193, 194] and biology [195–197].

6

Tuning fork AFM

The current method of preparing samples for tungsten tip tomography is by evaporation and annealing (see Section 2.1.3). This limits the versatility of the sample as some materials are not easily formed on the tip's apex. The current method also cannot control precisely where the nanoparticle is formed. These limitations motivate the design of a tuning fork atomic force microscope (AFM) for versatile sample preparation for tungsten tip tomography.

If the sample is placed on a substrate, an etched tungsten tip can be used to 'pick up' a nanoparticle in a scanning probe microscope (SPM) instrument. Two common modes of SPM are scanning tunnelling microscopy (STM) and AFM. STM requires a conductive substrate as the feedback signal is the tunnelling current between the tip and the substrate. It is not as flexible as AFM which does not depend on a tunnelling current signal, and thus does not limit the substrate to be a conductive surface. Traditional AFM requires a laser pointing directly at the cantilever. The deflection of the cantilever changes the location of the laser on a photodiode detector, and the feedback signal controls the movement of a piezoelectric motor. These components do not allow the flexibility of using an etched tungsten tip.

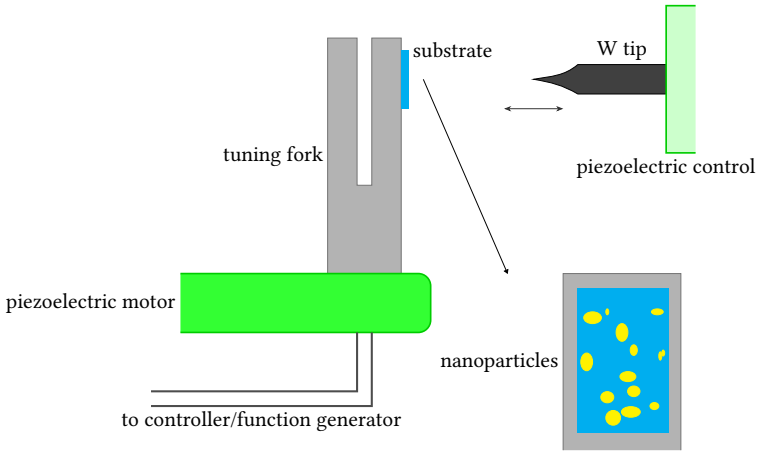


Figure 6.1: A schematic diagram of the design of the tuning fork AFM for picking up nanoparticles for tungsten tip tomography is shown here. A de-capped tuning fork is attached on the piezoelectric motor and it is excited by signals from a function generator. A small piece of substrate is glued on one of the two prongs of the de-capped tuning fork. Nanoparticles are drop-cast on its surface. An etched tungsten tip is mounted on a three-axis piezoelectric stage. By performing a raster scan, a suitable nanoparticle can be identified and it is picked up by the etched tungsten tip.

If the substrate is fixed on a quartz tuning fork oscillating at its resonant frequency, an etched tungsten tip can be approached and pick up the nanoparticle on the substrate. The equipment is not for characterisation purposes, so expensive and ultra-stable components are not needed. This chapter outlines the development of a tuning fork AFM for preparing tungsten tip samples for tomography.

6.1 Design

Conventional tuning fork AFMs usually have the tip attached on the tuning fork. For the purpose of preparing tungsten tip tomography's samples, the etched tungsten tip cannot be easily separated from the tuning fork if the tip is attached on it. Instead the substrate is attached on the tuning fork, while the etched tungsten tip is held by a hypodermic needle to allow easy exchange of tips. A schematic diagram showing the design of the tuning fork AFM is shown

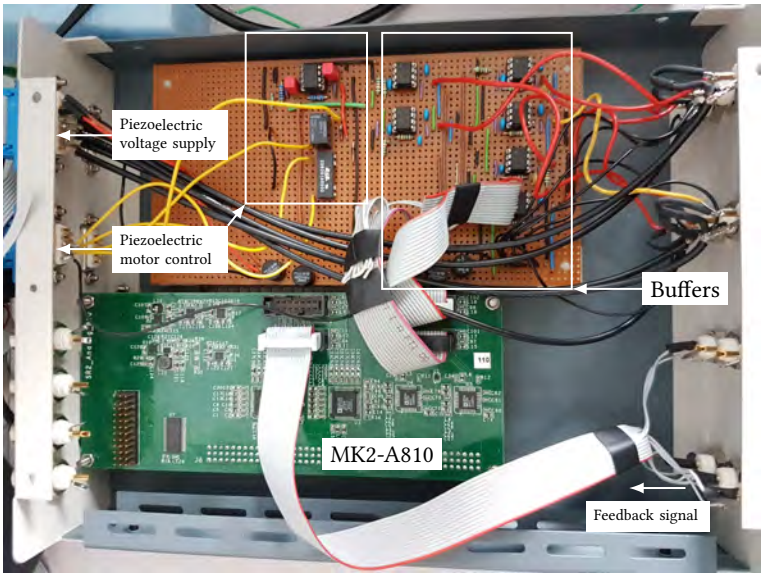


Figure 6.2: The MK2-A810 board is contained in a metal chassis together with external circuit controls. The feedback signal goes into the DSP and after processing, the output signals go to corresponding components with buffer controls.

in Figure 6.1. Different components of the tuning fork AFM are discussed in this section.

6.1.1 Controller

The control board is the Signal Ranger MK2 with A810 expansion board (MK2-A810) from Soft dB (see Figure 6.2). It includes a digital signal processor (DSP) operating at a rate of 300 MHz. DSPs are optimised for high-speed numeric operations by processing a continuous analog signal in digital form. The analog signals are converted to digital signals by analog-to-digital converters (ADCs), and any digital signals can be output as analog signal by digital-to-analog converters (DACs). The output voltages from the DACs are ±10 V. The full specification of MK2-A810 can be found in [198].

To interact with the MK2-A810 controller, Gnome X Scanning Microscopy (GXSM) is used [199]. It is an open-source project which has a close integration with MK2-A810 and other variants. Different operating modes such as

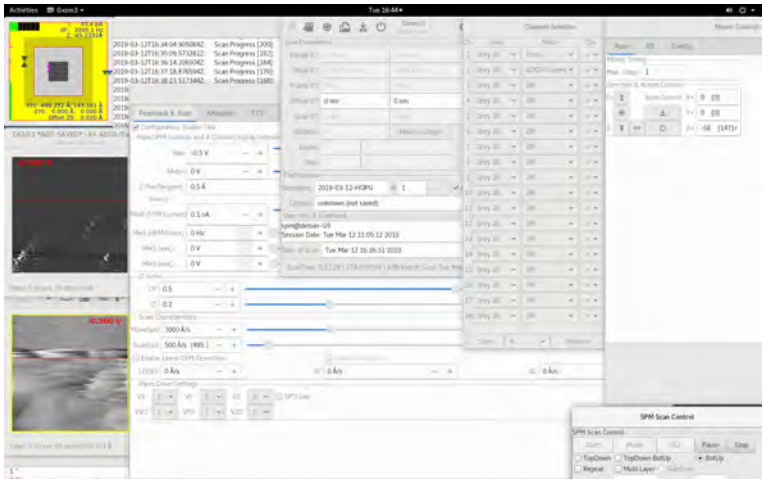


Figure 6.3: A screenshot of GXSM3 software while it is scanning in its STM mode. The software allows all typical control for SPM scanning, for example, different data acquisition channels, scanning and feedback control.

AFM or STM can be selected in the software. It allows all typical scanning controls, for example, scanning range, scanning speed, rotation, bias voltage control, PI feedback control, tip positioning and all kinds of spectroscopy. The GXSM3 (3.47.1, SVN 4963 with DSP firmware 20.35) software is installed in Debian Jessie (kernel version 4.9.82) and a screenshot of the software is shown in Figure 6.3. Python scripting can be used for advanced automation and control, and some scripts are written to control the movement of tip (see Sections 6.1.3 and 6.2.4).

6.1.2 Piezoelectric control

The piezoelectric control of the tuning fork AFM is vital for tip approach and scanning. It is realised by 5 surface-mounted piezoelectric buzzers (RS, 724-3162). The piezoelectric buzzer can convert electrical energy to mechanical movement and vice versa. It is called a 'buzzer' because it can vibrate in the audible frequency range (3.7 kHz–4.7 kHz). If an inappropriately high feedback is used, the Z piezoelectric buzzer will vibrate in high frequency to produce a sound.

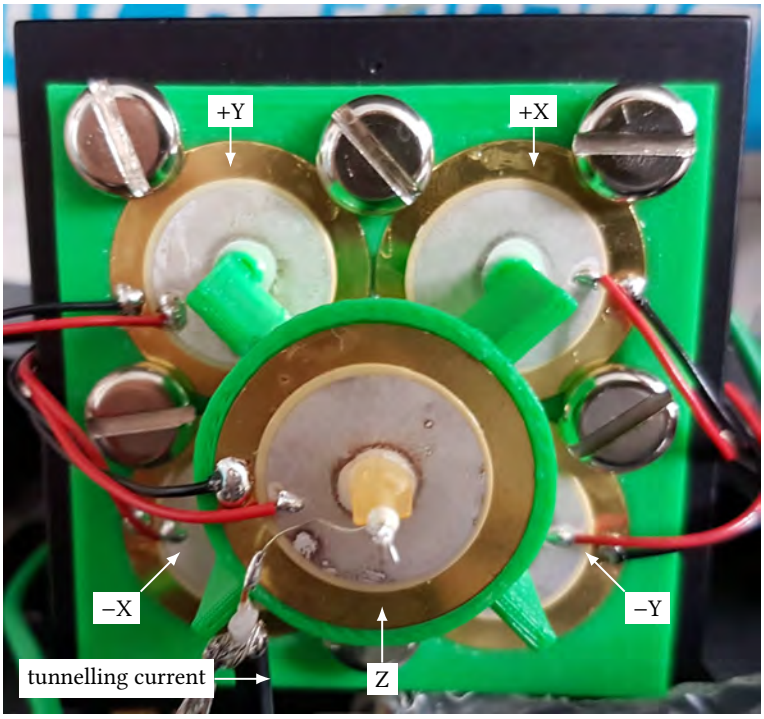


Figure 6.4: The piezoelectric buzzers in the tuning fork AFM is shown in the photo with the Z buzzer supported on X and Y buzzers. Voltages with opposite polarity are supplied to the X or Y piezoelectric buzzers at the same time to control the horizontal movement. A 25 G hypodermic needle is attached on the centre of the Z piezoelectric buzzer, and a very thin wire for tunnelling current (for STM mode) is attached on the external wall of the hypodermic needle by silver epoxy.

The arrangement of the 5 piezoelectric buzzers is shown in Figure 6.4. The voltage supplied to each piezoelectric buzzer is controlled by the DSP in the MK2-A810 board, which has a limit of ±10 V (the manufacturer's specified maximum range of the piezoelectric buzzer is ±15 V). The output to the X, Y and Z piezoelectric buzzers are from channel 3, 4 and 5 of the DACs from the MK2-A810 board respectively. To avoid any unwanted voltage flowing into the MK2-A810 board due to mechanical movement of the piezoelectric buzzers, operational amplifiers are used as buffers (see Figure 6.5). The polarity of the output voltage is opposite to that from DACs as the buffers operate in a negative feedback configuration with unity gain.

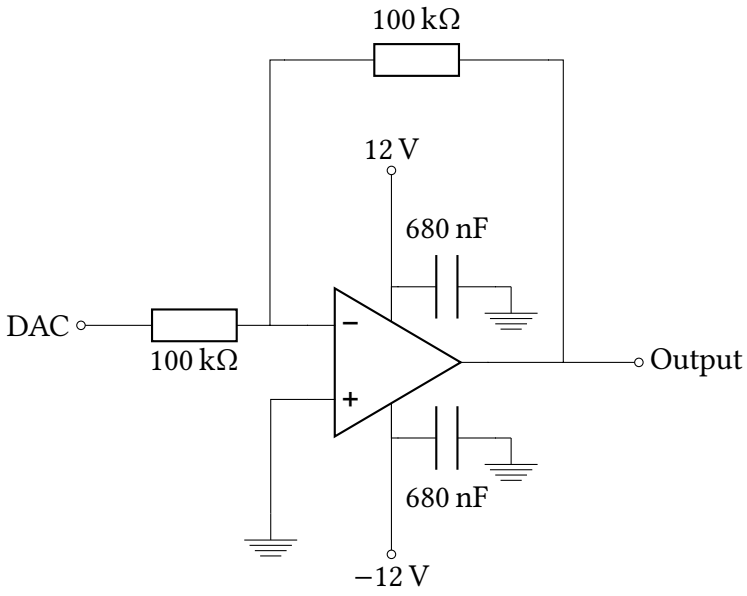
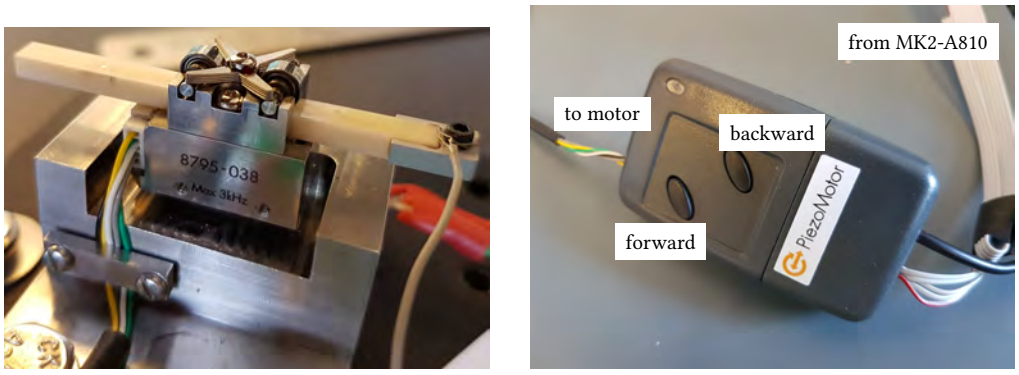


Figure 6.5: The MK2-A810 board is protected by the buffer shown here. The large impedance of the operational amplifier prevents any fluctuation in the output going into the DAC channel. The operational amplifier used is TL081.



(a) (b)

Figure 6.6: The linear piezoelectric motor and its handheld driver modified for the control from MK2-A810 are shown in (a) and (b) respectively.

6.1.3 Tip approach

The etched tungsten tip is held by a cut-out 25 G hypodermic needle which has an inner diameter of 260 µm. The hypodermic needle is attached to the centre of the Z piezoelectric buzzer. The tuning fork, which is mounted on a linear piezoelectric motor (PiezoMotor, LL10) [200] (see Figure 6.6a), is slowly moved towards the etched tungsten tip. The movement of the linear piezoelectric

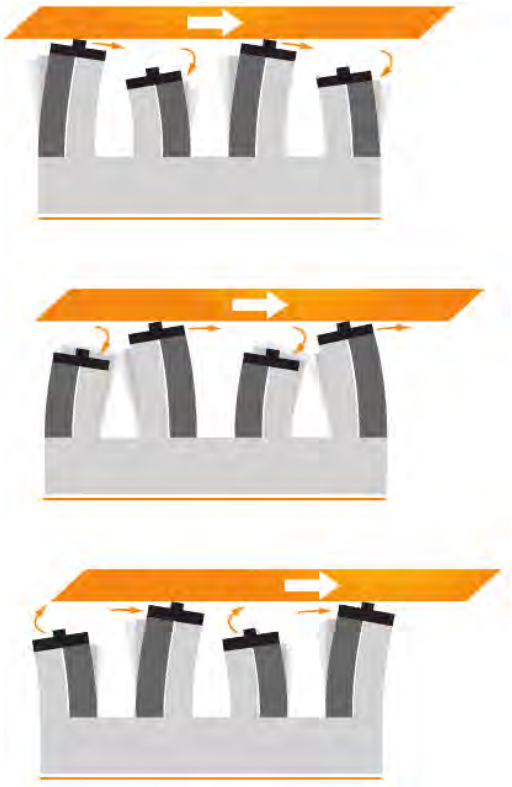


Figure 6.7: The movement of the linear piezoelectric motor is achieved by the contact between the driver rod and the 'piezo legs' underneath. When a set of 'piezo legs' is electrically activated, it moves the driver rod while another set of 'piezo legs' retract. Those two sets of 'piezo legs' are alternatively activated and the driver rod is moved forwards. The figure and caption are adapted from [200].

motor can be controlled by a handheld push button driver (see Figure 6.6b). The movement of the rod is driven by a combination of electrically activated 'piezoelectric legs' and a schematic illustration is provided in Figure 6.7. The detailed description of its working principle can be found in [201].

Channel 7 of the DAC in MK2-A810 can output a ±5 V pulse. The pulses are used to control the forward and backward movements of the linear piezoelectric motor. There are two press buttons on the handheld driver which control the forward and backward movements, and mechanical relays are used to emulate the switch-on behaviour to control the movements. A circuit diagram used to illustrate this emulation is shown in Figure 6.8. The pulse is a single-step pulse

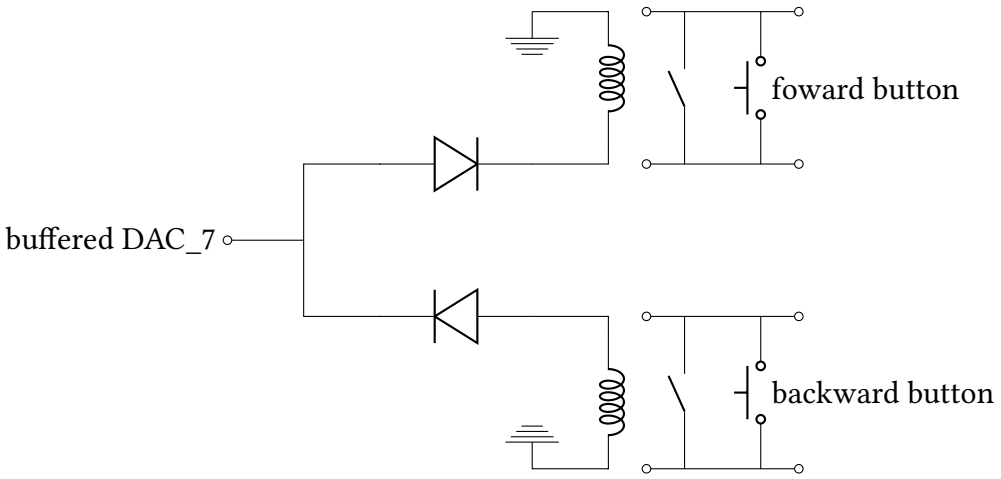


Figure 6.8: The output from channel 7 of DAC from the MK2-A810 board can be used to control the movement of the linear piezoelectric motor. Without any voltage from the DAC, the motor can be moved by pressing the button on the handheld driver which shorts the circuit of either the forward or backward circuit. When a negative pulse is output from the DAC, the forward circuit is shorted by the mechanical relay as the buffered signal is positive.

with a pulse width of 10 ms.

After a single forward step of the linear piezoelectric motor, the Z piezoelectric buzzer goes through a full extend-retract cycle (−10 V to 10 V from the DSP). The above steps repeat and the Z piezoelectric buzzer stops moving immediately once the setpoint of the feedback signal is reached, which is the phase difference in the AFM mode and tunnelling current in the STM mode. A settling time of 1 s is needed after reaching the setpoint for a successful approach. The speed of the tip approach can be controlled by adjusting the proportional and integral gain.

6.1.4 Image acquisition

After the tip is approached, scanning in the X and Y directions can acquire topographical information by enabling the 'constant-current' mode. The feedback is active and the Z piezoelectric buzzer is moved to maintain a constant current in STM mode or a constant phase difference in AFM mode. The movement

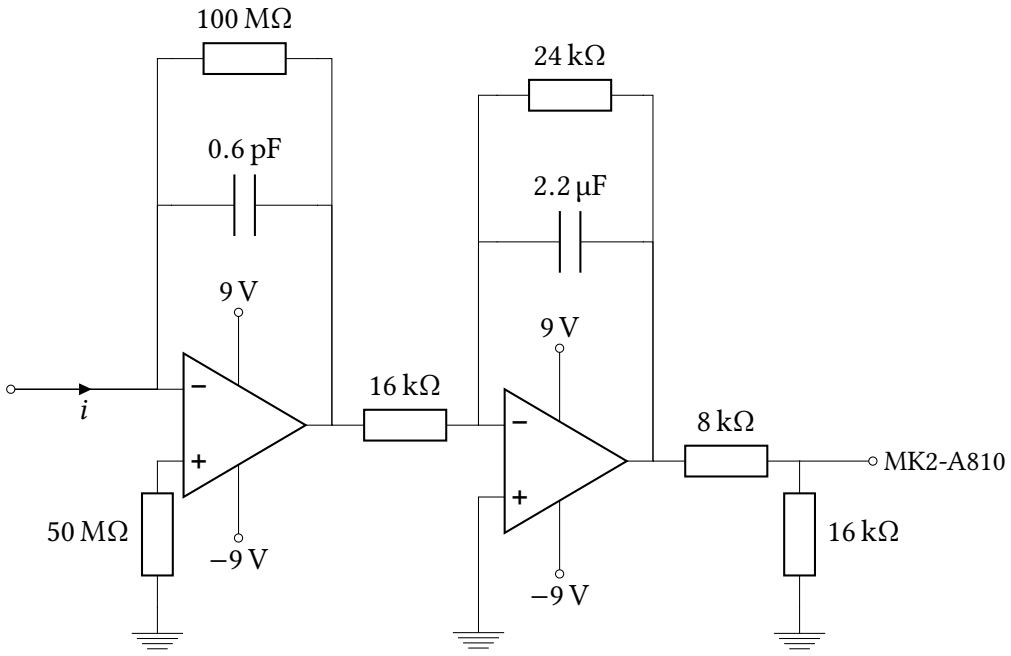
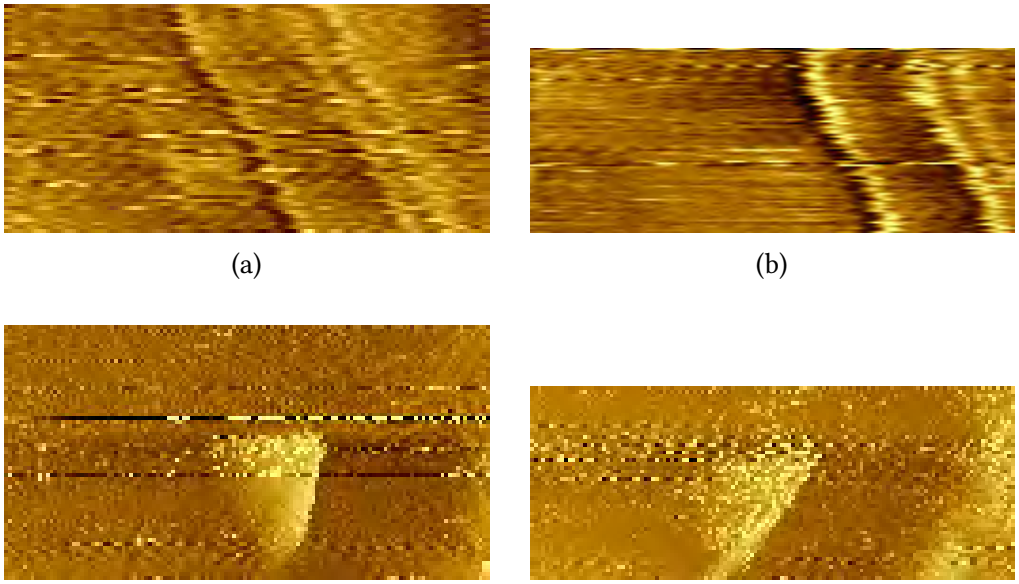


Figure 6.9: The preamplifier in STM mode consists of two TL081 operational amplifiers. The first operational amplifier (left) acts as a transimpedance amplifier which amplifies the small tunnelling current to a voltage by a factor of 10^8. A compensation resistor is added to cancel out any input bias current (≈30 pA). The purpose of the second operational amplifier (right) and the voltage divider is to protect the MK2-A810 board if the operational amplifiers are powered by its maximum voltage (±15 V). The cut-off frequency of the preamplifier is observed to be ≈2.2 kHz.

of the Z piezoelectric buzzer defines the topographical information during the raster scan.

STM mode is used to test the capability of image acquisition here because of its relative simplicity. The tunnelling current in the sample-tip junction is usually of the order of nA and it is amplified by a preamplifier before being converted to digital signal. The circuit of the preamplifier is shown in Figure 6.9. The bias voltage is applied to the sample through channel 6 of the DAC in the MK2-A810 board.

A piece of freshly cleaved highly oriented pyrolytic graphite (HOPG) is scanned using STM mode and two images are shown in Figures 6.10a and 6.10b. Graphite steps can clearly be seen in the topographical images. Few drops of colloidal



(a) (b)

(c) (d)

Figure 6.10: Some topographical images acquired by the SPM operating in STM mode are shown here. (a) and (b) are images from a bare HOPG surface. These images are filtered to remove high-frequency noise. (c) and (d) are images after gold nanoparticles were cast on the HOPG surface. These images are not filtered. All images are levelled by plane subtraction.

gold nanoparticles with an average diameter of 5 nm (BBI) were also cast on a freshly cleaved HOPG and dried in the air. Features corresponding to nanoparticles can be identified (see Figures 6.10c and 6.10d).

Once nanoparticles can be identified in the topographical information, the etched tungsten tip can be moved to the respective location and pick it up. The distance moved per unit voltage of the X, Y and Z piezoelectric buzzers are not calibrated from Figure 6.10 as the quality (interfered by noise) and the accuracy (the colloidal gold nanoparticles do not have a fixed size) are not sufficient for a definite and reliable calibration; however, the distinct features identified in Figure 6.10 are sufficient for the current purpose of the equipment.

6.1.5 Phase detector

When the quartz tuning fork is driven by an AC voltage, the tuning fork is oscillating and there is a small output current from the tuning fork because of its piezoelectric properties. There is a phase difference between the input AC voltage and the output current from the tuning fork, and this can be modelled as an RLC circuit. The phase difference is 90° when the frequency of the AC voltage is at the resonant frequency of the tuning fork. The resonant frequency of the quartz tuning fork will change upon a change in external environment, such as removing the cap of the tuning fork. Also, when the quartz tuning fork is moved closer to an etched tungsten tip, the interaction between the tuning fork and the tip changes the resonant frequency of the tuning fork. As the tuning fork is operated at a fixed frequency, the change in resonant frequency results in a change in phase difference. This phase difference is used as the feedback signal to the MK2-A810 board to determine whether the tip is approached and this is found by the phase detector.

The phase detector consists of several stages. A transimpedance amplifier (current-to-voltage converter) converts the output current of the tuning fork to a voltage signal (see Figure 6.11). The transimpedance amplifier is contained in a small diecast enclosure adjacent to the tuning fork to minimise noise and capacitive coupling of long cables. A resistive component is in parallel to the tuning fork and the transimpedance amplifier to cancel out any residual current when the tuning fork is not at resonance.

The current signal from the tuning fork is used to produce a DC voltage relating to the phase difference with the excitation voltage[1]. The resonant frequency of the tuning fork is determined by observing the frequency at which the maximum oscillating voltage or the maximum RMS voltage occurs in an

[1]This part of the circuit was built by Thibault Decoster.

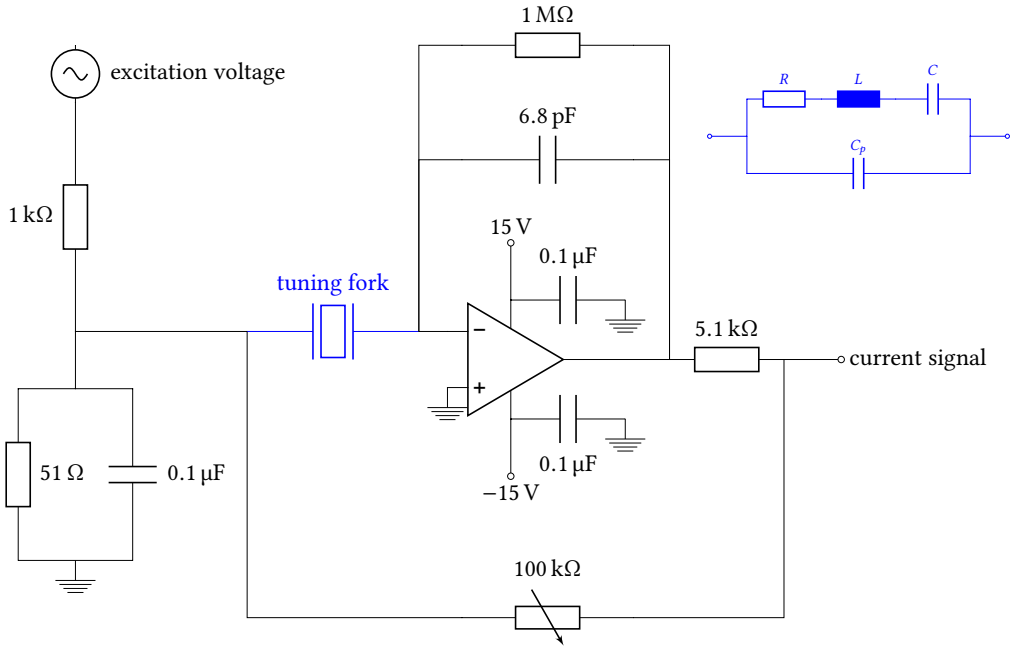
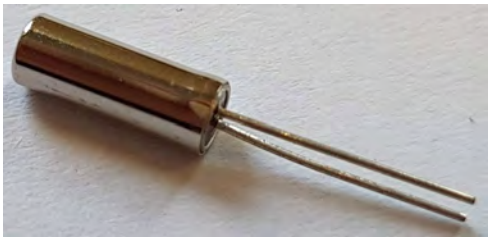


Figure 6.11: A voltage divider is used to decrease the magnitude of the excitation signal feeding into the tuning fork so that a smaller fork vibration can be achieved. The transimpedance amplifier is used to amplify the output current of the tuning fork, which has a phase shift of 90°. A further 90° phase shift is introduced after the signal is amplified and results in a 180° phase difference. The variable resistor is tuned to cancel out any current when the tuning fork is not at resonance which are from parasitic capacitances. The tuning fork is electrically equivalent to an RLC circuit in parallel to a capacitive circuit, which is shown in blue.

oscilloscope. A change in resonant frequency results in a change in phase, and the output voltage from the phase detector is changed accordingly. The voltage change caused by a phase change can be calibrated, for example, the voltage changes by 0.55 V when there is a 10° phase change as observed in the oscilloscope. The feedback system in the MK2-A810 board acts on this voltage and adjust the Z piezoelectric buzzer to compensate the change in resonant frequency.

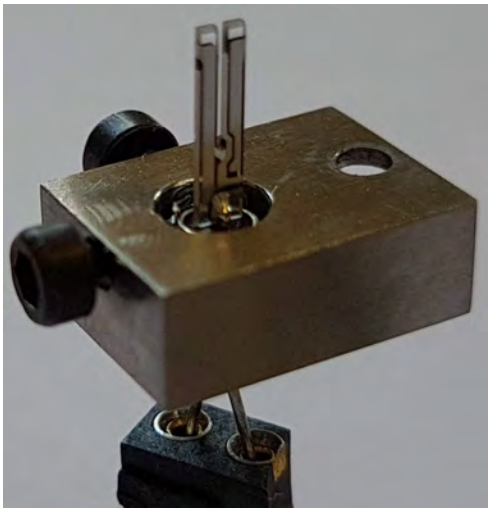
6.1.6 Samples on tuning fork

The tuning fork (RS, 547-6985) is de-capped by a rotary saw carefully to avoid damaging the lead wire and the prongs of the tuning fork. The de-capped

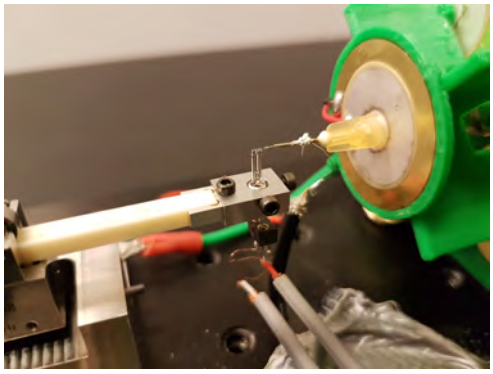


(a)

(b)



(c)



(d)

Figure 6.12: Some photos of the tuning fork and its metal holder are shown here. (a) the commercial tuning fork comes with a protective cap such that the tuning fork vibrates in a vacuum environment. (b) the prongs of the tuning fork are exposed by removing the cap and (c) the tuning fork is fixed by two screws in a metal holder with two pin connectors attached to the two prongs of the tuning fork, where the excitation signal and output current pass through. (d) shows the metal holder on the piezoelectric motor moving towards an etched tungsten tip.

tuning fork is mounted on a metal holder where the tuning fork is fixed by two screws, with one side of the prongs facing the tip. The excitation signal from the function generator (Tektronix, AFG3230) is applied to the lead wire of the tuning fork through two pin connectors (see Figure 6.12).

The sample needs to be attached on one of the two prongs of the tuning fork. This can be achieved by glueing the substrate on the prong. HOPG is chosen as the substrate in all the tests as it is atomically flat, inert in an ambient envir-

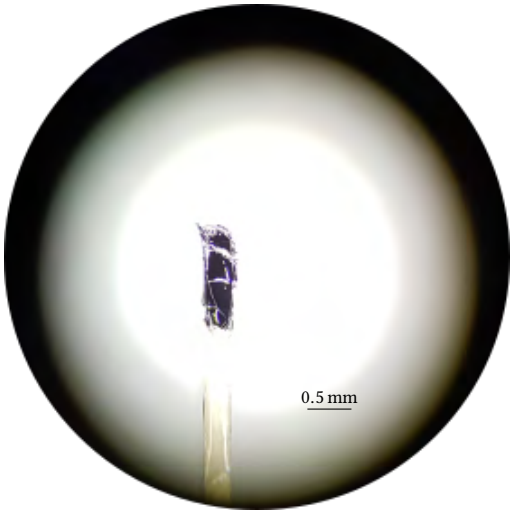


Figure 6.13: A piece of HOPG is glued on one of the prongs of a de-capped tuning fork and they are observed under an optical microscope.

onment and can be easily cleaved to obtain a clean surface. To glue the HOPG substrate on the prong, a tiny amount of 2-part epoxy glue (Araldite®) is applied to the prong. The prong with the glue is brought into contact with the HOPG substrate on a flat surface, and a gentle pressure is then applied to the prong. After complete drying, the tuning fork is separated meticulously from the HOPG substrate and a tiny piece of HOPG is fixed on the prong of the tuning fork. Residual HOPG is removed and a fresh layer of HOPG is obtained by removing the top layers by tape. A tuning fork with HOPG attached is shown in Figure 6.13.

6.2 Experimental data

The equipment was tested and data, including the changes of phase difference upon approach, was acquired. This section reports the procedures and the data acquired with the current design of the tuning fork AFM.

6.2.1 Z-spectroscopy

Z-spectroscopy allows probing the change of a property when the distance of the sample-tip junction is changing. The phase difference depends on the distance between the sample and the tip, so Z-spectroscopy is used to investigate its behaviour.

With the tuning fork excited at its resonant frequency, the phase difference remains 90° when the tip is far away from the surface of the tuning fork. While the tip is moving closer and closer, the tip-sample interaction is attractive in nature. The attractive interaction decreases the resonant frequency of the tuning fork; when the tip is moved even closer, the interaction becomes repulsive and this increases the resonant frequency.

The potential between two molecules and their separation can be represented by the Lennard-Jones potential model. The equation of Lennard-Jones potential can be written as:

$$V(r) = 4\epsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right] \tag{6.1}$$

r is the separation between two molecules, σ is the distance when the potential between the two nuclei is zero and ϵ is the depth of the potential well (see Figure 6.14).

The relationship of the phase difference and its distance between a bare tuning fork and a Pt–Ir tip was obtained by Z-spectroscopy and some acquired data are shown in Figure 6.15. The depth of the well is larger when the amplitude of the excitation signal is smaller as the chance to stay in the attractive regime is higher. When the tip is moved forwards and then backwards, the relationship of phase difference is, most of the time, not exactly the same (see Figure 6.16). The reason for the discrepancies could be the change of the geometry of the

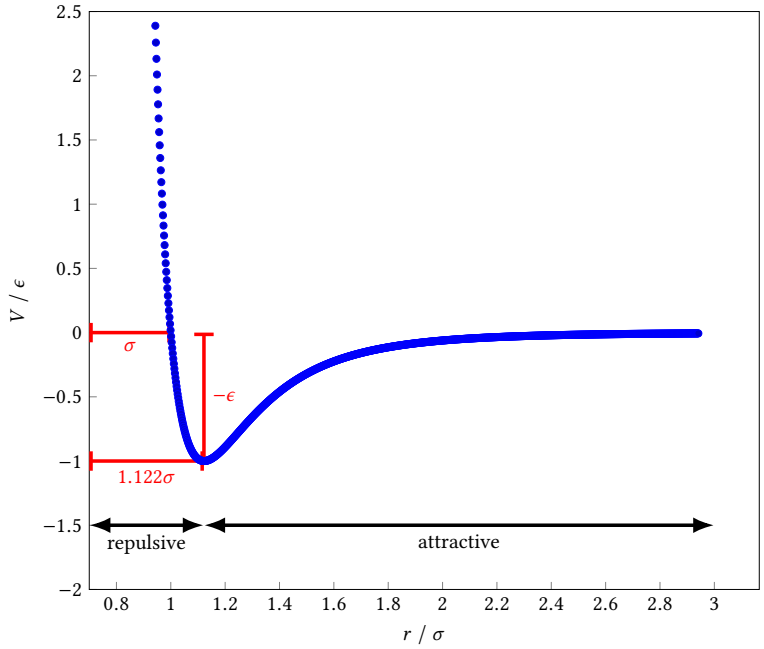
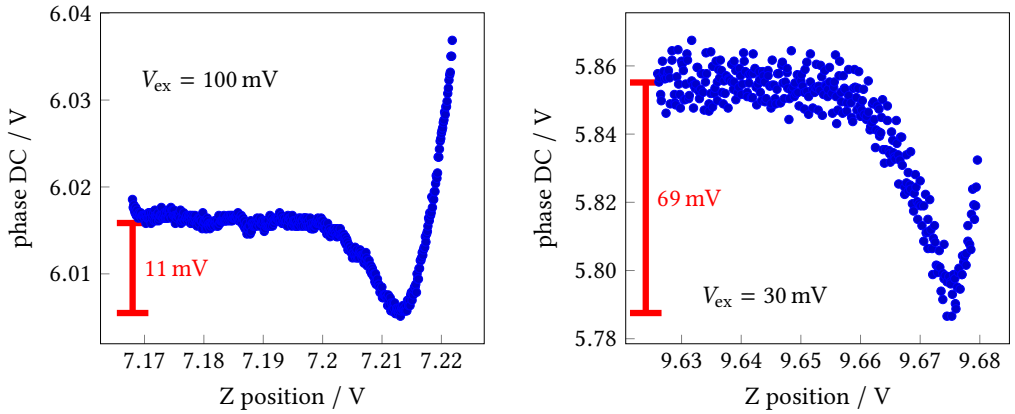


Figure 6.14: The Lennard-Jones potential is shown in term of σ and ϵ. When two nuclei are further apart, they are in attractive region while the two nuclei get closer and closer, they enter the repulsive region.

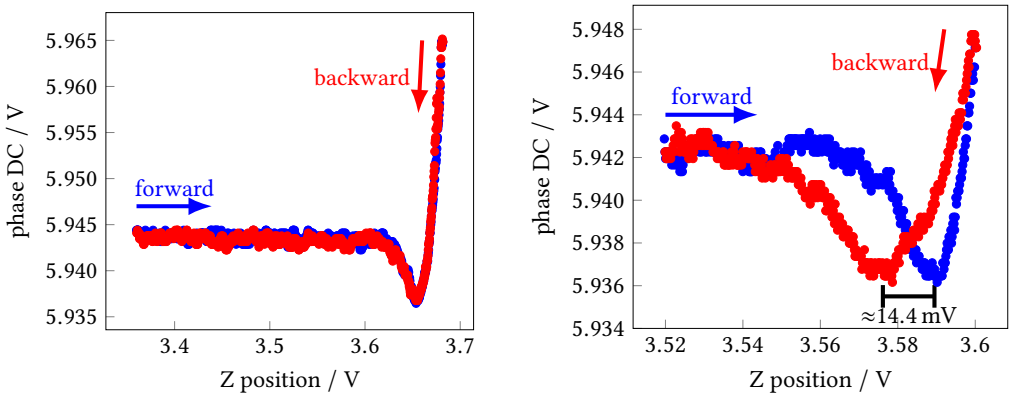


(a) excitation amplitude: 100 mV

(b) excitation amplitude: 30 mV

Figure 6.15: The relationship of phase difference between a tuning fork and a Pt–Ir tip with (a) 100 mV and (b) 30 mV excitation amplitude. A clear characteristic of Lennard-Jones potential can be seen. The depth of the well is larger if the excitation amplitude is smaller as the interaction in the attractive regime is stronger.

tip and/or the surface; the interacting area then becomes different when it is moving forwards or backwards.



(a) no change

(b) changed

Figure 6.16: The Pt–Ir tip is moved forwards and then backwards to the tuning fork and in (a), the recorded phase difference can be treated as identical which indicates the tip-sample interaction does not change; in (b), the position of the well is shifted by about 14.4 mV in this case and the shapes of the wells do not match. This indicates a slight change in the geometry of tip-sample junction.

6.2.2 Approach of tips

The primary purpose of the tuning fork AFM is to prepare samples for tungsten tip tomography. It is of utmost importance that the etched tungsten tips do not get damaged during approach and the pick-up of the nanoparticles. In view of this, etched tungsten tips before and after approach are examined.

To approach the tips, the excitation frequency of the tuning fork is tuned down very slightly such that the phase difference is above 90°, for example, at 90.1°. The reason to approach with a slightly offset excitation frequency is that the phase difference will decrease upon entering the attractive regime. It is desirable for the tip to stay in the attractive region and not to get too close to the surface; otherwise, the tip will suffer damage. With a phase difference slightly higher than 90° and the setpoint set at 90°, the phase difference will decrease until reaching the setpoint when the tip gets closer and closer.

The GXSM3 software does not support such approach strategy. The feedback can be set at negative to circumvent this, but after successful approach the tip

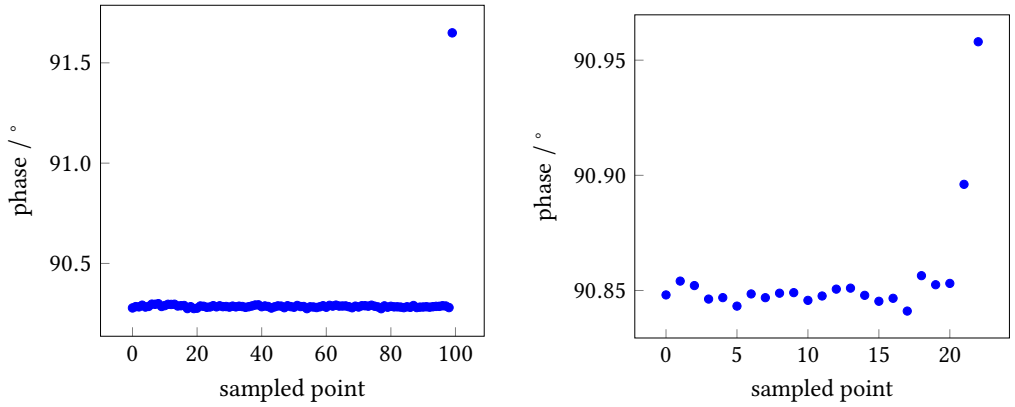


Figure 6.17: The phase difference is recorded when the tuning fork is approaching the etched tungsten tips. In (a), the phase difference abruptly jumped to a high value (≈91.65°) and this indicates repulsive region is reached immediately. The case in (b) entered the repulsive region gradually without showing a noticeable well. The sampling rate depends on the gains of the feedback.

will immediately crash into the surface if it is still at negative feedback; hence, a custom script was written to perform the approach. The aim is to approach without damaging the tip, so the repulsive regime should be avoided. The script will abort the approach when either the setpoint is reached or a particular phase difference is exceeded; in other words, it will abort the approach if it is going into the repulsive region. Figure 6.17 shows some examples of the change of phase difference during approach.

The tip usually cannot settle at the setpoint as the depth of the well is very small and varies with different tips; instead the approach is aborted when the tip is going into the repulsive region. The script cannot stop the approach immediately before going into the repulsive region. Figure 6.18 shows typical images of etched tungsten tips after using the script to approach, and it is either hooked or broken. The parameters in the script, including the speed of the approach and the abort condition, were adjusted; however, the tips were still damaged.

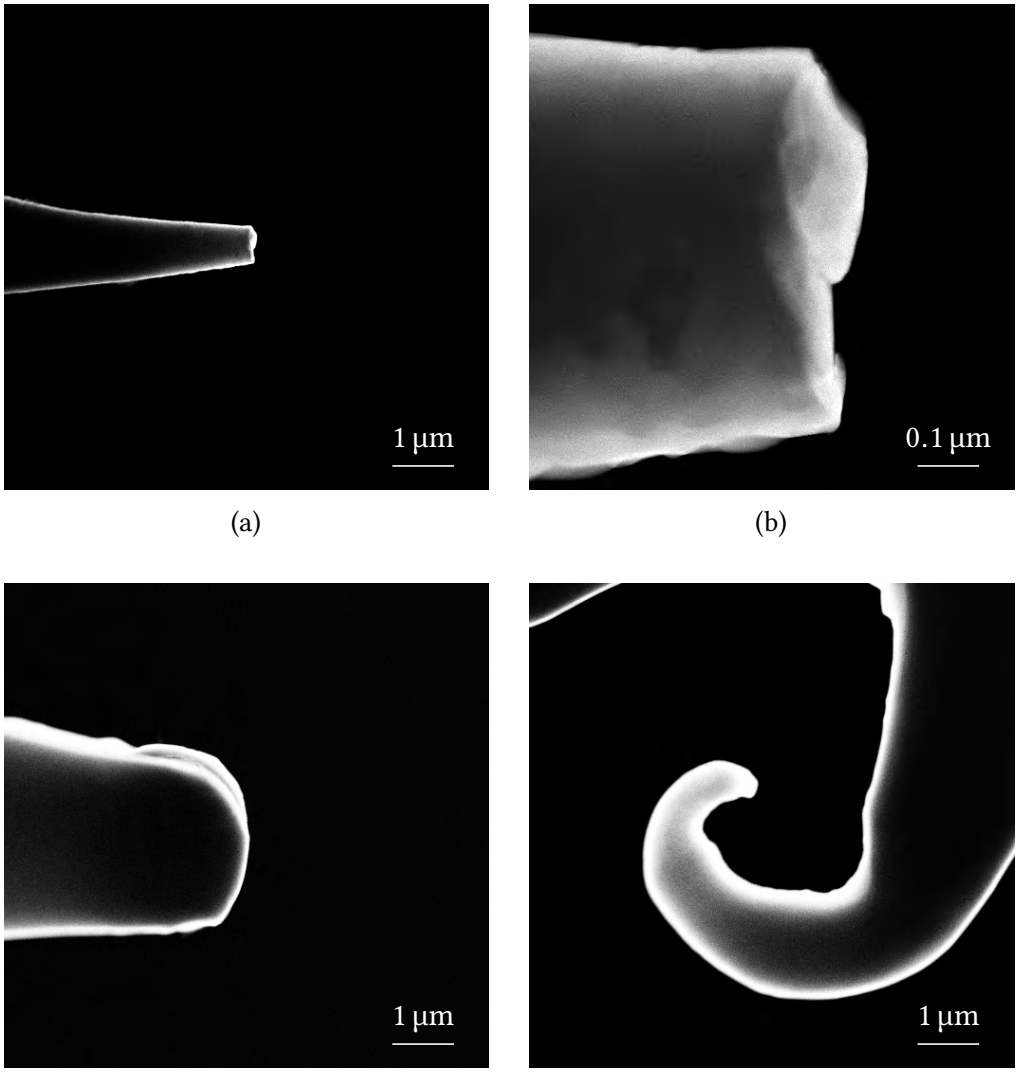


Figure 6.18: Images of etched tungsten tips after tuning fork AFM's approach are shown here. (a) and (b) show etched tungsten tips after approaching a bare tuning fork, and (c) and (d), to HOPG. All the tips are either broken or hooked.

There could be several reasons for the damage of the tips. The script is written in Python, which controls commands in a lower level (in DSP level). The additional layer could cause delay and may not be quick enough to abort the approach; therefore the approach only stops after crashing. The smallest step of the current linear piezoelectric motor is of the order of a few hundreds nm, and a single step may already crash the tip. Lastly, the phase difference fluctuates a lot in ambient condition and the tip may crash into the surface intermittently.

The issue about the stability of the phase difference is addressed in the following section.

6.2.3 Effects of ambient environment

For a successful approach and scanning, the phase difference cannot fluctuate more than 0.01° ideally as the phase difference is very sensitive to the interaction between the tip and the sample. However, the phase difference is very sensitive to the surrounding environment and this poses a serious problem on approach and scanning, especially after a long period of time. This section discusses the stability of the phase difference.

After the tuning fork is de-capped, the tolerance to temperature and the ambient condition decreases. Figure 6.19 shows the variation of phase difference when the as-received tuning fork and the phase detector are both inside a large thermal isolation box. The phase difference does not fluctuate more than 0.05°. There is a shielding box to prevent electromagnetic interference, and when this box is closed or opened with a de-capped tuning fork inside, the phase difference changes abruptly (see Figure 6.20). This shows that the electronics of the phase detector and the resonant frequency of the de-capped tuning fork are sensitive to the surrounding temperature and ambient conditions.

To address this issue, temperature control was introduced. A 2.7 Ω heating resistor was placed inside the shielding box in a bid to provide a stable temperature environment around the de-capped tuning fork. When a current passes through the heating resistor, the fluctuation of the phase difference is shown in Figure 6.21a. The period of the fluctuation was around 20 s, and this was related to the circulation of warm air inside the shielding box. To prevent the fluctuation, the heating resistor was mounted on a long metal rod further away from the de-capped tuning fork. The larger surface area of the metal rod can

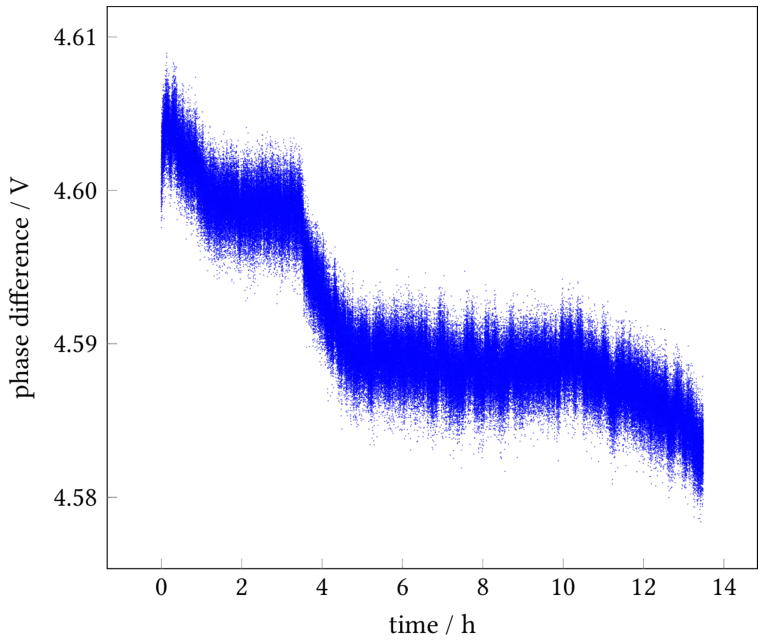


Figure 6.19: The phase difference was recorded at a rate of 7.5 Hz overnight while the as-received tuning fork and phase detector were placed in a thermal insulation box. The phase difference can be regarded as stable for a period of several hours. The decreasing trend is caused by the slight change of temperature inside the box and it does not fluctuate more than about 0.05° (equivalent to about 3 mV) after a stable temperature environment is reached.

provide more even heating inside the shielding box. The variation of phase difference with time in this case is shown in Figure 6.21b and the 20 s-fluctuation disappears. When the phase difference was followed for a longer duration, another fluctuation with a period of ≈90 min was observed. This was related to the air circulation inside the laboratory as the tuning fork AFM cannot be sealed off completely from its surrounding.

As a single step of the linear piezoelectric motor is very small, the approach could take around an hour and it could take another hour to identify a suitable nanoparticle on the substrate. It is essential the phase difference is stable for at least 2 h–3 h; otherwise, the tip will intermittently crash into the surface or even disengage from the substrate. The phase difference with a heating resistor was followed for more than 20 h and it is shown in Figure 6.22. A thermistor was placed nearby the tuning fork to measure the temperature (in V). With

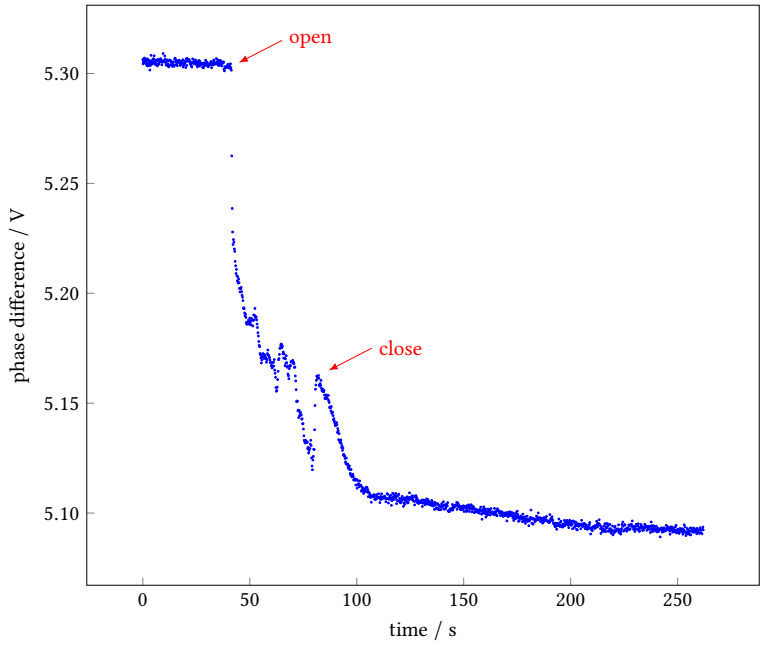
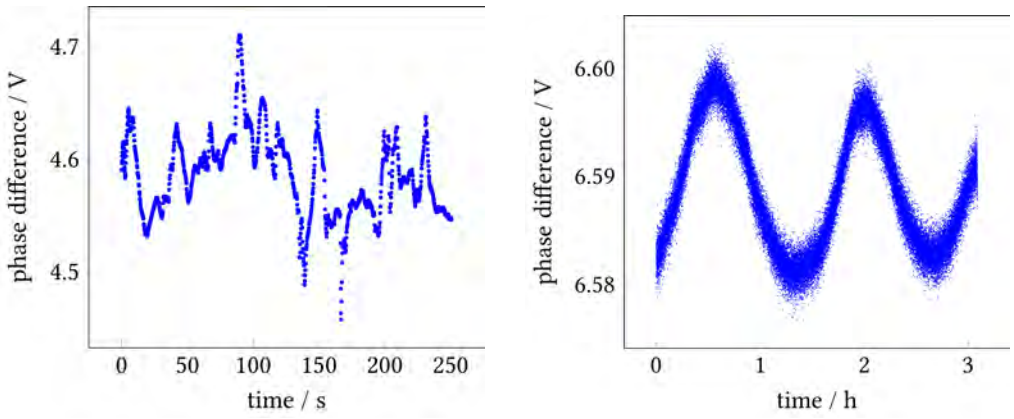


Figure 6.20: The phase difference changes abruptly when the shielding box is opened, as the air outside burst into the environment around tuning fork. When the box is closed, the phase change is less and eventually it settles over time.



(a) (b)

Figure 6.21: In (a), the heating resistor was placed beside the tuning fork and the fluctuation of the phase difference is due to the warm air circulation inside the shielding box (a period about 20 s). In (b), the heating resistor was placed on a long metal rod to distribute the heat evenly. The 20 s-period fluctuation disappeared when the tuning fork was also shielded. Another fluctuation of about 90 min was observed and this is due to air circulation of the room.

the heating resistor mounted on a metal rod inside the shielding box, the phase difference changes with the temperature which varies during a long period of

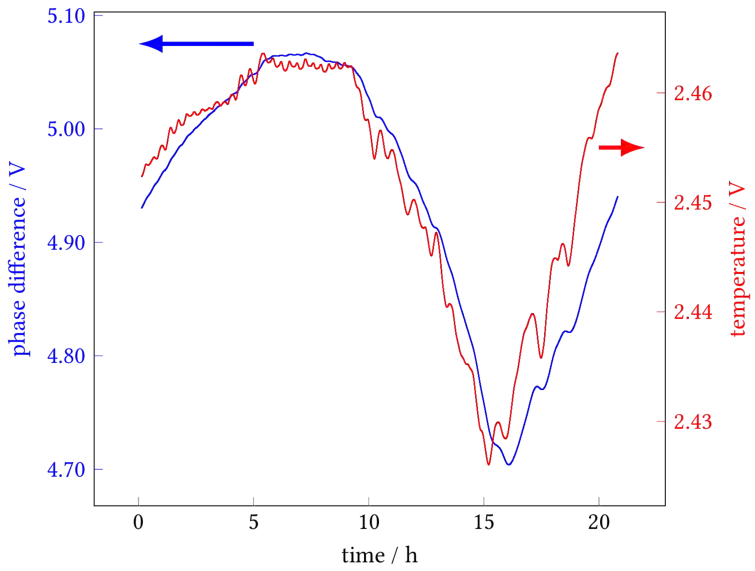


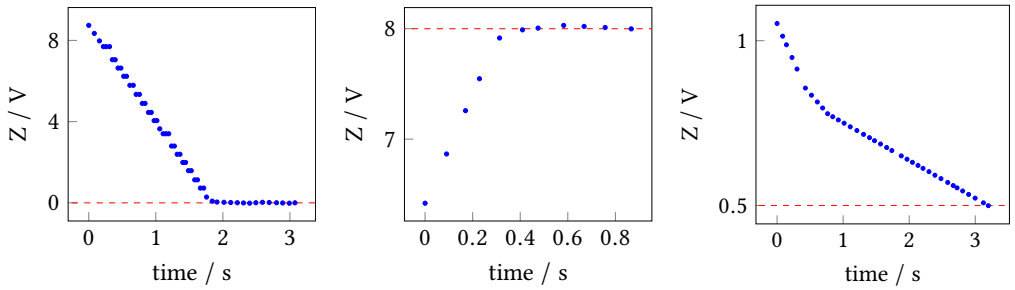
Figure 6.22: The phase difference was recorded when a heating resistor was placed on a metal rod and the tuning fork was also shielded from the heating resistor. The blue dots are the change of phase difference and the red dots are its temperature variation. The variation of phase difference is due to external temperature variation, as the shielding box cannot isolate the tuning fork completely and the heating resistor cannot maintain a temperature stable enough for the tuning fork.

time. A more isolated environment is needed to prevent the influence of room temperature fluctuation on the tuning fork.

6.2.4 Control of Z piezoelectric buzzer

Once a nanoparticle is identified by scanning, the tip needs to be brought closer to the nanoparticle so that the tip can pick it up. After moving the tip above the nanoparticle by adjusting the voltage to the X and Y piezoelectric buzzers, the Z piezoelectric buzzer needs to extend to bring the tip closer to the nanoparticle. The voltage to the Z piezoelectric buzzer cannot be controlled directly as it depends on the feedback control, unlike the X and Y piezoelectric buzzers.

The setting of Z-spectroscopy in the GXSM3 software only allows a fixed endpoint and does not abort the movement of Z piezoelectric buzzer if a certain phase difference is reached; therefore, a script was written to bring the voltage



(a) (b) (c)

Figure 6.23: The voltage of the Z piezoelectric buzzer can be changed to any arbitrary value so that it can extend or retract for picking up a nanoparticle. The dashed line shows the target which the Z piezoelectric buzzer is supposed to move to, and (a), (b) and (c) show a movement of Z from a large, intermediate and small distance from the target respectively. The mode of feedback is LIN.

of Z piezoelectric buzzer to a particular value while the phase difference is continuously monitored. This is achieved by changing the direction and magnitude of the feedback control which depend on the difference between the current and target voltage of Z piezoelectric buzzer. Figure 6.23 shows several examples of moving the Z piezoelectric buzzer. The script can reach the target voltage within a specified tolerance, and the speed to reach the target can be changed by using a different mode of feedback; 'LOG' mode is quicker while 'LIN' mode is slower. To move the tip towards the nanoparticle, 'LIN' mode is recommended as its movement is more gentle so that damage to the tip can be prevented.

6.3 Outlook

A tuning fork AFM is designed for the purpose of preparing samples for tungsten tip tomography and its operations are reported in Section 6.1 and Section 6.2. The tips suffered from damage after approach in the current design, no matter how the approach parameters were tuned. Although scanning is possible with damaged tips, these tips cannot be used to pick up a nanoparticle

for tomography. This section outlines some suggestions for the future design of the tuning fork AFM, aiming at approaching the tip without any damage.

The smallest step of the linear piezoelectric motor is of the order of a few hundreds nm [200]; after a full extend-retract cycle of the Z piezoelectric buzzer, a single step of the linear piezoelectric motor may just crash the tip into the substrate. Instead of using the handheld driver (see Figure 6.6b) to control the coarse movement of the linear piezoelectric motor, more precise controllers with sub-nanometer resolution could be used [202]. The reduced coarse step size during the approach of the tip can reduce the chance of crashing into the substrate. If the amplitude of oscillation of the tuning fork is too large, the chance of crashing into the etched tungsten tip is also increased. The typical excitation signal should be in the range of 1 mV–10 mV for an oscillation less than 1 nm [203, 204]. Although a voltage divider is used to decrease the magnitude of the excitation signal to ≈30 mV (see Figure 6.15), a lower magnitude of excitation signal is desired to minimise the chance of crashing; however, this is limited by the current function generator and electronic circuit due to a low signal-to-noise ratio when the excitation signal is at ≈10 mV.

Another main concern is the stability of the phase difference as its fluctuation will intermittently crash the tip and cause damage. Temperature control was introduced in a bid to stabilise the electronics of the phase detector and the surrounding environment of the tuning fork (see Section 6.2.3). Although the fluctuation of phase difference decreases significantly with temperature control (see Figure 6.21), it is still not at the scale of the depth of the potential well when the tip is moving towards the substrate, which is of the order of mV (see Figure 6.15). It is also shown in Figure 6.20 that the external environment can have a significant impact on the phase difference of a de-capped tuning fork.

The reason for the fluctuating phase difference is due to the change of resonant frequency when the surrounding temperature and environment are changing. This suggests a more isolated environment for the tuning fork AFM is needed. According to the data sheet, the variation of the resonant frequency of a tuning fork with temperature is $0.04\,\mathrm{ppm}/\Delta^{\circ}\mathrm{C}^2$ [205] — a change of 1 °C alters the resonant frequency by 1.31 mHz. The value in the data sheet is applicable to as-received tuning forks where the prongs of the tuning fork are in a vacuum environment; when the tuning fork is de-capped, the resonant frequency is much more sensitive to the change of temperature. The thermal stability of the components in the phase-detecting circuit can also affect the accuracy of the measurement of phase difference [206].

High-resolution topographical images can be obtained by tuning fork AFMs operated in an ultra-high vacuum (UHV) environment and often in cryogenic condition [207–209]. While our tuning fork AFM is not designed for high-resolution imaging, some of the design features in tuning fork AFMs for high-resolution imaging could be applicable to our tuning fork AFM. By housing the tuning fork AFM and the electronics inside a UHV environment, the fluctuation of the resonant frequency due to thermal fluctuation and ambient environmental change can certainly be reduced; however, to attain a UHV environment ($<10^{-9}$ mbar), the entire tuning fork AFM needs to be redesigned to ensure it is compatible with a UHV environment. As atomic resolution is not needed for this tuning fork AFM, a high vacuum (HV) environment (10^{-3}–10^{-9} mbar) should be sufficient. A more isolated environment can be achieved by constructing a vacuum chamber to house the tuning fork AFM and the electronics for detecting the phase difference.

The ‘qPlus’ configuration, where one prong of the tuning fork is fixed and the other prong is freely oscillated, is commonly used for high-resolution topo-

graphical imaging [210–213]. The qPlus configuration avoids the imbalance of mass when the substrate/tip is attached to one of the prongs of the tuning fork, which breaks the symmetry when it oscillates; by fixing one of the prongs, the tuning fork acts like a single oscillator and the problem of imbalance of mass is avoided [214–217]. The qPlus configuration can also be adapted to the design of the tuning fork AFM to increase the stability of oscillation, and no substantial modification is needed from the current design — the holder for the tuning fork could be redesigned to provide a hard surface to which one prong of the tuning fork can be fixed.

These suggestions aim to reduce the fluctuation of resonant frequency of the tuning fork such that the chance of tip damage can be minimised. It is expected that once the etched tungsten tip is not damaged during approach, a nanoparticle can be picked up by gently lowering the tip above the nanoparticle. A successful pick-up is indicated by a sudden change of the phase difference as the interaction between the tip and the substrate is changed when the nanoparticle is on the apex of the tip.

7

Conclusion

This thesis presents the method of tungsten tip tomography (see Chapter 2) for obtaining datasets in the full ±90° tilt range because it allows a free-standing nanoparticle to be imaged. This solves the long-standing 'missing wedge' problem in electron tomography (ET) and provides a better signal-to-noise ratio for the data. The procedures of tungsten tip tomography — tip and sample preparation, experimental set-up and data acquisition — are presented and demonstrated by a gold nanoparticle on the apex of the tungsten tip. The procedures of data processing before reconstruction are discussed in Chapter 3. The processing ensures scanning artefacts such as random drift and scanning-coil distortion are minimised, and the dataset is accurately aligned. These steps are essential for an accurate and reliable reconstruction with atomic precision. In Chapter 6, the development of a tuning fork atomic force microscope (AFM) for improving the sample preparation in tungsten tip tomography is presented. The design of the tuning fork AFM — the controller, piezoelectric control, tip approach — is discussed, and data including topographical information and Z-spectroscopy are presented. The issue of the stability of the tuning fork AFM is also discussed.

A new reconstruction algorithm, atomic site tomography (AST), is introduced in Chapter 4. AST is specifically designed for atomic electron tomography (AET) which aims to identify the precise locations of individual atoms in nanoparticles. The principles and steps of AST — construction of search grid, least absolute shrinkage and selection operator (lasso) fitting and refinement — are demonstrated by a simulated dataset in Chapter 4. In Chapter 5, an experimental dataset of a tungsten tip (published in [12, 189]) is selected to validate AST. Finally, a gold nanoparticle dataset acquired by tungsten tip tomography is reconstructed by AST and is presented in Chapter 5.

This work makes major contributions to AET by using a tungsten tip as a platform for acquiring high-quality datasets of nanoparticles, and developing a new reconstruction algorithm. The techniques in tungsten tip tomography are applicable to a wide variety of metallic nanoparticles, and the development of the tuning fork AFM aims to extend the range of samples that can be investigated. This has not been realised yet as the current stability of the tuning fork AFM is not sufficient for preventing the tips from damage upon approach. Section 6.3 makes some suggestions to address this issue.

Tungsten tip tomography opens up the possibility of investigating the interfacial atomic layers between tungsten and a nanoparticle in three-dimensional space. New datasets are needed to achieve this. It is also beneficial to reconstruct more datasets using AST to further demonstrate the applicability of AST in the field of AET.

List of references

[1] A. Stevenson, ed., *Oxford dictionary of English*, 3rd ed. (Oxford University Press, Oxford, 2011).

[2] S. Webb, 'A brief history of tomography and CT', in Proceedings 7th Asian & Oceanian congress of radiology (1995), p. 429.

[3] G. N. Hounsfield, 'A method of and apparatus for examination of a body by radiation such as X-or gamma-radiation', British Patent No. 1,283,915 (1972).

[4] P. Seynaeve and J. Broos, 'The history of tomography', Journal belge de radiologie **78**, 284–288 (1995).

[5] E. Ruska, 'The development of the electron microscope and of electron microscopy', Reviews of Modern Physics **59**, 627–638 (1987).

[6] D. De Rosier and A. Klug, 'Reconstruction of three dimensional structures from electron micrographs', Nature **217**, 130–134 (1968).

[7] M. Falke, U. Falke, A. Bleloch, S. Teichert, G. Beddies and H.-J. Hinneberg, 'Real structure of the CoSi2/Si(001) interface studied by dedicated aberration-corrected scanning transmission electron microscopy', Applied Physics Letters **86**, 203103 (2005).

[8] S. Bals, K. J. Batenburg, D. Liang, O. Lebedev, G. V. Tendeloo, A. Aerts, J. A. Martens and C. E. A. Kirschhock, 'Quantitative three-dimensional modeling of zeotile through discrete electron tomography', Journal of the American Chemical Society **131**, 4769–4773 (2009).

[9] I. MacLaren and G. Richter, 'Structure and possible origins of stacking faults in gamma-yttrium disilicate', Philosophical Magazine **89**, 169–181 (2009).

[10] I. MacLaren, L. Wang, A. J. Craven, Q. M. Ramasse, B. Schaffer, K. Kalantari and I. M. Reaney, 'The atomic structure and chemistry of Fe-rich steps on antiphase boundaries in Ti-doped Bi0.9Nd0.15FeO3', APL Materials **2**, 066106 (2014).

[11] M. C. Scott, C.-C. Chen, M. Mecklenburg, C. Zhu, R. Xu, P. Ercius, U. Dahmen, B. C. Regan and J. Miao, 'Electron tomography at 2.4-ångström resolution', Nature **483**, 444–447 (2012).

[12] R. Xu, C.-C. Chen, L. Wu, M. C. Scott, W. Theis, C. Ophus, M. Bartels, Y. Yang, H. Ramezani-Dakhel, M. R. Sawaya, H. Heinz, L. D. Marks, P. Ercius and J. Miao, 'Three-dimensional coordinates of individual atoms in materials revealed by electron tomography', Nature Materials **14**, 1099–1103 (2015).

[13] L. De Broglie, 'Recherches sur la théorie des Quanta', Theses (Migration - université en cours d'affectation, Nov. 1924).

[14] C. Davisson and L. H. Germer, 'Diffraction of electrons by a crystal of nickel', Physical Review **30**, 705–740 (1927).

[15]L. Reimer, *Transmission electron microscopy: physics of image formation and microanalysis*, edited by P. Hawkes, 4th ed., Vol. 36, Springer series in optical sciences (Springer-Verlag, Berlin Heidelberg, 1997).
[16]S. V. Aert, J. Verbeeck, R. Erni, S. Bals, M. Luysberg, D. V. Dyck and G. V. Tendeloo, 'Quantitative atomic resolution mapping using high-angle annular dark field scanning transmission electron microscopy', Ultramicroscopy **109**, 1236–1244 (2009).
[17]A. D. Backer, G. Martinez, K. MacArthur, L. Jones, A. Béché, P. Nellist and S. V. Aert, 'Dose limited reliability of quantitative annular dark field scanning transmission electron microscopy for nano-particle atom-counting', Ultramicroscopy **151**, 56–61 (2015).
[18]J. M. LeBeau, S. D. Findlay, L. J. Allen and S. Stemmer, 'Standardless atom counting in scanning transmission electron microscopy', Nano Letters **10**, 4405–4408 (2010).
[19]J. M. LeBeau, S. D. Findlay, L. J. Allen and S. Stemmer, 'Quantitative atomic resolution scanning transmission electron microscopy', Physical Review Letters **100**, 206101 (2008).
[20]A. Amali and P. Rez, 'Theory of lattice resolution in high-angle annular dark-field images', Microscopy and Microanalysis **3**, 28–46 (1997).
[21]S. Pennycook, 'Z-contrast stem for materials science', Ultramicroscopy **30**, 58–69 (1989).
[22]S. Pennycook and D. Jesson, 'High-resolution Z-contrast imaging of crystals', Ultramicroscopy **37**, 14–38 (1991).
[23]S. Pennycook, B. Rafferty and P. Nellist, 'Z-contrast imaging in an aberration-corrected scanning transmission electron microscope', Microscopy and Microanalysis **6**, 343–352 (2000).
[24]N. Tanaka, ed., *Scanning transmission electron microscopy of nanomaterials* (Imperial College Press, London, 2015).
[25]B. K. E. MacArthur, 'The use of annular dark-field scanning transmission electron microscopy for quantitative characterisation', Johnson Matthey Technology Review **60**, 117–131 (2016).
[26]S. Lee, Y. Oshima, H. Sawada, F. Hosokawa, E. Okunishi, T. Kaneyama, Y. Kondo, S. Niitaka, H. Takagi, Y. Tanishiro and K. Takayanagi, 'Counting lithium ions in the diffusion channel of an LiV2O4 crystal', Journal of Applied Physics **109**, 113530 (2011).
[27]S. Findlay, S. Azuma, N. Shibata, E. Okunishi and Y. Ikuhara, 'Direct oxygen imaging within a ceramic interface, with some observations upon the dark contrast at the grain boundary', Ultramicroscopy **111**, 285–289 (2011).
[28]E. Okunishi, H. Sawada and Y. Kondo, 'Experimental study of annular bright field (ABF) imaging using aberration-corrected scanning transmission electron microscopy (STEM)', Micron **43**, 538–544 (2012).
[29]O. Scherzer, 'Über einige Fehler von Elektronenlinsen', Zeitschrift für Physik **101**, 593–603 (1936).
[30]O. Scherzer, 'The theoretical resolution limit of the electron microscope', Journal of Applied Physics **20**, 20–29 (1949).

[31]A. Bleloch and A. Lupini, 'Imaging at the picoscale', Materials Today **7**, 42–48 (2004).
[32]H. Koops, G. Kuck and O. Scherzer, 'Erprobung eines elektronenoptischen Achromators', Optik **48**, 225–236 (1977).
[33]M. Haider, G. Braunshausen and E. Schwan, 'Correction of the spherical aberration of a 200 kV TEM by means of a hexapole-corrector', Optik **99**, 167–179 (1995).
[34]M. Haider, H. Rose, S. Uhlemann, E. Schwan, B. Kabius and K. Urban, 'A spherical-aberration-corrected 200kV transmission electron microscope', Ultramicroscopy **75**, 53–60 (1998).
[35]H. H. Rose, 'Historical aspects of aberration correction', Journal of Electron Microscopy **58**, 77–85 (2009).
[36]V. Lučić, F. Förster and W. Baumeister, 'Structural studies by electron tomography: from cells to molecules', Annual Review of Biochemistry **74**, 833–865 (2005).
[37]S. Brenner and R. Horne, 'A negative staining method for high resolution electron microscopy of viruses', Biochimica et Biophysica Acta **34**, 103–110 (1959).
[38]K. P. Ryan, 'Cryofixation of tissues for electron microscopy: a review of plunge cooling methods', Scanning Microscopy **6**, 715–743 (1992).
[39]M. Parthasarathy, 'Chapter 5 freeze-substitution', in *Methods in cell biology* (Elsevier, 1995), pp. 57–69.
[40]P. Penczek, M. Marko, K. Buttle and J. Frank, 'Double-tilt electron tomography', Ultramicroscopy **60**, 393–410 (1995).
[41]Thermo Fisher Scientific, *Single-Tilt Tomography Holder*, (2013) https://www.fei.com/accessories/Single-Tilt-Tomography-Holder/ (visited on 27/05/2019).
[42]Fischione Instruments Inc., *Model 2040 | Fischione*, (2013) https://www.fischione.com/products/holders/model-2040-dual-axis-tomography-holder (visited on 27/05/2019).
[43]Gatan Inc., *TEM Holders for Tomography*, (2008) http://www.gatan.com/products/tem-specimen-holders/tomography-holders (visited on 27/05/2019).
[44]J. Frank, ed., *Electron tomography: methods for three-dimensional visualization of structures in the cell*, 2nd ed. (Springer-Verlag, New York, 2006).
[45]J. F. Barrett and N. Keat, 'Artifacts in CT: recognition and avoidance', RadioGraphics **24**, 1679–1691 (2004).
[46]P. A. Midgley, M. Weyland, J. M. Thomas and B. F. G. Johnson, 'Z-contrast tomography: a technique in three-dimensional nanostructural analysis based on Rutherford scattering', Chemical Communications **10**, 907–908 (2001).
[47]Z. Saghi, X. Xu and G. Möbus, 'Electron tomography of regularly shaped nanostructures under non-linear image acquisition', Journal of Microscopy **232**, 186–195 (2008).

[48]X. Xu, Z. Saghi, R. Gay and G. Möbus, 'Reconstruction of 3D morphology of polyhedral nanoparticles', Nanotechnology **18**, 225501 (2007).
[49]W. V. den Broek, A. Rosenauer, B. Goris, G. Martinez, S. Bals, S. V. Aert and D. V. Dyck, 'Correction of non-linear thickness effects in HAADF STEM electron tomography', Ultramicroscopy **116**, 8–12 (2012).
[50]M. C. Altunbas, C. C. Shaw, L. Chen, C. Lai, X. Liu, T. Han and T. Wang, 'A post-reconstruction method to correct cupping artifacts in cone beam breast computed tomography', Medical Physics **34**, 3109–3118 (2007).
[51]R. Aveyard, Z. Zhong, K. Batenburg and B. Rieger, 'Optimizing experimental parameters for the projection requirement in HAADF-STEM tomography', Ultramicroscopy **177**, 84–90 (2017).
[52]J. Radon, 'Über die bestimmung von Funktionen durch ihre Integralwerte längs gewisser Mannigfaltigkeiten', Ber. Verh. Sächs. Akad **69**, 262–277 (1917).
[53]L. A. Shepp and B. F. Logan, 'The Fourier reconstruction of a head section', IEEE Transactions on Nuclear Science **21**, 21–43 (1974).
[54]R. Bracewell, *The Fourier transform and its applications*, 3rd ed., McGraw-Hill Series in Electrical and Computer Engineering. Circuits and Systems (McGraw-Hill Higher Education, United States of America, June 1999).
[55]R. A. Crowther, D. J. DeRosier and A. Klug, 'The reconstruction of a three-dimensional structure from projections and its application to electron microscopy', Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences **317**, 319–340 (1970).
[56]P. R. Smith, T. M. Peters and R. H. T. Bates, 'Image reconstruction from finite numbers of projections', Journal of Physics A: Mathematical, Nuclear and General **6**, 361–382 (1973).
[57]A. C. Kak and M. Slaney, *Principles of computerized tomographic imaging*, Classics in Applied Mathematics (Society for Industrial and Applied Mathematics, New York, Jan. 2001).
[58]G. N. Ramachandran and A. V. Lakshminarayanan, 'Three-dimensional reconstruction from radiographs and electron micrographs: application of convolutions instead of Fourier transforms', Proceedings of the National Academy of Sciences **68**, 2236–2240 (1971).
[59]E. Tanaka, 'Generalised correction functions for convolutional techniques in three-dimensional image reconstruction', Physics in Medicine and Biology **24**, 157–161 (1979).
[60]B. Schorr and D. Townsend, 'Filters for three-dimensional limited-angle tomography', Physics in Medicine and Biology **26**, 305–312 (1981).
[61]G. Harauz and M. van Heel, 'Exact filters for general geometry three dimensional reconstruction', in Proceedings of the IEEE computer vision and pattern recognition conf, Vol. 73 (1986), pp. 146–156.
[62]S. Karczmarz, 'Angenaherte auflosung von systemen linearer glei-chungen', Bull. Int. Acad. Pol. Sic. Let., Cl. Sci. Math. Nat. **3**, 355–357 (1937).

[63]R. Gordon, R. Bender and G. T. Herman, 'Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and X-ray photography', Journal of Theoretical Biology **29**, 471–481 (1970).
[64]K. Tanabe, 'Projection method for solving a singular system of linear equations and its applications', Numerische Mathematik **17**, 203–214 (1971).
[65]R. S. Ramakrishnam, S. K. Mullick, R. K. S. Rathore and R. Subramanian, 'Orthogonalization, bernstein polynomials, and image restoration', Applied Optics **18**, 464–468 (1979).
[66]M. Slaney and A. C. Kak, *Imaging with diffraction tomography*, tech. rep. 540 (Department of Electrical and Computer Engineering, Purdue University, 1985).
[67]R. Gordon, 'A tutorial on ART (algebraic reconstruction techniques)', IEEE Transactions on Nuclear Science **21**, 78–93 (1974).
[68]K. T. Smith, D. C. Solmon and S. L. Wagner, 'Practical and mathematical aspects of the problem of reconstructing objects from radiographs', Bulletin of the American Mathematical Society **83**, 1227–1270 (1977).
[69]P. Gilbert, 'Iterative methods for the three-dimensional reconstruction of an object from projections', Journal of Theoretical Biology **36**, 105–117 (1972).
[70]A. H. Andersen and A. C. Kak, 'Simultaneous algebraic reconstruction technique (SART): a superior implementation of the Art algorithm', Ultrasonic Imaging **6**, 81–94 (1984).
[71]M. Jiang and G. Wang, 'Convergence of the simultaneous algebraic reconstruction technique (SART)', IEEE Transactions on Image Processing **12**, 957–961 (2003).
[72]M. Yan, 'Convergence analysis of SART by bregman iteration and dual gradient descent', UCLA CAM report, 10–27 (2010).
[73]Y. Censor, D. Gordon and R. Gordon, 'Component averaging: an efficient iterative parallel algorithm for large and sparse unstructured problems', Parallel Computing **27**, 777–808 (2001).
[74]Y. Censor, T. Elfving, G. T. Herman and T. Nikazad, 'On diagonally relaxed orthogonal projection methods', SIAM Journal on Scientific Computing **30**, 473–504 (2008).
[75]K. J. Batenburg and J. Sijbers, 'DART: a practical reconstruction algorithm for discrete tomography', IEEE Transactions on Image Processing **20**, 2542–2553 (2011).
[76]G. Wang and M. Jiang, 'Ordered-subset simultaneous algebraic reconstruction techniques (OS-SART)', Journal of X-ray Science and Technology **12**, 169–177 (2004).
[77]V. Israel-Jost, 'FA-SART: a frequency-adaptive algorithm in pinhole SPECT tomography', SIAM Journal on Scientific Computing **30**, 819–836 (2008).
[78]R. Leary, Z. Saghi, P. A. Midgley and D. J. Holland, 'Compressed sensing electron tomography', Ultramicroscopy **131**, 70–91 (2013).
[79]D. Donoho, 'Compressed sensing', IEEE Transactions on Information Theory **52**, 1289–1306 (2006).

[80]D. L. Donoho, 'For most large underdetermined systems of linear equations the minimal l1-norm solution is also the sparsest solution', Communications on Pure and Applied Mathematics **59**, 797–829 (2006).
[81]E. J. Candés and M. Wakin, 'An introduction to compressive sampling', IEEE Signal Processing Magazine **25**, 21–30 (2008).
[82]Y. C. Eldar and G. Kutyniok, *Compressed sensing: theory and applications*, 1st ed. (Cambridge University Press, New York, 2012).
[83]Z. Saghi, D. J. Holland, R. Leary, A. Falqui, G. Bertoni, A. J. Sederman, L. F. Gladden and P. A. Midgley, 'Three-dimensional morphology of iron oxide nanoparticles with reactive concave surfaces. a compressed sensing-electron tomography (CS-ET) approach', Nano Letters **11**, 4666–4673 (2011).
[84]B. Goris, W. V. den Broek, K. Batenburg, H. H. Mezerji and S. Bals, 'Electron tomography based on a total variation minimization reconstruction technique', Ultramicroscopy **113**, 120–130 (2012).
[85]N. Monsegue, X. Jin, T. Echigo, G. Wang and M. Murayama, 'Three-dimensional characterization of iron oxide (α-Fe2O3) nanoparticles: application of a compressed sensing inspired reconstruction algorithm to electron tomography', Microscopy and Microanalysis **18**, 1362–1367 (2012).
[86]H. Rauhut, K. Schnass and P. Vandergheynst, 'Compressed sensing and redundant dictionaries', IEEE Transactions on Information Theory **54**, 2210–2219 (2008).
[87]J. Mairal, M. Elad and G. Sapiro, 'Sparse representation for color image restoration', IEEE Transactions on Image Processing **17**, 53–69 (2008).
[88]G. Peyre, 'A review of adaptive image representations', IEEE Journal of Selected Topics in Signal Processing **5**, 896–911 (2011).
[89]B. Liu, H. Yu, S. S. Verbridge, L. Sun and G. Wang, 'Dictionary-learning-based reconstruction method for electron tomography', Scanning **36**, 377–383 (2014).
[90]A. Al-Afeef, W. P. Cockshott, I. MacLaren and S. McVitie, 'Electron tomography image reconstruction using data-driven adaptive compressed sensing', Scanning **38**, 251–276 (2015).
[91]P. E. Batson, N. Dellby and O. L. Krivanek, 'Sub-ångstrom resolution using aberration corrected electron optics', Nature **418**, 617–620 (2002).
[92]R. Erni, M. D. Rossell, C. Kisielowski and U. Dahmen, 'Atomic-resolution imaging with a sub-50-pm electron probe', Physical Review Letters **102**, 096101 (2009).
[93]J. S. Barnard, 'High-resolution three-dimensional imaging of dislocations', Science **313**, 319–319 (2006).
[94]H. L. Xin, P. Ercius, K. J. Hughes, J. R. Engstrom and D. A. Muller, 'Three-dimensional imaging of pore structures inside low-κ dielectrics', Applied Physics Letters **96**, 223108 (2010).
[95]S. V. Aert, K. J. Batenburg, M. D. Rossell, R. Erni and G. V. Tendeloo, 'Three-dimensional atomic imaging of crystalline nanoparticles', Nature **470**, 374–377 (2011).

[96]C.-C. Chen, C. Zhu, E. R. White, C.-Y. Chiu, M. C. Scott, B. C. Regan, L. D. Marks, Y. Huang and J. Miao, 'Three-dimensional imaging of dislocations in a nanoparticle at atomic resolution', Nature **496**, 74–77 (2013).
[97]M. Azubel, J. Koivisto, S. Malola, D. Bushnell, G. L. Hura, A. L. Koh, H. Tsunoyama, T. Tsukuda, M. Pettersson, H. Hakkinen and R. D. Kornberg, 'Electron microscopy of gold nanoparticles at atomic resolution', Science **345**, 909–912 (2014).
[98]G. Haberfehlner, P. Thaler, D. Knez, A. Volk, F. Hofer, W. E. Ernst and G. Kothleitner, 'Formation of bimetallic clusters in superfluid helium nanodroplets analysed by atomic resolution electron tomography', Nature Communications **6**, 8779 (2015).
[99]B. Goris, J. D. Beenhouwer, A. D. Backer, D. Zanaga, K. J. Batenburg, A. Sánchez-Iglesias, L. M. Liz-Marzán, S. V. Aert, S. Bals, J. Sijbers and G. V. Tendeloo, 'Measuring lattice strain in three dimensions through electron microscopy', Nano Letters **15**, 6996–7001 (2015).
[100]Y. Yang, C.-C. Chen, M. C. Scott, C. Ophus, R. Xu, A. Pryor, L. Wu, F. Sun, W. Theis, J. Zhou, M. Eisenbach, P. R. C. Kent, R. F. Sabirianov, H. Zeng, P. Ercius and J. Miao, 'Deciphering chemical order/disorder and material properties at the single-atom level', Nature **542**, 75–79 (2017).
[101]K. Batenburg, 'A new algorithm for 3D binary tomography', Electronic Notes in Discrete Mathematics **20**, 247–261 (2005).
[102]J. Jinschek, K. Batenburg, H. Calderon, R. Kilaas, V. Radmilovic and C. Kisielowski, '3-D reconstruction of the atomic positions in a simulated gold nanocrystal based on discrete tomography: prospects of atomic resolution electron tomography', Ultramicroscopy **108**, 589–604 (2008).
[103]D. V. Dyck and M. O. de Beeck, 'A simple intuitive theory for electron diffraction', Ultramicroscopy **64**, 99–107 (1996).
[104]P. Nellist and S. Pennycook, 'Incoherent imaging using dynamically scattered coherent electrons', Ultramicroscopy **78**, 111–124 (1999).
[105]L. Allen, S. Findlay, M. Oxley and C. Rossouw, 'Lattice-resolution contrast from a focused coherent electron probe. part I', Ultramicroscopy **96**, 47–63 (2003).
[106]S. Findlay, L. Allen, M. Oxley and C. Rossouw, 'Lattice-resolution contrast from a focused coherent electron probe. part II', Ultramicroscopy **96**, 65–81 (2003).
[107]C. Rossouw, L. Allen, S. Findlay and M. Oxley, 'Channelling effects in atomic resolution STEM', Ultramicroscopy **96**, 299–312 (2003).
[108]T. Morimura and M. Hasaka, 'Bloch-wave-based STEM image simulation with layer-by-layer representation', Ultramicroscopy **109**, 1203–1209 (2009).
[109]J. Miao, F. Förster and O. Levi, 'Equally sloped tomography with oversampling reconstruction', Physical Review B **72**, 052103 (2005).
[110]L. Almeida, 'The fractional Fourier transform and time-frequency representations', IEEE Transactions on Signal Processing **42**, 3084–3091 (1994).

[111]A. Averbuch, R. R. Coifman, D. L. Donoho, M. Israeli and Y. Shkolnisky, 'A framework for discrete integral transformations I—the pseudopolar Fourier transform', SIAM Journal on Scientific Computing **30**, 764–784 (2008).
[112]J. Miao, D. Sayre and H. N. Chapman, 'Phase retrieval from the magnitude of the Fourier transforms of nonperiodic objects', Journal of the Optical Society of America A **15**, 1662–1669 (1998).
[113]J. Miao, J. Kirz and D. Sayre, 'The oversampling phasing method', Acta Crystallographica Section D Biological Crystallography **56**, 1312–1315 (2000).
[114]J. Miao, T. Ishikawa, B. Johnson, E. H. Anderson, B. Lai and K. O. Hodgson, 'High resolution 3D X-ray diffraction microscopy', Physical Review Letters **89**, 088303 (2002).
[115]J. Miao, H. N. Chapman, J. Kirz, D. Sayre and K. O. Hodgson, 'Taking X-ray diffraction to the limit: macromolecular structures from femtosecond X-ray pulses and diffraction microscopy of cells with synchrotron radiation', Annual Review of Biophysics and Biomolecular Structure **33**, 157–176 (2004).
[116]R. W. Gerchberg and W. O. Saxton, 'A practical algorithm for the determination of phase from image and diffraction plane pictures', Optik **35**, 237–246 (1972).
[117]J. R. Fienup, 'Reconstruction of an object from the modulus of its Fourier transform', Optics Letters **3**, 27–29 (1978).
[118]E. Lee, B. P. Fahimian, C. V. Iancu, C. Suloway, G. E. Murphy, E. R. Wright, D. Castaño-Díez, G. J. Jensen and J. Miao, 'Radiation dose reduction and image enhancement in biological imaging through equally-sloped tomography', Journal of Structural Biology **164**, 221–227 (2008).
[119]Y. Mao, B. P. Fahimian, S. J. Osher and J. Miao, 'Development and optimization of regularized tomographic reconstruction algorithms utilizing equally-sloped tomography', IEEE Transactions on Image Processing **19**, 1259–1268 (2010).
[120]Y. Zhao, E. Brun, P. Coan, Z. Huang, A. Sztrokay, P. C. Diemoz, S. Liebhardt, A. Mittone, S. Gasilov, J. Miao and A. Bravin, 'High-resolution, low-dose phase contrast X-ray tomography for 3D diagnosis of human breast cancers', Proceedings of the National Academy of Sciences **109**, 18290–18294 (2012).
[121]B. P. Fahimian, Y. Zhao, Z. Huang, R. Fung, Y. Mao, C. Zhu, M. Khatonabadi, J. J. DeMarco, S. J. Osher, M. F. McNitt-Gray and J. Miao, 'Radiation dose reduction in medical X-ray CT via Fourier-based iterative reconstruction', Medical Physics **40**, 031914 (2013).
[122]A. Pryor, Y. Yang, A. Rana, M. Gallagher-Jones, J. Zhou, Y. H. Lo, G. Melinte, W. Chiu, J. A. Rodriguez and J. Miao, 'GENFIRE: a generalized Fourier iterative reconstruction algorithm for high-resolution 3D imaging', Scientific Reports **7**, 10409 (2017).
[123]R. Franke, 'Scattered data interpolation: tests of some methods', Mathematics of Computation **38**, 181–200 (1982).

[124]D. Shepard, 'A two-dimensional interpolation function for irregularly-spaced data', in Proceedings of the 1968 23rd ACM national conference (1968), pp. 517–524.
[125]J. Zhou, Y. Yang, Y. Yang, D. S. Kim, A. Yuan, X. Tian, C. Ophus, F. Sun, A. K. Schmid, M. Nathanson, H. Heinz, Q. An, H. Zeng, P. Ercius and J. Miao, 'Capturing nucleation at 4D atomic resolution', arXiv preprint arXiv:1807.10709 (2018).
[126]X. Tian, D. S. Kim, S. Yang, C. J. Ciccarino, Y. Gong, Y. Yang, Y. Yang, B. Duschatko, P. M. Ajayan, J.-C. Idrobo, P. Narang and J. Miao, 'Correlating 3D crystal defects and electronic properties of 2D materials at the single-atom level', arXiv e-prints, arXiv:1901.00633, arXiv:1901.00633 (2019).
[127]D. Lei, A. E. Marras, J. Liu, C.-M. Huang, L. Zhou, C. E. Castro, H.-J. Su and G. Ren, 'Three-dimensional structural dynamics of DNA origami bennett linkages using individual-particle electron tomography', Nature Communications **9**, 592 (2018).
[128]D. Jung, L. M. A. Saleh, Z. J. Berkson, M. F. El-Kady, J. Y. Hwang, N. Mohamed, A. I. Wixtrom, E. Titarenko, Y. Shao, K. McCarthy, J. Guo, I. B. Martini, S. Kraemer, E. C. Wegener, P. Saint-Cricq, B. Ruehle, R. R. Langeslay, M. Delferro, J. L. Brosmer, C. H. Hendon, M. Gallagher-Jones, J. Rodriguez, K. W. Chapman, J. T. Miller, X. Duan, R. B. Kaner, J. I. Zink, B. F. Chmelka and A. M. Spokoyny, 'A molecular cross-linking approach for hybrid metal oxides', Nature Materials **17**, 341–348 (2018).
[129]P. Bühlmann and S. Van De Geer, *Statistics for high-dimensional data: methods, theory and applications* (Springer-Verlag, Berlin Heidelberg, 2011).
[130]R. Tibshirani, 'Regression shrinkage and selection via the lasso', Journal of the Royal Statistical Society: Series B (Methodological) **58**, 267–288 (1996).
[131]B. Efron, T. Hastie, I. Johnstone and R. Tibshirani, 'Least angle regression', The Annals of Statistics **32**, 407–499 (2004).
[132]M. R. Osborne, B. Presnell and B. A. Turlach, 'A new approach to variable selection in least squares problems', IMA Journal of Numerical Analysis **20**, 389–403 (2000).
[133]J. Friedman, T. Hastie, H. Höfling and R. Tibshirani, 'Pathwise coordinate optimization', The Annals of Applied Statistics **1**, 302–332 (2007).
[134]D. L. Donoho and I. M. Johnstone, 'Adapting to unknown smoothness via wavelet shrinkage', Journal of the American Statistical Association **90**, 1200–1224 (1995).
[135]R. J. Tibshirani, 'The lasso problem and uniqueness', Electronic Journal of Statistics **7**, 1456–1490 (2013).
[136]F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and E. Duchesnay, 'Scikit-learn: machine learning in Python', Journal of Machine Learning Research **12**, 2825–2830 (2011).

[137]J. Friedman, T. Hastie and R. Tibshirani, 'Regularization paths for generalized linear models via coordinate descent', Journal of Statistical Software **33**, 1–22 (2010).
[138]G. Binnig, H. Rohrer, C. Gerber and E. Weibel, 'Surface studies by scanning tunneling microscopy', Physical Review Letters **49**, 57–61 (1982).
[139]C. J. Chen, *Introduction to scanning tunneling microscopy*, edited by M. Lapp, J.-I. Nishizawa, B. B. Snavely, H. Stark, A. C. Tam and T. Wilson, Vol. 4, Oxford series in optical and imaging sciences (Oxford University Press, New York, 1993).
[140]J. Tersoff and D. R. Hamann, 'Theory and application for the scanning tunneling microscope', Physical Review Letters **50**, 1998–2001 (1983).
[141]G. Binnig, C. F. Quate and C. Gerber, 'Atomic force microscope', Physical Review Letters **56**, 930–933 (1986).
[142]F. J. Giessibl, 'Advances in atomic force microscopy', Reviews of Modern Physics **75**, 949–983 (2003).
[143]C.-E. Hsieh, A. Leith, C. A. Mannella, J. Frank and M. Marko, 'Towards high-resolution three-dimensional imaging of native mammalian tissue: electron tomography of frozen-hydrated rat liver sections', Journal of Structural Biology **153**, 1–13 (2006).
[144]A. AlAfeef, J. Bobynko, W. P. Cockshott, A. J. Craven, I. Zuazo, P. Barges and I. MacLaren, 'Linear chemically sensitive electron tomography using DualEELS and dictionary-based compressed sensing', Ultramicroscopy **170**, 96–106 (2016).
[145]T. Yaguchi, M. Konno, T. Kamino and M. Watanabe, 'Observation of three-dimensional elemental distributions of a Si device using a 360°-tilt FIB and the cold field-emission STEM system', Ultramicroscopy **108**, 1603–1615 (2008).
[146]E. Biermans, L. Molina, K. J. Batenburg, S. Bals and G. V. Tendeloo, 'Measuring porosity at the nanoscale by quantitative electron tomography', Nano Letters **10**, 5014–5019 (2010).
[147]X. Ke, S. Bals, D. Cott, T. Hantschel, H. Bender and G. V. Tendeloo, 'Three-dimensional analysis of carbon nanotube networks in interconnects by electron tomography without missing wedge artifacts', Microscopy and Microanalysis **16**, 210–217 (2010).
[148]M. Andrzejczuk, A. Roguska, M. Pisarek, M. Hołdyński, M. Lewandowska and K. Kurzydłowski, 'Morphology of TiO2 nanotubes revealed through electron tomography', Micron **95**, 35–41 (2017).
[149]E. Padgett, R. Hovden, J. C. DaSilva, B. D. A. Levin, J. L. Grazul, T. Hanrath and D. A. Muller, 'A simple preparation method for full-range electron tomography of nanoparticles and fine powders', Microscopy and Microanalysis **23**, 1150–1158 (2017).
[150]S. Subramaniam, 'The SIV surface spike imaged by electron tomography: one leg or three?', PLoS Pathogens **2**, e91 (2006).
[151]P. Ercius and W. Theis, 'Tungsten tips as a sample platform for single atom resolution S/TEM tomography of clusters and interfaces', Microscopy and Microanalysis **19**, 532–533 (2013).

[152]M. Scott, W. Theis, R. Xu, L. Wu, C.-C. Chen, C. Ophus, P. Ercius and J. Miao, 'Implementation of atomic resolution electron tomography of a needle sample', Microscopy and Microanalysis **21**, 1523–1524 (2015).
[153]H. Gu, G. Li, C. Liu, F. Yuan, F. Han, L. Zhang and S. Wu, 'Considerable knock-on displacement of metal atoms under a low energy electron beam', Scientific Reports **7**, 184 (2017).
[154]R. Egerton, P. Li and M. Malac, 'Radiation damage in the TEM and SEM', Micron **35**, 399–409 (2004).
[155]R. Egerton, R. McLeod, F. Wang and M. Malac, 'Basic questions related to electron-induced sputtering in the TEM', Ultramicroscopy **110**, 991–997 (2010).
[156]M. Greiner and P. Kruse, 'Recrystallization of tungsten wire for fabrication of sharp and stable nanoprobe and field-emitter tips', Review of Scientific Instruments **78**, 026104 (2007).
[157]E. O. Hall, 'The deformation and ageing of mild steel: III discussion of results', Proceedings of the Physical Society. Section B **64**, 747–753 (1951).
[158]N. Petch, 'The cleavage strength of polycrystals', Journal of the Iron and Steel Institute **174**, 25–28 (1953).
[159]D. B. Snow, 'The recrystallization of commercially pure and doped tungsten wire drawn to high strain', Metallurgical Transactions A **10**, 815–821 (1979).
[160]S. Leber, J. Tavernelli, D. White and R. Hehemann, 'Fracture modes in tungsten wire', Journal of the Less Common Metals **48**, 119–133 (1976).
[161]P. Zhao, J. Riesch, T. Höschen, J. Almanstötter, M. Balden, J. Coenen, R. Himml, W. Pantleon, U. von Toussaint and R. Neu, 'Microstructure, mechanical behaviour and fracture of pure tungsten wire after different heat treatments', International Journal of Refractory Metals and Hard Materials **68**, 29–40 (2017).
[162]G. Rieck, 'Growth and preferred orientations of crystals in tungsten wires', Acta Metallurgica **6**, 360–366 (1958).
[163]S. Yerra, B. Verlinden and P. van Houtte, 'On crystallographic texture of as-drawn doped-W wires', Materials Science Forum **495-497**, 913–918 (2005).
[164]W. Hosford, 'Microstructural changes during deformation of [011] fiber-textured metals', Transactions of the Metallurgical Society of AIME **230**, 12–15 (1964).
[165]A. J. Melmed, 'The art and science and other aspects of making sharp tips', Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures **9**, 601–608 (1991).
[166]G. S. Kelsey, 'The anodic oxidation of tungsten in aqueous base', Journal of The Electrochemical Society **124**, 814–819 (1977).
[167]J. P. Ibe, P. P. Bey, S. L. Brandow, R. A. Brizzolara, N. A. Burnham, D. P. DiLella, K. P. Lee, C. R. K. Marrian and R. J. Colton, 'On the electrochemical etching of tips for scanning tunneling microscopy', Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films **8**, 3570–3575 (1990).

[168]C. Kisielowski, B. Freitag, M. Bischoff, H. Van Lin, S. Lazar, G. Knippels, P. Tiemeijer, M. van der Stam, S. von Harrach, M. Stekelenburg, M. Haider, S. Uhlemann, H. Müller, P. Hartel, B. Kabius, D. Miller, I. Petrov, E. Olson, T. Donchev, E. Kenik, A. Lupini, J. Bentley, S. Pennycook, I. Anderson, A. Minor, A. Schmid, T. Duden, V. Radmilovic, Q. Ramasse, M. Watanabe, R. Erni, E. Stach, P. Denes and U. Dahmen, 'Detection of single atoms and buried defects in three dimensions by aberration-corrected electron microscope with 0.5-Å information limit', Microscopy and Microanalysis **14**, 469–477 (2008).

[169]U. Dahmen, R. Erni, V. Radmilovic, C. Ksielowski, M.-D. Rossell and P. Denes, 'Background, status and future of the transmission electron aberration-corrected microscope project', Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences **367**, 3795–3808 (2009).

[170]A. K. Schmid and N. Andresen, 'Motorized manipulator for positioning a TEM specimen', U.S. pat. 7,851,769 (The Regents of the University of California (Oakland, CA) LBNL, 14th Dec. 2010).

[171]J. V. D. Water, J. V. D. Oetelaar, R. Wagner, H. N. Slingerland, J. W. Bruggers, A. H. D. Ottevanger, A. Schmid, E. A. Olson, I. G. P. T. I. Donchev and T. Duden, 'Manipulator for rotating and translating a sample holder', U.S. pat. 7,884,326 (FEI Company (Hillsboro, OR); The Board of Trustees of the University of Illinois (Urbana, IL); The Regents of the University of California (Oakland, CA) LBNL, 8th Feb. 2011).

[172]P. Ercius, M. Boese, T. Duden and U. Dahmen, 'Operation of TEAM I in a user environment at NCEM', Microscopy and Microanalysis **18**, 676–683 (2012).

[173]T. Duden, 'Lateral displacement and rotational displacement sensor', U.S. pat. 8,704,536 (The Regents of the University of California (Oakland, CA), 22nd Apr. 2014).

[174]H. von Harrach, 'Instrumental factors in high-resolution FEG STEM', Ultramicroscopy **58**, 1–5 (1995).

[175]A. Muller and J. Grazul, 'Optimizing the environment for sub-0.2 nm scanning transmission electron microscopy', Microscopy **50**, 219–226 (2001).

[176]N. Nakanishi, 'Retrieval process of high-resolution HAADF-STEM images', Journal of Electron Microscopy **51**, 383–390 (2002).

[177]A. Rečnik, G. Möbus and S. Šturm, 'IMAGE-WARP: a real-space restoration method for high-resolution STEM images using quantitative HRTEM analysis', Ultramicroscopy **103**, 285–301 (2005).

[178]N. Braidy, Y. L. Bouar, S. Lazar and C. Ricolleau, 'Correcting scanning instabilities from images of periodic structures', Ultramicroscopy **118**, 67–76 (2012).

[179]L. Jones and P. D. Nellist, 'Identifying and correcting scan noise and drift in the scanning transmission electron microscope', Microscopy and Microanalysis **19**, 1050–1060 (2013).

[180]B. Berkels, P. Binev, D. A. Blom, W. Dahmen, R. C. Sharpley and T. Vogt, 'Optimized imaging using non-rigid registration', Ultramicroscopy **138**, 46–56 (2014).
[181]X. Sang and J. M. LeBeau, 'Revolving scanning transmission electron microscopy: correcting sample drift distortion without prior knowledge', Ultramicroscopy **138**, 28–35 (2014).
[182]C. Ophus, J. Ciston and C. T. Nelson, 'Correcting nonlinear drift distortion of scanning probe and scanning transmission electron microscopies from image pairs with orthogonal scan directions', Ultramicroscopy **162**, 1–9 (2016).
[183]A. H. Zewail, 'Four-dimensional electron microscopy', Science **328**, 187–193 (2010).
[184]D. Li, Z. L. Wang and Z. Wang, 'Revealing electron–phonon interactions and lattice dynamics in nanocrystal films by combining in situ thermal heating and femtosecond laser excitations in 4D transmission electron microscopy', The Journal of Physical Chemistry Letters **9**, 6795–6800 (2018).
[185]M. J. Stern, L. P. R. de Cotret, M. R. Otto, R. P. Chatelain, J.-P. Boisvert, M. Sutton and B. J. Siwick, 'Mapping momentum-dependent electron-phonon coupling and nonequilibrium phonon dynamics with ultrafast electron diffuse scattering', Physical Review B **97**, 165416 (2018).
[186]A. Kutty and S. Vaidya, 'Mean-square atomic displacements in f.c.c. crystals', Journal of Physics and Chemistry of Solids **41**, 1163–1170 (1980).
[187]R. C. Shukla and G. A. Heiser, 'Mean-square atomic displacement of alkali-metal atoms: a comparison of the lattice-dynamics and molecular-dynamics results', Physical Review B **33**, 2152–2157 (1986).
[188]V. M. Gundyrev and V. I. Zel'dovich, 'Root-mean-square displacements of atoms in the B2- and R-phases of TiNi single crystal', Materials Science and Engineering: A **481-482**, 231–234 (2008).
[189]B. D. Levin, E. Padgett, C.-C. Chen, M. Scott, R. Xu, W. Theis, Y. Jiang, Y. Yang, C. Ophus, H. Zhang, D.-H. Ha, D. Wang, Y. Yu, H. D. Abruña, R. D. Robinson, P. Ercius, L. F. Kourkoutis, J. Miao, D. A. Muller and R. Hovden, 'Nanomaterial datasets to advance tomography in scanning transmission electron microscopy', Scientific Data **3**, 160041 (2016).
[190]G. Hasson and C. Goux, 'Interfacial energies of tilt boundaries in aluminium. experimental and theoretical determination', Scripta Metallurgica **5**, 889–894 (1971).
[191]H. Ichinose and Y. Ishida, 'Observation of [110] tilt boundary structures in gold by high resolution HVEM', Philosophical Magazine A **43**, 1253–1264 (1981).
[192]U. von Luxburg, 'A tutorial on spectral clustering', Statistics and Computing **17**, 395–416 (2007).
[193]S. Zhang, X. Zhang, G. Jiang, H. Zhu, S. Guo, D. Su, G. Lu and S. Sun, 'Tuning nanoparticle structure and surface strain for catalysis optimization', Journal of the American Chemical Society **136**, 7734–7739 (2014).
[194]C. Friend and F. Xu, 'Perspectives on the design of nanoparticle systems for catalysis', Faraday Discussions **208**, 595–607 (2018).

[195]P. K. Jain, K. S. Lee, I. H. El-Sayed and M. A. El-Sayed, 'Calculated absorption and scattering properties of gold nanoparticles of different size, shape, and composition: applications in biological imaging and biomedicine', The Journal of Physical Chemistry B **110**, 7238–7248 (2006).
[196]M. De, P. S. Ghosh and V. M. Rotello, 'Applications of nanoparticles in biology', Advanced Materials **20**, 4225–4241 (2008).
[197]H. Daraee, A. Eatemadi, E. Abbasi, S. F. Aval, M. Kouhi and A. Akbarzadeh, 'Application of gold nanoparticles in biomedical and drug delivery', Artificial Cells, Nanomedicine, and Biotechnology **44**, 410–422 (2014).
[198]Soft dB, *Soft dB digital signal processing board signal ranger MK2*, (2016) https://www.softdb.com/dsp_boards/sr-mk2/ (visited on 15/03/2019).
[199]The GXSM Team, *Gxsm*, (2015) http://gxsm.sourceforge.net/ (visited on 15/03/2019).
[200]PiezoMotor, *LL10 - PiezoMotor*, (2019) https://piezomotor.com/produkt/ll10/ (visited on 15/03/2019).
[201]E. Golubovic, Z. Zhakypov, T. Uzunovic and A. Sabanovic, 'Piezoelectric motor driver: design and evaluation', in IECON 2013 - 39th annual conference of the IEEE industrial electronics society (Nov. 2013), pp. 3964–3969.
[202]PiezoMotor, *PMD301 Controller - PiezoMotor*, (2018) https://piezomotor.com/produkt/pmd301/ (visited on 15/03/2019).
[203]R. D. Grober, J. Acimovic, J. Schuck, D. Hessman, P. J. Kindlemann, J. Hespanha, A. S. Morse, K. Karrai, I. Tiemann and S. Manus, 'Fundamental limits to force detection using quartz tuning forks', Review of Scientific Instruments **71**, 2776–2780 (2000).
[204]A. Crottini, J. Staehli, B. Deveaud, X. Wang and M. Ogura, 'Ultra stable tuning fork sensor for low-temperature near-field spectroscopy', Ultramicroscopy **90**, 97–101 (2002).
[205] RS Components Ltd., *NC38LF-327 - Fox Electronics*, (2019) https://uk.rs-online.com/web/p/crystal-units/5476985/ (visited on 15/03/2019).
[206]H.-P. Rust, M. Heyde and H.-J. Freund, 'Signal electronics for an atomic force microscope equipped with a double quartz tuning fork sensor', Review of Scientific Instruments **77**, 043710 (2006).
[207]R. Pawlak, S. Kawai, S. Fremy, T. Glatzel and E. Meyer, 'High-resolution imaging of C60 molecules using tuning-fork-based non-contact atomic force microscopy', Journal of Physics: Condensed Matter **24**, 084005 (2012).
[208]K. W. Clark, S. Qin, X.-G. Zhang and A.-P. Li, 'Nanoscale periodic modulations on sodium chloride surface revealed by tuning fork atomic force microscopy', Nanotechnology **23**, 185306 (2012).
[209]H. Ooe, T. Sakuishi, M. Nogami, M. Tomitori and T. Arai, 'Resonance frequency-retuned quartz tuning fork as a force sensor for noncontact atomic force microscopy', Applied Physics Letters **105**, 043107 (2014).
[210]F. J. Giessibl, 'High-speed force sensor for force microscopy and profilometry utilizing a quartz tuning fork', Applied Physics Letters **73**, 3956–3958 (1998).

[211]F. J. Giessibl, 'Atomic resolution on Si(111)-(7×7) by noncontact atomic force microscopy with a force sensor based on a quartz tuning fork', Applied Physics Letters **76**, 1470–1472 (2000).
[212]F. J. Giessibl and M. Reichling, 'Investigating atomic details of the CaF2(111) surface with a qPlus sensor', Nanotechnology **16**, S118–S124 (2005).
[213]A. Bettac, J. Koeble, K. Winkler, B. Uder, M. Maier and A. Feltz, 'QPlus: atomic force microscopy on single-crystal insulators with small oscillation amplitudes at 5 K', Nanotechnology **20**, 264009 (2009).
[214]F. J. Giessibl, 'Principles and applications of the qPlus sensor', in *Noncontact atomic force microscopy* (Springer Berlin Heidelberg, 2009), pp. 121–142.
[215]F. J. Giessibl, F. Pielmeier, T. Eguchi, T. An and Y. Hasegawa, 'Comparison of force sensors for atomic force microscopy based on quartz tuning forks and length-extensional resonators', Physical Review B **84**, 125409 (2011).
[216]O. E. Dagdeviren and U. D. Schwarz, 'Optimizing qPlus sensor assemblies for simultaneous scanning tunneling and noncontact atomic force microscopy operation based on finite element method analysis', Beilstein Journal of Nanotechnology **8**, 657–666 (2017).
[217]F. J. Giessibl, 'The qPlus sensor, a powerful core for the atomic force microscope', Review of Scientific Instruments **90**, 011101 (2019).