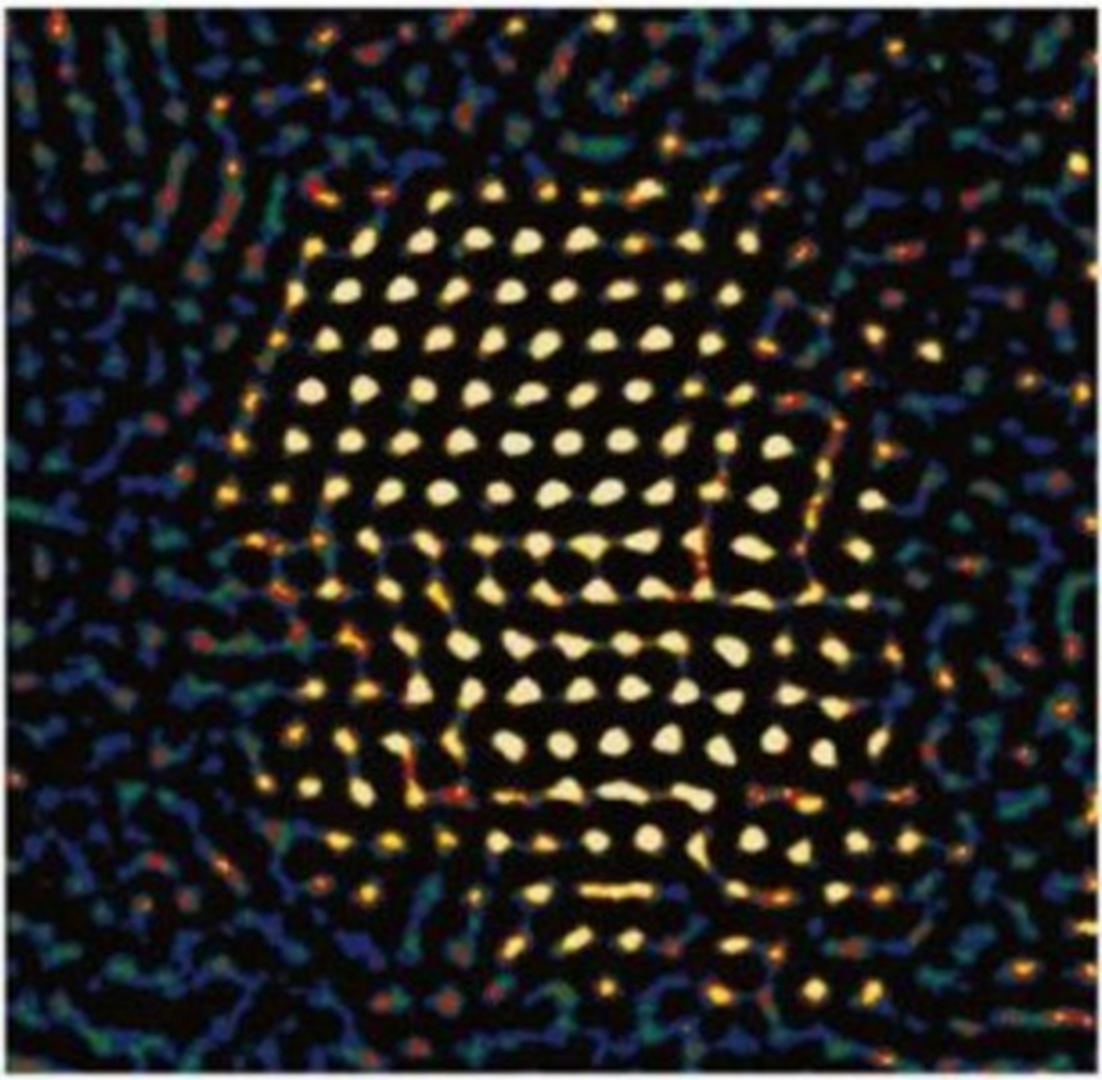
RSC Nanoscience & Nanotechnology

Edited by Angus I Kirkland and John L Hutchison

Nanocharacterisation



RSC Publishing

Nanocharacterisation

Nanocharacterisation

Edited by

Augus I Kirkland and John L Hutchison
Department of Materials, Oxford University, Oxford, UK

RSC Publishing

ISBN: 978-0-85404-241-8

A catalogue record for this book is available from the British Library

Published by The Royal Society of Chemistry,
Thomas Graham House, Science Park, Milton Road,
Cambridge CB4 0WF, UK

Registered Charity Number 207890

For further information see our web site at www.rsc.org

Preface

In presenting these chapters with the common theme "Nanocharacterisation", it is instructive to note the explosive development of nanotechnology over the past few years. Although the term itself is relatively recent, its origins can be traced back to the late 19th century, with the development of colloidal science. The medieval craftsmen who produced the stained glass that adorns many of Europe's great cathedrals unwittingly used nanotechnology: their addition of gold chloride to molten glass produced tiny, uniformly sized gold spheres ($\sim$50 nm in diameter) that impart the deep ruby-red colouration that is a characteristic feature of many famous windows; similarly, silver nanocrystals produced an intense yellow colouration.

The use of nanoscale particles with high surface areas has also been exploited for many years in industrial catalysis, and again the preparation of these materials has used "nanotechnology" in various forms.

The anticipation of many of the potential benefits of nanotechnology may be traced back to a lecture "*There's Plenty of Room at the Bottom*" delivered by Richard Feynman to a meeting of the American Physical Society in 1959. Describing a process by which the ability to manipulate individual atoms and molecules might be developed, Feynman conjectured:

"*What could we do with layered structures with just the right layers? What would the properties of materials be if we could really arrange the atoms the way we want them? They would be very interesting to investigate theoretically. I can't see exactly what would happen, but I can hardly doubt that when we have some control of the arrangement of things on a small scale we will get an enormously greater range of possible properties that substances can have, and of different things that we can do*".

In the course of his lecture, Feynman also noted that scaling issues would arise from changes in the magnitude of various physical phenomena and suggested that electron microscopy would play a key role in the characterisation of these materials.

The term "nanotechnology" was coined and defined in the mid-1970s in a paper presented by Norio Taniguchi (N. Taniguchi, "On the Basic Concept of 'Nano-Technology'," *Proc. Intl. Conf. Prod. Eng. Tokyo*, Part II, 1974, Japan Society of Precision Engineering) in which it was suggested that it "*consists of*

the processing of, separation, consolidation, and deformation of materials by one atom or one molecule.''

Subsequently, this basic definition was explored in greater depth by Eric Drexler, who vigourously promoted the technological significance of nanoscale phenomena and devices. [''*Engines of Creation: The Coming Era of Nanotechnology and Nanosystems: Molecular Machinery, Manufacturing, and Computation*'', (ISBN 0-471-57518-6).]

At the time of writing ''nanotechnology'' embraces a large (and growing number) of techniques that are used to create and characterise structures on a size scale below $\sim$100 nm. Included in these are methods used for the fabrication of nanowires, lithographic techniques routinely used in semiconductor fabrication, focused ion beam (FIB) machining and related techniques used for the production of micro-electro-mechanical systems (MEMS) and molecular self-assembly.

In the early 1980s two significant developments had a major impact on characterisation aspects of nanotechnology: the invention of the scanning tunneling microscope and the synthesis of small carbon clusters. The development of this instrument in 1981, for which Gerd Binning and Heinrich Rohrer were awarded the Nobel Prize in 1986, followed on exactly 50 years after the development of the first transmission electron microscope in 1931 by Max Knoll and Ernst Ruska, the latter sharing (somewhat belatedly) the same Nobel Prize. This has in turn spawned a plethora of other scanned probe microscopy techniques including atomic force microscopy, magnetic force microscopy, electrostatic force microscopy and others.

The discovery in 1985 of single, closed-shell carbon nanostructures (e.g. C_{60}) known collectively as fullerenes, by Harold Kroto and his colleagues, for which they also were awarded the Nobel Prize in 1996, and the recognition of carbon nanotubes, provided much of the impetus for what would become an unprecedented and explosive worldwide development and interest in nanostructures.

These developments have driven a need for ever-more powerful techniques for structural and chemical characterisation on the ''nanoscale'' and ''nanocharacterisation'' – the theme of this book, now encompasses a diverse range of techniques that are the essential tools underpinning much of nanotechnology.

In attempting to give an overview of the different techniques now in regular use in nanocharacterisation we have been necessarily selective and we note that other microscopies are still developing that will certainly find a place in the armoury with which nanotechnologists will pursue their quest for ever-smaller and tighter-controlled structures.

In the chapter devoted to high-resolution (transmission) electron microscopy David Smith illustrates how this technique has evolved into a widely used tool in modern nanostructure research, highlighting the problems relating to interpretation of image contrast, as well as providing numerous examples showing how HRTEM has been used to provide unique insights into local microstructure of nanocrystalline materials.

The development of the scanning transmission electron microscope in the 1970s by Albert Crewe and colleagues produced instruments that could use the

signals generated by scanning an electron beam across a thin specimen in a controlled manner. The chapter by Andrew Lupini, Stephen Pennycook and their coworkers outlines this technique and demonstrates how the recent application of aberration-correcting devices in STEM instrumentation is now capable of producing sub-Angstrom imaging and spectroscopy of nanostructures.

Complementing the above are the various scanned probe microscopy techniques that have emerged following the initial development of the scanning tunneling microscope. These instruments rely on a variety of surface/probe interactions and delicate rastering mechanisms to obtain images with extreme depth sensitivity to various parameters thus revealing many physical and electronic properties with atomic resolution. The additional ability to interact with *individual* atoms on surfaces in a precisely controlled manner also opens up an exciting range of opportunities for nanoscale manipulation and fabrication. Martin Castell in his chapter provides an overview of this field, and explores in detail the use of these techniques in nanoscience.

In addition to structural information, local chemical composition, bonding and electronic states are key aspects of nanostructures. A range of spectroscopic techniques is now available for probing local structures on the atomic scale. Among these, electron energy-loss spectroscopy and energy dispersive X-ray analysis have emerged as the most powerful and widely used. These techniques probe, respectively, the energy loss of electrons passing through a thin specimen and the X-rays generated by the incident electron beam. The chapter by Rik Brydson provides an overview of the physical principles involved in these and demonstrates how they are now being used to probe local chemistry and bonding at the atomic level.

In many applications of nanotechnology the measurement of the magnetic and electrical properties of individual nanoscale objects is becoming increasingly important. Relatively recent instrumental developments have enabled the technique of electron holography, first proposed by Gabor in 1947, to be used, to obtain measurement of these. Rafal Dunin-Borkowski and coworkers describe the theory and practice of this technique and its application to a range of important nanostructured magnetic and electronic materials.

As nanostructures become smaller, their 3-dimensional shape assumes much greater importance. This is particularly the case for supported catalysts, where active sites may be determined by such entities. Although electron tomography had been previously applied successfully to biological structures it has only recently been extended to for applications in nanoscience. Paul Midgley and Mathew Weyland survey the instrumental, theoretical and computational requirements of this important technique and illustrate how it can be used to gain 3-dimensional information from a wide range of technologically important materials.

The ability to carry out or induce chemical reactions on a very small scale is also crucial to many aspects of nanotechnology and the ability to observe these *in situ* is a particularly powerful combination. This can be achieved with specially modified instrumentation that allows a controlled, gaseous environment around the specimen. In the concluding chapter, Pratibha Gai gives an

account of the design and use of an environmental transmission electron microscopy facility including a description of the physical principles involved in its design together with numerous case histories describing some of the insights gained into some important chemical processes.

In summary, our aim in preparing this book has been to gather together a selection of articles written by internationally recognised experts that describe some of the characterisation techniques that are currently used in studies of nanostructured materials.

We hope that the handbook will provide a valuable resource to all involved in the characterisation of nanomaterials.

Contents

CHAPTER 1

Characterisation of Nanomaterials Using Transmission Electron Microscopy

D. J. SMITH

Department of Physics, Arizona State University, Tempe, AZ 85287 USA

1.1 Introduction

The Transmission Electron Microscope (TEM) has evolved over many years into a highly sophisticated instrument that has found widespread application across the scientific disciplines. Because the TEM has an unparalleled ability to provide structural *and* chemical information over a range of length scales down to the level of atomic dimensions, it has developed into an indispensable tool for scientists who are interested in understanding the properties of nanostructured materials and in manipulating their behaviour.

The resolution of the optical microscope is restricted by the wavelength of visible light, which thus precludes atomic-scale imaging. In contrast, an energetic electron has a wavelength of much less than 1 Å (where $1\ \text{Å} = 10^{-10}$ m), so that an enormous improvement in resolution can be achieved, at least in principle, by using a beam of fast electrons for imaging. A suitable combination of (magnetic) electron lenses is required, both for focusing the electron beam onto the object and also for providing an enlarged image. Maximum magnifications at the microscope are typically close to or exceed one million times, so that details of the nanoscale object are clearly visible on the final viewing screen or recording medium.

Image formation in the TEM is more complicated in practice than is the case for the optical microscope. Strong magnetic fields are needed for focusing the electron beam, and these cause electrons to take a spiral trajectory through the lens field. In addition, a major restriction on ultimate microscope performance

results from unavoidable aberrations of round electron lenses. Primarily, due to the need for a compromise between small-angle diffraction effects and wide-angle spherical-aberration limits, the resolution, d can be roughly expressed by an equation of the form

$$d = A\, C_s^{1/4}\, \lambda^{3/4} \tag{1.1}$$

where C_S is the spherical aberration coefficient of the objective lens, λ is the electron wavelength, and A is a constant with a value ranging from 0.43 to 0.7 depending on the type of imaging (coherent, incoherent, or phase contrast). Values of d typically range from about 3.0 Å down to 1.0 Å as electron energies are increased from 100 to 1250 keV. Modern-day TEMs operating at 200 or 300 keV have resolution limits well below 2.0 Å, which is comparable to the spacing between atoms. Individual columns of atoms can thus be resolved in crystalline materials, which must first, however, be oriented so that the incident electron beam is aligned along some major crystallographic zone axis of the sample.

The power of the technique is illustrated by the example in Figure 1.1, which shows the boundary region between two Al crystals, both of which are oriented so that the electron beam is parallel with the [001]-type zone axis. Each black spot in the image marks the position of a column of Al metal atoms viewed in an end-on geometry. It is obviously straightforward to visualise the periodic array of misfit dislocations (arrowed) that accommodate the angular misfit of 6° between the two crystals, and further analysis would enable the detailed atomic structure around the dislocation core to be determined.

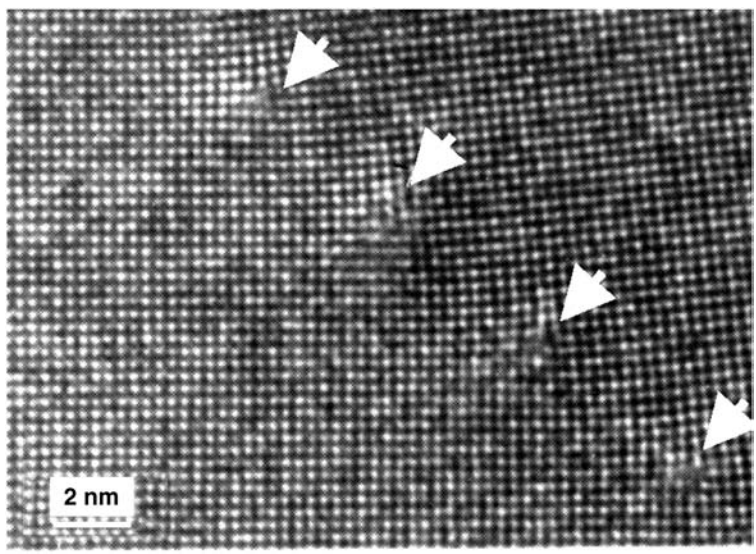


Figure 1.1 Atomic-resolution electron micrograph of Al 6° [001] symmetric tilt grain boundary with misfit accommodation by [110]/2 edge dislocations (arrowed). Each black spot corresponds to projection of an individual Al atomic column.

This chapter begins by providing a brief introduction to the TEM and some of the key aspects of high-resolution imaging. Applications to nanostructured materials are then described in greater detail, and some emerging trends and unresolved issues are briefly discussed. For further information about microscope operation and more details about applications to a broader range of materials, the interested reader is referred to the review articles and monographs listed at the end of the chapter.[1–9]

1.2 Imaging

1.2.1 Transmission Electron Microscopy

In the standard TEM operating mode, which is commonly referred to as amplitude or diffraction contrast imaging, only a fraction of those electrons that have passed through the sample are used to form the highly magnified final image. Most of the scattered (or diffracted) electrons are prevented from reaching the image plane by positioning a small objective aperture located in the back focal plane of the objective lens. This aperture thus serves to determine the image contrast. For the case of crystalline samples, the electron diffraction pattern (EDP) is used to ensure that the orientation of the specimen relative to the direction of the incident electron beam will satisfy a strongly diffracting condition. Many common structural defects have a highly characteristic appearance under such diffraction contrast conditions. The spacings and angles between crystal lattice planes can also be determined if the EDP is first calibrated using a known material. In addition, the availability of a crystalline substrate or support can provide a convenient method for sample orientation during observation. By using the substrate EDP for reference purposes, internal interfaces can be aligned perpendicular to the electron-beam direction so that any changes in the microstructure of thin films and multilayers can then be determined as a function of film thickness. As an example, Figure 1.2 shows a multilayered Magnetic Tunneling Transistor (MTT) deposited directly on the native oxide of a Si substrate.[10] The individual layers of the MTT can be clearly recognised, and their thickness uniformity is easily confirmed. Finally, it should be appreciated by the reader that examination of such complex samples with the TEM can represent a serious challenge to the electron microscopist. Because of considerable differences in thinning rates, it will often be difficult to prepare samples that are electron transparent across the entire region of interest simultaneously. Descriptions of different approaches for preparing electron-transparent specimens can be found elsewhere.[11,12]

1.2.2 High-Resolution Electron Microscopy

In the technique of High-Resolution Electron Microscopy (HREM), a much larger objective aperture (or sometimes none at all) is used. The directly transmitted beam can then interfere with one or more diffracted beams, and

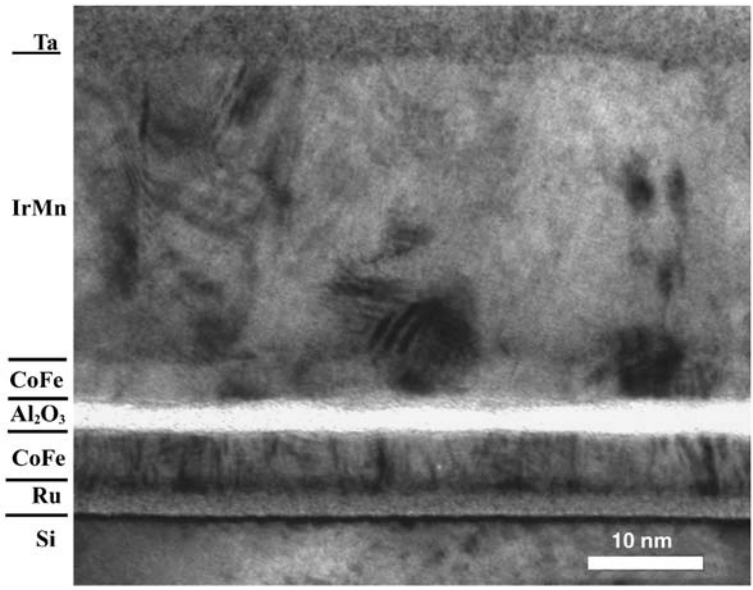


Figure 1.2 Bright-field electron micrograph showing the cross section of a typical MTT device structure with a Ru seed layer between the Si(001) collector substrate and the CoFe base layer. Reproduced from Ref. 10.

the contrast across the image will depend on the relative phases of the various beams. This imaging mode is thus often referred to as phase contrast imaging. When the microscope imaging conditions are properly adjusted (lens defocus, image astigmatism, incident beam alignment) it is possible to interpret phase-contrast images in terms of the projected crystal potential provided that the specimen thickness is not too great (less than 10 nm preferred). Indeed, individual atomic columns can be separately resolved in many crystalline inorganic materials using the latest generations of HREM instruments.[1] High electron doses, typically $\sim$500–2000 electrons per square Å, are required to record such images, which means that specimens intended for high-resolution studies must be relatively resistant to electron-irradiation effects. It is impossible to examine most organic materials and polymers directly under such intense imaging conditions. By using a specimen-heating holder, and by adding a TV rate image pickup system to the base of the electron microscope lens column, dynamic events can be followed in real time without significant loss of spatial resolution.[13,14]

Over the past 40 years, HREM has been used to characterise a wide range of inorganic materials. Important applications include determining the microstructure of crystalline defects, interfaces and grain boundaries, investigating nanocrystalline features in amorphous films, and studying small particles in heterogeneous catalysts. The characterisation of magnetic thin films and multilayers, for example, continues to be very important, since layer continuity and defect microstructure are crucial to the viability of recording media. High

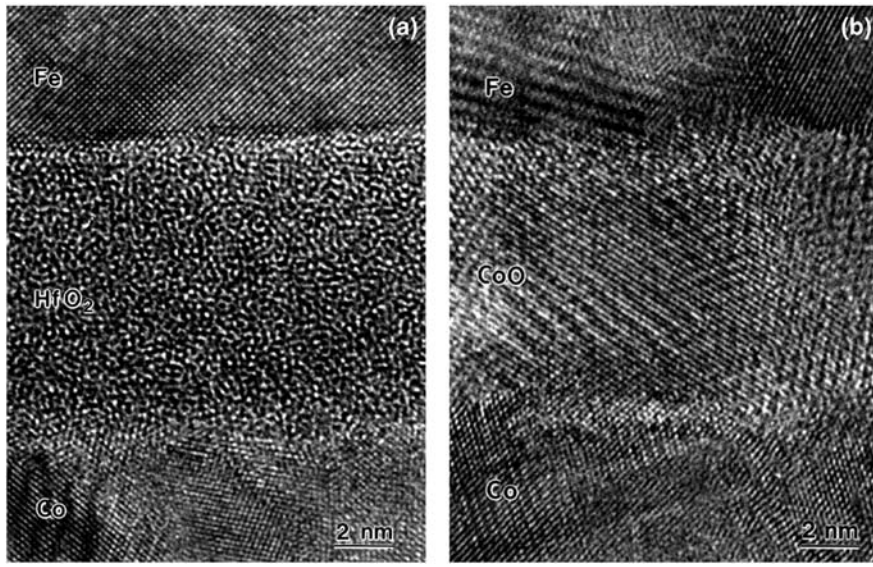


Figure 1.3 Cross-sectional high-resolution electron micrographs of magnetic tunnel junctions showing: (a) amorphous, and (b) polycrystalline, barrier layers. Reproduced from Ref. 15.

resolution images are able to provide specific details that are usually unavailable using other techniques. As an illustration, Figure 1.3(a) and (b) compare two high-resolution electron micrographs that reveal the amorphous or polycrystalline nature of the barrier layers in simple magnetic tunnel junctions grown by dc reactive sputtering.[15] The layer sequences in the images are: (a) Co (50 nm)/HfO_2 (10 nm)/Fe (50 nm), and (b) Co (50 nm)/CoO (10 nm)/Fe (50 nm). Further high resolution images of nanomaterials are presented in later sections.

1.2.3 Basis of High-Resolution Imaging

Image formation in the electron microscope occurs in two stages. Electrons of the incident beam interact with the specimen, undergoing both elastic and inelastic scattering. The electron wavefunction emerging from the exit surface of the specimen passes through the objective lens and additional magnifying lenses are used to form the final image. Electrons that are elastically scattered mainly contribute to the high-resolution bright-field image. Note that the inelastically scattered electrons can provide valuable information about sample composition via the technique of Electron Energy-Loss Spectroscopy (EELS), while electrons scattered to very large angles can be used for Z-contrast Annular Dark-Field (ADF) imaging in the Scanning Transmission Electron Microscope (STEM). These possibilities are described in other chapters.

Unlike X-ray or neutron scattering, electron scattering is strongly dynamical, meaning that the kinematical scattering approximation will be inadequate for

understanding image formation except for the very thinnest of samples. Multiple electron scattering with large phase changes is far more typical, so that knowledge about the relative heights and locations of different atoms in the specimen becomes important for quantitative interpretation of image features. Indeed, image simulations are considered as essential for extracting detailed information about atomic arrangements at dislocations and interfaces. Several approaches to image simulation have been developed over the years,[16,17] with the most widespread, commonly known as the multislice method, being based on an n-beam dynamical theory of electron scattering.[18,19] In this approach, atoms in the specimen are considered as being located on narrowly separated planes (or slices), normal to the beam direction. The electron wavefunction is then propagated slice-by-slice through the sample to eventually form the exit-surface wavefunction. This iterative process lends itself to convenient computer algorithms that enable rapid computations to be carried out, and these simulations are especially useful during the refinement of unknown defect structures.[20,21] Further information about different but equivalent theories of electron scattering can be found in the monograph by Cowley.[22]

The electron wavefunction at the exit surface of the specimen must still be transferred to the final viewing screen or recording medium. This process is determined primarily by the properties of the objective lens. The effect of this lens can be conveniently understood by reference to what is termed the Phase-Contrast Transfer Function (PCTF), as described by Hanszen.[23] The basic form of the PCTF is independent of both specimen and microscope so that a single set of universal curves can be used to describe the transfer characteristics of all objective lenses. Electron microscopes with different objective lenses, or operating at different electron energies, are then easily compared by using suitable scaling factors. Figure 1.4 shows PCTFs for the optimum defocus of the objective lens of a typical 400 kV HREM. The two curves correspond to: (a) coherent, and (b) partially coherent, incident electron illumination.

It is important to appreciate that the PCTF has an oscillatory nature, as visible in Figure 1.4, which means that electrons scattered to different angles undergo reversals in phase. These phase oscillations will thus cause artefactual detail in the final image that may be misinterpreted. The PCTF is also focus dependent, meaning that further phase changes occur when the focus is changed, and these will also affect the appearance of the image. Thus, much of the detail visible in the recorded micrograph could be uninterpretable unless the lens defocus is accurately known. Also note that the incident electron beam is ideally a coherent, monochromatic plane wave, whereas in practice some loss of coherence results from focal spread (temporal coherence) and finite beam divergence (spatial coherence).[24,25] These effects of partial coherence are conveniently represented by envelope functions that cause dampening of the PCTF at larger scattering angles. Specimen information scattered to higher spatial frequencies, equivalent to higher image resolution, is therefore lost. These incoherent effects are illustrated by the curve labeled (b) in Figure 1.4, where it should also be noted that the positions of the PCTF zeroes are not affected by the envelopes. Finally, note that additional specimen information may

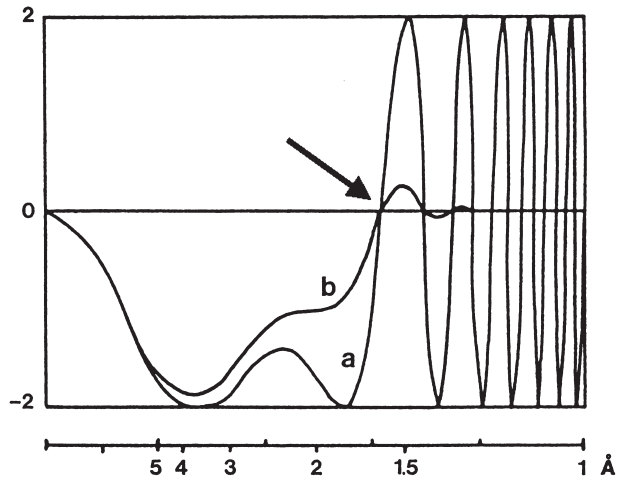


Figure 1.4 Phase Contrast Transfer functions for high-resolution imaging at optimum defocus for objective lens ($C_S = 1.0$ mm) of 400 kV HREM: (a) coherent illumination; and (b) partially coherent illumination, with focal spread of 8 nm and incident beam convergence half-angle of 0.5 mrad. Arrow indicates the interpretable resolution limit.

become available through the use of the highly coherent Field-Emission electron Gun (FEG).

1.2.4 Resolution Limits

The resolution of any imaging system is closely coupled to the illumination wavelength. Thus, resolution limits on the picometre scale might reasonably be expected for high-energy electrons. As mentioned earlier, the compromise between diffraction and spherical aberration gives an approximate estimate of the image resolution. In practice, high-resolution imaging is considerably more complicated, and there are several alternative definitions that are applicable depending on the sample and the microscope operating conditions.[26] These resolution limits are most easily understood by considering the PCTF of the objective lens.

The *interpretable* image resolution, which is sometimes referred to as the *structural* or *point* resolution, is defined only at the optimum or Scherzer defocus, where the PCTF has the largest possible band of spatial frequencies without any phase reversal. The corresponding first zero crossover, as indicated by the arrow in Figure 1.4, gives the interpretable resolution. C_S values increase slightly at higher electron energies. However, because of the reduction in λ, improvements in theoretical resolution limits are obtained. Typical interpretable resolutions are in the range of 2.5 Å down to 1.2 Å for corresponding accelerating voltages of 200 kV up to 1000 kV. The size and cost of higher-voltage electron microscopes, as well as the increasing likelihood of electron

irradiation damage for higher-energy electrons, are further practical factors that need to be taken into account. Intermediate-voltage HREMs, operating in the 200 to 400 kV range, have become widespread because of these considerations.

The envelope functions define the *instrumental* resolution or *information limit* of the HREM. A value of roughly 15% [*i.e.* exp(–2)] is usually taken as the resolution cutoff since this level is commonly regarded as the minimum acceptable for image-processing requirements.[27] This resolution limit can extend well beyond the interpretable resolution for 200 or 300 kV HREMs equipped with an FEG electron source, as illustrated in Figure 1.5. Very fine detail is thus often present but it is not easily related to specimen features because of the PCTF oscillations mentioned earlier. An objective aperture of suitable diameter can be used to prevent beams with inverted phase from contributing to the image. Alternatively, the phase modulations caused by the PCTF can be removed by *a posteriori* image processing when the defocus and C_S values are known well enough. Improved image resolution can then be achieved, as demonstrated by the pioneering studies of Coene *et al.*[28] who were able to resolve columns of oxygen atoms in a high-temperature superconductor for the first time using an approach based on focal-series reconstruction.

The expression *lattice-fringe* resolution refers to the very finest spacings that can be obtained as a result of interference between two or more diffracted beams. Here, the instrumental stability of the HREM is critical, as well as freedom of the microscope environment from adverse external factors such as acoustic noise, mechanical vibrations, and stray magnetic fields. The lattice-fringe resolution was formerly regarded as an important figure of merit for

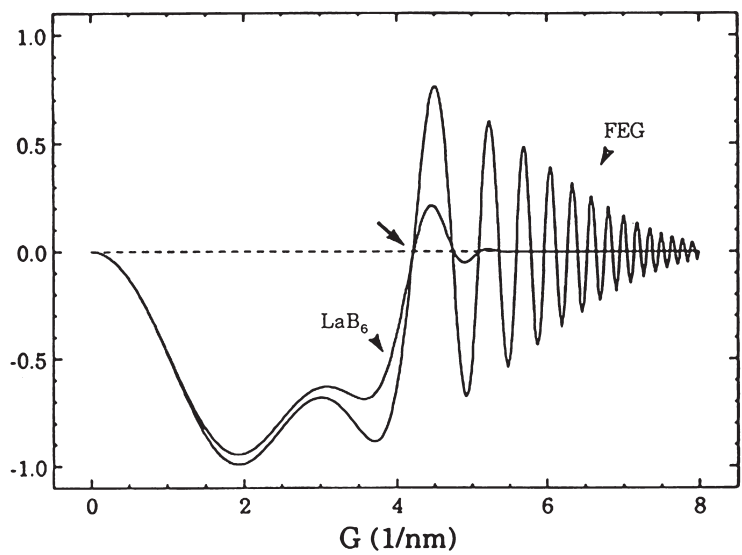


Figure 1.5 Transfer functions for 200 kV HREM comparing effect of the temporal coherence (energy spread) envelope on information transfer beyond the interpretable resolution limit.

comparing microscope performance but the interpretable and instrumental resolution limits are nowadays considered more relevant. Note that very fine lattice fringes do not usually contain any useful local information about atomic structure. The interfering diffracted beams can originate from comparatively large specimen areas, and the lattice-fringe images may be recorded at significant underfocus conditions when there are many PCTF oscillations.

1.2.5 Lattice Imaging or Atomic Imaging

Lattice-fringe images are relatively easy to obtain for many materials when imaged along major low-index zone axes with modern-day HREMs. However, comparatively very few lattice images are directly interpretable in terms of atomic arrangements because of the phase reversals that are caused by PCTF oscillations when the defocus is changed. These can also make it difficult to recognise and choose the optimum image defocus. The unique appearance of a crystal defect or Fresnel fringe along the edge of the sample is then needed, and knowledge of the focal step size(s) becomes essential. For materials with small unit cells, when only a few diffracted beams contribute to the final image, the situation becomes more complicated. In such cases, identical images, usually referred to as *Fourier* or *self-images*, recur periodically with changes in defocus, with a period given by $2d^2/\lambda$ where d is the corresponding lattice spacing.[29,30] When the defocus is changed by a half-period, the contrast of the image features will be reversed so that black spots become white and *vice versa*.

Dynamical multiple scattering causes further complications in thicker crystals. Intensity is progressively lost from the directly transmitted beam as the sample thickness is increased. A thickness is eventually reached, in the vicinity of the thickness extinction contour, where the direct beam is relatively low in intensity, and the resulting image is dominated by (second-order) interference processes between the different diffracted beams. Pairs of white spots, graphically referred to as *dumbbells*, are often visible for many small-unit-cell materials at these thicknesses.[31] The separation between these spots is often different from the projected atomic separations for the material being imaged, except for highly specific thickness and defocus values.[32] Thus, these effects from thicker specimen regions cannot usually be interpreted in terms of atomic positions.

1.2.6 Instrumental Parameters

The HREM is a highly sophisticated instrument, with an array of controls, and many adjustable parameters, of which only a small number need to be known accurately. Details about the experimental determination of these instrumental parameters can be found elsewhere.[26] The accelerating voltage must be highly stable, preferably to within one part per million (ppm), but the exact value is relatively unimportant. The current used to excite the objective lens windings will determine its focal length and other key imaging parameters, including the

spherical aberration coefficient. It is thus desirable to operate with a fixed current. Indeed, due to the extreme imaging sensitivity of the incident-beam tilt alignment and the objective lens astigmatism, it is strongly recommended that the lens current should be monitored continuously. The sample height should then be adjusted whenever the sample is tilted or another field of view is selected. The sensitivity of the image appearance to defocus means that knowledge of this parameter is essential for correct image interpretation. The image appearance of complicated structures may be highly characteristic and easily recognisable at the optimum defocus in some special cases. However, an alternative means of selecting defocus is often required when imaging unknown aperiodic features such as a dislocation or grain boundary and it can be helpful to simulate a through-thickness-through-focus set of images of the perfect crystal structure before commencing microscopy.[33] Alternatively, determination of both defocus and local specimen thickness on a local scale can be achieved using either cross correlation,[34] or nonlinear least squares[35] methods for matching experimental and simulated images.

1.3 Survey of Applications

In this section, the evolution of HREM performance is first summarised, and then applications to different classes of nanomaterials are briefly described. Representative examples from the work of the author and his colleagues are chosen as illustrations but the reader should appreciate that the scientific literature is literally bursting with other examples, far too many to summarise within the confines of a short chapter.

1.3.1 Developments in HREM

The performance of the HREM has evolved over many years, mostly as a result of improved mechanical, thermal and electrical stabilities. The resolution of the electron microscope first exceeded that of the optical microscope in the mid-1930s.[36] The first "structure images" of large unit cell block oxides were recorded in the early 1970s,[37,38] and interpretable resolution limits first surpassed the 2.0 Å barrier in the early 1980s.[39] At this level of resolution, it becomes straightforward to deduce structural models directly from observation of electron micrographs recorded at the optimum defocus. Figure 1.6(a) shows a striking example of pentagonal bipyramidal columnar defects in slightly nonstoichiometric tungsten oxide.[40] Each patch of dark contrast in the electron micrograph corresponds to a column of tungsten atoms viewed end-on. The corresponding structural model, as based on the appearance of the micrograph, is shown in Figure 1.6(b).

Evolution of microscope performance has continued through the ensuing years. The highly coherent electron beam provided by the FEG enabled information limits approaching 1.0 Å to be achieved in 300 kV TEMs in the mid-1990s.[41] High-voltage HREMs were also able to achieve the same

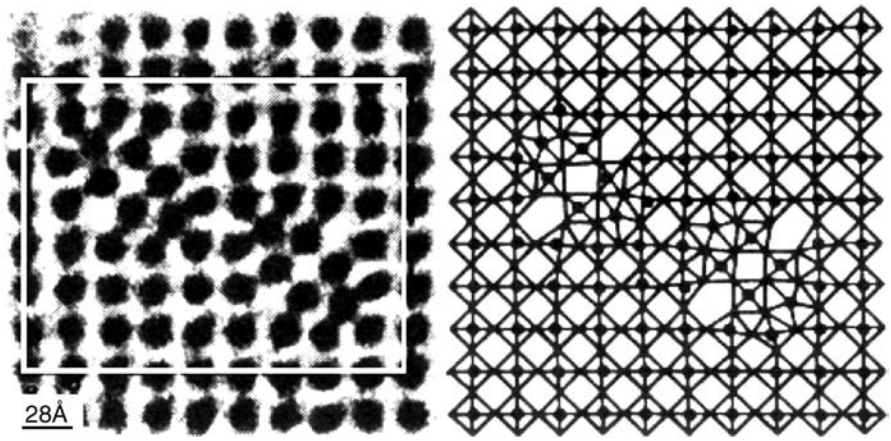


Figure 1.6 (a) Atomic-resolution electron micrograph of nonstoichiometric (W, Nb)$O_{2.93}$ showing pairs of pentagonal bipyramidal columnar defects. (b) Corresponding structural model. Occupied tunnel sites are located by direct visual inspection. Reproduced from Ref. 40.

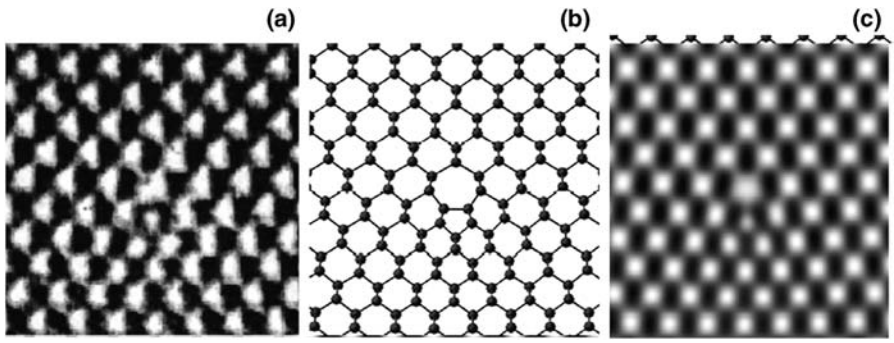


Figure 1.7 (a) Atomic-resolution electron micrograph showing core structure of symmetrical Lomer edge dislocation at Ge/Si(001) heterointerface, as recorded with 1.25 MeV atomic-resolution electron microscope; (b) corresponding structural model; (c) image simulation based on structural model. Reproduced from Ref. 43.

resolution level without any PCTF reversals,[42] making image interpretability much easier than with FEG imaging (see Figure 1.5). Figure 1.7(a) shows an example of a symmetric Lomer dislocation core observed at a Ge/Si(001) heterointerface with the 1250 kV HREM in Stuttgart.[43] The corresponding ''ball-and-stick'' structural model shown in Figure 1.7(b) identifies this dislocation core as having a seven-fold ring structure, which was unlike the large majority of asymmetrical defects that were observed in this study. The image simulation based on this model is shown in Figure 1.7(c), and the close similarity, except for photographic noise, with the experimental micrograph is readily apparent.

1.3.2 Small Particles and Precipitates

Nanometre-sized particles and precipitates are commonly observed in many different types of materials, and supported metal particles in particular play a pivotal role in the catalysis arena.[44] For these latter applications, information is needed about particle size, shape, and any surface layers or absorbates, and it is also important to know whether any changes have taken place as a result of catalytic reactions involving oxidation and reduction processes. Extraction of reliable information about the morphology and chemistry of precipitates is far from easy. Image overlap is an obvious problem during observation, since the surrounding matrix will always tend to obscure signals from the precipitates. In some special cases, however, the existence of an epitaxial relationship between the matrix and the precipitate will mean that their lattice planes are aligned, and size and shape information can sometimes become available. Detection of small particles can also be difficult, again because of the problem of obtaining sufficient contrast against background "noise" originating from the support. Eliciting information about particle size, shape and surfaces can be even more challenging, depending on the imaging geometry. In this regard, surface-profile imaging, wherein the particle edge is located adjacent to vacuum and is thus free of interference from the underlying support during imaging, is particularly useful.[45] An additional complication is that highly dispersed metal species may be mobile, even at room temperature, depending on the ambient gases. Moreover, small metal particles are often mobile under the electron-beam irradiation conditions normally used for high-resolution imaging. As an interesting example, Figure 1.8 shows a series of images for a small Au particle recorded on videotape during an observation period of 20 s.[46] Shape changes and different

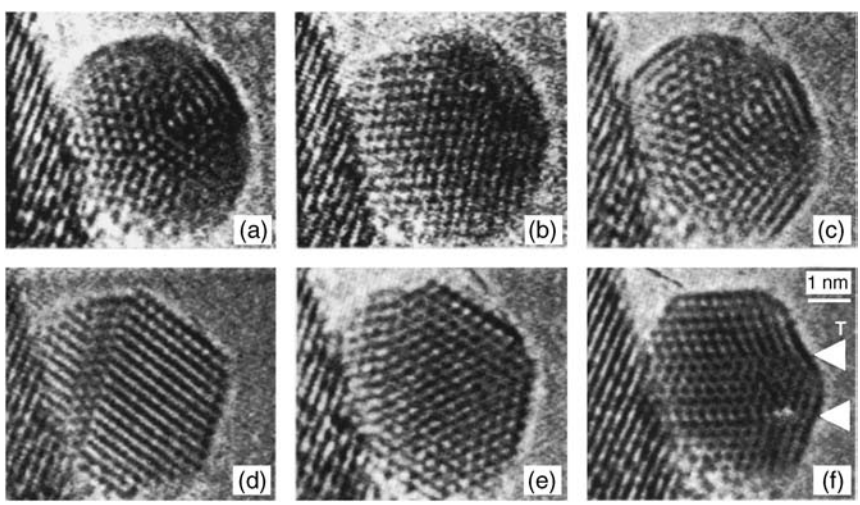


Figure 1.8 Shape changes of a nanoscale multiply twinned particle of Au observed during a 20 s time interval during observation at very high magnification in a 400 kV HREM. Reproduced from Ref. 46.

structural defects (see examples arrowed in frame f) are clearly evident. Similar structural rearrangements under electron irradiation have been documented for several different metallic nanoparticles.[1]

1.3.3 Two-Dimensional Objects

There are several different types of two-dimensional objects within the class of nanomaterials that have been extensively studied by electron microscopy. Particular mention needs to be made here of the recent discovery of nanobelts of semiconducting oxides[47] that appear to provide an ideal geometry for investigating transport phenomena in a confined system. In the field of semiconductors, the development of very well controlled deposition methods, such as molecular beam epitaxy, has led to the emergence of single and multiple quantum wells (QWs). These structures are especially useful for light-emitting and laser diode applications because of the possibility of tailoring the bandgap by control of the layer thickness(es). In the development of these QW and related superlattice (SL) structures, electron microscopy has played an invaluable role in assessing microstructure and also providing measurements of interface abruptness, especially after annealing. Superlattices of HgTe/(Hg)CdTe are of interest for possible infrared detector materials. Figure 1.9 shows a typical SL structure after annealing at 225 °C.[48] In this system, such low-temperature annealing was found to induce substantial interdiffusion, leading to changes in well/ barrier widths that impacted on the optical and electronic properties.

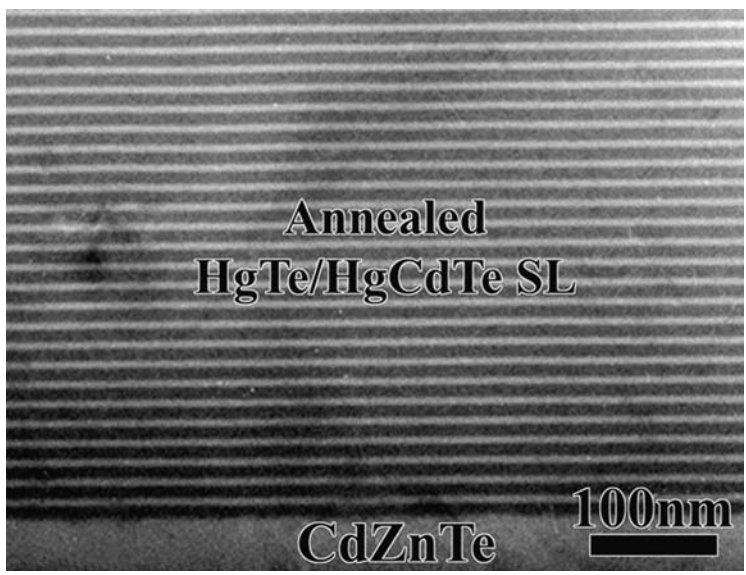


Figure 1.9 Cross-sectional bright-field electron micrograph showing the microstructure of the HgTe/HgCdTe(211)B superlattice after annealing at 225 °C. Reproduced from Ref. 48.

1.3.4 One-Dimensional Objects

Interesting and novel physical and electronic properties emerge as specimen size and dimensionality are further reduced. Nanotubes and nanowires are examples of one-dimensional objects that have attracted extraordinary attention from the scientific community during the last several years. Carbon nanotubes (CNTs) could perhaps even be considered as providing the original seeds for the ongoing nanotechnology revolution.[49] CNTs have remarkable electrical and mechanical properties, making them ideal for many demanding device applications. Our recent studies have focused on the possibility of using multiwalled CNTs to encapsulate single-crystal FeCo nanowires, which could then have potential applications in high-density magnetic storage devices. Figure 1.10 shows an example of an FeCo-filled nanotube produced at 700 °C.[50] Magnetisation measurements revealed that the coercive fields of this sample were very much larger at room temperature when compared to the bulk FeCo phase.

New methods for growing nanowires consisting of ceramic oxides, semiconductors, and metallic materials, are being reported in the scientific literature on an almost daily basis. We have recently contributed to a comprehensive study of silicide nanowires, which are potentially of interest for low-resistance interconnects and nanoscale contacts to quantum dots. Self-assembled nanowires of cobalt silicide with very large length:width aspect ratios were found to grow via an endotaxial mechanism, with growth into the substrate along inclined Si{111}

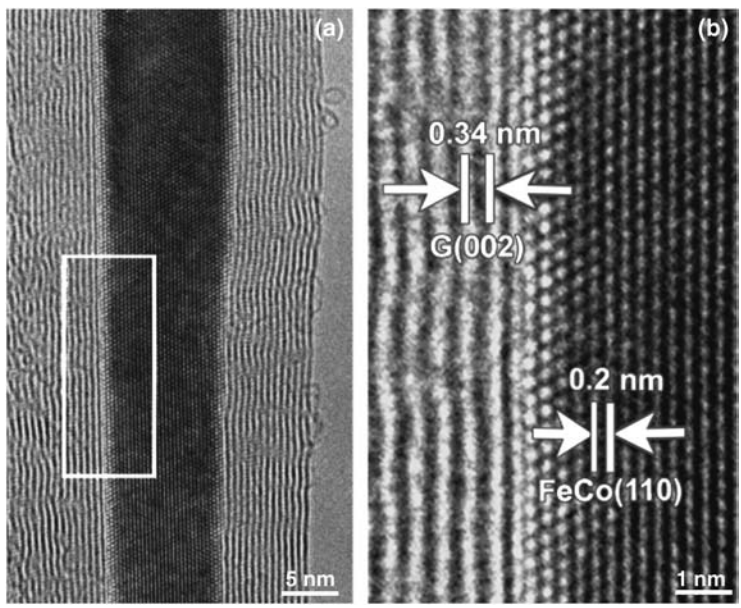


Figure 1.10 (a) High-resolution electron micrograph of FeCo-filled multiwalled carbon nanotube produced at 700 °C; (b) Magnified view of FeCo alloy inside the carbon nanotube. Reproduced from Ref. 50.

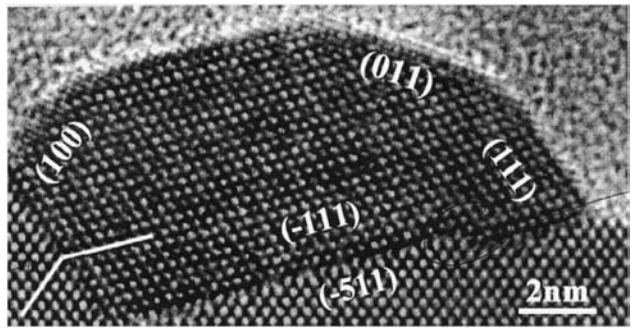


Figure 1.11 Cross section of $CoSi_2$ nanowire grown by reactive epitaxy on the Si(100) surface at 750 °C. Reproduced from Ref. 51.

planes. Figure 1.11 shows an example of a nanowire formed after reactive deposition of 1 monolayer of Co onto a Si(100) substrate at 750 °C.[51] Note the B-type interface indicated by the white line at the left, and the higher order {511}-type buried interface.

1.3.5 Zero-Dimensional Objects

The unique properties of the zero-dimensional object, or "quantum dot", represent an exciting area of research, especially in the field of semiconductors where changes in the size of the quantum dot (QD) will impact on optical properties, and incorporation of the QD into a circuit results in quantised conductance. QDs range in size from $\sim$1 nm up to perhaps 50 nm in some cases, depending on the particular application and the growth method. Many techniques are being used to fabricate QDs and much attention is being devoted to the challenge of achieving uniform size and shape. High-resolution imaging is ideal for characterising QDs because of their small dimensions. Figure 1.12 shows two examples of epitaxial Ge/Si(100) nanoscale islands from a sample with 14 monolayers of Ge grown by molecular beam epitaxy at 400 °C.[52] The very small island ($d \sim 25$ nm) in Figure 1.12(a) is free of defects and interfacial dislocations, whereas the larger dislocated dome cluster ($d \sim 65$ nm) shown in Figure 1.12(b) has several {111} stacking faults (arrowed) extending through to the surface. Recent studies of this system have used lithographic patterning to achieve better periodicity and size uniformity.[53]

The III-nitride system offers the prospect of optoelectronic devices spanning a range of wavelengths from infrared to deep-ultraviolet depending on the relative (Al, Ga, In) concentrations. QDs are of interest in this alloy system because they can be used as the active device region but the challenge for the crystal grower is again to achieve QDs with high levels of size uniformity and spatial distribution. Our TEM studies of GaN QDs, as illustrated by the truncated pyramidal dot visible in Figure 1.13, show distinct differences in dot morphology compared with the Ge/Si system.[54] These QD shapes can be

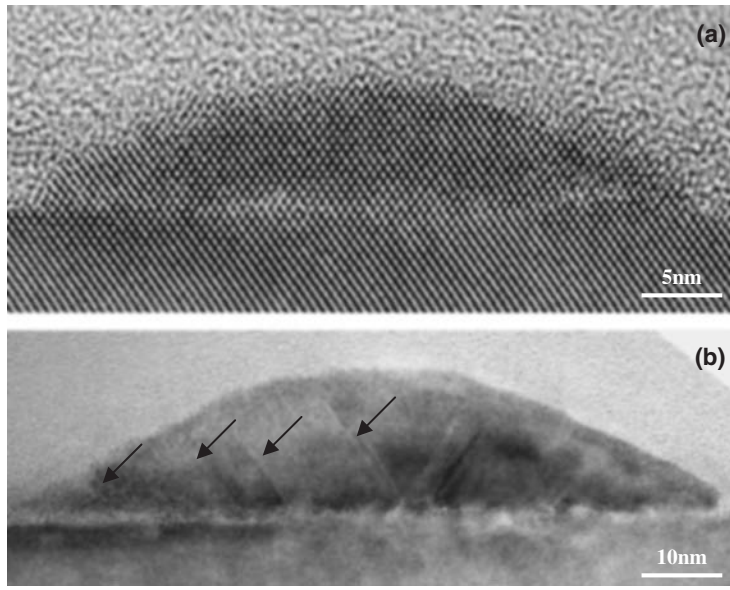


Figure 1.12 High-resolution electron micrographs showing cross sections of Ge/Si islands from a sample with 14 monolayers of Ge grown by molecular beam epitaxy at 400 °C. Note stacking faults arrowed in (b). Reproduced from Ref. 52.

attributed primarily to the nitride material having the wurtzite rather than the zincblende crystal structure.

1.3.6 Surfaces and Interfaces

Strictly speaking, interfaces and surfaces do not fit within the classification of nanomaterials. In both cases, however, there is an abrupt structural change, either from one material to another, or from bulk to vacuum, on a nanometre length scale or less. HREM again represents a powerful tool for elucidating unique structural information about interfaces and surfaces.

There are several types of interfaces between dissimilar materials, depending on their amorphous or crystalline nature and whether any (inter-) diffusion has occurred. Imaging of a planar interface in the edge-on geometry provides details about interface abruptness, smoothness, and flatness, and establishes the occurrence of any preferred orientational relationship (*i.e.* epitaxy) between the layers. The presence and nature of any defects or any further crystalline or amorphous phase(s) can also be determined. Grain boundaries are commonly observed in many monocrystalline systems, and it is often found that the misfit between grains with specific angular misfits is accommodated by well-defined, and usually nanoscale, structural units. In the case of the $\Sigma = 5$, (310) symmetric tilt grain boundary in NiO shown in Figure 1.14, two distinct structural units

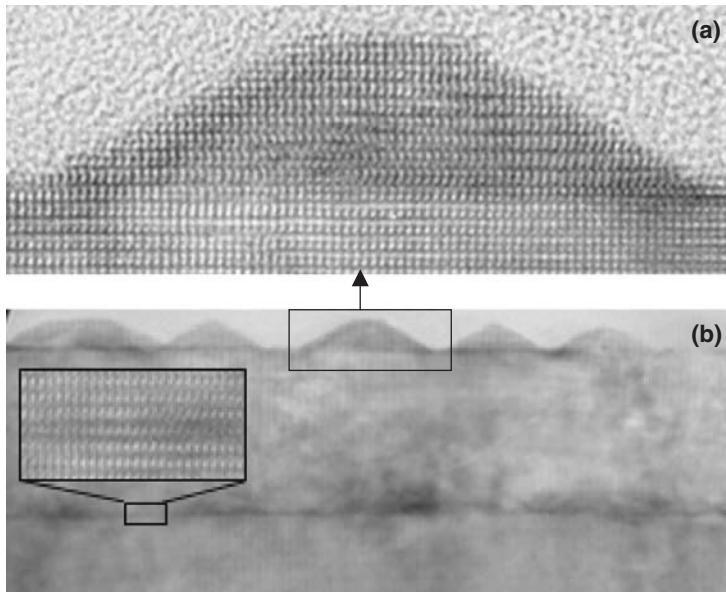


Figure 1.13 (a) High-resolution electron micrograph showing cross section of GaN quantum dot with truncated pyramidal shape; (b) lower magnification view showing GaN wetting layer (inset) and GaN QDs grown on AlN barriers. Reproduced from Ref. 54.

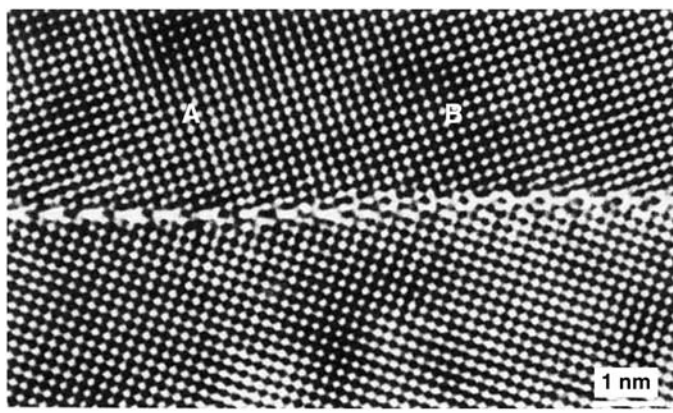


Figure 1.14 Atomic-resolution electron micrograph showing a $\Sigma = 5$, (310) symmetric tilt grain boundary in NiO recorded close to optimum defocus (black atomic columns). Note the change in geometry of structural units as boundary translates from region A to region B. Reproduced from Ref. 55.

are visible.[55] Their clearly asymmetrical nature differed from predictions based on earlier modeling calculations, leading eventually to refined structural models.

Examination of surfaces by electron microscopy has a long history, with many different materials and imaging geometries.[56] The surface-profile imaging

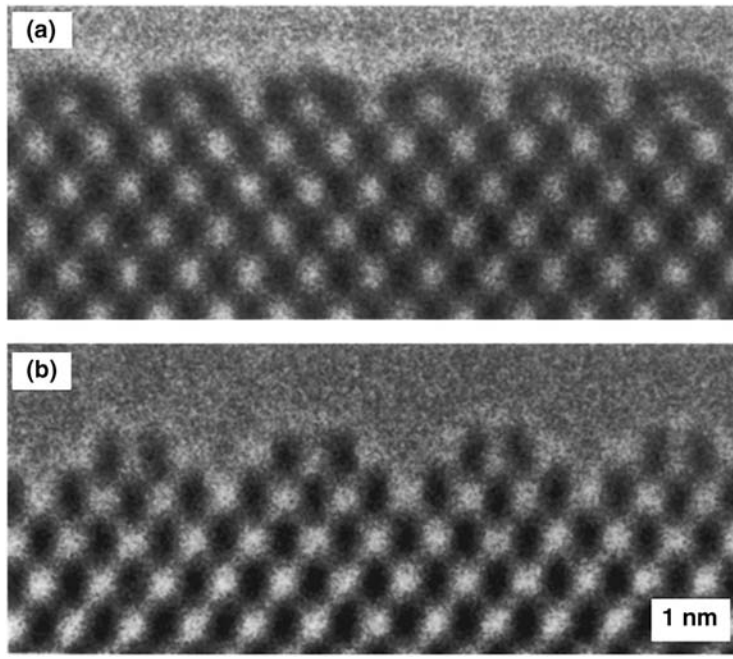


Figure 1.15 Surface-profile images showing a reconstructed CdTe (001) surface at different temperatures: (a) 2×1 at 140 °C; (b) 3×1 at 240 °C. Reproduced from Ref. 57.

geometry mentioned earlier, makes use of the full resolving power of the HREM to provide atomic-scale information about surfaces. Cleanliness is an obvious issue that could invalidate image interpretation but characterisation using a dedicated ultrahigh-vacuum instrument removes this concern. Abrupt surface termination of a bulk semiconductor will lead to a high surface energy associated with unsatisfied or ''dangling'' bonds, in turn driving a surface reconstruction. As shown in Figure 1.15, the CdTe (001) surface experiences a reversible phase transformation from a 2×1 structure at temperatures below about 200 °C to a 3×1 structure at more elevated temperatures.[57] Careful studies showed that the surface-terminating atoms could be identified, and the positions of atomic columns at the surface could be located to within about 0.01 nm.

1.4 Emerging Trends and Practical Concerns

1.4.1 Atomic Location and Quantitative Imaging

The determination of atomic arrangements at defects has historically been based around qualitative comparisons between experimental micrographs and image simulations derived from structural models.[1] Specific models were later regarded as more acceptable when the image-matching process involved

multiple members from a focal series.[58] Alternative approaches have used projected atomic column positions overlaid on the experimental and/or simulated images,[59] while other studies have utilised superposition[60] or subtraction[61] of simulated and experimental images. A nonlinear least-squares optimisation approach has even been used to refine atomic positions.[62]

One concern with this refinement process has been the issue of somehow estimating the "goodness of fit" between experimental micrographs and simulated images. The reliability or R-factors used in X-ray or neutron diffraction studies are applicable to the entire unit cell, whereas the aperiodic defects normally studied by HREM, such as grain boundaries or dislocation cores, do not usually have such well-defined discrete entities. Several Image Agreement Factors (IAFs) have been proposed: a useful summary of those most commonly used can be found in the Appendix of Ref. [63]. Slightly different results are obtained depending on the particular IAF used, which could be due to the fact that different IAFs weight bright and dark contrast areas differently. By using the complex exit-surface wavefunction of the specimen as a basis for comparison, defocus and other microscope parameters can be eliminated in the refinement process. Furthermore, it might be considered as advantageous in some studies that determination of the exit-surface wavefunction makes phase and amplitude information available. Because this wavefunction reflects electron scattering by the object, the process of solving unknown object structures could be enhanced. The refinement of atomic locations can, however, be demanding experimentally, and it can also be extremely time consuming because of tedious trial-and-error fitting of microscope and specimen parameters. Therefore prior knowledge of the essential experimental parameters can greatly facilitate the interactive structure refinement process.

1.4.2 Detection and Correction of Aberrations

As resolution limits steadily improve, higher order objective lens aberrations start to play a more dominant role in determining the overall integrity or faithfulness of the imaging process. For example, image reconstruction schemes that involve either focal series[28] or off-axis electron holography[64] require accurate knowledge of the spherical aberration coefficient with error bars of less than 1%. Two-fold image astigmatism results in focal length differences in orthogonal image directions, and will have a seriously detrimental impact on image appearance. However, this type of astigmatism can be detected using a thin film of amorphous material, such as carbon or silicon, since Fourier transformation of the recorded image will reveal concentric rings that are elliptical depending on the amount of astigmatism.[65]

The second order aberrations of axial coma (beam tilt) and three-fold astigmatism are less easily measured since they only visibly affect images of amorphous materials recorded with tilted incident illumination.[66,67] Three-fold astigmatism was first quantified during an early study of coma free alignment.[68] However, its impact on high-resolution imaging at the 1.0 Å level was not fully

appreciated until the Automatic Diffractogram Analysis (ADA) method for autotuning was implemented.[69] The effect of three-fold astigmatism is an asymmetrical shift of phase information so that its presence is not apparent in an axial bright field image of an amorphous material. Image simulations have shown, however, that very high resolution images of crystalline materials will be greatly affected, depending on the relative orientations of the three-fold astigmatism and the crystal symmetry axes.[70,71] The magnitude of the three-fold astigmatism can be estimated based on diffractogram analysis of images recorded with four mutually orthogonal beam tilt directions. Correction is then best achieved using hexapole stigmator coils located near the back focal plane of the objective lens. Reasonable adjustments can be reached using fixed correction currents in existing objective lens stigmator coils provided that the coils are energised separately rather than being wired in pairs.[72] Microscopists purchasing new microscopes for high resolution studies without aberration correctors, as described below, should at least insist that correction of three-fold astigmatism is carried out during microscope installation and testing so that its impact on their high-resolution imaging studies is reduced to the negligible level. Moreover, many microscopes are nowadays easily equipped for autotuning, meaning that focus, astigmatism and beam alignment (coma) can be carried out automatically using a CCD camera as well as computer control of relevant power supplies. As demonstrated in Figure 1.16, this process can eliminate two-fold astigmatism (two parameters) and coma/beam-tilt (two more parameters).

It is well known that spherical aberration cannot be avoided in rotationally symmetric electron lenses.[73,74] However, early attempts at C_S correction were reported to have failed because of insufficient electrical stability and lack of alignment precision.[74] Correction of spherical aberration becomes a critical issue as information limits approach and surpass the 1.0 Å barrier. Image detail can be made directly interpretable without needing to remove any artefactual detail as normally caused by PCTF oscillations. A double-hexapole system enabled the normal 2.3 Å interpretable resolution limit of a 200 keV FEG-TEM to be surpassed, and an information limit of $\sim$1.3 Å was finally reached once extraneous instabilities were removed.[75] Probe sizes of less than 1.0 Å were achieved in a 100 keV STEM by using multiple quadrapole-octapole elements to correct the spherical aberration of the probe-forming objective lens.[76] Note that both TEM and STEM approaches to aberration correction are completely dependent on computer supported analysis of the imaging conditions and also require high-precision feedback and control for the numerous deflector and corrector power supplies. An additional benefit of aberration corrected imaging is that the impact of image delocalisation, which causes severe distortions at discontinuities such as interfaces and surfaces, can be markedly reduced.[75,76] Another benefit is that other imaging aberrations can also be substantially reduced, which should simplify the process of exit wave retrieval using through focal reconstruction[77] and the accuracy needed for sample tilting is also alleviated. In practice, having an adjustable spherical aberration can provide additional flexibility to the microscopist interested in solving particular

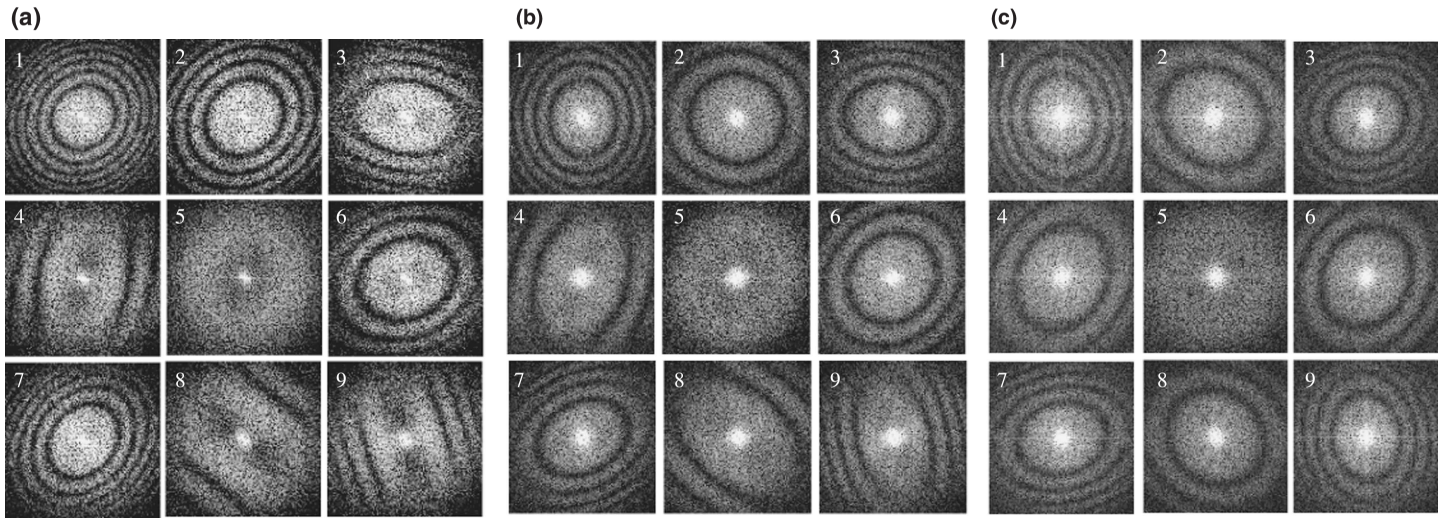


Figure 1.16 Diffractogram tableaus from thin amorphous carbon film recorded (a) before; (b) after one cycle; (c) after two cycles of an autotuning sequence based on Automated Diffractogram Analysis. Initial tilt misalignment reduced successively from 4 mrad (a) to 0.4 mrad (b) and <0.1 mrad (c). Courtesy of M.R. McCartney.

materials problems. For example, a small negative C_S with a slightly overfocus condition enabled clear imaging of oxygen-atom columns in a perovskite ceramic.[78] At the present stage of corrector development, it is clear that much careful work is still required to explore the full range of possibilities for aberration-corrected HREM and eventually to reach some consensus about standard imaging conditions.

Chromatic aberration does not affect the interpretable resolution of the HREM, meaning that knowledge of C_C, the chromatic aberration coefficient, is not essential for most high-resolution imaging. Moreover, the envelope for temporal coherence is determined by an effective focal spread, which can be estimated empirically.[25] Thus C_C impacts temporal coherence, so that any reduction or correction of chromatic aberration is considered as desirable. In the energy range applicable for HREM (upward from about 100 kV) C_C correction has not so far been reported, although it has been achieved in a low-voltage scanning electron microscope.[79] Alternatively, a monochromator located immediately after the electron gun can also be used to reduce chromatic effects. The energy spread can be reduced to about 0.1 eV,[80] but this improvement can only be achieved at the expense of reduced beam current.[81]

1.4.3 The Stobbs' Factor

The absolute intensity levels of high-resolution electron micrographs could not be easily measured in early studies using photographic film, so that microscopists tended to scale the contrast range of their simulations in a purely qualitative manner to agree with experimental micrographs. With the advent of the CCD camera, quantitative image recording has become widespread in high-resolution studies. However, careful comparisons between simulated and experimental micrographs have revealed significant discrepancies in contrast levels. Factors as large as 6 to 8 have been reported,[82] although factors of about three are reported to be more typical.[83] These differences are highly perplexing since careful investigations of possible sources of error, such as inelastic scattering and surface-contamination overlayers, have been unable to positively identify the cause(s).[84] Further concentrated efforts have been equally unsuccessful in fully accounting for this so-called Stobbs' factor, named after the microscopist who first alerted the microscopy community to the extent of the problem.[82] Recent experiments combining off-axis electron holography with energy-filtered imaging,[85] suggest that thermal diffuse scattering is at least a strong contributing factor to the contrast loss in experimental images.

1.4.4 Radiation Damage

Beam–specimen interactions are obviously the source of the multiplicity of signals used by electron microscopists to characterise their samples. At the same time, the microscopist must be ever vigilant to the possibility that these very same interactions can lead to permanent irreversible change in local

chemistry and microstructure. It is well known[86] that beam damage can be classified into two basic types: i) "knock-on" displacements, which take place above characteristic energy thresholds, involve ballistic collisions between incident electrons and atoms of the specimen; and ii) radiolytic processes, which are often referred to as "ionisation damage", involving interactions with electrons of the sample. These latter processes will affect most covalent and ionic solids. An incident beam energy of 200 keV is enough to cause atomic displacements for elements in the Periodic Table up to and including silicon ($Z = 14$). It should also be appreciated that the energy needed for diffusion at interfaces, or for sputtering from the exit surface of the sample, will be much less than for the bulk material. For example, by making careful image comparisons, it was shown that preferential damage relative to the bulk had occurred at a Cu/sapphire interface.[87] In this study (at 1250 kV, magnification of 600 000×, specimen current density of ~ 1.6 A cm^{-2}), the useful viewing time was limited to ~ 10 min. In another study of maximally valent, transition-metal oxides (at 400 kV, 5–15 A cm^{-2}), oxygen depletion was caused by electron-stimulated desorption from regions near the sample surface. As shown in Figure 1.17, a thin layer of reduced oxide then coated the surface.[88] Higher and higher magnifications are used for imaging as resolution limits are improved. While better signal statistics are obtained, because of the higher current densities, damage rates will be greater (the current density at the level of the sample will increase with the *square* of the magnification for constant screen brightness). Thus, it is strongly recommended that the image magnification and the beam current density should be limited whenever possible. Moreover, periodic checks should be made for signs of structural damage. Possibly erroneous results can be discounted once changes start to become apparent.[89] Finally, it should be appreciated that the intense focused probe used for microanalysis in the STEM is much more likely to cause localised beam damage and this possibility should always be closely monitored.

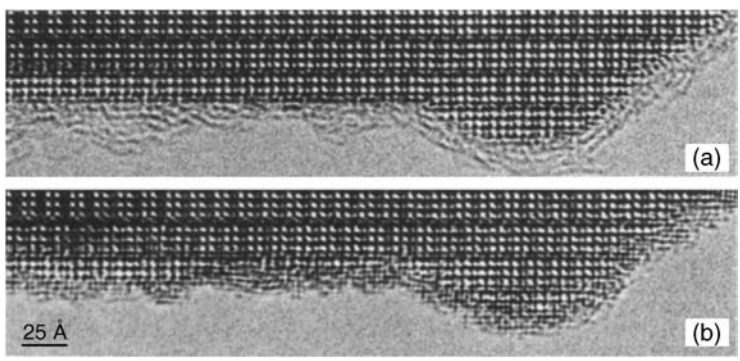


Figure 1.17 Profile images of $Ti_2Nb_{10}O_{29}$ recorded at 400 keV showing changes of the near-surface region due to electron-stimulated desorption of oxygen caused by continuous electron irradiation of 15 A cm^{-2} for period of $\sim$ 20 min. Reproduced from Ref. 88.

1.5 Prospects

This chapter has provided a brief overview of the transmission electron microscope, with the primary objective of highlighting some of its applications and achievements in the field of nanomaterials. The electron microscope clearly has the potential to have a huge impact in this emerging area of research because of its unique capacity to determine atomic structures with very high accuracy on a routine basis. The challenge for the HREM experts is to ensure that their colleagues who are more knowledgeable about the properties of nanomaterials are provided with the assistance with microscope operation that is required to capitalise on the exciting possibilities. Many novel applications can be anticipated to emerge over the next several years.

On the microscopy front, further developments can be expected. Aberration-corrected microscopy is in its infancy and much hard work is required to determine standard imaging conditions. Further gains in instrumental resolution might also be achieved through correction of higher-order aberrations and reduced chromatic effects. Digital recording and online microscope control (''autotuning'') are likely to become standard on next-generation instruments. Environmental electron microscopy with *in-situ* sample treatment is another likely area of concentrated activity as discussed in Chapter 7. These developments in instrumentation will also benefit the field of nanomaterials.

Acknowledgments

The author is pleased to acknowledge friends and colleagues for their involvement in the research briefly described here. Many of these studies involved use of the facilities in the John M. Cowley Center for High Resolution Electron Microscopy at Arizona State University.

References

1. D.J. Smith, *Rep. Prog. Phys.*, 1997, **60**, 1513.
2. F. Ernst and M. Rühle, *Curr. Opin. Solid State Mater. Sci.*, 1997, **2**, 469.
3. J.C.H. Spence, *Mater. Sci. Eng.*, 1999, **R26**, 1.
4. Z.L. Wang, *Adv. Mater.*, 2003, **15**, 1497.
5. M. Rühle, F. Phillipp, A. Seeger and J. Heydenreich, ed., *High-Voltage and High-Resolution Electron Microscopy*, *Ultramicroscopy*, **56**, Nos. 1–3, November 1994.
6. M. Sarikaya, H.K. Wickramasinghe and M. Isaacson, ed., *Determining Nanoscale Physical Properties of Materials by Microscopy and Spectroscopy, Mater. Res. Soc. Symp. Proc.*, 1995, **332**, Materials Research Society, Pittsburgh.
7. J.C.H. Spence, *Experimental High Resolution Electron Microscopy*, Clarendon, Oxford, 1988.

8. S. Horiuchi, *Fundamentals of High Resolution Transmission Electron Microscopy*, North Holland, Amsterdam, 1994.
9. D. Shindo and K. Hiraga, *High Resolution Electron Microscopy for Materials Science*, Springer, Tokyo, 1998.
10. H. Wang, M.R. McCartney, D.J. Smith, X. Jiang, R. Wang, S. van Dijken and S.S.P. Parkin, *J. Appl. Phys.*, 2005, **97**, 104514.
11. J.C. Bravman and R. Sinclair, *J. Electron. Microsc. Technol.*, 1984, **1**, 53.
12. S.B. Newcomb, C.B. Boothroyd and W.M. Stobbs, *J. Microsc.*, 1985, **140**, 195.
13. P.L. Gai, *Top. Catal.*, 2002, **21**, 161.
14. R. Sharma, *J. Mater. Res.*, 2005, **20**, 1695.
15. D.J. Smith, M.R. McCartney, C.L. Platt and A. Berkowitz, *J. Appl. Phys.*, 1998, **83**, 5154.
16. D. Van Dyck, *J. Microsc.*, 1983, **132**, 31.
17. P.G. Self and M.A. O'Keefe, in *High-Resolution Transmission Electron Microscopy and Associated Techniques*, ed. P.R. Buseck, J.M. Cowley and L. Eyring, Oxford University Press, New York, 1988, ch. 8.
18. J.M. Cowley and A.F. Moodie, *Acta Crystallogr.*, 1957, **10**, 609.
19. P. Goodman and A.F. Moodie, *Acta Crystallogr.*, 1974, **30**, 280.
20. P.A. Stadelmann, *Ultramicroscopy*, 1987, **21**, 131.
21. P.G. Self, M.A. O'Keefe, P.R. Buseck and A.E.C. Spargo, *Ultramicroscopy*, 1983, **11**, 35.
22. J.M. Cowley, *Diffraction Physics*, North Holland, Amsterdam, 3rd revised edn, 1995.
23. K.-H. Hanszen, *Adv. Opt. Electron Microsc.*, 1971, **4**, 1.
24. J. Frank, *Optik*, 1973, **38**, 519.
25. W.O. Saxton, *Optik*, 1977, **49**, 51.
26. D.J. Smith, *Adv. Opt. Electron Microsc.*, 1988, **11**, 1.
27. M.A. O'Keefe, *Ultramicroscopy*, 1992, **47**, 282.
28. W. Coene, G. Janssen, M. Op de Beeck and D. van Dyck, *Phys. Rev. Lett.*, 1992, **69**, 3743.
29. J.M. Cowley and A.F. Moodie, *Proc. Phys. Soc.*, 1957, **70**, 486.
30. S. Iijima and M.A. O'Keefe, *J. Microsc.*, 1979, **17**, 347.
31. D.J. Smith, R.A. Camps and L.A. Freeman, *Inst. Phys. Conf. Ser.*, 1982, **61**, 381.
32. J.L. Hutchison, T. Honda and E.D. Boyes, *JEOL News*, 1986, **24E**, 9.
33. M.A. O'Keefe, U. Dahmen and C.J.D. Hetherington, *Mater. Res. Soc. Symp. Proc.*, 1990, **159**, 453.
34. A. Thüst and K. Urban, *Ultramicroscopy*, 1992, **45**, 23.
35. W.E. King and G.H. Campbell, *Ultramicroscopy*, 1994, **56**, 46.
36. E. Ruska, *Z. Phys.*, 1934, **85**, 580.
37. S. Iijima, *J. Appl. Phys.*, 1971, **42**, 5891.
38. J.L. Hutchison and J.S. Anderson, *Phys. Status Solidi A*, 1972, **9**, 207.
39. D.J. Smith, R.A. Camps, V.E. Cosslett, L.A. Freeman, W.O. Saxton, W.C. Nixon, C.J.D. Catto, J.R.A. Cleaver, K.C.A. Smith and A.E. Timbs, *Ultramicroscopy*, 1982, **9**, 203.

40. L.A. Bursill and D.J. Smith, *Nature*, 1984, **309**, 319.
41. D. van Dyck, H. Lichte and K.D. vander Mast, *Ultramicroscopy*, 1997, **64**, 1.
42. F. Phillipp, R. Höschen, M. Osaki, G. Möbus and M. Rühle, *Ultramicroscopy*, 1994, **56**, 1.
43. J.N. Stirman, P.A. Crozier, D.J. Smith, F. Phillipp, G. Brill and S. Sivananthan, *Appl. Phys. Lett.*, 2004, **84**, 2530.
44. A.K. Datye and D.J. Smith, *Catal. Rev. Sci. Eng.*, 1992, **34**, 129.
45. D.J. Smith, R.W. Glaisher, P. Lu and M.R. McCartney, *Ultramicroscopy*, 1989, **29**, 123.
46. D.J. Smith, A.K. Petford-Long, L.R. Wallenberg and J.-O. Bovin, *Science*, 1986, **233**, 872.
47. Z.W. Pan, Z.R. Dai and Z.L. Wang, *Science*, 2001, **291**, 1947.
48. T. Aoki, M. Takeguchi, P. Boieriu, R. Singh, C. Grein, Y. Chang, S. Sivananthan and D.J. Smith, *J. Cryst. Growth*, 2004, **271**, 29.
49. S. Iijima, *Nature*, 1991, **354**, 56.
50. A.L. Elias, J.A. Rodriguez-Manzo, M.R. McCartney, D. Golberg, A. Zamudio, S.E. Baltazar, F. Lopez-Urias, E. Munoz-Sandoval, L. Gu, C.C. Tang, D.J. Smith, Y. Bando, H. Terrones and M. Terrones, *Nano Lett.*, 2005, **5**, 467.
51. Z. He, D.J. Smith and P.A. Bennett, *Phys. Rev. Lett.*, 2004, **93**, 256102.
52. D.J. Smith, D. Chandrasekhar, S.A. Chaparro, P.A. Crozier, J. Drucker, M. Floyd, M.R. McCartney and Y. Zhang, *J. Cryst. Growth*, 2003, **259**, 232.
53. J. Drucker, *IEEE J. Quantum. Electron.*, 2002, **38**, 975.
54. L. Zhou, D.J. Smith, T. Xu and T.D. Moustakas, unpublished results.
55. K.L. Merkle and D.J. Smith, *Phys. Rev. Lett.*, 1987, **59**, 2887.
56. D.J. Smith, in *Chemistry and Physics of Solid Surfaces VI*, ed. R. Vanselow and R. Howe, Springer, Berlin, 1986, ch. 15.
57. P. Lu and D.J. Smith, *Surf. Sci.*, 1991, **254**, 119.
58. J.C. Barry, *Philos. Mag.*, 1991, **A64**, 111.
59. A. Bourret, J.L. Rouviere and J. Spendeler, *Phys. Status Solidi A*, 1988, **107**, 481.
60. J.M. Penisson, T. Nowicki and M. Biscondi, *Philos. Mag. A*, 1988, **58**, 947.
61. T. Hoche, P.R. Kenway, H.-J. Kleebe and M. Rühle, in *Atomic Scale Imaging of Surfaces and Interfaces*, ed. D. Biegelsen, D.J. Smith and S.-Y. Tong, *Mater. Res. Soc. Symp. Proc.*, 1993, **295**, Materials Research Society, Pittsburgh, p. 115.
62. W.E. King and B.S. Lamver, in *Microbeam Analysis*, ed. D.G. Howitt, San Francisco Press, San Francisco, 1991, p. 217.
63. G. Möbus and M. Rühle, *Ultramicroscopy*, 1994, **56**, 54.
64. A. Orchowski, W.D. Rau and H. Lichte, *Phys. Rev. Lett.*, 1995, **74**, 399.
65. O.L. Krivanek, *Optik*, 1976, **45**, 97.
66. O.L. Krivanek, *Ultramicroscopy*, 1994, **55**, 419.
67. D.J. Smith, W.O. Saxton, M.A. O'Keefe, G.J. Wood and W.M. Stobbs, *Ultramicroscopy*, 1983, **11**, 263.

68. F. Zemlin, K. Weiss, P. Schiske, W. Kunath and K.-H. Herrmann, *Ultramicroscopy*, 1978, **3**, 49.
69. O.L. Krivanek and P.E. Mooney, *Ultramicroscopy*, 1993, **49**, 95.
70. W.O. Saxton, *Ultramicroscopy*, 1995, **58**, 239.
71. O.L. Krivanek and P.A. Stadelmann, *Ultramicroscopy*, 1995, **60**, 103.
72. M.H.F. Overwijk, A.J. Bleeker and A. Thust, *Ultramicroscopy*, 1997, **67**, 163.
73. O. Scherzer, *J. Appl. Phys.*, 1949, **20**, 20.
74. H. Rose, *Ultramicroscopy*, 1994, **56**, 11.
75. M. Haider, H. Rose, S. Uhlemann, E. Schwan, B. Kabius and K. Urban, *Ultramicroscopy*, 1998, **75**, 53.
76. O.L. Krivanek, N. Dellby and A.R. Lupini, *Ultramicroscopy*, 1999, **78**, 1.
77. J.H. Chen, H.W. Zandbergen and D. van Dyck, *Ultramicroscopy*, 2004, **98**, 91.
78. C.L. Jia, M. Lentzen and K. Urban, *Science*, 2003, **299**, 870.
79. M. Haider and J. Zach, In: *Proc. Microscopy and Microanalysis 1995*, ed. G.W. Bailey, M.H. Ellisman, R.A. Henninger and N. Zaluzec, Jones and Begall, New York, 1995, p. 596.
80. H. Rose, *Optik*, 1990, **85**, 95.
81. H.W. Mook and P. Kruit, *Ultramicroscopy*, 2000, **81**, 129.
82. M.J. Hÿtch and W.M. Stobbs, *Ultramicroscopy*, 1994, **53**, 191.
83. C.B. Boothroyd, *J. Microsc.*, 1998, **190**, 99.
84. C.B. Boothroyd, *Ultramicroscopy*, 2000, **83**, 159.
85. C.B. Boothroyd and R.E. Dunin-Borkowski, *Ultramicroscopy*, 2004, **98**, 115.
86. L.W. Hobbs, in *Quantitative Electron Microscopy*, ed. J.N. Chapman and A.J. Craven, Scottish Universities Summer School in Physics, Edinburgh, 1984, p. 399.
87. G. Dehm, K. Nadarzinski, F. Ernst and M. Rühle, *Ultramicroscopy*, 1996, **63**, 49.
88. D.J. Smith, L.A. Bursill and M.R. McCartney, *Ultramicroscopy*, 1987, **23**, 299.
89. D. Hofmann and F. Ernst, *Ultramicroscopy*, 1994, **53**, 205.

CHAPTER 2

Scanning Transmission Electron Microscopy

A. R. LUPINI, S. N. RASHKEEV, M. VARELA,
A. Y. BORISEVICH, M. P. OXLEY, K. VAN BENTHEM,
Y. PENG, N. DE JONGE, G. M. VEITH, S. T. PANTELIDES,
M. F. CHISHOLM AND S. J. PENNYCOOK

Materials Science and Technology Division, Oak Ridge National Lab, Oak Ridge, TN, 37831–6031, USA

2.1 Introduction

The Scanning Transmission Electron Microscope[1] (STEM) is one of the most useful tools in many areas of atomic-scale materials science and nanocharacterisation. A STEM has the ability to generate local maps of the chemical composition and electronic structure at atomic resolution, even in complex or unknown samples. In this chapter we describe the important components of a state-of-the-art aberration-corrected STEM and discuss the application of STEM imaging to a variety of nanoanalysis examples in catalysis and solid-state materials problems.

In his famous lecture ''There's plenty of room at the bottom,'' Richard Feynman[2] argued that the best way to solve many outstanding fundamental problems in physics and biology would be to ''*just look at the thing*''. We use the STEM because we are literally able to *see* nanostructures. Importantly STEM images are often amenable to direct or intuitive interpretation, which can be difficult for conventional Transmission Electron Microscope (TEM) images. This ability also contrasts to bulk techniques, such as diffraction or bulk spectroscopy, where translating the available information into a real-space model can be extremely difficult. Later in that same lecture, Feynman put forward a challenge, ''*Is there no way to make the electron microscope more powerful?*'' Aberration correction directly addresses this challenge and is currently one of the most exciting developments in electron optics. We will outline

how aberration correction is implemented in a STEM and applied in practice. We will base our discussion on the aberration corrected STEMs at Oak Ridge National Laboratory (ORNL) and use examples drawn from our recent results published elsewhere.

2.1.1 Basic Description

Like the conventional TEM, the STEM primarily uses transmitted electrons to form an image. However, like the Scanning Electron Microscope (SEM) a STEM scans a very small probe over a sample. In its basic form the STEM consists of an electron source, several lenses to focus these electrons into a small probe, a scanning unit to scan this probe across the sample and a detector that collects a signal after the electrons have interacted with the specimen. An image is formed by recording a signal of interest as a function of the probe position. The image therefore relates to the part of the sample that the probe interacts with at each position. The probe in a modern instrument can be about the size of an atom (or smaller!), although the relevant interaction volume may be much larger. There is also considerable variety in the nature of images that can be formed by recording different signals.

Although the preceding discussion covered the important components, a practical instrument is rather more complex than the simple concept presented here. Figure 2.1 shows a more detailed, but still simplified, schematic based on the aberration-corrected STEMs at ORNL originally manufactured by VG microscopes.

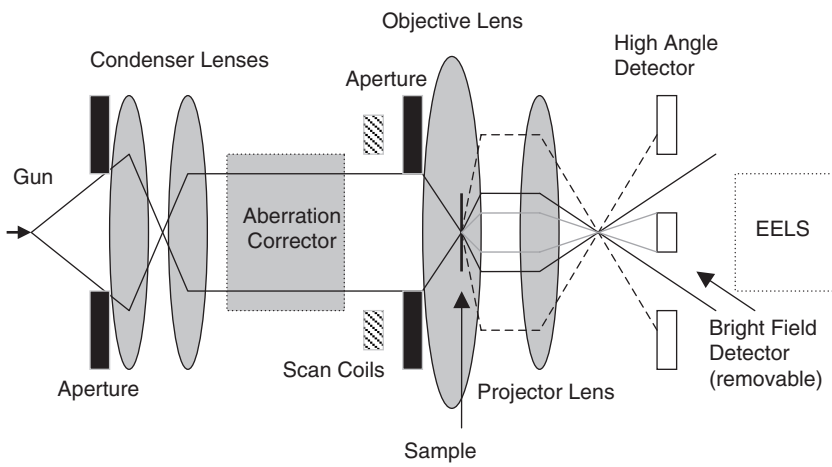


Figure 2.1 Schematic of an aberration-corrected STEM. Electron trajectories at the edge of the apertures are indicated with solid lines. High-angle scattering used to form the Z-contrast image is indicated with dashed lines and low-angle scattering used to form the bright-field image is indicated with grey lines.

The electron source in a STEM is often referred to as the ''gun''. The emitted electrons are accelerated through the column by a high voltage, which is in the range of 100–300 kV for the instruments considered here. There are three main choices of electron source: cold field emitter,[3] thermally assisted field emitter,[4] and heated filaments. The VG Microscopes' STEMs that we employ use a cold field emitter, which produces a low energy spread in the electron beam ($\sim$0.3 eV) and high brightness ($\sim 10^8$ A sr^{-1} m^{-2} V^{-1}), defined as the current per unit area, per unit solid angle, divided by the beam energy. Brightness is important, because it cannot be increased and will limit the current in a given size probe. The total current in the probe will ultimately determine the image acquisition time and signal-to-noise ratio; two of the key factors for practical microscopy. Recently, it was shown that electron sources made from carbon nanotubes can be extremely bright[5] and it can be expected that ultra-high brightness sources will be available in the future.

All electron sources emit electrons that have a spread of energies. This spread is undesirable because electrons with different energies will not all be focused equally, due to chromatic aberration, degrading the energy resolution in spectroscopy and spatial resolution in imaging. A monochromator can be used to decrease the energy spread of the electrons. Several different designs for these exist, but the general concept is to disperse the electrons based on energy and select the desired range.[6] Clearly this process will reduce the available current and the principal difficulty is to reduce the energy spread without losing too much current.

The purpose of the lenses (condenser and objective) between the gun and the sample in a STEM is to project an image of the source onto the sample to form the probe. The virtual size[3] of the source is broadened by vibrations and aberrations, but can be smaller than the physical size. The condenser and objective lenses are used to demagnify the source in order to limit the contribution to the final spot size. However, there is a reciprocal relationship between the angular and spatial magnifications; as the size in real space is reduced, the angles are increased. Since the angles are ultimately limited by the apertures used, this process results in a reduction in the beam current. However, brightness, as defined above, is conserved. It is possible to include the effect of an extended source size in simulations of STEM images by convolving the simulated image with a source function, normally a Gaussian of appropriate width. Thus a large source will degrade the image resolution. Unfortunately, because demagnifying the source reduces the beam current, the optimal probe will involve a compromise between these two factors.

2.1.2 Detectors

A wide variety of detectors can be used in a STEM, each of which will generate an image that provides different information. The collected signal could be emitted X-rays, light, backscattered or transmitted electrons, and others (a useful summary is presented by Williams and Carter[7]). Here we will first consider axial bright field and annular dark-field detectors.

A bright-field detector can be used to collect the transmitted electrons on axis. In this configuration, a STEM looks very much like a TEM run backwards (in Figure 2.1, imagine interchanging the source with the bright-field detector). In fact, the principle of reciprocity[8,9] tells us that if we keep everything the same (apertures, angles, lens strengths, *etc.*), for *elastic* scattering, we obtain an identical image upon interchanging the source and detector. The principle of reciprocity does not apply to inelastic scattering, where the sample absorbs energy from the beam and so the initial and final states of the sample are different.[10] This principle, arising because of time-reversal symmetry, is extremely useful for understanding bright-field imaging. For purely elastically scattered electrons, we obtain the same image in bright-field TEM as in bright-field STEM. The collection angle in a STEM is equivalent to the illumination angle in TEM, the difference being that in a TEM all of the image points are recorded in parallel, whereas in a STEM the points are recorded serially. So a bright-field STEM will be less efficient; but has the advantage that a wide variety of other signals that can be collected simultaneously.

Perhaps the most useful detector for a STEM is a High-Angle (HA) Annular Dark-Field (ADF) detector, which collects the transmitted electrons scattered to high angles. This detector provides a Z-contrast image, which is so-called because the contrast depends upon the thickness and (approximately) the square of the atomic number Z. A simple model would suggest that an ADF detector would record Rutherford scattering ($\sim Z^2$, if one neglects atomic recoil and other factors). The bright spots correspond to the locations of atomic columns (to a good approximation) and the brightness gives an indication of their composition. Thus an atomic resolution Z-contrast image is frequently simpler to interpret than a high-resolution bright-field image. Z-contrast imaging is therefore well suited to unknown or unexpected structures, and is almost unrivalled for locating single dopant atoms in or on a lighter support. To illustrate this we will give practical examples including heterogeneous catalysts, semiconductors, and complex oxides taken from published work.

2.1.3 Electron Energy-loss Spectroscopy

One of the benefits of a STEM is that Z-contrast images can be acquired simultaneously with Electron Energy-Loss (EEL) spectra[11,12] because the Z-contrast image relies upon high-angle scattering, while the EEL spectrometer collects low-angle electrons. When the incident electron beam is transmitted through the sample, some of the incident electrons will transfer energy to the electrons in the sample and a spectrometer can be used to record the energy lost by the transmitted electrons. In its simplest form, a spectrometer is a magnetic field that bends the electrons with a bending radius that depends on their energy. Thus, recording the energy-dispersed electrons on a Charge-Coupled Device (CCD) allows the energy lost to be quantified.

To understand what the EEL spectrum represents, we need to consider the energy absorbed by the sample. Low energy losses are a bulk excitation

(e.g. valence-electron excitations, plasmons, *etc.*) and so depend on the dielectric properties of the sample. In a core-loss process, an inner-shell electron in the sample is excited from an occupied state to an unoccupied state above the Fermi level. The initial state is a bound core state and the final state depends on the band structure of the solid. Thus, the EEL spectrum contains information on the *chemical composition* of the sample and also probes the local *electronic structure*.[11,12] Because the STEM probe can be similar in size to an atomic column it is in principle possible to extract this information with atomic resolution. However, in practice, atomic-resolution EEL spectra can be difficult to interpret.

In a thick specimen, electrons undergo multiple scattering. Electrons that lose energy can undergo another inelastic scattering event. The multiply scattered spectrum can therefore be regarded as a convolution (in energy), which can degrade the energy resolution and alter the energy loss near-edge fine structure. Various deconvolution techniques exist but these often amplify noise and produce artifacts.

The EEL signal is ''delocalised'' because the coulomb interaction has a range that allows the incident electron to excite a transition at a distance. Non-locality[13,14] arises from the interference of different components of the incident electron wavefunction through the mixed dynamical form factor.[15] Furthermore, the probe spreads out due to the finite convergence angle, is attracted to the nuclei and is scattered as it propagates through the specimen. Thus, the probe within the crystal may have a very distorted shape.[16] A full dynamical calculation[17–19] including the propagation of the probe through the sample and the effective non-local potential is therefore essential to reproduce some of these details. In the small angles normally used in EEL spectroscopy the dipole approximation is extremely useful and leads to selection rules (only some transitions are allowed). In addition, the EEL spectrum depends on the band structure around the atom that is excited; the *local* density of states. However, the band structure is further complicated because the excitation of a core electron to a different state also generates a hole at that atom, which is most important for nonmetals. The $Z + 1$ approximation attempts to simulate this effect by replacing the atom that loses an electron with an element of one atomic number higher.[20] One could go another step further and also include the dynamical interaction between the hole and the excited electron. There is thus a wealth of information to be extracted, although detailed interpretation can require extensive band-structure[21] and scattering calculations.[17–19]

The ability to obtain EEL spectra from a single atom in its bulk environment has recently been achieved by Varela and coworkers[22] who were able to detect a single dopant atom inside a $CaTiO_3$ sample. The sample was doped with very carefully controlled concentrations of La atoms in different layers. The single La atoms were found by locating the layers with the appropriate doping, using layers with higher concentrations as markers and calibration standards. Atoms could be identified using the Z-contrast image; La is heavier than Ca for which it substitutes and so makes the column appear brighter. The identity of the dopant is confirmed by the EEL spectra.

Figure 2.2 shows a Z-contrast image and the spatially resolved spectra taken from the columns on and around the dopant atom site. Remarkably, it was found that even when only a single atom was present, a very faint signal could be detected on nearby columns. In order to understand why a signal from a single atom can be detected on neighbouring columns, it was necessary to perform a full dynamical simulation.

The results of such a calculation are also shown in Figure 2.2. The faint signal on adjacent columns is dependent on the depth of the dopant La atom within the crystal. The dynamical propagation of the probe through the sample affects the signal recorded. We can thus estimate that this particular dopant is of the order of 10 nm beneath the surface of the sample. Note, that the signal drops to almost zero beyond the nearest-neighbour columns. It is therefore apparent that we need to take delocalisation into consideration, but also that it does not preclude our obtaining atomically resolved EEL spectra. Indeed experimental demonstration of atomic-column-resolved EEL spectra and dynamical simulations were given by Allen and coworkers.[23] Even though the probe spreads out and the signal is "delocalised", different columns generate different EEL spectra. The key experimental factors were found to be preparing a good, thin sample and using a large collection angle to collect as much and as localised[18] a signal as possible.

Atomically resolved EEL spectra can be effectively applied to complex oxides, which are very relevant to many areas of materials science and solid-state physics. These materials exhibit a fascinating array of properties including

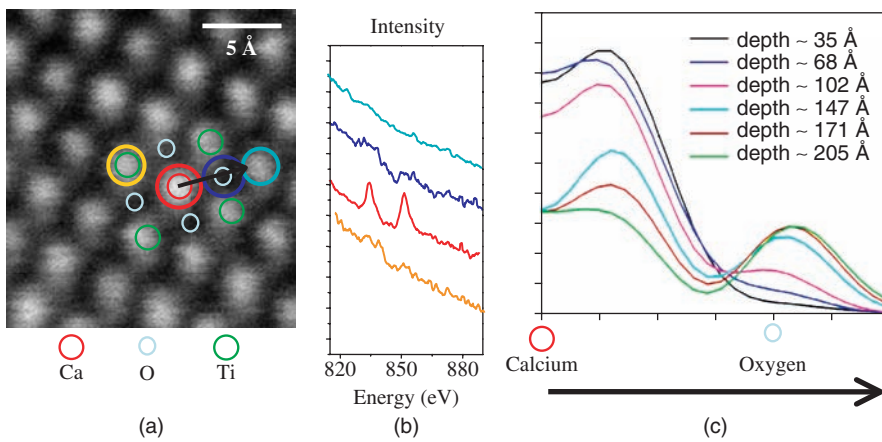


Figure 2.2 Spatially resolved EEL spectra from a single atom. (Left) Z-contrast image; bright columns are Ti (green), dimmer columns are Ca (red), O columns (blue) are not resolved, the single brightest column contains one La atom (red open circle). (Middle) Spectra from individual columns, corresponding to the open circles on the Z-contrast image. (Right) The *integrated* EELS intensity expected on and between the Ca and O columns for single La atoms inside the crystal at the depths indicated. Data reproduced from reference 22.

charge-ordering, superconductivity, colossal magnetoresistance, and ferromagnetism, and have a wide variety of applications of both scientific interest and practical utility. In complex oxides, particularly at interfaces, the nanoscale structure can have a dramatic effect on the bulk properties. This is especially true for high-Tc superconductors. In Figure 2.3, we show a Z-contrast image of an interface in a $YBa_2Cu_3O_7/La_{0.67}Ca_{0.33}MnO_3$ superconducting/ferromagnetic multilayer structure and atomically resolved EEL spectra across the interface.[24] This figure reveals that there are oscillations in the intensity of the Ba, Cu and Mn edges that peak at the relevant atomic plane positions. Note that the Ba and Cu signals peak at different locations, which confirms that this oscillation is not just a modulation of the signal by the increased scattering to high angles at heavy columns; it is not just an incoherent bright-field image. Thus, we are able to determine atomic plane stacking sequences from EEL spectrum images.

It is widely believed that in $YBa_2Cu_3O_7$, transfer of electrons from the CuO_2 planes to the CuO chains generates the holes responsible for superconductivity. Hence the ability to identify planes from EEL spectra is very important. The EEL spectra reveal that the stacking sequence at the interface is ...BaO-CuO_2-Y-CuO_2-BaO-MnO_2-LaO..., which can be confirmed by comparison with the Z-contrast image.[25] The stacking sequence in the perfect superconductor would be ... BaO-CuO_2-Y-CuO_2-BaO-CuO ... , which means

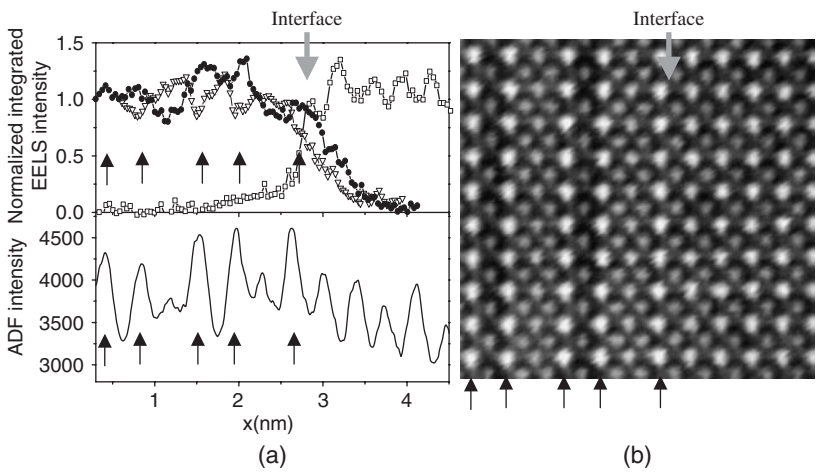


Figure 2.3 EELS of a $YBa_2Cu_3O_7/La_{0.67}Ca_{0.33}MnO_3$ superconducting/ferromagnetic interface. (Left) The integrated EELS intensity and ADF intensity averaged over several cells in a linescan across the interface (grey arrows) for: Ba M edge (solid circles), the Cu L edge (open triangles), the Mn L edge (open squares). (Right) Z-contrast image at the same magnification. Black arrows are used to indicate the BaO layers in the $YBa_2Cu_3O_7$; as Ba is the heaviest element, these are the brightest columns. The BaO layers can also be located from the EELS line trace. Data adapted from reference 24.

that a CuO chain is missing at the interface, being replaced with a MnO_2 plane. Thus, EELS is able to directly provide an explanation for some puzzling experimental results from these multilayer structures: The thinnest layers (1 or 2 unit cells) are not found to be superconducting.[26,27] Z-contrast imaging and X-ray scattering were able to demonstrate that the layers were well ordered and essentially perfect.[24–25,27] However, the stacking sequence as obtained from the EEL spectra above provides further answers. Single layers of $YBa_2Cu_3O_7$ are not superconducting because the CuO planes are missing and even double layers have too few CuO planes. Thicker layers have sufficient CuO planes that at least some of the layers are complete and superconducting.

This example has demonstrated that the combination of EELS and Z-contrast imaging is able to solve important fundamental materials problems. We are able to gain unique insight into what is really going on at the nanoscale. Significantly, both of the above examples relied on the enhanced resolution and sensitivity provided by aberration correction.

2.2 Aberration Corrected STEM

Aberration correction is perhaps the most exciting development taking place in electron optics at the present time. Even though the basic concept has been known for more than 50 years, it is only recently with modern, computer-controlled electronics that it has become a practical reality.

2.2.1 The Aberration Function

In a microscope, lens imperfections are described by aberrations, the most significant of which in a round electron lens are spherical aberration (C_3) and chromatic aberration (C_c), which will be described below. An aberration function is expressed as a sum over various aberration coefficients in a series expansion, which in the most general case is rather complicated.[28,29] However, in STEM for most purposes, we can make the approximation that the aberration function is a function only of angle to the optic axis θ, which is small, so the high-order terms can be neglected. The aberration function gives the phase shift imparted by the lens as a function of angle. The aberration function for the first four round aberrations, when considering pure geometric terms only, is given by:

$$\chi(\theta) = \frac{1}{2}C_1\theta^2 + \frac{1}{4}C_3\theta^4 + \frac{1}{6}C_5\theta^6 + \frac{1}{8}C_7\theta^8 \tag{2.1}$$

where C_1 is the defocus, C_3 the spherical aberration, C_5 the fifth-order round aberration (which is important *after* C_3 correction) and C_7 the seventh-order round aberration. Obviously a similar notation can be extended to nonround aberrations, such as astigmatism, by introducing a dependence on the azimuthal angle, which increases the number of terms needed. We have listed

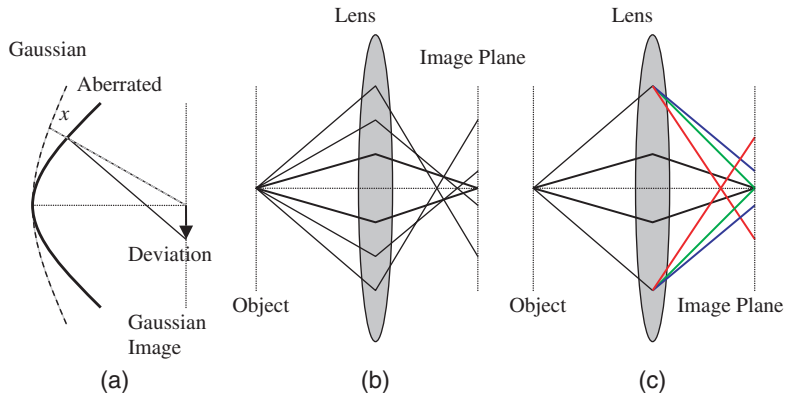


Figure 2.4 Simple illustrations of aberrations. (a) definition of general aberrations by an aberration surface. The aberration function is the distance from the aberrated wavefront to the Gaussian wavefront. (b) Electrons traveling at higher angles to the axis are focused too strongly by spherical aberration, C_3. (c) Chromatic aberration, C_C, causes electrons with different energies (indicated by colour) to be focused differently.

C_7, which can normally be neglected, to illustrate how the series continues; if the fifth- and sixth-order aberrations are corrected then this term would be significant. As used here, the aberration function has units of length and when multiplied by $2\pi/\lambda$ defines a phase change across the aperture plane of the objective lens.

An equally useful description would be to consider the probe to be made up of a series of rays at different angles. In that case the aberration function describes the distance between the surface of constant phase and the spherical surface of constant phase in an ideal case with all rays focused to the same point (Figure 2.4). Since rays propagate normally to such surfaces, the ray deviation (where the ray intersects the Gaussian focal plane) will depend on the *gradient* of the aberration function. Ray deviations are hence one order lower than the corresponding phase changes: C_3 is a third-order aberration because the ray deviation is $C_3\theta^3$ and has a fourth-order phase change: $\frac{2\pi}{\lambda}\frac{1}{4}C_3\theta^4$.

2.2.2 Spherical and Chromatic Aberration

To describe a general lens imperfection to fifth order would need 25 aberration coefficients to be included in the expression for the aberration function.[30,31] Fortunately, most of these coefficients are small in a round lens because of symmetry, and can be kept small by suitable lens design and column alignment. However, in 1936, Scherzer[32] proved that any lens that is rotationally symmetric, produces a real image, has a constant field, and has no charge on the axis will always suffer from both positive spherical and chromatic aberration. Therefore the practical resolution limit in a modern instrument will normally depend on these two aberrations and is about 50 times worse than the

wavelength (λ) of the electrons used (which is 2.0–3.7 pm for 300–100 kV electrons). The effects of spherical and chromatic aberration are shown in Figure 2.4. Spherical aberration causes rays (or electrons) travelling at higher angles to the optical axis to be focused too strongly. Chromatic aberration causes the focusing to depend on the energy (and there is always some energy spread in the beam). Thus with a real lens, even a perfect point-like object would generate a blurred image.

Round lenses are the primary means of focusing the electrons in a STEM. A simple round lens is shown in Figure 2.5. Most of the field is in the z-direction (along the electron trajectory). Because the magnetic field produces a force perpendicular to the velocity, focusing is a second-order effect. Due to their small transverse component of velocity, the electrons rotate about the optic axis and due to that tangential velocity, the magnetic field produces a radial force and the electrons spiral in towards the axis. A round lens is therefore always going to be relatively inefficient. However, the aberrations in round lenses are small on an absolute scale; the objective lens in a conventional TEM or STEM is of a very high quality. Round lenses also share the rotational symmetry that is desirable for the probe in most applications. Most importantly a round lens

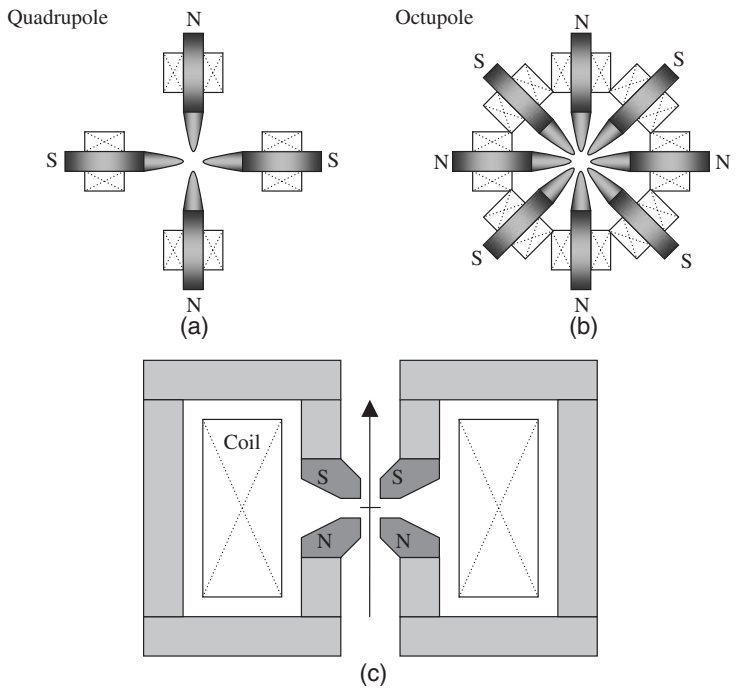


Figure 2.5 Simplified illustrations of some electron optical elements. (a,b) Top view of magnetic multipoles, which consist of poles with alternating polarities. The beam direction is out of the page. (a) Quadrupole and (b) Octupole. (c) Cross section through a round magnetic lens. The electron beam travels upwards (arrowed) and most of the field is in the same direction.

allows a very short focal length, of the order of 1 mm, allowing instabilities and imperfections elsewhere in the column to be conveniently demagnified. Therefore, even in an aberration corrected system, the round lenses are important.

2.2.3 Aberration Correctors

The general route to correct lens aberrations was given by Scherzer[33] in 1947 and to eliminate spherical or chromatic aberration, we must break one of the conditions of his proof. The most practical condition to exploit seems to be rotational symmetry. Instead of round lenses, we therefore use nonround elements, known as multipoles. Some example multipoles are shown in Figure 2.5. Multipoles are named after their rotational symmetry: dipoles, quadrupoles, sextapoles (or hexapoles), octupoles and so on. The magnetic field inside a multipole has to be a solution to the Laplace equation so unlike light optics, where lenses can have arbitrary shapes, the radial field depends on the rotational symmetry. Thus in order to obtain the desired, radially symmetric, correction an aberration corrector has to consist of a series of different elements (Figure 2.6).

At present, there are two main types of aberration corrector, which both have a long history:[29,34,35] The quadrupole-octupole corrector[30,31] and the round lens-hexapole corrector.[36–39] Both classes of corrector consist of correcting elements (octupoles or hexapoles), which do the correction, and first-order elements (quadrupoles or round lenses), which primarily steer the electron trajectories. In a quadrupole-octupole corrector, the octupoles have a field with a suitable radial dependence to correct C_3, but have an undesirable 4-fold rotational symmetry (see Hawkes and Kasper[29] for a complete derivation). Quadrapoles are used to deform the beam in order to pick up the desired components of the phase change from three suitably excited octapoles, but

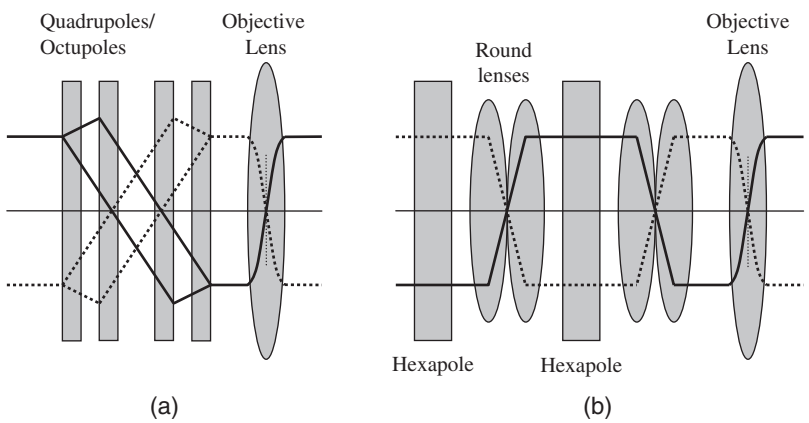


Figure 2.6 Simplified illustrations of the two main sorts of aberration corrector and their first-order trajectories. Solid lines *xz*-plane and dotted lines *yz*-plane, neglecting rotations. (a) a quadrupole-octupole corrector in which trajectories are astigmatic. (b) A round-lens-hexapole corrector.

leave a round beam after correction. One advantage of this type of corrector is that adding electrostatic elements allows the focusing to depend on the energy, which is exactly what is required for C_C correction.[40]

In a hexapole corrector,[38] the extended hexapoles correct C_3, but have an undesired 3-fold symmetry. The correction by hexapoles relies on *extended* multipoles where the field changes the trajectories, which means that the electrons go through different parts of the field and are further deflected.[37] Pairs of round lenses are used to project the beam from one hexapole to the next and into the objective lens, such that the correction is retained and the final beam is round. This type of corrector can be relatively simple whilst retaining good high-order aberrations.[41]

It was about 50 years after Scherzer[33] first showed the outline of how to correct spherical aberration before researchers were able to demonstrate a working corrector that actually improved the resolution in a high-resolution microscope.[31,36] The practical problem is that an aberration corrector is complicated; even the simplified designs illustrated here would require about 40–60 controls in practice. In order to actually improve the resolution, each element has to be aligned to within a few micrometres of the electron beam, which is extremely difficult to achieve mechanically.[42] Thus, many alignment, trim, and rotation controls are needed. The difficulty then is to set so many power supplies to the correct values with the required stability (better than 1 ppm) and in practice, computer control turns out to be essential in an aberration-corrected system.

2.2.4 What Do We See in a STEM?

In order to understand why we need aberration correctors, we first review what a STEM image represents by considering the limits of very small angle (coherent bright field) and very high-angle (annular dark field) detectors. In both images, which are acquired *simultaneously*, the electron probe is formed in the same way. We assume that the objective aperture is uniformly illuminated and the phase change across the aperture is given by the aberration function. The 2-dimensional probe amplitude incident on the sample surface is then given by a Fourier transform as:

$$P(\mathbf{R}, \mathbf{R}_0) = \int H(\mathbf{K}_i) \mathrm{e}^{2\pi i \chi(\mathbf{K}_i)/\lambda} \mathrm{e}^{-2\pi i \mathbf{K}_i.(\mathbf{R}_0 - \mathbf{R})} \mathrm{d}\mathbf{K}_i \tag{2.2}$$

where we have omitted the z-direction dependence (so vectors are 2-dimensional). $H(\mathbf{K}_i)$ is simply a top-hat function that describes the range of angles admitted by the objective aperture (with $|\mathbf{K}_i| = \theta/\lambda$). The probe position is $\mathbf{R}_0$, while $\mathbf{R}$ is just a coordinate on the sample. As we will see, in the STEM we do not record anything as a function of $\mathbf{R}$, which is regarded as a dummy variable. The intensity in the probe is given by $|P(\mathbf{R},\mathbf{R}_0)|^2$, which depends on the aberrations (including defocus), and changes as it propagates through space or through a sample.

Figure 2.7 shows the calculated STEM probe intensity incident on the sample. In Figure 2.7(c), the intensities of the before and after aberration correction probes have been normalised to the same total current. After aberration correction, clearly the probe not only has a smaller diameter, but has a sharper, more intense peak. Before aberration correction, significant current is ''wasted'' in the probe tails, so after correction we might hope for not only better resolution, but also improved contrast.

The probe then interacts with the sample. Equation (2.2) reveals that the probe is made up of a sum of waves at different angles. If only a single angle were used, for a crystalline sample we might expect a diffraction pattern to be formed in the far field. Because of the range of angles, each diffraction spot is broadened into a

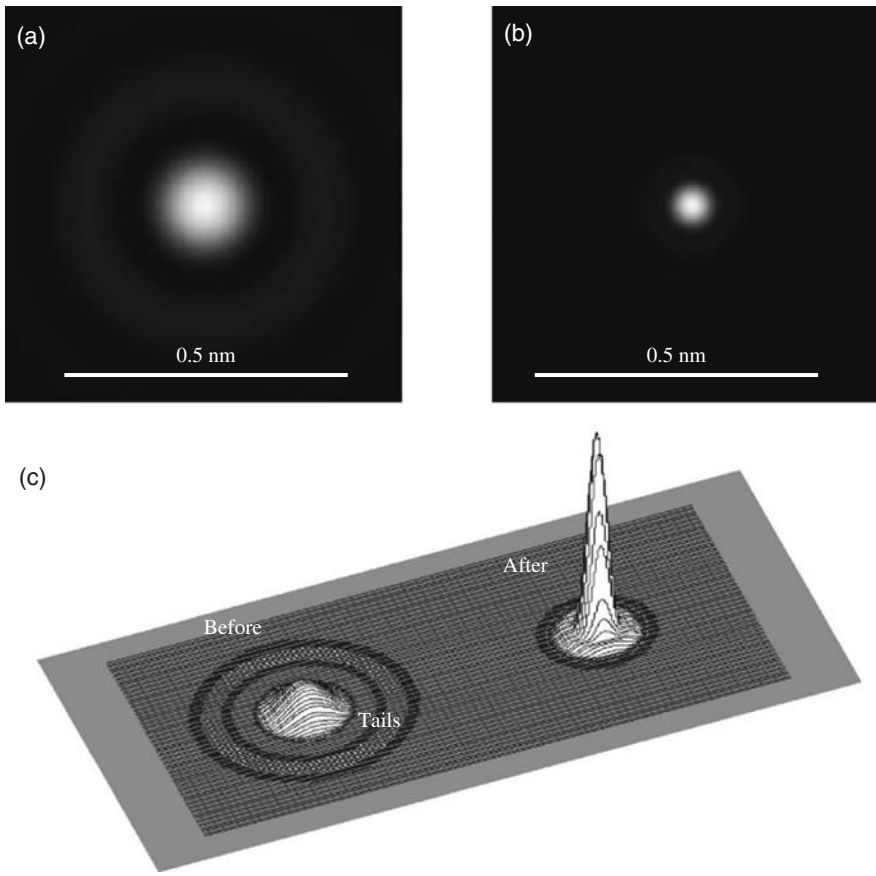


Figure 2.7 Calculated probes due to purely geometric aberrations for typical operating conditions of the ORNL VG microscopes HB603. (a) Before correction: Aperture 10 mrad, $C_3 = 1.3$ mm, Defocus -59 nm, FWHM $\approx$ 1.2 Å. (b) After aberration correction: Aperture 22 mrad, $C_5 = 100$ mm, $C_3 = -40$ μm, Defocus 5 nm, FWHM $\approx$ 0.55 Å. (c) 3D-representation of the probes with height representing intensity, normalised to the same integrated intensity in each.

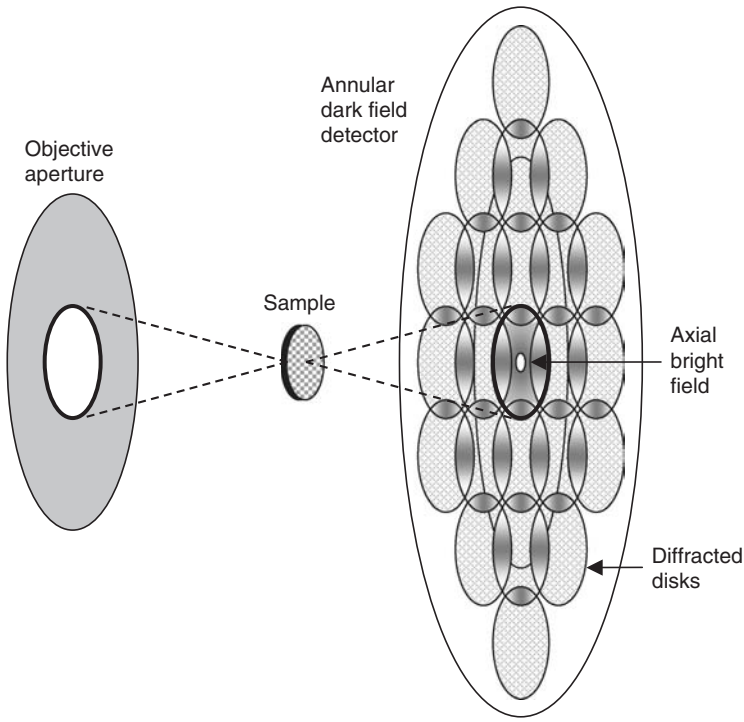


Figure 2.8 Schematic of the formation of STEM image contrast for an aligned crystal. The range of angles incident on the sample causes the diffraction pattern to become broadened into an array of overlapping disks. The detectors record contrast generated by this overlap.

disk of the same angular size as the objective aperture (Figure 2.8). The amplitude in the far field is given by the convolution of the objective aperture and the diffraction pattern of the sample $\tilde{V}(\mathbf{K}_i)$. Mathematically, this amplitude is:[45]

$$M(\mathbf{K}, \mathbf{R}_0) = \int H(\mathbf{K}_i) \mathrm{e}^{2\pi i \chi(\mathbf{K}_i)/\lambda} \mathrm{e}^{-2\pi i \mathbf{K}_i . (\mathbf{R}_0)} \tilde{V}(\mathbf{K} - \mathbf{K}_i) \, \mathrm{d}\mathbf{K}_i \qquad (2.3)$$

where $\mathbf{K}_\mathrm{i}$ is used for the incident angles and $\mathbf{K}$ for the angles in the far field. This pattern is known as the electron Ronchigram, after V. Ronchi who studied the optical equivalent extensively.[43] Some of these disks overlap and interfere, both constructively and destructively. The detector integrates over a range of this interference and the variation of the integrated intensity as a function of probe position $\mathbf{R}_0$ gives an image. Although this model is a simplification, it gives some valuable insights.

In the phase-object approximation for coherent imaging, the sample only changes the phase of the beam, which also depends on the aberrations. The image contrast will depend on these phases and is therefore known as a phase-contrast image (a useful summary is given by Spence[44]). As we record the

intensity collected by the detector, which is given by the interference of the diffracted disks, the image is not generally a simple projection of the sample. For an infinitesimally thin sample, and point-sized (delta function) axial detector, the recorded image intensity reduces to:

$$I_{\mathrm{BF}}(\mathbf{R}_0) = \left| \int V(\mathbf{R}) P(\mathbf{R} - \mathbf{R}_0) \mathrm{d}\mathbf{R} \right|^2 \tag{2.4}$$

where $V(\mathbf{R})$ represents the sample in real space. Thus the bright-field image is the intensity of the convolution of a complex object function with a complex probe function. In order to understand what this means for the image, we consider the weak phase object approximation (WPOA).

In the WPOA, we further assume that the diffracted disks are much weaker than the central undiffracted disk and that interference between diffracted disks is unimportant. After Fourier transforming eqn (2.4), these assumptions eventually lead us to the idea of the phase-contrast transfer function (PCTF). We can express the image contrast as a function of spatial frequency ρ (for ρ within the objective aperture) as:

$$I_{\mathrm{BF}}(\rho) \propto \partial(\rho) + 2\sigma \sin\left(2\frac{\pi}{\lambda}\chi\left(\frac{\rho}{\lambda}\right)\right)\tilde{V}(\rho) \tag{2.5}$$

where χ is the aberration function defined earlier, σ is the interaction constant,[44] and the factor λ converts ρ to an angle. This function oscillates (examples are shown in Figure 2.9 and by Hawkes and Kasper[29]), which means that different spatial frequencies contribute to the image in very different ways. Note that only disks that overlap can contribute to the image; the size of the objective aperture therefore defines the maximum possible resolution. The point resolution is defined as the first zero of eqn (2.5); after this frequency the contrast reverses and it becomes difficult to interpret an image. An advantage of aberration correction is that it allows us to increase the spatial frequency at which this first zero occurs.

In the approximations used here, a bright-field detector records the interference between the zero order, undiffracted disk and diffracted disks. An on-axis bright-field image therefore has a lower potential resolution than an off-axis or annular detector. A small detector can admit only a very small (and therefore noisy) signal, while a larger detector will have the potential to collect more signal, but will necessarily integrate over more of the intersections between disks. The phase changes across each disk (due primarily to aberrations) so the integrated intensity will lose some of the fine variation.[45] The integrated signal is damped by an amount that depends on the gradient of the aberration function $\nabla\chi$ and the collection angle θ_{c}. By approximating the aperture as a Gaussian, (then θ_{c} is 0.77 times the real convergence angle for a better fit[46]) in the same way that the damping is derived for conventional TEM,[47] we arrive at:

$$D_{\mathrm{S}}(\rho) = \exp\left(-\theta_{\mathrm{c}}\left|\nabla\chi\left(\frac{\rho}{\lambda}\right)\right|^2\right) \tag{2.6}$$

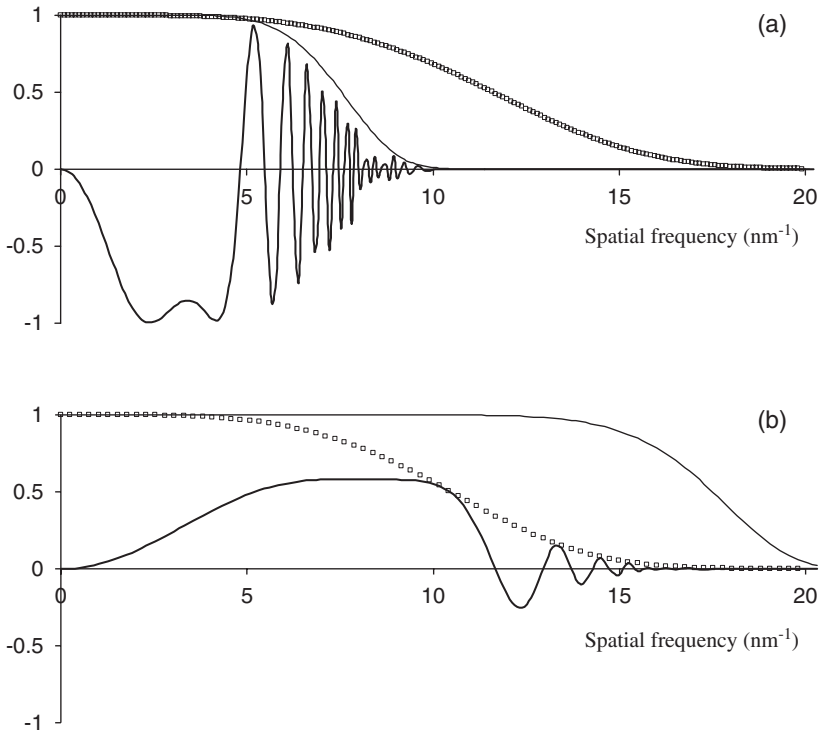


Figure 2.9 Bright-field Contrast Transfer Functions (CTFs) for the ORNL VG Microscopes' HB603. The thick line is the CTF, multiplied by the damping envelopes: spatial (thin line) and chromatic (squares). (a) before C_3 correction at the Scherzer condition, $C_3 = 1.3$ mm, Defocus -59 nm, $C_C = 1.3$ mm, stabilities 1 ppm, collection angle 0.2 mrad. (b) After C_3 correction, $C_5 = 100$ mm, $C_3 = -40$ μm, Defocus 5 nm, $C_C = 1.6$ mm, stabilities 1 ppm, collection angle 0.5 mrad.

This result is significant for aberration-corrected systems where it leads to the prospect of using larger collection angles because the gradient of the aberration function is smaller. Note that in Figure 2.9, before aberration correction, the significant damping is due to the spatial coherence, while after aberration correction, even with a larger aperture, the most significant damping is due to temporal coherence (essentially chromatic aberration). Bright-field STEM will therefore be more efficient after aberration correction and an increase by a factor of ten in angle would increase the collection area by a factor of 100.

Chromatic aberration will also introduce a phase change. Again, the assumption is that energy spread is Gaussian in form and integration over the range of energy spread leads[48,49] to:

$$D_E(\rho) = \exp\left(-\frac{1}{2}\pi^2\lambda^2|\Delta f|^2\rho^4\right) \tag{2.7}$$

where Δf is the Gaussian standard deviation of the range of focus. If the energy spread of the beam were the only contribution, then we would have $\Delta f = C_c \Delta E / E_0$, where ΔE describes the energy spread and E_0 is the beam energy. Normally other contributions, such as instabilities, are assumed to add in quadrature.[44]

These two damping envelopes multiply the PCTF and ultimately lead to an information limit beyond which it is not possible to extract any more information from the image. Note that eqn (2.7) only applies to axial bright-field imaging. For tilted bright-field or for ADF imaging, the damping envelope due to energy spread (limited temporal coherence) is less severe. Nellist and Pennycook[50] have given an alternative form for ADF imaging and shown that there the damping does not restrict the highest resolution, but reduces contrast in mid-range spatial frequencies. A simple way to consider this result is that at the limit of resolution in ADF we detect interference between the outer edges of disks, which have received the same phase change from the (rotationally symmetric) chromatic aberration. However, in axial bright-field imaging, we detect interference between the centre of the zero-order disk and the edges of diffracted disks, which have different phase changes.

The alternative limit to the perfectly coherent image is an incoherent image, which is a good approximation for ADF imaging. In the perfectly incoherent case, the image intensity is given as a convolution between a real object function $O(\mathbf{R})$ and the probe intensity:

$$I_{\mathrm{ADF}}(\mathbf{R}_0) = \int O(\mathbf{R}) |P(\mathbf{R} - \mathbf{R}_0)|^2 \mathrm{d}\mathbf{R} \tag{2.8}$$

Thus, using the convolution theorem, the Contrast Transfer Function (CTF) for incoherent imaging is given in this approximation by:

$$I_{\mathrm{ADF}}(\rho) = \tilde{O}(\rho) T(\rho) \tag{2.9}$$

where $T(\rho)$ is the Fourier transform of the probe intensity and $\tilde{O}(\rho)$ the Fourier transform of the object function. Since the probe intensity is real and even, the resulting CTF for incoherent imaging is simple and for a thin sample will have the general form shown in Figure 2.10.

In Figure 2.9, we show the CTFs for bright- and dark-field STEM imaging before and after C_3 correction under the above assumptions and for C_3 corrected Z-contrast imaging in Figure 2.10. We can immediately see one reason why Z-contrast STEM is simpler to interpret than bright-field STEM; there are no contrast reversals. After aberration correction, both situations are improved. Phase-contrast imaging is by definition sensitive to the phase, which in turn is sensitive to even very small structural changes. The problem is that a small change in either specimen thickness or defocus can give contrast reversals; we cannot say whether atomic columns for an unknown sample will be light or dark without additional information. If one does not know the specimen structure, thickness, or defocus, from some other means, bright-field images are hard to interpret. Incoherent Z-contrast images do not undergo contrast reversals with changing sample thickness or defocus over a *reasonable*

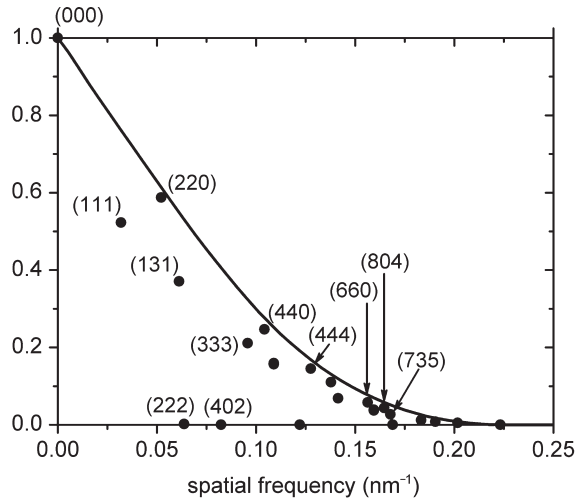


Figure 2.10 An example Z-contrast CTF for the ORNL VG Microscopes' HB603 after aberration correction. The solid line is the Fourier transform of the probe intensity. The circles are intensities of spots in the Fourier transform of a simulated image of a Si [112] sample 1 nm thick.

range and thus offer advantages for the examination of unknown specimen structures.

We have noted that it is the detector size that determines whether an image is best described as coherent or incoherent. A useful guide is the van Cittert–Zernicke theorem, which, when applied to STEM, tells us that the coherence width on the sample is the Fourier transform of the detector. A small detector gives a more coherent image; a larger detector gives a more incoherent one. Nellist and Pennycook[51] have shown how incoherent imaging can arise from purely elastic scattering with a large detector. In practice no image fits perfectly into either category and no real detector is infinitely small or large. However, in STEM it is remarkable that even though both the bright-field and Z-contrast images are formed using the same probe, we can obtain such different images.

An example is given in Figure 2.11, where we show simultaneously acquired ADF and bright-field images of a gold film on activated carbon catalyst. The activated carbon has a high surface area and forms a semi-amorphous structure. Because the Au atoms are heavy, even single atoms are easily detectable in the ADF image. However, in the bright-field image, the structure of the activated carbon support is revealed more clearly, but the single atoms are almost impossible to locate on the rough support.

2.2.5 Measuring Aberrations

In practice, to operate an aberration corrector on a real sample, it is necessary to measure the aberrations. Several techniques exist to automatically measure

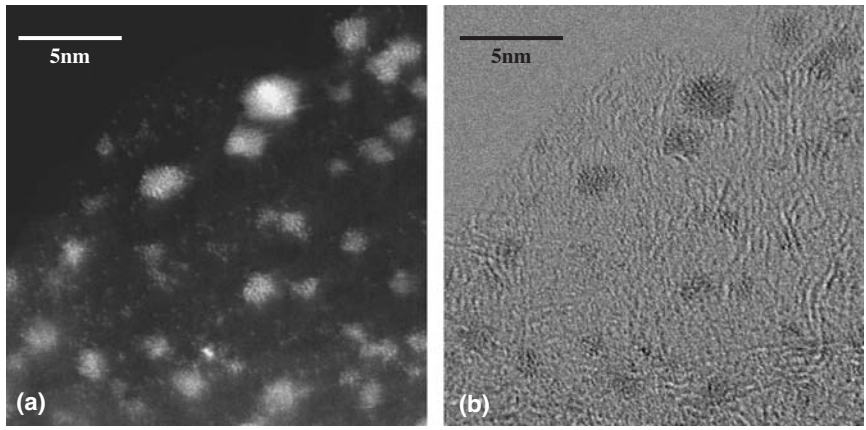


Figure 2.11 Nanoparticle gold on activated carbon catalyst sample. (Left) Z-contrast image in which single Au atoms can be seen. (Right) A simultaneously acquired bright-field image in which the single Au atoms cannot be easily located, but some of the structure of the carbon support can be seen. Scale bar is approximate. Data reproduced from Veith et al., Catal. Today, in press, 2007.

the aberration function. One can observe changes in the bright-field images as the detector is moved. By the principle of reciprocity, these methods are equivalent to tilting the illumination angle in TEM. The image will shift, and the apparent aberrations will change.[52] To understand this idea, consider expanding the series describing the aberration function of eqn (2.1) about a different point, including all nonround terms. The shift depends on the gradient of the aberration function. The measured defocus and astigmatism, which for an amorphous material can be determined from the diffractogram,[53] will depend on the second derivatives of the aberration function. Once these values are known, it is a fairly simple matter to fit a suitable aberration function.

Another approach relies on the Ronchigram. For a crystalline sample,[54] the overlap regions contain fringes that depend both on the sample and the aberrations and so these patterns can be used.[55,56] For an amorphous sample, the Ronchigram is a shadow image of the sample in which the magnification depends on the gradient of the aberration function. Local distortions depend on the second derivative of the aberration function. Therefore, determining the magnification in the Ronchigram of an amorphous material again allows the aberration function to be determined. A convenient way to determine the magnification is to move the sample by predetermined amounts and measure the apparent image shifts at several different angles within the Ronchigram.[30,31]

2.2.6 Phonons

At this point, it should be remembered that the simple picture given in Figure 2.8 only applies to elastic scattering. An important contribution to

the image is given by inelastic scattering. As the electrons pass through the sample, they interact with phonons. The details of this interaction can be quite complicated because multiple phonons are involved and may be intrinsically nonlocal. The thermal motion of the atoms causes them to be displaced slightly from their "ideal" positions, which causes the scattering to be diffuse and so it is known as thermal diffuse scattering (TDS). TDS is extremely important for Z-contrast imaging because almost all of the electrons at high angles have been thermally scattered. The TDS is responsible for breaking the coherence in the z-direction.[57] The nearly random time-varying displacements mean that the Z-contrast image intensity does not oscillate with depth in the same way as a coherent bright-field image.

2.2.7 Resolution

The primary aim of aberration correction has been to improve the resolution of the electron microscope. However, resolution can be difficult to define.[58–60] As described above, there are defined point and information resolution limits, an absolute cutoff imposed by the aperture and the practical limit of signal-to-noise, which in principle can all be different. A good general indicator of incoherent resolution when diffraction is the primary limitation is given by the Rayleigh criterion (R_r), which for a round aperture is:[28]

$$R_r = 0.61\lambda/\theta \tag{2.10}$$

Clearly in order to improve the resolution it is necessary to increase the aperture size θ. Equivalently, the uncertainty principle shows that better localisation of the probe requires larger transverse momentum. For our 300 kV STEM, before aberration correction the Rayleigh resolution with a 10 mrad aperture was about 1.2 Å. The aperture was limited to this size by spherical aberration. After aberration correction, the aperture could be increased to about 23 mrad giving a Rayleigh resolution limit of close to 0.5 Å. However, in practice, other factors (noise and instabilities) provide the effective resolution limit.

Figure 2.12 shows an example image[61] from the aberration-corrected VG Microscopes' 603. The sample is a silicon crystal, viewed down the [112] zone axis. The "dumbbells" of atoms have a projected spacing of only 0.78 Å. Taking the Fourier transform (diffractogram) reveals the spatial frequencies that are present in the image. In an incoherent image, this provides a good indication of the resolution. Note that in a coherent image, one sees sum and difference spots, so this technique should not be used for coherent bright-field images. The image in Figure 2.12 has information transfer detectable above the noise to 0.63 Å. However, one needs to take some care when examining the diffractogram of an image. The presence of a spatial frequency in the diffractogram is a necessary, but not sufficient condition for proving that such a resolution has been achieved. The image has been amplified and digitized and any "clipping" of the signal by incorrect setting of any amplifier levels or gain, or other nonlinearity in this process will introduce extra spatial

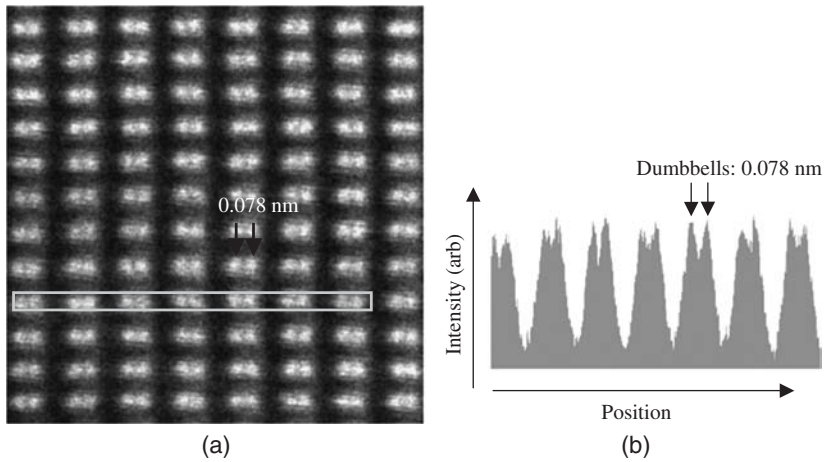


Figure 2.12 (a) Z-contrast STEM image of Si viewed down the [112] axis. Image has been smoothed to print and a slight scan distortion removed. (b) A line trace through a typical row shows that the Si dumbbells (0.78 Å spacing) are resolved. Adapted from reference 6.

frequencies. Instabilities during the acquisition process will also introduce spots or streaks in the diffractogram that can be misinterpreted. In practice these spots are a useful diagnostic; because the image is scanned, their position in the diffractogram relates to the frequency of the instability.

2.2.8 Three-Dimensional Microscopy

One remarkable consequence of aberration correction is that the larger objective aperture size means that the depth of field also decreases. Recalling that the Rayleigh criterion for resolution is the distance from the centre to the first zero of the point response function suggests that we can define the equivalent of the Rayleigh criteria in the vertical direction[62,63] (Figure 2.13). The vertical resolution is then given by:

$$R_z = 2\lambda/\theta^2 \tag{2.11}$$

which is twice the depth of field[28] (FW80%M). Note that this result applies for incoherent imaging; if the image is coherent, then whether two point-like objects can be resolved will depend on their relative phases. Note that in Figure 2.13, the vertical and horizontal axes have different scales. At present for our aberration-corrected 300 kV STEM, the vertical resolution (around 8 nm[62]) is significantly worse than the transverse resolution (below 0.8 Å[61]).

However, the vertical resolution improves faster than the transverse resolution with increasing aperture size. This result is significant because the depth resolution is moving from a regime where it is bigger than many interesting nano-objects to one in which it is just smaller. Note, that for a sample that is thicker

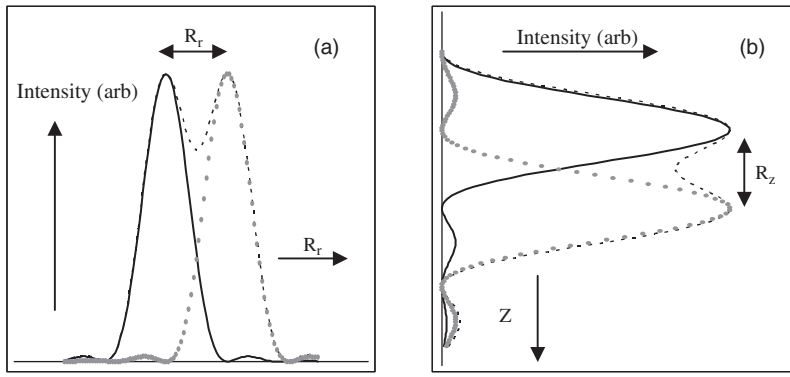


Figure 2.13 Intensity on axis in transverse (a) and vertical (b) directions for a probe limited by diffraction only, demonstrating the Rayleigh criteria for incoherent resolution from a round aperture. Two points of equal intensity are just resolved when the second point response function (grey dots) is centred on the first zero of the first point response function (black line). The sum is shown as a dashed line. Note that the linear axes scales and the point-response shapes are different in the two directions. Adapted from references 62 and 63.

than the depth of focus, a STEM image cannot be regarded as a projection. A through-focal series now becomes a through-depth series and changing focus makes it possible to obtain information from different depths within a sample. It has been proposed that it is possible to depth section through a sample and reconstruct a 3-dimensional model of the sample from the resulting dataset. A useful comparison is to confocal optical microscopy,[64,65] although we do not have a pinhole aperture in a conjugate focal plane, so it is closer to a 3-dimensional wide-field microscope than a true ''confocal'' microscope.

van Benthem and coworkers[63] have shown that for the case of Hf atoms in a $Si\text{-}SiO_2\text{-}HfO_2$ structure, individual Hf atoms can be found within the silica layer, Figure 2.14. This example shows the benefit of a three-dimensional technique; a single two-dimensional image might not prove that the Hf atoms are in the bulk; conceivably they might be on the surface of the sample, perhaps as a consequence of sample preparation. However, the three-dimensional technique clearly shows that the Hf atoms are located inside the volume of the silica layer and provides an indication of the distribution in all three dimensions. This distribution depends on the annealing treatment of the sample and will have consequences for the dielectric properties of the device.[63]

One remarkable aspect of this result is that it demonstrates that the precision with which the Hf atoms can be located is not the same as the apparent vertical resolution. Even though the thickness of the sample ($\sim$6 nm) was similar to the depth resolution, the Hf atoms could be located inside the sample. The principle is simple; given sufficient slices and signal-to-noise ratio, it is possible to determine the location of the maximum of a smooth curve with a precision that is far better than the resolution (or FWHM).

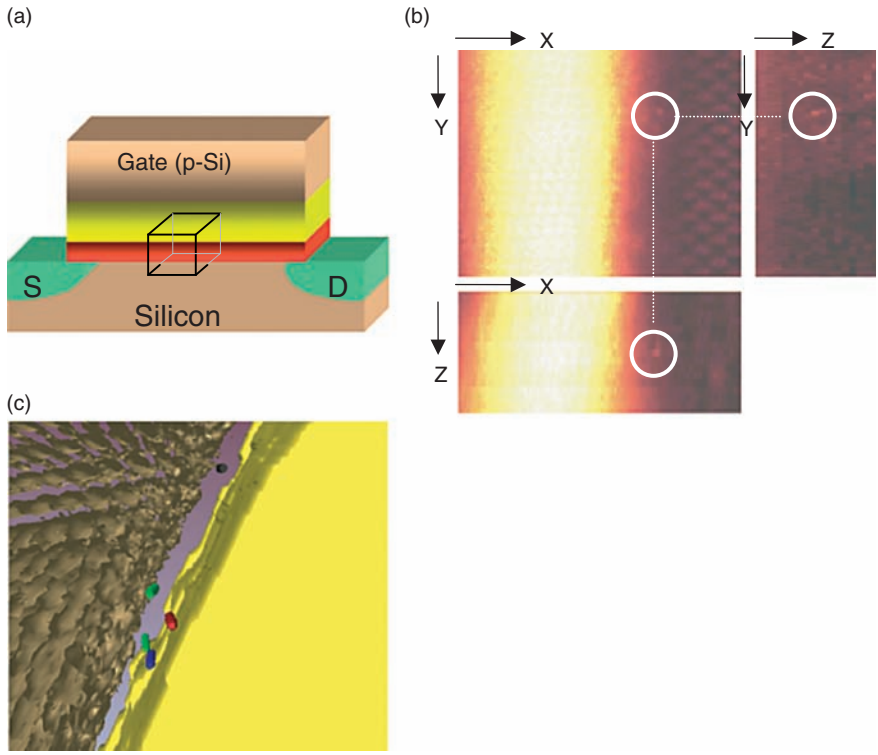


Figure 2.14 3-dimensional location of single Hf atoms. (a) A schematic view of an alternative gate dielectric stack with a SiO_2 layer shown in red and the alternative dielectric, HfO_2, shown in yellow. (b) The acquired three-dimensional ADF image stack is represented by slice views. One isolated single Hf atom is marked by the white circles. This representation demonstrates that the Hf atom is located *inside* the TEM sample. (c) A 3D reconstruction of part of the dataset showing single Hf atoms. Reproduced from reference 63.

2.2.9 Channeling

The above resolution criterion only holds for well-separated point-like objects in a lighter matrix (which is ideal for Hf atoms in SiO_2). For extended or closely spaced objects it will not be possible to obtain this resolution. Furthermore, in an aligned crystal, channelling,[16] which is the dynamical interaction of the incident electron beam with the aligned columns of atoms, makes such results more difficult to interpret.

A simple model for Z-contrast imaging would be to assume that the image can be represented as the 2-dimensional convolution of eqn (2.8) above. Taking the 3-dimensional probe shape into account would suggest that the convolution should be performed in 3-dimensional space. This model would require that a local model for the scattering cross section can be used and that all of the

electrons that are scattered to high angles reach the detector. Neither of these assumptions is strictly accurate, but again they give a useful approximation. Figure 2.15 illustrates how the depth sensitivity in free space or in aligned crystals varies with the aperture size. For a 10 mrad aperture (before correction), the depth of field is quite large; we have to move the focus a long way to see much change in the probe. However, when the aperture size is increased to 22 mrad, the depth of focus is dramatically reduced.

For simplicity, this calculation uses all aberrations set to zero at 300 kV for Si [110], with no phonons included. However, as the intensity within the crystal is not what is observed, some care needs to be taken when interpreting such images. It is also apparent that the probe shape changes depending on the coordinates on the sample and a convolution model would have to include different probe shapes at different positions. Figure 2.15 reveals that the channeling makes the probe sharper than in free space and it is pulled deeper into the crystal.[16,66] However, the sharpening of the intensity distribution in real space (**R**) does not correspond to improved resolution. Resolution in a STEM image depends on the intensity recorded as a function of probe position ($\mathbf{R}_0$), so we cannot use channeling to increase the resolution although it can modify the signal from a single atom.

Figure 2.15 also shows that although channeling makes it harder to obtain depth resolution inside an aligned crystal, it does not preclude it altogether. In fact, increasing the convergence angle further may allow single dopant atoms to be located even within an aligned crystal.[67] The larger aperture also means that many of the electrons are travelling at higher angles to the axis and so higher-order Bloch states become more important.[68]

2.3 Applications to Nanostructure Characterisation in Catalysis

One area in which nanotechnology already affects everyday life is in catalysis. Catalysts are used at some stage in almost every modern manufacturing process. A prerequisite to the non-empirical design and refinement of improved catalysts is the identification of the atomic-scale structure and properties of the catalytically active sites.

2.3.1 Anomalous Pt–Pt Distances in Pt/alumina Catalytic Systems

The Pt/γ-alumina system is representative of many common heterogeneous catalysts, which consist of transition metals dispersed on a high surface area support. This system is often used for catalytic reduction and oxidation of automotive pollutants.

Even before aberration correction, application of Z-contrast STEM to catalytic systems had revealed the presence of intriguing atomic-scale

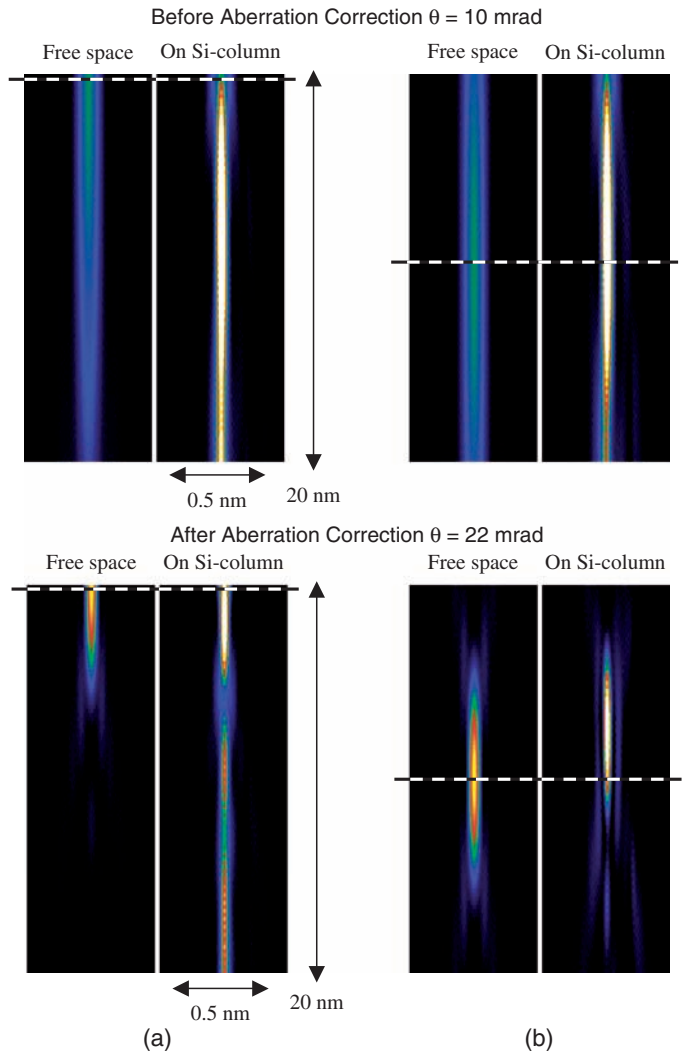


Figure 2.15 Comparisons of probe intensity in free space and in a Si [110] crystal, illustrating the focal depth and channelling with different convergence angles and defocus values. The plane at which the probe is set to be focused (in free space) is indicated by a dashed line. The channeling by the crystal changes the probe intensity and the location of the maximum. All other aberrations set to zero, 300 kV, with no phonons. (a) Focused at the surface. (b) Focused 9.5 nm into crystal. Scales are approximate.

structures.[69] After aberration correction, we can see quite clearly that in addition to isolated Pt atoms, trimers of Pt are observed with preferred shapes. The benefit of aberration correction is that the enhanced resolution and signal-to-noise allowed this observation to be verified and the Pt–Pt distances to be

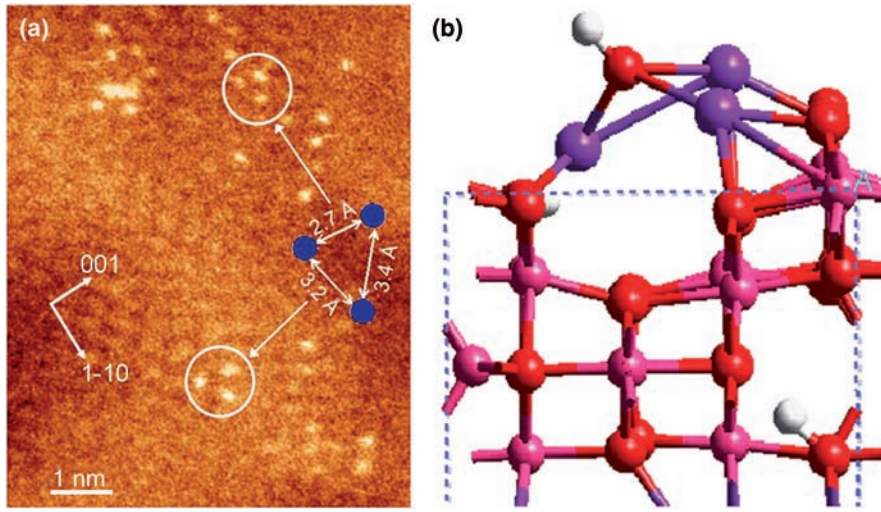


Figure 2.16 (a) Z-contrast STEM image of Pt on the surface of γ-Al_2O_3 close to (110) orientation. Two Pt_3 trimer structures are circled; inset – measured interatomic distances. A hint of the alumina lattice is also visible. (b) Schematic of the configuration for the Pt_3OH unit on the (110) surface of γ-Al_2O_3, determined by first-principles calculations, with closely matching interatomic distances. Data reproduced from reference 70.

measured (Figure 2.16). It is apparent that the three Pt atoms do not form an equilateral triangle with regular Pt–Pt distances.

The explanation for the distorted shape comes through density-functional total-energy calculations.[70] Placing bare Pt trimers on a γ-Al_2O_3 $<110>$ surface and relaxing the structure to equilibrium results in an almost equilateral triangle with bond lengths of 2.59, 2.65, and 2.73 Å, close to the interatomic spacings in metallic Pt. The longer bonds found experimentally can be explained by adding an OH group to the top of the trimer: Two of the bonds lengthen to 3.1 and 3.6 Å, in better agreement with the experimentally determined values (3.2 and 3.4 Å). The addition of the OH group also changes the electron density on the Pt atoms. There is a clear indication of depletion of electron density from the Pt–Pt bonds, which may explain the catalytic activity.[70] Although directly detecting OH groups on a single Pt trimer on a thick support would be extremely difficult, this example demonstrates that the enhanced sensitivity due to aberration correction provides vital information about the structure. Better images from new instruments will provide a wealth of information from many more systems.

2.3.2 La Stabilisation of Catalytic Supports

γ-alumina is one of many polytypes of Al_2O_3 that are used extensively as catalytic supports. A porous form of γ-alumina is used extensively because it

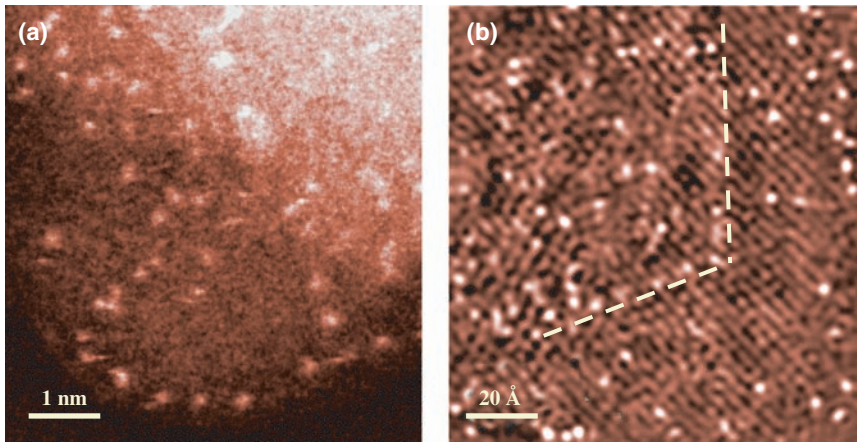


Figure 2.17 Z-contrast image of a sample of La-doped γ-Al_2O_3. Bright spots correspond to La atoms. (a) An off-axis image suggests that the La atoms are clustered at edges. (b) Close to the (110) axis, the support structure is revealed and the La atoms appear to be clustered on step edges. Data reproduced from reference 71.

has a large specific surface area. However, at temperatures in the range 1000–1200°C, γ-alumina transforms rapidly into the thermodynamically stable α-alumina phase (corundum); the pores close and catalytic activity is degraded. A small amount of La doping increases the temperature at which this transition occurs. Many explanations have been proposed for this change, but until recently it was not clear which was correct. A combination of STEM, EXAFS and first-principles density-functional calculations has been able to locate the La and explain this remarkable result.[71]

A demonstration that the La is present as single atoms was provided by Z-contrast imaging. In Figure 2.17 single La atoms are seen superposed on the crystal lattice planes formed by the substrate. Furthermore, this image shows that they are found most often on step edges, which suggests that the La atoms are on the surfaces. Three-dimensional microscopy provided an essential confirmation: By changing the focus of the beam through the sample as shown in Figure 2.18, it was found that the La-atom intensities peak at only two planes (the top and bottom surfaces); they are not evenly distributed throughout the bulk. Experimentally, the La atoms are found at the surfaces of the sample.

Theory independently suggests that La atoms favour the surfaces over bulk sites by substantial energy differences ($\sim$4 eV), with several possible configurations, having comparable energies. Figure 2.19 shows a schematic indicating that a La atom on a (100) surface induces a significant local reconstruction event. In addition, theory has demonstrated that it is not energetically favourable for La atoms to cluster on alumina surfaces; they remain as single adatoms. La atoms have a much larger binding energy on γ-alumina (7.5–9 eV) than on α-alumina (4–5 eV), which is the final product of the undesirable phase transformation. Thus, the extra 3–5 eV binding energy

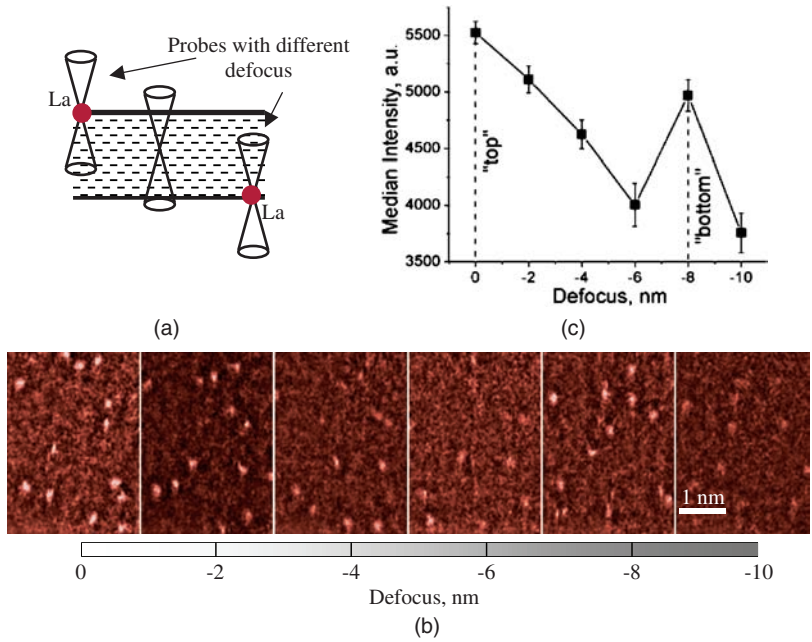


Figure 2.18 Z-contrast images of a La-doped γ-Al_2O_3 sample obtained by changing the focus of the beam from the top surface of the sample to the bottom surface. (a) Illustration of the depth-sectioning technique. By changing the focus, we can focus the probe to different depths within the sample. (b) The through focus series of Z-contrast images. (c) The La atom intensities peak at the top and bottom surfaces. Data reproduced from reference 71.

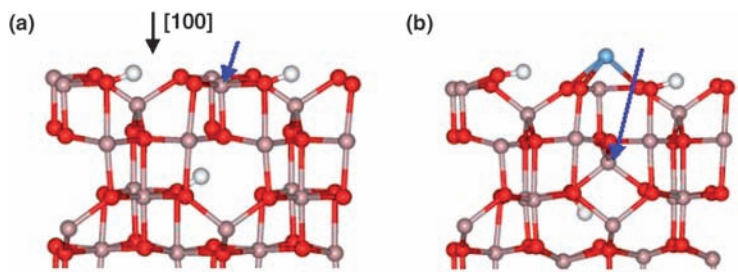


Figure 2.19 Schematics of the configurations for the (100) surface of γ-Al_2O_3, determined by first-principles calculations: (a) undoped, and (b) La-doped. The Al, O, H, and La atoms are shown in grey, red, white, and blue, respectively. When the La atoms is present on the surface, a surface Al atom relaxes from the surface into a cation vacancy as indicated by the blue arrow. The La atom occupies a site close to the initial location of the Al atom and binds to the nearest three oxygen atoms. Data reproduced from reference 71.

required per La atom is the driving force for the higher transition temperature. In simple terms, the La atoms prevent the phase change effectively by stabilising the surface of the γ-alumina phase.[71]

2.3.3 CO Oxidation by Supported Noble-Metal Nanoparticles

One of the fascinating things about the nanoworld is that properties change at very small length scales. For example, gold (Au) in bulk form is not a very active catalyst but when prepared as supported nanoparticles it becomes one of the most active catalysts for several reactions including CO oxidation.[72] Low-temperature CO oxidation has important practical applications and it is also interesting as a model reaction. Furthermore, the enhanced activity of nanosize Au is exciting because it contrasts with most other noble metals for this reaction. As a specific example, supported platinum (Pt) nanoparticles[73] are less active than low-index Pt surfaces. There have been a wide variety of explanations for the high activity of small Au nanoparticles.

First, it is obvious that the smaller that Au nanopaticles become, the number of particles per gram of Au will increase (scaling roughly as $1/d^3$ for diameter d). The total surface area per gram will scale roughly as $1/d$, while the perimeter (where a small reactant molecule can interact with both the support and the nanoparticle) will scale as $1/d^2$. Most authors therefore give a turnover frequency (a rate normalised by the number of sites) to help distinguish more exotic effects from this kind of simple scaling.

Haruta and coworkers[74] have proposed that the perimeter must be important. In apparent contradiction of this perimeter model, Goodman and coworkers[75] have claimed that a bilayer structure, with effectively zero perimeter, is the most active structure and that negatively charged Au plays an important role. Landman and coworkers[76] also found evidence to support the idea that the catalytic activity is related to negatively charged Au on MgO. However, Guzman and Gates[77] have shown evidence that suggests the active site involves cationic Au. Norksov and coworkers[78] have demonstrated that in calculations coordination number has the strongest effect on the activity. Changes in coordination number seem to produce rather larger changes in the calculated reaction barriers than other factors (such as charge). In order to clarify this situation, we have examined an array of nanoparticles, prepared using both chemical[79] and physical[80] techniques and performed calculations on an ensemble of nanoparticles.

We constructed[83] an ensemble of Au nanoparticles on TiO_2 substrates and optimised their geometries using density-functional theory. We found that at least one substrate defect is needed to anchor Au particles and previous work has shown that large Au particles are likely anchored by many O vacancies.[81] To compare to other noble metals, in a subset of our ensemble of nanoparticles, we replaced all Au atoms with Pt atoms and reoptimised geometries (Figure 2.20). The adsorption of O_2 and CO molecules on both types of nanoparticles was then optimised at various sites and the binding energies (or desorption energies) and reaction barriers were calculated. On virtually all

Au particles, O_2 adsorbs as a molecule. Adjacent O_2 and CO molecules then react to produce CO_2, which desorbs, leaving a bonded O atom. On Pt surfaces and nanoparticles, O_2 adsorbs strongly with a stretched O–O bond that dissociates easily, allowing adjacent O atoms and CO molecules to react.

Figure 2.20 reveals a crucial difference between model Au and Pt nanoparticles. Upon adsorption, the neighbouring Au–Au bonds are weakened, making the rotation and stretching needed for the CO + O_2 reaction less energetically costly (Figure 2.21). Such weakening (akin to the structural fluxionality noted by Landman and coworkers[76]) does not occur at the corresponding Pt–Pt bonds. The Pt nanoparticles are more rigid, resulting in higher reaction barriers because it is more difficult method for the strongly adsorbed molecules to move and interact.

In Figure 2.21, we show reaction barriers (E_r) obtained by the nudged elastic band method for CO oxidation and O_2 desorption energies (E_d). CO desorption energies were generally found to be larger and are so are not shown. The most striking result for Au nanoparticles is the crossing of the E_r and E_d curves as a function of the average coordination, n, of the two Au atoms to which O_2 is bonded. Catalysis is favourable over desorption only at sites with average coordination $n < 5$, when $E_r < E_d$. At perimeter sites, O_2 molecules bind more

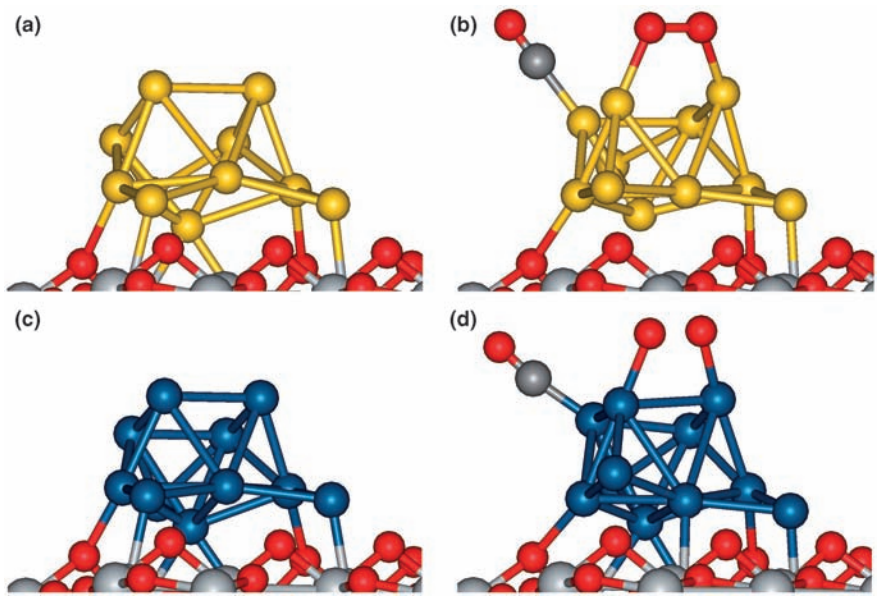


Figure 2.20 Adsorption of O_2 and CO molecules on TiO_2-supported Au (a,b) and Pt (c,d) nanoparticles. Ti is shown in light grey, O in red, Au in yellow, Pt in blue, and C in dark grey. The relaxed configuration of 11-atom supported Au nanoparticle before and after adsorption is shown in (a) and (b). The O–O distance increases to 1.39 Å here. (c,d) The corresponding figures for Pt nanoparticles. The O–O distance is 1.48 Å. Data reproduced from reference 83.

strongly and catalytic activity is favoured even with $n = 6$ or 7 because the bridge bond to the substrate helps weaken the O–O bond. Reaction barriers on the perimeter also decrease with decreasing coordination.

The behaviour of Pt nanoparticles is distinctly different. Figure 2.21 shows that we always have $E_r < E_d$. In principle, catalysis is always favourable over desorption (for large nanoparticles, our E_r values approach the known value at Pt surfaces $\sim$0.8 eV), but the absolute values become larger with decreasing coordination. In addition, perimeter sites for Pt particles were not catalytically active in the simulation as O_2 molecules preferred other nanoparticle sites. The

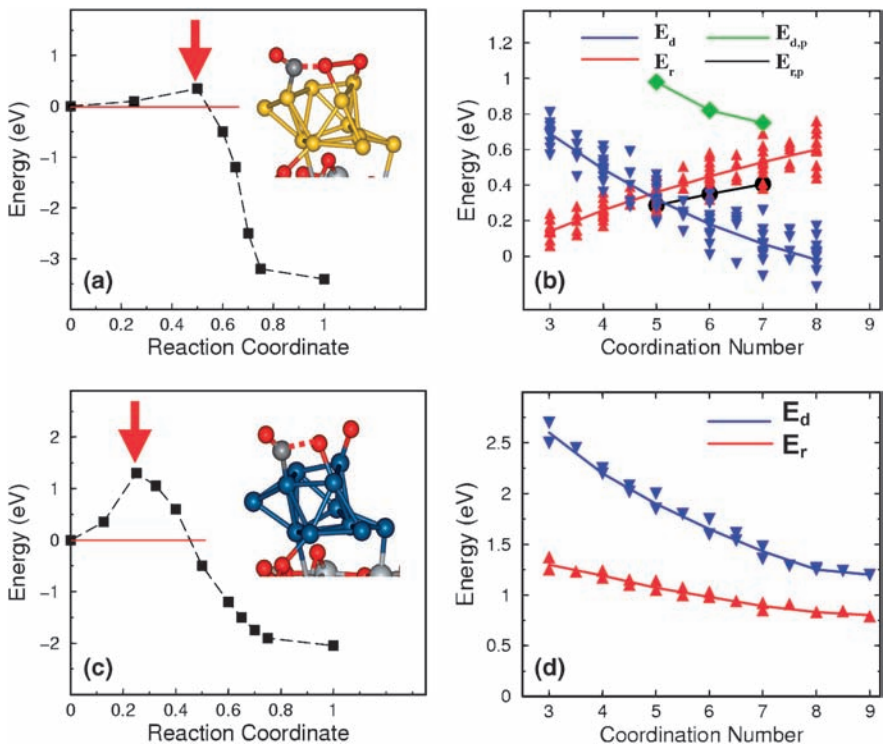


Figure 2.21 Reaction of O_2 and CO molecules on TiO_2-supported Au (a,b) and Pt (c,d) nanoparticles. (a,c) The reaction energy profiles and schematics of the Au and Pt nanoparticles at the transition state (arrowed). (b,d) The desorption energy, E_d, of an O_2 molecule and the reaction barrier, E_r, as a function of the average coordination number n of the two Au (or Pt) atoms to which the O_2 molecule is attached. Points correspond to different adsorption sites and/or different nanoparticles located over oxygen vacancies in rutile (110) or anatase (101) surfaces. The average curves of E_d and E_r (bold solid lines) are shown. The lines $E_{d,p}$ and $E_{r,p}$ correspond to the desorption energy and the reaction barrier for the bridge-bond perimeter sites of Au clusters. The Pt–Pt bonds are more rigid, resulting in a higher reaction barrier. Data reproduced from reference 83.

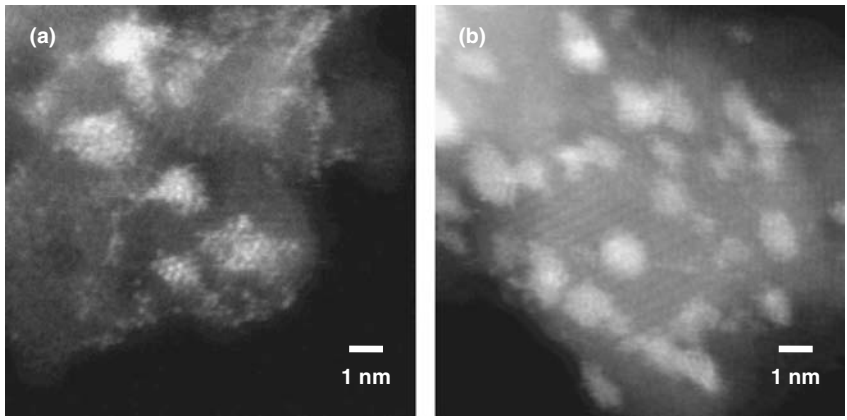


Figure 2.22 High-magnification Z-contrast micrographs showing 10% wt-loaded Au on anatase after two stages of preparation. (a) In the precursor state following deposition-precipitation of Au, individual Au atoms are sharply resolved, suggesting stabilised nanoparticles. (b) In the most active form, after mild reduction in 12%-H_2 at 423 K, individual Au atoms are not resolved, suggesting large structural fluxionality. Data reproduced from reference 83.

larger binding energies of CO and O_2 to Pt nanoparticles relative to flat surfaces, imply longer residence times, effectively blocking active sites. Thus, from calculations it appears that coordination number is indeed the driving factor behind the contrasting changes in activity with particle size in both Au and Pt nanoparticles.

Images of supported Au nanoparticles provide some tantalising evidence for the structural fluxionality,[76] necessary for high activity. Figure 2.22 shows atomic-resolution Z-contrast micrographs of as-synthesised and reduced nanoparticles, prepared as part of an experimental investigation of CO oxidation by Au nanoparticles on TiO_2.[79] In Figure 2.22(a) Au atoms are sharply resolved, whereas in Figure 2.22(b) Au nanoparticles appear blurred. The difference can be attributed to Au hydroxide present on the as-prepared samples, stabilising the nanoparticle. The fuzziness of Figure 2.22(b) provides evidence for large Au-atom motions, consistent with dynamic structural fluxionality[76] and the low melting point of Au nanoparticles in this size range.[82]

2.4 Summary and Outlook

In this chapter we have shown how aberration corrected STEM can be applied to a wide variety of nanoscale problems. We have illustrated how EELS allows us to obtain chemical information with atomic resolution and single-atom sensitivity. We have also shown that in order to understand the images and spectra obtained, it is sometimes necessary to perform dynamical scattering calculations of how the beam interacts with the sample, although Z-contrast

images are frequently directly interpretable. We have also seen how density-functional theory can provide additional insight into many problems and allows us to understand the physical consequences of what we see. Aberration correction is improving both the resolution and sensitivity of STEM imaging and new, improved instruments with higher-order or chromatic-aberration correctors will allow these advances to continue.

One exciting prospect is that it may become possible to overcome the historic limitations of the electron microscope image as a two-dimensional projection of a three-dimensional world. Some of the first steps towards three-dimensional microscopy have been taken and true atomic-resolution three-dimensional imaging appears to be within reach through a combination of tilt and depth slicing methods. Three-dimensional STEM may become routine on a wide variety of samples. In addition, we can anticipate extension of these methods to *in-situ* examinations, for example seeing changes in the morphology of catalysts under reaction conditions or watching the nucleation of nanotubes on a catalyst nanoparticle. Aberration-corrected STEM looks set for a very exciting and informative future.

Acknowledgements

We express our most sincere acknowledgements to all of our collaborators, including: J. Santamaria, H. Christen, S.H. Overbury, N.J Dudney, N. Dellby, O.L. Krivanek, P.D. Nellist, K. Sohlberg, Sanwu Wang, R. Adams and M. Amiridis. Research sponsored by the Division of Materials Sciences and Engineering, Office of Basic Energy Sciences, U.S. Department of Energy, under contract DE-AC05 00OR22725 with Oak Ridge National Laboratory, managed and operated by UT-Battelle, LLC.

References

1. A.V. Crewe, J. Wall and L.M. Welter, A High-Resolution Scanning Transmission Electron Microscope, *J. Appl. Phys.*, 1968, **39**, 5861.
2. R.P. Feynman, There's Plenty of Room at the Bottom, *Eng. Sci.*, 1960 **23**, 22.
3. J.F. Hainfeld, Understanding and Using Field Emission Sources, *Scan. Electron Microsc.*, 1977, **1**, 591.
4. L.W. Swanson, G.A. Schwind and J. Orloff, in *Handbook of Charged Particle Optics*, CRC Press, New York, 1997.
5. N. de Jonge, Y. Lamy, K. Schoots and T.H. Oosterkamp, High Brightness Electron Beam from a Multi-Walled Carbon Nanotube, *Nature*, 2002 **420**, 393.
6. P.C. Tiemeijer, J.H. Lin and A.F. de Jong, First Results of a Monochromized 200 kV TEM, *Microsc. Microanal.*, 2001, **7**(sup. 2), 1130.

7. D.B. Williams and C.B. Carter, *Transmission Electron Microscopy: A Textbook for Materials Science*, Plenum, New York, 1996.
8. J.M. Cowley, Image Contrast in a Transmission Scanning Electron Microscope, *Appl. Phys. Lett.*, 1969, **15**, 58.
9. A.P. Pogany and P.S. Turner, Reciprocity in Electron Diffraction and Microscopy, *Acta Crystallogr. Section A*, 1968, **24**, 103.
10. H. Kohl and H. Rose, Theory of Image-Formation by Inelastically Scattered Electrons in the Electron-Microscope, *Adv. Imaging Electron Phys.*, 1985, **65**, 173.
11. L.M. Brown, Scanning-Transmission Electron-Microscopy - Microanalysis for the Microelectronic Age, *J. Phys. F - Met. Phys.*, 1981, **11**, 1.
12. R.F. Egerton, *Electron Energy-Loss Spectroscopy in the Electron Microscope*, Plenum, New York, 1986.
13. E.C. Cosgriff, M.P. Oxley, L.J. Allen and S.J. Pennycook, The Spatial Resolution of Imaging Using Core-Loss Spectroscopy in the Scanning Transmission Electron Microscope, *Ultramicroscopy*, 2005, **102**, 317.
14. M.P. Oxley, E.C. Cosgriff and L.J. Allen, Nonlocality in Imaging, *Phys. Rev. Lett.*, 2005, **94**, 203906.
15. H. Rose, Image Formation by Inelastically Scattered Electrons in Electron Microscopy, *Optik*, 1976, **45**, 187.
16. J. Fertig and H. Rose, Resolution and Contrast of Crystalline Objects in High-Resolution Scanning-Transmission Electron-Microscopy, *Optik*, 1981, **59**, 407.
17. L.J. Allen, S.D. Findlay, M.P. Oxley and C.J. Rossouw, Lattice Resolution Contrast From a Focused Coherent Probe I, *Ultramicroscopy*, 2003, **96**, 47.
18. L.J. Allen and T.W. Josefsson, Inelastic Scattering of Fast Electrons by Crystals, *Phys. Rev. B*, 1995, **52**, 3184.
19. S.L. Dudarev, L.-M. Peng and M.J. Whelan, Correlations in Space and Time and Dynamical Diffraction of High-Energy Electrons by Crystals, *Phys. Rev. B*, 1993, **48**, 13408.
20. T. Fujikawa, Theory of the X-Ray Absorption Near Edge Structure (XANES) at a Deep L2,3 Edge Studied by the Short-Range Order Multiple-Scattering Theory, *J. Phys. Soc. Jpn.*, 1983, **52**, 4001.
21. P. Rez, J. Bruley, P. Brohan, M. Payne and L.A. Garvie, J. Review of Methods for Calculating near-Edge Structure, *Ultramicroscopy*, 1995, **59**, 159.
22. M. Varela, S.D. Findlay, A.R. Lupini, H.M. Cristen, A.Y. Borisevich, N. Dellby, O.L. Krivanek, P.D. Nellist, M.P. Oxley, L.J. Allen and S.J. Pennycook, Spectroscopic Imaging of Single Atoms Within a Bulk Solid, *Phys. Rev. Lett.*, 2004, **92**, 095502.
23. L.J. Allen, S.D. Findlay, A.R. Lupini, M.P. Oxley and S.J. Pennycook, Atomic-Resolution Electron Energy Loss Spectroscopy Imaging in Aberration Corrected Scanning Transmission Electron Microscopy, *Phys. Rev. Lett.*, 2003, **91**, 105503.
24. M. Varela, T.J. Pennycook, W. Tian, D. Mandrus, S.J. Pennycook, V. Pena, Z. Sefrioui and J. Santamaria, Atomic Scale Characterisation of Complex Oxide Interfaces, *J. Mater. Sci.*, 2006, **41**, 4389.

25. M. Varela, A.R. Lupini, S.J. Pennycook, Z. Sefrioui and J. Santamaria, Nanoscale Analysis of $YBa_2Cu_3O_{7-x}/La_{0.67}Ca_{0.33}MnO_3$ Interfaces, *Solid-State Electron.*, 2003, **47**, 2245.
26. V. Pena, Z. Sefrioui, D. Arias, C. Leon, J. Santamaria, M. Varela, S.J. Pennycook and J.L. Martiuez, Coupling of Superconductors Through a Half-Metallic Ferromagnet: Evidence for a Long-Range Proximity Effect, *Phys. Rev. B*, 2004, **69**, 224502.
27. Z. Sefrioui, M. Varela, V. Pena, D. Arias, C. Leon, J. Santamaria, J.E. Villegas, J.L. Martiuez, W. Salldarriaga and P. Prieto, Superconductivity Depression in Ultrathin $YBa_2Cu_3O_{7-delta}$ Layers in $La_{0.7}Ca_{0.3}MnO_3/YBa_2Cu_3O_{7-delta}$ Superlattices, *Appl. Phys. Lett.*, 2002, **81**, 4568.
28. M. Born and E. Wolf, *Principles of Optics*, Cambridge University Press, Cambridge, 1997.
29. P.W. Hawkes and E. Kasper, *Principles of Electron Optics*, Academic Press, London, 1989.
30. N. Dellby, O.L. Krivanek, P.D. Nellist, P.E. Batson and A.R. Lupini, Progress in Aberration-Corrected Scanning Transmission Electron Microscopy, *J. Electron Microsc.*, 2001, **50**, 177.
31. O.L. Krivanek, N. Dellby and A.R. Lupini, Towards Sub-Angstrom Electron Beams, *Ultramicroscopy*, 1999, **78**, 1.
32. O. Scherzer, Uber Einige Fehler von Elektonenlinsen, *Z. Phys.*, 1936 **101**, 593.
33. O. Scherzer, Sparische und Chromatische Korrektur von Elektronen-Linsen, *Optik*, 1947, **2**, 114.
34. A. Bleloch and A. Lupini, Imaging at the Picoscale, *Mater. Today*, 2004 **7**, 42.
35. H. Koops, Test of a Chromatically Corrected Objective Lens of an Electron-Microscope, *Optik*, 1978, **52**, 1.
36. M. Haider, S. Uhlemann, E. Schwan, H. Rose, B. Kabius and K. Urban, Electron Microscopy Image Enhanced, *Nature*, 1998, **392**, 768.
37. H. Rose, Correction of Aberrations, a Promising Means for Improving the Spatial and Energy Resolution of Energy-Filtering Electron-Microscopes, *Ultramicroscopy*, 1994, **56**, 11.
38. V.D. Beck, Hexapole Spherical-Aberration Corrector, *Optik*, 1979, **53**, 241.
39. A.V. Crewe and D. Kopf, Sextupole System for the Correction of Spherical-Aberration, *Optik*, 1980, **55**, 1.
40. D.F. Hardy, Combined Magnetic and Electrostatic Quadrupole Electron Lenses, PhD thesis, Cambridge University, UK, 1967.
41. Z.F. Shao, On the 5th Order Aberration in a Sextupole Corrected Probe Forming System, *Rev. Sci. Instrum.*, 1988, **59**, 2429.
42. V. Beck, Experiments with a Quadrupole Octupole Corrector in a STEM, *Proc. 32nd EMSA*, 1977, 90.
43. V. Ronchi, Forty Years of History of a Grating Interferometer, *Appl. Opt.*, 1964, **3**, 437.
44. J.C.H. Spence, *Experimental High-Resolution Electron Microscopy*, Oxford University Press, New York, 1988.

45. P.D. Nellist and J.M. Rodenburg, Beyond the Conventional Information Limit - the Relevant Coherence Function, *Ultramicroscopy*, 1994, **54**, 61.
46. J.-O. Malm and M.A. O'Keefe, in *51st Annual proceedings of the MSA*, ed. G.W. Bailey and C.L. Rieder, San Francisco Press, 1993, pp. 974.
47. J. Frank, Envelope of Electron Microscopic Transfer Functions for Partially Coherent Illumination, *Optik*, 1973, **38**, 519.
48. P.L. Fejes, Approximations for Calculation of High Resolution Electron Microscope Images of Thin Films, *Acta Crystallogr. Section A*, 1977, **33**, 109.
49. R.H. Wade and J. Frank, Electron Microscope Transfer Functions for Partially Coherent Axial Illumination and Chromatic Defocus Spread, *Optik*, 1977, **49**, 81.
50. P.D. Nellist and S.J. Pennycook, Sub-angstrom Resolution by Underfocused Incoherent Transmission Electron Microscopy, *Phys. Rev. Lett.*, 1998, **81**, 4156.
51. P.D. Nellist and S.J. Pennycook, Incoherent Imaging Using Dynamically Scattered Coherent Electrons, *Ultramicroscopy*, 1999, **78**, 111.
52. S.C. McFarlane, Imaging of Amorphous Specimens in a Tilted Beam Electron Microscope, *J. Phys. C - Solid-State Phys.*, 1975, **8**, 2819.
53. O.L. Krivanek, Method for Determining Coefficient of Spherical Aberration from a Single Electron Micrograph, *Optik*, 1976, **45**, 97.
54. J.M. Cowley, Electron Microdiffraction, *Adv. Electron. Electron Phys.*, 1978, **46**, 1.
55. J.A. Lin and J.M. Cowley, Calibration of the Operating Parameters for an HB5 STEM Instrument, *Ultramicroscopy*, 1986, **19**, 31.
56. Q.M. Ramasse and A.L. Bleloch, Diagnosis of Aberrations from Crystalline Samples in Scanning Transmission Electron Microscopy, *Ultramicroscopy*, 2005, **106**, 37.
57. D.E. Jesson and S.J. Pennycook, Incoherent Imaging of Crystals Using Thermally Scattered Electrons, *Proc. Roy. Soc. Lond. Series A - Math. Phys. Sci.*, 1995, **449**, 273.
58. A.J. den Dekker and A. van den Bos, Resolution: A survey, *J. Opt. Soc. Am. A*, 1997, **14**, 547.
59. M.A. O'Keefe, L.F. Allard and D.A. Blom, HRTEM Imaging of Atoms at Sub-Angstrom Resolution, *J. Electron Microsc.*, 2005, **54**, 169.
60. D. Van Dyck, S. Van Aert and den A.J. Dekker, Physical Limits on Atomic Resolution, *Microsc. Microanal.*, 2004, **10**, 153.
61. P.D. Nellist, M. Chisholm, N. Dellby, O.L. Kriranek, M.F. Murtitt, Z.S. Szilagyi, A.R. Lupini, A. Boriserich, W.H. Sides and S.J. Pennycook, Direct Sub-Angstrom Imaging of a Crystal Lattice, *Science*, 2004, **305**, 1741.
62. A.Y. Borisevich, A.R. Lupini and S.J. Pennycook, Depth Sectioning With the Aberration Corrected Scanning Transmission Electron Microscope, *Proc. Nat. Acad. Sci. USA*, 2006, **103**, 3044.
63. K. van Benthem, A.R. Lupini, M. Kim, H.S. Baik, S. Doh, J.H. Lee, M.P. Oxley, S.D. Findlay, L.J. Allen, J.T. Luck and S.J. Pennycook, Three Dimensional Imaging of Individual Hafnium Atoms Inside a Semiconductor Device, *Appl. Phys. Lett.*, 2005, **87**, 034104.

64. M. Minsky, Memoir on Inventing the Confocal Scanning Microscope, *Scanning*, 1988, **10**, 128.
65. T. Wilson, *Confocal Microscopy*, Academic Press, San Diego, 1990.
66. C. Dwyer and J. Etheridge, Scattering of Angstrom Scale Electron Probes in Silicon, *Ultramicroscopy*, 2003, **96**, 343.
67. A.Y. Borisevich, A.R. Lupini, S. Travaglini and S.J. Pennycook, Depth Sectioning of Aligned Crystals With the Aberration Corrected Scanning Transmission Electron Microscope, *J. Electron Microsc.*, 2006, **55**, 7.
68. Y.P. Peng, P.D. Nellist and S.J. Pennycook, HAADF-STEM Imaging With Sub-Angstrom Probes: a Full Bloch Wave Analysis, *J. Electron Microsc.*, 2004, **53**, 257–266.
69. P.D. Nellist and S.J. Pennycook, Direct Imaging of the Atomic Configuration of Ultradispersed Catalysts, *Science*, 1996, **274**, 413.
70. K. Sohlberg, S. Rashkeev, A.Y. Borisevich, S.J. Pennycook and S.T. Pantelides, Origin of Anomalous Pt–Pt Distances in the Pt/alumina Catalytic System, *ChemPhysChem*, 2004, **5**, 1893.
71. S.W. Wang, A.Y. Borisevich, S.N. Raskkeev, M.V. Glazott, K. Sohlberg, S.J. Pennycook and S.T. Pantelides, Dopants Adsorbed as Single Atoms Prevent Degradation of Catalysts, *Nature, Mater.*, 2004, **3**, 143.
72. M. Haruta, S. Tsubota, T. Kobayashi, H. Kageyama, M.J. Genet and B. Delmon, Low-Temperature Oxidation of CO over Gold Supported on TiO_2, Alpha-Fe_2O_3 and Co_3O_4, *J. Catal.*, 1993, **144**, 175.
73. G.R. Bamwenda, S. Tsubota, T. Nakamura and M. Haruta, The Influence of the Preparation Methods on the Catalytic Activity of Platinum and Gold Supported on TiO_2 for CO Oxidation, *Catal. Lett.*, 1997, **44**, 83.
74. M. Haruta, Nanoparticulate Gold Catalysts for Low-Temperature CO Oxidation, *J. New Mater. Electrochem. Systems*, 2004, **7**, 163.
75. M.S. Chen and D.W. Goodman, The Structure of Catalytically Active Gold on Titania, *Science*, 2004, **306**, 252.
76. B. Yoon, H. Hakkinen, U. Landman, A.S. Worz, J.M. Autonietti, S. Abbet, K. Judai and U. Heiz, Charging Effects on Bonding and Catalyzed Oxidation of CO on Au-8 Clusters on MgO, *Science*, 2005, **307**, 403.
77. J. Guzman and B.C. Gates, Catalysis by Supported Gold: Correlation Between Catalytic Activity For CO Oxidation and Oxidation States of Gold, *J. Am. Chem. Soc.*, 2004, **126**, 2672.
78. N. Lopez, T.V.W. Janssens, B.S. Clausen, Y. Xu, M. Marrikakis, T. Bliggard and J.K. Norskov, On the Origin of the Catalytic Activity of Gold Nanoparticles For Low-Temperature CO Oxidation, *J. Catal.*, 2004, **223**, 232.
79. W.F. Yan, B. Chen, S.M. Mahuriu, V. Schwartz, D.R. Mullius, A.R. Lupini, S.J. Pennycook, S. Dai and S.H. Overbury, Preparation and Comparison of Supported Gold Nanocatalysts on Anatase, Brookite, Rutile and P25 Polymorphs of TiO_2 For Catalytic Oxidation of CO, *J. Phys. Chem. B*, 2005, **109**, 10676.
80. G.M. Veith, A.R. Lupini, S.J. Pennycook, G.W. Ownby and N.J. Dudney, Nanoparticles of Gold on Gamma-Al_2O_3 Produced by DC Magnetron Sputtering, *J. Catal.*, 2005, **231**, 151.

81. E. Wahlstrom, N. Lopez, S. Schaub, P. Thostrup, A. Ronnan, C. Africh, E. Laegsgaard, J.K. Norskov and F. Besenbacher, Bonding of Gold Nanoclusters to Oxygen Vacancies on Rutile TiO_2(110), *Phys. Rev. Lett.*, 2003, **90**, 026101.
82. P. Buffat and J.P. Borel, Size Effect on Melting Temperature of Gold Particles, *Phys. Rev. A*, 1976, **13**, 2287.
83. S.N. Rashkeev, A.R. Lupini, S.H. Overbury, S.J. Pennycook and S.T. Pantelides, The Role of the Nanoscale in Catalytic CO Oxidation By Supported AU and PT Nanostructures, *Manuscript in Preparation*, 2007.

CHAPTER 3

Scanning Tunneling Microscopy of Surfaces and Nanostructures

M. R. CASTELL

Department of Materials, University of Oxford, Parks Road, Oxford OX1 3PH, UK

3.1 History of the STM

The Scanning Tunneling Microscope (STM) was invented in 1981 by a team including Gerd Binnig and Heinrich Rohrer at the IBM Zurich Research Laboratories. The first publications showing atomic resolution appeared in 1982,[1] and the pair were awarded the Nobel Prize for their invention in 1986 together with the inventor of the transmission electron microscope, Ernst Ruska. It is perhaps a sign of the immediate impact of the STM that its inventors only had to wait 5 years for their Nobel Prize, whereas Ruska had waited 55 years. Binnig and Rohrer's account of the birth of the STM is given in their Nobel Prize Lecture,[2] and it is interesting to note that they started the development of the STM without being aware of related work performed in the 1970s. The topografiner, invented by Russell Young,[3] used field-emitted electrons between a metal tip and a surface to map sample topography with the tip scanned using piezoelectric crystals. Clayton Teague's experiments on vacuum tunneling of electrons between two metal electrodes were also highly relevant because the STM is in essence a scanned tunnel junction.

There are two main reasons why STM became such a popular technique so rapidly after its invention. Firstly, it was the only technique to provide atomic resolution images of large flat surfaces, and secondly once adequate vibration isolation systems had been developed, it was relatively easy to implement. There were a vast number of home-built STMs in the 1980s until commercial manufacturers started to offer instruments and these are now widely used. The increased availability of STMs brought with it ever-larger dedicated conference attendances and publication output. This expansion accelerated further

following the invention of the atomic force microscope by Binnig *et al.*[4] in 1986 and these new instruments would from then on be referred to as scanning probe microscopes. There are now a number of authoritative books[5–8] and many specialised review articles[9–11] on STM.

The basic principle of the STM is to scan a sharp metal tip over a surface whilst maintaining a gap of a few Å (10^{-10}m) between the sample surface and the tip. A voltage is applied between the tip and the sample, which gives rise to a tunneling current, so called because it relies on the quantum mechanical phenomenon of electron tunneling. This tunneling current is of the order of a nA (10^{-9} Ampere) and can reveal the positions of individual atomic locations on the sample surface.

Many STMs are housed in ultrahigh vacuum (UHV) chambers. This is not a microscope requirement, but instead is necessary for the cleanliness of the surface. To image the top atomic layer of a sample it needs to be free of contamination and this can only be achieved in a vacuum environment of pressures below 10^{-8} Pa. STMs can operate in ambient conditions, and using suitable tips have been employed to image surface structure in solution, but these conditions preclude atomic resolution imaging. This chapter will focus solely on STM imaging in a UHV environment.

3.2 The Tunneling Interaction and Basic Operating Principles of STM

When two conductors are separated by a macroscopic vacuum gap, then electrons will not travel from one to the other unless a voltage above the field-emission threshold is applied. If, however, the conductors are brought into the proximity of less than a few nanometres, then their quantum-mechanical wavefunctions overlap to an extent that allows electrons to be shared between them. Classically, this interaction is not allowed because the electrons do not have sufficient kinetic energy to overcome the vacuum barrier. The movement of electrons across the barrier is therefore known as tunneling because the electrons "tunnel" through the vacuum barrier.

In an STM the tunneling current that flows between the sharp metallic tip and the sample is the central mechanism that allows the microscope to function. Figure 3.1 is a diagram showing the tip and sample atoms. The tunneling current density has been shown in grey and is most intense where the tip and sample are closest. The tunneling current can be described as

$$I_t \propto V\phi^2\rho_s\,\rho_t\,e^{-kd} \tag{3.1}$$

where I_t is the tunneling current, V is the voltage applied between the sample and the tip, ϕ is the average barrier height, ρ_s is the density of sample states, ρ_t is the density of tip states, k is a constant related to the decay length for the wavefunctions in a vacuum and d is the tip–sample separation. This tunneling current equation emerges from the derivation obtained by Tersoff and

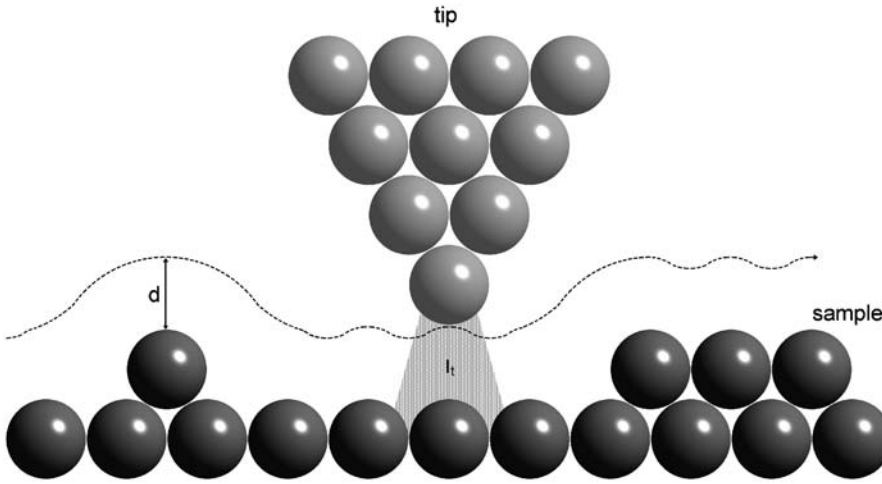


Figure 3.1 Schematic atomic level diagram of the STM tip and sample. The tip-to-sample separation is d, and the tunneling current is I_t. In the constant-current mode, shown here, the tip is scanned over the surface at constant I_t resulting in an atomic topography map indicated by the dashed line.

Hamman,[12] although other theoretical treatments have also been developed.[10] Given that k is approximately 2 Å^{-1} it can be seen from eqn (3.1) that an increase in tip–sample separation from 5 Å to 6 Å will cause the tunneling current to drop by almost an order of magnitude. It is this extreme sensitivity of the tunneling current to tip–sample separation that gives the STM its extremely high vertical resolution.

The inverse exponential relationship between distance and tunneling current means that the atom at the apex of the tip is the one from which the overwhelming majority of the electrons tunnel. In turn, electrons will only tunnel into the closest sample atoms, and this is the reason why STM images show only the structure of the topmost atomic layer of a sample.

The relative energy levels of the tip and sample are shown in the diagrams of Figure 3.2. In Figure 3.2(a) the tip and sample are separated by a macroscopic distance. The exponentially decaying tip wavefunction is shown, and the local density of electronic states of the sample are indicated. When the tip and sample are brought within a few nanometres of each other electrons can tunnel between them, which allows the Fermi levels to equalise as shown in Figure 3.2(b). When a steady state has been reached as many electrons tunnel from the tip into the sample and *vice versa*.

If the sample is biased by a positive voltage, V_s with respect to the tip, then the energy levels of the tip will move up by eV_s. This situation is shown in Figure 3.2(c). Electrons tunnel from the filled tip states into the empty sample states (indicated in grey), and a dc tunneling current is established. The magnitude of the tunneling current depends amongst other things on the local density of empty states of the sample. By biasing the sample negatively with

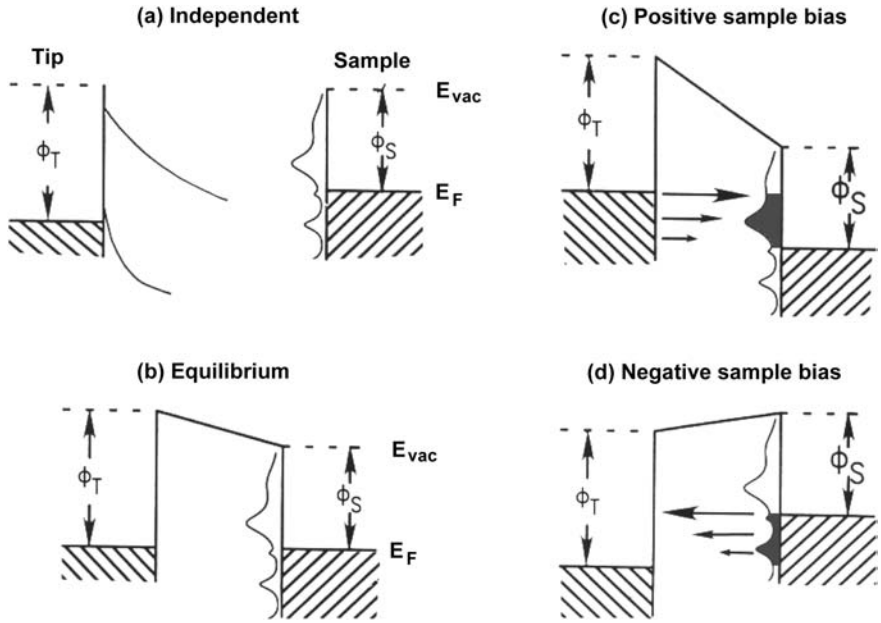


Figure 3.2 Diagrams of the relative energy levels of the STM tip and sample. (a) The tip and sample do not interact and the vacuum levels (E_{vac}) are the same. (b) The small vacuum gap allows tunneling between tip and sample leading to equalisation of the Fermi energies (E_F). (c) The sample is biased positively resulting in a tunneling current from the tip into the empty electronic states of the sample. (d) The sample is biased negatively causing electrons from filled sample states to tunnel into the tip. Reprinted with permission from The Annual Review of Physical Chemistry, Volume 40 © 1989 by Annual Reviews.[13]

respect to the tip the tunneling current flows in the opposite direction, as shown in Figure 3.2(d). For a negatively biased sample the tunneling current is a function of the filled density of sample states. It is therefore possible to use the STM to perform local electron spectroscopy of filled and empty sample states. This model makes the reasonable assumption that the density of empty and filled electronic states close to the Fermi energy of the metallic tip is rather featureless. It also requires there to be a strong energy dependence of the density of sample states. For example, a system where these restrictions would be satisfied is a W tip scanning a GaAs crystal surface.

Once a tip and sample have formed a suitable tunnel junction, the tip can be scanned in 3D to allow an image to be created. Scanning the tip is achieved by attaching it to piezoelectric crystals that change their length when a voltage is applied along them. The simplest form of scanning involves having three separate piezoelectric crystals responsible for movements in the *x*-, *y*-, and *z*-directions. It is more common, however, to use so-called tube and shear scanning geometries. The great advantage of piezoelectric scanners is that their

expansion or contraction, and hence the movement of the tip, can be controlled very accurately.

There are two main modes used to create a surface topography map. The first is to scan the tip in a plane over the surface and record the tunneling current as a function of x and y. This mode is called constant-height mode, and the image is displayed with the brightest regions being those where the tunneling current is greatest. This mode is sometimes used for rapid scanning over very flat surfaces. The second mode, and by far the most widely used, is the constant-current mode shown in Figure 3.1. In this operating mode the tunneling current is determined by the user, and the tip height is adjusted by the computer to maintain a constant tunneling current. When the tip is scanned, the height of the tip is recorded, and the higher the tip the brighter that region in the image. The constant-current mode can operate over much greater z distances than the constant-height mode, and it is this increased dynamic range that makes it the more usual choice for STM operation.

The resolution of the STM is ultimately determined by the nature of the electronic structure of the atom at the tip apex, which tends to be outside the operator's control. There are also other factors such as vibrations, noise in the control electronics, noise in the tunneling current, and the electronic structure of the sample surface that will limit the resolution. Generally the x and y lateral resolution of a good STM is around 1 Å, and the vertical z resolution is better than 0.1 Å. The transmission electron microscope has seen a steady resolution improvement over the years due to the enhancement of electron optics and advances in image processing. However, the STM's resolution has remained virtually unchanged since its original inception because the limit to the resolution, the apex atom, cannot be "improved".

3.3 Atomic-Resolution Imaging of Surface Reconstructions

When a crystal is terminated and forms a surface, the bonding arrangement of the surface atoms differs significantly from those in the bulk. The result is that the location of the surface atoms changes and when this happens the surface is termed "reconstructed". The periodicity of the new arrangement of the surface atoms give rise to a surface unit cell that is defined as the multiple of the unit cell of a bulk termination of the crystal. A good example of a reconstruction is the Si (111) surface, as illustrated in Figure 3.3.

A (111) slice through a Si crystal exposes a high density of dangling bonds that extend into the vacuum. The surface can reduce the number of dangling bonds by a complicated atomic reordering process involving adatoms, a stacking fault, dimerisation, and a deep vacancy, which all coexist to form the (7×7) reconstruction. In Figure 3.3(a) a reconstructed terrace is shown with the (7×7) surface unit cell indicated on the image. Numerous single-atom defects can be seen in the image. A higher magnification image of the atomic

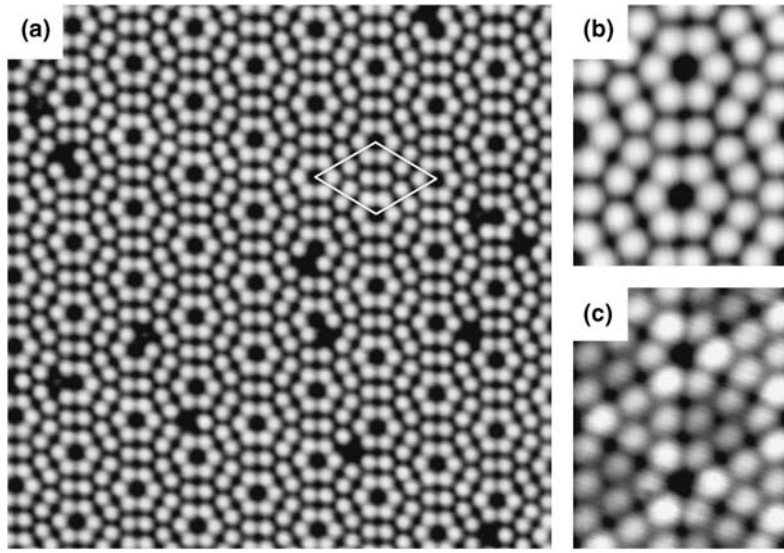


Figure 3.3 STM images of the Si (111)-(7 × 7) reconstruction. (a) An image of a terrace is shown. The (7 × 7) unit cell is indicated and individual point defects can be seen. The image width is 21 nm. (b) A high magnification empty states image is shown that was taken at a sample bias of +2 V. All adatoms are equally bright. (c) A filled-states image taken at a sample bias of –2 V. The adatoms in half the unit cell are brighter than in the other half. Image widths in (b) and (c) are 4.7 nm.

structure is shown in Figure 3.3(b) where it can be seen that only 12 bright spots per unit cell are imaged. These spots are the adatoms that form the outermost atomic layer of the silicon crystal. As mentioned above, electron tunneling overwhelmingly takes place between the tip atom and surface atoms that are in closest proximity. This is why the STM images in Figure 3.3 are composed solely of the adatoms.

Although STM images of the Si (111)-(7 × 7) reconstruction show the detailed arrangement of the adatoms, the complicated atomic structure is a number of monolayers deep. These deeper layers do not lend themselves to STM analysis. The full atomic structure was actually solved by Takayanagi *et al.*[14] using transmission electron diffraction, who proposed the Dimer-Adatom-Stacking fault (DAS) structure. Other surface-sensitive diffraction techniques such as Low-Energy Electron Diffraction (LEED), Reflection High-Energy Electron Diffraction (RHEED), or X-ray diffraction can also be used to detect the (7 × 7) ordering of the surface. However, diffraction by its nature averages over very many surface unit cells, so although the surface periodicity is measured, all information on individual defects is averaged out. The quality that sets the STM apart from these diffraction techniques is that it images the surface in real space, and therefore all defects are captured in the image. The image in Figure 3.3(a) illustrates this, where 13 adatoms appear to be missing.

As discussed in the previous section, by applying negative sample biases filled electronic states are measured and by applying positive sample biases empty electronic states are measured. The STM images in Figures 3.3(b) and (c) demonstrate the influence of the electronic structure. In Figure 3.3(b) a voltage of +2 V has been applied to the sample with respect to the tip, and the image has therefore been created through electron tunneling into empty silicon states. In this image all the adatoms appear equally bright. In Figure 3.3(c) a negative voltage has been applied to the sample, so that electrons tunnel from the sample to the tip, which results in a filled-states image. All the bright spots in the image are still adatoms, but there is now a subtle difference in the relative brightness of the adatoms in the unit cell. This change in brightness is due to the relative differences in the filled local density of states for the various adatom locations.

The ability of the STM to detect single atom vacancies, as in Figure 3.3(a), is mainly a topographic effect. Given that the capability exists to image differences in local electronic structure, one would expect to be able to see defects that do not disrupt the surface topography but that are electronically distinct. This is indeed possible and as shown in a later section the location of individual dopant atoms can be imaged on certain semiconductor surfaces.

An example of a reconstruction on a metal surface is shown in Figure 3.4. This is an STM image of a Pt (001) surface, which undergoes a Pt (001)-hex-R0.7°

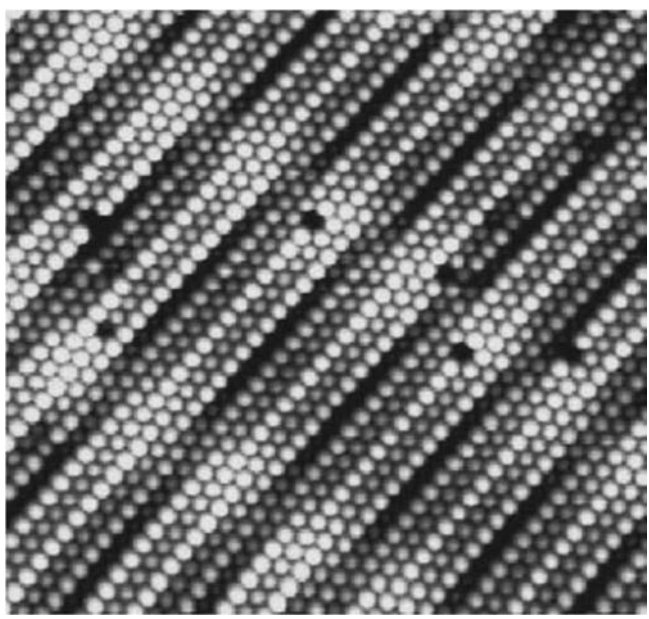


Figure 3.4 STM image of the Pt (001)-hex-R0.7° reconstructed surface. The corrugated topography of the surface and single atom vacancies can be seen. The image size is 10 nm × 10 nm, the tunneling current is 6.1 nA, and the sample bias voltage is 5.1 mV. Reprinted from Borg *et al.*[15] © 1994 with permission from Elsevier.

reconstruction. This reconstruction is rather complicated because the surface layer of atoms forms a quasi hexagonal structure on top of the square Pt lattice. The driving force for this reconstruction is that a close-packed hexagonal structure can accommodate more atoms than an open square lattice, with an associated lowering of the surface energy. The mismatch between the hexagonal surface atomic layer and the layer below gives rise to the complicated topography shown in Figure 3.4. The most obvious features of the surface are the linear corrugations that run diagonally across the image. Individual atomic vacancies can also be seen.

Another metal crystal where reconstructions are seen is the Au (111) surface. If the atoms in the (111) plane of the face centred cubic (fcc) crystal retained their bulk positions a flat hexagonal surface lattice would result. This is what happens to all fcc metal crystal (111) surfaces apart from Au. On Au (111) surfaces a complicated interaction of short-range and long-range forces allows the surface atoms to shorten their nearest-neighbour distances to compensate for the lower surface atom coordination. The surface layer is uniaxially compressed by 4.4%, which gives rise to changes in the atomic stacking sequence across the surface varying from unfaulted fcc stacking to faulted hcp stacking. The regular arrangements of the stacking faults produce a $\sqrt{3} \times 23$ surface unit cell as shown in the STM image of Figure 3.5.

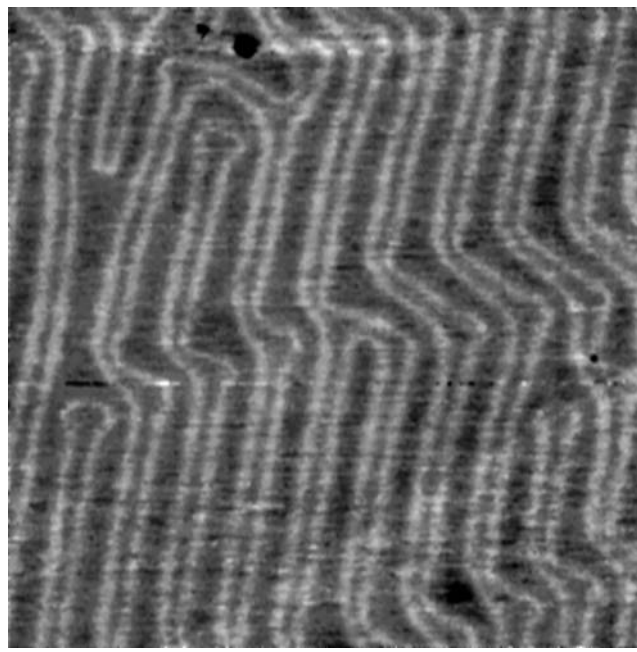


Figure 3.5 STM image of the reconstructed Au (111) surface. The pairs of bright lines are the hcp to fcc stacking boundaries. Image size 81×81 nm^2, sample bias 0.1 V, tunneling current 3 nA. Image courtesy of Dr Quanmin Guo, University of Birmingham, UK.

The pairs of bright lines in Figure 3.5 are the regions in the crystal where the surface atoms have been forced to occupy incommensurate bridge sites that form the boundary between regions of fcc and hcp stacking. The atoms on the bridge sites stand proud from the surface, and therefore appear bright in the STM images. The reconstruction is characterised by 60° kinks in the fcc–hcp boundaries. These kinks are also referred to as elbows and can order to form what is known as the herringbone reconstruction. Kinks may act as preferential adsorption sites for adsorbed atoms or molecules.

3.4 Imaging of Surface Nanostructures

A nanostructure is virtually any piece of material of nanometre dimensions. There are essentially two ways to create nanostructures, either by the top-down or the bottom-up approach. The top-down method employs techniques such as electron beam or UV lithography followed by chemical etching and materials evaporation onto a patterned substrate to produce small structures. This approach has held the microelectronics and data storage industries in good stead since the early 1960s because the technology for miniaturising the relevant processes has kept improving. These days the processes are so refined that highly regular features of tens of nanometres in size can be created. However, when this technology finally reaches its limit it will be necessary to look to bottom-up methods to take over.[16]

The basic building blocks of bottom-up technologies are atoms, atomic clusters, and molecules. The STM has the ability to move around individual atoms and molecules, with the result that one could in theory build nanostructures atom by atom. In practice this would take far too long and is therefore not commercially viable, though it is still amazing that it can be done at all. A technological solution to the bottom-up approach requires that inherent molecular and atomic interactions between the building blocks be exploited to create nanostructures. The simplest interaction is due to the interatomic potential that causes atoms to condense together to become crystals. Atoms can be evaporated onto a substrate, and given sufficient thermal energy, they will diffuse to form regular structures that are determined by their interatomic potential and their bonding to the substrate. In practice, relative surface energy effects tend to cause either thin-film wetting to take place or for nanocrystals to form.

In Figure 3.6 Ge nanocrystals grown on a Si (001) substrate are shown. The interesting aspect of Ge on Si growth is that the first few monolayers of Ge wet the Si surface, but because the lattice parameter of Ge is greater that that of Si, strain builds up in the Ge layer. This strain can be relieved by the formation of small isolated islands. This process is known as Stranski–Krastanov growth. A particularly interesting feature of the Ge islands is that their shape changes as they grow in size. In Figure 3.6(a) the bimodal distribution of small pyramids and larger domes can be seen in the STM image. The atomic structure of a (105) side facet of a pyramid is shown in Figure 3.6(b). These images

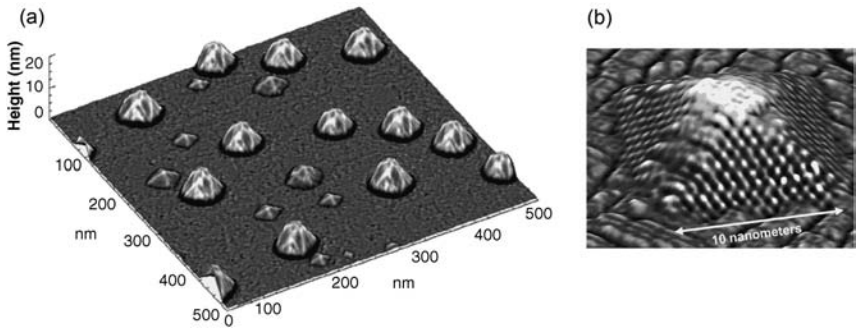


Figure 3.6 STM images of Ge nanocrystals on Si (001). (a) The two crystal shapes that can be seen are small pyramids and larger domes. Image rendering has been adjusted to show the curvature of the topography. (b) A high magnification image of a single pyramid in which the atomic scale structure of the (105) side facets is resolved. Image (a) is taken from *Science*, 1998, **279**, 353, reprinted with permission from AAAS.[17] Image (b) © (1999) Hewlett-Packard Development Company, L.P. Reproduced with permission.

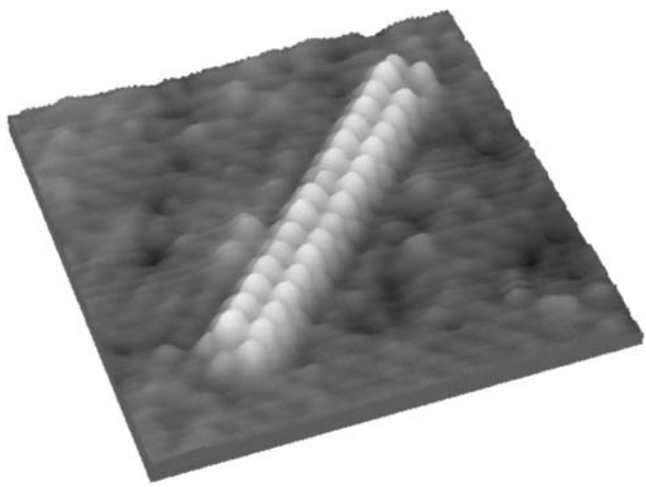


Figure 3.7 3D rendering of an STM image of a nanoline on a $SrTiO_3$ (001) surface. The image is 14 nm × 14 nm, and the nanoline is 12 nm long and 0.2 nm high. Individual paired atoms run along the length of the nanoline. Adapted from Castell.[19]

demonstrate the capability of the STM to image the surface topography of surface nanostructures.

Other examples of nanostructures, which form spontaneously by localised phase separation, are nanolines on the surface of $SrTiO_3$ (001). An isolated 12 nm long line is shown in Figure 3.7 where the individual paired atoms that form

the repeat unit of the nanoline are resolved. The image demonstrates atomic-resolution capability but also poses the question of chemical sensitivity. Are the raised dots of the nanoline Ti, O, or Sr? The STM cannot answer that question, and other surface-science techniques are needed to make inferences about the chemical properties of nanostructures. Auger electron spectroscopy has been used on $SrTiO_3$ samples covered in nanolines and compared with a stoichiometric sample.[18] An enhanced Ti peak is observed for the nanoline samples, which allows one to infer that they are TiO_x rich. This information is useful, but it would be far better to be able to distinguish directly between different chemical elements in the STM image. Tunneling spectroscopy can be used to obtain information about the local electronic structure, but it is difficult to correlate this data with specific elements.

The examples in Figures 3.6 and 3.7 above rely on interatomic potentials to cause nanostructure formation. The process variables available are therefore inherently limited, and rely on controlling the complicated kinetic processes during nanostructure nucleation and growth. An alternative is to use molecules as the building blocks for the creation of ordered patterns and surface nanostructures. This provides new possibilities because molecules can link to each other via noncovalent interactions such as hydrogen bonding. The molecules can be synthesised in such a way as to control the bonding sites as well as the molecular size and shape. By combining different types of molecules a vast number of different architectures are possible. Self-assembled patterns of molecules into networks are known as supramolecular nanostructures.

An example of how a supramolecular network may be achieved is by coadsorption of 1,3,5-triazine-2,4,6-triamine (melamine) and perylene tetracarboxylic di-imide (PTCDI) onto a Ag terminated Si (111) surface. Surface diffusion of the melamine and PTCDI molecules allows them to interact with each other through the formation of hydrogen bonds. Three PTCDI molecules can bond with a melamine molecule, as shown in the inset of Figure 3.8. This nucleus then allows further attachment of melamine and PTCDI molecules, with the result that a honeycomb network is created, as seen in the majority of the image in Figure 3.8. The underlying Ag/Si(111) substrate can also be seen in the top left and bottom right parts of the image.

The hydrogen bonding between the linear PTCDI and triangular melamine molecules gives rise to a relatively open network with large pores. These pores may be used to capture other molecules such as C_{60} fullerenes. In Figure 3.8 the bright flower like features are pores that have captured seven C_{60} molecules to form heptomers. It is interesting to note that isolated C_{60} heptomers are not usually seen in other systems and that this arrangement has been stabilised by the underlying PTCDI–melamine network. This is particularly exciting because the pores in such networks have the potential to act as nanoscale vessels that can promote local chemical interactions.

Another area where STM has contributed to the understanding and development of technologically important materials is in heterogeneous catalysis. Industrial catalyst nanoparticles are usually supported on disordered, porous, and insulating substrates, which makes them unsuitable for STM analysis.

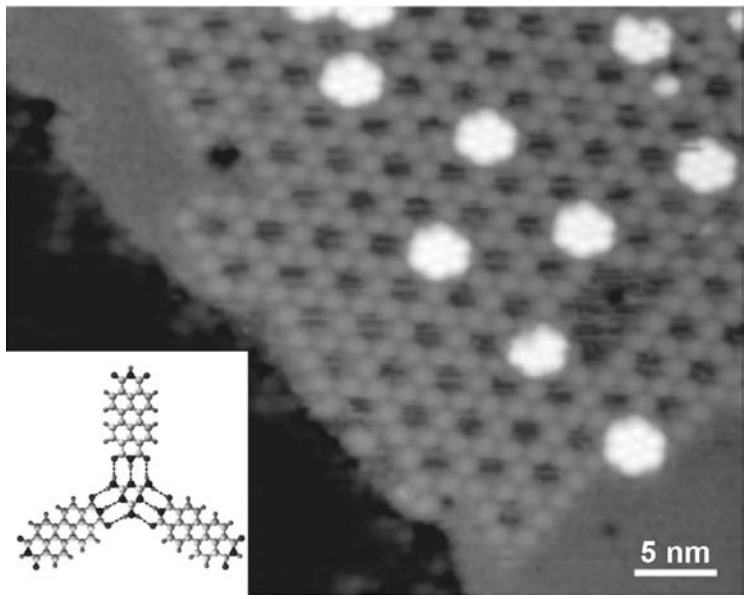


Figure 3.8 STM image of C_{60} heptamers confined within the gaps of a PTCDI-melamine network. The open honeycomb network is stabilised by hydrogen bonding between the molecules as indicated by dotted lines in the inset. Melamine is the central triangular molecule in the figure inset and PTCDI are the three larger molecules. Reprinted by permission from Macmillan Publishers Ltd: *Nature*, **424**, 1029, © 2003.[20]

Hence, in order to study catalysts, model systems have been developed where the nanoparticles are dispersed on single-crystal substrates. Much STM work has been carried out on supported metal particles on TiO_2 crystals and Al_2O_3 thin films.[21–24] The example discussed below is the MoS_2-based hydrodesulfurisation catalyst, which is used as a way of creating cleaner fuels. Naturally there is a significant interest in relating the catalytic activity of the MoS_2 nanoclusters with their atomic and chemical structure.

An STM image of MoS_2 nanoclusters grown on a Au (111) surface is shown in Figure 3.9(a). The Au (111) support is used because it is chemically inert, and the properties of the nanoclusters can therefore be studied in isolation from the support. Au (111) also lends itself particularly well to the growth of the clusters. As discussed previously, the Au (111) surface adopts $\sqrt{3} \times 23$ ordering, which can have regularly spaced 60° kinks that result in the herringbone reconstruction. When Mo is evaporated onto the surface, Mo islands tend to nucleate in the elbows of the 60° kinks. This gives rise to a regularly dispersed arrangement of Mo islands. Upon sulfidation they become flat, monolayer-thick S–Mo–S crystals, with a characteristic triangular shape, as shown in Figure 3.9(a). The high magnification image of a single cluster in Figure 3.9(b) shows that it has a bright edge. This brightening is not due to increased rim height, but rather it shows the existence of metallic localised electron edge states. It is the metallic

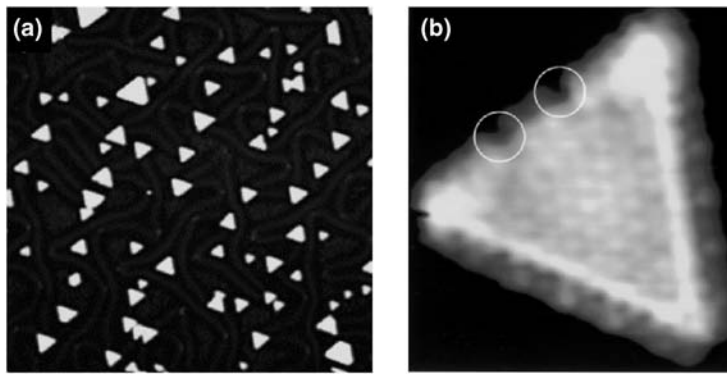


Figure 3.9 STM images showing MoS_2 nanoclusters on a Au (111) surface. (a) The distribution of clusters (image size 75 nm × 75 nm). (b) A close-up of a cluster after exposure to hydrogen which results in S vacancies as indicated with circles. In (b) the bright edge around the rim of the cluster is due to an increased local density of states. Image (a) reprinted from Lauritsen *et al.*[25] © 2006 with permission from Elsevier. Image (b) reprinted from Helveg *et al.*[26] (*Phys. Rev. Lett.*, **84**, 951) © 2000 with permission from The American Physical Society.

character of these states, which allows electrons to be donated or accepted, that gives rise to the catalytic behaviour of the MoS_2 nanoclusters.

Modification of the edges of the MoS_2 nanoclusters can be achieved by exposing them to atomic hydrogen. Figure 3.9(b) shows two circled edge locations where S atoms have been removed following hydrogen dosing. Detailed information of this type about the local atomic and electronic structure of the nanoclusters has been used to manufacture a new type of hydrotreating catalyst. The new catalyst material maximises the number of metallic edge states, which gives it its enhanced hydrogenation properties.

3.5 Manipulation of Adsorbed Atoms and Molecules

The previous section illustrated how molecules can adsorb onto a substrate and self-assemble into functionally interesting structures. This self-assembly relies on sufficient thermal energy in the system to allow the molecules to diffuse once they have been adsorbed on the surface. However, if molecular or atomic adsorption is carried out at low temperatures then the molecules are fixed in place. The STM tip can then be used to drag or push the adsorbates to predetermined locations on the surface. This procedure was first performed in Don Eigler's group at the IBM Almaden Research Center in 1990. The result is shown in Figure 3.10 where Xe atoms have been used to spell out the letters IBM.[27]

The atomic manipulation procedure that led to the IBM logo was carried out in a liquid-He-cooled STM operating at 4 K. Cryogenic STMs have the

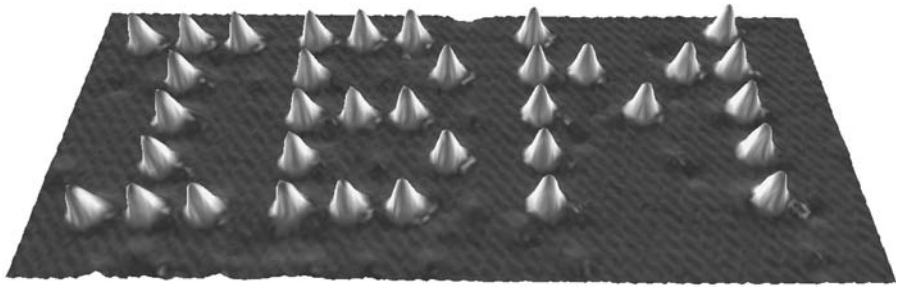


Figure 3.10 Three-dimensionally rendered STM image of manipulated Xe atoms on a Ni (110) surface spelling IBM. Each letter is 5 nm from top to bottom. The STM was operated at liquid He temperatures. Reprinted by permission from Macmillan Publishers Ltd: *Nature*, **344**, 524, © 1990.[27] Reprint courtesy of International Business Machines Corporation, © 1990, IBM.

advantage that the sample and microscope are very stable over long periods of time allowing single-atom experiments to be performed over a series of days. After preparing a clean and ordered Ni (110) surface Xe atoms are dosed onto the surface and form a random distribution. The Xe adsorbates are imaged with a sample bias of –0.01 V and a tunneling current of 1 nA, and under these conditions they appear as bumps 0.16 nm in height in the STM images. For atomic manipulation, the tip is placed above an atom and the tunneling current is set to typically between 10 and 60 nA. The increase in tunneling current is achieved by reducing the tip–atom distance, which also has the effect of strengthening the van der Waals interaction between the tip and the Xe atom. The tip is then moved to a desired location, and due to the bond between tip and Xe atom, the Xe atom is dragged along with the tip. The Xe–tip bond is then broken by decreasing the tunneling current setpoint and moving the tip away from the Xe atom. The letters IBM were written by sequentially moving randomly positioned atoms to specific locations on the surface and took many hours to write. Atomic manipulation of this kind can be automated, but it is still a lengthy process.

One might argue that atomic manipulation resulting in the letters IBM is not particularly technologically useful. There are easier ways to store information. However, Eigler's group have taken a step towards nanologic circuitry by using atomic manipulation to create logic gates made from complicated arrangements of CO molecules on a Cu (111) surface.[28] The molecular cascades work in a similar way to toppling dominoes. A series of end-on dominoes with multiple pathways can be set up to act as logic gates, such as AND and OR gates, and Eigler's group did something analogous with CO molecules to demonstrate mechanical computation on the nanometre length scale. CO molecules turn out to be more stable if they are separated by a $\sqrt{3}$ lattice spacing than if they are in a nearest-neighbour configuration. A series of CO molecules can be set up so that when the first molecule is pushed into a nearest-neighbour site by the STM tip its neighbour eventually hops to another site.

The CO sequence is arranged in such a way that the hopping molecule ends up in a nearest-neighbour site and causes yet another molecule to hop. This results in the molecular cascade. By building two or more cascade pathways it is possible to construct complicated logic circuits.

Experiments of this type highlight the versatility of the STM in that it can be used not only to image surfaces at atomic resolution, but it also has the capability to manipulate matter atom by atom. It is therefore the ultimate bottom-up tool for materials processing. However, in practice, unless atomic manipulation can be made to be highly automated with hundreds of STMs operating simultaneously it is unlikely to become a widespread manufacturing device. However, specific experiments such as those illustrating the quantum nature of electrons can be carried out using atomically manipulated surface structures.

One of the first examples of the use of the STM to directly image quantum-mechanical eigenstates was also carried out in Eigler's group.[29] Quantum mechanics shows that a quantum particle trapped in a potential well can only have discrete eigenfunctions and adopt discrete energy eigenvalues. In practice, such a scenario can be created by laterally confining electrons in a surface state. A surface state is limited only to the top layer of atoms, and hence the electrons in the surface state are also confined to the top layer. If an atom is adsorbed on the surface then the surface-state electrons will scatter from it. An enclosure, or corral, of sufficiently densely packed adsorbates will confine the surface-state electrons within it.

Figure 3.11 shows the building of a quantum corral using atomic manipulation of 48 Fe atoms on a Cu (111) surface. The sequence of low temperature (4 K) STM images shows snapshots of the building process. Quantum confinement effects can already be seen in Figure 3.11(b) where there is only a semicircle of Fe atoms. In Figure 3.11(c) the standing wave pattern is well developed, and it appears near perfect in Figure 3.11(d).

Anyone with doubts about the wavelike nature of quantum particles must surely have them eliminated with the images in Figure 3.11. However, it is important to point out that the STM images the local density of states, and not the quantum mechanical wavefunction itself. Electrons trapped in the quantum corral create a standing wave pattern, which results in a high average electron density at the maxima, and a low average electron density at the nodes. What we see in the STM images is the variation of the electron density. This is equivalent to the envelope of the standing wave pattern created by the trapped electrons.

In related experiments, Eigler's group were able to use quantum corrals to create what they termed a quantum mirage.[30] To achieve this they built elliptical corrals from Co atoms on a Cu (111) surface. An elliptical shape was used because this geometry has two foci. For example, if light is emitted from one of the foci inside an elliptical mirror, all the light will be focused to the other focal point. They then placed a single Co atom inside the corral at one of the foci of the ellipse. Because of the magnetic nature of Co, isolated Co atoms induce a Kondo resonance. This is because the electrons near the Fermi energy

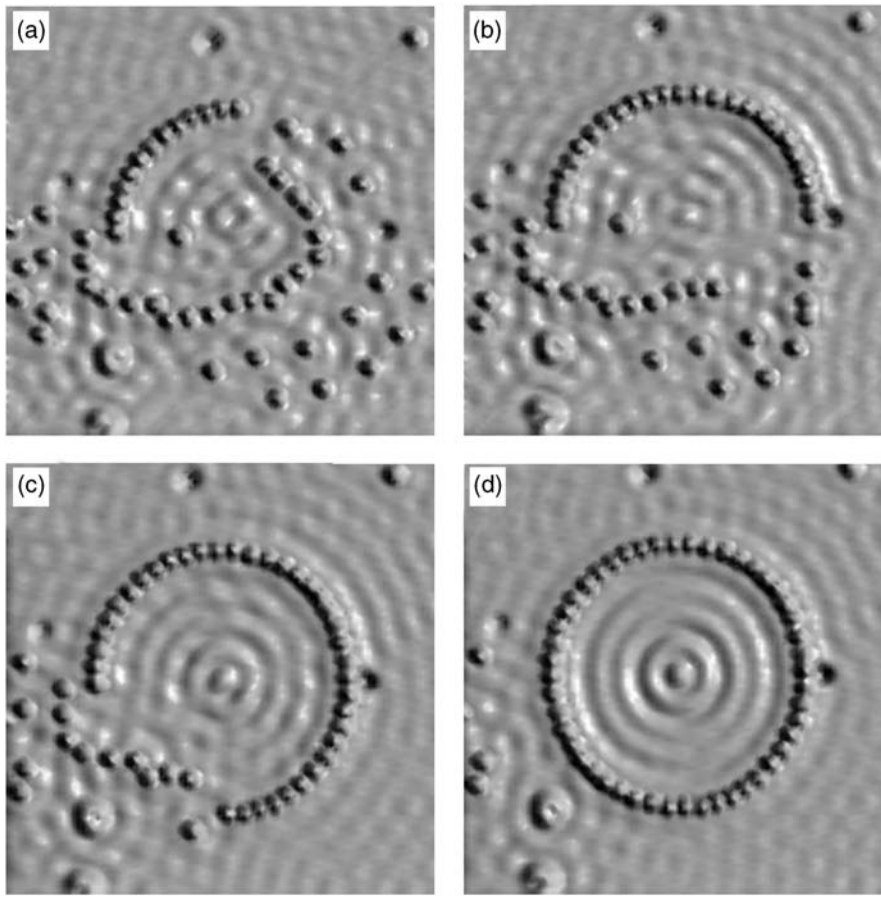


Figure 3.11 Sequence of STM images showing the building of a quantum corral out of Fe atoms on a Cu (111) surface. The Fe atoms are deposited randomly and then manipulated into a circle. (a) A quarter of the circle is seen, (b) Half of the circle, and (c) Three quarters. The standing wave pattern within the circle in (d) is created by the eigenstates of the surface state electrons trapped within the corral. The corral is made from 48 Fe atoms and has a diameter of 14.26 nm. Image reproduced from *Science*, 1993, **262**, 218 reprinted with permission from AAAS.[29] Reprint courtesy of International Business Machines Corporation, © 1993, IBM.

of the nonmagnetic host, Cu in this case, screen the magnetic moment of the Co impurity. This Kondo resonance can be detected using the STM because of the way in which it perturbs the local density of states around the Fermi energy. The Cu surface state electrons that scatter from the Co atom within the corral then rescatter from the corral wall and are coherently focused onto the other focus of the ellipse. It was shown that the Kondo resonance was also present at the opposite empty focus point from where the Co atom was placed. We are familiar with light scattering off an object and focusing it to create its

image, but to achieve something equivalent with electrons was a significant breakthrough.

An ever-present issue is that atomic manipulation is usually carried out at cryogenic temperatures and this adds another technological hurdle, though for some systems such as Br on Cu (001) room-temperature manipulation has been shown to be possible.[31] Atomic manipulation at room temperature requires that the adsorbate atoms be sufficiently strongly bonded to the substrate so that uncontrolled diffusion due to thermal motion does not occur. However, at the same time the bond must not be too strong to prevent the atom being moved by the tip. These criteria immediately limit the adsorbate–substrate systems where atomic manipulation at room temperature is possible.

In contrast to atomic manipulation, the substantial armoury of synthetic chemistry can be used to create molecules that are ideally suited to molecular manipulation. One such molecule is Cu-tetra(3,5 di-tertiary-butyl-phenyl)-porphyrin with four di-tertiary butyl phenyl substituents, which are rotated out of the plane of the porphyrin ring. The molecule itself is quite rigid and is therefore not damaged when manipulated, but because only the ends of the legs are in contact with the surface it can be pushed around by the STM tip.

Figure 3.12 shows a sequence of STM images during the molecular manipulation procedure of the porphyrin molecules on a Cu (100) surface. Due to

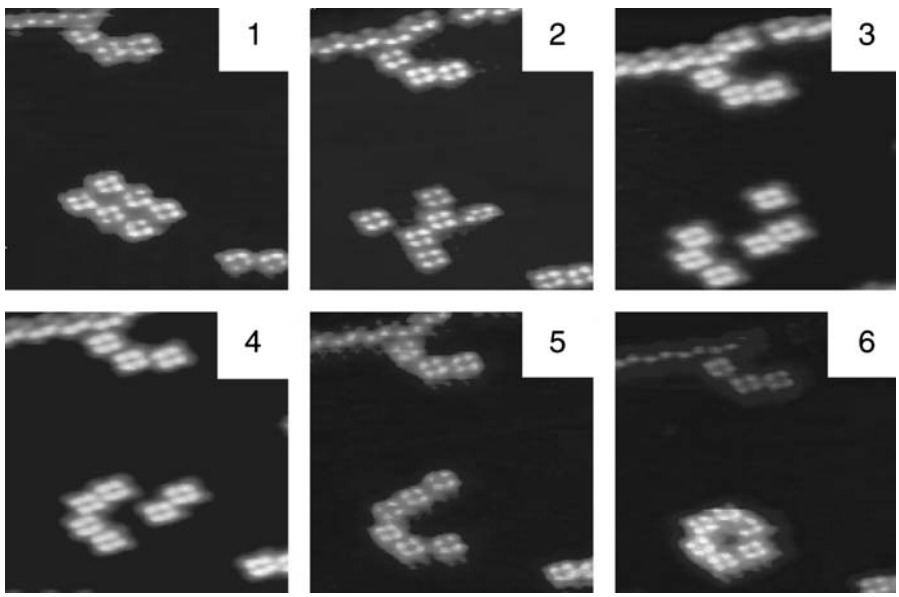


Figure 3.12 Sequence of STM images (26×26 nm^2) showing the construction of a ring of 6 Cu-TBP-porphyrin molecules in the lower half of each frame. The molecules in the upper half of the frames are unperturbed. Manipulation is performed at room temperature with the tip pushing the molecules into place. Image reproduced from *Science*, 1996, **271**, 181 reprinted with permission from AAAS.[32]

the shape of the molecules they always appear as 4 bright spots clustered together. The bright spots are the tops of the individual legs. In the lower part of panel 1 of Figure 3.12 there are 6 molecules clustered together. This image is taken with a sample bias of typically 2.2 V and a tunneling current of 0.08 nA. To push the molecules around the bias voltage is reduced to 0.03 V and the tunneling current is increased to 2 nA, *i.e.* in manipulation mode the tip is much closer to the molecule and a tip movement can therefore push the molecule across the surface. This is illustrated in panel 2 of Figure 3.12 where the tip in manipulation mode has been dragged through the previously ordered 2 by 3 array of molecules which are now disordered. Subsequent images show how the individual molecules are then arranged into a hexagonal ring.[32]

Although there have been astonishing recent breakthroughs, atomic and molecular manipulation are still in their infancy. One might for example expect there to be more work on custom assembly of molecules atom by atom. Eigler's group started this process by assembling CsI and NaI chains from their constituent atoms on a Cu (111) surface, but more complicated molecular assembly has not been reported so far. One drawback is the lack of technological push. Self-assembly of nanostructures is viewed as a potentially very important field because the scale up from the laboratory to the factory floor is not inherently difficult. However, assembly on a single atom by atom scale using an STM still has the problems of speed and mass parallelisation to overcome before it can become a realistic manufacturing technology.

3.6 Influence of the Surface Electronic States on STM Images

Equation (3.1) shows that the tunneling current between the sample and the STM tip depends on the local density of states of the tip and sample as well as on the inverse exponential separation of the tip and sample. How then is it possible to distinguish between changes in topography and changes in the local density of states? The way these factors are separated in practice is by taking images under different bias conditions. Theoretical modelling of the surface electronic structure is also an invaluable tool.

A particularly good example of the influence of the local density of states on STM images occurs in doped semiconductors. Almost all STM imaging of dopants in semiconductors has been done on (110) cleavage surfaces of zinc-blende III-V compounds. This surface is particularly suitable because it does not have any electronic surface states in the bandgap, so the Fermi level is not pinned by surface states. Dopant atoms have electronic states in the bandgap that pin the Fermi level, so they appear as depressions or protrusions in STM images. Figure 3.13 shows a Ga vacancy and Si dopant atom point defects on the GaAs (110) surface.

To interpret the STM images of Figure 3.13 it is necessary to understand the nature of the atomic lattices in the background of the images. The GaAs (110)

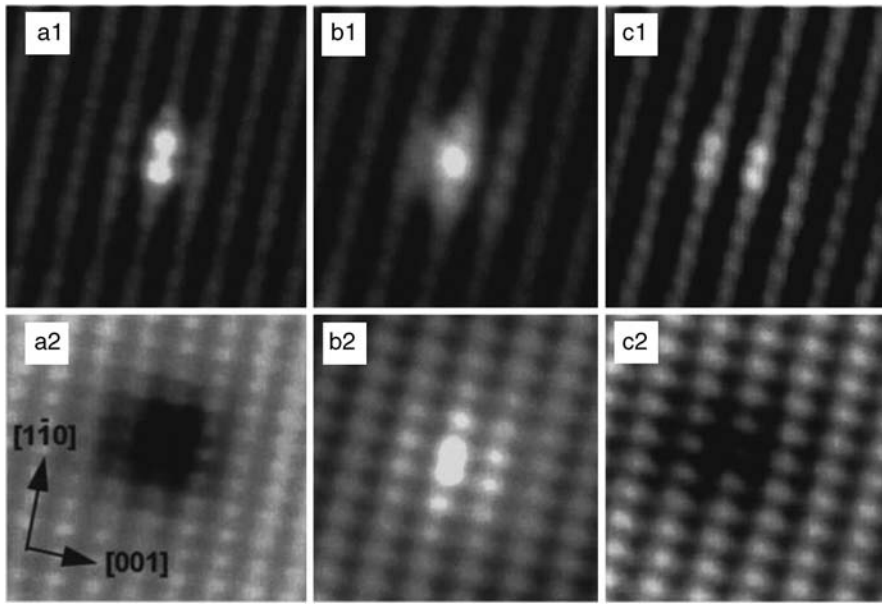


Figure 3.13 STM images of filled states (upper panels) and empty states (lower panels) of point defects on the Si doped GaAs (110) surface. (a1 and a2) A Ga vacancy. (b1 and b2) A Si donor dopant located substitutionally on a subsurface Ga site. (c1 and c2) A Si acceptor dopant located substitutionally on a subsurface As site. Sample biases were (a1) −2.4 V, (a2) + 1.8 V, (b1) −2.0 V, (b2) + 1.4 V, (c1) −2.0 V, (c2) + 1.4 V. Image reprinted from Domke *et al.*[33] (*Phys. Rev. B*, **54**, 10288) © 1996 with permission from The American Physical Society.

surface consists of rows of Ga and As ions. Due to the electronic structure of GaAs, when filled states are imaged they reveal the locations of the As atoms (upper panels in Figure 3.13), and when empty states are imaged the Ga atoms are seen (lower panels in Figure 3.13). Atomic-scale defects on these surfaces have a distinctly different appearance depending on whether filled or empty states are imaged.

Figures 3.13(a1) and (a2) are STM images of a surface Ga vacancy. The defect causes an increase in the filled states around the defect and a reduction in the empty states, resulting in bright spots in (a1) and a dark depression in (a2). This contrast behaviour can be explained through the negative charge on the defect which gives rise to band bending. Figures 3.13(b1) and (b2) show a subsurface Si atom located substitutionally on a Ga site, called a Si_{Ga} donor. A Si atom on a Ga site acts as an n-type dopant and increases both the filled and empty local density of states. Si in GaAs is an amphoteric system, meaning that Si can either locate substitutionally on a Ga or an As site. If Si is located on an As site it will act as a p-type dopant. This atomic arrangement is shown in panels (c1) and (c2), which gives rise to a negatively charged defect, and displays similar contrast to a Ga vacancy.

Point defects on (110) surfaces of III-V compound semiconductors are good examples of how the local density of states can affect STM images. In contrast, silicon surfaces have surface states in the bandgap, with the result that dopants cannot be easily detected. That is why STM images of doped silicon surfaces never show any sign of their dopants. Other materials where individual electronically active point defects have been imaged include Li acceptors in NiO,[34] and impurities in high-T_c superconductors.[35]

Another example of significant surface electronic structure effects on STM images can be seen on the rutile TiO_2 (110) surface. The (1×1) termination of this crystal is shown in Figure 3.14(a). The prominent rows of bridging oxygen ions and rows of 5-fold coordinated Ti ions are indicated. The topography of this surface is therefore dominated by rows of prominent oxygen atoms. The corresponding empty states STM image of this surface is shown in Figure 3.14(b), which is dominated by bright rows. Using a purely

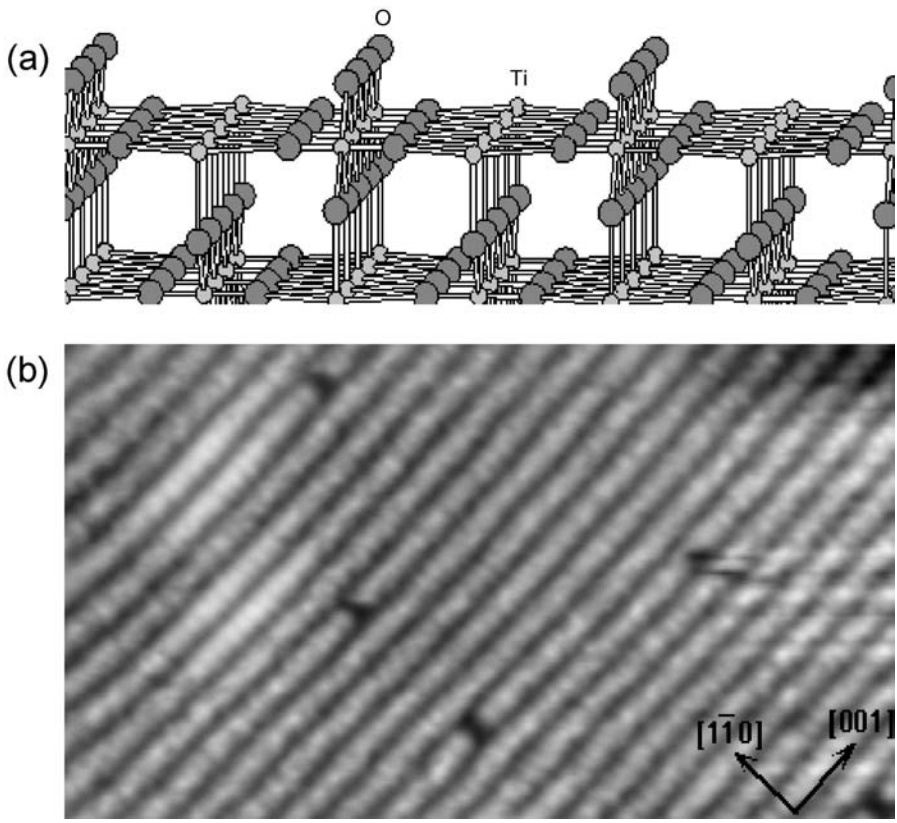


Figure 3.14 Schematic diagram and STM image of the rutile TiO_2 (110) surface. (a) Model of the (1×1) crystal termination which has rows of O (large dark spheres) and Ti ions (small light spheres) on the surface. The STM image in (b) shows characteristic bright rows, which are associated with the Ti ions (15×8 nm^2, $V_s = +1.0$ V, $I_t = 0.1$ nA).

topographical interpretation of the STM image one would assume that the bright rows in the STM image are due to the prominent rows of O atoms. In fact, this interpretation is incorrect. Surface electronic-structure calculations reveal that the contribution to the empty density of states of the under-coordinated Ti ions are so large that more electrons tunnel into the Ti ions even though they are further away from the STM tip.[36] Surprisingly, the bright rows in the STM image in Figure 3.14b are therefore due to the Ti ions.

The examples of local density of states effects on STM images described above are intended to illustrate that an understanding of electronic surface structure is no less important than surface topography when interpreting the image contrast due to atomic-scale features. Of course for larger features of nanometre dimensions, such as that shown in Figure 3.6(a), the influence of the electronic structure is diminished, and the STM images are dominated by sample topography. For samples where there is uncertainty about the relative contributions of electronic structure and surface topography it is advisable to image over a large range of voltages. By changing the imaging voltage electrons will tunnel into (or out of) different energy states and changes in contrast will be evident in the images if there is a strong energy dependence of the local density of states.

3.7 Tunneling Spectroscopy

Given that the STM tunneling current is dependent on the local electronic structure, and that the STM tip is a high-resolution probe, one is rapidly drawn to the exciting prospect of using the STM as a high-resolution electron spectroscopy tool. This is indeed possible, and can be illustrated by examination of the tunneling diagram in Figure 3.2. In Figure 3.2(c) electrons tunnel from the tip into the empty sample states. The electrons at the Fermi energy in the tip experience the lowest barrier height and therefore have a higher tunneling probability, as indicated in the figure through the use of a larger arrow. By keeping the tip to sample distance constant and varying the sample bias one can thus measure the local density of empty sample states.

Tunneling spectroscopy is less advantageous for probing filled electronic states. Electrons near the Fermi energy of the sample always contribute most to the tunneling current. This prevents a particular energy window from being selected and means that the tunneling data is sensitive to changes in the tip empty density of states. However, filled-states spectroscopy can still be useful if an electronically featureless tip is used in conjunction with a sample that has a strong energy dependence of its density of filled states as illustrated in Figure 3.2(d).

The main method for performing tunneling spectroscopy is to measure bias voltage versus tunneling current at a constant tip–sample separation. The differential tunneling conductance $\mathrm{d}I/\mathrm{d}V$ divided by the conductance I/V is related to the local density of states.[6] It is therefore possible to numerically process an *I versus V* spectrum to obtain the surface local density of states. In

reality the poor signal-to-noise ratio of *I versus V* curves tends to require that numerous scans are averaged. A method to improve the signal involves superimposing a high frequency sinusoidal voltage modulation on top of the imaging bias. This allows the differential tunneling conductance dI/dV to be measured directly as a function of imaging bias using a lock-in amplifier. Additionally, signal-to-noise levels are generally improved significantly by performing the spectroscopy experiments at low temperatures.

An example of a tunneling spectrum from a p-type Ge (111) crystal surface is shown in Figure 3.15 (from Feenstra *et al.*[37]). In the graph a gap devoid of states can be seen extending from –0.1 V to + 0.5 V. As one would expect in p-type material, the Fermi energy at 0 V lies towards the bottom of the bandgap. The peak labelled "A" at + 0.7 V is due to the empty states associated with the adatoms that form part of the c(2 × 8) reconstructed surface. Beyond this band above + 1.0 V lie the conduction band states labeled "C". The filled states peaks in Figure 3.15 are due to the valence band (V) and the rest atoms (R) associated with the reconstruction. Overall, the tunneling spectrum reveals that the surface bandgap of 0.6 eV has bulk-related valence-band states as the lower boundary and surface adatom states as the upper boundary.

Tunneling spectra of the quality of Figure 3.15 are not easily obtained because numerous experimental challenges need to be overcome. However, once an instrument is set up correctly one can probe both filled and empty states at atomic resolution. This is the equivalent of an extremely high spatial resolution combined photoemission and inverse photoemission experiment, but in reality tunneling spectroscopy suffers from a number of drawbacks. One limitation is that only the states due to wavefunctions protruding into the vacuum are detected, and this may be only part of the electronic structure

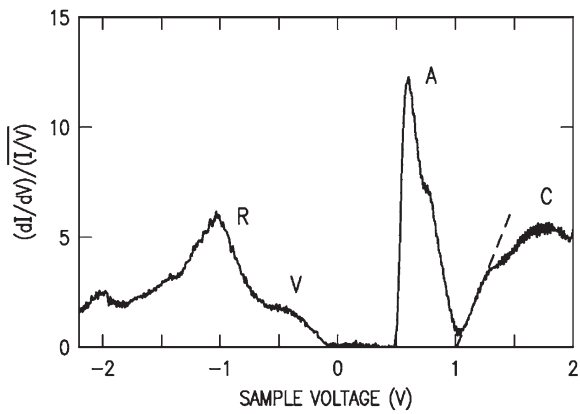


Figure 3.15 Normalised tunneling spectrum acquired at 61 K of the Ge(111)-c(2 × 8) surface. The features in the spectrum are due to surface rest atoms (R), the valence band (V), surface adatoms (A), and the conduction band (C). Image reprinted from Feenstra *et al.*[37] (*Phys. Rev. B*, **73**, 035310) © 2006 with permission from The American Physical Society.

relevant to the surface region. A more serious problem is that tunneling spectroscopy is limited to a few eV around the Fermi energy. This precludes the detection of core states, which in turn prevents chemically specific information from being recorded. Additionally, the tip shape can affect the tunneling spectra, and it is generally found that blunt tips produce better results than sharp ones.

3.8 Tip Artefacts in STM Imaging

The assumption in this chapter so far has been that the STM tip is sharp and terminated by a single metal atom. But what if the apex atom is something other than metallic? Much of the work in this area is rather speculative because it is not possible to measure the chemical identity of the apical atom. In addition, effects that are unusual and might be due to foreign tip atom adsorption are frequently observed by STM operators, but rarely reported in the scientific literature. Of all the good quality images taken in an experimental STM session less than 1% will ever be seen by the wider community. This results in an element of self-censorship, where the microscopist decides which images are worthy of dissemination and which are not. Images containing features due to tip artefacts or that display unusual contrast tend to be rejected. An exception to this are experiments involving ordered O monolayers on metal surfaces, because the effect is so striking and reproducible. Depending on whether the tip is a metal or an O atom, a contrast inversion takes place. A metallic tip shows the adsorbed O atoms as depressions, whereas an O tip shows them as protrusions. The origin of the effect has been revealed through studies on the Ru (0001) surface covered with an ordered O (2×2) monolayer, where STM experiments were compared with *ab-initio* modeling.[38] The results can be explained by a lowering of the local density of states by the adsorption of an O atom onto the tip apex.

Although contrast reversals of atomic periodicities are interesting, the most usual tip artefacts are due to blunt or irregularly shaped tips. An STM image where a particular feature has been reproduced all over the image is probably of dubious quality. Figure 3.16 is a schematic diagram demonstrating the effect of tip shape on STM images. In Figure 3.16(a) a perfectly sharp tip is imaging a sample with a nanoscale pyramidal protuberance on it and the resultant STM image is shown below. The height of the pyramid is recorded accurately, but even with a perfect tip, the pyramid in the STM image appears wider than it should. This is because the vacuum gap between the tip and sample must be maintained when the tip approaches the side of the sample. In the STM image the pyramid will therefore appear to have an additional width of twice the tip–sample separation. Of course no tip can be perfectly sharp because even a single atom apex has a finite end-radius associated with it.

A more realistic tip shape is shown in Figure 3.16(b). The end of the tip is slightly blunt, and the resultant STM image reflects this through a widening

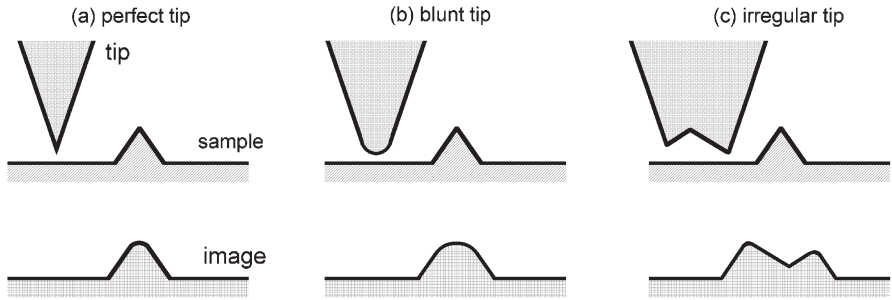


Figure 3.16 Schematic diagrams demonstrating the effect of tip shape when imaging a sharp nanoscale surface feature. (a) A perfect tip is used to produce an image that closely reflects the shape of the pyramidal protuberance. (b) The blunted tip gives rise to a rounded appearance of the pyramid. The irregular tip shape shown in (c) results in an image that is a convolution of the shapes of the tip and sample.

and rounding of the pyramidal feature. It should be noted, however, that even with a blunt tip the pyramid height is still accurate in the STM image. In general, the larger the tip end radius the more sample features will suffer from broadening in the image.

Figure 3.16(c) shows a tip with an irregular termination. The pyramidal feature on the sample is sharper than the end of the tip, with the result that the STM image displays a complicated convolution of the pyramidal feature and the end of the tip. In a sense the pyramidal feature of the sample is imaging the tip, and not the other way around. A tip of this type could still be used to successfully image flat surfaces at atomic resolution because only the right-hand side of the tip would be close enough to the sample to take part in the tunneling process. However, to image nanostructures the tip in Figure 3.16(c) is unsuitable.

An example of tip–sample convolution is shown in Figure 3.17. Fe islands have been grown on a $SrTiO_3$ (001) surface through vapour phase deposition in ultra high vacuum.[39] The pointed shape of the Fe islands is reproduced accurately when a sharp tip is used, as shown in Figure 3.17(a). Around 50 Fe islands can be seen situated on terraces divided by 3 step edges. During STM imaging the tip can sometimes crash into the surface, a small particle may be picked up by the end of the tip, or the sharp end of the tip can break off. These processes often result in an irregular tip termination shown schematically in Figure 3.16(c). The result of imaging following a tip crash can be seen in Figure 3.17(b). In this image the Fe islands are no longer pointed compact shapes, but are imaged as identical irregular tooth-shaped features. An STM image like this, showing repeated identical features, is a clear indication of tip–sample convolution. An experienced STM operator will immediately recognise a convolution, and reject the image, but surprisingly quite a lot of such images appear regularly in the scientific literature.

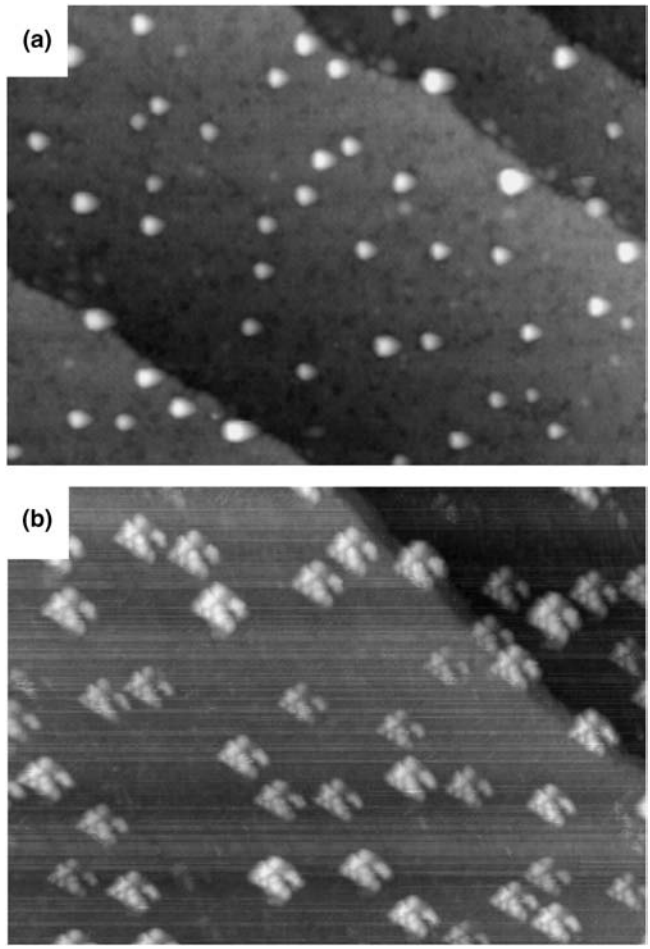


Figure 3.17 STM images of Fe islands on a $SrTiO_3$ (001) support. (a) The tip is sharp and images the substrate terraces and around 50 Fe islands. The image in (b) was taken minutes later from the same sample under the same conditions, but following a tip crash. The irregular tip end shape that resulted from the crash has been imaged by the sharp Fe islands and is reproduced all over the image. Image width = 100 nm, $V_s = 1.1$ V, $I_t = 0.1$ nA.

3.9 Conclusions

There is no doubt that scanning tunneling microscopy and atomic force microscopy have revolutionised our fundamental understanding of atomic and electronic surface structure. Prior to the inception of the STM it was experimentally extremely challenging to determine the nature of surface defects, the dynamics of terrace growth and step behaviour, the location of molecular

adsorption sites, and the evolution of nanostructure growth and shape. Surface science had previously been limited to diffraction techniques so that when the STM provided atomic-scale real-space imaging, it cast a new light on the subject.

Moreover, the use of STM as a way of manipulating individual atoms and molecules demonstrates its potential as an atomic-scale lithography tool. This has resulted in exotic structures such as quantum corrals that allow us to image electron standing waves. These types of experiments are an important step towards demonstrating the possibility of harnessing the power of the wave-mechanical nature of nanoscale particles and structures. With the use of scanning tunneling spectroscopy it is now also possible to obtain electronic structure information on the atomic scale.

The invention of the STM led to the development of the AFM, which in turn spawned a wide variety of other scanning probe microscopes. Apart from the undoubted utility of these instruments to scientific research, they can also be regarded as one of the drivers of nanotechnology. Without the ability to image what has been made or perform local property measurements it is difficult to implement changes in processing conditions that might improve a device or structure.

Developments in local probe methods since the invention of the STM in 1981 have been truly remarkable. In parallel, other nanoscale imaging and analysis methods have also been vastly improved, as described in the other chapters of this book. The challenge for researchers in the field on nanoscience now is no longer the development of new analysis techniques, but rather how to combine existing methods to maximise their relative value for any given project.

References

1. G. Binnig, H. Rohrer, C. Gerber and E. Weibel, *Surface studies by scanning tunneling microscopy*, *Phys. Rev. Lett.*, 1982, **49**, 57.
2. G. Binnig and H. Rohrer, *Scanning tunneling microscopy-from birth to adolescence*, *Rev. Mod. Phys.*, 1987, **59**, 615.
3. R. Young, J. Ward and F. Scire, *The topografiner: an instrument for measuring surface microtopography*, *Rev. Sci. Instrum.*, 1972, **43**, 999.
4. G. Binnig, C.F. Quate and C. Gerber, *Atomic Force Microscope*, *Phys. Rev. Lett.*, 1986, **56**, 930.
5. C.J. Chen, *Introduction to Scanning Tunneling Microscopy*, Oxford University Press Inc., USA, 1993.
6. R. Wiesendanger, *Scanning Probe Microscopy and Spectroscopy: Methods and Applications*, Cambridge University Press, UK, 1994.
7. R. Wiesendanger and H.-J. Güntherodt, *Scanning Tunneling Microscopy I-III*, Springer Verlag, Germany, 2nd edn, 1996.
8. D. Bonnell, *Scanning Probe Microscopy and Spectroscopy*, John Wiley and Sons Inc., USA, 2nd edn, 2001.

9. G.A.D. Briggs and A.J. Fisher, *STM experiment and atomistic modelling hand in hand: individual molecules on semiconductor surfaces*, *Surf. Sci. Rep.*, 1999, **33**, 1.
10. A.D. Gottlieb and L. Wesoloski, *Bardeen's tunneling theory as applied to scanning tunneling microscopy: a technical guide to the traditional interpretation*, *Nanotechnology*, 2006, **17**, R57.
11. R.G. Egdell and F.H. Jones, *Structure and reactivity of oxide surfaces: new perspectives from scanning tunneling microscopy*, *J. Mater. Chem.*, 1998, **8**, 469.
12. J. Tersoff and D.R. Hamann, *Theory and application for the scanning tunneling microscope*, *Phys. Rev. Lett.*, 1983, **50**, 1998.
13. R.J. Hamers, *Atomic-Resolution Surface Spectroscopy with the Scanning Tunneling Microscope*, *Annu. Rev. Phys. Chem.*, 1989, **40**, 531.
14. K. Takayanagi, Y. Tanishiro, S. Takahashi and M. Takahashi, *Structure analysis of Si(111)-7×7 reconstructed surface by transmission electron diffraction*, *Surf. Sci.*, 1985, **164**, 367.
15. A. Borg, A.M. Hilmen and E. Bergene, *STM studies of clean, CO- and O_2-exposed Pt (100)-hex-R0.7°*, *Surf. Sci.*, 1994, **306**, 10.
16. J.V. Barth, G. Costantini and K. Kern, *Engineering atomic and molecular nanostructures at surfaces*, *Nature*, 2005, **437**, 671.
17. G. Medeiros-Ribeiro, A.M. Bratkovski, T.I. Kamins, D.A.A. Ohlberg and R.S. Williams, *Shape transition of germanium nanocrystals on a Si (001) surface from pyramids to domes*, *Science*, 1998, **279**, 353.
18. D.S. Deak, F. Silly, D.T. Newell and M.R. Castell, *Ordering of TiO_2-based nanostructures on $SrTiO_3$ (001) surfaces*, *J. Phys. Chem. B*, 2006, **110**, 9246.
19. M.R. Castell, *Nanostructures on the $SrTiO_3$ (001) surface studied by STM*, *Surf. Sci.*, 2002, **516**, 33.
20. J.A. Theobald, N.S. Oxtoby, M.A. Phillips, N.R. Champness and P.H. Beton, *Controlling molecular deposition and layer structure with supramolecular surface assemblies*, *Nature*, 2003, **424**, 1029.
21. M. Baumer and H.J. Freund, *Metal deposits on well-ordered oxide films*, *Prog. Surf. Sci.*, 1999, **61**, 127.
22. C.R. Henry, *Morphology of supported nanoparticles*, *Prog. Surf. Sci.*, 2005, **80**, 92.
23. C.T. Campbell, *Ultrathin metal films and particles on oxide surfaces: structural, electronic and chemisorptive properties*, *Surf. Sci. Rep.*, 1997, **27**, 1.
24. U. Diebold, *The surface science of titanium dioxide*, *Surf. Sci. Rep.*, 2003, **48**, 53.
25. J.V. Lauritsen, R.T. Vang and F. Besenbacher, *From atom-resolved scanning tunneling microscopy (STM) studies to the design of new catalysts*, *Catal. Today*, 2006, **111**, 34.
26. S. Helveg, J.V. Lauritsen, E. Laensgaard, I. Stensgaard, J.K. Norskov, B.S. Clausen, H. Topsoe and F. Besenbacher, *Atomic-scale structure of single-layer MoS_2 nanoclusters*, *Phys. Rev. Lett.*, 2000, **84**, 951.

27. D.M. Eigler and E.K. Schweizer, *Positioning single atoms with a scanning tunneling microscope*, *Nature*, 1990, **344**, 524.
28. A.J. Heinrich, C.P. Lutz, J.A. Gupta and D.M. Eigler, *Molecule cascades*, *Science*, 2002, **298**, 1381.
29. M.F. Crommie, C.P. Lutz and D.M. Eigler, *Confinement of electrons to quantum corrals on a metal surface*, *Science*, 1993, **262**, 218.
30. H.C. Manoharan, C.P. Lutz and D.M. Eigler, *Quantum mirages formed by coherent projection of electronic structure*, *Nature*, 2000, **403**, 512.
31. T.W. Fishlock, A. Oral, R.G. Egdell and J.B. Pethica, *Manipulation of atoms across a surface at room temperature*, *Nature*, 2000, **404**, 743.
32. T.A. Jung, R.R. Schlitter, J.K. Gimzewski, H. Tang and C. Joachim, *Controlled room-temperature positioning of individual molecules: molecular flexure and motion*, *Science*, 1996, **271**, 181.
33. C. Domke, P. Ebert, M. Heinrich and K. Urban, *Microscopic identification of the compensating mechanisms in Si-doped GaAs*, *Phys. Rev. B*, 1996, **54**, 10288.
34. M.R. Castell, P.L. Wincott, N.G. Condon, C. Muggelberg, G. Thornton, S.L. Dudarev, A.P. Sutton and G.A.D. Briggs, *Atomic resolution STM of a system with strongly correlated electrons: NiO (001) surface structure and defect sites*, *Phys. Rev. B*, 1997, **55**, 7859.
35. E.W. Hudson, K.M. Lang, V. Madhavan, S.H. Pan, H. Eisaki, S. Uchida and J.C. Davis, *Interplay of magnetism and high-T_c superconductivity at individual Ni impurity atoms in $Bi_2Sr_2CaCu_2O_{8+\delta}$*, *Nature*, 2001, **411**, 920.
36. U. Diebold, J.F. Anderson, K. Ng and D. Vanderbilt, *Evidence for the tunneling site on transition-metal oxides: TiO_2 (110)*, *Phys. Rev. Lett.*, 1996, **77**, 1322.
37. R.M. Feenstra, J.Y. Lee, M.H. Kang, G. Meyer and K.H. Rieder, *Bandgap of the Ge(111)c(2×8) surface by scanning tunneling spectroscopy*, *Phys. Rev. B*, 2006, **73**, 035310.
38. F. Calleja, A. Arnau, J.J. Hinarejos, A.L. Vazquez de Parga, W.A. Hofer, P.M. Echenique and R. Miranda, *Contrast reversal and shape changes of atomic adsorbates measured with scanning tunneling microscopy*, *Phys. Rev. Lett.*, 2004, **92**, 206101.
39. F. Silly and M.R. Castell, *Fe nanocrystal growth on $SrTiO_3$ (001)*, *Appl. Phys. Lett.*, 2005, **87**, 213107.

CHAPTER 4

Electron Energy-loss Spectroscopy and Energy Dispersive X-Ray Analysis

R. BRYDSON

Leeds Electron Microscopy and Spectroscopy Centre, Institute for Materials Research, SPEME, University of Leeds, Leeds LS2 9JT, UK

4.1 What is Nanoanalysis?

Generally the term nanoanalysis, as applied to the investigation of solids, encompasses the determination of a number of interdependent types of information about a specimen, all of which ultimately govern the resultant physical and chemical properties. These may be conveniently classified as follows:

(i) the morphology (the microstructural or nanostructural architecture);
(ii) the crystal structure (the detailed atomic arrangement within the chemical phases contained within the microstructure);
(iii) the chemistry (the elements and possibly molecular groupings present);
(iv) the electronic structure (the nature of the bonding between atoms).

Strictly, the resolution of analysis should be divided into lateral or spatial resolution (*i.e.* from what area on a sample surface or from what volume within a sample a particular signal originates) and also depth resolution (*i.e.* how far below the sample surface a signal originates). Very simplistically a large number of techniques may be classified as either surface or bulk (*i.e.* volume) analytical techniques. Different types of nanostructures, *e.g.* thin films, quantum dots, nanotubes or nanoparticles, *etc.*, require different combinations of lateral and depth resolutions.

In this chapter, I will confine myself to analytical techniques that probe the chemistry and electronic structure of specimens *in situ* within the environment of the Transmission Electron Microscope (TEM). Consequently these

techniques arise from excitation of atoms in a thin sample by focused electrons and generally probe what we may term "bulk" properties (with the obvious proviso that thin sample preparation has not altered the material in any way). The nanoanalytical capability then arises from the fact that both the volume irradiated by the electrons and/or the volume from which the stimulated signal arises is of nanodimensions. Note that I have deliberately omitted any discussion of analysis methods employed in conjunction with Scanning Electron Microscopy (SEM) or Focused Ion Beams (FIB) operated at very low accelerating voltages where the penetration depths and associated interaction volumes can approach nanodimensions.

4.2 Nanoanalysis in the Electron Microscope

The high-energy incident electron beam of the TEM interacts with the sample in a number of ways. Low-angle, coherent elastic scattering of electrons (through 1–10°) occurs *via* the interaction of the incident electrons with the electron cloud associated with atoms in a solid. High-angle, incoherent elastic (back)scattering (through 10–180°) occurs *via* interaction of the negatively charged electrons with the nuclei of atoms. The cross section or probability for elastic scattering varies as the square of the atomic number of the element, whereas inelastic scattering, which provides the analytical signal, generally occurs through smaller angles than for elastic scattering and the cross section varies linearly with atomic number. Inelastic scattering of electrons by solids predominantly occurs *via* four major mechanisms:

(i) Phonon scattering, where the incident electrons excite phonons (atomic vibrations) in the material. Typically the energy loss is <1 eV, the scattering angle is quite large ($\sim 10°$) and for carbon, the mean free path, Λ, between scattering events is ~ 1 μm. This is the basis for heating of the specimen by an electron beam.

(ii) Plasmon scattering, where the incident electrons excite collective, "resonant" oscillations (plasmons) of the valence (bonding) electrons associated with a solid. Here the energy loss from the incident beam is between 5–30 eV and Λ is ~ 100 nm, causing this to be the dominant scattering process in electron–solid interactions.

(iii) Single-electron excitation, where the incident electron transfers energy to single atomic electrons resulting in the ionisation of atoms. The mean free path for this event is of the order of μm. Lightly bound valence electrons may be ejected from atoms and if they escape from the specimen surface, may be used to form secondary electron images in scanning electron microscopy (SEM). Energy losses for such excitations typically range up to 50 eV. If inner-shell electrons are removed, the energy loss can be up to keV. For example, the energy loss required to ionise carbon 1s (*i.e.* K shell) electrons is 284 eV. The energy loss of the incident beam can be used in electron energy-loss spectroscopy (EELS)

analysis and the secondary emissions (*e.g.* X-ray or Auger electron production) produced when the ionised atom relaxes can also be used for analytical purposes (energy dispersive or wavelength dispersive X-ray (EDX/WDX) spectroscopy) as discussed in Section 4.3.

(iv) Direct radiation losses, the principal of which is Bremstrahlung X-ray emission caused by the deceleration of electrons by the solid; this forms the background in the X-ray emission spectrum upon which characteristic X-ray peaks produced by single-electron excitation and subsequent relaxation are exhibited. The Bremstrahlung energy losses can take any value and can approach the total incident beam energy in the limit of full deceleration.

Based on the description above, the two major techniques for chemical nanoanalysis in the transmission electron microscope are both concerned with inelastic interactions and are based on the analysis of either the energy or wavelength of the emitted X-rays (EDX or WDX), or the direct energy losses of the incident electrons (EELS).

4.2.1 General Instrumentation

A comprehensive description of TEM instrumentation is given by Williams and Carter.[1] A schematic diagram of a Conventional TEM (CTEM), consisting of a number of standard electron-optical components, is shown in Figure 4.1. Firstly, the electron source or gun consists of a filament, which is typically tungsten or a ceramic material with a low work function such as lanthanum hexaboride (LaB_6), and a focusing electrode (Wehnelt). Emission of electrons is by either thermionic emission (W or LaB_6) or by field emission (single crystal W) or, in recent machines, a combination of the two (a Schottky emitter). The type of electron gun determines the total emission current and the source size (these two are related by the quantity the brightness: the current density per unit solid angle) as well as the spread of electron energies contained in the electron beam. The electron beam is accelerated towards the anode, typically the acceleration is in a number of stages and the final beam energy, E_0, can be between 100 and 400 keV in most standard commercial instruments.

Two or more electromagnetic condenser lenses demagnify the probe to a size typically between a few micrometres and a few nanometres, the excitation of these lenses controls both the beam diameter and the beam divergence/convergence. For these condenser-lens systems the first condenser (C1 or *spot size*) controls the demagnification of the source, while the second (C2 or *intensity*) controls the size of the spot at the specimen and hence the beam divergence/convergence.

The specimen is in the form of a thin (<100 nm) 3 mm disc of either the material itself or the material supported on an electron transparent film. The specimen is usually inserted into the vacuum of the TEM *via* an airlock and fixed into a side-entry specimen rod that can be translated or tilted (about one or two axes).

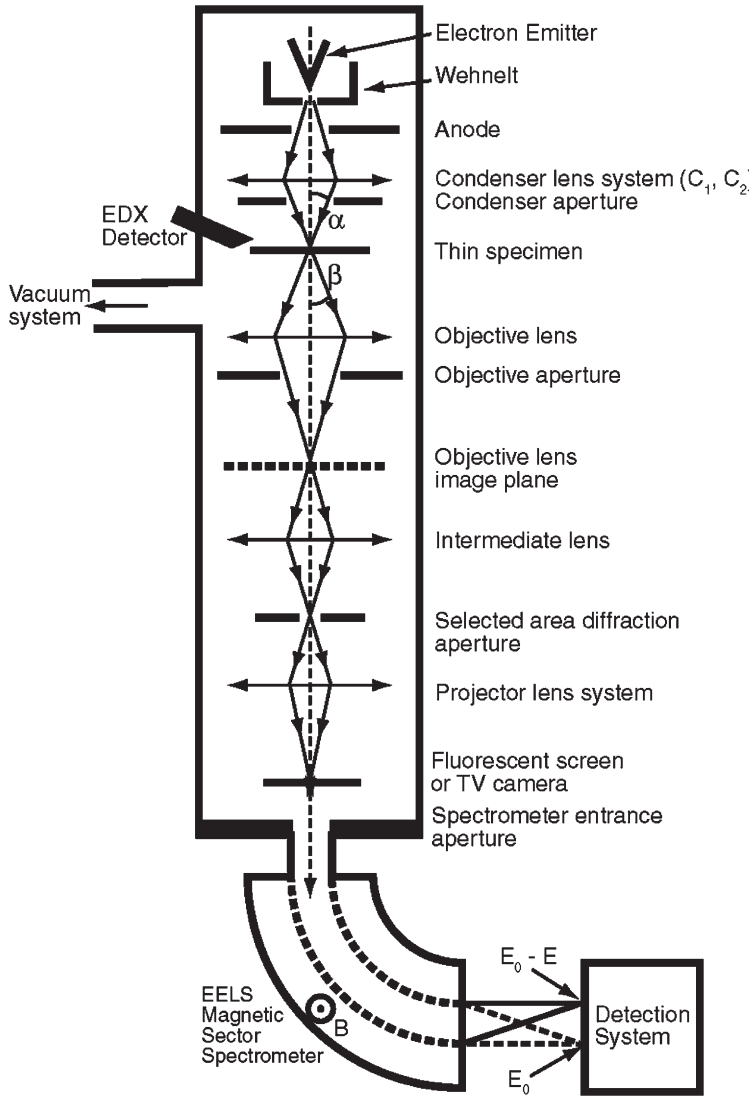


Figure 4.1 Schematic diagram of the layout of an analytical conventional transmission electron microscope (CTEM) fitted with an energy dispersive X-ray (EDX) detector and an electron energy loss (EEL) spectrometer, α and β denote the convergence and collection angles, respectively.

The main electromagnetic objective lens forms the first intermediate, real-space projection image of the illuminated specimen area (in the image plane of the lens) as well as the corresponding reciprocal space diffraction pattern (in the back focal plane of the lens). Here the image magnification relative to the specimen is typically 50–100 times. An objective aperture can be inserted in the back focal plane of the objective lens to limit beam divergence in

reciprocal space of the transmitted electrons contributing to the magnified image. Typically there are a number of circular objective apertures ranging from 10 to 100 μm in diameter.

The projector lens system consists of a first projector or intermediate lens that focuses on either the objective lens image plane (microscope operating in imaging mode) or the back focal plane (microscope in diffraction mode). The first projector lens is followed by a series of three or four further projector lenses – each of which magnify the image or diffraction pattern by typically up to 20 times. The Selected Area Electron Diffraction (SAED) aperture usually lies in the image plane of one of the projector lenses (due to space considerations) and if projected back to the first intermediate image and hence the specimen effectively allows the selection of a much smaller area on the specimen for the purposes of forming a diffraction pattern. The overall microscope system can provide a total magnification of up to a few million times on the electron fluorescent microscope viewing screen or, below this, the camera.

X-rays produced when the electron probe interacts with the specimen are most commonly detected from the incident surface using a low take-off angle Energy Dispersive X-ray (EDX) detector (*i.e.* the detector is approximately in the same plane as the sample, some 20° to the horizontal), so as to allow the detector to be brought close to the sample and also to minimise the predominantly forward-peaked Bremstrahlung background contribution to the X-ray emission spectrum. Even though the detector is inserted to within a few mm of the sample surface, it collects only a small proportion of the isotropically emitted X-ray signal owing to the limited solid collection angle of the detector. Generally, the specimen is tilted towards the detector (typically through *ca.* 15°) so as to provide a clear X-ray trajectory between the irradiated area and the detector. The volume of the specimen that produces X-rays is controlled by the electron probe size (and hence condenser-lens currents) as well as beam broadening within the specimen that increases with (amongst other things) thickness and average atomic number and decreases with microscope accelerating voltage. High take-off angle X-ray detectors also exist and these do not require the specimen to be tilted towards the detector.

Below the microscope viewing screen, the self-contained EELS spectrometer (which almost always possesses a variable entrance aperture itself) and detection system collects the transmitted electron signal that is composed of both elastically and inelastically scattered electrons. In Figure 4.1, α and β are known as the convergence and collection semiangles, respectively. Note certain commercial EELS systems (particularly those initially developed for the purposes of energy-filtered imaging – see Section 4.5) employ an incolumn design with the EELS spectrometer placed between the objective and projector lenses. Further details are given in Section 4.4.1.

An important variant on this transmission imaging system is the Scanning TEM (STEM)[2] where a small diameter (typically a nanometre or less) probe, produced by an electron gun/lens system, is serially scanned in a two-dimensional raster across the specimen. At each point the transmitted beam intensity is measured – thus building up a serial image of the specimen, the

resolution of which is determined primarily by the probe size on the specimen. Two intensities are usually recorded, that falling on both an on-axis bright field (BF) detector that collects electrons that have undergone relatively small angles of scattering (principally undiffracted, Bragg diffracted and inelastically scattered electrons) as well as a high-angle annular dark-field (HAADF) detector that principally collects higher-angle incoherently elastically scattered electrons. STEMs are ideal analytical machines since they easily allow the simultaneous recording of images and X-ray emission spectra, furthermore retraction of the bright-field detector allows electrons to enter an EELS spectrometer whilst still simultaneously recording the HAADF image. Generally, STEMs can collect analytical EDX and EELS data in one of two ways: firstly, by scanning the beam over an area and collecting the signal from the whole scanned area and, secondly, by scanning the beam slowly and recording the analytical signal serially at each point (known as spectrum imaging). Dedicated STEMs employ extremely small probe sizes produced by cold field-emission electron sources that can provide extremely high energy resolution (EELS) and high spatial resolution (EELS and EDX) measurements. There are also a number of hybrid TEM/STEM instruments which can operate in both modes. A schematic diagram of a STEM is shown in Figure 4.2.

4.3 X-Ray Analysis in the TEM

4.3.1 Basics of X-Ray Analysis

In a TEM sample, atoms that have undergone inner-shell ionisation by the primary electron beam and have been promoted into a higher-energy excited state, relax back to their ground state by a process in which electrons from higher energy levels drop into the hole created in the vacant inner shell. This relaxation process results in the release of excess energy corresponding to the difference between the electron energy levels involved in the transition (ΔE) and this occurs *via* the creation of either a low-energy (100–1000 eV) *Auger* electron or, alternatively, an X-ray or visible photon of wavelength $\lambda = hc/\Delta E$, as shown in Figure 4.3. The Auger yield is generally small, except for the case of light elements. However, it is extremely useful for the surface-specific chemical analysis of thick, bulk specimens in SEM-based machines known as scanning Auger microscopes and also in some specialised TEMs. Conversely, the alternative de-excitation mechanism involves X-ray emission and analysis of the energies or wavelengths of the emitted X-rays, which are characteristic of the atom involved and is most commonly used for elemental analysis in the TEM (Jones[3] Williams and Carter[1]).

The energy of the X-ray photon emitted when a single outer electron drops into the inner shell hole is given by the difference between the energies of the two excited states involved. For example, if a K-shell electron is ionised from a molybdenum atom and is replaced by a higher-energy L-shell electron falling into the vacant state the energy difference, ΔE, is 17400 eV which is emitted as a

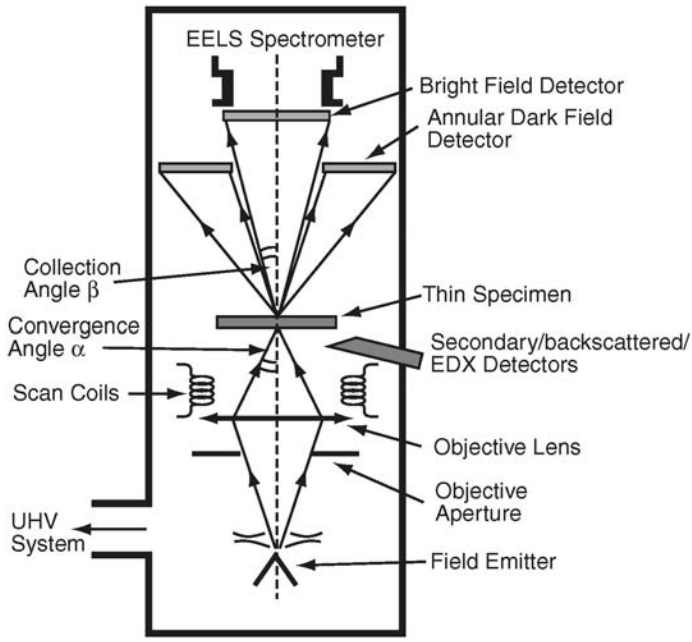


Figure 4.2 Schematic diagram of the layout of a dedicated analytical scanning transmission electron microscope (STEM) fitted with an energy dispersive X-ray (EDX) detector and an electron energy-loss (EEL) spectrometer, α and β denote the convergence and collection angles, respectively.

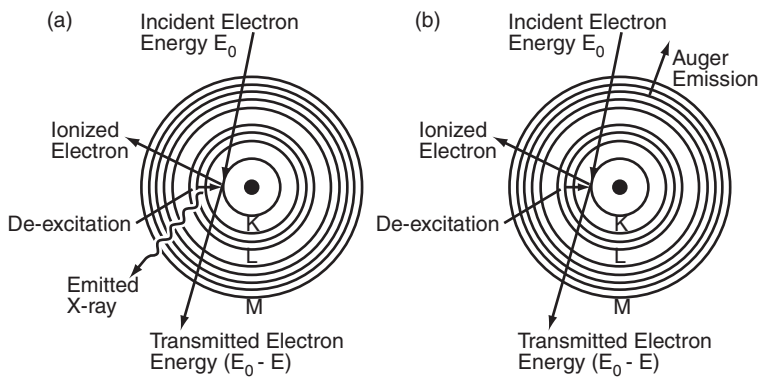


Figure 4.3 De-excitation mechanisms for an atom that has undergone K-shell ionisation by primary electrons: (a) emission of a characteristic Kα X-ray and (b) emission of a KLM Auger electron.

Kα X-ray of Mo; the X-ray wavelength, given by $\lambda = hc/\Delta E$, is 0.071 nm. Alternatively, the inner K-shell hole could be filled by an electron dropping from the higher-energy Mo M-shell, which would result in the emission of a Mo Kβ X-ray of energy 19600 eV and wavelength 0.063 nm. The exact

nomenclature of X-ray emission spectroscopy involves the use of the term K, L, M, N, O, *etc*., which denotes the inner (Bohr) shell that has undergone ionisation. There is a further subsequent term, α, β, γ . . . *etc*., as well as a subscript, 1, 2, *etc*., which denotes the family of lines formed by all the possible electron transitions. At first sight, it would appear that there would be a large set of characteristic X-rays for each atom, since electron transitions between all possible energy levels would appear to be allowed. However, a set of dipole selection rules determine which transitions are observed; more specifically the change in angular momentum (quantum number) between the two electron states involved in transition must be equal to ± 1. This prohibits many of the expected transitions, such as a transition between a 2s state and a 1s state. Numerous general electron microscopy texts such as Jones[3] or Williams and Carter[1] list the most useful X-rays for microanalysis in the TEM.

Due to the well-defined nature of the various atomic energy levels, it is clear that the energies and associated wavelengths of the set of dipole-allowed X-rays that are emitted will have characteristic values for each of the atomic species present in the irradiated specimen volume. By measuring either the energies or wavelengths of the X-rays emitted from the (top) surface of the sample, it is possible to determine which elements are present at the particular position of the electron probe; this is the basis for energy dispersive and wavelength dispersive X-ray analysis (EDX and WDX, respectively). WDX spectrometers use crystal monochromators to disperse the emitted X-ray spectrum in terms of Bragg diffraction angle and hence wavelength. Detectors then move along the arc of a circle centred on the specimen and collect the spectrum serially. EDX detectors collect X-rays in a near-parallel fashion and rely on the creation of electron–hole pairs in a biased silicon or germanium *pn* junction; the number of electron–hole pairs and hence the current is directly proportional to the energy of the incident X-ray. Fast electronics allow separate pulses of X-rays to be discriminated and measured. The current generation of EDX detectors employ ultrathin windows in order to isolate them from the vacuum in the microscope column; these polymer windows permit detection of elements as light as boron, although extensive absorption of low-energy X-rays in the specimen may severely limit accurate quantification of first row elements in the Periodic Table. EDX detectors should be positioned as close to the TEM specimen as possible so as to maximise the number of X-ray counts during the analysis of specific areas in a thin specimen. However, EDX detectors often need to be withdrawn or isolated *via* a shutter mechanism to prevent the detector being flooded with X-rays during the general illumination of larger specimen areas.

Compared to EDX, WDX spectra are slow to acquire but have an increased spectral resolution (typically 10 eV for WDX compared to 100–150 eV for EDX) that gives much enhanced peak-to-background ratios for X-ray lines resulting in a greater sensitivity to all elements, particularly light elements. However, instrumentally, WDX is usually confined to dedicated analytical SEMs known as Electron Probe Microanalysers (EPMA) as well as some very specialised TEMs, whilst EDX detectors are fitted as a common add-on attachment to most SEMs and TEMs. The remainder of this discussion will

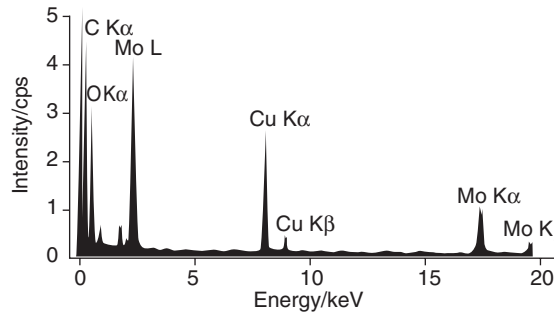


Figure 4.4 Schematic energy dispersive X-ray (EDX) emission spectrum from a thin specimen of molybdenum oxide on a carbon support film. The copper Kα and Kβ X-ray peaks are due to the support grid and TEM specimen holder.

therefore be confined to the technique of EDX. The poor energy resolution of current EDX detectors may give rise to peak overlap problems at low X-ray energies and also effectively precludes any chemical-state information being extracted from any small changes X-ray energies arising from slight shifts in inner-shell binding energies due to differing chemical environments. However, the recent advent of high energy resolution detectors based on microcalorimetry[4] may alter this position in the future.

Figure 4.4 shows a typical electron-generated X-ray emission (EDX) spectrum from a thin plate-like crystal of molybdenum oxide on a carbon support film. The Mo Kα, Kβ and the series of Mo L X-ray lines as well as the O Kα line are clearly observed and are superimposed upon the Bremstrahlung background. The latter X-rays are not characteristic of any particular atom but depend principally on specimen thickness.

4.3.2 Analysis and Quantification of X-Ray Emission Spectra

Identifying which elements are present from an EDX spectrum such as that shown in Figure 4.4, is relatively routine. The positions and relative heights of the various X-ray peaks are either tabulated in references or, more commonly, are stored in a database associated with the software package of the EDX system. Working from the high-energy end of the spectrum, for internal consistency, it is necessary to confirm manually that if the K lines for a particular element are present, then the corresponding L (and possibly M lines) are also evident, *i.e.* if a Mo Kα peak is present then both an Mo Kβ and the series of Mo L X-ray lines must also be observed.

Preprocessing of the spectrum prior to quantification involves both background subtraction and correction for escape peaks due to X-ray absorption in the surface of the detector and sum peaks due to the near simultaneous arrival (and hence detection) of two X-rays at high count rates. This may be performed using deconvolution techniques involving Fourier transformation and filtering,

or alternatively both the background and escape and sum fractions may be modeled mathematically. Such facilities are an integral part of many currently available EDX software packages.

After the spectrum has been processed, the characteristic X-ray peaks are matched, using least squares methods, either to stored spectra or to computed profiles. Such a procedure can also deal with the problem of overlapping peaks. Alternatively, simple integration of areas under the peaks may be performed. Once the characteristic line intensities have been extracted the next step is to turn these into a chemical composition present in the irradiated sample volume.

For quantification it is necessary to know the relevant cross sections for X-ray excitation by the accelerated electrons as well as the absorption characteristics of the window, electrode and dead layer of the X-ray detector, so as to correct the measured X-ray intensities. The most usual approach to tackling this problem is to use a proportionality factor known as a *k-factor*; this may be calculated from first principles or, more usually, it is measured experimentally. The latter approach employs a standard compound of known composition that contains the elements of interest. The basic equation for the analysis of inorganic materials is:

$$C_A/C_B = k_{AB} I_A/I_B \tag{4.1}$$

where C denotes the concentration in wt.% and I denotes the characteristic X-ray peak intensity above the background, for both elements A and B. k_{AB} is the appropriate proportionality or k-factor, also known as a Cliff–Lorimer factor, which is independent of specimen composition and thickness but varies with accelerating voltage. Frequently, k-factors are measured for several element pairs that have a common element such as Si or Fe; k_{AB} is then simply given by $k_{AB} = k_{ASi}/k_{BSi}$. EDX software packages often provide a set of k-factors, obtained for a particular detector system at a given electron beam accelerating voltage, known as a *virtual standards pack*.

If the specimen is very thin (a few tens of nanometres), it is possible to neglect the following phenomena: (i) the differing absorption, within the material itself, of the various characteristic X-rays generated in the specimen as they travel to the specimen exit surface *en route* to the detector, and (ii) fluorescence of one characteristic X-ray by another higher-energy characteristic X-ray. However, if the specimen is thicker, or for quantification of the concentration of a light element present in a heavy element matrix then, for accuracy, we have to correct for these effects, in particular absorption. The absorption correction employs an iterative procedure that initially assumes a starting composition for the specimen based on the uncorrected X-ray intensities. The absorption of different energy X-rays in this specimen *en route* to the detector is then accounted for using a Beer–Lambert-type expression to give a new composition that is subsequently used as input to a further absorption correction. This procedure is repeated until the change in composition falls below some preset level of required accuracy. Fluorescence by characteristic X-rays is only really significant when there are two elements of

interest with similar X-ray energies and the energy of one X-ray is just above the absorption edge of another X-ray (*e.g.* elements close to each other in the periodic table). Fluorescence is made considerably worse by the presence of a large concentration of the fluorescing element combined with a small concentration of the fluoresced species.

With careful measurement and analysis, EDX can detect levels of elements down to 0.1 atom% with an accuracy of roughly ±5%. However, the absorption of low-energy X-rays becomes a severe problem for elements of $Z < 11$ and this can make light-element quantification extremely unreliable without the use of very carefully and individually determined k-factors.

4.3.3 Application to the Analysis of Nanometre Volumes in the S/TEM

Practically, EDX analysis can be performed in one of three differing ways:

(i) a single-point analysis;
(ii) a one-dimensional spectrum image (*i.e.* either a whole or partial EDX spectrum recorded at each probe position along a one-dimensional line) that may be processed to form one dimensional linescans of elemental distributions;
(iii) a two-dimensional spectrum image that may be processed to form two-dimensional maps of elemental distributions.

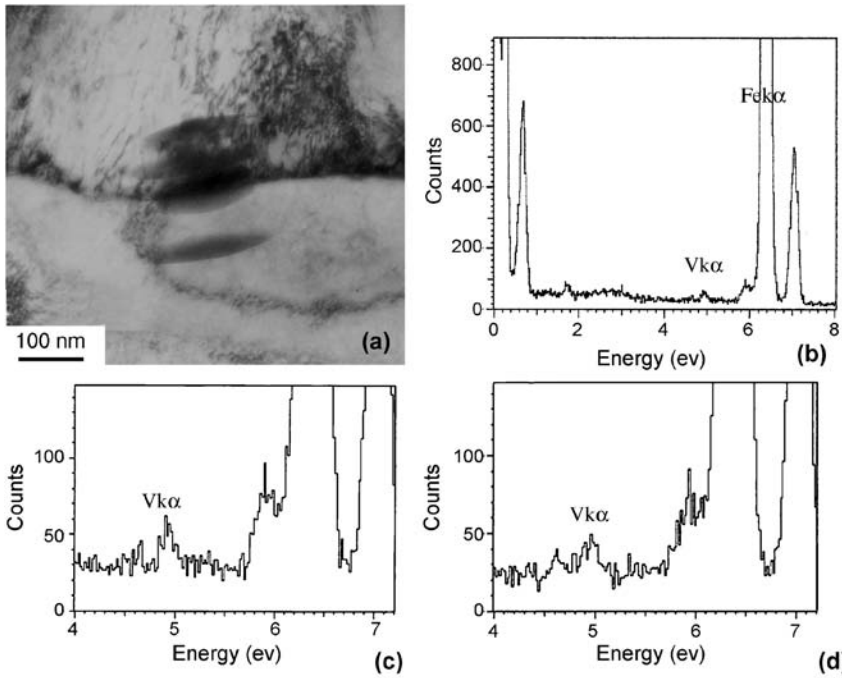
The latter two options are facilitated by the beam scanning capabilities present in a STEM or hybrid TEM/STEM. Figures 4.5 to 4.7 provide some examples of EDX analyses from nanostructures.

In the majority of S/TEMs, typical achievable minimum probe sizes are in the range of a few nanometres to even subnanometre, although it is important to appreciate that reducing the probe size also reduces probe current and hence decreases the emitted X-ray signal. A specimen volume irradiated with a focused electron probe in the TEM or STEM will emit characteristic X-rays isotropically of which typical EDX detectors collect roughly only 1%. This makes the detection process highly inefficient which often necessitates quite long recording times (*e.g.* minutes) in order to achieve good statistics in EDX spectra obtained from small volumes.

In terms of spatial resolution, an additional consideration is that there may be some broadening of the beam as it passes through the thin TEM specimen. For a foil of thickness, t (nm), of density, ρ (g cm^{-3}), and atomic weight, A, the beam broadening (in nanometres) can be approximated by

$$b = 0.198\,(\rho/A)^{1/2}(Z/E_0)t^{3/2} \tag{4.2}$$

where E_0 is the incident beam energy in keV and Z is the atomic number (Jones[2]). At 100 keV a 100 nm thick sample gives a typical beam broadening of



Scanning transmission electron microscopy energy dispersive X-ray (STEM-EDX) analysis results: bulk V content is 0.053 (mole fraction) for 0.5% V steel

Number	$\{V\}_b$	$\{V\}_\alpha$	Scanned Area Width nm	Γ_b^V atoms / nm^2	Monolayer (V)
1	0.74	0.61	75	8.77	0.45
2	0.83	0.68	37.5	4.64	0.24
3	0.55	0.44	37.5	3.57	0.18
4	0.68	0.40*	20	4.98	0.26
5	0.84	0.52*	20	5.62	0.29
6	0.63	0.52*	20	1.93	0.10
7	0.74	0.36*	20	6.75	0.35
8	0.74	0.70*	20	0.81	0.04
9	0.93	0.55*	20	6.77	0.35
10	0.58	0.43*	20	2.71	0.14
11	0.81	0.56*	20	4.49	0.23
Average	0.73	0.52		4.64	0.24

* Average of analysis of ferrite plates on both sides of ferrite phase boundary.

Figure 4.5 STEM/EDX point analysis of the vanadium content at a ferrite grain boundary in a steel of composition 0.1C-0.27Si-1.2Mn(wt%) using a letter-box of length ~180 nm and width 20 nm visible as contamination stripes on (a) TEM BF image of the boundary. (b) EDX spectrum collected from grain boundary; expanded section of the EDX spectrum showing the V Kα X-ray peak (c) on and (d) off the grain boundary. The table summarises the analytical results for interfacial vanadium segregation, for more details see K. He and D.V. Edmonds, "Formation of acicular ferrite and the influence of vanadium alloying", *Mater. Sci. Technol.*, 2002, **18**, 289–296.

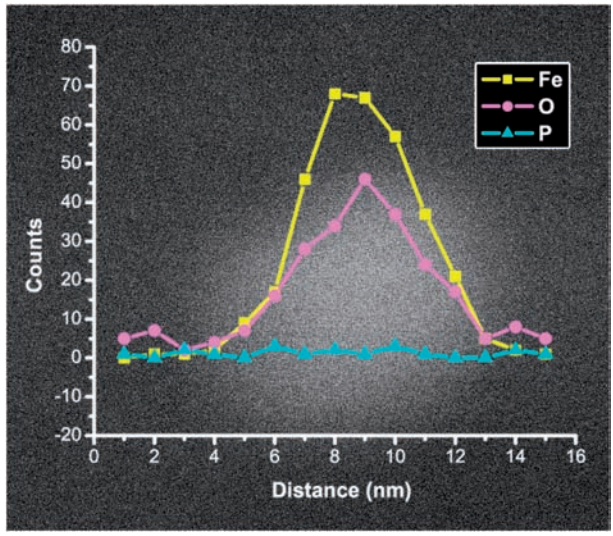


Figure 4.6 STEM/EDX linescans using the Fe Kα, O Kα and P Kα X-ray peaks across a mineral core of a single ferritin iron storage protein *in situ* within biological tissue; the linescans are superimposed on the corresponding STEM HAADF image of the sample region.

the order of 10 nm. This beam broadening will limit the ultimate (lateral) spatial resolution for the analysis of nanoparticles and nanostructures using small probes, particularly for EDX analysis where we collect all X-rays produced isotropically within the beam-broadened volume. EDX maps of multilayers and nanoscale precipitates have demonstrated resolutions of, at best, a few nanometres in ultrathin sample areas (Figure 4.7).

Another essential aspect of EDX spectrum acquisition is the collection of X-rays solely from the area of interest on the specimen and not elsewhere. Problems can arise when electrons from outside the primary beam as well as X-rays generated in the TEM illumination system produce X-rays away from the region of analysis and even from areas that are not part of the specimen. It is always necessary to remove the TEM objective aperture because otherwise it will constitute a large source of extraneous and unwanted X-rays. Stray radiation may be also minimised by tilting the sample towards the detector, as discussed in Section 4.2.1, and by the use of spray apertures beneath the condenser aperture as well as coating the lower part of the objective-lens polepiece and specimen stage with low atomic number materials. Special analytical TEM sample holders are most often used for quantitative EDX analysis. Instead of the normal brass holders that give rise to characteristic Cu and Zn X-rays, the near-sample region in the holder is constructed of a low atomic number material – often graphite or beryllium, which reduces absorption and produces X-rays that generally do not interfere with the analysis. If the sample is supported on a grid and support film, these materials can also often

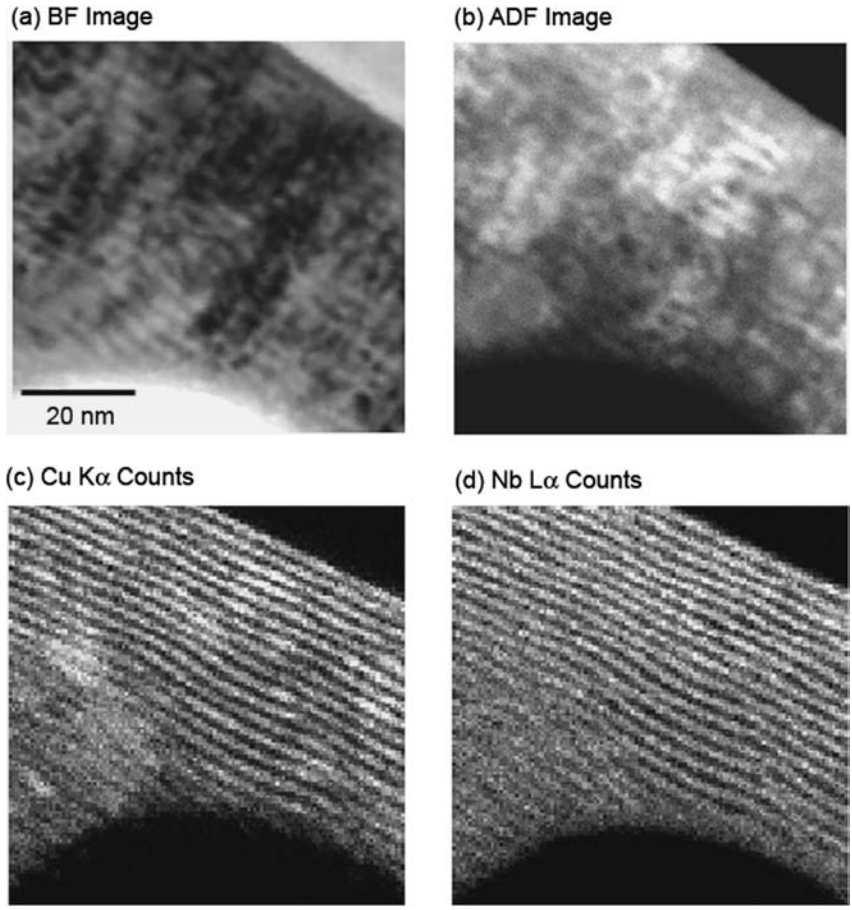


Figure 4.7 STEM (a) Bright-Field (BF) and (b) Annular Dark-Field (ADF) images from a 1.5 nm Cu-Nb multilayer. EDX Elemental maps acquired using (c) Cu Kα peak and (d) Nb Lα peaks. Reproduced from V. J. Keast, A. Misra, H. Kung and T. E. Mitchell, Compositional Distributions in Nanoscale Metallic Multilayers Studied Using X-ray Mapping, *J. Mater. Res.*, 2001, **16**, 2032–2039.

be chosen so as to avoid problems of interference with the analysis of elements present in the specimen. Since the principal source of spurious X-rays is often thicker regions of the specimen surrounding the thin area of analysis, the use of uniformly thin samples is desirable. Some of these extraneous effects mentioned above may be circumvented by displacing the beam off the edge of the specimen and collecting a *hole spectrum* under the same conditions and counting time. This may then be subtracted from the original spectrum before subsequent processing. Higher take-off angle X-ray detectors generally have lower hole counts and have less problems in terms of self-absorption and self-fluorescence in the specimen.

4.3.4 Related Photon Emission Techniques in the TEM

Cathodoluminescence (CL) of visible and ultraviolet photons is generated by the recombination of electron–hole pairs produced by electron-beam-induced ionisation of valence electrons in a semiconducting or insulating sample. The exact CL energies (and therefore wavelengths) depend on the width of the bandgap of the material. Any changes in temperature, crystal structure (*e.g.* different polymorphs), impurity levels or defect concentrations will modify the bandgap and thus the CL wavelength. Spectroscopic analysis of the CL emission can therefore provide information on the electronic structure of the sample. The emitted photon intensity is low and therefore low scan rates and high probe currents (large probe diameters) are required, limiting the resolution to 1–10 μm in the SEM – the most common environment. However, in the STEM the resolution can be improved considerably,[5] making the technique viable for nanostructural analysis of single semiconducting nanoparticles or quantum dots.

4.4 Basics of EELS

Electron Energy-Loss Spectroscopy (EELS) in a S/TEM involves analysis of the inelastic scattering suffered by the transmitted electron beam *via* measurement of the electron-energy distribution. The various inelastic scattering processes have already been outlined in Section 4.2 and each provides valuable information on the sample area that is irradiated by the electron probe. The transmitted electrons are dispersed according to their energy loss using an electron spectrometer, as shown in Figure 4.8, and recorded in parallel using a scintillator and either a one- or two-dimensional photodiode array detection system. This technique can provide high-resolution elemental analysis and mapping as well as a means of determining the local electronic structure, in crude terms the local chemical bonding, present in a nanoparticle or a precipitate, or at an interface or defect in an extended solid.

4.4.1 Instrumentation for EELS

There have been a number of types of electron spectrometer used in conjunction with TEMs. The so-called Ω filter, though commercially available, usually forms an integral part of the microscope column and is mainly employed in dedicated energy-filtered imaging TEMs. The sector magnet spectrometer, shown in Figure 4.8, can be added below an existing TEM column and is therefore considerably more common. Hence, the following discussion is limited to sector magnets although more detail concerning incolumn filters may be found in the references.[6]

A typical sector magnet consists of a homogeneous magnetic field normal to the electron beam. This causes electrons of a given kinetic energy and hence velocity, to follow trajectories that are essentially arcs of circles. Usually, for

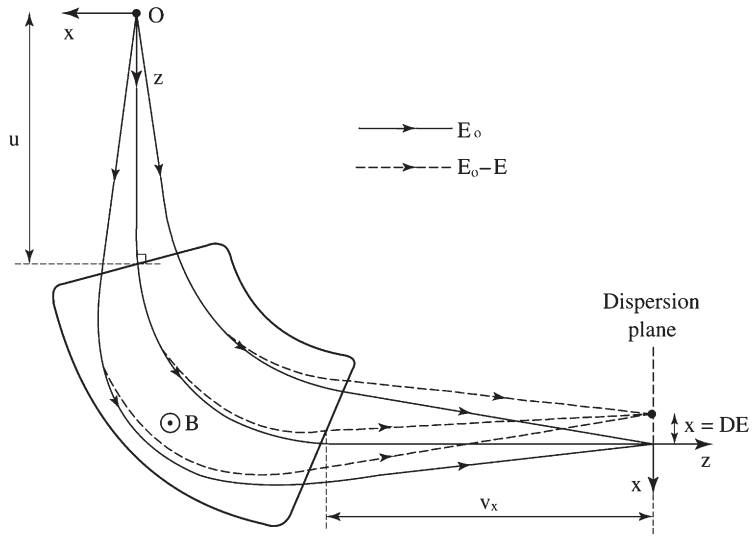


Figure 4.8 Schematic diagram showing the first-order focusing properties and dispersion for a 90° magnetic-sector EEL spectrometer.

mechanical convenience the electrons are bent through a right angle and the magnetic force causes electrons of different energy (and therefore energy loss) to emerge from the spectrometer spatially dispersed. The spectrometer magnetic field is provided by two parallel polepieces with two current-carrying coils providing the excitation. In practice, the polepieces and coils are at atmospheric pressure, while an evacuated drift or flight tube, connected to the microscope vacuum system, passes between the polepieces. An offset voltage may be applied to this drift tube that will alter the kinetic energy of all the electrons and rigidly shift the whole energy-loss spectrum across the detector system. In fact the spectrometer, together with a set of manually adjustable quadrupole and sextupole lenses, has a double focusing action that results in the formation of a line spectrum (each energy loss is separately focused to a point) at the spectrometer image plane. The spectrometer dispersion is the displacement of the focus per unit change in electron energy and is typically a few μm eV^{-1}. The energy resolution of the resultant EEL spectrum is determined not only by the energy spread of the electron source, but also by the resolution of the spectrometer itself, the effect of external interference as well as by the resolution of the detection system. To minimise aberrations, the entrance angle of the spectrometer must be limited, which is achieved by use of an adjustable spectrometer entrance aperture (SEA) at the position where the spectrometer bolts onto the TEM camera chamber or, alternatively, by the use of an objective aperture in the microscope to limit the collection angle of scattered electrons. The electron-beam path must be also adequately screened from oscillating magnetic fields from both the mains supply and from high-frequency oscillations in the microscope; besides enclosing the camera chamber and

spectrometer in a soft magnetic material, this may be further achieved by adding an additional ac compensation signal into the spectrometer excitation coils.

The magnetic sector spectrometer can be experimentally coupled to a microscope in a number of ways. The simplest case is that of a dedicated STEM with the specimen in field-free space, as shown in Figure 4.9(a). Here, the electron probe is focused onto the specimen with a semiangle of convergence, α typically in the range 2–15 mrad. In most dedicated STEMs, there are no lenses after the specimen and the specimen is in the object plane of the spectrometer. The area irradiated by the probe becomes the spectrometer object and its size is determined by the size of the electron probe, which can be as low as 0.1 nm for recent instrumentation and this effectively determines the spatial resolution for analysis. The collection angle is either defined by the spectrometer entrance aperture or alternatively, in a dedicated STEM, a postspecimen collector aperture may be employed. If the magnetic field due to the STEM objective lens is strong and immerses the specimen then there will be magnetic induction after the specimen and this will act as a postspecimen lens, giving rise to the situation in Figure 4.9(b). This is the same as in many hybrid TEM/STEM instruments operated in STEM mode. Here, the angular range (collection angle), β, of electrons leaving the specimen can be compressed into a spectrometer of smaller angular acceptance by forming a suitably magnified (virtual) image of the specimen which acts as the spectrometer object. β will then be determined by the spectrometer entrance aperture and the camera length (effectively the magnification of the diffraction pattern).

By using additional lenses after the specimen, it becomes possible to vary the angular compression over a wide range that is similar to a conventional TEM operating in diffraction mode as shown in Figure 4.9(c). Here, the spectrometer entrance aperture is close to the microscope viewing screen on which the diffraction pattern is displayed. The spectrometer object is then the demagnified image of the specimen in the final projector lens and so the spectrometer is said to be *image coupled* (even though, when lowered, the TEM viewing screen would display a diffraction pattern). The angle of acceptance at the specimen, β, is proportional to the radius of the selected spectrometer entrance aperture divided by the camera length of the diffraction pattern displayed on the viewing screen. The area from which the spectrum is taken is limited either by the area of the illuminated region or by the size of the selected area diffraction aperture (which can be as low as 100 nm in some instruments). However, both spherical and more importantly, chromatic aberrations in both the TEM objective and projector lenses can affect the precision of the area selected using the SAED aperture and for high spatial resolution analysis, limiting the area illuminated using a highly convergent probe (*e.g.* microprobe or nanoprobe mode) is generally the preferred option.

The alternative TEM-based mode for EELS is *diffraction coupling* as illustrated in Figure 4.9(d). This is the experimental arrangement used in EFTEM with a postcolumn imaging filter. Here, the microscope is operated in image mode and the spectrometer object is then the diffraction pattern at the

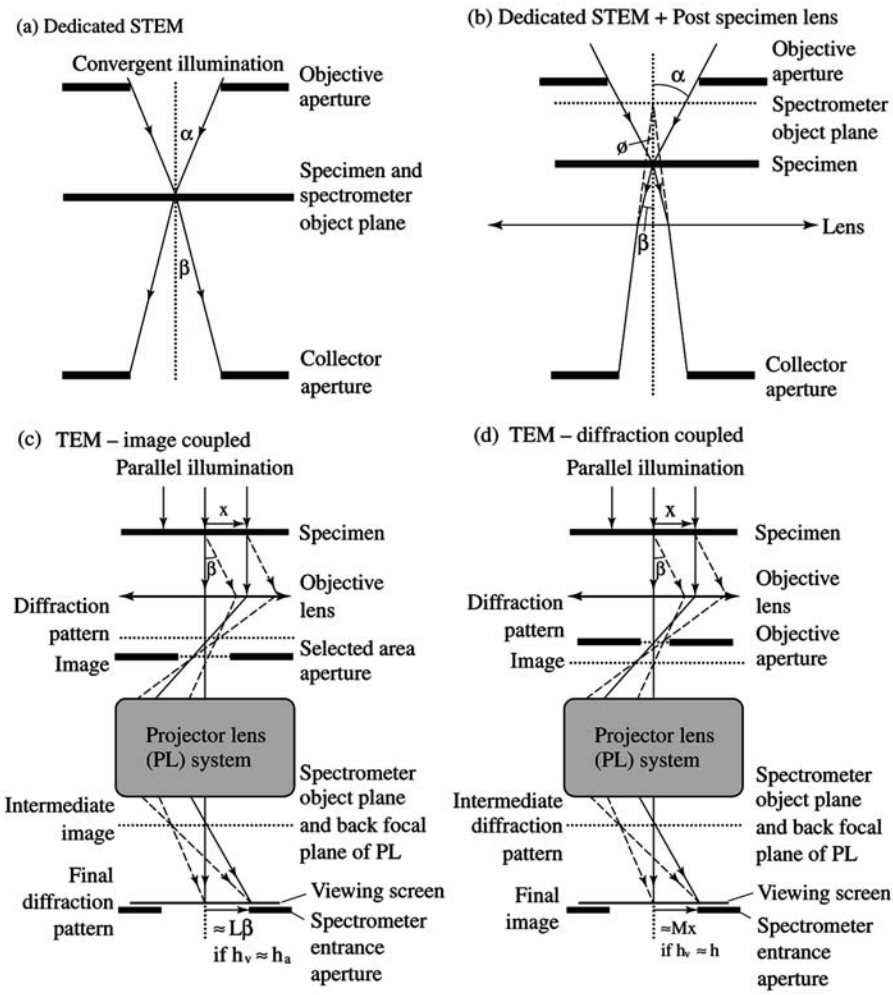


Figure 4.9 Schematic ray diagram of coupling between a magnetic prism spectrometer and: (a) dedicated STEM; (b) dedicated STEM with a post specimen lens; (c) TEM – image coupled and (d) TEM – diffraction coupled. The diagram indicates the parameters that determine the area of analysis and the collection angle.

crossover of the final projector lens. The diameter of the analysis area selected on the specimen is usually controlled by the part of the image selected by the spectrometer entrance aperture and this depends inversely on the image magnification. However, if the irradiated area in the image is smaller than that selected by the spectrometer entrance aperture, this will ultimately control the area of analysis but may cause problems due to chromatic aberrations affecting the collection efficiency of the system at different energy losses. The collection semiangle can normally only be defined accurately by inserting the

objective aperture – this can be calibrated for the range of objective apertures in a TEM using diffraction patterns from a known crystalline specimen. However, use of the objective aperture may present problems when performing simultaneous EELS and EDX measurements owing to the stray X-ray signal generated from the aperture. Because of the effects of aberrations in the postspecimen imaging lenses, recording an EELS spectrum without an objective aperture may cause problems in terms of both spatial and energy resolution.

The exact operating mode of the microscope and hence coupling to the spectrometer will depend on the nature of the problem being investigated, however, some general guidelines are discussed in the references.[6,7] For EELS analysis at nanometre spatial resolution, employing a STEM is the most straightforward option, whilst the most usual mode of operation in TEM is to employ a focused probe and to work in the diffraction coupled mode.

Having obtained an EEL spectrum in the dispersion plane of the spectrometer, it is necessary to record the intensity at each particular energy loss as a sequence of digital values over a large number of detector channels (typically 1024 or 2048). Early systems employed a single-element detector, where the spectrum was sequentially scanned (using a progressively ramped drift tube voltage) across the detector and recorded in a *serial* fashion, whereas most modern systems use a multielement detector where the whole spectrum is recorded in parallel. Modern one dimensional linear photodiode arrays (PDAs) and two-dimensional arrays of charge coupled devices (CCDs), the latter identical to those used for the digital recording of TEM images, are both well suited to the task of parallel recording. These systems rely on the measurable discharge (over a certain integration time) of a large array of self-scanning, cooled silicon diodes by photons created by the direct electron irradiation of a suitable scintillator. This signal is superimposed on that due to thermal leakage currents as well as inherent electronic noise from each individual diode and together these are known as *dark current* that can be subtracted from the measured spectrum; gain variations and cross-talk between individual diode elements can also be measured and corrected for. One major problem with such recording arrays is their limited dynamic range, which may require at least two or three suitable integration periods. Hence, separate measurements, to record accurately and efficiently the entire EEL spectrum (which can show a total variation in signal of up to 10^8) is collected, stored and then processed on a multichannel analyser software package.

4.4.2 Basics of the EEL Spectrum

The various energy losses observed in a typical EEL spectrum are shown schematically in Figure 4.10(a), which displays the scattered electron intensity as a function of the decrease in kinetic energy (the energy loss, E) of the transmitted fast electrons and represents the response of the electrons in the solid to the electromagnetic disturbance introduced by the incident electrons.

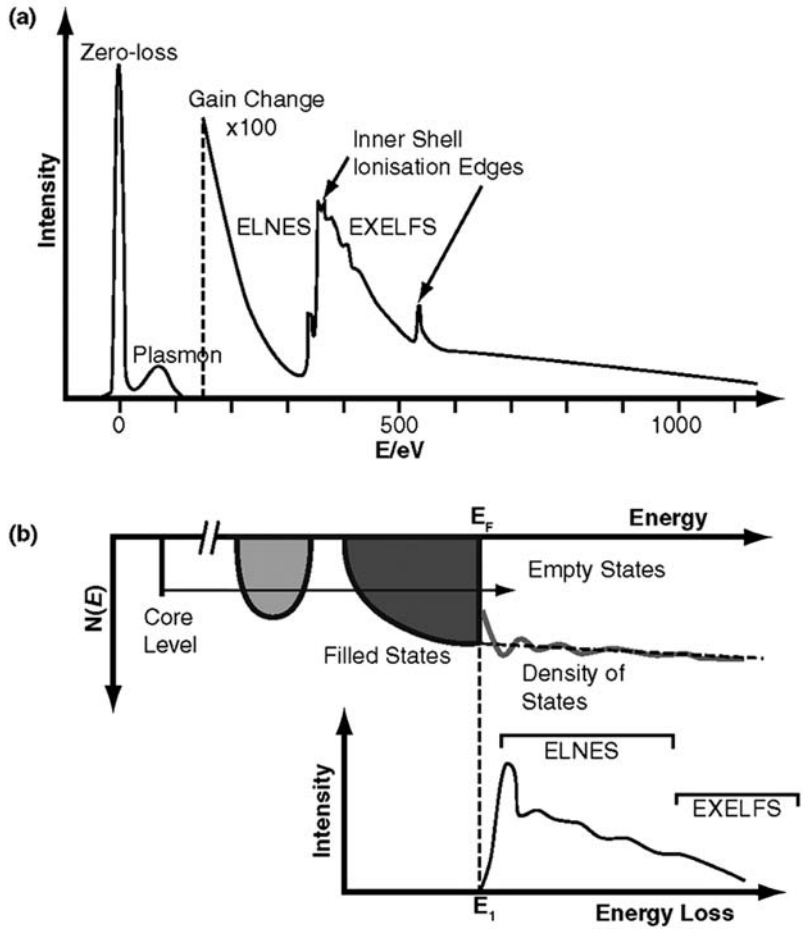


Figure 4.10 Schematic diagram of (a) A general EELS spectrum (with a linear intensity scale and a gain change at high energy loss) showing all of the general observable features and (b) The ELNES intensity that reflects the unoccupied DOS above the Fermi level.

Note that the intensity at 2000 eV is typically eight orders of magnitude less than that at the zero-loss peak and therefore, for clarity, in Figure 4.10(a) a gain change has been inserted in the linear intensity scale at 150 eV. In a specimen of thickness less than the mean free path for inelastic scattering (roughly 100 nm at 100 keV), by far the most intense feature in the spectrum is the zero-loss peak at 0 eV energy loss that contains all the elastically and quasielastically (*i.e.* vibrational- or phonon-) scattered electron components. Neglecting the effect of the spectrometer and detection system, the Full Width Half-Maximum (FHWM) of the zero-loss peak is usually limited by the energy spread inherent in the electron source. In a TEM, the energy spread will generally lie between $\sim$0.3–3 eV, depending on the type of emitter (cold field

emission < Schottky emitter < LaB_6 thermionic emitter < tungsten thermionic emitter), and this parameter often determines the overall spectral resolution. In recent years, the use of electron monochromators has improved achievable resolutions to $\sim$0.1 eV.[8]

The *low-loss* region of the EEL spectrum, extending from 0 to about 50 eV, corresponds to the excitation of electrons in the outermost atomic orbitals that are delocalised in a solid due to interatomic bonding and may extend over several atomic sites. This region therefore reflects the solid-state character of the sample. The smallest energy losses (10–100 meV) arise from phonon excitation, but these are usually subsumed in the zero-loss peak. The dominant feature in the low-loss spectrum arises from collective, resonant plasmon oscillations of the valence electrons. The energy of the plasmon peak is governed by the density of the valence electrons, and its width by the rate of decay of the resonant mode. In a thicker specimen ($\geq$100 nm) there are additional peaks at multiples of the plasmon energy, corresponding to the excitation of more than one plasmon; the intensities of these multiple plasmon peaks follow a Poisson statistical distribution. A further feature in the low-loss spectra of insulators are peaks, known as interband transitions, which correspond to the excitation of valence electrons to low-energy unoccupied electronic states above the Fermi level. These single-electron excitations may lead to a shift in the energy of the plasmon resonance. The low-loss region is used mainly to determine the relative (or absolute) specimen thickness (*via* the logarithm of the ratio of the total spectral intensity to the zero-loss intensity) and to correct for the effects of multiple inelastic scattering when performing quantitative microanalysis on thicker specimens (see Section 4.4.3). In a more detailed analysis, the overall shape of the low-loss region may be related to the dielectric response function of the material, which allows a correlation with optical measurements including reflectivity and bandgap determination in insulators and semiconductors.

The *high-loss region* of the EEL spectrum extends from about 50 eV to several thousand electron volts and corresponds to the excitation of electrons from localised orbitals on a single atomic site to extended, unoccupied electron energy levels just above the Fermi level of the material. This region therefore reflects the atomic character of the specimen. As the energy loss progressively increases, this region exhibits steps or edges superimposed on the monotonically decreasing background intensity that usually follows an inverse power law, $I = AE^{-r}$. These edges correspond to excitation of inner-shell electrons and are therefore known as ionisation edges. The various EELS ionisation edges are classified using the standard spectroscopic notation similar to that employed for labeling X-ray emission peaks; *e.g.*, K excitation for ionisation of 1s electrons, L_1 for 2s, L_2 for $2p_{1/2}$, L_3 for $2p_{3/2}$ and M_1 for 3s, *etc*. The subscript, in for example $2p_{1/2}$, refers to the total angular momentum quantum number, j, of the electron that is equal to the orbital angular momentum, l, plus the spin quantum number, s. For a 2p electron, $l = 1$, and this can couple to the spin of the electron in one of two ways, *i.e.* $j = l + s = 1 + 1/2 = 3/2$ (L_3) or $j = 1–1/2 = 1/2$ (L_2).

Since the energy of the ionisation edge threshold is determined by the binding energy of the particular electron subshell within an atom – a characteristic value, the atomic type may be easily identified with reference to a tabulated database. The signal under the ionisation edge extends beyond the threshold, since the amount of kinetic energy given to the excited electron is not fixed. The intensity or area under the edge is proportional to the number of atoms present, scaled by the cross section for the particular ionisation process, and hence this allows the technique to be used for quantitative analysis. EELS is particularly sensitive to the detection and quantification of light elements ($Z < 11$) as well as transition metals and rare earths. Problems can arise if the sample thickness is greater than the mean free path for inelastic scattering, in this case multiple inelastic scattering will occur. This will significantly increase the plasmon intensities, leading to an increase in the background contribution making it difficult to identify the presence of edges in a spectrum, a further effect is the transfer of intensity away from the edge threshold towards higher energy losses due to the increase of double scattering events involving a plasmon excitation followed by an ionisation event or *vice versa*.

It is possible to remove this multiple inelastic scattering contribution (at the expense of some added noise) from either the whole EEL spectrum or a particular spectral region by Fourier transform deconvolution techniques. Two techniques are routinely employed, one, known as the Fourier-log method, requires the whole spectrum over the whole dynamic range as input data. This large signal dynamic range can be a problem with data recorded in parallel. The second, known as the Fourier-ratio method, requires a spectrum containing the feature of interest (*i.e.* an ionisation edge) that has had the preceding spectral background removed. A second spectrum containing the low-loss region from the same specimen area is then used to deconvolute the ionisation edge spectrum.

If electrons are scattered *via* inelastic collisions with the K-shell electrons of free atoms (*e.g.* such as in the case of a gas) the core-loss edges are sharp, sawtooth-like steps displaying no features. Other core-loss excitations in free atoms display a variety of basic edge shapes that are essentially determined by the degree of overlap between the initial- and final-state electronic wavefunctions that may be determined by quantum-mechanical calculations of free atoms. In solids, however, the unoccupied electronic states near the Fermi level (the highest occupied electronic energy level) may be appreciably modified by chemical bonding leading to a complex Density Of States (DOS) and this is reflected in the electron-energy loss near-edge structure (ELNES) that modifies the basic atomic shape within the first 30–40 eV above the edge threshold (Figure 4.10(b)). The ELNES effectively represents the available electronic states above the Fermi level, in the environment of the atom(s) being ionised, and hence gives information on the local crystallographic structure and chemical bonding.[9] Although the exact interpretation is somewhat complex, ELNES can be modeled using electronic-structure calculations of the unoccupied DOS using, for instance, Density Functional Theory (DFT) calculations. However, experimentally this DOS is usually a local DOS

representative of the environment(s) of the particular atom undergoing ionisation and also usually a DOS of a particular angular momentum symmetry (*i.e.* s, p, d, f, *etc.*) relative to the angular momentum of the ionised inner shell owing to the fact that the optical dipole selection rule operates for the electronic transitions observed in spectra collected with a small collection aperture and hence scattering angle. Besides providing a direct means of probing electronic structure, ELNES also allow the determination of simpler chemical concepts such as local coordinations and valence states of atomic species from regions of nanodimension *via* the use of "spectral fingerprints". Beyond the near-edge region, superimposed on the gradually decreasing tail of the core-loss edge, a region of weaker, extended oscillations are observed, known as the extended energy-loss fine structure (EXELFS). As in the more commonly known X-ray synchrotron technique of EXAFS, the period of the oscillations may be used to determine bond distances, while the amplitude reflects the coordination number of the particular atom. In terms of scattering theory, the main distinction between the ELNES and EXELFS regions lies in the fact that the low kinetic energy of the ejected electron in the near-edge region means that it samples a greater volume (the mean free path is large) and multiple elastic scattering occurs – so providing geometrical information on bonding arrangements. In the EXELFS regime, the higher kinetic energy results in predominantly single elastic scattering of the ejected electron thus giving short-range information.

If the transmitted electrons are collected over small scattering angles (corresponding to small momentum transfers upon collision), then as discussed above the (optical) dipole selection rules apply to the various transitions observed at the core-loss edges, as is the case for X-ray emission. This limits the observed electron transitions to those in which the angular momentum quantum number, l, changes by ± 1. This results in different edges of the same element probing different symmetries of the final state, *i.e.* a K-edge will probe the unoccupied p-like DOS, whereas an $L_{2,3}$-edge will probe the unoccupied s- and d-like DOS. At larger scattering angles the dipole selection rules break down and other transitions are observed. At very large angles, a new regime is encountered in which the electrons in the sample may be regarded as if they were free and a hard-sphere collision occurs with an associated large momentum transfer resulting in the electron Compton profile. The width of this feature represents a Doppler broadening of the energy of the scattered electrons due to the initial state momentum of the electrons in the sample and can, in principle, also give bonding information.

4.4.3 Quantification of EELS – The Determination of Chemical Composition

After some initial data processing, such as dark-current subtraction, gain correction (which are both a function of the response of the detection system outlined in Section 4.4.1) and also possibly deconvolution to remove the effects

of multiple inelastic scattering in thicker specimen regions, in order to quantify the elemental analysis it is necessary to measure the intensities under the various edges. This is achieved by fitting a background (in this case a power-law $A.E^{-r}$) to the spectrum immediately before the edge. This is then subtracted (Figure 4.11) and the intensity is measured in an energy window, Δ, which begins at the edge threshold and usually extends some 50 to 100 eV above the edge. The next step is to compute the inelastic partial cross section, σ, for the particular inner-shell scattering event under the appropriate experimental conditions, *i.e.* $\sigma(\alpha, \beta, \Delta, E_o)$; where the bracketed terms in italics simply represent the variables on which the partial cross section depends. This partial cross section is calculated for the case of a free atom using simple hydrogenic or Hartree–Fock–Slater wavefunctions and is generally an integral part of the analysis software package (Figure 4.11). The measured edge intensity is normalised (*i.e.* divided) by the partial cross section so that either different edge intensities can be compared, or the intensity can be directly interpreted in terms of an atomic concentration within the specimen volume irradiated by the electron probe. For the latter case, the measured edge intensity is divided by both the partial cross section and the combined zero loss and low loss intensity measured over the same energy window, Δ; the result is usually expressed in terms of an areal density in atoms/nm^2 multiplied by the specimen thickness.

Apart from the case of light elements as well as many of the transition metal and rare earth elements, detection limits for EELS are generally worse than

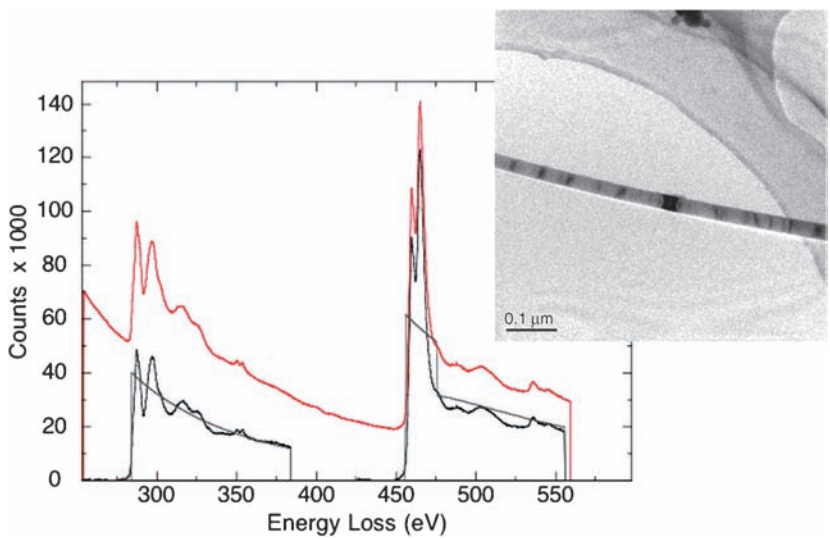


Figure 4.11 EELS from a single-crystal titanium carbide nanorod (TEM image inset) shows the EELS carbon K-edge at ca. 282 eV and the titanium $L_{2,3}$-edge at ca. 455 eV both before (red) and after (black) subtraction of the extrapolated background intensities, an integration window of 100 eV is shown. The background subtracted EELS edges are superimposed on the theoretical free-atom partial cross sections.

those for EDX and typically lie between 0.1 and 1 atom%. This is principally due to the limited spectral range of the technique relative to EDX as well as the steep and intense background signal upon which the ionisation signal lies. Meanwhile, analytical accuracies in elemental quantification usually lie in the range 5–10%. An example of relative EELS elemental quantification is provided in Figure 4.11, for the case of single-crystal titanium carbide nanorods fabricated from multiwalled carbon nanotubes (prepared by catalytic chemical vapour deposition) and a LiCl, KCl and KF molten salt system containing titanium metal powder.[10] Figure 4.11 shows the EELS carbon K-edge at ca. 282 eV and the titanium $L_{2,3}$-edge at ca. 455 eV both before and after subtraction of the extrapolated background intensities, an integration window of 100 eV is shown. This data was recorded from a single-crystal nanorod (with a diameter 35 nm and length some tens of micrometres as shown inset in Figure 4.11) using a $\sim$200 nm SAED aperture and the TEM operating in diffraction mode. Following background subtraction, integration and normalisation of the integrated intensities by the relevant partial cross sections calculated for the experimental conditions employed, this gives a Ti/C atomic ratio of 0.80 ± 0.12 indicating that this particular crystal is somewhat substoichiometric.

An example of absolute EELS elemental quantification of an EELS spectrum image (Section 4.5) is shown in Figure 4.12, for the case of a 6 nm diameter inorganic core of ferritin molecules contained *in situ* within a microtomed thin section of human liver biopsy. Ferritin cores contain iron in the form of a hydrated iron oxide and are surrounded by a protein shell, the overall ferritin molecule allows iron to be temporarily stored during the human body's daily iron cycle. One interesting question is the capacity of such ferritin molecules for iron storage. STEM spectrum imaging using a 1 nm probe over individual cores allows both the low-loss spectrum and the iron $L_{2,3}$-edge to be recorded and the absolute number of iron atoms in the core to be determined. For accuracy, in this study, the Fe $L_{2,3}$-edge partial cross section was independently calibrated using a separate set of measurements on individual cube-octahedral nanoparticles of maghemite (γ-Fe_2O_3) that have a known shape and density. Overall, the results obtained suggest ferritin cores can contain between a few hundred and around 2000 Fe atoms, which is substantially lower than the 4500 atoms previously thought.[11]

4.4.4 Determination of Electronic Structure and Bonding

4.4.4.1 Low-loss Measurements

Generally the low-loss region is dominated by the bulk plasmon excitation and this may be thought of as a resonant oscillation of the valence electron gas of the solid (as pictured in the Drude–Lorentz model for metals) stimulated by the fast incident electron. Within the free-electron model of solids an expression for the bulk plasmon energy, E_p, is given by

$$E_p = h/2\pi[Ne^2/(m_e\varepsilon_o)]^{1/2} \tag{4.3}$$

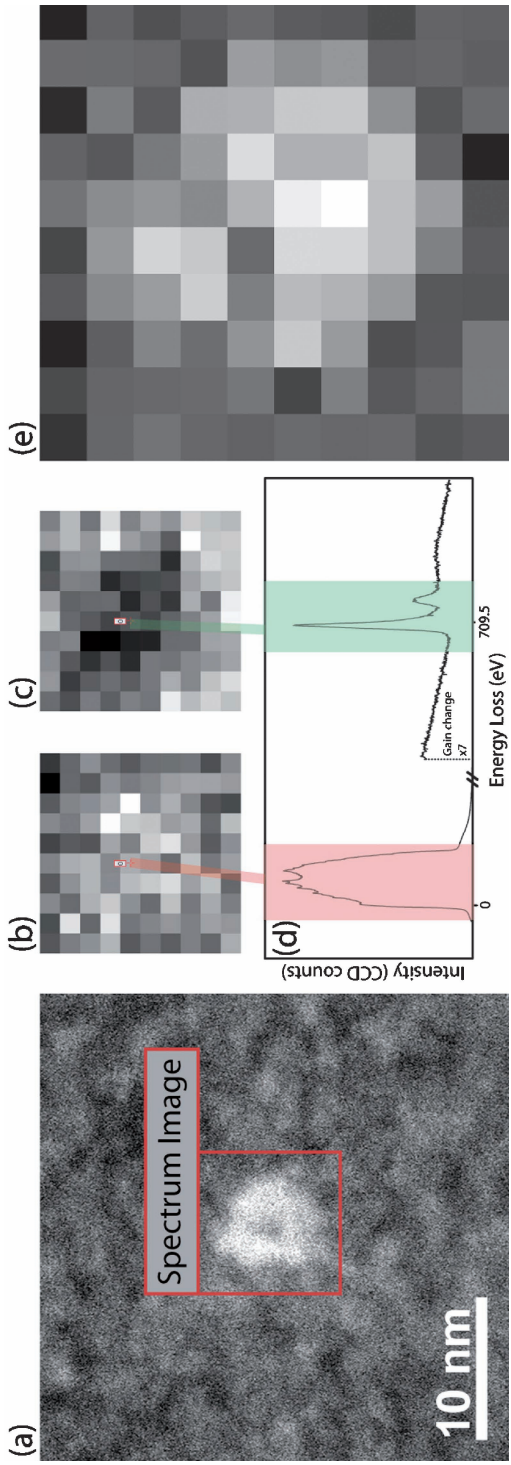


Figure 4.12 (a) HAADF image of a single ferritin molecule core. The spectrum imaging box is indicated in the figure with size 10 pixels by 10 pixels (10 nm × 10 nm); (b, c) Low-loss and core-loss spectrum imaging, respectively. Each pixel represents one spectrum as shown in (d). In order to record both low-loss and core-loss spectra with same amount of current, the EELS spectrometer was strongly defocused for low-loss acquisition, while core-loss (Fe $L_{2,3}$-edges) was correctly focused; (e) Density map showing the number of Fe atoms nm^{-2} in each pixel.

where N is the valence electron density and ε_o the permittivity of the free space. This free-electron formula works surprisingly well for a range of elements and compounds. Free-electron metals, such as aluminium, show very sharp plasmons, while those in insulators and semiconductors are considerably broader since the valence electrons are damped by scattering with the ion-core lattice. In addition to three-dimensional bulk plasmons it is also possible to excite two dimensional surface or interface plasmons, these may be apparent in very thin nanolayers as well as small nanoparticles or nanorods.

As the plasmon energy and hence plasmon peak position is sensitive to the valence-electron density, any changes in this quantity, such as those due to alloying in metals or general structural rearrangement in different microstructural phases, can be detected as a shift in the plasmon energy allowing a means of chemical phase identification. In a nanostructure, changes in properties such as electrical or thermal conductivity and even elastic modulus are a function of the local valence-electron densities, hence determination of the plasmon energy can provide a powerful tool for highly localised property determination in solid microstructures. As an example, Daniels *et al.*[12] have studied a set of pitch-derived graphitising carbons heat treated to different temperatures and have correlated the position of the volume plasmon peak with the heat treatment temperature; this data is shown graphically in Figure 4.13(b). The change in plasmon peak position mirrors the change in mass density (and hence valence-electron density) that occurs as long-range structural order (graphitic character) progressively develops in the sample as it is heated to higher temperatures. As a result, the plasmon peak position can be used to map graphitic character in a carbonaceous sample using either energy-filtered TEM and a plasmon image ratio technique[13] or directly using STEM spectrum imaging. In principle, this data can also be correlated with variations in physical properties such as elastic modulus and thermal or electronic conductivity with a spatial resolution of ca. 1 nm.

As well as plasmon oscillations, the low-loss region may also exhibit interband transitions, *i.e.* single electron transitions from the valence band to unoccupied states in the conduction band, which appear as peaks superimposed on the main plasmon peak. In a solid state physics picture, these interband transitions represent a Joint Density Of States (JDOS) – a convolution between the valence and conduction band DOS. The presence of these single-electron excitations can lead to a shift in the energy of the plasmon resonance. As an example, Figure 4.13(a), the low-loss spectrum of graphite exhibits a π to π^* antibonding interband transition at 6.2 eV in addition to the volume plasmon peak at 26 eV. In a carbonaceous sample, the intensity of this interband transition (when normalised to the intensity obtained from perfectly crystalline graphite) is indicative of the proportion of carbon atoms involved in graphitic (sp^2-like) bonding. Figure 4.13(c) shows the variation in the percentage of sp^2-bonded carbon atoms in a series of pitch-derived graphitising carbons heat treated to different temperatures. A similar analysis may be performed using the intensity of the π^* peak at the carbon K-edge as is shown in Section 4.4.4.[12] (Figure 4.18(a)).

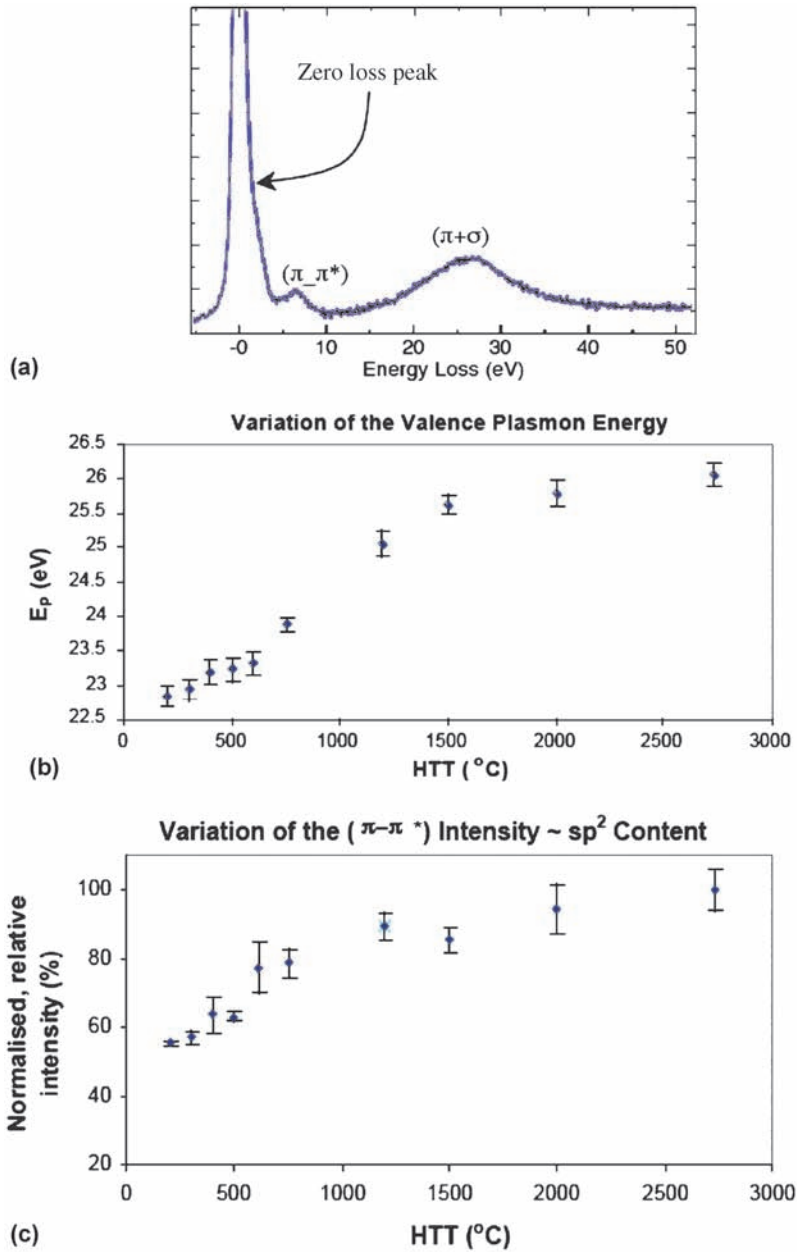


Figure 4.13 (a) EELS low-loss spectrum measured from a sample of heat-treated carbonaceous mesophase pitch showing a feature at 6.5 eV arising from an interband transition between the π bonding and antibonding orbitals, and the bulk valence plasmon ($\sim$26 eV) that arises from a collective oscillation of the valence electrons. (b) Variation of bulk valence plasmon energy with heat treatment temperature. (c) Relative carbon sp^2 content obtained from the low-loss region by measuring the relative intensity of the feature at 6.5 eV.

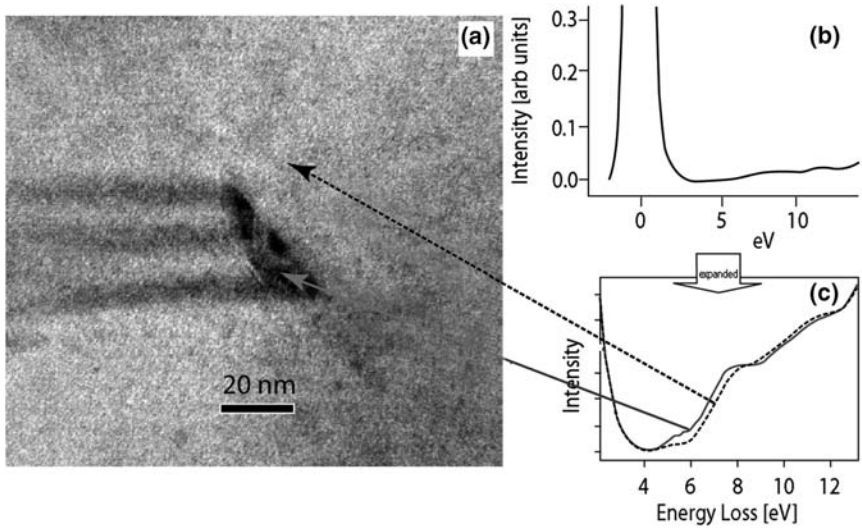


Figure 4.14 (a) STEM BF image of a stacking-fault arrangement in CVD single-crystalline diamond comprising of primary and secondary faults. (b) EEL spectra taken at locations of the arrows, *i.e.* in the perfect crystal (dashed) and near the stacking fault (solid). The difference can be seen more clearly in the expanded view in (c). Enhancement in the joint density of states in the vicinity of the partial dislocations bounding the stacking fault occurs due to energy states below the conduction band edge ($\sim$5.5 eV) and due to contribution of sp^2 bonding ($\sim$7 eV). Data courtesy of Dr Uschi Bangert, Manchester.

In an insulator with a bandgap there should be no interband transitions below the bandgap energy and thus the bandgap is associated with an initial rise in intensity in the low-loss region, as seen in Figure 4.14 for the case of diamond. Accurate removal of the zero-loss peak from the rest of the spectrum allows the determination of bandgap energies and whether the gap is direct or indirect. However, a more accurate procedure is based on extraction of the dielectric function of the material, described below. Figure 4.14 shows the variation of bandgap energy at a defect in diamond.

A more sophisticated analysis of the low-loss region is based upon the concept of the dielectric function, ε, of the material. This is a complex quantity, dependent on both energy loss and momentum, which represents the response of the entire solid to the disturbance created by the incident electron. The same response function, ε, describes the interaction of photons with a solid and this means that energy-loss data may be correlated with the results of optical measurements in the visible and UV regions of the electromagnetic spectrum, including quantities such as refractive index, absorption and reflection coefficients. Owing to their resolution limitations, optical measurements are by necessity averages of an ensemble of individual nano-objects, whereas low-loss EELS data can in principle be obtained from single nanostructures.

4.4.4.2 Near-edge Structure measurements

As described in Section 4.4.2 and Figure 4.10(b), the change in the cross section for inner-shell ionisation with energy (the energy differential cross section) and therefore the detailed ionisation edge shape, is proportional to a site and symmetry projection of the unoccupied DOS. Since the exact form of both the occupied and unoccupied DOS will be appreciably modified by the presence of bonding between atoms in the solid, this will therefore be reflected in the detailed ionisation edge structure known as ELNES.

In many cases it is found that, for a particular elemental ionisation edge, the observed ELNES exhibits a structure that, principally, is specific to the arrangement, *i.e.* the number of atoms and their geometry, as well as the type of atoms solely within the first coordination shell. This occurs whenever the local DOS of the solid is dominated by atomic interactions within a molecular unit and is particularly true in many nonmetallic systems such as semiconducting or insulating metal oxides where we can often envisage the energy band structure as arising from the broadened molecular orbital levels of a giant molecule. If this is the case, we then have a means of qualitatively determining nearest-neighbour coordinations using characteristic ELNES shapes known as coordination fingerprints. A wide range of cations in different coordinations (*e.g.* aluminium, silicon, magnesium, various transition metals, *etc.*) and anion units (*e.g.* borate, boride, carbonate, carbide, sulfate, sulfide, nitrate, nitride, *etc.*) in inorganic solids show this behaviour that can be of great use in phase identification and local structure determination.[9,14,15] An example of this for the case of aluminium is provided in Figure 4.15 and this technique has been employed to identify the nanostructure in hydrated cement pastes; here EELS can be used to identify the substitutional chemistry of aluminium in inner and outer product regions within a cement microstructure (inner product is the region originally occupied by the original unhydrated cement grain, while outer product is the region into which the hydrating grain expands). In outer product regions, Al substitutes for Si present in silicate tetrahedra, while in inner product regions a substantial proportion of Al substitutes into octahedral sites in a Mg,Al hydroxide phase.[15] This coordination fingerprinting technique has also recently been applied to the early stages of nucleation of both synthetic and *in-vivo* biomineralised aluminosilicates.[16]

For certain edges, as well as for certain compounds, the concept of a local coordination fingerprint breaks down. In these cases, the unoccupied DOS cannot be simply described on such a local level and ELNES features are found to depend critically on the arrangement of the atoms in outerlying coordination shells allowing medium-range structure determination. This opens up possibilities for the differentiation between different structural polymorphs, as well as the accurate determination of lattice parameters, vacancy concentrations and substitutional site occupancies in complex structures.[17] An example of this behaviour is provided by the case of the carbon K-edge of cubic titanium carbide shown in Figure 4.16.

As noted in Section 4.4.2, this figure shows that it is possible to model the unoccupied DOS and compare this directly with the ELNES. In this particular

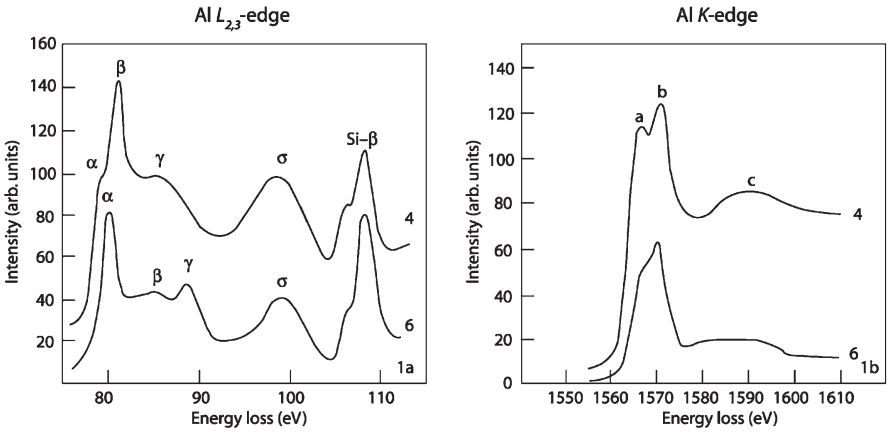


Figure 4.15 Al $L_{2,3}$- and Al K-edges of the minerals orthoclase (tetrahedral coordination of aluminium to oxygen) and beryl (octahedral coordination of aluminium to oxygen). The pre-edge shoulder at the Al $L_{2,3}$-edge and the lower relative intensity of peaks a and b at the Al K-edge represent a coordination fingerprint for aluminium in fourfold tetrahedral coordination. Reproduced from P. Hansen, R. Brydson and D. McComb, EELS fingerprint of Al-co-ordination in silicates, *Microsc. Microanal. Microstruc.*, 1994, **5**, 173–182.

case the C K-ELNES can be shown to be very sensitive to the long-range crystalline cubic rocksalt structure and the fact that the C K-edge from the carbide nanorods, shown in Figure 4.11, is virtually identical to the spectrum in Figure 4.16 suggests that the nanorods have the same cubic TiC structure (which can be confirmed by electron diffraction) and are relatively stoichiometric.

An additional chemical use of ELNES is for the spatially resolved determination of the formal valency of elements in a solid. The valence or oxidation state of the particular atom undergoing excitation influences the ELNES in two distinct ways. Firstly, changes in the effective charge on an atom lead to shifts in the binding energies of the various electronic energy levels (both the initial core level and the final state) that often manifests itself in an overall chemical shift of the edge onset. Secondly, the valence of the excited atom can affect the intensity distribution in the ELNES. This predominantly occurs in edges that exhibit considerable overlap between the initial and final states and hence a strong interaction between the core hole and the excited electron leading to the presence of quasiatomic transitions (so called since the observed ELNES is essentially atomic in nature and only partially modified by the crystal field due to the nearest-neighbour atoms). Examples of such spectra are provided by the $L_{2,3}$-edges of the 3d and 4d transition metals and their compounds and the $M_{4,5}$-edges of the rare-earth elements. These spectra exhibit very strong, sharp features known as white lines which result from transitions to energetically narrow d or f bands. This makes detection and quantification of these elements extremely easy. Such spectra may be modeled using atomic multiplet theory in the presence of a crystal field of the appropriate ligand field symmetry (Figure 4.17).

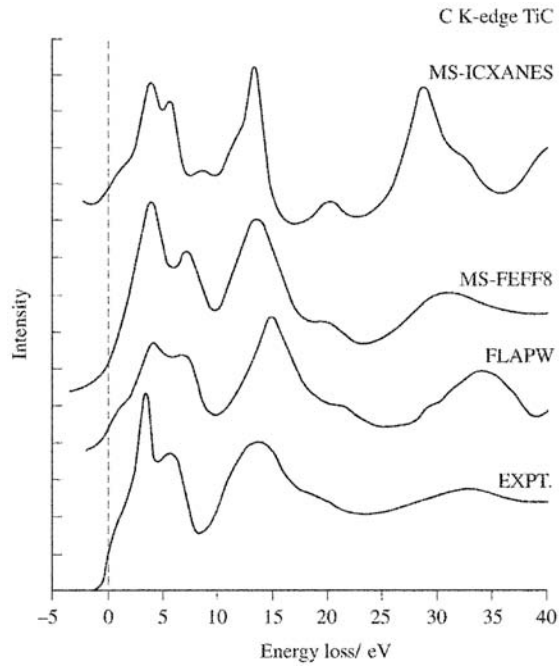


Figure 4.16 Comparison of the experimental C K-ELNES from TiC (*Expt.*) with the results of theoretical modeling calculations using both a DFT augmented plane-wave band-structure code (*FLAPW*), and two different real-space multiple scattering codes (*FEFF8* and *ICXANES*).

As an example, Figure 4.17 shows the $L_{2,3}$-edges from a number of iron oxides, containing iron in either the [+ III] or [+ II] oxidation state. Transitions from the Fe 2p shell are actually split into two components separated by the spin orbit splitting of the ionised 2p core level: an L_3-edge followed at higher energy loss by a broader L_2-edge. It is clear that both the L_3- and L_2-edges exhibit a chemical shift to higher energy loss with increasing iron valency (this remains true when elemental iron is also considered). Additionally, the relative intensities in the two separate white line components also clearly depend on the valence of the excited atom thus leading to the possibility of oxidation-state identification in an unknown sample. Such a procedure can be employed to identify the iron valence state in single nanometre-sized ferritin cores *in-situ* within liver biopsies (exclusively [+III] in healthy livers) and to also investigate any deviation from this in diseased livers *via* determination of the ratio of Fe[+III]/total Fe.[18,19]

Finally, EELS may be used to extract local structural information such as the length of chemical bonds. There are two approaches to bond-length determination: the first is to analyse the weak EXELFS oscillations occurring some 40–50 eV above the edge onset. Since these oscillations are weak, high statistical accuracy (*i.e.* high count rates and long acquisition times) is required if useful information is to be extracted. The second procedure employs the

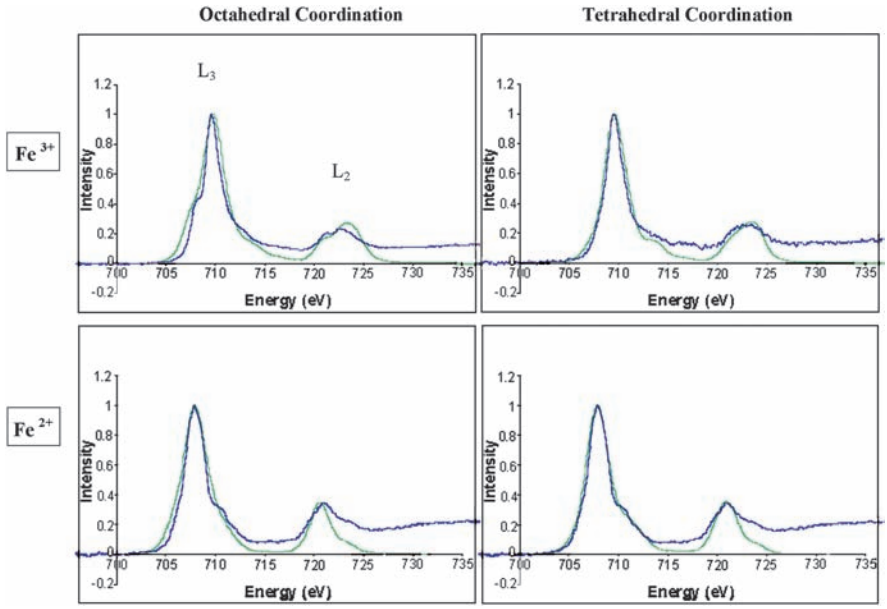


Figure 4.17 Fe $L_{2,3}$-edges from the minerals haematite and orthoclase (containing Fe^{3+} in octahedral and tetrahedral coordination, respectively), and hedenbergite and hercyanite (both containing Fe^{2+} in octahedral and tetrahedral coordination, respectively); the blue curves are the experimental EELS results, while the green curves are the results of theoretical modeling using atomic multiplet theory.[7,14] Note the shift to higher energy (from 707.5 eV to 709.5 eV) with increasing oxidation state of Fe. Theoretical data calculated by Derek Revill and Andrew Scott, University of Leeds.

energy position of the broad ELNES peaks some 20–30 eV above the edge onset, known as multiple scattering resonances (MSR). As their name suggests, these features arise from a resonant scattering event involving the excited electron and a particular shell of atoms. The energies of these features above the edge onset have been shown to be proportional to $1/R^2$, where R is the bond length from the ionised atom. Identification of such MSR permits a semiquantitative determination of nearest neighbour, and in some cases second nearest-neighbour bond lengths. Figure 4.18 shows bond-length data derived from the carbon K-edge in a series of pitch-derived graphitising carbons heat treated to different temperatures (the same set of samples shown in Figure 4.13). The energy position of the MSR provides information on the second nearest-neighbour shell, from which the C–C bond length can be derived. This is shown to decrease gradually with increasing temperature. Furthermore, as discussed in Section 4.4.4.1, the relative intensity of the π^* peak at the C K-edge (relative to the intensity of the higher-energy σ^* structure and normalised to that in graphite) can be used to extract the proportion of sp^2-bonded carbon atoms.

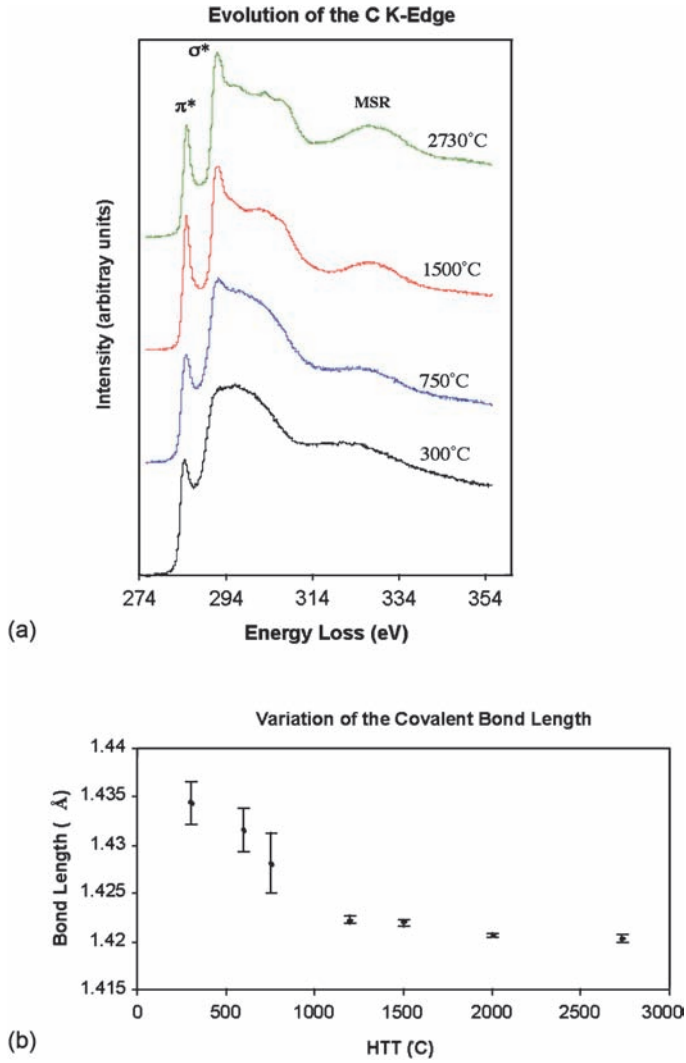


Figure 4.18 (a) Evolution of the C K-ELNES as a function of heat treatment for a carbonaceous mesophase pitch. Note the change in the relative intensity of the π antibonding peak that can be used to derive the relative carbon sp^2 content as well as the shift in the position of the multiple scattering resonance (msr) peak that can be used to infer (b) the C–C bond length.

4.4.5 Application to the Analysis of Nanometre Volumes in the S/TEM

There are a number of strategies for high spatial resolution analysis applicable to both EELS and EDX. The first is to form a small probe either in STEM or TEM mode. As discussed in Section 4.3.3, there will be broadening of the beam as it passes through the thin TEM specimen and this will limit the ultimate

(lateral) spatial resolution for microanalysis using small probes, particularly for EDX analysis where all X-rays produced isotropically within the total beam-broadened volume are collected. However, as shown in Figure 4.19, in the case of EELS in transmission where the signal is highly forward-peaked, the signal collection angle (β) and hence analytical volume can be defined using a collector aperture that can significantly improve the ultimate spatial resolution compared to that obtained with EDX. When analysing the signal from an isolated feature such as a dopant or a defect, generally, (for the same probe current) the smaller the probe size the higher the signal to background ratio (and hence detection sensitivity) and this rapidly increases with the inverse square of the probe size.

A second strategy is the *spatial difference* technique that employs larger probe sizes and correspondingly lower electron doses. Essentially this is a difference spectrum formed by monitoring changes in EELS (or EDX) intensity

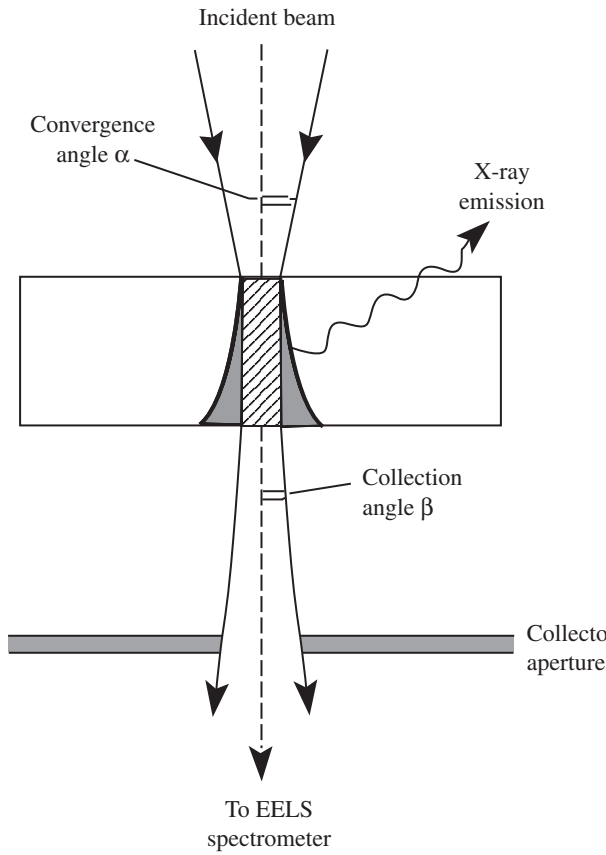


Figure 4.19 Schematic diagram of the shape of the interaction volume in a thin TEM specimen showing the regions from which the different analytical signals are produced. Note that the use of angle-limiting EELS collector aperture significantly improves the spatial resolution for analysis over that obtained with EDX.

with respect to changes in beam position or spatial coordinate. Generally probe sizes in TEM, or scanned probe areas in STEM, are of the order of a few square nanometres. The spatial difference spectrum is generated by numerically determining the difference between two (or more) spectra recorded at different locations in the sample and can reveal changes in composition and bonding. A schematic diagram showing the procedure for extracting the ELNES signal at a heterophase interface is shown in Figure 4.20. Such an approach has allowed the detection and analysis of low levels of segregant at grain boundaries and defects and been extended to monitor changes in bonding across a boundary at nanometre resolution with better than monolayer sensitivity.[20]

The ultimate in high spatial resolution EELS analysis occurs when it is performed in combination with STEM High Angle ADF imaging. With the

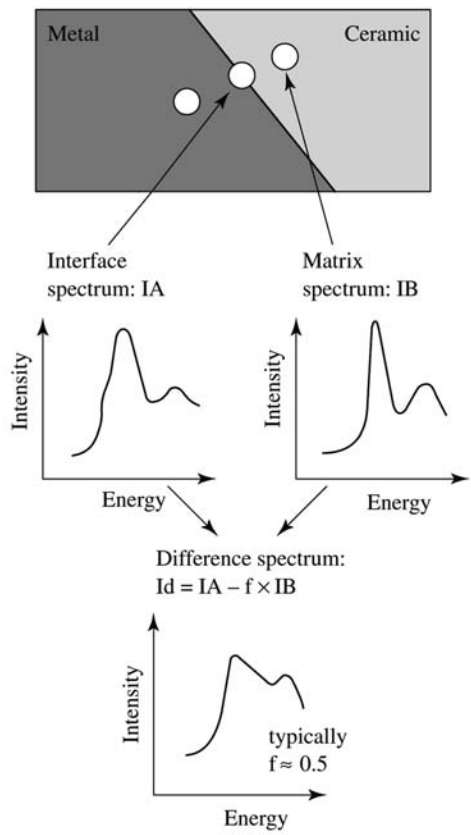


Figure 4.20 Schematic diagram of the spatial difference method for the determination of the ELNES at a metal ceramic interface. With a beam located at points A and B, the measured intensities are IA and IB, respectively. Scaling IB by a factor, *f*, allows modeling of the background of IA to produce the spatial difference I_d, which represents the ELNES contribution of atoms at or near the interface and hence their local chemical bonding environment.

incident beam oriented along a major zone axis, the beam can be channeled along an atomic column that can limit the broadening of a small STEM probe as it traverses a specimen. For a probe of sub-Angstrom dimensions (such as that formed in aberration corrected machines such as SuperSTEM – www.superstem.co.uk) this channeling allows the formation of an atomic resolution Z-contrast image of the specimen and, if the probe is accurately placed over selected atomic columns in the image, simultaneous EELS microanalysis from single atomic columns becomes possible. Figure 4.21 shows an example of this atomic column analysis on a dislocation core in gallium nitride.

When analysing small volumes of a crystalline material that is crystallographically anisotropic, the EEL spectrum (particularly the ELNES) may exhibit a

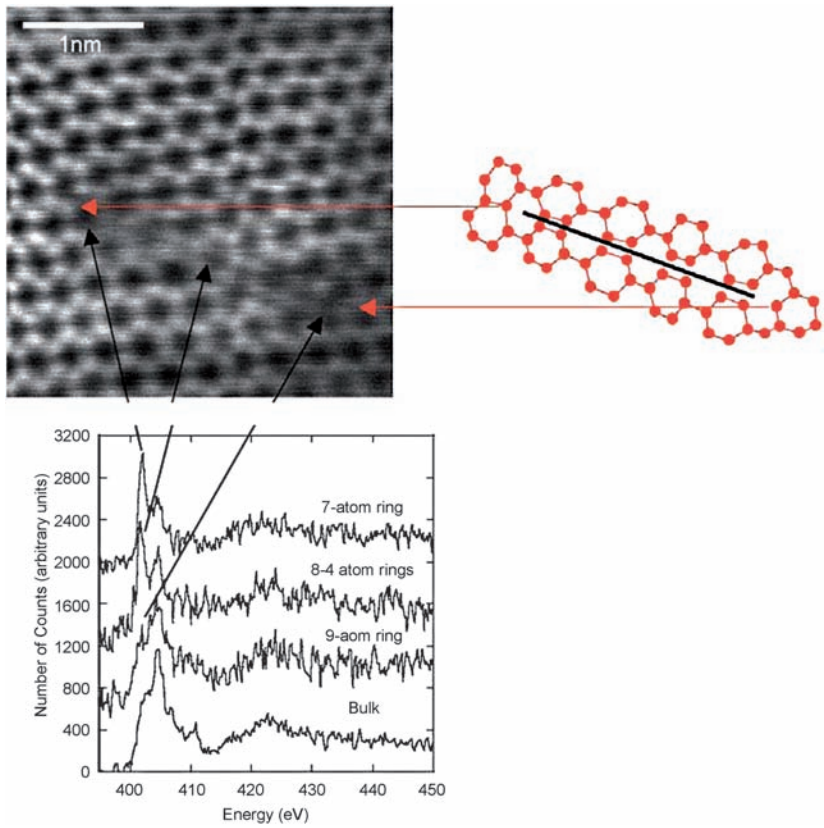


Figure 4.21 STEM HAADF image of a dissociated mixed dislocation in GaN imaged along the c-axis together with a schematic of the atomic columns. EELS nitrogen K-edges are shown from three positions along the dislocation and reveal marked differences in the N-K edge at either end and at the stacking fault between the partials. Reproduced from I. Arslan, A. Bleloch, E. A. Stach and N. D. Browning, Atomic and Electronic Structure of Mixed and Partial Dislocations in GaN, *Phys. Rev. Lett.*, 2005, **94**, 025504.

directional dependence that reflects the bonding anisotropy present in the sample. By carefully choosing the sample orientation and the collection conditions, this anisotropy can be intentionally studied or, alternatively, it may be averaged so as to provide a signal that is independent of sample orientation.[7,21] The structure and bonding in graphite is anisotropic and the data from graphitic-like regions shown in Figures 4.13, 4.18 and 4.24 has been experimentally recorded under conditions so as to average any orientation dependence.

4.5 EELS Imaging

Owing to the inherently high spatial resolution of the EELS analytical technique, using electrons of a specific energy loss or range of energy losses to form images of specimen thickness, elemental composition, chemical phase composition or even electronic structure represents an extremely attractive and exciting prospect for nanoanalysis.

Energy Filtered TEM (EFTEM) involves the selection of a specific energy loss value, or narrow range of energy losses (''an energy window'') from the transmitted electron beam *via* use of energy selecting slit after the EEL spectrometer. Using only zero loss (elastically) scattered electrons to form images and diffraction patterns increases contrast and resolution, allowing easier interpretation than with unfiltered data. Chemical mapping may be achieved by acquiring and processing images formed by electrons that have undergone either a specific low energy loss event (such as an interband transition or a plasmon event) or an inner-shell ionisation event, using either the whole ionisation edge or a specific ELNES feature indicative of a certain bonding characteristic. The resolution of such EFTEM images depends on a number of factors (see Chapter 6 in Ref. 14), however, generally it can approach around one nanometre.

For the mapping of elemental distributions using EFTEM, two approaches are commonly employed: (a) jump-ratio mapping employs two energy windows and hence energy-filtered images, one positioned just before the ionisation edge and one positioned just after the edge; the post-edge image is divided by a pre-edge image to produce a semiquantitative map of the elemental distribution that is sensitive to small concentrations and is relatively free of noise and also diffraction contrast present in crystalline materials. (b) true, quantitative elemental mapping (where image intensity is proportional to concentration) usually employs three energy windows, two pre-edge windows and a post edge window; the two pre-edge images are used to extrapolate the background contribution to the post-edge image and this extrapolated background image is then subtracted from the post-edge image to produce a quantitative elemental map where image intensity is directly related to areal density. Figure 4.22 shows an example of EFTEM jump-ratio imaging of the particle surface of mixed chromium oxide – titanium dioxide material which forms a Ti-rich surface nanophase on sintering in air;[22] the corresponding surface layer is absent in material sintered in a reducing atmosphere where a true oxide solid solution is formed.

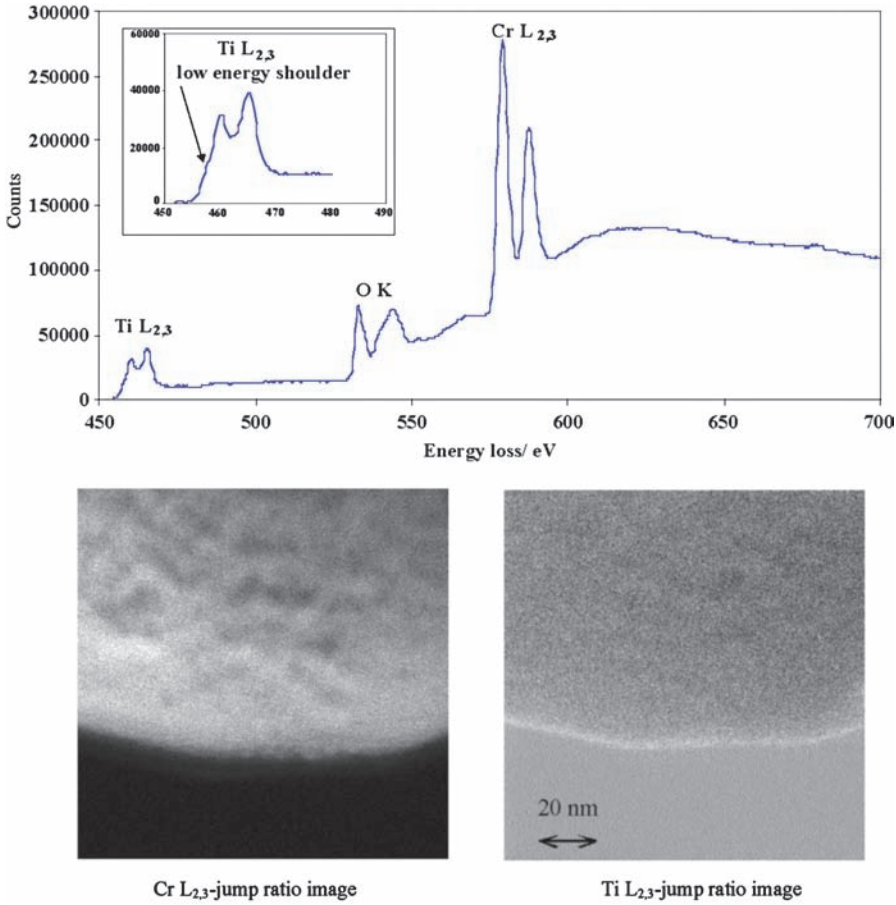


Figure 4.22 Background subtracted EELS spectrum from a crystalline grain of air-sintered $Cr_{1.8}Ti_{0.2}O_{3+x}$, corresponding energy filtered jump-ratio images using the Cr $L_{2,3}$-edge and the Ti $L_{2,3}$-edge. Notice the enrichment of titanium at the particle surface.

An alternative approach, used when mapping is performed with an EEL spectrometer and detector attached to an STEM, is to raster the electron beam across the specimen and record an EELS spectrum at every point (x,y) – this technique being known as "spectrum imaging". The complete dataset may then be subsequently processed to form a 2D quantitative map of the sample using either standard elemental quantification procedures, or the position and/or intensity of characteristic low-loss or ELNES features. The inherent high spatial resolution of the EELS spectrum imaging technique has allowed maps of elemental distributions to be formed at subnanometre resolution. In addition, the possibility of plasmon or ELNES chemical bonding maps has also been demonstrated. An example of a spectrum image is shown in Figure 4.12 for the case of the quantification of iron in ferritin cores.

Processing of any series of linescans or maps, whether obtained using STEM/EDX, STEM/EELS spectrum imaging or EFTEM may be done using a number of techniques including the construction of simple scatter plots and identification of data clusters or employing more detailed multivariate statistical analysis methods to determine the significant principal components. Such methods can be used to highlight distinct chemical phases within a sample region either in terms of elemental concentrations, or even local coordinations or valence states of particular elements.[14]

4.6 Radiation Damage

A major concern in analytical electron microscopy is that the interaction with the electron beam does not alter or damage the sample in any way. This is of particular concern for measurements made at high spatial resolution such as those on nanostructures. Generally, radiation damage can manifest itself in one of two main ways: firstly, atomic displacement and even removal of atoms resulting in structural damage and also mass loss, secondly ionisation damage (radiolysis) that can cause chemical changes as well as often subsequently causing a change in structure and also mass loss. In many cases the electron-beam-induced damage is proportional to the electron dose (the current density in the probe multiplied by the exposure time), although there may be certain cases where the dose rate is important.[23]

In terms of the analyses already discussed, measurements with a stationary electron probe can obviously lead to specimen damage, particularly if the current density is high. Scanning a focused probe in a STEM to form an EELS or EDX linescan or spectrum image generally alleviates damage (particularly where heating of the specimen is an important damage mechanism) but, in some cases, this can lead to elements being driven by the electron beam and to erroneous results. Possible solutions to these problems include the use of cooled specimen holders as well as changing the S/TEM accelerating voltage and probe current density during measurement.

A further effect can arise from specimen contamination by hydrocarbons either present on the specimen or within the microscope. The electron beam polymerises these hydrocarbons on the surface of the TEM specimen and this can lead to mass gain and severe problems with analysis. Specimen cleanliness is the key to avoiding such effects.

An example of the effect of electron-beam-induced damage on measurements of ferritin cores *in-situ* within biological tissue (as discussed in Section 4.4.3) is shown graphically in Figure 4.23.[11] Here, measurements have been made as a function of electron dose and have been analysed so as to extract the local chemistry (valence state) and coordination of iron in the ferritin cores. It can be seen that these quantities change as a function of dose owing to the fact that a complex damage mechanism leads to a change in iron coordination from octahedral to tetrahedral, as well as a reduction in iron from Fe^{3+} to Fe^{2+}. As the data shows it is possible either to choose a safe dose for measurement where the iron coordination or

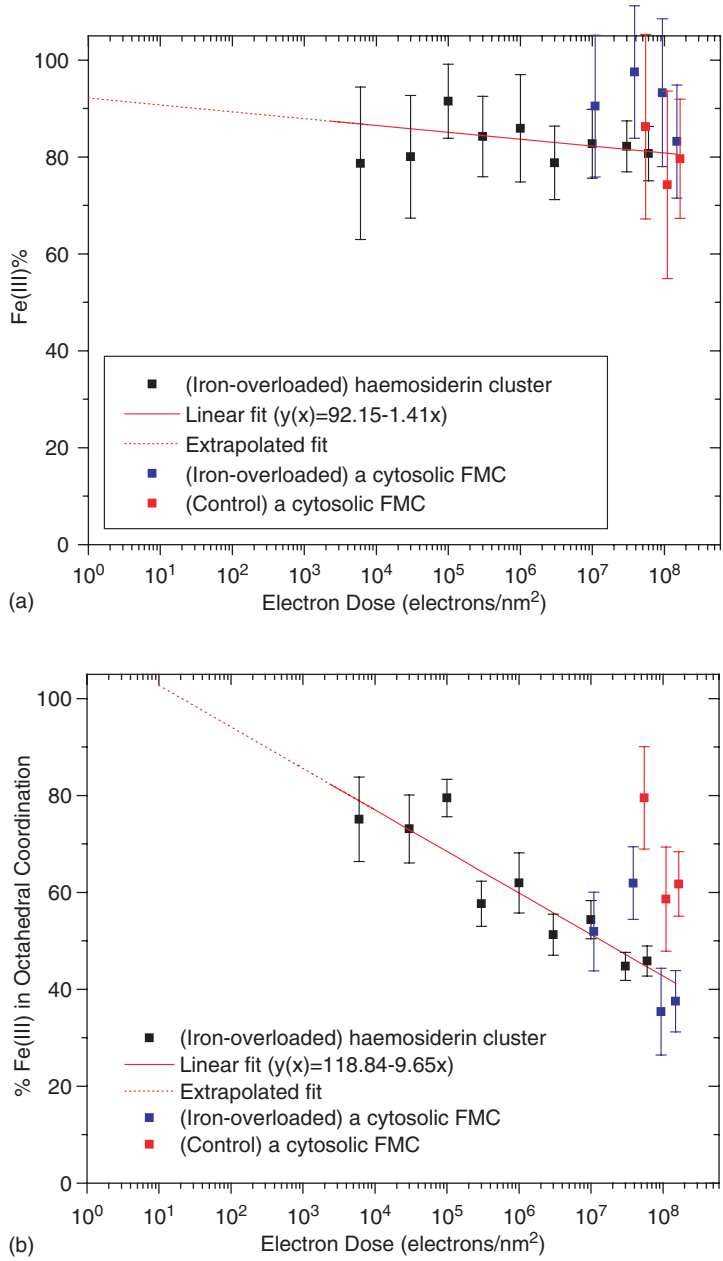


Figure 4.23 Graphs showing the relative proportion of (a) Fe^{3+} and (b) Fe^{3+} in octahedral coordination in ferritin cores as a function of electron dose in a 100 kV STEM. This data was obtained from both normal (control) and diseased (iron overloaded) liver biopsies (the latter from two different cellular locations). The data was derived using curve fitting of measured ELNES spectra with Fe $L_{2,3}$-ELNES from reference materials (haematite, orthoclase, hedenbergite and hercyanite).

valency is relatively unaffected, or to extrapolate the data back to a very low dose where the specimen remains completely damage free.

4.7 Emerging Techniques

An emerging technique in the TEM that could become potentially important when performed in conjunction with EELS is the use of a combined *in-situ* TEM-STM holder.[24] The STM tip may be used to extract current-voltage characteristics from individual nano-objects, whilst EELS can provide an alternative indication of the density of unoccupied states and hence chemistry. Additionally, the STM tip may be used either to deform nanostructures such as carbon nanotubes whilst simultaneously monitoring both images and EELS spectra. Figure 4.24 shows the changes in the carbon K-edge as a bundle of single-walled carbon nanotubes is deformed; generally deformation of a CNT results in a decrease and broadening of the initial π^* peak intensity at the C K-edge together with similar loss in detail and intensity in the higher-energy σ^* structure. Initial results suggest that reversible (elastic) deformation leads to an increase in the $\pi^*/(\pi^* + \sigma^*)$ intensity ratio as σ^* bonds are most sensitive to variations from linearity induced by bending of the tubes, whereas

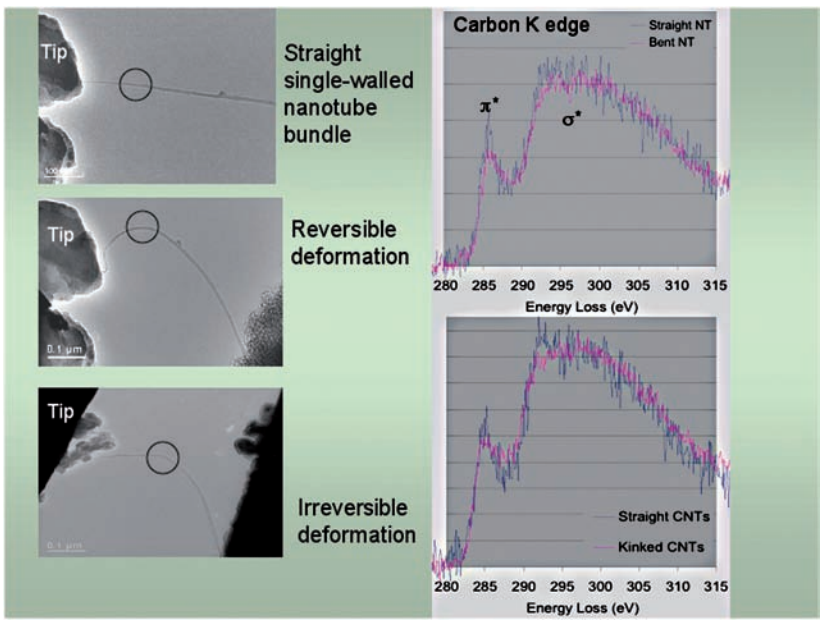


Figure 4.24 TEM bright field image of an *in-situ* STM tip being used to deform a bundle of single-walled carbon nanotubes both reversibly and irreversibly (with the introduction of kinks). The corresponding C K-ELNES spectra taken from the areas circled are shown (top) straight bundle compared with reversibly deformed bundle and (bottom) straight bundle compared with irreversibly deformed bundle.

irreversible deformation that introduces kinks leads to a decrease in the $\pi^*/(\pi^* + \sigma^*)$ intensity ratio and a C K-edge spectrum more similar to C_{60} fullerene owing to the introduction of pentagonal defects at the kink.

4.8 Conclusions

This review of the chemical analysis of nanostructures *in-situ* within the transmission electron microscope has highlighted the two major techniques of energy dispersive X-ray analysis and electron energy-loss spectroscopy. Details of instrumentation have been discussed including the practical usage of the techniques as applied to nanostructured objects together with the types of chemical information that can be extracted. This has been facilitated by the provision of a limited set of examples, although it should be stressed that the interested reader should consult the references and further reading material for a much more complete list of potential applications.

References

1. D.B. Williams and C.B. Carter, *Transmission Electron Microscopy*, Plenum Press, New York, 1997.
2. R.J. Keyse, A.J. Garrett-Reed, P.J. Goodhew and G.W. Lorimer, *Introduction to Scanning Transmission Electron Microscopy*, BIOS, Oxford, 1998.
3. I.P. Jones, *Chemical Microanalysis using Electron Beams*, Institute of Materials, London, 1992.
4. J. Höhne, M. Bühler, T. Hertuch and U. Hess, High Resolution X-Ray Spectrometers for Materials Analysis, *Inst. Phys. Conf. Ser.*, 1999, **161**, 249.
5. S.J. Pennycook, A.J. Craven and L.M. Brown, Cathodoluminescence on a Scanning Transmission Electron Microscope, *Inst. Phys. Conf. Ser.*, 1977, **36**, 69.
6. R.F. Egerton, *Electron Energy-loss spectroscopy in the Electron Microscope*, Plenum Press, New York, 1996.
7. R. Brydson, *Electron Energy-loss Spectroscopy*, BIOS (now Taylor Francis), Oxford, 2001.
8. P.C. Tiemeiger, Operation Modes of a TEM Monochromator, *Inst. Phys. Conf. Ser.*, 1999, **161**, 249.
9. V.J. Keast, A.J. Scott, R. Brydson, D.B. Williams and J. Bruley, Electron Energy Loss Near Edge Structure – a Tool for the Investigation of Electronic Structure on the Nanometre Scale, *J. Microsc.*, 2001, **203**, 135–175.
10. X. Li, Z. Dong, G. Yuan, A. Brown, A. Westwood, R. Brydson and B. Rand, The Synthesis and Characterisation of Transition Metal Carbide Nanophases using Molten Salts, *Proc. carbon 2006, Extended Abstract*, **11C4**, 2006.
11. Y. Pan, A. Brown, R. Brydson, A. Warley, A. Li and J. Powell, Electron Beam Damage Studies of Synthetic 6-line Ferrihydrite and Ferritin Molecule Cores Within a Human Liver Biopsy, *Micron.*, 2006, **37**, 403–411.

12. H. Daniels, A. Brown, R. Brydson and B. Rand, Investigating the Graphitization of Carbon using Analytical FEGTEM, *Inst. Phys. Conf. Ser.*, 2003, **179**, 27.
13. H. Daniels, A. Brown, R. Brydson and B. Rand, Quantitative Valence Plasmon Mapping in the TEM: Viewing Physical Properties at the Nanoscale, *Ultramicroscopy*, 2003, **96**, 547–558.
14. C.C. Ahn, (ed.), *Transmission EELS in Materials Science*, Wiley, Weinheim, 2004, 2nd edn.
15. (*a*) I.G. Richardson, A. Brough, R. Brydson, G.W. Groves and C.M. Dobson, The Location of Aluminium in Substituted Calcium Silicate Hydrate Gels as Determined by ^{29}Si and ^{27}Al NMR and EELS, *J. Am. Ceram. Soc.*, 1993, **76**, 2285–2288; (*b*) L.A.J. Garvie, A.J. Craven and R. Brydson, The Use of Electron Energy Loss Near-Edge Structure in the Study of Minerals, *Amer. Min.*, 1994, **79**, 411.
16. A. Brown, R. Brydson, R. Jugdaohsingh, A. Bleloch, K. White, C. McCrohan, R. Walton and J. Powell, Analytical Electron Microscopy of the Earliest Identified Alumino-silicates: from Dilute Solutions to Freshwater Snails, *Proceedings of International Microscopy Congress IMC16*, 2006, **1**, 1895.
17. A.J. Scott, R. Brydson, M. MacKenzie and A.J. Craven, A Theoretical Investigation of the ELNES of Transition Metal Carbides and Nitrides for the Extraction of Structural and Bonding Information, *Phys. Rev. B*, 2001, **63**, 245105.
18. C.C. Calvert, A. Brown and R. Brydson, Determination of the Local Chemistry of Iron in Inorganic and Organic Materials, *J. Electron Spectrosc. Relat. Phenom.*, 2005, **143**, 173.
19. Y. Pan, A. Brown, R. Brydson, A. Warley, A. Li, J. Powell, A. Bleloch, U. Falke, M. Falke and C.C. Calvert, TEM, SuperSTEM, and EELS Studies of the Structure, Chemistry, and Biological Function of Ferritin in Human Liver Biopsies, *European Microscopy Congress 2004 Proceedings*, 2004, III, 167–168.
20. C. Scheu, G. Dehm, G. Mobus, M. Ruhle and R. Brydson, Electron Energy-Loss Spectroscopic Studies of Copper/Alumina Interfaces Grown by MBE, *Philos. Mag. A*, 1998, **78**, 439–465.
21. H. Daniels, A. Brown, A. Scott, T. Nichells, B. Rand and R. Brydson, Experimental and Theoretical Evidence for the Magic Angle in Transmission Electron Energy-Loss Spectroscopy, *Ultramicroscopy*, 2003, **96**, 523–534.
22. S. McBride and R. Brydson, Analytical TEM and Surface Spectroscopy of Ceramics: The Microstructural Evolution in Titanium-Doped Chromia Polycrystals as a Function of Sintering Conditions, *J. Mater. Sci.*, 2004, **39**, 1–12.
23. R.F. Egerton, P. Li and M. Malac, Radiation Damage in the TEM and SEM, *Micron*, 2004, **35**, 399–409.
24. Z. Aslam, M. Abraham, R. Brydson, A. Brown and B. Rand, Initial Studies Using a Combined TEM-Scanning Tunnelling Microscopy (STM) Side Entry Sample Holder, *J. Phys. Conf. Ser.*, 2006, **26**, 54–58.

CHAPTER 5

Electron Holography of Nanostructured Materials

R. E. DUNIN–BORKOWSKI,[1] T. KASAMA[2] AND R. J. HARRISON[3]

[1] *Department of Materials Science and Metallurgy, University of Cambridge, Pembroke Street, Cambridge CB2 3QZ, UK, and Center for Electron Nanoscopy, Technical University of Denmark, DK-2800 Kongens Lyngby, Denmark*

[2] *Frontier Research System, The Institute of Physical and Chemical Research, Hatoyama, Saitama 350–0395, Japan, and Department of Materials Science and Metallurgy, University of Cambridge, Pembroke Street, Cambridge, CB2 3QZ, UK*

[3] *Department of Earth Sciences, University of Cambridge, Downing Street, Cambridge, CB2 3EQ, UK*

5.1 Introduction

The technique of electron holography, which was originally described by Gabor,[1] is based on the formation of an interference pattern or "hologram" in the Transmission Electron Microscope (TEM). In contrast to most conventional TEM techniques, which only allow the spatial distribution of image intensity to be recorded, electron holography also allows the phase shift of the high-energy electron wave that has passed through a specimen to be measured directly. The phase shift can, in turn, be used to provide information about local variations in magnetic induction and electrostatic potential in and around the specimen. This chapter begins with an outline of the theoretical background and experimental procedures that are required to obtain and analyse phase information from electron holograms. Medium-resolution applications of the off-axis, or "side-band", TEM mode of electron holography to the characterisation of nanostructured materials are then described, followed by a description of high-resolution electron holography and alternative forms of electron holography.

5.1.1 Basis of Off-Axis Electron Holography

The TEM mode of off-axis electron holography involves the examination of an electron-transparent specimen using defocused illumination from a highly coherent Field-Emission Gun (FEG) electron source. The region of interest is positioned so that it covers approximately half the field of view. The application of a voltage to an electron biprism results in the overlap of a "reference" electron wave that has passed through vacuum with the electron wave that has passed through the specimen, as shown schematically in Figure 5.1(a). If the electron source is sufficiently coherent, then an interference fringe pattern (an electron hologram) is formed in the overlap region, in addition to an image of the specimen. The amplitude and the phase shift of the specimen wave are recorded in the intensity and the position, respectively, of the fringes. For studies of magnetic materials, a Lorentz lens (a high-strength minilens) allows the microscope to be operated at high magnification with the objective lens switched off and the sample in magnetic-field-free conditions.

Figure 5.2(a) shows a representative electron hologram of a thin crystal of the mineral hematite. Changes in the positions of the holographic interference fringes are visible in the inset to Figure 5.2(a).

An expression for the intensity distribution in an off-axis electron hologram can be obtained by considering the addition of a tilted plane reference wave to a complex specimen wave, in the form

$$I_{\mathrm{hol}}(\boldsymbol{r}) = |\psi_i(\boldsymbol{r}) + \exp[2\pi \mathrm{i}\boldsymbol{q_c}.\boldsymbol{r}]|^2 \tag{5.1}$$

where $\boldsymbol{r}$ is a two–dimensional vector in the plane of the sample, the tilt of the reference wave is specified by the two-dimensional reciprocal space vector $\boldsymbol{q} = \boldsymbol{q_c}$, and the electron wavefunction in the image plane is given by the expression

$$\psi_i(\boldsymbol{r}) = A_i(\boldsymbol{r}) \exp\left[\mathrm{i}\phi_i(\boldsymbol{r})\right] \tag{5.2}$$

where A and ϕ refer to amplitude and phase, respectively. By combining these two expressions, eqn (5.1) can be rewritten in the form

$$I_{\mathrm{hol}}(\boldsymbol{r}) = 1 + A_i^2(\boldsymbol{r}) + 2A_i(\boldsymbol{r})\cos[2\pi \mathrm{i}\boldsymbol{q_c}.\boldsymbol{r} + \phi_i(\boldsymbol{r})] \tag{5.3}$$

From eqn (5.3), it can be seen that there are three separate contributions to the intensity distribution in a hologram: the reference image intensity, the specimen image intensity, and a set of cosinusoidal fringes, whose local phase shifts and amplitudes are equivalent to the phase and amplitude, respectively, of the electron wavefunction in the image plane.

In order to extract phase and amplitude information, an off-axis electron hologram is first Fourier transformed. From eqn (5.3), it can be shown that the

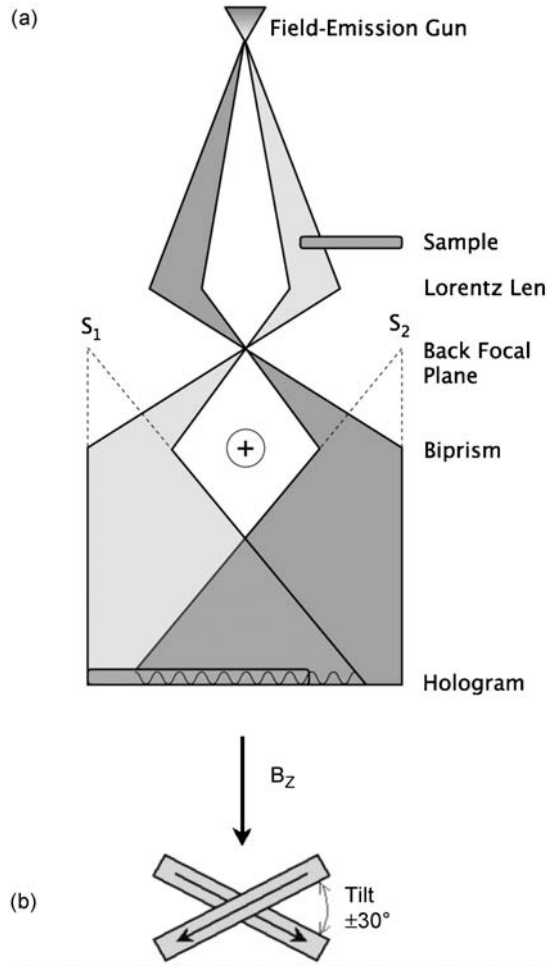


Figure 5.1 (a) Schematic illustration of the experimental setup used for generating off-axis electron holograms. The specimen occupies approximately half the field of view. Essential components are the field-emission gun electron source, which provides coherent illumination, and the electron biprism, which causes overlap of the object and (vacuum) reference waves. The biprism is usually a wire, below 1 μm in diameter, located in place of one of the conventional selected-area apertures. The sample and reference waves can be considered as originating from two virtual sources, S_1 and S_2. The Lorentz lens allows imaging of magnetic materials in close-to-field-free conditions. (b) Schematic diagram illustrating the use of specimen tilt to provide an inplane component of the vertical magnetic field, H of the conventional electron microscope objective lens for *in situ* magnetization-reversal experiments.

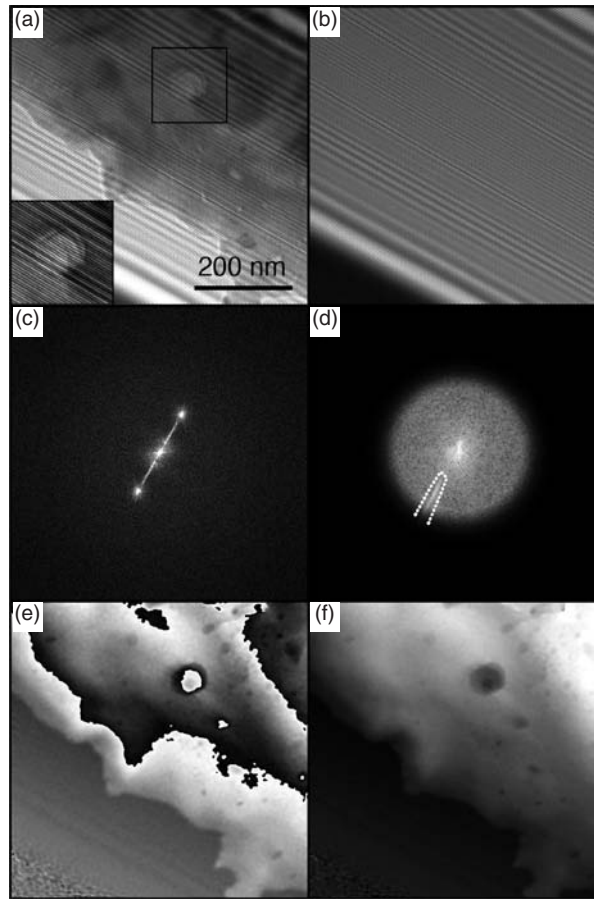


Figure 5.2 Sequence of image processing steps used to convert an off-axis electron hologram into a phase image. (a) Representative off-axis electron hologram of the region of interest on a specimen (the edge of a thin foil of hematite containing nanoscale inclusions of maghemite). The inset shows a magnified image of the holographic interference fringes in the outlined region. Coarser fringes, resulting from Fresnel diffraction at the edges of the biprism wire, are visible at the upper right and lower left corners of the original hologram. (b) Reference hologram obtained from a region of vacuum immediately after recording the hologram of the specimen. (c) Modulus squared of the Fourier transform of the electron hologram shown in (a), comprising a central peak, two sidebands, and a diagonal streak arising from the presence of the Fresnel fringes in the hologram. (d) One of the sidebands is extracted from the Fourier transform of the hologram, and a circular mask with smooth edges is used to reduce its intensity radially to zero. If required, the streak from the Fresnel fringes can be removed in a similar manner, by assigning a value of zero to pixels inside the region shown by the dashed line. (e) Analysis of the inverse Fourier transforms of sidebands from the specimen hologram and the reference hologram yields the complex image wave of the specimen, from which the phase is initially calculated modulo 2π. (f) Phase unwrapping algorithms are used to remove the 2π phase discontinuities from (e) to yield a final unwrapped phase image.

Fourier transform of a hologram can be written in the form

$$\begin{aligned}\mathrm{FT}[I_{\mathrm{hol}}(\boldsymbol{r})] = {} & \delta(\boldsymbol{q}) + \mathrm{FT}\left[A_i^2(\boldsymbol{r})\right] \\ & + \delta(\boldsymbol{q}+\boldsymbol{q_c}) \otimes \mathrm{FT}[A_i(\boldsymbol{r})\exp[\mathrm{i}\phi_i(\boldsymbol{r})]] \\ & + \delta(\boldsymbol{q}-\boldsymbol{q_c}) \otimes \mathrm{FT}[A_i(\boldsymbol{r})\exp[-\mathrm{i}\phi_i(\boldsymbol{r})]]\end{aligned} \tag{5.4}$$

Equation (5.4) describes a peak at the reciprocal-space origin corresponding to the Fourier transform of the reference image, a second peak centred at the origin corresponding to the Fourier transform of a conventional bright-field TEM image of the sample, a peak centred at $\boldsymbol{q} = -\boldsymbol{q_c}$ corresponding to the Fourier transform of the desired image wavefunction, and a peak centered at $\boldsymbol{q} = +\boldsymbol{q_c}$ corresponding to the Fourier transform of the complex conjugate of the wavefunction. Figure 5.2(c) shows the Fourier transform of the hologram shown in Figure 5.2(a). In order to recover the complex electron wavefunction, one of the two "sidebands" in the Fourier transform is selected, as shown in Figure 5.2(d), and then inverse Fourier transformed.

As phase information is stored in the lateral displacement of the holographic interference fringes, long-range phase modulations arising from inhomogeneities in the charge and the thickness of the biprism wire, as well as from lens distortions and charging effects (*e.g.*, at apertures) can introduce artifacts into the reconstructed wavefunction. In order to take these effects into account, a reference hologram is usually obtained from vacuum alone by removing the specimen from the field of view without changing the optical parameters of the microscope (Figure 5.2(b)). Correction is then possible by performing a complex division of the recovered sample and vacuum wavefunctions in real space, and then calculating the phase (the arctangent of the ratio of the imaginary and real parts) of the resulting complex wavefunction, to obtain the distortion-free phase of the image wave[2] (Figure 5.2(e)). The acquisition of a reference hologram has the additional advantage that the use of the same locations for the sidebands determined from the sample and vacuum holograms removes any tilt of the recorded wave that might be introduced by an inability to locate the exact (subpixel) position of the sideband frequency in Fourier space. Figure 5.2(e) illustrates the fact that a phase image that is calculated digitally is initially evaluated modulo 2π, meaning that 2π phase discontinuities that are unrelated to specimen features appear at positions where the phase shift exceeds multiples of this amount. If required, the phase image may then be "unwrapped" by using suitable algorithms,[3] as shown in Figure 5.2(f).

The recorded phase shift can be used to measure the electrostatic potential and also the inplane component of the magnetic induction in the specimen. Neglecting dynamical diffraction (*i.e.* assuming that the specimen is thin and weakly diffracting), the phase shift can be expressed in the form

$$\phi(x) = C_E \int V(x,z)\,\mathrm{d}z - \left(\frac{e}{\hbar}\right) \iint B_{\perp}(x,z)\,\mathrm{d}x\mathrm{d}z \tag{5.5}$$

where

$$C_E = \left(\frac{2\pi}{\lambda}\right)\left(\frac{E + E_0}{E(E + 2E_0)}\right) \tag{5.6}$$

z is the electron-beam direction, x is a direction in the plane of the specimen, $B_\perp$ is the component of the magnetic induction perpendicular to both x and z, V is the electrostatic potential, λ is the (relativistic) electron wavelength, and E and E_0 are, respectively, the kinetic and rest mass energies of the incident electron.[4] C_E takes values of 7.29×10^6, 6.53×10^6 and 5.39×10^6 rad V^{-1} m^{-1} at accelerating voltages of 200 kV, 300 kV and 1 MV, respectively. If neither V nor $B_\perp$ varies along the electron-beam direction within a sample of thickness t, then eqn (5.5) can be simplified as

$$\phi(x) = C_E V(x)t(x) - \left(\frac{e}{\hbar}\right)\int B_\perp(x)t(x)\,\mathrm{d}x \tag{5.7}$$

Hence, by making use of eqns (5.5)–(5.7), information about V and $B_\perp$ can be recovered from a measured phase image, as described subsequently.

5.1.2 Experimental Considerations

Several practical issues must be addressed in order to record and analyse an electron hologram successfully. A key experimental requirement is the availability of a reference wave that can be overlapped onto the region of interest on the specimen. This restriction, which usually implies that the hologram must be recorded from a region close to the specimen edge, can be relaxed if a thin, clean, weakly diffracting region of electron-transparent support film, rather than vacuum, can be overlapped onto the region of interest.

Electron holograms have traditionally been recorded on photographic film. However, digital acquisition using charge coupled device (CCD) cameras is now used widely due to their linear response, high dynamic range and high detection quantum efficiency, as well as the immediate accessibility of the recorded information.[5] Whether a hologram is recorded on film or digitally, the field of view is typically limited to approximately 5 μm, either by the dimensions of the recording medium or by the sampling and the contrast of the holographic fringes.

The high electron beam coherence required for electron holography necessitates the use of a FEG electron source, a small spot size, a small condenser aperture size and a low gun extraction voltage. The coherence may be maximised by adjusting the condenser-lens stigmators in the microscope to provide elliptical illumination that is wide in a direction perpendicular to the biprism when the condenser lens is overfocused.[6] The contrast of the holographic fringes is also determined by the mechanical stability of the biprism and the point spread function of the recording medium. Should the fringe contrast decrease too much, reliable reconstruction of the image wavefunction will not be possible. A phase detection limit for electron holography[7] can be determined from the effect on a

recorded hologram of Poisson-distributed shot noise, the detection quantum efficiency and the point spread function of the CCD camera, and the fringe contrast. In practice, averaging of the phase is often implemented, particularly if the features of interest vary slowly across the image or in one direction.

A final artifact results from the presence of Fresnel diffraction from the biprism wire, which is visible in Figures 5.2(a) and (b), and causes the diagonal streak at the lower left of Figure 5.2(d), as well as subsequent phase and amplitude modulations of both the specimen wave and the reference wave.[8] These effects can be removed to an extent by using a reference hologram (Figure 5.2(b)), and by Fourier-filtering the sideband before reconstruction of the image wave (Figure 5.2(d)). More advanced approaches for removing Fresnel fringes from electron holograms based on image analysis[9] and double-biprism electron holography[10] have recently been introduced. Great care should also be taken to assess the effect on the reference wave of long-range electromagnetic fields that may extend outside the sample and affect both the object wave and the reference wave.[11]

Further details about the theoretical background and practical aspects of electron holography can be obtained from recent books[12–14] and review papers.[15–19]

5.2 The Mean Inner Potential Contribution to the Phase Shift

If a phase image is associated solely with variations in mean inner potential and specimen thickness, then electron holography can be used to characterise local variations in specimen morphology and composition. For a specimen that has uniform structure and composition in the electron-beam direction, in the absence of magnetic and long-range electrostatic fields (such as those at depletion regions in semiconductors), eqn (5.7) takes the simpler form

$$\phi(x) = C_E V_0(x)\, t(x) \tag{5.8}$$

where the mean inner potential, V_0, is the volume average of the electrostatic potential in the specimen. Theoretical values for V_0 can be obtained by assuming that the specimen can be described as a collection of free atoms and using the expression

$$V_0 = \left(\frac{h^2}{2\pi m e \Omega}\right) \sum_{\Omega} f_{\mathrm{el}}(0) \tag{5.9}$$

where $f_{\mathrm{el}}(0)$ is the electron scattering factor at zero scattering angle (with dimensions of length), Ω is the unit cell volume, and the sum is performed over all atoms in the unit cell.[4] Although calculated values for $f_{\mathrm{el}}(0)$ have been tabulated,[20,21] the use of eqn (5.9) typically leads to overestimates for V_0 by approximately 10% because the redistribution of electrons due to bonding, which results in a contraction of the electron density around each atom is neglected.[22,23]

Experimental measurements of V_0 have been obtained using electron holography, in combination with independent measurements of specimen thickness profiles. Values for V_0 of 13.0 ± 0.1, 14.5 ± 0.2, 17.2 ± 0.1 and 14.3 ± 0.2 V have been obtained from cleaved wedges and cubes of MgO, GaAs, PbS and Ge, respectively.[24–26] By analysing GaAs, InAs, GaP and InP wedges, values of 14.2 ± 0.2, 14.5 ± 0.2, 14.4 ± 0.2, 14.5 ± 0.1 V, respectively, were obtained,[27] while wedge-shaped Si samples with stacked Si oxide layers on their surfaces were used to measure the mean inner potentials of the oxide layers.[28] Si nanospheres coated in layers of amorphous SiO_2 have been used[29] to determine values of 12.1 ± 1.3 V, 11.9 ± 0.9 V and 10.1 ± 0.6 V for crystalline Si, amorphous Si and amorphous SiO_2, respectively. Similar measurements obtained from spherical latex particles embedded in vitrified ice were used to provide values of 8.5 ± 0.7 and 3.5 ± 1.2 V for the two materials, respectively,[30] while cylindrical GaN nanowires suspended over holes in a carbon support film were used[31] to measure a value for V_0 of GaN of 16.7 ± 0.3 V. Care is always required when measuring V_0 using electron holography, as dynamical contributions to the phase shift can complicate the determination of V_0 from crystalline samples. Equation (5.8) is no longer valid when the sample is tilted to a strongly diffracting orientation, as the phase shift can then vary nonlinearly with sample thickness, and is also very sensitive to small changes in sample orientation. Such effects can be simulated by using either multislice or Bloch-wave algorithms. Additional experimental factors that may affect measurements of V_0 include the chemical and physical state and the crystallographic orientation of the specimen surface[32] and specimen charging.[33–35]

If V_0 is known, then measurements of phase shift can be used to determine the local specimen thickness, t. Examples of the measurement of specimen shapes from phase images include the characterisation of faceted ZrO_2 crystals,[36] carbon nanotubes[37] and bacterial flagellae.[38] Such measurements can, in principle, be extended to three dimensions by combining electron holography with electron tomography, as demonstrated by the analysis of tilt series of electron holograms of latex particles.[39]

The specimen thickness can also be inferred from a holographic amplitude image in units of λ_{in}, the mean free path for inelastic scattering, by making use of the expression

$$\frac{t(x)}{\lambda_{\text{in}}} = -2 \ln\left(\frac{A_i(x)}{A_r(x)}\right) \tag{5.10}$$

where $A_{\text{i}}(x)$ and $A_r(x)$ are the amplitudes of the sample and reference holograms, respectively.[40] If desired, the thickness dependence of both the phase and the amplitude image can then be removed by combining eqns (5.8) and (5.10) as

$$\frac{\phi(x)}{-2C_E \ln\left(\frac{A_i(x)}{A_r(x)}\right)} = V_0(x)\lambda_{\text{in}}(x) \tag{5.11}$$

to generate an image, in which the contrast is the product of the local values of the mean inner potential and the inelastic mean free path. These parameters depend only on the local composition of the sample. Equation (5.11) can therefore be useful for interpreting images obtained from samples with varying composition and thickness.[41]

5.3 Measurement of Magnetic Fields

When examining magnetic materials using electron holography, the conventional TEM objective lens is usually switched off, as its strong magnetic field would be likely to saturate the magnetisation in the sample in the electron-beam direction. Instead, a high-strength minilens (a Lorentz lens) located below the objective lens can be used to provide high magnification ($\sim$50–75 k$\times$) with the sample either in a magnetic-field-free environment or in a chosen (precalibrated) vertical magnetic field provided by using an intermediate setting of the TEM objective lens.

5.3.1 Early Experiments

Early examples of the examination of magnetic materials using electron holography involved the reconstruction of electron holograms using a laser bench, and included the characterisation of horseshoe magnets,[42] magnetic recording media[43] and flux vortices in superconductors.[44–46] The most elegant of these experiments involved the confirmation of the Aharonov–Bohm effect,[47] which states that when an electron wave from a point source passes on either side of an infinitely long solenoid the relative phase shift that occurs between the two parts of the wave results from the presence of a vector potential. The Aharonov–Bohm effect provides the only observable confirmation of the physical reality of gauge theory. Electron holography experiments were carried out on 20 nm thick permalloy toroidal magnets that were covered with 300 nm thick layers of superconducting Nb, which prevented electrons from penetrating the magnetic material and confined the magnetic flux by exploiting the Meissner effect. The observations showed that the phase difference between the centre of the toroid and the region outside it was quantised to a value of 0 or π when the temperature was below the Nb superconducting critical temperature (5 K), *i.e.* when a supercurrent was induced to circulate in the magnet.[48,49] Although the observed quantisation of magnetic flux, and the measured phase difference with the magnetic field entirely screened by the superconductor, appeared to provide confirmation of the Aharonov–Bohm effect, alternative explanations for these observations are still proposed.[50]

5.3.2 Experiments Involving Digital Acquisition and Analysis

Recent applications of electron holography to the characterisation of magnetic fields in nanostructured materials have almost invariably made use of digital

recording. The off-axis mode of electron holography, combined with digital image processing, is ideally suited to the characterisation of such materials because unwanted contributions to the contrast from local variations in composition and specimen thickness can be removed from a phase image more easily than from images recorded using other phase-contrast techniques. For example, the Fresnel and Foucault modes of Lorentz microscopy and differential phase contrast (DPC) imaging provide signals that are approximately proportional to the first or the second differential of the phase shift. These techniques enhance contributions to the contrast from rapid variations in specimen thickness and composition, as compared to the weak and slowly varying magnetic signal of primary interest.

Of particular interest for magnetic materials is the digital determination of the gradient of a phase image. If V and $B_\perp$ do not vary in the electron-beam direction in the specimen, then the phase gradient can be written in the form

$$\frac{\mathrm{d}\phi(x)}{\mathrm{d}x} = C_E \frac{\mathrm{d}}{\mathrm{d}x}\{V(x)t(x)\} - \left(\frac{e}{\hbar}\right) B_\perp(x)t(x) \tag{5.12}$$

According to eqn (5.12), for a specimen of uniform thickness and composition, the phase gradient is directly proportional to the inplane component of the magnetic induction in the specimen, *i.e.*

$$\frac{\mathrm{d}\phi(x)}{\mathrm{d}x} = -\left(\frac{et}{\hbar}\right) B_\perp(x) \tag{5.13}$$

A graphical representation of the strength and direction of the local magnetic induction can therefore be obtained simply by adding contours to a recorded phase image. A phase difference of 2π between such contours then corresponds to an enclosed magnetic flux of 4×10^{-15} Wb. Unfortunately, even if the first term in eqn (5.12) is negligible, great care is required when interpreting contoured phase images acquired from general specimens, in which the region of interest may have variable thickness and finite lateral extent, because the gradient of the magnetic contribution to the phase shift then contains contributions from the internal demagnetising field and stray magnetic fields outside the sample.

The primary advantage of the digital analysis of electron holograms is that the magnetic and mean inner potential contributions to the measured phase shift can be separated, particularly at the edges of magnetic nanoparticles, where rapid variations in specimen thickness can dominate both the phase and the phase gradient. Several approaches can be used to achieve this separation. The sample may be inverted to change the sign of the magnetic contribution to the phase and a second hologram recorded. The sum and the difference of the two phase images can then be used to provide twice the magnetic contribution, and twice the mean inner potential contribution, respectively.[51,52] Alternatively, two holograms may be acquired from the same region of the specimen at two different microscope accelerating voltages. In this case, the magnetic signal is independent of accelerating voltage, and subtraction of the two phase images can be used to

provide the mean inner potential contribution. A more practical method often involves performing a magnetisation reversal experiment *in situ* in the electron microscope, and subsequently selecting pairs of holograms that differ only in the (opposite) direction of the magnetisation in the specimen. Just as when turning the specimen over, the magnetic and mean inner potential contributions to the phase shift are calculated by taking half the difference, and half the sum, of the resulting phase images. The mean inner potential contribution can then be subtracted from all other phase images acquired from the same specimen region.[53] *In situ* magnetisation reversal, which is required both for this purpose and for performing magnetisation reversal experiments in the TEM, can be achieved by exciting the conventional microscope objective lens slightly and tilting the specimen to apply known inplane magnetic fields, as shown schematically in Figure 5.1(b). In practice, if the two remanent magnetic states are not exactly equal and opposite to each other, then it may be necessary to repeat the switching process several times so that nonsystematic differences between switched pairs of phase images average out. Such differences, which can lead to artifacts in the final magnetic-induction map, are often identified by inspection. By varying the applied field, it is possible to record a series of images that correspond to any desired point on a remanent hysteresis loop or magnetisation reversal cycle.

5.3.2.1 *Isolated Magnetic Nanoparticles*

Analytical expressions for the phase shift of a magnetic particle, based on eqn (5.5), can be derived for a uniformly magnetised sphere of radius a, magnetic induction $B_\perp$ (along y) and the mean inner potential V_0 in the form

$$\phi(x,y)\big|_{x^2+y^2\le a^2} = 2C_E V_0\sqrt{a^2-(x^2+y^2)} + \left(\frac{e}{\hbar}\right)B_\perp a^3\left(\frac{x}{x^2+y^2}\right)\left\{1-\left(1-\left(\frac{x^2+y^2}{a^2}\right)\right)^{\frac{3}{2}}\right\} \tag{5.14}$$

$$\phi(x,y)\big|_{x^2+y^2>a^2} = \left(\frac{e}{\hbar}\right)B_\perp a^3\left(\frac{x}{x^2+y^2}\right) \tag{5.15}$$

Graphical representations of eqns (5.14) and (5.15) are shown in Figure 5.3 for a uniformly magnetised 100 nm diameter spherical particle of iron, on the assumption that $V_0 = 22$ V and $B_\perp = 2.2$ T. The total phase shift (Figure 5.3(c)) is the sum of mean inner potential (Figure 5.3(a)) and magnetic (Figure 5.3(b)) contributions. Line profiles, generated from Figures 5.3(a)–(c) along a line passing through the centre of the particle in a direction perpendicular to its magnetisation direction, are shown in Figures 5.3(d)–(f). As mentioned above, the inplane component of the magnetic induction integrated in the electron-beam direction can be visualised by adding contours to the magnetic contribution to the phase shift ϕ_{MAG}, as shown in Figure 5.3(h) in the form of the

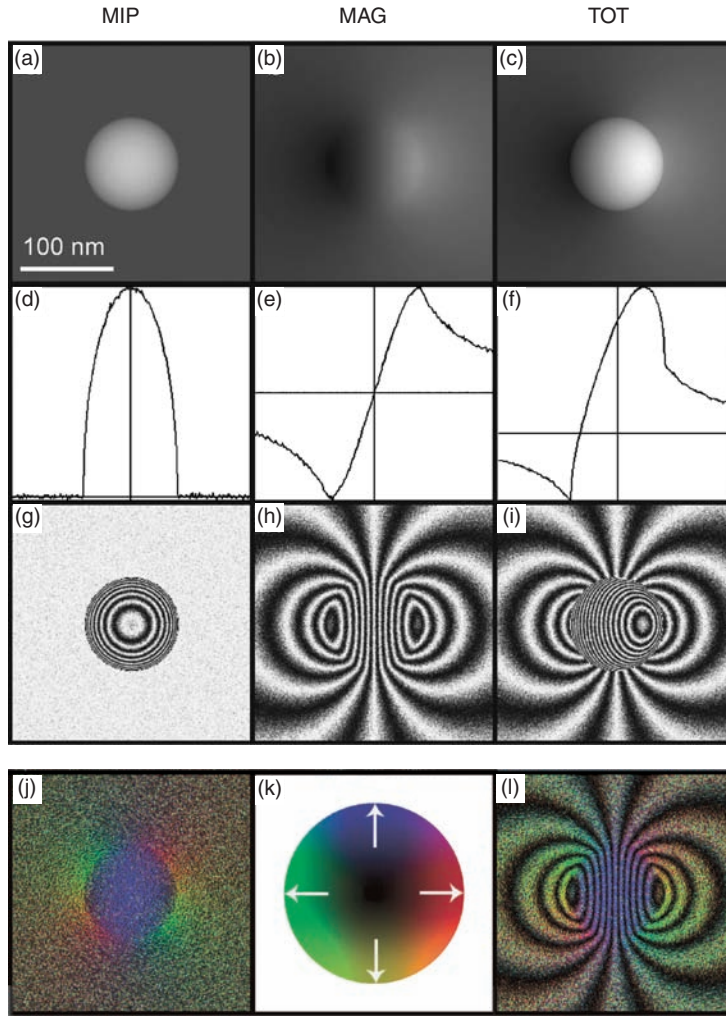


Figure 5.3 Simulations of the phase shift associated with a 100 nm diameter spherical particle of iron. The particle is assumed to be magnetised uniformly in the vertical direction, with $V_0 = 22$ V and $B_\perp = 2.2$ T. The mean inner potential contribution to the phase shift is shown in (a), the magnetic contribution in (b), and the sum of the two contributions in (c). A small amount of noise has been added to the simulations. (d)–(f) show line profiles, taken from images (a)–(c) horizontally through the centre of the particle (*i.e.* in a direction normal to the magnetisation direction). The analytical forms of these graphs are given by eqns (5.14) and (5.15). In (d) min = −0.3 max = 14.5; in (e) min = −9.2 max = 9.1; in (f) min = −8.7 max = 18.7. (g)–(i) Magnetic phase contours (4× amplification; 1.57 rad spacing), corresponding to the cosine of four times each of the phase images shown in (a)–(c). (j) Colour map derived from the gradient of the magnetic contribution to the phase shift shown in (b). The hue and intensity of the colour are used to label the direction and magnitude, respectively, of the inplane component of the magnetic induction integrated in the electron-beam direction, using the colour wheel shown in (k) (red = right, yellow = down, green = left, blue = up). The colours can be combined with the contour map, as shown in (l).

cosine of four times the phase image. The horizontal and vertical derivatives of the magnetic contribution to the phase shift ($d\phi_{MAG}/dx$ and $d\phi_{MAG}/dy$) can be used to generate either an arrow map or a colour map (Figure 5.3(j)), in which the direction and magnitude of the projected inplane magnetic induction are represented by the hue and intensity of the colour, respectively, according to the colour wheel shown in Figure 5.3(k). The colours can also be combined with the cosine image if desired (Figure 5.3(l)). Equations (5.14) and (5.15) have been used to interpret phase images recorded from Co particles suspended over a hole in a carbon support film.[54] In this study, the magnetic induction and mean inner potential of each particle were determined by fitting experimental line traces to simulations similar to that shown in Figure 5.3(f).

Figure 5.4 shows experimental results obtained using electron holography from an isolated 50 nm diameter single crystal of magnetite (Fe_3O_4) from a bacterial cell. Figure 5.4(a) shows a high-resolution TEM image of the crystal. The three-dimensional morphology and orientation of the crystal (Figure 5.4(b)) were determined by applying electron tomography[56] to a series of High-Angle Annular Dark-Field (HAADF) images of the particle taken over a high range of specimen tilt angles. The tomographic reconstruction reveals that the particle is elongated slightly in the [111] direction in the plane of the specimen, as indicated by the white arrow in Figure 5.4(a). Electron holograms of the magnetite crystal were acquired in magnetic-field-free conditions, both at room temperature (Figure 5.4(c)) and at 90 K (Figure 5.4(d)). The magnetic contribution to the phase shift was isolated by performing a series of *in situ* magnetisation reversal experiments, as described previously. The direction of the inplane component of the applied field is indicated by the black double arrow. Both images show uniformly magnetised single-domain states, including the characteristic return flux of an isolated magnetic dipole (Figure 5.3(h)). In both cases, the remanent magnetisation direction appears to make a large angle to the applied field direction. At room temperature, the phase contours in the crystal make an angle of $\sim 30^\circ$ to its [111] elongation direction (Figure 5.4(c)). The contours appear to be parallel to the [111] elongation direction at 90 K (below the Verwey transition for magnetite; Figure 5.4(d)). Figure 5.4(e) shows a line profile generated from the magnetic phase image that was used to create Figure 5.4(c), taken along a line passing through the centre of the crystal in a direction perpendicular to the phase contours. A least-squares fit of the experimental phase profile to eqns (5.14) and (5.15) yielded a value for $B_\perp$ of 0.6 ± 0.12 T. This value is equal to the room-temperature saturation induction of magnetite, suggesting that the magnetisation direction of the particle lies in the plane of the specimen, close to the [131] crystallographic direction. This direction corresponds to the longest diagonal dimension of the particle, which is consistent with shape anisotropy dominating the magnetic state of the crystal at room temperature. The 90 K phase profile yielded a value for $B_\perp$ of 0.46 ± 0.09 T. This value is lower than the saturation induction of magnetite at 90 K, suggesting that, at remanence, the magnetisation direction in the crystal is tilted out of the plane by $\sim 40^\circ$ to the horizontal. This direction is close to either [210] or [012]. Below the Verwey transition, the

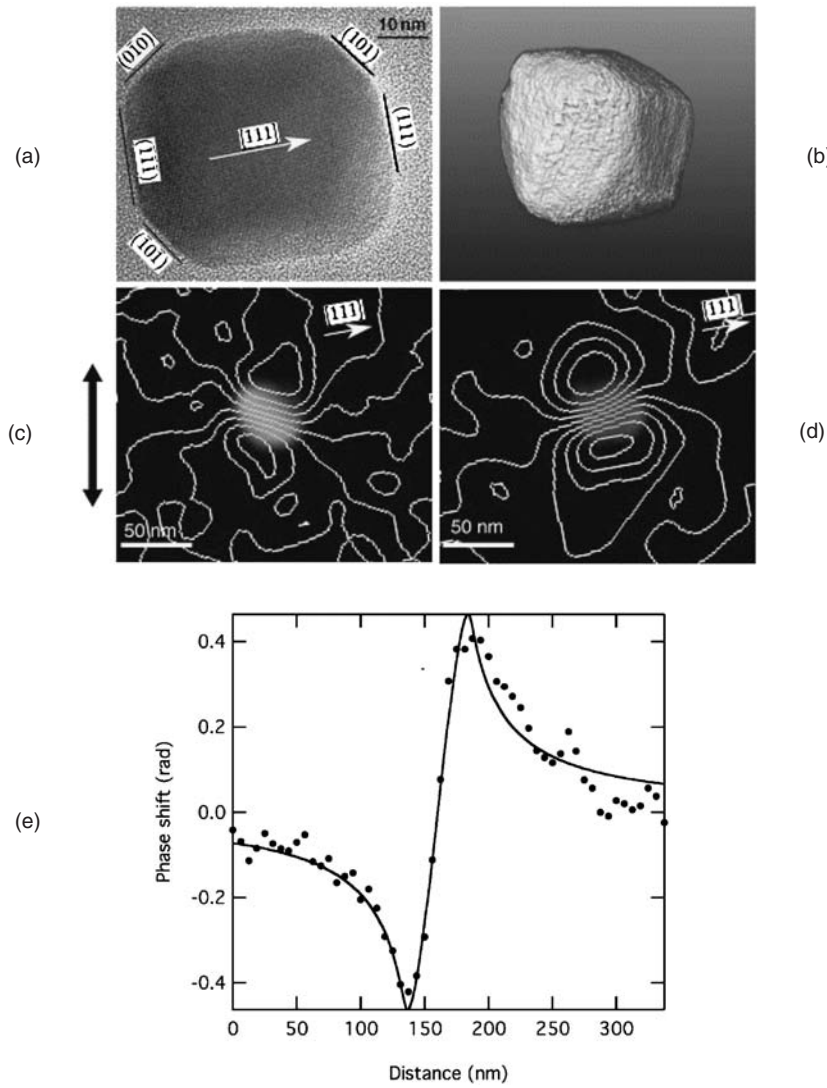


Figure 5.4 (a) High resolution image of a faceted 50 nm diameter magnetite crystal in a magnetotactic bacterium. (b) Visualisation of a high-angle annular dark-field tomographic reconstruction of the same particle. (c) and (d) Magnetic-induction maps showing remanent states recorded from the particle at room temperature and at 90 K, respectively. The images were recorded after tilting the sample by $\pm 30^{\circ}$ and applying a vertical 2 T magnetic field using the TEM objective lens. The inplane component of the applied field was directed along the black double arrow. The mean inner potential contribution to the phase shift has been removed from each image (see text for details). (e) Line profile of the magnetic contribution to the phase shift recorded from the magnetite particle at 90 K (filled circles), taken through the centre of the particle in a direction normal to the contours shown in (d). The solid line shows a least-squares fit of eqns (5.14) and (5.15) to the data. Such fits were used to provide values for the inplane component of B of 0.60 ± 0.12 T at room temperature and 0.46 ± 0.09 T at 90 K.

magnetocrystalline anisotropy of magnetite is known to increase considerably in magnitude, and to switch from $<111>_{\text{cubic}}$ to $[001]_{\text{monoclinic}}$. The $[001]_{\text{monoclinic}}$ easy axis can lie along any one of the original $<100>_{\text{cubic}}$ directions. Both the [100] and the [001] direction of the original cubic crystal lie close to the observed remanence direction, suggesting that magnetocrystalline anisotropy has a more significant effect on the remanence direction at 90 K than at room temperature. The fact that the remanence direction in Figure 5.4(d) is perpendicular to the applied field direction suggests that this choice may be influenced by the morphology of the crystal.

When acquiring electron holography results similar to those shown in Figure 5.4, it is important to recognise that a sample with uniaxial anisotropy constant, K and saturation magnetisation, M_S that is tilted by an angle of 30° is expected to reverse magnetically when the vertical field reaches 0.52 B_K, where $B_K = 2K/M_S$ is the coercivity for fields applied along the anisotropy axis,[55] as shown in the form of a schematic diagram in Figure 5.5. It is also important to relate experimentally acquired magnetic-induction maps, such as those shown in Figures 5.4(c) and (d), to theoretical predictions of the effect of the size and shape of a particle on its magnetic state. In Figure 5.6, the upper solid line shows the theoretical boundary between single domain and two domain states for magnetite nanoparticles.[57] The dashed line shows the boundary between single-domain and single-vortex states, as predicted by micromagnetic simulations.[58]

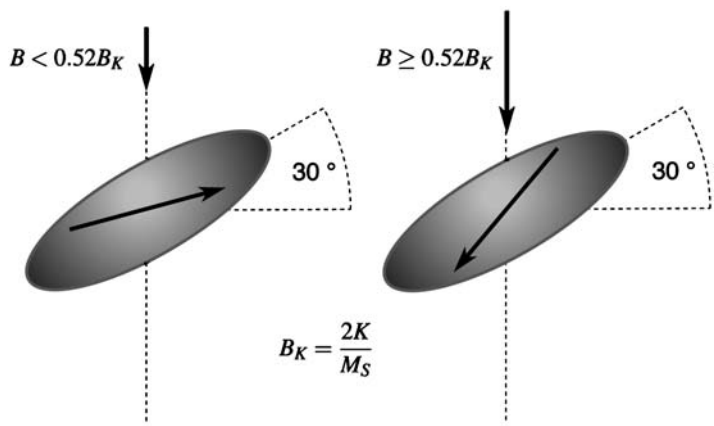


Figure 5.5 Schematic illustration of magnetic switching of a nanoparticle in the TEM, achieved by tilting the sample by 30° and applying a magnetic field using the conventional microscope objective lens. A uniaxial particle with anisotropy constant, K and saturation magnetisation, M_S, initially oriented horizontally and magnetised to the right, is tilted by an angle of 30° to the horizontal. A chosen current is passed through the objective lens of the TEM, exposing the sample to a downward magnetic field of up to 2 T. The direction of magnetisation in the particle is predicted to switch when the vertical field reaches 0.52 B_K, where $B_K = 2K/M_S$. The objective lens is then switched off and the sample tilted back to the horizontal.

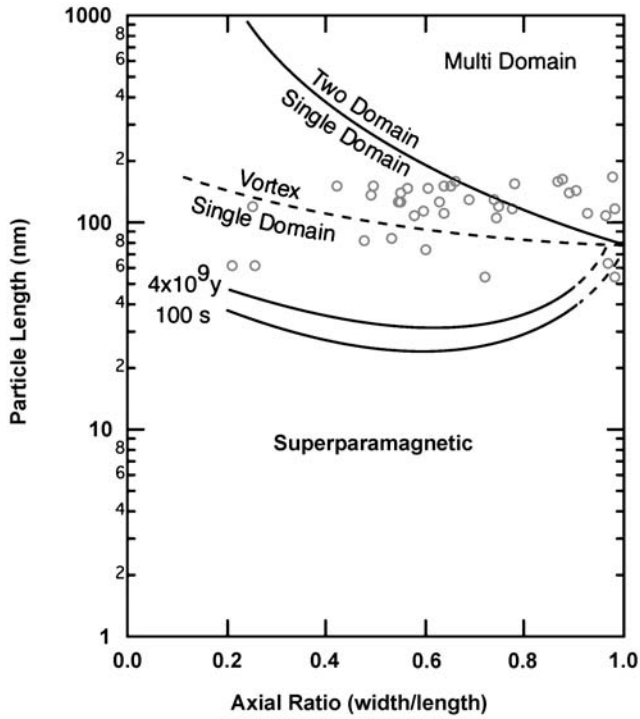


Figure 5.6 Calculated equilibrium threshold sizes for superparamagnetic, single-domain, vortex and two-domain magnetic states in magnetite particles, shown as a function of particle length and axial ratio. The upper solid line shows the boundary between single-domain and two-domain states.[57] The lower solid lines show the sizes for the onset of superparamagnetic behaviour, with relaxation times of 4×10^9 years and 100 s.[57] The dashed line shows the boundary between single-domain and vortex states for uniaxial ellipsoidal particles, calculated using finite element micromagnetic methods.[58] Open circles show the sizes and aspect ratios of closely spaced magnetite blocks in an exsolved titanomagnetite inclusion imaged using electron holography by Harrison *et al.*[74] Reproduced from Ref. 57.

For isolated equidimensional particles, the equilibrium single-domain–single-vortex transition is predicted[59,60] to occur at a particle size of 70 nm, while the transition to a superparamagnetic state is expected[61] to occur below a particle size of 25–30 nm. The observation of a stable single-domain state in the approximately equidimensional 50 nm diameter crystal shown in Figure 5.4 is in agreement approximately with the expected behaviour.

5.3.2.2 Nanoparticle Rings

An illustration of the characterisation of magnetostatic interactions between particles that each contain single magnetic domains is provided by the

examination of rings of 20 nm diameter crystalline Co particles. Such rings are appealing candidates for high-density information-storage applications because they are expected to form chiral domain states that exhibit flux closure. The magnetisation directions of the present nanoparticle rings cannot be reversed by applying an inplane external field. As a result, phase images were obtained both before and after turning the specimen over. The resulting pairs of phase images were aligned in position and angle, and their sum and difference calculated as described previously. Figure 5.7(a) shows a low-magnification bright-field image of the Co rings.[62] Many self-assembled structures are visible, including five- and six-particle rings, chains and closely packed aggregates. Figures 5.7(b)–(d) show magnetic flux closure states in four different rings, measured using electron holography at room temperature in magnetic-field-free conditions.[63] The magnetic flux lines, which are formed from the cosine of 128 times the magnetic contribution to the measured phase shift, reveal the inplane induction within each ring ensemble. Subsequent electron holography experiments have been used to show that the chirality of each flux closure state can be switched *in situ* in the TEM by using an out-of-plane magnetic field.

5.3.2.3 Nanoparticle Chains

There are few experimental measurements of the critical sizes at which magnetic particles that are arranged in chains are large enough to support magnetic vortices rather than single domains. Previous electron holography studies of magnetic nanoparticle chains[64,65] did not provide direct images of such vortex states. Figures 5.8(a) and (b) show the remanent magnetic states of two chains of $Fe_{0.56}Ni_{0.44}$ particles, whose average diameter of 50 nm is expected to be close to the critical size for vortex formation.[66] For a 75 nm $Fe_{0.56}Ni_{0.44}$ particle sandwiched between two smaller particles (Figure 5.8(a)), closely spaced contours run along the chain in a channel of width 22 ± 4 nm. A comparison of this result with micromagnetic simulations[66] indicates that the particle contains a vortex, whose axis is parallel to the chain axis, as shown schematically in Figure 5.8(c). In Figure 5.8(b), a vortex can be seen end-on in a 71 nm particle at the end of a chain. The positions of the particle's neighbours determine the handedness of the vortex, with the flux channel from the rest of the chain sweeping around the core to form concentric circles (Figure 5.8(d)). The vortex core, which is perpendicular to the chain axis, is only 9 ± 2 nm in diameter. The larger value of 22 nm observed in Figure 5.8(a) results from magnetostatic interactions along the chain. Figure 5.8(e) shows similar particles with an alloy concentration of $Fe_{0.10}Ni_{0.90}$, which contain wider flux channels of diameter $\sim$70 nm, and single-domain states when the particles are below $\sim$100 nm in size. The complexity of such vortex states highlights the importance of controlling the shapes, sizes and positions of closely spaced magnetic nanocrystals for storage and recording applications.

Magnetic-induction maps similar to those shown in Figure 5.8 have been obtained from chains of magnetite (Fe_3O_4) and greigite (Fe_3S_4) crystals in

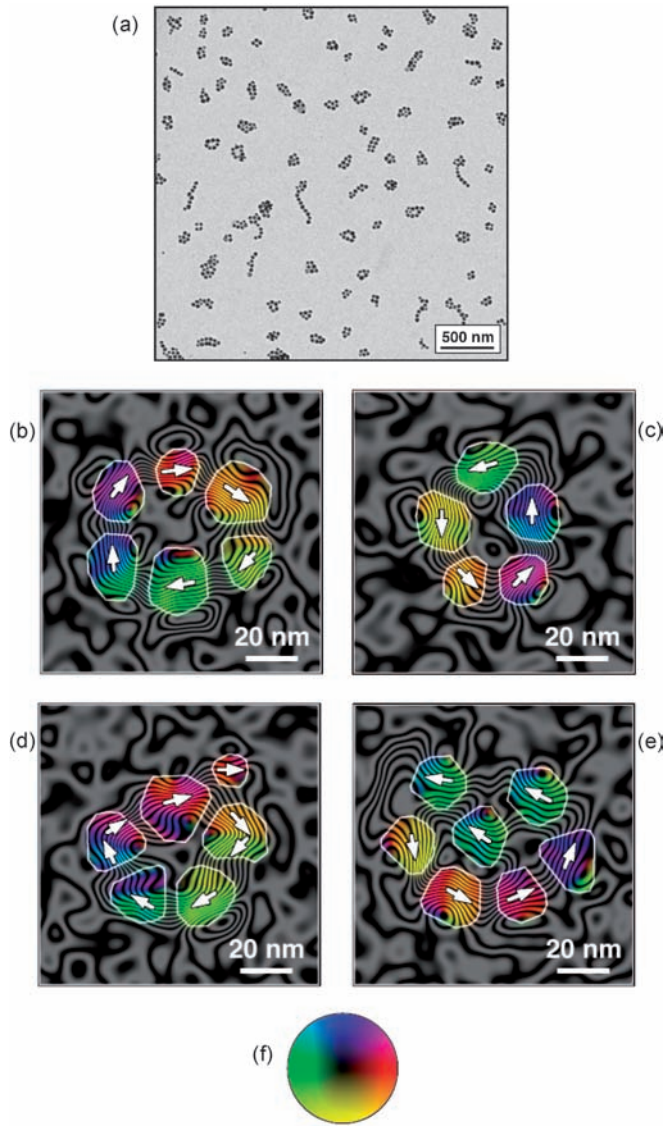


Figure 5.7 (a) Low-magnification bright-field image of self-assembled Co nanoparticle rings and chains that were deposited directly onto an amorphous carbon support film. Each Co particle has a diameter of between 20 and 30 nm. (b)–(e) Magnetic-induction maps showing remanent states in four different nanoparticle rings, in the form of magnetic phase contours (128× amplification; 0.049 rad spacing) formed from the magnetic contribution to the measured phase shift. The mean inner potential contribution to the phase shift has been subtracted from each image (see text for details). The outlines of the nanoparticles are marked in white, while the direction of the measured magnetic induction is indicated both using arrows and according to the colour wheel shown in (f) (red = right, yellow = down, green = left, blue = up). Reproduced from Ref. 63.

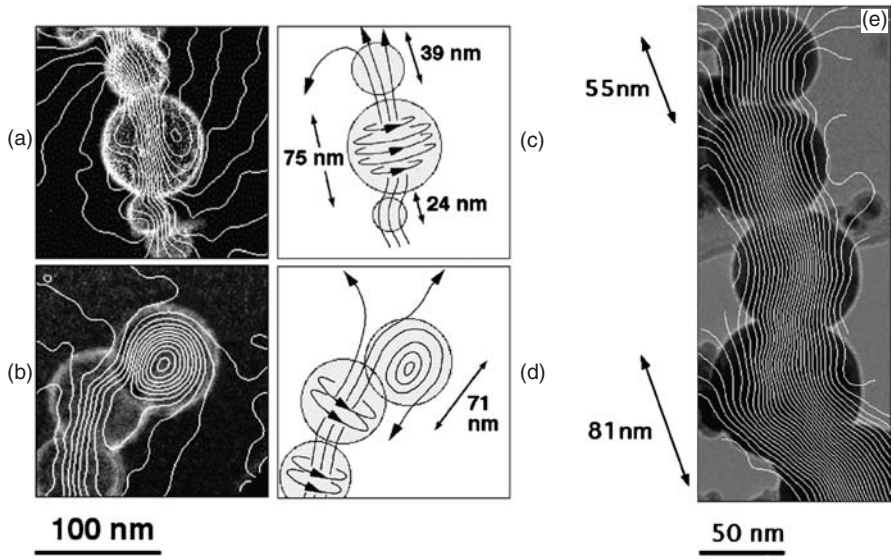


Figure 5.8 (a) and (b) Magnetic phase contours, overlaid onto oxygen chemical maps, showing the inplane magnetic induction (integrated in the electron-beam direction) in chains of $Fe_{0.56}Ni_{0.44}$ particles, recorded with the TEM objective lens switched off. The contour spacings are 0.083 and 0.2 rad for (a) and (b), respectively. The mean inner potential contribution to the phase shift has been removed from each image. (c) and (d) show schematic diagrams of the magnetic microstructure in the chains, in which vortices spin about flux tubes that lie along the chain axes. A vortex, whose axis lies perpendicular to the chain axis, is also visible in (d). (e) A similar map (contour spacing 0.125 rad, overlaid onto a bright-field image) recorded from a chain of $Fe_{0.10}Ni_{0.90}$ particles. Reproduced from Ref. 66.

magnetotactic bacteria.[67–70] The magnetic properties of the bacterial iron oxide and sulfide nanocrystals, which are typically 30 to 120 nm in size and arranged in linear chain configurations, depend on crystal structure, morphology, orientation and spacing. For highly aligned chains of crystals, magnetostatic interactions move the boundary between single domain and vortex states to larger particle sizes, and promote the stability of single-domain states. This effect enables bacteria to grow single-domain crystals to larger sizes than would otherwise be possible, optimising the magnetic moment of the chain. Similarly, crystals that are small enough to be superparamagnetic at room temperature, if they were isolated, are constrained to be single domain by magnetostatic interactions with neighbouring crystals in the chain. Shape anisotropy appears to be the most important factor controlling the magnetic microstructure of these bacterial crystals, followed by interparticle interactions, with magneto-crystalline anisotropy being least important. The strict biological control that results in the presence of uniform morphologies, sizes and orientations of magnetite crystals appears to be much weaker in greigite-producing bacteria.

5.3.2.4 Two-dimensional Nanoparticle Arrays

The magnetic behaviour of the rings and chains of nanomagnets described above contrasts with that of a two-dimensional array of closely spaced crystals. Figure 5.9(a) shows a chemical map, obtained using three-window background-subtracted elemental mapping, of a crystalline naturally occurring magnetite-ulvöspinel (Fe_3O_4-Fe_2TiO_4) inclusion in clinopyroxene,[71,72] in which the distribution of Fe is shown in blue and Ti in red. The sample exsolved during slow cooling yielding an intergrowth of ferrimagnetic magnetite-rich (Fe_3O_4) blocks separated by paramagnetic ulvöspinel-rich (Fe_2TiO_4) lamellae.[73] Magnetic-induction maps obtained from this region, which are shown in Figures 5.9(b) and (c), were recorded by tilting the specimen in zero field and then turning the objective lens on fully to saturate the sample, in order to provide a known starting point from which further fields could be applied. The objective lens was then turned off, the specimen tilted in zero field in the opposite direction and the objective lens excited partially to apply a known inplane field to the specimen in the opposite direction. Finally, the objective lens was switched off and the sample tilted back to 0° in zero field to record each hologram. Figures 5.9(b) and (c) show that the magnetic-domain structure in this sample is both complicated and unexpected. The remanent state of such a system might be expected to involve adjacent blocks being magnetised in an

Figure 5.9 (a) Chemical map of a finely exsolved naturally occurring titanomagnetite inclusion within pyroxene, acquired using a Gatan imaging filter. The blue regions show the positions of magnetite (Fe_3O_4) blocks, which are separated from each other by paramagnetic ulvöspinel (Fe_2TiO_4) lamellae. (b) and (c) show magnetic-phase contours, measured using electron holography from the same region after magnetising the sample using two different values of the magnetic field provided by the conventional microscope objective lens. The inplane component of the applied field was always vertical on the page. The black contour lines show the direction and magnitude of the projected inplane magnetic induction, which can be correlated with the positions of the magnetite blocks (outlined in white). The direction of the measured induction is also indicated using colours and arrows. Each image was acquired with the specimen in magnetic-field-free conditions. The outlines of the magnetite-rich regions are marked in white, while the direction of the measured magnetic induction is indicated using arrows and colours. Reproduced from Ref. 72.

alternating manner along their elongation directions. However, strong interactions between adjacent magnetite blocks (which are separated by ~ 15 nm) constrain their remanent magnetisation direction to lie almost perpendicular to their elongation direction and to the applied-field direction. The chains of blocks form superstates that are magnetised parallel to the chain axes but perpendicular to the easy axes of the individual blocks, even though the inplane component of the applied field was parallel to their elongation directions. In an earlier electron holography study, arrays of magnetite blocks in a similar sample that were expected to display single vortex states at remanence were found to be primarily single-domain states.[74] The majority of the blocks examined in this study, whose dimensions are plotted in Figure 5.6, would be expected to display single vortex states at remanence if they were isolated and at equilibrium. In an array of strongly interacting particles, the demagnetising energy, which normally destabilises the single-domain state with respect to the vortex state in an isolated particle, is reduced greatly.

It is interesting to note that, in the titanomagnetite system, mean inner potential contributions to the measured phase shift can be removed using a different procedure to that used for the rings and chains of particles described above. Although both thickness and composition vary in the region of interest, the different compositions of magnetite and ulvöspinel are compensated by their densities in such a way that their mean inner potentials are equal. As a result, only a thickness correction may be required. The local specimen thickness across the region of interest can be determined in units of inelastic mean free path by using energy–filtered imaging. This thickness measurement can then be used to determine the mean inner potential contribution to the phase shift, which was in turn used to establish the magnetic contribution to the phase shift.

5.3.2.5 *Lithographically Patterned Magnetic Nanostructures*

Specimen preparation can present a challenge for electron holography of nanomagnet arrays. In order to examine square arrays of Ni pillars that were fabricated on Si substrates using interferometric lithography,[75] specimens were prepared for TEM examination using focused ion beam (FIB) milling so that a single row of pillars could be imaged in cross section.[76] In order to protect the pillars from the damaging effects of implantation and physical damage during FIB milling, the sample was first coated with a 200–500 nm layer of amorphous carbon. A 1 μm thick Pt strip was then deposited onto the carbon layer. A single row of pillars was cut out of the array using a 30 kV Ga^+ ion beam by the lift-out method, and placed on a carbon-coated Cu TEM grid. The presence of the layer of amorphous carbon under the protective Pt strip meant that during the final stage of the milling process the Pt fell away from the sample without the need for milling from the top surface. Figure 5.10 shows representative examples of magnetic-induction maps recorded from a single row of 57 nm diameter 115 nm high Ni pillars that were originally fabricated on Si in a square array of side 100 nm. At remanence, the pillars are each in close-to-single-domain magnetic

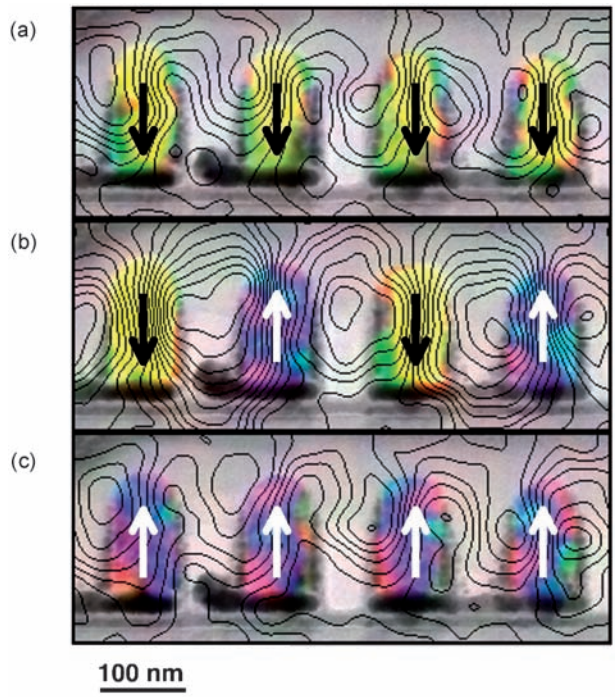


Figure 5.10 Magnetic-induction maps acquired from a single row of Ni pillars prepared using focused ion beam milling from a square array of pillars grown directly onto a Si substrate. Each pillar is 57 nm in diameter, 115 nm in height and separated from its neighbours by 100 nm. The images were acquired in magnetic-field-free conditions, after saturating the pillars downwards by using the conventional TEM objective lens and then applying reverse inplane fields (upwards) of (a) 0 Oe, (b) 450 Oe and (c) 920 Oe. The phase contours have a spacing of 0.049 rad. The mean inner potential contribution to the phase shift has been removed from each image (see text for details). Reproduced from Ref. 76.

states. Pillars that are magnetised parallel to their neighbours appear to show a stronger return flux and a weaker stray magnetic field than pillars that are magnetised antiparallel to their neighbours. Care is clearly required when preparing specimens using FIB milling, as the magnetic moments of the pillars shown in Figure 5.10 are smaller than expected for pure Ni, possibly because of oxidation or damage sustained during specimen preparation. In a similar study, square arrays of Co disks were prepared in plan-view geometry using FIB milling.[77]

Similar results to those shown in Figure 5.10 have been obtained from a wide range of larger lithographically patterned structures that were prepared directly onto electron-transparent substrates, many of which show multidomain behaviour and complicated switching mechanisms.[78–81] Few phase contours are visible outside such elements when they support magnetic flux closure states. Electron holography has also been used to provide information about magnetic

interactions between closely separated ferromagnetic layers within individual spinvalve elements.[82–84] The presence of two different phase contour spacings at different applied fields in such elements is associated with the reversal of the magnetisation directions of the ferromagnetic layers in each element at different fields. The switching field of each layer may also be influenced strongly by magnetostatic interactions between the two layers.

5.3.2.6 Nanowires

An important question relates to the minimum size of a nanostructure in which magnetic fields can be characterised successfully using electron holography. Although 4 nm diameter single-crystalline Co nanowires have been characterised successfully,[85] this measurement is challenging because the mean inner potential contribution to the phase shift at the centre of a 4 nm wire relative to that in vacuum is 0.57 rad (assuming a value for V_0 of 22 V), whereas the step in the magnetic contribution to the phase shift across the wire is only 0.032 rad (assuming a value for B of 1.6 T). Although the magnetic contribution to the measured phase shift for an isolated wire was measured successfully in this study, the closely spaced contours along the wire were not straight, resulting from smoothing a signal that is noisy and weak, as also demonstrated in a more recent study.[86]

5.3.2.7 Quantitative Measurements, Resolution and Micromagnetic Simulations

A particular strength of electron holography is its ability to provide quantitative information about magnetic properties. For example, the magnetic moment of a nanoparticle can be obtained from the expression

$$m_x = \left(\frac{\hbar}{e}\right) \int_{y=-\infty}^{y=+\infty} \int_{x=-\infty}^{x=+\infty} \frac{\partial}{\partial y} \phi_{\mathrm{MAG}}(x,y)\,\mathrm{d}x\mathrm{d}y \tag{5.16}$$

where, ϕ_{MAG} is the magnetic contribution to the phase shift and y is a direction perpendicular to x in the plane of the specimen. According to eqn (5.16), the magnetic moment in a given direction can be obtained by measuring the area under the first differential of ϕ_{MAG} evaluated in the perpendicular direction. The contribution of stray magnetic fields to the moment is included in this calculation if the integration is carried out over a large enough distance from the particle.

The spatial resolution that can be achieved in a phase image is determined primarily by the spacing of the holographic interference fringes. However, the contrast of these fringes decreases as their spacing is reduced, and the recording process is also dominated by Poisson-distributed shot noise.[87] These parameters are affected by the illumination diameter, exposure time and biprism voltage. The final phase resolution[7] and spatial resolution are therefore

inherently linked, as a small phase shift can be measured with high precision and poor spatial resolution, or with poor precision and high spatial resolution. In each of the examples described above, the recorded phase images were smoothed slightly to remove noise, and the spatial resolution of the magnetic information is typically between 10 and 20 nm. This procedure is necessarily subjective, and care is required to ensure that artifacts are not introduced. Higher spatial and phase resolution may also be achieved by recording several holograms of each area of interest and subsequently averaging the resulting phase images.

It is therefore necessary to compare electron holographic measurements with micromagnetic simulations due to the sensitivity of the magnetic domain structure in nanoscale materials and devices to their detailed magnetic history and to their microstructure and chemistry. Differences between starting magnetic states that are too small to be distinguished visually, as well as inter-element coupling and the presence of out-of-plane magnetic fields, are important for the formation of subsequent domain states, and in particular for determining the sense (the handedness) with which magnetic vortices unroll.[88] The sensitivity of the domain structure to such effects, as well as to the three-dimensional physical and chemical microstructure of the specimen, emphasises the need to correlate high-quality experimental holographic measurements with micromagnetic simulations.

5.3.2.8 Cross-sectional Layered Specimens

One of the most challenging problems for electron holography of magnetic materials is the quantitative measurement of the magnetic properties of nm-scale magnetic layers when examined in cross section. In such samples, the weak magnetic signal must be separated from rapid and unknown variations in both composition and thickness. By rearranging eqns (5.8) and (5.12), it can be shown that specimen thickness effects may be eliminated by calculating the difference between the phase gradients of images, between which the magnetisation has been reversed, divided by the average of their phases and multiplied by a constant and by the value of the mean inner potential of each magnetic layer separately. Formally, this procedure can be written as

$$\left(\frac{C_E\hbar V_0(x,y)}{e}\right)\left\{\frac{\Delta[\mathrm{d}\phi(x,y)/\mathrm{d}x]}{\langle\phi(x,y)\rangle}\right\} = \frac{\Delta B_\perp(x,y)}{(1-(e/C_E\hbar V_0(x,y)))\left\{\left\langle\int B_\perp(x,y)t(x,y)\mathrm{d}x\right\rangle/t(x,y)\right\}} \tag{5.17}$$

According to eqn (5.17), by combining appropriate phase profiles and their gradients (evaluated in a direction perpendicular to the layers), both the magnitude and the sign of $\Delta B_\perp(x, y) = 2B_\perp(x, y)$ are obtained exactly if the magnetisation reverses exactly everywhere in the sample (the denominator on the right-hand side of the equation is then unity). Furthermore, nonzero values are returned only in regions where the magnetisation has changed. Equation

(5.17) has been applied successfully to the characterisation of a cross-sectional magnetic tunnel junction sample.[89] Similar approaches have also been used to remove specimen thickness and mean inner potential contributions from phase shifts measured from holograms of $La_xCa_{1-x}MnO_3$ acquired at different temperatures.[90,91]

5.4 Measurement of Electrostatic Fields

In this section, the application of electron holography to the characterisation of electrostatic fields is reviewed. Examples are taken from the characterisation of fringing fields outside biased nanowires, dopant contrast at depletion layers in semiconductors and the characterisation of interfaces at which both charge redistribution and changes in chemistry are possible.

5.4.1 Electrically Biased Nanowires

Early experiments on tungsten microtips demonstrated that electron holography could be used to measure electrostatic fringing fields in biased samples.[92] Further studies were carried out on pairs of parallel 1 μm diameter Pt wires held at different potentials[93] and on single conducting wires.[94] More recent experiments have involved the use of electron holography to map electrostatic potentials around the ends of electrically biased carbon nanotubes. In one such study, a three-axis manipulation electrode was used to position a multi-walled carbon nanotube approximately 6 μm from a gold electrode.[95] Contoured phase images recorded before a bias was applied between the nanotube and the electrode showed a featureless phase gradient profile around the nanotube, whereas closely spaced contours were observed around the nanotube tip when a voltage of up to 120 V was applied. The threshold voltage for field emission was observed to be approximately 70 V. Comparisons with simulations provided a value of 1.22 V nm^{-1} for the electric field at the nanotube tip when the bias voltage was 120 V. This field was stable over time, even when the emission current varied. Results obtained from a similar experiment are shown in Figure 5.11. Figure 5.11(a) shows a design drawing of a side-entry TEM specimen holder that allows specimens to have up to three electrical contacts applied to them *in situ* in the TEM, while allowing tilt angles in excess of 70° to be used for electron tomography. The end of the holder, which contains a removable cartridge that allows two independent electrical contacts to be made to the specimen, is shown in the inset to Figure 5.11(a). A third electrical contact can be moved towards the specimen using micrometers and piezoelectric drives. The removable cartridge is compatible with a FIB workstation, and allows the same specimen to be examined under an applied bias in a scanning electron microscope. Figure 5.11(b) shows a defocused bright-field image of bundles of single-walled carbon nanotubes that were placed in the cartridge shown in Figure 5.11(a). The application of a voltage resulted in attraction of the nanotubes towards a gold electrode and deflection of the incident electrons

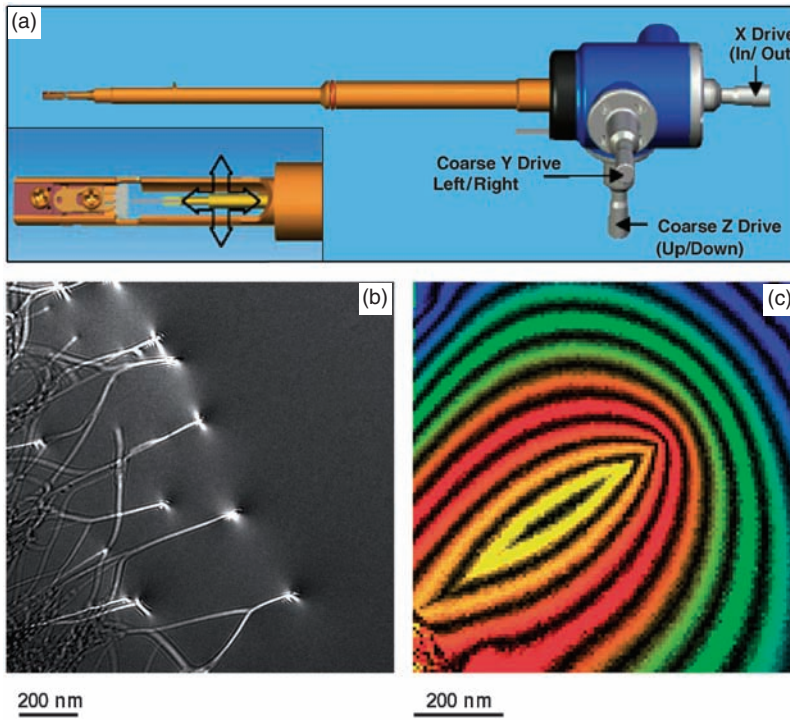


Figure 5.11 (a) Design drawing of an ultrahigh-tilt three-contact cartridge-based electrical biasing nanopositioning TEM specimen holder, with coarse and fine three-axis motion of the moveable contact provided by micrometers and piezoelectric crystals, respectively. The inset shows the location of the cartridge and the position and motion of the third contact. (b) Defocused bright-field image of a specimen containing bundles of single-walled carbon nanotubes, placed in the cartridge shown in (a), with a voltage applied between the tubes and a gold needle that was brought to within 1–2 μm of them in the nanopositioning specimen holder. (c) Contoured phase images obtained from an off-axis electron hologram of a carbon nanotube bundle that has a voltage applied to it *in situ* in the TEM. Reproduced from Ref. 96.

by the electric field at the end of each nanotube bundle. Figure 5.11(c) shows a contoured holographic phase image of the end of a nanotube bundle. The electric field is strongest where the contours are most closely spaced at the end of the bundle.[96]

Under certain circumstances, it may be useful to apply a voltage to a sample to separate a parameter of interest from other, unwanted contributions to the contrast.[96] An example of how this separation may be achieved is shown in Figure 5.12. Figure 5.12(a) shows a holographic amplitude image of two needles, one of Fe and the other of W, which have been brought to a separation of 560 nm using the nanopositioning specimen holder shown in Figure 5.11(a). The acquisition of phase images with different voltages applied between the two

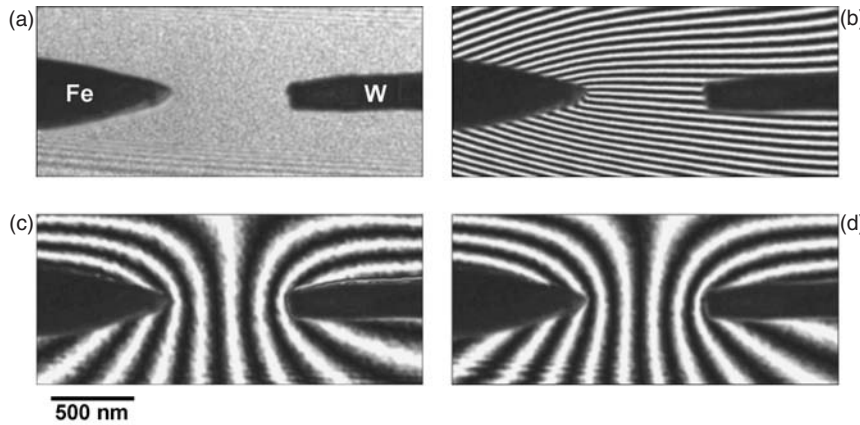


Figure 5.12 (a) Electron holographic amplitude image of the ends of Fe and W needles, which have been positioned with a separation of 560 nm using the nanopositioning specimen holder. (b) Phase contours measured using electron holography showing the inplane component of the magnetic induction integrated in the electron-beam direction. The contour spacing is 2π rad. (c) Phase contours arising from the projected electrostatic potential alone, integrated in the electron-beam direction, with the W needle at $+5$ V with respect to the Fe needle. The contour spacing is 6π rad. (d) As for (c) but with the W needle at -5 V. The slight asymmetry about the horizontal axis in the contours shown in (c) and (d) almost certainly arises because of the long-range electrostatic field, which perturbs the wave that is overlapped onto the sample to form the hologram. Reproduced from Ref. 96.

needles allows the contribution to the phase shift from the electric field between the needles to be separated from magnetic and mean inner potential contributions to the contrast, as well as from any diffraction contrast that may be present (Figure 5.12).

5.4.2 Dopant Potentials in Semiconductors

Off-axis electron holography promises to fulfil the requirements of the semiconductor industry by providing a technique that can be used to provide quantitative information about dopant potentials in semiconductors with nanometer spatial resolution. In a specimen of uniform thickness, the measured phase shift is expected to provide a quantitative measure of the variation in potential associated with the presence of dopant atoms. Attempts to tackle this problem have been made since the 1960s using many forms of electron interferometry, both experimentally[97,98] and theoretically.[99,100] However, TEM specimen preparation is now known to have a profound influence on phase shifts measured from doped semiconductors. In addition to surface depletion resulting from the presence of the specimen surfaces, the electrostatic potential in the specimen may be affected by oxidation, physical damage and the

implantation of Ar and Ga during preparation of the sample for electron microscopy, as well as by irradiation by high-energy electrons during TEM examination. The effects of specimen preparation, and in particular the electrical state of near-surface regions, may account for many of the anomalous results in early experiments. However, recent studies indicate that it may be possible to resolve these problems.

The first unequivocal demonstration of two-dimensional mapping of the electrostatic potential in an unbiased doped semiconductor using electron holography was achieved for metal-oxide-semiconductor (MOS) Si transistors.[101] The source and drain regions were visible in phase images with a spatial resolution of 10 nm and an energy resolution of close to 0.10 eV. Differential thinning was discounted as a cause of the observed phase shifts and an optimal specimen thickness of 200–400 nm was identified for such experiments. The transistors were prepared for TEM examination using conventional mechanical polishing and Ar-ion milling. A 25 nm thick electrically altered layer was identified on each surface of the specimen, which resulted in measured phase shifts across each *p-n* junction that were lower than predicted. More recently, electron holography studies of transistors have been compared with process simulations.[102] In this study, specimens were prepared primarily using tripod wedge polishing, followed by limited low-angle Ar-ion milling at 3.5 kV. Surprisingly, no electrically altered surface layer had to be taken into account to quantify the results.

The electrostatic potential across an abrupt *p-n* junction is shown schematically in Figures 5.13(a) and (b). In Figure 5.13(a), the effect of the unknown electrical state of the specimen surface on the potential is included phenomenologically by assuming the presence of electrically altered layers of uniform thickness on the specimen surfaces. The true junction potential is then assumed to lie within specimen thickness t_{el}, which is smaller than the total specimen thickness, t. The graph in Figure 5.13(b) is drawn on the assumption that the ''transition regions'' on each side of the depletion region are negligibly small – an assumption that is unlikely to be true in practice, but that can be assessed experimentally by calculating the charge density profile from a measured phase profile. It should be noted that care is required both with specimen preparation and with interpretation of measured phase shifts because a small change in specimen thickness (of 5–10%) can result in similar contrast to that across a *p-n* junction.

Recently, electrostatic potential profiles have been measured from reverse-biased Si *p-n* junctions that were prepared for TEM examination using FIB milling.[103–106] FIB milling is the technique of choice for preparing TEM specimens from site-specific regions of integrated circuits. Specimens for *in situ* electrical biasing were prepared by using a 30 kV FEI 200 FIB workstation to machine parallel-sided electron-transparent membranes at the corners of 1×1 mm 90° cleaved squares of wafer. This geometry allowed electrical contacts to be made to the front and back surfaces of each specimen using a modified single-tilt holder.[103] Care was taken to expose the region of interest to the focused beam of Ga ions only at a glancing angle to its surface.

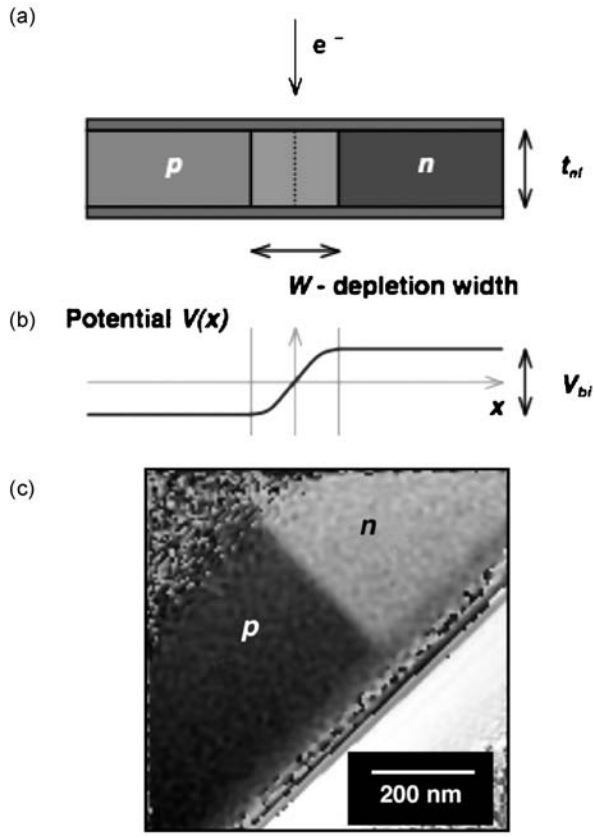


Figure 5.13 (a) Schematic diagram showing the cross-sectional geometry of a TEM specimen of uniform thickness that contains a symmetrical semiconductor *p-n* junction. t_{el} is the "electrically active" specimen thickness and W is the width of the depletion region over which the potential changes. The layers at the top and bottom surfaces of the specimen represent electrically passivated or depleted layers, whose physical and electrical nature is affected by TEM specimen preparation. (b) Schematic diagram of the electrostatic potential profile across the *p-n* junction. V_{bi} is the built-in voltage. The sign convention for the potential is consistent with the mean inner potential of the specimen being positive relative to vacuum. (c) Representative phase image reconstructed from an off-axis electron hologram of a FIB-milled Si *p-n* junction sample. The sample edge is at the lower right of the image. No attempt has been made to remove the phase "wraps" lying along this edge. (d) and (e) Line profiles of the measured phase shift across the Si *p-n* junction, shown as a function of (d) specimen thickness for three different unbiased FIB-milled "trench" specimens and (e) reverse-bias voltage for a single specimen whose crystalline thickness was measured to be 390 nm. The phase images were averaged over a distance of approximately 100 nm on the specimen to form the profiles, in which the zero of phase has been chosen to lie on the left (the *p*-) side of the junction. (f) Schematic diagram showing, in cross section, the physical and electrical structure of a FIB-milled TEM specimen inferred from results such as those shown in (d) and (e). Reproduced from Ref. 104.

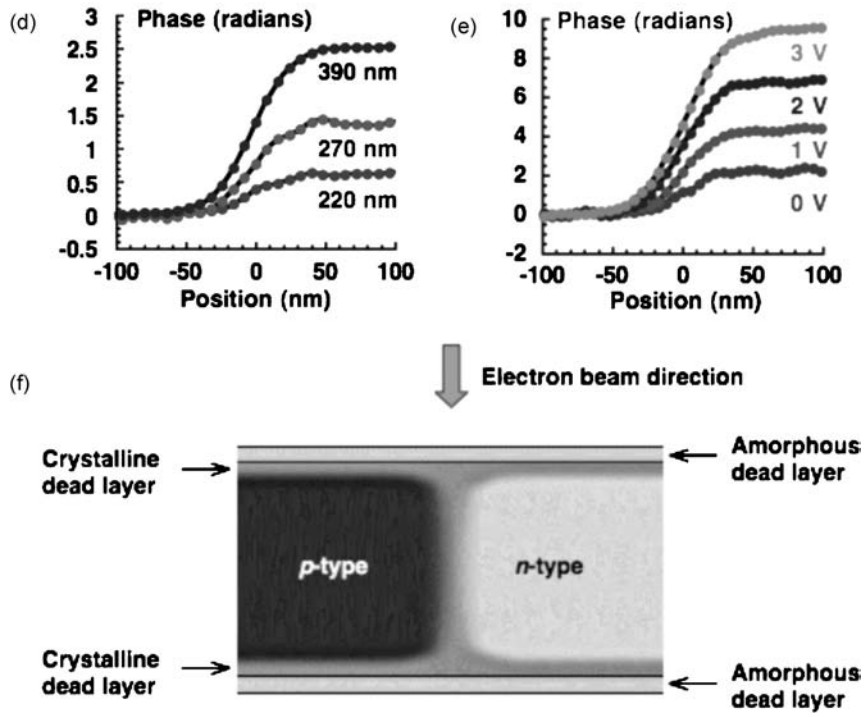


Figure 5.13 Continued

Figure 5.13(c) shows a representative phase image recorded from an unbiased Si *p-n* junction sample, whose thickness was measured to be 550 nm using convergent beam electron diffraction. The *p*-type and *n*-type regions are delineated as areas of darker and lighter contrast, respectively. The additional "grey" band at the specimen edge is likely to be associated with the presence of an electrically altered layer, which is visible in cross section but is thought to extend around the entire specimen surface. Line profiles across the junction were obtained from phase images acquired with different reverse bias voltages applied to a specimen of 390 nm crystalline thickness (Figure 5.13(d)), as well as from several unbiased specimens (Figures 5.13(d) and (e)). Each profile is qualitatively consistent with the expected potential profile for a *p-n* junction in a specimen of uniform thickness.[107] In Figure 5.13(e) the height of the potential step across the junction, $\Delta\phi$, increases linearly with reverse bias voltage V_{appl} according to the equation

$$\Delta\phi = C_E\left(V_{\text{bi}} + V_{\text{appl}}\right)t_{\text{el}} \tag{5.18}$$

where C_E is defined in eqn (5.6) and the *p-n* junction is contained in an electrically active layer of thickness, t_{el} in a specimen of total thickness, t. Analysis indicates that 25 ± 5 nm of the crystalline thickness on each surface

of the TEM specimen is electrically altered. Depletion widths across the junction measured from the line profiles are higher than expected, suggesting that the electrically active dopant concentration in the specimen is lower than it's nominal value. External electrostatic fringing fields were never observed outside any FIB-milled specimens, indicating that their surfaces are equipotentials in both unbiased and biased specimens. Analysis of the results shown in Figures 5.13(c)–(e) suggests that a layered structure may be present in the TEM membrane, with amorphous outer surface layers surrounding inner, crystalline electrically altered surface layers, themselves surrounding crystalline electrically active material, as shown schematically in Figure 5.13(f).

Figure 5.14(a) shows a holographic phase image of a GaAs *p-n* junction that was prepared for TEM examination using FIB milling. In contrast to results obtained from FIB-milled Si specimens of similar thickness, such as that shown

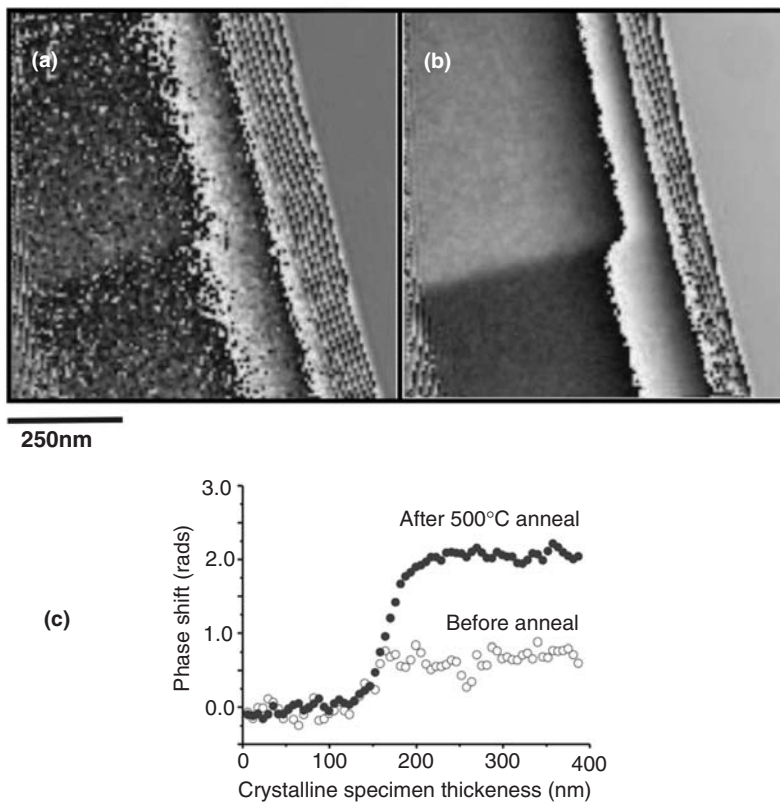
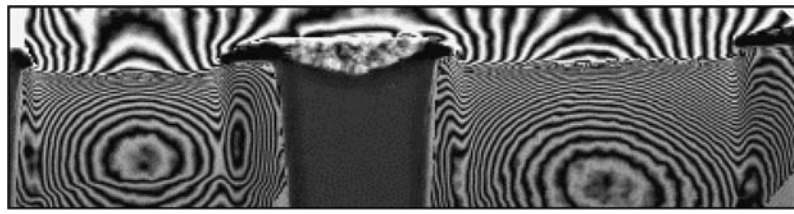


Figure 5.14 (a) and (b) Wrapped electron holographic phase images of a focused ion beam milled GaAs *p-n* junction of crystalline thickness 240 nm, recorded at room temperature (a) before and (b) after *in situ* annealing the specimen at 500°C for 30 min. (c) Phase profiles across a focused ion beam milled GaAs *p-n* junction in a specimen of crystalline thickness 300 nm, measured both before (open circles) and after (closed circles) annealing at 500°C. Reproduced from Ref. 108.

in Figure 5.13(c), the step in phase across the GaAs junction is indistinct and the phase image is noisy. Figures 5.14(b) and (c) show that *in situ* annealing of this specimen in the TEM can be used to increase the phase shift across the junction, while at the same time decreasing noise in the recorded phase image. These results suggest that annealing can be used to remove defects resulting from Ga^+ implantation and to reactivate dopant atoms in the thin specimen. A similar, although smaller, improvement is seen for FIB-milled Si specimens.[108] Although Figure 5.14(a) shows that *in situ* annealing improves recorded phase images and decreases the thickness of the electrically inactive surface layers, the measured built-in voltage across the junction, V_{bi} is still incorrect. Recent work has shown that V_{bi} may be improved for FIB-prepared GaAs specimens by taking care in setting the incident electron beam current during examination (and therefore to the rate at which charge is dissipated from the area of interest) and by providing high-quality electrical contacts to the region of interest. All of these factors, in combination with simulations of electrostatic potentials in TEM specimens,[109,110] must be understood in order to develop electron holography into a technique that can be used to characterise semiconductor dopant potentials reliably.[111–114] Details about the use of electron holography to map transistor structures reliably have recently been presented.[115,116]

Charging effects during electron irradiation of a semiconductor can be seen directly in two dimensions in Figure 5.15, which shows contoured phase images acquired from a specimen that contains a linear array of transistors located several μm below the surface of a wafer and separated from its surface by metallisation and oxide layers. Conventional "trench" FIB milling was used to prepare the specimen, with a nominal thickness of 400 nm. Figure 5.15(a) shows eight-times-amplified phase contours obtained from the edge of the specimen, which contains an array of tungsten contacts separated by regions of silicon oxide. Instead of the expected phase distribution, which should be proportional to the mean inner potential multiplied by the specimen thickness, elliptical contours are visible in each silicon oxide region, and an electrostatic fringing field is present outside the specimen (at the top of Figure 5.15(a)). Both the elliptical contours and the fringing field are associated with the build-up of positive charge in the oxide layers. Figure 5.15(b) shows a similar phase image obtained after coating the specimen on one side with approximately 20 nm of carbon. The effects of charging are now absent, there is no fringing field outside the specimen edge and the phase contours follow the change in specimen thickness. If the charge is assumed to be distributed through the thickness of the specimen, then the electric field in the oxide is approximately 2×10^7 V/m. This value is just below the breakdown electric field[107] for thermal SiO_2 of 10^8 V m^{-1}. If FIB milling from the substrate side of the wafer is used to prepare the specimen, then charging no longer occurs, presumably as a result of Si redeposition onto the specimen surface.[117]

Although questions still remain about phase contrast observed at simple *p-n* junctions, electron holographic data have recently been obtained from more complicated semiconductor device structures, in which changes in composition as well as doping concentration are present, such as a strained



(a)

(b)

Figure 5.15 (a) Eight-times-amplified phase contours (0.785 rad spacing) obtained from a cross-sectional semiconductor device specimen of nominal thickness 400 nm prepared using conventional "trench" focused ion beam milling. The amorphous layers on either side of the W contact are formed from Si oxides of different densities. Charging of the oxide layers results in the presence of electrostatic fringing fields in the vacuum region outside the specimen edge, as well as elliptical phase contours within the oxide layers between adjacent W contacts. (b) shows an equivalent phase image obtained after coating the specimen on one side with approximately 20 nm of carbon to remove the effects of charging. The phase contours now follow the expected mean inner potential contribution to the phase shift in the oxide layers, and there is no electrostatic fringing field outside the specimen edge. Reproduced from Ref. 35.

n–$Al_{0.1}Ga_{0.9}N$/ $In_{0.1}Ga_{0.9}N$/ *p*–$Al_{0.1}Ga_{0.9}N$ heterojunction diode, in which strong piezoelectric and polarisation fields are used to induce high two-dimensional electron gas concentrations[118] and internal electrostatic potentials across InGaN quantum wells with thicknesses ranging from 2 to 10 nm.[119]

5.4.3 Space-Charge Layers at Grain Boundaries

Electron holography has also been used to characterise space-charge layers at doped and undoped grain boundaries in electroceramics. At such boundaries, several contributions to the electron holographic phase shift can complicate interpretation. The space-charge distribution that is predicted to form at such a grain boundary[120] is often described as a double (back-to-back) Schottky barrier. For Mn-doped and undoped grain boundaries in $SrTiO_3$, a decrease

in the measured phase shift at the boundary relative to that in the specimen was observed using electron holography.[121] The changes in phase measured at the doped boundaries were larger in magnitude and spatial extent than at similar undoped boundaries. Possible contributions to the contrast from changes in density, composition, specimen thickness, dynamical diffraction and electrostatic fringing fields were considered and the remaining contributions to the measured phase shifts at the doped boundaries were attributed to space charge. The results were finally interpreted in terms of a narrow (1–2 nm) region of negative grain boundary charge and a wider (3–5 nm) distribution of positive space charge. Defocus contrast has been used[122] to assess possible space-charge contributions to electrostatic potential profiles across grain boundaries in doped and undoped $SrTiO_3$. The contrast observed in these experiments was not consistent with a dominant contribution to the signal from space charge. In similar electron holography studies, variations in electrostatic potential at undoped and Ca-doped grain-boundary dislocations in $YBa_2Cu_3O_{7-x}$ were studied[123] and space charge layers with widths of approximately 150 nm were measured at grain boundaries in ZnO.[124] In an earlier study, defocus contrast recorded from delta–doped layers in Si and GaAs was also attributed to the presence of space charge.[125] Related experiments have been performed[126,127] to measure polarisation distributions across domain boundaries in ferroelectric materials such as $BaTiO_3$ and $PbTiO_3$. There are many opportunities for further work in this area.

5.5 High-Resolution Electron Holography

Aberrations of the objective lens, which result in modifications to the amplitude and phase shift of the electron wave, rarely need to be taken into account when characterising magnetic and electrostatic fields at medium spatial resolution, as described in Sections 5.3 and 5.4. However, these aberrations must be considered when interpreting electron holograms that have been acquired at atomic resolution, in which lattice fringes are visible.

The back focal plane of the objective lens contains the Fraunhofer diffraction pattern, *i.e.* the Fourier transform, of the specimen wave $\psi_s(\boldsymbol{r}) = A_s(\boldsymbol{r}) \exp[i\phi_s(\boldsymbol{r})]$, denoted $\psi(\boldsymbol{q}) = \mathrm{FT}[\psi_s(\boldsymbol{r})]$. Transfer from the back focal plane to the image plane is then represented by an inverse Fourier transform. For a perfect thin lens, neglecting magnification and rotation of the image, the complex image wave would be equivalent to the object wave $\psi_s(\boldsymbol{r})$. Modifications to the electron wave that result from objective-lens aberrations can be represented by multiplication of the electron wavefunction in the back focal plane by a transfer function of the form

$$T(\boldsymbol{q}) = B(\boldsymbol{q}) \exp[i\chi(\boldsymbol{q})] \tag{5.19}$$

In eqn (5.19), $B(\boldsymbol{q})$ is an aperture function that takes a value of unity for q within the objective aperture and zero beyond the edge of the aperture. The

effects of two primary objective lens aberrations, defocus and spherical aberration, are included in the phase factor in the form

$$\chi(\boldsymbol{q}) = \pi\Delta z\lambda q^2 + \frac{\pi}{2}C_S\lambda^3 q^4 \tag{5.20}$$

where Δz is the defocus of the lens and C_S is the spherical aberration coefficient. The complexity of eqn (5.20) increases rapidly as further aberrations are included. The complex wave in the image plane can then be written in the form

$$\psi_i(\boldsymbol{r}) = \mathrm{FT}^{-1}[FT[\psi_s(\boldsymbol{r})] \times T(\boldsymbol{q})] \tag{5.21}$$

$$= \psi_s(\boldsymbol{r}) \otimes t(\boldsymbol{r}) \tag{5.22}$$

where $t(\boldsymbol{r})$ is the inverse Fourier transform of $T(\boldsymbol{q})$, and the convolution $\otimes$ of the specimen wave $\psi_s(\boldsymbol{r})$ with $t(\boldsymbol{r})$ represents the smearing of information that results from lens imperfections. Since both $\psi_s(\boldsymbol{r})$ and $t(\boldsymbol{r})$ are in general complex, the intensity of a conventional bright-field image, which can be expressed in the form

$$I(\boldsymbol{r}) = |\psi_s(\boldsymbol{r}) \otimes t(\boldsymbol{r})|^2 \tag{5.23}$$

is not related simply to the structure of the specimen.

The effects of lens aberrations can be removed by multiplying the complex image wave by a suitable phase plate corresponding to $T^*(\boldsymbol{q})$ to provide the amplitude and the phase shift of the *specimen* wave $\psi_s(\boldsymbol{r})$ rather than the *image* wave $\psi_i(\boldsymbol{r})$. In this way, the interpretable resolution of the image can be improved beyond the point resolution of the electron microscope. The optimal defocus that maximises the resolution of the reconstructed specimen wave after correction of aberrations[128–131] is given by the expression

$$\Delta z_{\mathrm{opt}} = -\frac{3}{4}C_S(\lambda q_{\max})^2 \tag{5.24}$$

where $q_{\max}$ is the maximum desired spatial frequency.

Figure 5.16 illustrates the application of aberration correction to a high-resolution electron hologram of crystalline Si imaged along the [110] zone axis, at which characteristic "dumbbell" contrast, of spacing 0.136 nm, is expected.[132] The original hologram was acquired using an interference fringe spacing of 0.05 nm on a CM30 FEGTEM, which has a point resolution of 0.198 nm and an information limit of 0.1 nm at 300 kV. Figures 5.16(a) and (b) show, respectively, the reconstructed amplitude and phase shift of the hologram after aberration correction using a phase plate. The phase image reveals the expected white "dumbbell" contrast, at a spatial resolution that is considerably better than the point resolution of the microscope, after lens aberrations, including residual astigmatism and off-axis coma, have been measured and removed. The projected atom column positions are visible as dark contrast in the amplitude image. High-resolution electron holography is clearly an exciting area of research, with many recent developments and applications of the technique to a wide range of materials problems.[133,134]

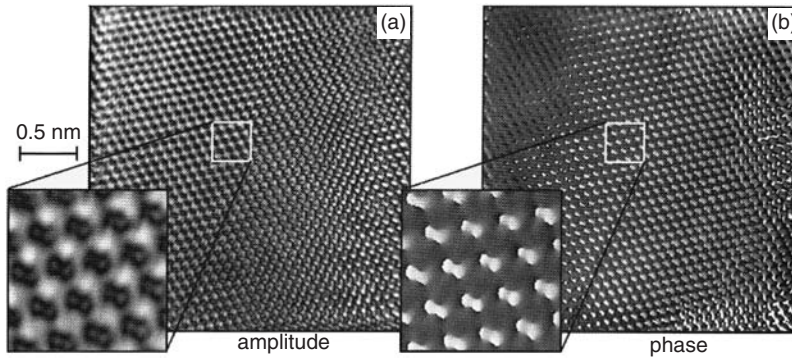


Figure 5.16 High-resolution (a) amplitude and (b) phase images of the aberration-corrected specimen wave reconstructed from an electron hologram of [110] Si, obtained at 300 kV on a CM30 FEGTEM. The spacing of the holographic fringes was 0.05 nm. The sideband contains {111}, {220}, {113} and {004} reflections, corresponding to lateral information of 0.136 nm. The characteristic Si dumbbell structure is visible only after aberration correction. Reproduced from Ref. 132.

5.6 Alternative Forms of Electron Holography

Many different forms of electron holography can be implemented both in the TEM and in the Scanning TEM (STEM).[135] There are also several ways in which the off-axis mode of TEM electron holography can be implemented. A full discussion of these various schemes, which include interferometry in the diffraction plane of the microscope,[136] and reflection electron holography,[137] is beyond the scope of this chapter. However, in this section, a few of the more important developments in this area are reviewed.

The need for a vacuum reference wave is a major drawback of the standard off-axis mode of TEM holography since this requirement restricts the region that can be examined to near the specimen edge. In many applications, the feature of interest is not so conveniently located. The implementation of a Differential Phase Contrast (DPC) mode of electron holography in the TEM enables this restriction to be overcome. DPC imaging is well established as a technique in the STEM, involving the use of various combinations of detectors to obtain magnetic contrast.[138] DPC contrast can also be obtained using far-out-of-focus STEM electron holography[139] (see below). An equivalent TEM configuration can be achieved by using an electron biprism located in the condenser aperture plane of the microscope.[140] The hologram is acquired under out-of-focus conditions, and is in effect the superposition of a pair of Fresnel images. For characterisation of both components of the inplane induction without removing the sample from the microscope, a rotating biprism or a rotating sample holder is required.

An alternative scheme that is conceptually similar to the differential mode of electron holography in the TEM, but which does not require the use of an

electron biprism or a field-emission electron gun, is termed amplitude-division electron holography. Division of the amplitude of the electron wave can be achieved by using a crystal film located before the specimen. The lattice fringes of the crystal film are then used as carrier fringes. The specimen can be inserted into the normal object plane by placing a single-crystal thin film and the sample of interest on top of each other.[141] The single crystal film is tilted to a strong Bragg condition and used as an electron-beam splitter. The hologram plane contains two defocused images of the specimen that are shifted laterally with respect to one another. One of these images is carried by the direct beam and the other by the Bragg-reflected beam. When the distance between the two images is greater than the size of the object, the images separate perfectly and interfere with adjacent plane waves to form an off-axis electron hologram. The defocus of the object can be corrected at the reconstruction stage by using a phase plate, although high coherence of the incident illumination is then required. The coherence used when forming the image therefore determines the spatial resolution of the final reconstructed image.

An approach that can be used to increase the sensitivity of electron holography is termed phase-shifting electron holography. This approach is based on the acquisition of several off-axis holograms while the phase offset (the initial phase) of the image is changed.[142] Electron holograms are recorded at successive values of the incident-beam tilt, such that the phase is shifted by at least 2π over the image series. The advantages of the phase-shifting approach are greatly improved phase sensitivity and spatial resolution. Care is required if the object is out of focus, as tilting the beam will also induce an image shift between successive images. Very small phase shifts have been observed from individual unstained ferritin molecules using this approach.[143]

A real-time approach for acquiring and processing holograms has been demonstrated by using a liquid-crystal panel to reconstruct holograms.[144] Holograms were recorded at video rate and transferred to a liquid-crystal spatial light modulator located at the output of a Mach–Zender interferometer. The liquid crystal panel was illuminated using a He-Ne laser, and interference micrographs were observed at video rate on the monitor beside the microscope as the specimen was examined. In an alternative configuration, a liquid-crystal panel was used as a computer-controlled phase plate to correct for aberrations.

Whereas an off-axis electron hologram is formed by the interference of an object and a reference wave that propagate in different directions in the electron microscope, the simplest way of recording an electron hologram without using an electron biprism involves using the transmitted wave as the reference wave to form an inline hologram. Gabor's original paper described the reconstruction of an image by illuminating an inline hologram with a parallel beam of light and using a spherical aberration correcting plate and an astigmatism corrector. The reconstructed image is, however, disturbed by the presence of a "ghost" or "conjugate" twin image. If the hologram is recorded and subsequently illuminated by a plane wave, then the reconstructed image and a defocused conjugate image of the object are superimposed on each other. The most effective method of separating the twin images is to use Fraunhofer

inline holography. Here, inline holograms are recorded in the Fraunhofer diffraction plane of the object. Under this condition, the conjugate image is so blurred that its effect on the reconstructed image is negligible.[145]

The STEM holographic mode used for DPC imaging, which has similarities with the TEM differential mode of electron holography described above, is a point projection technique in which a stationary beam in a STEM is split by a biprism preceding the sample so that two mutually coherent electron point sources are formed just above the specimen. In this operating mode, the objective lens is excited weakly so that the hologram is formed in the diffraction plane rather than the image plane.[146] By defocusing the objective lens greatly, a shadow image of the object is formed, which has the appearance of a TEM hologram, although it is distorted by spherical aberration and defocus. The image magnification and the separation of the sources relative to the specimen are flexible in this configuration, and can be adjusted by changing the biprism voltage and/ or the objective or post-specimen lens settings. The far-out-of-focus mode of STEM holography has been applied to the characterisation of a range of magnetic materials.[147]

A rapid approach that can be used to visualise equiphase contours involves superimposing a hologram of the specimen onto a vacuum hologram acquired under identical conditions, with the specimen removed from the field of view.[148] Interference between the holographic fringes in the two images then provides widely spaced, low-contrast bands that reveal phase contours directly and by defocusing the combined image, the unwanted finely spaced holographic fringes can be removed.

5.7 Discussion, Prospects for the Future and Conclusions

In this chapter, the technique of off-axis electron holography has been described, and its application to a variety of materials has been reviewed. A selection of results has been presented from the characterisation of magnetic fields in arrangements of closely spaced nanocrystals and patterned elements, to electrostatic fields in field emitters and semiconductors. *In situ* experiments, which allow magnetisation reversal processes to be followed and electrostatic fields in working semiconductor devices to be characterised, have been described, and the advantages of using digital approaches to analyse electron holograms have been highlighted. High-resolution electron holography and alternative modes of holography have also been described. Although the results that have been presented are specific to the dimensions and morphologies of the examples chosen, they illustrate the ways in which holography can be adapted to tackle a range of materials problems.

Future developments in electron holography are likely to include the development and application of new forms of electron holography and instrumentation, the introduction of new approaches for enhancing weak signals and the

formulation of a better understanding of the effect of sample preparation on phase images recorded from semiconductors and ferroelectrics.

A particularly exciting prospect involves the combination of electron holography with electron tomography to image electrostatic and magnetic fields *inside* nanostructured materials in three dimensions rather than simply in projection. This approach has been used to image magnetic fringing fields *outside* materials in three dimensions, by acquiring two ultrahigh-tilt series of electron holograms about orthogonal axes.[149] If each phase image is differentiated in a direction perpendicular to the tilt axis, then standard tomographic reconstruction algorithms can be used to calculate the three-dimensional distribution of the component of $\boldsymbol{B}$ that lies parallel to the tilt axis, based on the relations

$$\frac{\mathrm{d}}{\mathrm{d}y}\phi(x,y) = -\left(\frac{e}{\hbar}\right)\int_{z=-\infty}^{z=+\infty} B_x(x,y,z)\,\mathrm{d}z \tag{5.25}$$

and

$$\frac{\mathrm{d}}{\mathrm{d}x}\phi(x,y) = +\left(\frac{e}{\hbar}\right)\int_{z=-\infty}^{z=+\infty} B_y(x,y,z)\,\mathrm{d}z \tag{5.26}$$

After determining B_x and B_y in three dimensions in this way, B_z can be inferred by making use of the criterion that $\nabla.\boldsymbol{B}=0$. The application of this approach to the characterisation of magnetic fields *inside* nanostructured materials is complicated by the fact that the (often dominant) mean inner potential contribution to the measured phase shift must be removed at each sample tilt angle. This requirement can be achieved if each tilt series is recorded both before and after reversing the direction of magnetisation in the specimen (*e.g.*, using the microscope objective lens). Subsequently, half of the difference between pairs of reversed images acquired at each tilt angle can be used to provide the magnetic contribution to the phase shift. Four tilt series of holograms are therefore required. The fact that eqns (5.25) and (5.26) are expected to hold can be illustrated analytically for a uniformly magnetised sphere.[96] The region of interest must clearly lie close to a large enough hole to allow electron holograms to be acquired at high sample tilt angles about two axes, without either the region of interest or the hole being shadowed by other parts of the specimen. The difficulty of finding such a region means that the distribution of crystals that is imaged should be isolated and small, so that the magnetic signal from the region of interest decreases to close to zero at the edges of the field of view.

In conclusion, the unique capability of electron holography to provide quantitative information about magnetic and electrostatic fields in materials at a resolution approaching the nanometer scale, coupled with the increasing availability of field-emission-gun transmission electron microscopes and quantitative digital recording, ensure that the technique has a very promising future.

Acknowledgements

We are grateful to the Royal Society, the EPSRC, FEI and the Isaac Newton Trust for support, and to T. J. Bromwich, R. F. Broom, P. R. Buseck, R. K. K. Chong, D. Cooper, I. Farrer, J. M. Feinberg, P. E. Fischione, R. B. Frankel, K. Harada, M. J. Hÿtch, P. A. Midgley, S. B. Newcomb, P. Parameswaran, A. K. Petford-Long, M. Pósfai, G. Pozzi, D. A. Ritchie, A. C. Robins, C. A. Ross, M. R. Scheinfein, E. T. Simpson, E. Snoeck, A. C. Twitchett, A. Tonomura, S. L. Tripp, A. C. Twitchett and A. Wei for discussions and ongoing collaborations.

References

1. D. Gabor, *Proc. R. Soc. London A*, 1949, **197**, 454.
2. W.J. de Ruijter and J.K. Weiss, *Ultramicroscopy*, 1993, **50**, 269.
3. D.C. Ghiglia and M.D. Pritt, *Two-Dimensional Phase Unwrapping. Theory, Algorithms and Software*, Wiley, New York, 1998.
4. L. Reimer, *Transmission Electron Microscopy*, Springer-Verlag, Berlin, 1991.
5. W.J. de Ruijter, *Micron*, 1995, **26**, 247.
6. D.J. Smith and M.R. McCartney, in *Introduction to Electron Holography*, ed. E. Völkl, L.F. Allard and D.C. Joy, Kluwer Academic/Plenum Publishers, New York, 1998, p. 87.
7. A. Harscher and H. Lichte, *Ultramicroscopy*, 1996, **64**, 57.
8. K. Yamamoto, T. Tanji and M. Hibino, *Ultramicroscopy*, 2000, **85**, 35.
9. T. Fujita and M.R. McCartney, *Ultramicroscopy*, 2005, **102**, 279.
10. K. Harada, A. Tonomura, Y. Togawa, T. Akashi and T. Matsuda, *Appl. Phys. Lett.*, 2004, **84**, 3229.
11. G. Matteucci, M. Muccini and U. Hartmann, *Phys. Rev. B*, 1994, **50**, 6823.
12. A. Tonomura, L.F. Allard, G. Pozzi, D.C. Joy and Y.A. Ono, ed., *Electron Holography*, Elsevier, Amsterdam, 1995.
13. A. Tonomura, *The Quantum World Unveiled by Electron Waves*, World Scientific, Singapore, 1998.
14. E. Völkl, L.F. Allard and D.C. Joy, ed., *Introduction to Electron Holography*, Plenum Press, New York, 1998.
15. A. Tonomura, *Adv. Phys.*, 1992, **41**, 59.
16. R.E. Dunin-Borkowski, M.R. McCartney and D.J. Smith, in *Encyclopaedia of Nanoscience and Nanotechnology*, ed. H.S. Nalwa, American Scientific Publishers, Stevenson Ranch California, 2004, **3**, 41.
17. P.A. Midgley, *Micron*, 2001, **32**, 167.
18. H. Lichte, *Philos. Trans. R. Soc. Lond. A*, 2002, **360**, 897.
19. G. Matteucci, G.F. Missiroli and G. Pozzi, *Adv. Imag. Electron Phys.*, 2002, **122**, 173.
20. P.A. Doyle and P.S. Turner, *Acta Crystallogr. A*, 1968, **24**, 390.
21. D. Rez, P. Rez and I. Grant, *Acta Crystallogr. A*, 1994, **50**, 481.

22. G. Radi, *Acta Crystallogr. A*, 1970, **26**, 41.
23. M. Gajdardziska–Josifovska and A. Carim, in *Introduction to Electron Holography*, ed. E. Völkl, L.F. Allard and D.C. Joy, Kluwer Academic/Plenum Publishers, New York, 1998, p. 267.
24. M. Gajdardziska–Josifovska, M.R. McCartney, W.J. de Ruijter, D.J. Smith, J.K. Weiss and J.M. Zuo, *Ultramicroscopy*, 1993, **50**, 285.
25. W.J. de Ruijter, M. Gajdardziska, M.R. McCartney, R. Sharma, D.J. Smith and J.K. Weiss, *Scan. Microsc. Suppl.*, 1992, **6**, 347.
26. J. Li, M.R. McCartney, R.E. Dunin and D.J. Smith, *Acta Crystallogr. A*, 1999, **55**, 652.
27. P. Kruse, A. Rosenauer and D. Gerthsen, *Ultramicroscopy*, 2003 **96**, 11.
28. W.D. Rau, F.H. Baumann, J.A. Rentschler, P.K. Roy and A. Ourmazd, *Appl. Phys. Lett.*, 1996, **68**, 3410.
29. Y.C. Wang, T.M. Chou, M. Libera and T.F. Kelly, *Appl. Phys. Lett.*, 1997, **70**, 1296.
30. A. Harscher and H. Lichte, in *Electron Microscopy 98*, ed. H.A. Calderón Benavides and M.J. Yacamán, Institute of Physics Publishing, Bristol UK, 1998, **1**, 553.
31. A.S.W. Wong, G.W. Ho, R.E. Dunin-Borkowski, T. Kasama, R.A. Oliver, P.M.F.J. Costa and C.J. Humphreys, *Mater. Res. Soc. Symp. Proc.*, 2005, **892**, FF11.02.
32. M. O'Keeffe and J.C.H. Spence, *Acta Crystallogr. A*, 1994, **50**, 33.
33. S.J. Lloyd, R.E. Dunin-Borkowski and C.B. Boothroyd, *Inst. Phys. Conf. Ser.*, 1997, **153**, 113.
34. M.R. McCartney, M.A. Gribelyuk, J. Li, P. Ronsheim, J.S. McMurray and D.J. Smith, *Appl. Phys. Lett.*, 2002, **80**, 3213.
35. R.E. Dunin-Borkowski, S.B. Newcomb, T. Kasama, M.R. McCartney, M. Weyland and P.A. Midgley, *Ultramicroscopy*, 2005, **103**, 67.
36. L.F. Allard, E. Völkl, A. Carim, A.K. Datye and R. Ruoff, *Nano. Mater.*, 1996, **7**, 137.
37. X. Lin and V.P. Dravid, *Appl. Phys. Lett.*, 1996, **69**, 1014.
38. K. Aoyama and Q. Ru, *J. Microsc.*, 1996, **182**, 177.
39. G. Lai, T. Hirayama, K. Ishizuka, T. Tanji and A. Tonomura, *Appl. Opt.*, 1994, **33**, 829.
40. M.R. McCartney and M. Gajdardziska-Josifovska, *Ultramicroscopy*, 1994, **53**, 283.
41. J.K. Weiss, W.J. de Ruijter, M. Gajdardziska-Josifovska, D.J. Smith, E. Völkl and H. Lichte, in *Proceedings of the 49th Annual EMSA Meeeting*, ed. G.W. Bailey, San Francisco Press, San Francisco, 1991, p. 674.
42. T. Matsuda, A. Tonomura, R. Suzuki, J. Endo, N. Osakabe, H. Umezaki, H. Tanabe, Y. Sugita and H. Fujiwara, *J. Appl. Phys.*, 1982, **53**, 5444.
43. N. Osakabe, K. Yoshida, Y. Horiuchi, T. Matsuda, H. Tanabe, T. Okuwaki, J. Endo, H. Fujiwara and A. Tonomura, *Appl. Phys. Lett.*, 1983, **42**, 746.

44. T. Matsuda, S. Hasegawa, M. Igarashi, T. Kobayashi, M. Naito, H. Kajiyama, J. Endo, N. Osakabe and A. Tonomura, *Phys. Rev. Lett.*, 1989, **62**, 2519.
45. T. Matsuda, A. Fukuhara, T. Yoshida, S. Hasegawa, A. Tonomura and Q. Ru, *Phys. Rev. Lett.*, 1991, **66**, 457.
46. J.E. Bonevich, K. Harada, T. Matsuda, H. Kasai, T. Yoshida, G. Pozzi and A. Tonomura, *Phys. Rev. Lett.*, 1993, **70**, 2952.
47. Y. Aharonov and D. Bohm, *Phys. Rev.*, 1959, **115**, 485.
48. A. Tonomura, T. Matsuda, R. Suzuki, A. Fukuhara, N. Osakabe, H. Umezaki, J. Endo, K. Shinagawa, Y. Sugita and H. Fujiwara, *Phys. Rev. Lett.*, 1982, **48**, 1443.
49. A. Tonomura, H. Umezaki, T. Matsuda, N. Osakabe, J. Endo and Y. Sugita, *Phys. Rev. Lett.*, 1983, **51**, 331.
50. T.H. Boyer, *Found. Phys.*, 2000, **20**, 893.
51. D.J. Wohlleben, *in Electron Microscopy in Materials Science*, Academic Press, New York, 1971, **2**, 712.
52. A. Tonomura, T. Matsuda, J. Endo, T. Arii and K. Mihama, *Phys. Rev. B: Solid State*, 1986, **34**, 3397.
53. R.E. Dunin-Borkowski, M.R. McCartney, D.J. Smith and S.S.P. Parkin, *Ultramicroscopy*, 1998, **74**, 61.
54. M. de Graef, T. Nuhfer and M.R. McCartney, *J. Microsc.*, 1999, **194**, z84.
55. E.C. Stoner and E.P. Wohlfarth, *Philos. Trans. Roy. Soc. London A*, 1948, **240**, 599.
56. P.A. Midgley, M. Weyland, J.M. Thomas and B.F.G. Johnson, *Chem. Commun.*, 2001, **10**, 907.
57. R.F. Butler and S.K. Banerjee, *J. Geophys. Res.*, 1975, **80**, 4049.
58. W. Williams and A. Muxworthy, *private communication*, 2006. W. Williams, School of GeoSciences, University of Edinburgh, Edinburgh EH9 3JW, UK, Email wyn.williams@ed.ac.uk. A. Muxworthy, Southampton Oceanography Centre, European Way, Department of Earth Science and Engineering, Imperial College, London, SW7 2AZ, UK, Email: adrian.muxworthy@imperial.ac.uk.
59. K. Fabian, A. Kirchner, W. Williams, F. Heider, T. Leibl and A. Hubert, *Geophys. J. Int.*, 1996, **124**, 89.
60. W. Williams and T.M. Wright, *J. Geophys. Res.*, 1998, **103**, 30537.
61. D.J. Dunlop and Ö. Özdemir, *Rock Magnetism*, Cambridge University Press, Cambridge, UK, 1997.
62. S.L. Tripp, S.V. Pusztay, A.E. Ribbe and A. Wei, *J. Am. Chem. Soc.*, 2002, **124**, 7914.
63. S.L. Tripp, R.E. Dunin-Borkowski and A. Wei, *Angew. Chemie*, 2003 **42**, 5591.
64. S. Seraphin, C. Beeli, J.M. Bonard, J. Jiao, P.A. Stadelmann and A. Chatelain, *J. Mater. Res.*, 1999, **14**, 2861.
65. S. Signoretti, L. Del Blanco, L. Pasquini, G. Matteucci, C. Beeli and E. Bonetti, *J. Magn. Magn. Mater.*, 2003, **262**, 142.

66. M.J. Hÿtch, R.E. Dunin-Borkowski, M.R. Scheinfein, J. Moulin, C. Duhamel, F. Mazelayrat and Y. Champion, *Phys. Rev. Lett.*, 2003, **91**, 257207.
67. R.E. Dunin-Borkowski, M.R. McCartney, R.B. Frankel, D.A. Bazylinski, M. Pósfai and P.R. Buseck, *Science*, 1998, **282**, 1868.
68. R.E. Dunin-Borkowski, M.R. McCartney, M. Pósfai, R.B. Frankel, D.A. Bazylinski and P.R. Buseck, *Eur. J. Mineral.*, 2001, **13**, 671.
69. M.R. McCartney, U. Lins, M. Farina, P.R. Buseck and R.B. Frankel, *Eur. J Mineral.*, 2001, **13**, 685.
70. E.T. Simpson, T. Kasama, M. Pósfai, P.R. Buseck, R.J. Harrison and R.E. Dunin-Borkowski, *J. Phys. Conf. Ser.*, 2005, **17**, 108.
71. J.M. Feinberg, H.R. Wenk, P.R. Renne and G.R. Scott, *Am. Mineral.*, 2004, **89**, 462.
72. J. M. Feinberg, R. J. Harrison, T. Kasama, R. E. Dunin-Borkowski, G. R. Scott and P. R. Renne, *J. Geophys. Res.*, 2006, **111**, B12S15.
73. G.D. Price, *Am. Mineral.*, 1981, **66**, 751.
74. R.J. Harrison, R.E. Dunin-Borkowski and A. Putnis, *Proc. Nat. Acad. Sci. USA*, 2002, **99**, 16556.
75. C.A. Ross, *Ann. Rev. Mater. Res.*, 2001, **31**, 203.
76. T.J. Bromwich, T. Kasama, R.K.K. Chong, R.E. Dunin-Borkowski, A.K. Petford-Long, O.G. Heinonen and C.A. Ross, *Nanotechnology*, in press.
77. T.J. Bromwich, A. Kohn, A.K. Petford-Long, T. Kasama, R.E. Dunin-Borkowski, S.B. Newcomb and C.A. Ross, *J. Appl. Phys.*, 2005, **98**, 053909.
78. R.E. Dunin-Borkowski, M.R. McCartney, B. Kardynal, B.S.S.P. Parkin, M.R. Scheinfein and D.J. Smith, *J. Microsc.*, 2000, **200**, 187.
79. M. Heumann, T. Uhlig and J. Zweck, *Phys. Rev. Lett.*, 2005, **94**, 077202.
80. H. Hu, H. Wang, M.R. McCartney and D.J. Smith, *J. Appl. Phys.*, 2005, **97**, 054305.
81. H. Hu, H. Wang, M.R. McCartney and D.J. Smith, *Phys. Rev. B*, 2006, **73**, 153401.
82. D.J. Smith, R.E. Dunin-Borkowski, M.R. McCartney, B. Kardynal and M.R. Scheinfein, *J. Appl. Phys.*, 2000, **87**, 7400.
83. T. Kasama, P. Barpanda, R.E. Dunin-Borkowski, S.B. Newcomb, M.R. McCartney, F.J. Castaño and C.A. Ross, *J. Appl. Phys.*, 2005, **98**, 013903.
84. J.-B. Park, G.-S. Park, I.-Y. Song, J.-S. Bae, J.-E. Lee, J.-H. Yoo, Y. Murakami and D. Shindo, *J. Electron Microsc.*, 2006, **55**, 17.
85. E. Snoeck, R.E. Dunin-Borkowski, F. Dumestre, P. Renaud, C. Amiens, B. Chaudret and P. Zurcher, *Appl. Phys. Lett.*, 2003, **82**, 88.
86. Y. Gao, Y.P. Bao, A.B. Pakhomov, D. Shindo and K.M. Krishnan, *Phys. Rev. Lett.*, 2006, **96**, 137205.
87. H. Lichte, K.H. Herrmann and F. Lenz, *Optik*, 1987, **77**, 135.
88. R.E. Dunin-Borkowski, M.R. McCartney, B. Kardynal, D.J. Smith and M.R. Scheinfein, *Appl. Phys. Lett.*, 1999, **75**, 2641.

89. M.R. McCartney and R.E. Dunin-Borkowski, in *Electron Microscopy 98*, ed. H.A. Calderón Benavides and M.J. Yacamán, Institute of Physics Publishing, Bristol UK, 1998, **2**, 497.
90. J.C. Loudon, N.D. Mathur and P.A. Midgley, *Nature*, 2002, **420**, 797.
91. Y. Murakami, J.H. Yoo, D. Shindo, T. Atou and M. Kikuchi, *Nature*, 2003, **423**, 965.
92. G. Matteucci, G.F. Missiroli, M. Muccini and G. Pozzi, *Ultramicroscopy*, 1992, **45**, 77.
93. G. Matteucci, G.F. Missiroli and G. Pozzi, *Physica B*, 1988, **151**, 223.
94. T. Kawasaki, G.F. Missiroli, G. Pozzi and A. Tonomura, *Optik*, 1993 **92**, 168.
95. J. Cumings, A. Zettl, M.R. McCartney and J.C.H. Spence, *Phys. Rev. Lett.*, 2002, **88**, 056804.
96. T. Kasama, Y. Antypas, R.K.K. Chong and R.E. Dunin-Borkowski, *Mater. Res. Soc. Proc.*, 2005, **839**, P3.1.
97. J.M. Titchmarsh, A.J. Lapworth and G.R. Booker, *Phys. Status Solidi*, 1969, **34**, K83.
98. S. Frabboni, G. Matteucci and G. Pozzi, *Ultramicroscopy*, 1987, **23**, 29.
99. G. Pozzi and M. Vanzi, *Optik*, 1982, **60**, 175.
100. M. Beleggia, D. Cristofori, P.G. Merli and G. Pozzi, *Micron.*, 2000 **31**, 231.
101. W.D. Rau, P. Schwander, F.H. Baumann, W. Höppner and A. Ourmazd, *Phys. Rev. Lett.*, 1999, **82**, 2614.
102. M.A. Gribelyuk, M.R. McCartney, J. Li, C.S. Murthy, P. Ronsheim, B. Doris, J.S. McMurray, S. Hegde and D.J. Smith, *Phys. Rev. Lett.*, 2002, **89**, 025502.
103. A.C. Twitchett, R.E. Dunin-Borkowski and P.A. Midgley, *Phys. Rev. Lett.*, 2002, **88**, 238302.
104. A.C. Twitchett, R.E. Dunin-Borkowski, R.J. Hallifax, R.F. Broom and P.A. Midgley, *J. Microsc.*, 2004, **214**, 287.
105. A.C. Twitchett, R.E. Dunin-Borkowski, R.J. Hallifax, R.F. Broom and P.A. Midgley, *Microsc. Microanal.*, 2005, **11**, 66.
106. A.C. Twitchett, R.E. Dunin-Borkowski and P.A. Midgley, *Philos. Mag. A*, in press.
107. S.M. Sze, *Physics of Semiconductor Devices*, New York, Wiley, 2002.
108. D. Cooper, A.C. Twitchett, P.K. Somodi, P.A. Midgley, R.E. Dunin-Borkowski, I. Farrer and D.A. Ritchie, *Appl. Phys. Lett.*, 2006, **88**, 063510.
109. M. Beleggia, G.C. Cardinali, P.F. Fazzini, P.G. Merli and G. Pozzi, *Inst. Phys. Conf. Ser.*, 2001, **169**, 427.
110. P.K. Somodi, R.E. Dunin-Borkowski, A.C. Twitchett, C.H.W. Barnes and P.A. Midgley, *Mater. Res. Soc. Proc.*, 2005, **839**, P3.2.
111. Z. Wang, T. Hirayama, T. Kato, K. Sasaki, H. Saka and N. Kato, *Appl. Phys. Lett.*, 2002, **80**, 246.
112. Z. Wang, K. Sasaki, N. Kato, K. Urata, T. Hirayama and H. Saka, *J. Electron Microsc.*, 2002, **50**, 479.

113. Z. Wang, T. Kato, N. Shibata, T. Hirayama, N. Kato, K. Sasaki and H. Saka, *Appl. Phys. Lett.*, 2002, **81**, 478.
114. Z. Wang, T. Kato, T. Hirayama, N. Kato, K. Sasaki and H. Saka, *Surf. Int. Anal.*, 2005, **37**, 221.
115. A. Lenk, H. Lichte and U. Muehle, *J. Electron Microsc.*, 2005, **54**, 351.
116. P. Formanek and E. Bugiel, *Ultramicroscopy*, 2006, **106**, 292.
117. S.M. Schwarz, B.W. Kempshall, L.A. Giannuzzi and M.R. McCartney, *Microsc. Microanal.*, 2003, **9**(Suppl. 2), 116.
118. M.R. McCartney, F.A. Ponce, J. Cai and D.P. Bour, *Appl. Phys. Lett.*, 2000, **76**, 3055.
119. M. Stevens, A. Bell, M.R. McCartney, F.A. Ponce, H. Marui and S. Tanaka, *Appl. Phys. Lett.*, 2004, **85**, 4651.
120. J. Frenkel, *Kinetic Theory of Liquids*, Oxford University Press, Oxford, 1946.
121. V. Ravikumar, R.P. Rodrigues and V.P. Dravid, *Phys. Rev. Lett.*, 1995, **75**, 4063.
122. Z. Mao, R.E. Dunin-Borkowski, C.B. Boothroyd and K.M. Knowles, *J. Am. Ceram. Soc.*, 1998, **81**, 2917.
123. M.A. Schofield, M. Beleggia, Y.M. Zhu, K. Guth and C. Jooss, *Phys. Rev. Lett.*, 2004, **92**, 195502.
124. M. Elfwing and E. Olsson, *J. Appl. Phys.*, 2002, **92**, 5272.
125. R.E. Dunin-Borkowski, W.M. Stobbs, D.D. Perovic and Z.R. Wasilewski, in *Electron Microscopy 1994*, ed. B. Jouffrey and C. Colliex, Les Editions de Physique, Les Ulis, 1994, **1**, 411.
126. H. Lichte, *Cryst. Res. Technol.*, 2000, **35**, 887.
127. H. Lichte, M. Reibold, K. Brand and M. Lehmann, *Ultramicroscopy*, 2003, **93**, 199.
128. H. Lichte, *Ultramicroscopy*, 1991, **38**, 13.
129. H. Lichte, *Ultramicroscopy*, 1992, **47**, 223.
130. H. Lichte and W.D. Rau, *Ultramicroscopy*, 1994, **54**, 310.
131. K. Ishizuka, T. Tanji, A. Tonomura, T. Ohno and Y. Murayama, *Ultramicroscopy*, 1994, **55**, 197.
132. A. Orchowski, W.D. Rau and H. Lichte, *Phys. Rev. Lett.*, 1995, **74**, 399.
133. M. Lehmann, H. Lichte, D. Geiger, G. Lang and E. Schweda, *Mater. Charac.*, 1999, **42**, 249.
134. M. Lehmann and H. Lichte, *Cryst. Res. Technol.*, 2005, **40**, 149.
135. J.M. Cowley, *Ultramicroscopy*, 1992, **41**, 335.
136. R.A. Herring, G. Pozzi, T. Tanji and A. Tonomura, *Ultramicroscopy*, 1995, **60**, 153.
137. H. Banzhof and K.H. Herrmann, *Ultramicroscopy*, 1993, **48**, 475.
138. N.H. Dekkers and H. de Lang, *Optik*, 1974, **41**, 452.
139. M. Mankos, M.R. Scheinfein and J.M. Cowley, *J. Appl. Phys.*, 1994 **75**, 7418.
140. M.R. McCartney, P. Kruit, A.H. Buist and M.R. Scheinfein, *Ultramicroscopy*, 1996, **65**, 179.

141. Q. Ru, *J. Appl. Phys.*, 1995, **77**, 1421.
142. Q. Ru, J. Endo, T. Tanji and A. Tonomura, *Appl. Phys. Lett.*, 1992 **59**, 2372.
143. T. Kawasaki, J. Endo, T. Matsuda, N. Osakabe and A. Tonomura, *J. Electron Microsc.*, 1986, **35**, 211.
144. J. Chen, T. Hirayama, T. Tanji, K. Ishizuka and A. Tonomura, *Opt. Commun.*, 1994, **110**, 33.
145. T. Matsumoto, T. Tanji and A. Tonomura, *Ultramicroscopy*, 1994 **54**, 317.
146. J.M. Cowley, *Ultramicroscopy*, 1990, **34**, 293.
147. M. Mankos, J.M. Cowley and M.R. Scheinfein, *Mater. Res. Soc. Bull.*, 1995, **20**, 45.
148. G. Matteucci and M. Muccini, *Ultramicroscopy*, 1994, **53**, 19.
149. G. Lai, T. Hirayama, A. Fukuhara, K. Ishizuka, T. Tanji and A. Tonomura, *J. Appl. Phys.*, 1994, **75**, 4593.

CHAPTER 6

Electron Tomography

M. WEYLAND[1] AND P. A. MIDGLEY[2]

[1] *Department of Materials Science and Metallurgy, University of Cambridge, Pembroke St, Cambridge, CB2 3QZ, UK*
[2] *Monash Centre for Electron Microscopy, Monash University, Clayton, VIC 3800, Australia*

6.1 Introduction

Transforming promising nanoscience into practical nanotechnology is fraught with a number of major challenges, the foremost among these being, perhaps, the assembly and characterisation of functional nanostructures. While significant progress has been made in the growth of the building blocks of nanoscience, the act of assembly into functional devices is far less well developed. In addition, the length scales and structural complexity involved in nanostructures are such that traditional methods for materials characterisation are of limited use. One area that is particularly challenging is to visualise such complex structures in three dimensions. While a number of techniques, such as Atomic Force Microscopy (AFM)[1] or Scanning Electron Microscopy (SEM),[2] can achieve high spatial resolution in two dimensions, they are primarily surface analysis techniques. They have insufficient depth sensitivity to determine complex 3D structure, especially if the structure is buried well beneath the surface. Conventional Transmission Electron Microscopy (TEM)[3] and Scanning Transmission Electron Microscopy (STEM) are immediately advantageous in this respect: by their very nature they generate a transmitted image of the entire structure, including buried or surface details. While the overall specimen thickness is limited to a few hundred nanometres, this is precisely the length scale upon which true ''nanostructures'' exist. Electron microscopy techniques can also operate at spatial resolution up to, and even beyond, 0.1 nm. For some nanostructures such a transmitted image, or ''projection'', can be sufficient to characterise the structure of the specimen. There is,

however, an increasing trend, especially with complex nanostructured devices, for systems that show multiple overlapping structures within many TEM specimens. In such cases a projection can offer only a limited, and often misleading, assessment of the true 3D structure.

Interestingly, a similar problem on almost exactly the same length scales, exists when attempting to elucidate the structure and function of biological macromolecules. An example of the convergence between length scales in the physical and biological sciences is given in Figure 6.1.

Due to the projection problem conventional TEM data is insufficient to unravel the complex overlapping, low-contrast, protein structures that make up the bulk of cellular material. An elegant solution to this is electron tomography. The principle of which is identical to that of medical tomography, of which the X-ray Computer-Aided/Axial Tomography (CAT) is the most widely known example.[4] Indeed the development of the two techniques is contemporary. Cormack first suggested the radiological technique in 1963[5] and a patent was filed on the first CAT scanner by Hounsfield in 1972[6] while three papers published in 1968 covered the theory and first applications of electron tomography. The first experimental examples of electron tomography were demonstrated by de Rosier and Klug,[7] who made use of the helical symmetry of a bacteriophage tail to achieve a tilt series in a single micrograph, and by Hart,[8] who used a limited tilt series to improve the contrast from a single projection. However it was Hoppe who suggested that any asymmetrical biological structure would be amenable to being reconstructed from a series of electron micrographs taken at different angles.[9]

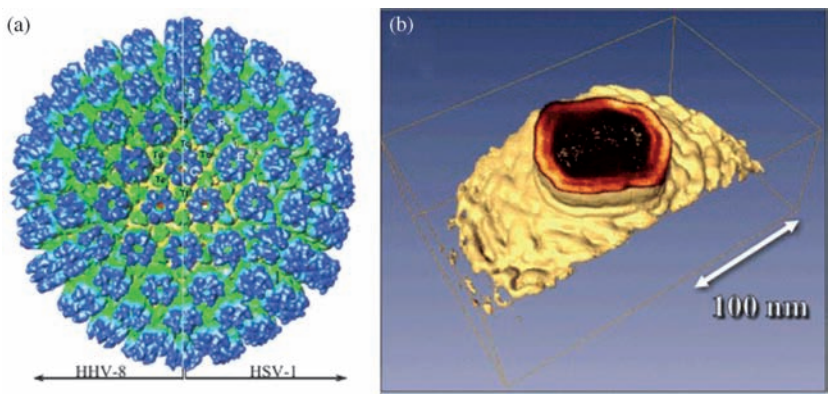


Figure 6.1 A comparison of the convergent length scales between biological and physical science structures (a) The herpes simplex virus (HSV), right, and the human herpes virus 8 (HHV-8), left, capsids as reconstructed by single-particle cryoelectron tomography.[190] The diameter of the capsid is ~130 nm. (b) A double-lined MOSFET transistor, the via width is ~150 nm as reconstructed by HAADF STEM tomography, showing the surface roughness of the tantalum liner (gold) and the interior double liner (red). Peter Ercius is acknowledged for providing (b).

While the conceptual development of electron tomography was rapid and much of the tomographic theory covered in the following sections has advanced very little in the last 30 years, the experimental implementation of the technique was stalled for important practical reasons, namely the susceptibility of biological sections to radiation damage under the electron beam[10] and the computationally intensive nature of tomographic reconstruction. The availability of inexpensive, powerful computers has not only speeded up reconstruction but it has also led to a comprehensive automation of electron microscopes that allow beam damage to be minimised, controlled, and quantified during tilt-series acquisition. While beam damage is still the most significant limitation to biological tomographic resolution, automation, combined with other experimental and theoretical advances, has allowed three-dimensional spatial resolution on the order of 2 nm to be achieved,[11] well beyond the pessimistic predictions of early studies.[10]

The first physical science application of electron tomography, investigating the structure of block copolymer systems, was demonstrated by Spontak *et al.*[12] in the late 1980s. These materials consist of polymers formed from several distinct monomer units with mutual immiscibilities and structural incompatibilities. Driven by these differences, and the exact details of the monomer mix, they self-assemble into a range of complex nanostructures from lamellae to 3D networks. These specimens exhibit complex ordering that requires detailed analysis by both bulk techniques, such as X-ray crystallography,[13] and local techniques such as TEM and electron tomography.[14] The complicated overlapping nanoscale features in many of these systems make them ideal candidates for electron tomography. Further application of electron tomography in the physical sciences was stimulated by a paper of Koster *et al.*[15] published in 2000 that demonstrated the reconstruction of porous zeolites. In the following five years, electron tomography has been applied to nearly all classes of nanostructured materials; from polymers,[16] through metals,[17] ceramics[18] and semiconductors.[19] In addition, alternative imaging modes for electron tomography, which are arguably more suited to materials than conventional Bright-Field (BF) TEM, have also been developed; including High-Angle Annular Dark-Field (HAADF) STEM[20] and Energy-Filtered TEM (EFTEM).[21]

While the basic methodology of electron tomography is much the same across all fields and all specimens the goals, and types of information that are of interest, in the physical and biological sciences are quite different. As such, existing reviews of the technique, while valuable, lack direct relevance to materials science. This text aims to bridge that gap by covering the theory and application of electron tomography with a particular emphasis on its application to nonbiological nanoscale systems. Sections 6.2–6.5 cover the theory and general practical application of tomography. Sections 6.6 and 6.7 are aimed at illustrating the applicability of the range of different imaging techniques to a number of nanoscale materials problems, reviewing the current state of the art in the physical sciences and suggesting some future directions for this new field of nanocharacterisation.

6.2 Theory of Electron Tomography

6.2.1 From Projections to Reconstructions

While tomography, both by electrons and X-rays, has its roots in the 1960s, the German mathematician Johan Radon first described the mathematics of projection in a landmark 1917 paper,[22] an English translation of which is available in the exhaustive review by Deans.[23] A projection of an object, $f(x,y)$ in real space, D, can be described by a Radon transform R as the sampling of the object by a line integral through all possible lines L:

$$Rf = \int_L f(x,y)\,\mathrm{d}s \tag{6.1}$$

where ds is the unit length of L. The geometry of this transform is shown in Figure 6.2. Sampling of an object by some form of projection is therefore equivalent to a selected sampling of the Radon integral. It follows that an estimation of the density of an object can be achieved through the inverse Radon transform of a series of set of projections, Rf, of that object. The application of the mathematics of the inverse radon transform therefore allows the reconstruction of higher-dimensionality information from lower-dimensionality data: 2D slices from a set of 1D lines, or 3D volumes from a set of 2D images.

To understand fully the reconstruction process requires a closer examination of the Radon transform and its consequences in more detail. Although the

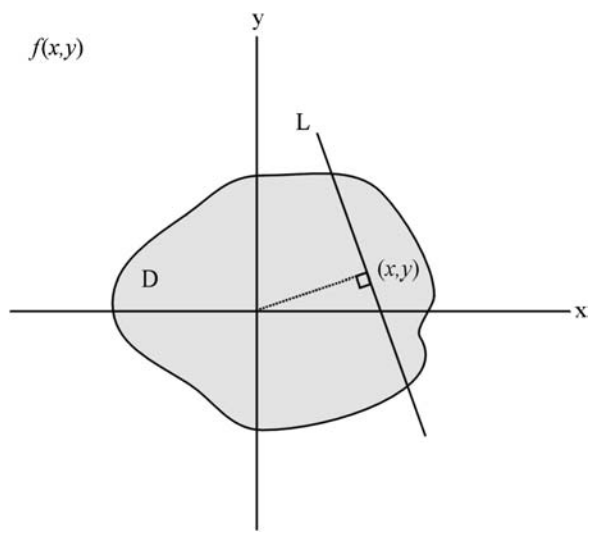


Figure 6.2 The Radon transform, R can be visualised as the integration through a body D in real space, $f(x,y)$ along all possible line integrals L, with its normal at an angle θ to the horizontal.

description presented here is mathematically equivalent to Radon's original, the more modern approaches of Deans[23] and Herman[24] are followed.

For clarity it is mathematically convenient to use polar, (r,ϕ) coordinates, related to Cartesian axes by: $r = \sqrt{x^2 + y^2}$ and $\phi = \tan^{-1}(y/x)$. For an explicit definition of the Radon transform it is also necessary to define a set of coordinates that are perpendicular, (l) and parallel, (z) to the transform direction defined by the ray integral, L. The angle θ is that from the horizontal to the line normal of L, usually termed the "projection angle". The geometry of these terms is illustrated in Figure 6.3(a). This description allows a more explicit statement of the radon transform in terms of polar coordinates as:

$$Rf(l,\theta) = \int_{-\infty}^{\infty} f\left(\sqrt{l^2 + z^2}, \theta + \tan^{-1}(z/l)\right) \mathrm{d}z \quad \text{if } l \neq 0$$

$$Rf(0,\theta) = \int_{-\infty}^{\infty} f(z, \theta + \pi/2) \tag{6.2}$$

The Radon transform operator R converts the coordinates of the data into "Radon space", (l,θ), where l is the line perpendicular to the projection direction and θ is the angle of the projection (Figure 6.3(b)). A point in real space is a line in Radon space linked through $l = r\cos(\theta - \phi)$, such as the one marked with a circle in Figure 6.3(b). A model example of this transformation is shown for a 2D test object in Figure 6.4. In the Radon-transformed object, (Figure 6.4) the bright spots have been converted into sine curves and as such this kind of Radon space image is often termed a "sinogram".

The relationship between real space and Radon space allows us to examine the experimental situation more explicitly. A single projection of the object, a discrete sampling of the Radon transform, is a line at constant θ in Radon space. A more complete sampling of Radon space is therefore achieved by recording projections over a range of angles. With a sufficient number of

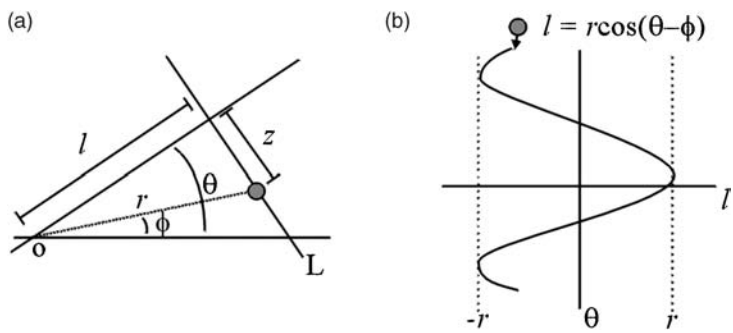


Figure 6.3 The relationship between polar, (r,ϕ) space and Radon (l,θ) space, (a) Geometry of the Radon transform in real space where L is the direction of the projection. (b) The point in real space marked with a circle becomes the line $l = r\cos(\theta - \phi)$ in Radon space.

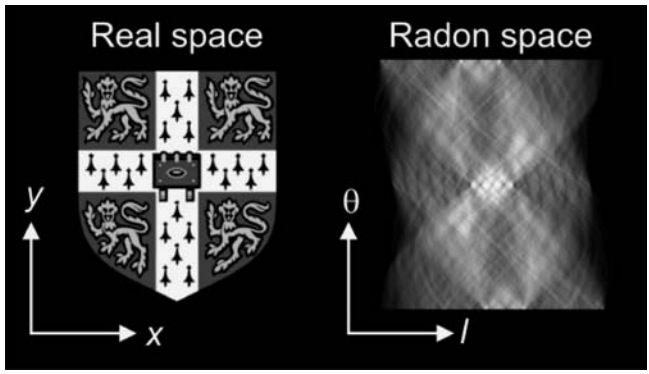


Figure 6.4 Visual representation of the transform between real space and Radon space. Here, the same image is viewed is real space and then transformed into radon space by projection at a range of angles θ. Here l is equivalent in direction to x.

projections spread across the entire range of θ, a reconstruction of an object can be achieved by simply applying the inverse Radon transform. However, as any sampling of Radon space (l,θ) is a discrete sampling of a continuous function such an inversion will be imperfect; attaining the most accurate reconstruction from discrete experimental data is hence the central challenge of tomographic reconstruction.

The relationship between real space and Radon space also allows an understanding of the nature of projection and its relationship with the structure of the original object. However, the theory of reconstruction, and especially the consequences of incomplete sampling, are aided by a description of the relationship between a projection in real space and a projection in Fourier space; the ''projection slice theorem''. This states that a projection at a given angle of an object is a central section, at that angle, through the Fourier transform of the 3D object. Although this relationship has long been known in the field of pure mathematics,[25] its implications for practical experimentation were first explored by Bracewell.[26] His 1956 study, reconstructing a 2D map of solar microwave emission from a series of 1D ''fan beam'' profiles measured by a radio telescope, was the first practical application of what would become to be known as tomography.

The relationship between the Fourier transform, F and the Radon transform, R, the heart of the projection slice theorem, is summarised in Figure 6.5. In operator form, this diagram is:

$$F_2 f = F_1 R f = F_1 \hat{f} \tag{6.3}$$

where $\hat{f}$ is the full Radon transform of object f. An intermediate interpolation step is required to convert from polar to Cartesian coordinates in Fourier space. This is the case for a complete sampling of the Radon transform, given

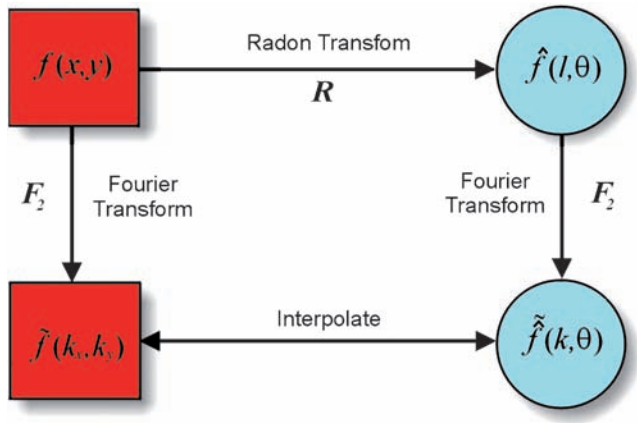


Figure 6.5 The relationship between the Radon and Fourier transforms, illustrating the projection-slice theorem in two dimensions. Interpolation is required to convert between polar (blue circles) and Cartesian (red squares) coordinates in Fourier space.[23]

all possible projections. A full exploration of the relationship between the Radon and the Fourier transforms is covered by Deans.[23]

This relationship allows an intuitive description of tomographic reconstruction. As a tilt series proceeds each acquired projection samples the object through a central section of Fourier space at the angle being projected. While each section contains only information from a single direction the whole tilt series combined will sample a broad range of Fourier space. By a superposition of the acquired projections in Fourier space and applying an inverse Fourier transform the 3D structure of the object may be reconstructed. This approach, known as ''direct Fourier reconstruction'', was used by Bracewell in radio astronomy and for the first tomographic reconstruction from electron micrographs.[7] This technique, also known as ''Fourier synthesis'', can also be used for atomic structure determination by X-ray crystallography[27] and by electron diffraction.[28]

The projection slice theorem can also explain some of the problems that arise in the experimental implementation of tomography. Even though all frequencies may be sampled in each image (and hence Fourier slice), a finite number of projections around a single tilt axis will lead to a skewed sampling of frequencies in the 3D Fourier space of the object and lead to a number of the reconstruction artefacts described later. Another limitation, especially in tomography performed in the TEM, is that projections cannot be acquired through a full tilt range of the specimen because of a finite pole piece gap or the geometry of the sample. It therefore follows that Fourier space will be undersampled in certain directions and hence the reconstruction will be of lower resolution in those directions.

While the Fourier theory of tomographic sampling and reconstruction acts as an extremely useful framework for conceptualising tomography, in practice,

Fourier methods have largely been superseded by faster, easier to implement, real-space reconstruction approaches based on backprojection.

6.2.2 Backprojection: Real-Space Reconstruction

Backprojection operates on a simple principle: any point object in three-dimensional space may be described by the intersection of two noncolinear "rays". A more complex object will require more "rays" to describe it completely. The projection operator can be thought of as summation through an object in a single direction, the opposite of such a "ray". Following this reasoning, inverting the projection, propagating the projection through a volume, generates a "ray" (more accurately a set of rays for 1D or 2D projections) that will describe an object along the direction of that ray. Combining a sufficient number of such rays from different directions will restore the actual measured density of the object. This real-space technique is known as direct backprojection.[29] Figure 6.6 is graphical description of the process. Several groups have independently formulated reconstruction by backprojection, including those in the CAT[24] and electron microscopy communities.[8,30] The first detailed description of its application using electron microscopy, was by Vainshtein.[31]

The backprojection operator is implemented by carrying out a discrete estimation, for example by the use of a Riemann sum, of the infinite integral of the inverse Radon transform, eqn (6.2). As Radon space is almost always undersampled successful reconstruction requires interpolation between missing angles in Radon space. Much like the interpolation between slices in a Fourier reconstruction, the quality of reconstruction is heavily influenced by the kind of interpolation required.

In practice, successful tomographic reconstruction by backprojection requires a further correction step due to the nature of the sampled tilt series.[31] Following the projection slice theorem each projection corresponds to an

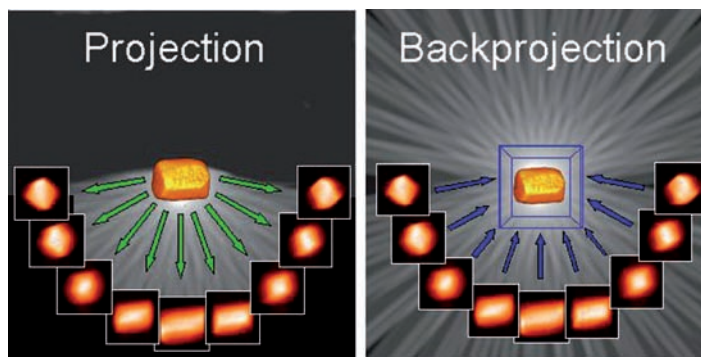


Figure 6.6 A schematic of reconstruction via backprojection. An object is sampled by projection from a range of angles and then reconstructed by summing these projections at the original sampling angles into the object space.

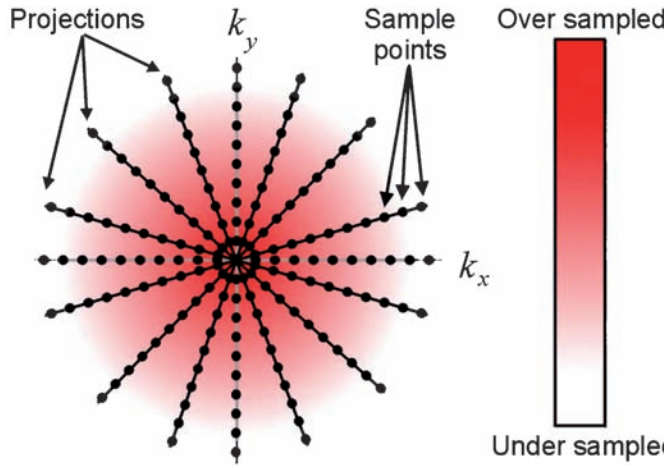


Figure 6.7 An illustration, in Fourier space, of the sampling density problem encountered in direct backprojection reconstruction. The larger number of sample points at low frequencies results in a heavily blurred reconstruction.

equally sampled line/plane in Fourier space. However, this is only a central section and thus will lead to an uneven sampling of the object space as a whole (Figure 6.7). High frequencies are significantly undersampled compared to low frequencies. This weights the reconstruction to enhance the low-frequency information and results in a blurred reconstruction. One solution is to reweight the data in Fourier space based on the sampling geometry; a technique known as weighted backprojection. The weighting filter can be generated in two ways, by either determining the average radial sampling and producing a simple ramp function[29] (zero at zero frequency, maximum towards the edges) or by producing an exact filter based on the full angular sampling.[29] While the former is the standard approach the latter is particularly useful with irregular angular sampling. Weighted backprojection is the standard reconstruction technique used for electron tomography; it is fast, robust, easy to implement and retains the relationship between sampling and resolution. In the limit, it gives reconstructions that should, in theory, be equivalent to those generated by Fourier techniques.

6.2.3 Constrained Reconstructions

In practice, the number of projections that can be sampled from an object is always limited. This is usually well below that needed for sufficient sampling at the Nyquist frequency to give a reconstruction with the same resolution as individual projections. In addition, limited angular sampling, the usual case in electron tomography, will lead to reconstruction "fan" artefacts that can be characterised by a point spread function (PSF) of the reconstruction.[32] Further

imperfections will be caused by the type of interpolation in the reconstruction and projection, the use of a nonoptimal weighting filter and by deficiencies in the signal-to-noise (SNR) ratio of the original data. Conventional reconstruction already makes assumptions regarding the mathematical nature of the projection but other assumptions can be applied as constraints to improve the quality of reconstructions. These include refining the reconstruction to match the original projections,[33] taking advantage of symmetry[34] and achieving a global ''most likely'' reconstruction using statistical techniques such as maximum entropy.[35]

One symptom of the problems of conventional reconstruction is that a reprojection of a reconstruction in the direction of an original image (projection) will not be identical to that original projection. However as each original image represents a ''perfect'' reference projection of the object, a refinement can be carried out by constraining the reconstruction to match each projection. This can be carried out by projecting the reconstruction, comparing to the original projections, backprojecting the differences into the reconstruction volume, and finally modifying the reconstruction by the ''difference'' reconstruction. As the difference also undergoes an imperfect transform during the backprojection it is necessary to iterate the process. Techniques that attempt to restore the correct reconstruction density based on the projections are collectively known as iterative reconstruction techniques. The simplest form of iterative reconstruction is the Algebraic Reconstruction Technique (ART).[36] It operates by comparing the reconstruction with a single projection, correcting in a single direction and then moving on to the next projection. The substantial resolution improvements initially offered by ART[36] soon proved to be spurious due to flawed methodology[37] and, when correctly formulated, the iteration process quickly diverges in the presence of even small levels of noise. However, a subtly different version of the technique that compares all the projections simultaneously rather than in isolation, the Simultaneous Iterative Reconstruction Technique (SIRT), is stable with real data.[33] Whilst, in theory, SIRT does not offer any resolution improvements over weighted backprojection, it can offer a significantly ''cleaner'' reconstruction that is easier to interpret.

The implementation of iterative reconstruction has been summarised in Figure 6.8 for an iterative reconstruction cycle of N iterations. The most important practical variables for such a loop are the number of iterations used, (N) and whether an additive or multiplicative system is applied. The number of iterations is usually chosen by a subjective principle and the improvement in the reconstruction tends to be minimised after 20–30 iterations. Multiplicative approaches calculate a difference ratio between the projections and reprojections and multiply that difference in the backprojection. Additive approaches calculate a difference by subtraction and add that difference, usually modified by a normalisation parameter, in the backprojection. The choice of whether multiplicative or additive techniques are used is covered briefly by Gilbert,[33] who finds there is very little difference between approaches for SIRT and only small differences for ART. A demonstration of the

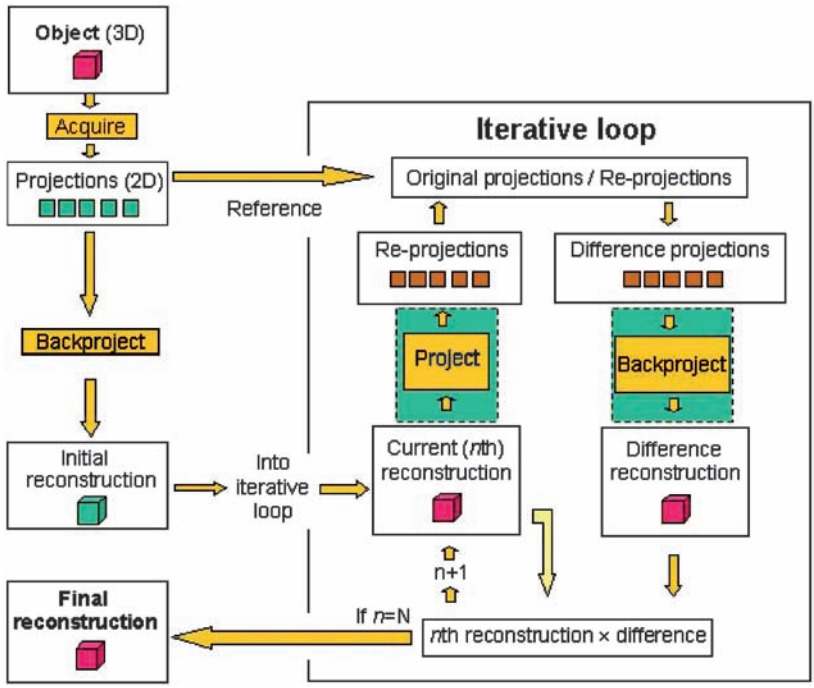


Figure 6.8 Diagram showing how iterative reconstruction is implemented for N iterations where n is the current iteration (starting with $n=0$). This schematic is for both multiplicative and additive reconstruction.

reconstruction improvement offered by iterative reconstruction, especially for datasets with limited sampling, shown in Figure 6.9.

Whilst iterative routines had been mostly ignored in single-tilt tomography, perhaps due to the initial controversy over the results generated,[37,38] there has recently been an renewed interest in iterative reconstruction in both biological[39] and materials electron tomography.[20] In the latter case, it appears that an assumption of positivity, implicit in iterative reconstructions, can offer significant gains in reconstruction fidelity when combined with high SNR data. These kinds of iterative techniques for structural refinement, such as ART and SIRT, can be described as solutions by "Projection Onto Convex Sets" (POCS).[40]

It is often the case that macromolecular assemblies show well defined structural symmetry. This suggests some degree of informational redundancy in Fourier space and as such a smaller number of projections are needed to describe the structure, a benefit that can be exploited both in acquisition (limited dose) and reconstruction. The first demonstration of electron tomographic reconstruction, of the tail of a T7 bacteriophage,[7] was carried out from a single projection by making use of the helical symmetry of the structure. Such structural relationships are used extensively in conventional tomography[41] and

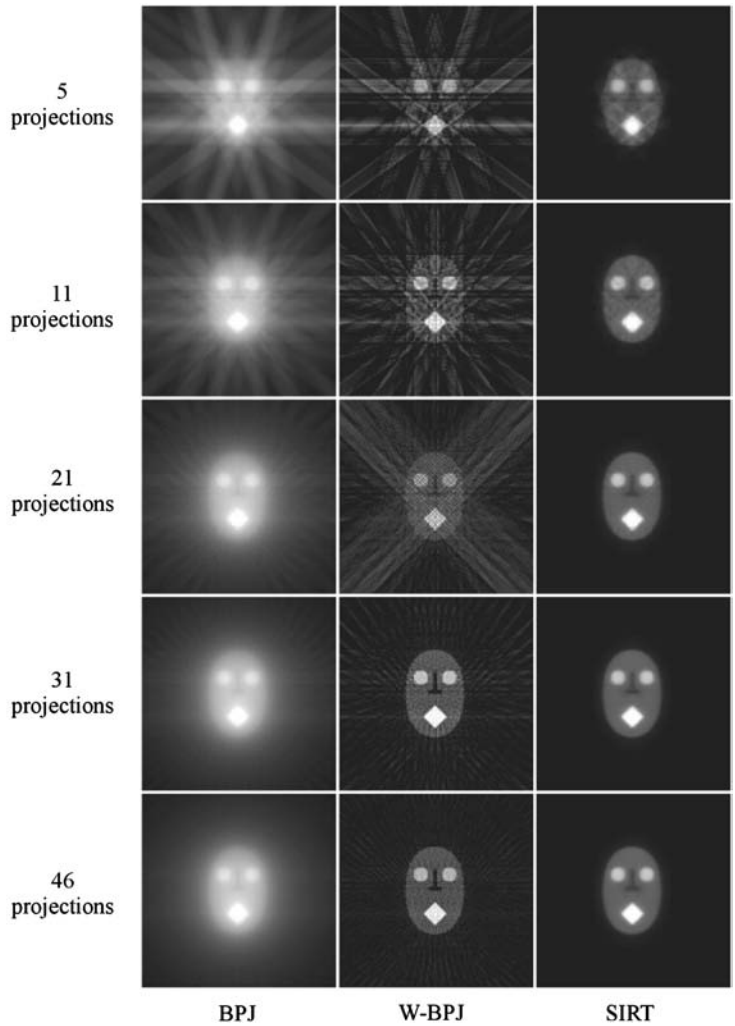


Figure 6.9 Illustration of the improvements offered by iterative reconstruction (SIRT) over weighted backprojection (W-BPJ) especially for small numbers of initial projections.

single-particle reconstruction, where both the preferential orientation of particles[34] and underlying symmetries of subunits determined by X-ray techniques[11] can lead to improved reconstructions. Whilst perfect structural symmetry on the nanoscale is rare in the physical sciences, there are occasions where such assumptions may be of value. One case in particular is the growth of self-assembled nanostructures where local symmetry may be present and aid the reconstruction of such complex systems.

Although optimisation by iterative correction and/or symmetry constraints can improve reconstruction quality, if used without sufficient rigour there is the

potential to introduce serious artefacts. A statistical approach, based on Bayesian methods, should be more robust as it offers the ''least constrained'' or ''most open'' solution given initial data.[42] In addition, a Bayesian solution can take advantage of additional knowledge about the experiment, such as the Contrast Transfer Function (CTF) of the imaging lens, the Point Spread Function (PSF) of the reconstruction, or the symmetry of the object itself, as mentioned above. This approach has been applied with varying success to electron microscopy in both 2D[43] and 3D.[44] Whilst several applications of the maximum entropy technique have been demonstrated in electron tomography, most recently the Constrained Maximum Entropy Tomography (COMET)[35] approach, they have yet to see broad application despite their promise. This may be due to a combination of their innate complexity and a scepticism of any such ''optimised'' reconstruction.

6.2.4 Reconstruction Resolution

The resolution of a single projection in a tilt series is, of course, simply that of the image formed by the microscope, characterised by the highest statistically significant frequency transferred in Fourier space. However, for tomographic reconstructions, the increased number of sampling points in the 3D volume means that achieving statistical significance at a particular frequency is correspondingly more difficult.[45] To achieve a high-resolution reconstruction requires the acquisition of as many projections as possible, over as wide a tilt range as possible. The relationship between the number of projections, (N) and the resolution, (d) attainable can be defined following Crowther,[30] as:

$$d = \frac{\pi D}{N} \tag{6.4}$$

where D is the diameter of the reconstruction volume. An implicit assumption in this equation is that the N projections are spread evenly though 180°. In electron tomography there is almost always an upper limit to the tilt angle, leading to anisotropic resolution brought about by this ''wedge'' of missing information. Most noticeably this leads to a loss of resolution in the least sampled direction, manifest as an ''elongation'' of objects in that direction (usually the optic axis). An estimation of this elongation, (e), as a function of the maximum tilt angle (α), can be made[46] as:

$$e = \sqrt{\frac{\alpha + \sin\alpha\cos\alpha}{\alpha - \sin\alpha\cos\alpha}} \tag{6.5}$$

For a maximum tilt angle of 70°, not uncommon in electron tomography, an elongation factor of 30% arises. Another assumption of Crowther's equation is that the volume to be reconstructed is equiaxed. However, TEM samples are often extended thin slabs with far larger dimension inplane, (x,y) than in depth, (z). This can be modeled by a modified diameter, (D) that is a factor of the

thickness of the slab, (T) and the maximum tilt angle, (α):[47]

$$D = T\cos\alpha \tag{6.6}$$

As a consequence of a reduced reconstruction volume, fewer projections are required to attain the same overall resolution. Taking account of eqn (6.6) tends to produce a less pessimistic resolution prediction than the basic Crowther equation.

Although an overall resolution prediction based on Fourier space sampling is a useful guide, it ceases to become accurate for constrained reconstruction techniques,[32] such as iterative POCS (ART, SIRT), symmetry guided reconstruction and Bayesian methods. For these, the reconstruction resolution is strongly governed by the noise characteristics of the original data, the shape of the object to be reconstructed and the nature of the constraints applied. As such, it is impossible to define a set resolution for constrained reconstruction as it will vary between different datasets, imaging techniques and samples.

6.2.5 Measuring Reconstruction Resolution

A prediction of the tomographic resolution as described above can be extremely valuable, especially when determining experimental conditions needed to resolve particular features. However, a measurement of the achieved reconstruction resolution is arguably more important. This will become increasingly important in the physical sciences where tomographic reconstructions will be used increasingly to generate quantitative metrological results. For quantitative results to have meaning will require a statistical estimation of random errors, the overall resolution limit and an account of anisotropy in the reconstruction.

The classical Rayleigh criterion, the ability to resolve the separation of two features, is often used as a basis to determine resolution in 2D imaging.[48] However, in tomography the large amount of 3D data, the ability to "slice" data in an arbitrary direction and the anisotropy of resolution all make such an approach highly subjective. If a specimen contains a hierarchy of features of a known size this can be used to estimate the reconstruction resolution, as demonstrated elegantly for axonems by McEwen *et al.*[49] However, in most cases the object will not contain such a convenient set of structures and assessment is often limited to a subjective identification of the smallest significant feature.

A more quantitative, and less subjective, approach to resolution measurement is possible by examining the intensity distribution of the object in Fourier space, and determining where this is above the threshold for noise. This is the approach used in Fourier Shell Correlation (FSC),[50] the Differential Phase Residual (DPR)[51] or Spectral Signal-to-Noise Ratio (SSNR) methods.[52]

These methods were developed originally for determining the resolution of single-particle reconstruction.[53] The number of projections involved in such techniques is very large, hundreds of thousands in some cases, as opposed to tomography where there are rarely more than 150 projections. The small

number of images from tomography leads to comparatively poor sampling and often low Signal-to-Noise Ratios (SNR) of resultant reconstructions. As these resolution-assessment techniques rely on the analysis of the sampling and noise statistics of the reconstructed volume, it is difficult to get these methods to give accurate, or more importantly reliable, results from conventional tomographic datasets. Some success has been reported[54] using SSNR for tomographic datasets but neither approach has been used widely on such data.

In order to use the FSC method, the data must be separated into two sets on the basis of odd and even tilt angles. The FSC is then calculated as:

$$\mathrm{FSC}(k) = \frac{\sum F_1(k)F_2^*(k)}{\sqrt{\left(\sum |F_1(k)|^2 |F_2^*(k)|^2\right)}} \tag{6.7}$$

where $F_n(k)$ is the sum is over all the images in series 1 and 2 and within a shell of small width at frequency k. Using this definition, by convention, the resolution of the dataset is determined at the 0.5 value of the FSC curve.[55] This is equivalent to a SNR of 1.

Care must be taken to avoid artificial similarities, such as sharp masks or discontinuities at the edges of reconstruction volumes. Interpolation can also lead to rounding errors. These can lead to correlated high-frequency components and artificially high FSC values. It has been argued[54] that the FSC (and DPR) measures are unsuitable for tomography because the two reconstructions from a single tilt series contain too few projections (generally 35–80, although in some cases slightly fewer). A more significant problem for the technique arises from the missing wedge. It is known that the resolution in the worst reconstruction direction is lower than in the other two directions, yet the FSC method yields a single value for the resolution. Despite these drawbacks, the FSC represents, at worst, an analytical method for comparing reconstructed objects and hence a tool to optimise the reconstruction techniques. At best, the FSC permits the quantitative determination of the resolution of a reconstruction.

A tilt series for a heterogeneous catalyst, porous silica with Pt/Ru particles (Section 6.5), was used for the comparing SIRT and weighted backprojection methods. It was split into two series of 77 images by picking alternate projections. Figure 6.10 shows how the resolution of the object was the same, regardless of the reconstruction algorithm used, because both traces cross the cut-off FSC = 0.5 at approximately the same resolution. This value is probably determined by the missing wedge. Below the ultimate resolution, however, the SIRT trace is consistently above that for weighted backprojection, which corresponds to a higher SNR and indicates a better reconstruction quality over these high frequencies.

6.2.6 The Projection Requirement

The mathematics of tomographic reconstruction is based on the premise that the intensity in the projection is a monotonic function of the physical quantity

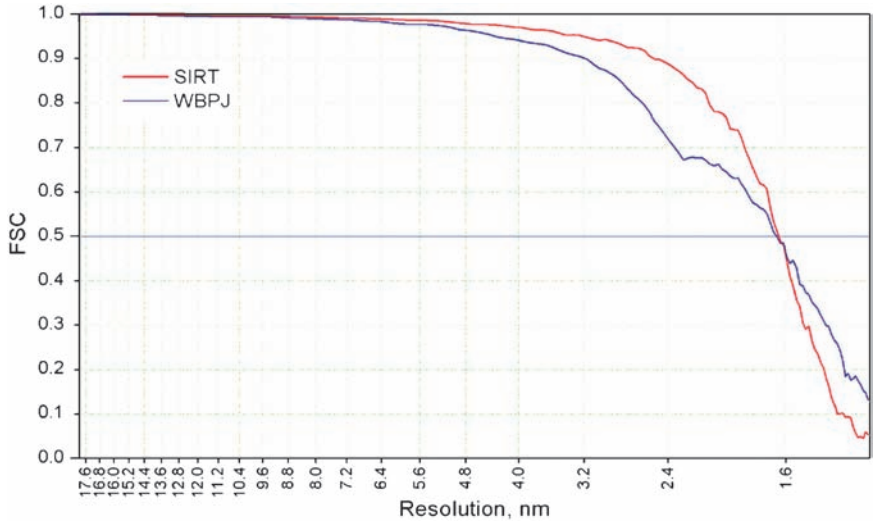


Figure 6.10 Fourier shell correlation for a dataset recorded from mesoporous silica, shown in Figure 6.25, reconstructed using Weighted Backprojection (WBPJ) and SIRT with 30 iterations as a function of resolution. The higher FSC at all meaningful frequencies indicates an improved fidelity for the SIRT reconstruction.

to be reconstructed.[56] This is known as the ''projection requirement''. In conventional X-ray tomography, image contrast is generated by changes in the X-ray absorption coefficient and the transmitted X-ray signal varies monotonically with thickness. In electron microscopy there are a number of competing contrast mechanisms that could be used, only some of which in general fulfil the projection requirement. For amorphous specimens, such as many biological and polymeric specimens,[12] conventional BF TEM contrast arises due to changes in mass-thickness (much like the X-ray case) or, with an applied defocus, due to the phase change induced by elastic scattering. For very weakly scattering, thin specimens phase-contrast images can be used for tomography. Ideally, the CTF of the objective lens should be corrected in the tilt images before reconstruction, as is usually the case in single-particle reconstruction,[57] however, typically the SNR in BF tilt series are too low to allow accurate CTF determination. There are several new approaches to CTF determination that suggest this correction, and a consequent improvement in reconstruction contrast and reliability, may become routine.[58,59]

However, crystalline materials are more problematic because coherent diffraction contrast violates the projection requirement. The complex dynamical interactions of the electron beam with the crystal potential ensures that there is no simple relationship between the BF intensity and the physical properties of the specimen. Nevertheless, BF TEM has been used for reconstructions of crystalline materials,[19,60,61] but results often show serious artefacts and poor reconstruction fidelity of internal structure.[62]

Table 6.1 A summary of the broad range of imaging techniques that have been, or could possibly be, used for electron tomography of materials.

Imaging technique	*Contrast mechanism*	*Suitable specimens/structures/ materials*
Bright-field TEM (STEM)	Phase, amplitude contrast	Biological specimens: cellular structure, macromolecular assemblies. Amorphous materials
Dark-field TEM	Angular-resolved scattering	Lattice defects (dislocations, precipitates) and nanocrystals
High angle annular dark field (HAADF) STEM	Atomic number (Z) contrast	Crystalline materials, Amorphous materials, high Z-contrast specimens, stained/ immunolabeling of biological sections
Core-loss energy-filtered TEM (EFTEM)	Inelastic scattering: core loss (chemistry) or low loss (bonding)	Chemically segregated specimens (bio or physical), Specimens with strong bond variations, beam-insensitive specimens
Low-loss energy-filtered TEM (EFTEM)	Differing plasmon frequencies	Large differences in bonding, structurally varying yet chemically similar specimens
Electron energy-loss spectroscopy (EELS) mapping	Inelastic scattering: core loss (chemistry) or low loss (bonding/ optical)	Chemically segregated specimens (biological or physical), Specimens with strong bond variations, possibly optical properties
Energy dispersive X-ray (EDX) mapping	Secondary X-ray generation	Chemically segregated, beam-insensitive, specimens
Electron holography	Reconstructed phase and/or amplitude contrast	Visualisation of mean-inner potential, electrostatic and magnetic fields. Doped semiconductors, nanomagnetic materials

For crystalline samples there are a growing number of alternative imaging and contrast modes, utilising both TEM and STEM, which satisfy the projection requirement. A summary of TEM/STEM imaging modes is presented in Table 6.1, including a description of the projected function and the types of material each has been, or can be, used for. As each technique has different experimental considerations and applications, Section 6.6 will cover each of them in more detail.

6.3 Acquiring Tilt Series

Sampling theory concludes that the best tomographic resolution is achieved by the acquisition of as many projections as possible, over as wide a tilt range as

possible. The two principal factors limiting this are instrumental design and the specimen itself. The former is a particular challenge in the physical sciences, where the goals of tomography are at odds to conventional microscope specification, and the latter the most challenging in biology, where beam damage is the central limiting factor.

6.3.1 Instrumental Considerations

6.3.1.1 Objective Lens Polepiece and Holder Design

The design of TEMs shows subtle differences for users in biology or the physical sciences due to the differing goals of each community. In the biosciences the most important considerations are maximising contrast, maintaining low temperatures to minimise damage, and achieving a high tilt range for tomography. In the physical sciences the emphasis has been on ultimate resolution and analytical microscopy. This is reflected in the design of two key components, the objective lens polepiece and the specimen holder, both of which are critical for electron tomography.

The objective lens, as the main image-forming lens for TEM and STEM, is central to the performance of the instrument. In almost all electron microscopy, the objective is an immersion lens, where the specimen is located between two magnetic polepieces and thus in the centre of the lens (Figure 6.11). In general, the smaller the distance between the upper and lower polepieces, the polepiece gap, the lower the spherical aberration (C_3) of the lens. A lower value of C_3 will change the shape of the CTF and yield an improved point resolution of the microscope, and will lead to a smaller focused probe in STEM. However, the narrower the polepeice gap, the less room there is to tilt a specimen, restricting the tilt range and hence degrading the tomographic resolution. Given a set polepiece gap, (g), the maximum tilt angle, (θ) will be dependent on the width, (w) and thickness, (t) of the specimen holder in the specimen region

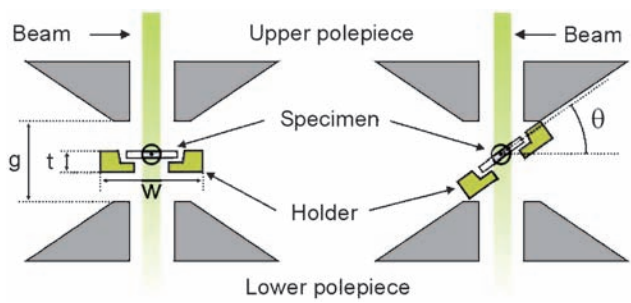


Figure 6.11 Relative geometries of the objective polepiece and specimen holder at zero tilt and at the contact-limited maximum tilt. This illustrates the importance of the thickness, (t) and width, (w) of the specimen holder and the polepiece gap, (g) in determining the maximum tilt angle, (θ). The rotation axis is marked in the centre of the specimen (eucentric).

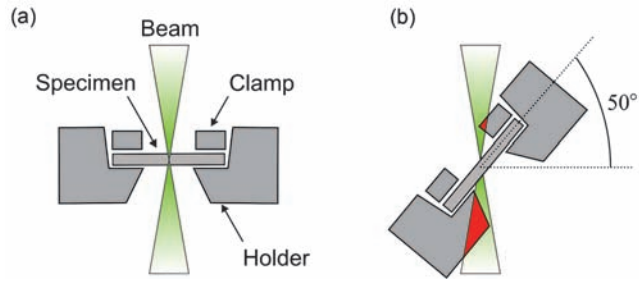


Figure 6.12 a) Shadowing of the illuminated area arising from the design of the specimen holder and clamp around the specimen. b) As the specimen holder is tilted these components of the holder obstruct the incident or transmitted beam.

(Figure 6.11). If the polepiece gap is sufficiently large, or conversely the specimen holder sufficiently small, the specimen holder may be rotated 360° within the gap. The usable tilt range will then be limited only by any shadowing from the specimen holder and specimen clamp. This is illustrated in Figure 6.12. As the holder is tilted the area of the specimen visible to the beam decreases; an object at the centre of the grid will have a larger usable tilt range than one at the edge of the grid.

In the biosciences the emphasis on high tilt and effective specimen cooling has led to the use of wide polepiece gap lenses. Conversely, in materials science the drive towards higher spatial resolution has led to the use of small (2–5 mm) gap lenses. A compromise is often reached to enable sufficient tilt to reach major zone axes and for the effective integration of an energy dispersive X-ray (EDX) detector. Given the standard TEM specimen diameter of 3 mm, a redesign of a single-tilt holder is required at the specimen-clamping position to allow freedom of rotation.[20] The usable tilt range is then limited by the shape of the holder and the specimen-clamping mechanism in the direction perpendicular to the tilt axis. Whilst this is essentially the same limitation faced by the biological community, the majority of specimens in the physical sciences do not require cooling. The need for cryotransfer, to study, for example, frozen hydrated specimens, complicates specimen-holder design considerably. Without the requirement for cooling, there has been rapid commercial development of room-temperature high-tilt holders for tomography.[63] A selection of holders for analytical microscopes is shown in Figure 6.13.

However, the design of high-tilt holders for microscopes with high-resolution polepieces is a considerably more difficult challenge. The small polepiece gap means it is physically impossible to achieve 360° tilt with a 3 mm grid specimen. The tilt range can be maximised by removing any clamping in the direction perpendicular to the tilt axis, but at the possible expense of sample stability and possible damage on contact with the polepiece.

An elegant solution to the problem of tilt range, at least for a certain subset of specimens, is the use of an in-holder tilt mechanism. Here, a narrow ''needle''

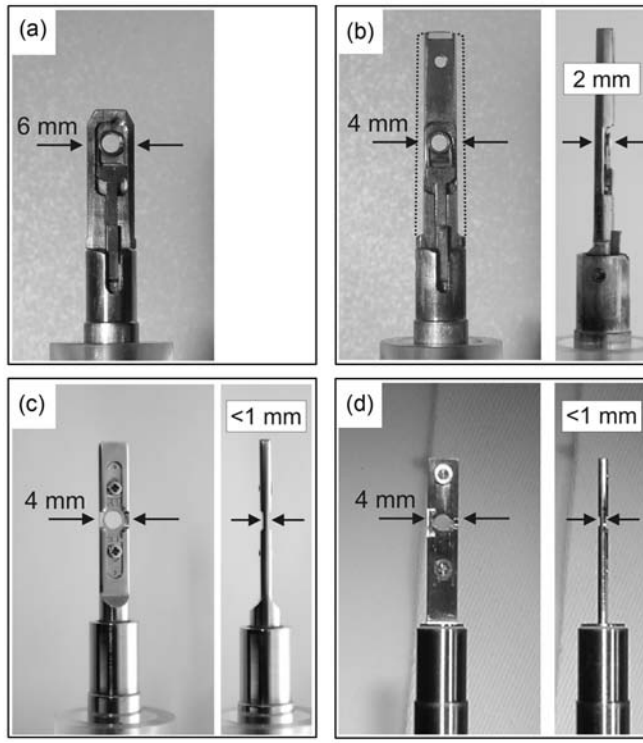


Figure 6.13 Room-temperature sample holders for analytical polepieces. (a) Tip of a standard FEI single tilt holder. (b) Tip of a Philips EM400 single-tilt holder modified to increase the maximum tilt angle. The width has been reduced (old width marked with dashed line) from 6 mm to 4 mm. (c) and (d) are commercially available specimen holders from Fischione, model 2020, and Gatan, model 916, respectively. Both of these holders have a narrow edge-on profile, reducing shadowing to allow tilting to at least $\pm 70^{\circ}$.

or pillar geometry specimen is used supported by macroscale bulk material.[64] This bulk specimen can then be carefully mounted in a specimen holder containing a rotating stub standing free at the end of the holder rod. This allows complete 180° tilt of the specimen without shadowing, and hence no ''missing wedge'', and should lead to isotropic reconstruction resolution. Preparation of a needle specimen is normally carried out by Focused Ion Beam milling (FIB) that allows the milling of site-specific self-supporting samples from almost any material.[65] However, a significant amount of damage results from FIB milling with severe ion damage (often complete amorphisation) seen to a depth of 20 nm or more.[66] For the needle geometry this will result in an amorphous coating surrounding the entire specimen. However, the promise of unrestricted tilting, the expanding application of FIB (possibly even to biological materials[67]), the introduction of low-energy ion milling to remove damaged layers and improved imagine techniques, suggests that further research in this

area is likely. The only likely limitation to needle-based morphologies is that they are not intrinsically eucentric, and it may be a challenge to keep the area under study in the field of view especially in high-magnification studies.

6.3.2 Specimen Support and Positioning

Although holder and polepiece design set limits to tilt range, the practical limit for a tilt series is often determined by the orientation of the specimen and its support. In the biosciences such support is almost always in the form of a mesh grid, with/without some from of additional low atomic number supporting film. For materials specimens this is not always the case and leads to specific issues to be addressed for each type of support.

Nanostructured materials that are available in powder form are usually dispersed on a thin film supported by a metallic mesh grid. For conventional 2D examination the specification of the film is simply to be sufficiently thin for adequate contrast. However, for electron tomography both the grid and support film should be chosen carefully. The smallest mesh size should be used that can rigidly support the carbon film. This is critical as when tilting to high angles the grid thickness becomes significant and will cause shadow. Very high tilts are possible using parallel bar grids, mounted with the bars perpendicular to the tilt axis, where few horizontal crossbars are present. For the support film, typically carbon, two factors are important, the film thickness and its planarity. The increase of projected thickness with tilt (at 70° the film will be $3\times$ thicker) can diminish the image contrast of weakly scattering materials and hence thinner films are preferred. For rigidity the thinnest films are often supported on a second, thicker, holey carbon film, although their height difference can make tomography problematic. While support films often look planar on initial inspection, heavily strained films, such as lacey carbon film, or films that have been carelessly bent or deformed, have irregularities that can cause a loss in contrast in one direction (due to increased thickness), mistakes in tracking, and in the worst case mechanically deformation or breakage during the tilt series.

6.3.3 Specimen Considerations

Central to the success of a tilt series is the specimen quality, or more accurately preserving the structure of the specimen through the tilt series while maximising the amount of data acquired. Three issues are therefore key: (i) the effects of electron-induced beam damage, (ii) specimen contamination, and (iii) automated, dose-efficient, approaches to acquisition.

(i) Beam damage

In biological microscopy, the importance of beam damage in electron tomography cannot be overstated and reconstruction resolution is dictated by the

total dose that a specimen can withstand. The theory of dose-fractionation developed by Hegerl and Hoppe[68] proposes that the total dose for a tilt series can be spread over as many projections as practicable without affecting the resultant SNR of the reconstruction. Resolution will be limited by the SNR of the original projections, the minimum being that required for adequate alignment of the images series[69] or the SNR required for statistically significant contrast in the tilted images.[49] The predicted resolution limit for unstained, frozen hydrated specimens is in the range of $\sim$2–4 nm, dependent on the thickness of the section under study. While stained biological specimens are both more beam stable and show higher contrast than frozen hydrated specimens, the spread of the stain provides a limit to the resolution and does not allow interrogation of the specimen in its hydrated, ''natural'', form.

The situation in the physical sciences is somewhat different. The relatively limited body of work applying electron tomography to (inorganic) materials specimens means the subject of beam damage during tomography experiments on such materials has yet to be examined fully. In general, however, inorganic specimens tend to be considerably more beam stable than their organic counterparts. The exact mechanism and effects of beam damage is specimen dependent and as such must be tackled on an individual basis. The microscopy of polymers will follow similar trends to those of biological structures but large differences between the chemistry of polymeric structures will lead to a broad range of beam sensitivities.

A recent review of electron beam effects on specimens is given by Egerton *et al.*[70] As a general rule, beam damage will be lessened by reducing the electron flux on the specimen and the imaging conditions should be tuned to a dose level to give sufficient SNR in the acquired projection. Of the many different mechanisms for damage the primary causes are either knock-on damage or ionisation due to beam heating (radiolysis).[71] Differing, indeed somewhat mutually exclusive, strategies for damage minimisation exist in both cases. Knock-on damage is caused when an incident electron imparts sufficient energy in an elastic collision to knock the atom from its equilibrium position, even out of the specimen itself. As such, there is a displacement threshold energy, and hence microscope acceleration voltage, above which this interaction will occur. For example, the threshold voltage for Si is $\sim$240 kV and therefore to avoid displacing atoms, one should ensure the microscope is operated below that threshold. Damage by electron heating, however, occurs when the specimen is heated *via* a large number of, predominantly of low-energy, inelastic scattering events that can provide sufficient energy to break interatomic bonds. This effect is minimised in materials with high conductivity such as a metal, but can be severe in insulating materials such as polymers. The cross section for all inelastic scattering events falls with increasing accelerating voltage and therefore damage can be reduced considerably by increasing the accelerating voltage. This is precisely the opposite approach to that for knock-on damage; highlighting the specimen-by-specimen nature of electron-beam damage.

(ii) Contamination

Contamination occurs due to the cracking of mobile hydrocarbons into amorphous carbon under the electron beam.[70] Temperature gradients at the specimen surface brought about by the electron interactions drive the diffusion of the hydrocarbons. A demonstration of the deposition of carbon in STEM is shown in Figure 6.14. One way to minimise contamination is to clean the specimen and the holder in an argon/oxygen plasma for a short time.[72] This is not suitable for many organic specimens, such as polymers, as they will be damaged by the plasma. If the specimen cannot be plasma cleaned, or contamination is still observed after cleaning, the problem can be minimised by carrying out a widespread "flood" of the specimen. This is achieved by illuminating a wide area, usually the entire grid square to be studied, using parallel illumination and a high electron flux for $\sim$30 min. This produces an even cracking of hydrocarbons across the entire illuminated area as well as sequestering the bulk of free hydrocarbons within diffusion range of the area of the specimen illuminated.

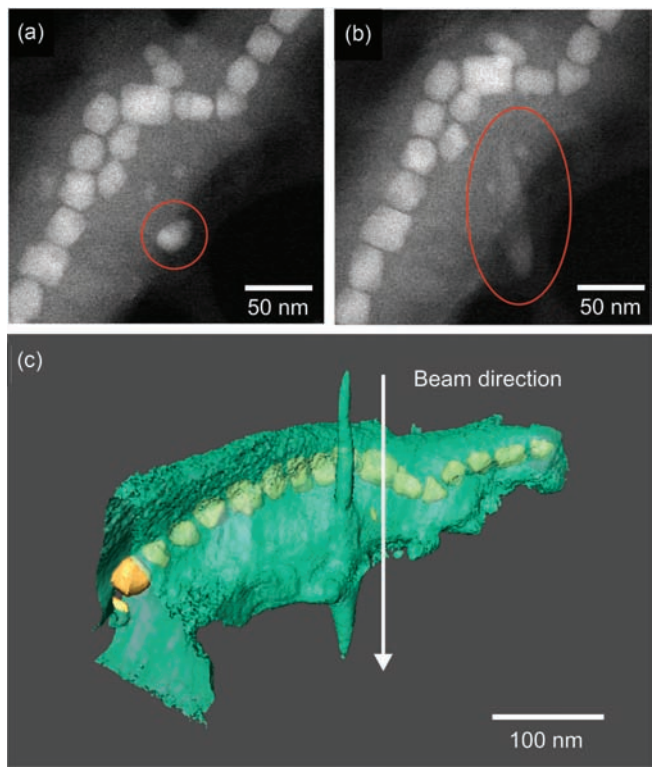


Figure 6.14 Effect of contamination STEM tomography. (a) and (b) show a carbon "pinnacle" grown accidentally by an electron beam left on the specimen prior to tilt-series acquisition. (c) A STEM tomographic reconstruction of the same pinnacle.

(iii) Automated acquisition

Acquiring a single tilt axis series is a simple concept; record a projection, tilt, record another projection and repeat until the end of the tilt range. In practice however, it is far more complex. A specimen holder is never fully eucentric and any tilt will introduce a lateral and vertical movement of the specimen area and due to mechanical imperfections this deviation is worse at high tilt. Therefore, the area of interest needs to be manually recentred in the field of view, and refocused due to specimen shift in the vertical direction, before each projection can be recorded. This increases the total exposure time, which can make this approach problematic, if not impossible, for beam-sensitive specimens.[73] In order to reduce the total dose automatic acquisition routines have been developed. By slaving the goniometer tilts/translates and microscope lens controls to a computer, the positional corrections described above can be carried out automatically.[74,75] A computer can carry out repositioning and refocusing tasks with a small number of images and in a shorter time than an operator, leading to a vast reduction in dose when compared to manual shifting and focusing. The time taken acquiring extra images for focusing and tracking are often known as the ''tilt overhead''; reducing these are the goal for any automation technique. Automated acquisition schemes usually make use of filtered cross-correlation algorithms to determine the shift between successive images. Focus can be corrected in BF TEM by measuring the image shift induced by a small beam tilt.[76] In STEM HAADF imaging, focus is determined by maximising image contrast. To minimise dose, the prime consideration for biological tomography, focusing and shift correction can also be carried out on a different area to the final acquisition,[74] a so-called ''low-dose'' methodology. Whilst still time consuming, such an approach allows a longer tilt series to be acquired with much lower total dose than required for manual acquisition. The time required to perform this automated approach is in the order of 2–4 h for a 1 degree, $\pm 70^{\circ}$, BF TEM tilt series.[73]

The recent availability of reliable computerised goniometers has enabled a more efficient approach to automatic acquisition. Goniometers can be precalibrated by acquiring a tilt series from a flat specimen with high contrast features, such as gold on a carbon film and recording image movement as a function of tilt. These calibrated movements can be applied during succeeding tilt series vastly reducing the tilt overhead.[77] As the calibration will vary with specimen height, both calibration and tilt series should be acquired at the eucentric height where the movement will be a minimum. A 140 image bright-field TEM tilt series can be acquired using this approach in as little as 30 min.[77] Following the dose fractionation theory this also increases the potential reconstruction resolution for a given series.

A new approach to automation[78,79] has recently been demonstrated based on prediction of the goniometer movement without precalibration. In both techniques, whilst using very different models for the stage predications, the sample movements are measured during the first few tilt increments and then these shifts are used to predict the movement for successive tilt increments. When

combined with checking the accuracy of these predictions, in case the character of movements changes drastically, this approach may offer even lower overheads than the precalibration method. Another apparent advantage to the prediction methodology over precalibration, is that there appears to be significant deviations in the calibrated stage movements dependent on the start x-y position of the goniometer.[78]

The image may also shift with tilt because of an offset between the optic axis of the microscope, defined by the objective lens and deflector alignment, and the eucentric position defined by the position of the goniometer. This ''optic axis offset'', typically a few micrometres in size,[80] can have a significant influence on the movement of an image through a tilt series. In theory, if this offset is known it can be corrected by the application of a beam shift above the objective lens, essentially shifting the optic axis onto the eucentric axis. Several automation schemes now calibrate this offset by measuring image movement in the direction perpendicular to the tilt axis and then build in compensating beam shifts during the acquisition to reduce any apparent image movement.[78–79,81]

A future direction for automated acquisition is suggested by Mastronade[78] in his paper on stage prediction methodology. This paper describes an ''intelligent'' fallback system, which tracks any large deviations from predicted stage movements and applies a series of corrective measures. By restricting fitting routines to measure only the newest data points, the system is reactive to rapid change in stage character. A dynamic retrack and refocus system falls back on more traditional automation techniques when a given prediction error threshold is reached. A combination of the precalibration approach with a ''passive'' prediction methodology based on acquired tilt images rather than extra images, to correct any deviations from the calibration, with an emergency fallback to conventional tracking modes, would seem a logical step forward. Such a ''dynamic'' approach to automation, combining several techniques, with a minimum tilt overhead, may be the closest to an ''optimal'' acquisition methodology, at least with current goniometer designs.

6.4 Alignment of Tilt Series

As described in the last section, the mechanical imperfections of the stage force the area of interest to be recentred for each tilt increment. This results in an approximate spatial alignment, sufficient to keep the object in the field of view. However, to achieve high-quality tomographic reconstructions all projections must be spatially aligned, preferably to subpixel accuracy, to a common tilt axis. Visually, this results in the tilt series flowing smoothly throughout the tilt range about a single axis with no sudden jumps or changes in apparent axis location. This step is the most important and often the most time-consuming stage in the reconstruction process. It is crucial to realise that there are an infinite number of common axes for a given tilt series, defined by different series of shifts in the image plane, (x_i,y_i). The alignment is performed on objects that are present throughout the tilt series, constraining the tilt axis to reside within

the reconstruction volume. Whilst the tilt axis may be moved to pass through any point of the reconstruction volume, the direction of the tilt axis is fixed and must also be determined with high accuracy for a successful tomographic reconstruction. A summary of the alignment process is given in Figure 6.15. There are two conventional approaches used to align tomographic tilt series: tracking of fiducial markers and cross-correlation. Given adequate image contrast both techniques should lead to high alignment accuracy. The advantage of the fiducial technique is that it can determine not only spatial alignment but simultaneously the direction of the tilt axis and any secondary distortions caused by optical effects and/or beam damage.[82] Cross-correlation alignment, however, makes use of the majority of the information in an image (rather than a few selected points in the fiducial technique), it does not require the subjective selection of markers, makes no assumptions about the shape of the supporting film and is perhaps easier to implement.[83] Another benefit to markerless alignment is that it avoids the reconstruction problems associated with high-contrast objects, such as colloidal gold, which can mask details in the

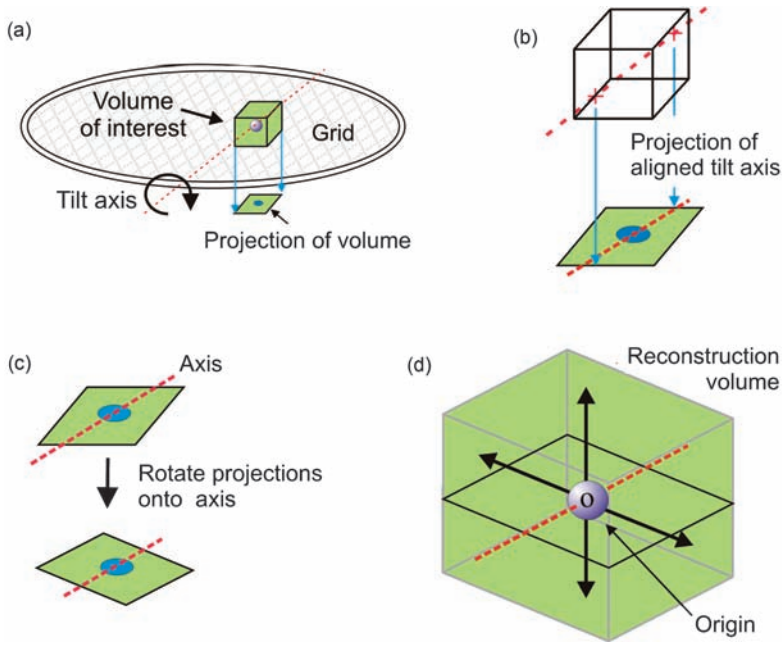


Figure 6.15 Schematic of the alignment of the tilt axis for tomographic tilt series. (a) The actual tilt axis usually away from the volume of interest. During acquisition, and subsequent alignment, this axis is reset (b) to rest inside the volume by the successive repositioning of the object/projections. The tilt axis, a line in the projection of the volume, is identified and the projections rotated (c) to bring the axis direction perpendicular to a major axis. (d) The spatial alignment can then effectively reposition the axis origin, marked o, to any point inside the reconstruction volume as long as the tilt axis remains perpendicular to this main axis.

reconstruction. Cross-correlation alignment, however, does not automatically determine the tilt-axis direction and additional techniques are needed to determine this. Alignment techniques are not mutually exclusive and are often used in combination; the technique or combination of techniques used is that which works best on each dataset.

6.4.1 Alignment by Tracking of Fiducial Markers

As an object is projected throughout a tilt series, its position and shape will change depending on the angle of rotation, the tilt-axis direction and its position. The shape of the object under study is usually not known and therefore using the object itself for tilt-series alignment is often problematic. The task of alignment can be simplified by introducing a dispersion of markers of known geometry onto the specimen grid, often spherical colloidal gold particles, which can be used as an alignment reference. The movement of these particles in each projection can be recorded and the tilt-axis direction, relative lateral shift, magnification change and image rotation determined by a least-squares tracking of fiducial markers with comparison to a reference projection.[82] This approach was first demonstrated by Hart[8] and is still the principal method for the alignment of tilt series of biological specimens, though modern methodology is based on the later research of Olins *et al.*,[84] Lawrence[85] and Beriman *et al.*[86] Figure 6.16 shows a marker alignment example in a modern alignment package, part of the TOM software toolbox.[87]

The tracking of fiducial markers is based on the solution of a series of linear equations, with functions representing all the major sources of shift, rotation and distortions including those due to the specimen and those due to the microscope. Like any solution to a series of linear equations, with a given number of unknowns, the accuracy of the solution is dependent on the amount of information provided/assumed. This information is the relative position of gold markers and the number of known factors in the imaging system. As such, the number of fiducial markers required for alignment will depend on the number of images, whether different tilt axes are used and on the accuracy of the goniometer tilt readings. In general, 15–20 markers are sufficient. Another challenge to the fiducial technique is that the selection of markers in each projection is manually intensive and prone to user error. Automated approaches to marker selection can be unreliable and often need manual adjustment,[88,89] however, the increased output of tomography data being produced due to automatic acquisition means this is still an area of ongoing research.[90]

6.4.2 Alignment by Cross-Correlation

Any two projections in a tilt series, given a finely spaced angular increment, will share many common image features offset by the relative shift between the two images. The major difference between the images will be their spatial offset that

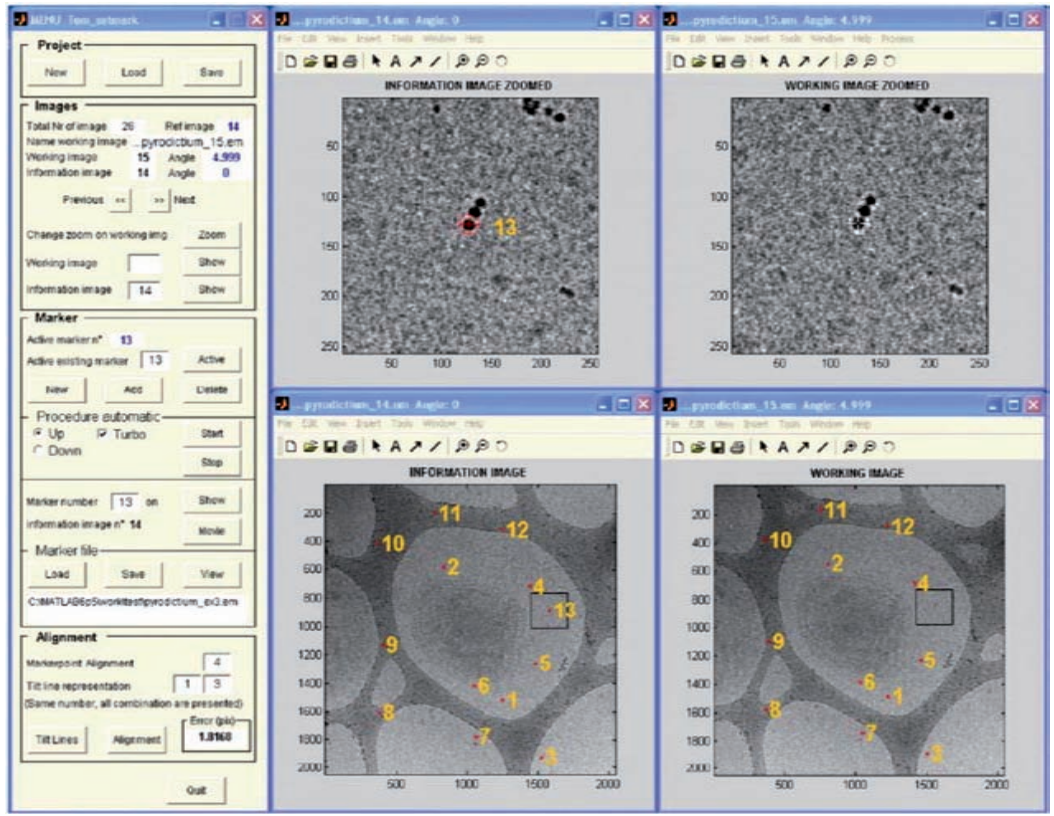


Figure 6.16 Fiducial marker tracking in a modern tomography processing environment, the TOM toolbox. The windows on the right show two successive images, at high (top) and low (bottom) magnification, in a BF tilt series. The numbers indicate identified markers in each image, and the high-magnification display allows user selection of the position of the next marker (no 13) to be selected manually. Reproduced from *J. Struct Biol.*, 149(3), Nickell, S., *et al.*, TOM software toolbox: acquisition and analysis for electron tomography, **227**, Copyright 2005, with permission from Elsevier.

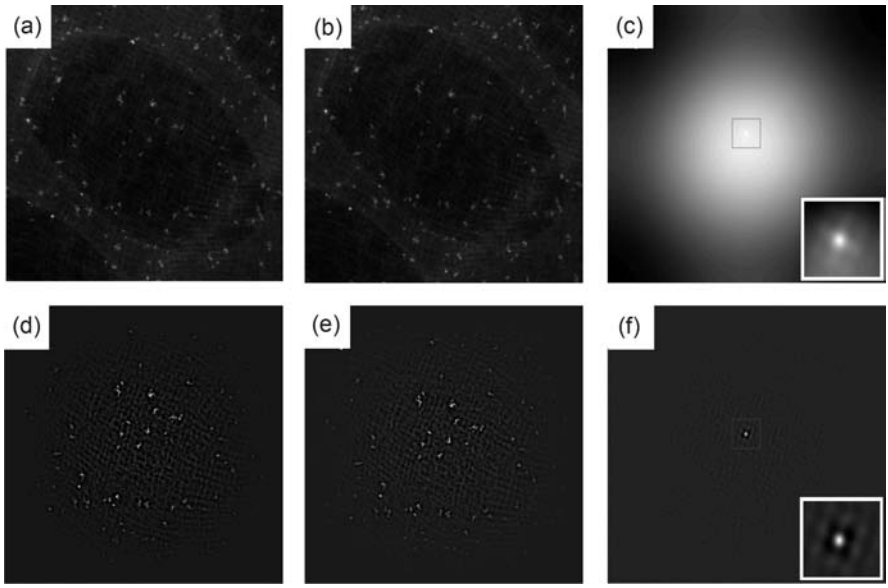


Figure 6.17 Example of cross-correlation to determine the relative shift between two images. (a) and (b) Two images of a nanostructured block-copolymer from a HAADF STEM tilt series. (c) the cross-correlation between the two images, showing the high intensity peak at (−2,19) pixels from the image centre. This corresponds to the shift between the two images. (d)–(f) are the same series of image, but show the improvement in the sharpness of the cross-correlation peak on application of a high-pass filter to the original images.

can be determined through the use of a cross-correlation algorithm.[83] The position of the intensity maximum in the cross-correlation gives the relative shift between the two images with respect to the image centre. An example of such a shift determination is given in Figure 6.17. A simple determination of the shift is given by the *x*-*y* position of the highest-intensity pixel in the correlation function but a more accurate, subpixel, measure of the shift can be determined by measuring the centre of mass of the whole peak area,[83] or by fitting the peak with a function such as a 2D Gaussian.

The quality of spatial alignment by cross-correlation is affected strongly by the contrast and noise of the images being correlated. When dealing with low-contrast and/or noisy images the calculated correlation peak may be indistinct, making the shift difficult to determine with any accuracy. In order to improve the correlation, projections are often filtered to enhance image features. Two filters are often used, one that enhances or suppresses particular frequencies, such as a high/band/low-pass filter, and one that highlights edges,[91] such as a Sobel filter. Figure 6.17 demonstrates the improvement in correlation achieved by prefiltering images. Filtered correlation is also used during tilt-series acquisition to improve the quality of automatic tracking and focusing.

Alignment of an image series by cross-correlation would ideally be achieved with reference to a single image. However, due to the change in shape of an object through tilt, cross-correlation on tomographic datasets is typically carried out sequentially. One problem of sequential cross-correlation is that accumulation of errors can occur as a consequence of small, usually subpixel, misalignments between each of the projections.[83] Over a large tilt series this build-up may be a significant source of errors in the reconstruction. In order to minimise this effect a tilt series can be split into two parts, each using the zero-tilt image as its first reference. All alignment steps can then be applied from the zero-tilt projection out to the largest tilt in both directions, which reduces the effect of cumulative shift error.

In general, consecutive images in a tilt series with small increments will contain similar information but they will not be identical. As such, the cross-correlation function between such images will show a broadened correlation peak and hence identifying the actual shift is prone to error. The difference between the two images in a tilt series is dependent on the angle (θ) between them, and following the approach of Guckenburger,[92] applying a linear "stretch" to the images of $1/\cos\theta$ perpendicular to the tilt axis has the effect of restoring the spatial relationship between successive projections.[93] In essence this converts orthogonal projections, which arise because the specimen is rotated with a fixed source and detector, into inclined projections, which would exist if only the source was rotated.

In theory, a single alignment pass using cross-correlation should be sufficient to provide accurate alignment, but in practice more than one alignment pass is necessary. This is because of the nature of rotational alignment, especially without fiducial markers, and the use of apodising filters, which tend to centre-weight image features that can subtly alter correlation, especially for large shifts. In addition, the distribution of intense image features in a tilt series, especially away from the main area of interest, may demand alignment on a subvolume in order to avoid inaccurate correlation.

6.4.3 Rotational Alignment without Fiducial Markers

Whilst the correction of an inplane rotation can be carried out by a number of different techniques[94] an out-of-plane rotation (or tilt), such as induced during a tomographic tilt series is a more difficult problem to solve. A number of approaches are presented here; solution by common lines, series summation, arc minimisation, and reconstruction optimisation.

6.4.3.1 Common-lines Approach

Between any two images of the same object, recorded at different tilts, there is some commonality. In the direction parallel to the tilt axis, the image and its Fourier transform will be identical. In the perpendicular direction, the correlation peak will be extended due to the foreshortening of image features and an estimate of the tilt angle can be made by measuring the direction of this spread.

The common-lines approach to axis alignment is based on measuring the trend in this spread throughout the tilt series. Between images that are close in tilt this spread is small and difficult to measure with certainty and between images that are far apart this spread can be masked by changes in object shape and/or poor noise statistics,[93,94] This was one of the first methods to be suggested for tilt-axis determination for both crystalline and asymmetrical specimens.[93,95] The commonality between images is also present in real space; there should be an equivalent line in each image of the tilt series. Searching for this line, in Radon (or sinogram) space, is the basis behind the most recent implementation of the common-lines approach by Liu *et al.*[96] Whilst in theory the common-lines approach is both straightforward and widely applicable, it has seen little use for real tilt series because of the difficulty of applying it to low-contrast images of biological specimens and the preference for the fiducial marker method.

6.4.3.2 Series Summation

The determination of the tilt-axis direction by least-squares tracking of fiducial markers is made possible because the movement of the markers with tilt is closely related to the direction of the tilt axis. This movement can also be used without the need for tracking to provide an estimate of the tilt-axis direction. For the single tilt-axis geometry, movement of objects through the tilt series should follow a path that is perpendicular to that tilt axis and perpendicular to the ''common line''. Assuming that the spatial, (x,y) alignment is close to optimal then a summed image over all, or some, of the tilt series should highlight the movement of any objects through the series.[97] This is illustrated for an experimental tilt series in Figure 6.18. Once the tilt-axis direction is determined the whole dataset can be rotated to place the axis parallel to the image x-axis. The accuracy of this approach is dependent on strong point image features and may be less effective for tilt-series images with low-contrast and/or large extended objects. As this approach is very rapid it is most often used as a ''first guess'' before more accurate, slower techniques are applied.

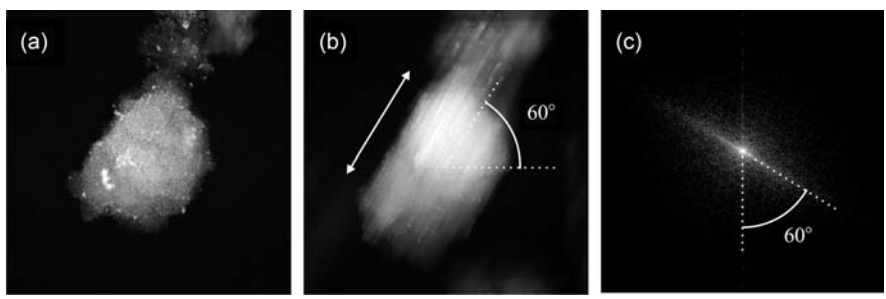


Figure 6.18 Tilt-axis direction determination by series summation. (a) A single STEM HAADF image, at zero tilt, from a tilt series acquired from a catalyst structure of a carbon matrix embedded with Pd particles. (b) The direction of the tilt axis is determined using a tilt series summation or (c) its Fourier transform.

6.4.3.3 Arc Minimisation

A misalignment of the tilt axis produces an inaccurate reconstruction that manifests itself as a smearing of the reconstruction intensity out of the correct object position into ''arcs'', the direction of which will depend on the direction of the misalignment away from the correct axis direction and the degree of ''spread'' is dependent on the magnitude of that misalignment. This misalignment is demonstrated for a test object in Figure 6.19. These distinctive distortions can provide a way to determine the axis direction by manually aligning, in real time, 2D reconstruction slices from the whole dataset. Ideally this would be carried out by rotating the tilt series before reconstructing the new slice, and for small angles this can be approximated by a shift of the projections in a single slice in the direction perpendicular to the tilt axis. The geometry used for this assumption is shown in Figure 6.20. The simple 2D shift is significantly less computationally demanding than a 3D rotation, and as such can be carried out in real time even for large datasets. The alignment is most easily carried out on three slices simultaneously, one slice in the centre of the volume and two near the edge of the volume of interest, perpendicular to the assumed tilt-axis direction.

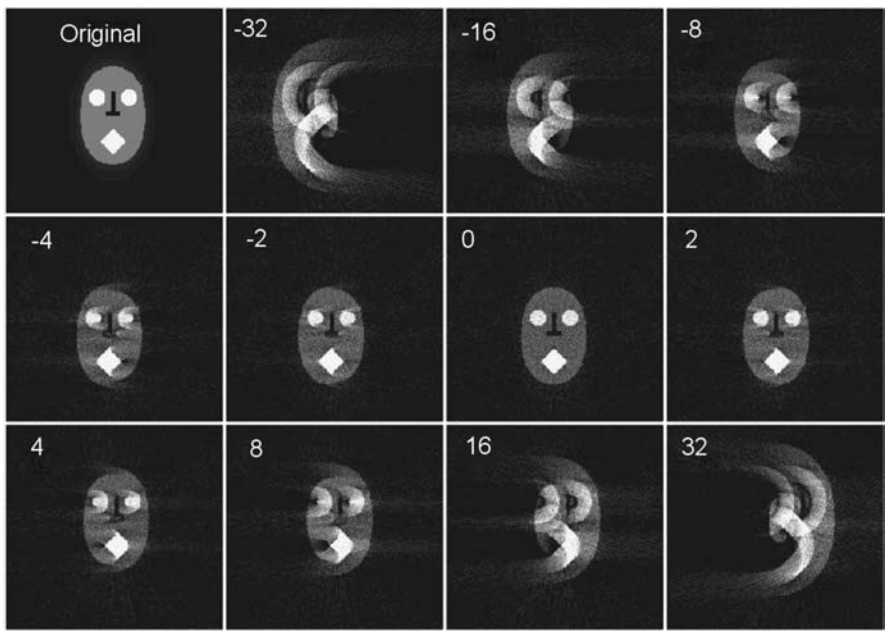
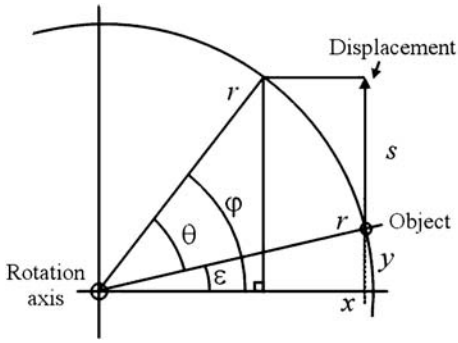


Figure 6.19 Effect of tilt axis misalignment on the reconstruction of a ''head phantasm'' test object. Distinctive spreading, ''arcing'', is observed, with a direction dependent on the direction of misalignment and the degree of spread directly related to the magnitude of misalignment. Here the misalignment introduced is in pixels perpendicular to the tilt axis, however, sufficiently small rotational alignments can be approximated by a simple shift.



$$r = \sqrt{x^2 + y^2}$$

$$\sin\varphi = \frac{y+s}{r}$$

$$\tan\varepsilon = \frac{y}{x}$$

$$\theta = \varphi - \varepsilon$$

$$\theta = \sin^{-1}\left(\frac{y+s}{\sqrt{x^2+y^2}}\right) - \tan^{-1}\left(\frac{y}{x}\right)$$

Figure 6.20 Geometry for fine rotational alignment of the projection stack. After setting the rotation axis for the central slice any rotational misalignment of an object should follow a circular path around this rotation axis. However, as only the vertical displacement, (s) rather than the path along the circle is being measured the actual correction angle, (θ) needs to be calculated as shown above.

Two variables can be adjusted: an overall shift perpendicular to the axis, which shifts all slices in the same direction, and a rotation, which shifts the two outlying projections in opposite directions and leaves the centre projection unchanged. The shift of the axis can be thought of as moving the whole dataset to the centre of rotation determined by the spatial alignment. A practical demonstration is given in Figure 6.21 for a HAADF STEM tilt series acquired from a magneto-tactic bacterium. Typically, a misalignment in the axis shift is characterised by arcing of all three slices in the same direction, (Figure 6.21(b)). If the axis shift is correct, (Figure 6.21(c)), the central slice shows no arcing. Any remaining rotational misalignment is characterised by arcing in the two outlying slices in opposite directions, (Figure 6.21(c)), the arcing will be minimised, or preferably removed, by interactive rotation of the dataset. The amount of shift needed to correct each of these images will be closely related to the misalignment of the tilt axis, which can be calculated by simple trigonometry, (Figure 6.20) given the number of pixels from the aligned object to the centre of the image, (x,y) and the number of pixels shifted, (s). The shifts should be symmetrical on either side of the rotation axis, but to allow for errors in each alignment and position of the original rotation axis, the angles determined from the two slices are averaged. This alignment is very sensitive to the slices and features within those slices and therefore should be repeated on different slices to verify the measured values.

As this alignment methodology is dependent on visual recognition of reconstruction features and manual adjustment, it is subjective in nature and susceptible to user bias. It is also more accurate when high-contrast image features are present and so its use for unstained low-contrast biological material is therefore questionable. However it can offer refinement of alignment for large continuous, as opposed to point, structures as demonstrated in Figure 6.21 that are not suited to other alignment techniques.

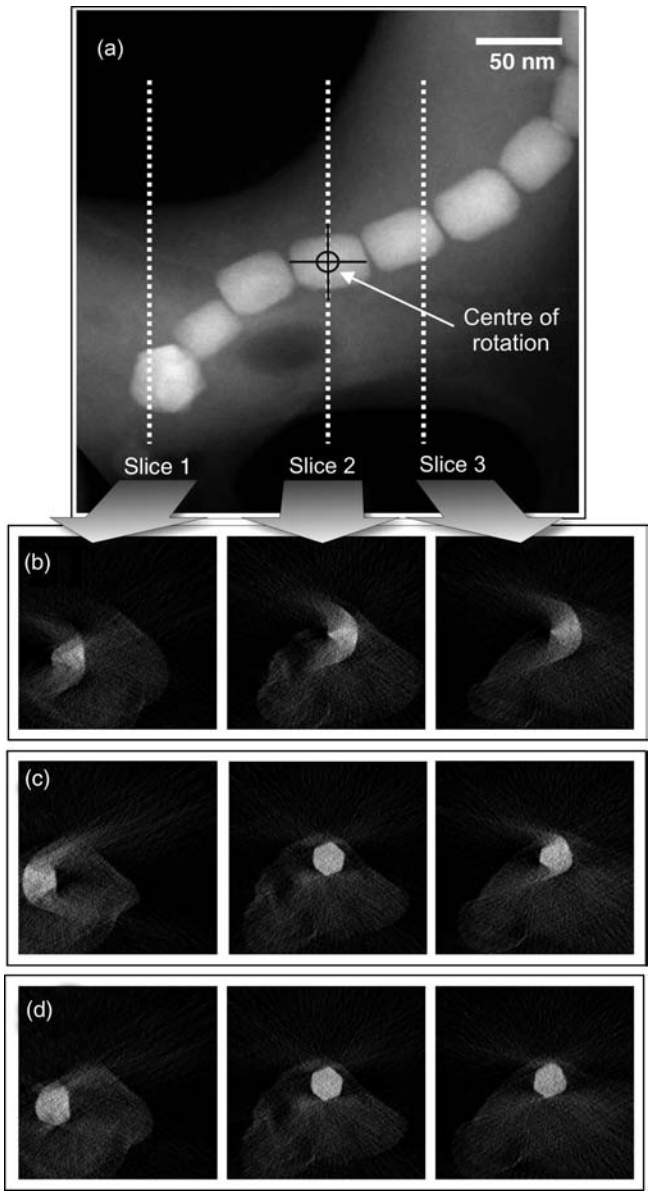


Figure 6.21 Example of fine-tilt axis alignment by arc minimisation, shown here for a HAADF STEM tilt series acquired from a magnetotactic bacterium. (a) The zero degree projection showing the location of the three slices used for axis alignment. (b) Arcing in one direction from all three slices indicates a misalignment of the tilt axis shift. (c) Here the shift is correct, the centre slice is well aligned, yet the two outlying slices show arcing in opposite directions. This indicates an error in the tilt axis angle. (d) All three slices are well aligned indicating the correct axis angle and shift have been found.

6.4.4 Other Markerless Alignment Techniques

Even though alignment by cross-correlation and some form of angular determination, by fiducial means, common lines, or observation of reconstruction, can produce good results, a more holistic automated approach to markerless alignment would be extremely useful. Several such techniques have been proposed but have yet to see broad application. These approaches include optimising the reconstruction by iterative alignment of projections and test reconstructions,[98] a new clearer approach to which has recently been published by Winkler and Taylor,[99] or by tracking of image features based on the expected geometric relationship of a tilt series.[100] None of these techniques have been applied widely, probably due to the complexity of implementation compared to existing techniques, but the promise of a robust, widely applicable, markerless alignment technique makes the likelihood of further research almost certain.

6.5 Visualisation, Segmentation and Data Mining

Achieving a tomographic reconstruction is only the middle stage of the process; the successful visualisation of a volumetric dataset can often be as great a challenge. The simplest form of visualisation is by slicing the data to examine a single 2D slice, or multiple slices, from the 3D dataset. A more attractive, if arguably less accurate, form of visualisation, is to produce a true 3D visualisation via surface rendering or a voxel projection.[101] All approaches to 3D visualisation methods have their drawbacks and which is favoured depends on the type of information required from the dataset: in general, surfaces are better at revealing distinct morphologies and topography, while voxels are better for visualising subtle variations and internal structure. Both can be made more representative and informative by combining them with some form of volumetric segmentation. Indeed, one of the most significant barriers to accurate visualisation is the difficulty of objective segmentation, especially in systems with low contrast and high noise levels.

Although visualisation is purely qualitative, there is an underlying desire to perform more quantitative analysis of volumetric data. This goal is of particular interest in the physical sciences where nanoscale quantitative 3D metrological data from a specimen can be used to predict macroscale properties and provide feedback to materials synthesis.

6.5.1 Visualisation Techniques

6.5.1.1 Slicing (Sectioning) of Volumes

It may seem counterintuitive to reduce the 3D dataset to a 2D one, but the value of slicing/sectioning cannot be underestimated. Central to this is the difference between a projection, as generated by the original microscopy, and a slice, a section through the volume only one voxel (volume pixel) thick.

A projection is compromised as the sum of all the structure projected in a single direction, whilst a slice reveals that structure unambiguously. Indeed, of all the visualisation techniques only slicing reveals the true distribution of intensities in the reconstruction volume; more advanced techniques can show an abstraction from the real data brought about by the subjective choice of visualisation methodology and/or details of the segmentation.

Slicing is mostly simply carried out along the three major axes of the reconstruction, (x,y,z). Following radiographic convention these are sometimes known as the axial, sagital, and coronal directions. Sections along these major axes are not always the most useful and modern software enables slices to be determined at any orientation within the reconstruction volume; inevitably this will involve 3D interpolation that may affect the quality of the data in that slice.

6.5.1.2 Surface Rendering

In this context, to "render" means to convert some form of data into a visual display by the addition of depth, texture, perspective and shadows, all of which help to convey the appearance of a 3D object on a 2D screen. Surface rendering produces a polygonal surface that can then be rotated and visualised from any angle and subjected to a variety of light sources, improving further the 3D effect. A distinct advantage over volumetric visualisation is that it is much less computationally intensive and navigating, rotating, translating, and zooming such a surface is very rapid. This approach is not without its drawbacks, the largest of which is the method used to select the position of the polygonal surface. A different surface will lead to an apparently different structure and hence greatly influence any conclusions made. A common approach to rapid visualisation is to use an isosurface in which a surface is generated by selecting a single threshold intensity within the dataset and generating a polygon that follows that intensity. An example of isosurface rendering is shown in Figure 6.22. Typically, tomographic reconstructions lack sharp intensity boundaries,

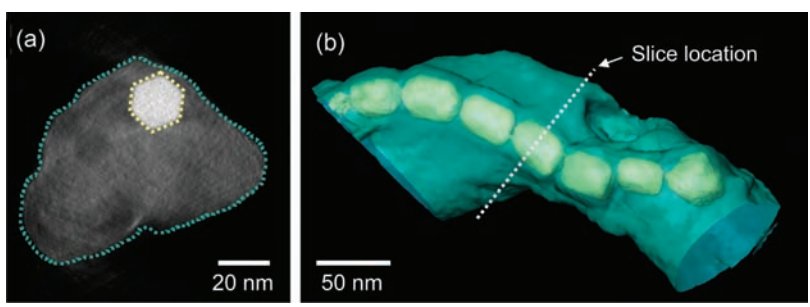


Figure 6.22 Visualisation of a magnetotactic bacterium, an ideal system for the use of isosurface reconstruction. Here, only two intensity distributions are present, the low intensity of the cellular membrane and the much higher intensity of the magnetite crystals between which there is a large, sharp, intensity step.

especially from low-contrast datasets, making the choice of the isovalue largely subjective. Another limitation is that whilst more than one surface can be visualised using transparencies, there is no simple way of using a surface to visualise a function that is changing constantly through the volume.

6.5.1.3 Volume Rendering

Volume rendering, or voxel projection, is in essence the reprojection, or directional summation, of the 3D dataset at any arbitrary angle, and as such is analogous to the original projection operation in the microscope. Since the volume is in the computer, rather than the instrument, we can modify its optical properties before projection in order to enhance the information of interest about the reconstruction. Before projection, voxels of differing intensity can be set to have specific optical characteristics such as colour, transparency or luminosity in order to differentiate them. This allows the removal of unwanted features and viewing of internal and external structures. In addition voxel projection allows any subtle changes in object density to be visualised, which is difficult, if not impossible, by isosurface visualisation. An example of volume rendering is shown in Figure 6.23. However voxel projection is computationally much slower, for an equivalent size dataset, than isosurface visualisation, as there is no significant data reduction involved.

6.5.1.4 Visualisation of Higher-dimensionality Data

Some modes of tomography will give reconstructions that have more information than simple *x*-*y*-*z* and intensity. Examples of extra information may include time, such as a time resolved series, energy, for a spectral reconstruction, or a vector, when reconstructing complex number properties such as magnetic/electrostatic fields. Despite the difficulty of using this data, static displays can still be achieved with a degree of clarity. The simplest, if least efficient, method of display is a simple series of volumes at different index (time, energy) values. However, if colour is employed to represent different indices or range of indices, all three types of information can be represented in a single, static, display. In time and energy fields these can simply be different colours for each time/energy unit, or in the latter case particular energies can be separated into different elements using a red-green-blue (RGB) visualisation. Three different chemical species can be visualised in this manner with different shades indicating degrees of mixing. An example of this is given subsequently in Figure 6.43 for EFTEM tomography of Ni, Fe and O combined into a colour volumetric visualisation. Vector fields can also be visualised by colour, with the different directions indicating different field directions, using a rainbow colour wheel, where the direction of the field is indicated by the colour of the pixel, in 2D for a slice or projection, which can be modified, with increased complexity, for 3D.[102] Due to the complexity of vector fields a series of other options exists for their visualisation. These include displaying the inplane component of the field for a given slice and slice direction (analogous to the conventional arrow

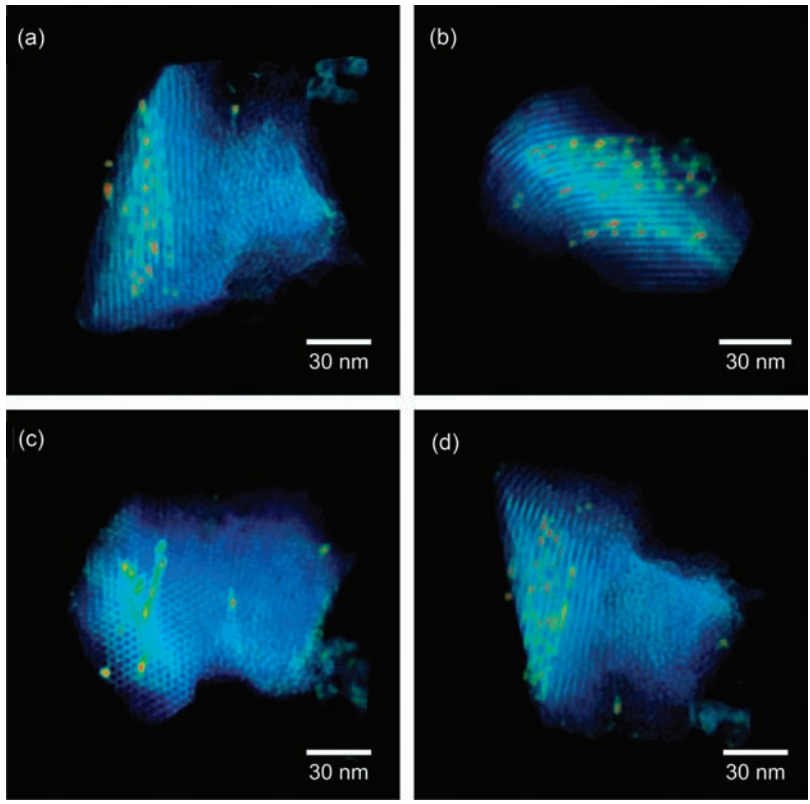


Figure 6.23 Volume visualisation of a heterogeneous catalyst, the matrix (blue) is nanostructured silica (MCM-41) embedded with iron oxide precursors (red-yellow). (a)–(c) The volumetric visualisation allows the hexagonal pores to be resolved along the three major axes. The reconstruction reveals regions of pores that are nonregular and do not incorporate the precursor, and the nonhomogeneous distribution of the precursor particles in the regular pores.

map), displaying field direction and magnitude by 3D arrows and displaying field lines as a texture on a surface.

Visualising the extra dimensionality of data can be greatly aided by generating movies, as opposed to static displays. These fit naturally for time-resolved data, but can also be extremely valuable for vector reconstruction by animating field lines.

6.5.2 Segmentation

Perhaps the most challenging aspect to the analysis of tomographic reconstructions is the general ambiguity in the boundaries between reconstruction features. While some systems show clear delineations, such as the example in Figure 6.22, this is the exception rather than the rule. The separation of

greyscale pixels, or voxels, into regions of structural, functional or chemical similarity is one of the central challenges of machine vision across many fields of research.[101] There have been many different methods posited to carry out this task of segmentation. The simplest and the most widely used approach, certainly in electron tomography, is manual segmentation by eye. In each slice, boundaries are drawn around each feature by hand and whilst this is both time consuming and subjective the resultant visualisations can be highly instructive. Two visualisations from manual segmentation are given in Figures 6.24 and 6.25, for biological and physical science reconstructions, respectively.)

A robust and successful automatic, or semiautomatic segmentation method would undoubtedly be of great value in a wide range of fields, including electron tomography. A wide range of such methods has been developed and some have recently been applied in electron tomography. These can be broadly classified into those that operate globally and those that operate locally. An isosurface is possibly the simplest form of global segmentation. However, far more advanced approaches to global segmentation have recently been demonstrated by Frangakis and Hegerl[103] who applied eigenvector analysis of the image grey

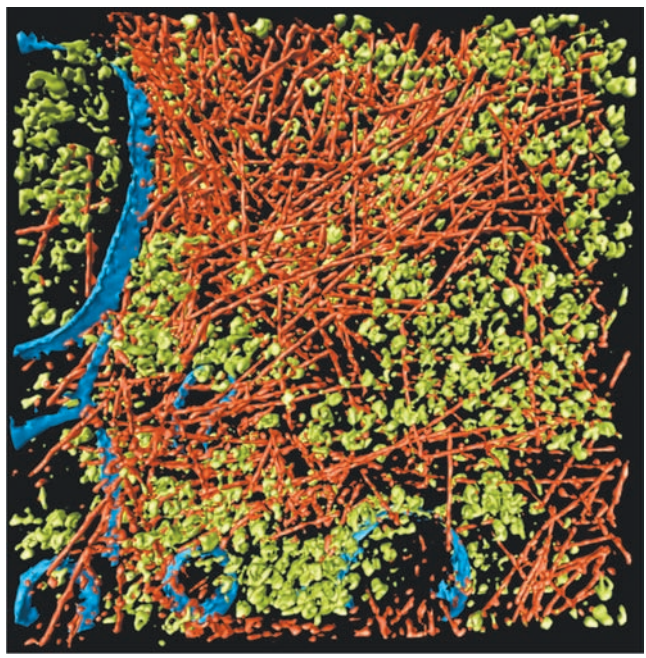


Figure 6.24 An example of volumetric segmentation applied in the biosciences, in this case a thin cryosection of eukaryotic cells from the organism dictyostelium discoideum.[191] The field of view is 815×870 nm and 97 nm thick. Here, cellular membranes (blue), actin fibres (red) and ribosomes (green) have been hand segmented and surface rendered. Each of these features shows far too low a contrast to be visualised adequately in any other way. Reprinted with full permission from O. Medalia, *et al.*, *Science*, **298**, 1209–1213 (8 November 2002). Copyright 2002 AAAS.

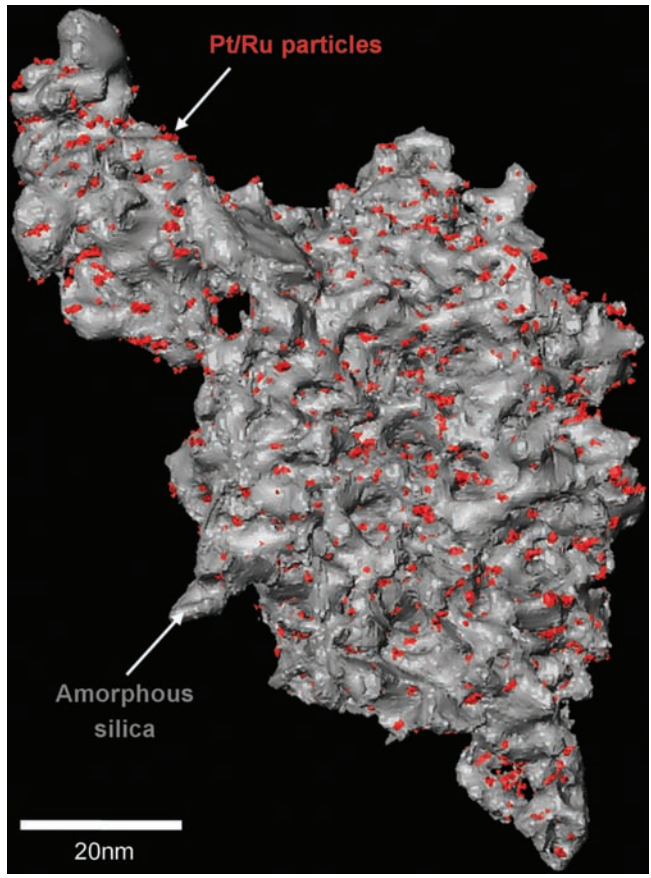


Figure 6.25 Demonstration of segmentation in the physical sciences: a surface render of a segmented volumetric reconstruction of a heterogeneous catalyst. Here, grey is the mesoporous silica support for the red Pt/Ru bimetallic nanoparticles.

values and their relative proximity to segment a volume. A larger number of studies have demonstrated the application of local approaches to segmentation, that usually rely on the detection of changes in volumetric intensity due to some feature of interest. An early application of local segmentation to electron tomography was use of Watershed transforms by Volkmann,[104] which treats voxel intensity as a relief map to be "filled" and sets thresholds, the "watersheds", above which connected areas are segmented into different volumes. Bajaj *et al.*[105] have adapted another general approach to segmentation, gradient vector diffusion,[106] to electron tomography. Gradient vector approaches are based not on the image intensities but on the local differential gradients within the image, and as such should be less prone to local irregularities in voxel intensity.

Practical demonstrations of automatic segmentation in electron tomography are very small in number, due in part to the difficulty of implementing such

algorithms but primarily on the unsuitability of tomographic reconstructions of biological material for automatic segmentation. The small density differences in biological structures, which are made up primarily of water and protein, result in reconstructions that show very small intensity differences between different structures and their surrounding medium. In addition poor sampling and high noise levels in the acquired projections result in a low SNR in the reconstruction. This gives any segmentation routine very little information, voxel by voxel, to separate out potential volumes. However, automatic approaches are at least highly useful as an initial "first pass"[107] or as part of a semiautomatic approach to segmentation.[108]

The high density of many physical science specimens, or more importantly large density differences between different material components and/or free space, suggest that segmentation may be more applicable to materials science tomographic reconstructions. This will be particularly true for imaging techniques that show strong contrast between phases such as HAADF STEM and EFTEM. Application of automatic segmentation in such systems has yet to be demonstrated but given the interest in quantitative analysis of volumetric properties, such as volume fraction it is an obvious direction for further research.

6.5.3 Quantitative Analysis

Whilst quantitative analysis of volumetric data is of value in both the biological and physical sciences, the nature of the properties of interest, and how they are measured, in each will differ to a large extent. Quantitative analysis in biological tomography is usually directed to the examination of the 3D morphology of macromolecular assemblies, usually by some form of structure matching. Such examples include the docking of high-resolution protein structures, often determined by X-ray diffraction, to a lower resolution, yet larger and more complex, tomographic reconstruction.[109,110] In materials science, however, the interest is typically not in the analysis of local structural arrangements, as in general this structure is not perfectly repeatable, but of global properties such as number density, roughness, curvature, connectivity and partitioning of species across/around an interface.

Despite the relatively recent developments of electron tomography in the physical sciences there are already a number of examples of quantitative volumetric analysis. The early study of Laurer *et al.*[14] demonstrated that characteristic lengths of repeating structures, in this case block copolymers, measured from reconstructions agreed very closely with bulk techniques such as small-angle X-ray scattering (SAXS), the advantage being that electron tomography can offer highly localised analysis. Further analysis of these reconstructions, covered in detail in Section 6.6, has been shown to produce insights into the physics behind their growth.[111] Given a segmented volume it is also relatively straightforward to measure the size distribution of structures as demonstrated by Ikeda *et al.*[112] for silica inclusions in natural rubbers. The

same group also analysed the nearest-neighbour distances between carbon black inclusions in rubbers.[113] Their results show a very close correlation between this nearest neighbour difference with the measured bulk volume resistivity; linking nanoscale quantitative tomography to macroscale physical properties. The activity of a catalyst is determined by its loading, the mass of active particles per unit area. Electron tomography is the only technique that can provide *direct* measurement of loading for many catalysts. The measurement of particle loading of a nanoscale heterogeneous catalyst has been demonstrated by Midgley *et al.*[114] using HAADF STEM tomography.

With electron tomography becoming increasingly routine in the physical sciences the extraction of volumetric numerical information is likely to become an area of significant research in the coming years.

6.6 Imaging Modes

There are a large number of imaging modes that can, in theory, be used for electron tomography. The most dominant of these is unquestionably bright-field (BF) TEM, using both amplitude and phase contrast. This is not because it is always the most suitable technique but more that it is the most widely available and easiest to implement. It is available in nearly all electron microscopes and as it has been used in the life sciences for many years, the majority of automated acquisition implementations, alignment tools and reconstruction algorithms are designed for BF TEM. The explosion of interest in the tomography of materials has, in part, been driven by the demonstration of the power of alternative imaging modes.[20] These techniques, however, are more difficult to implement and less widely available, but the potential suggests that their use should increase even to the levels of BF tomography. This section attempts to summarise some of the most important issues in applying each imaging technique to materials, including a discussion of specific experimental issues and, where possible, highlight applications of each.

6.6.1 Bright-Field TEM

The history of BF TEM tomography is the history of the electron tomography technique itself; from the first reconstructions by Hart[8] and De Rosier and Klug,[7] to the latest high-resolution reconstructions from frozen hydrated specimens.[115] As such, there is a wealth of literature and for a more complete picture the edited book by Frank,[116] combined with more recent reviews by Koster *et al.*[73] and Lucic *et al.*[117] are recommended.

The first demonstration of electron tomography of a materials specimen was that by Spontak *et al.*,[12] applying BF TEM tomography to block copolymer systems. This was followed by a number of studies by other research groups in the polymer field where the technique has seen considerable application in the study of these systems.[14,16,118,119] Many polymers share similarities with biological systems, including the lack of long-range order and low atomic mass,

and the application of electron tomography seems a logical step in their analysis. It was a study by Koster *et al.*[15] that suggested the wider applicability of electron tomography to inorganic materials. Originally working in the life sciences, Koster was able to apply state-of-the-art tomographic techniques to the examination of heterogeneous catalysts. Their most recent results show impressive reconstructions of both the nanostructured support, with regular and irregular pore structures, and the embedded active particles.[120] Following Koster's work BF tomography has been applied over a broad range of complex nanoscale materials, including for example the arrangement[121] and connectivity[122] of multiwall carbon nanotubes. Given the interest in complex nanostructures built around nanotubes it seems likely that tomography will be key in such developments in this field. In the same year as Koster's paper Mardinly *et al.*,[61] from Intel, demonstrated the earliest results of BF tomography applied to semiconductor device materials. Their study revealed the position of failures in metallic vias that were not visible in projection. With the size of semiconductor device features dropping every year, and with new designs becoming more complex, it is becoming apparent that defects will be harder to isolate and characterise using conventional 2D TEM. Both Mardinly, and Stegmann *et al.*,[19] recognise that some form of 3D characterisation, especially if it can be implemented as a standard failure analysis technique on the production line, will be of great value. The least likely application for BF TEM tomography would appear to be to metallic systems. The polycrystalline nature of most metals combined with their relative ductility, when compared to semiconductors, gives rise to a large amount of diffraction contrast. This arises due to the bending of the specimen, differences in orientation of various grains and large number of defects such as dislocations. However, tilt series from an extremely deformed metallic specimen has recently demonstrated that even in these cases fine structural information can be revealed.[123] This study shows highly sheared Guinier–Preston (GP) zones of silver precipitates in an aluminium alloy; the reconstruction is of sufficient quality to estimate local strain from the elongation of the GP zones. Success here was due in part to the much higher atomic number of silver compared to aluminium, which would lead to large underlying amplitude contrast.

6.6.1.1 Meeting the Projection Requirement

Contrast in bright-field images comes from a number of different mechanisms; depending on the material under study. In practice, some amplitude contrast is nearly always present, and is primarily the result of the elastic scattering of the incident electron with the sum of the atomic potentials. The intensity of scattered electrons will rise with specimen thickness and with atomic density and as such amplitude contrast is often known as mass-thickness contrast. For thin, weakly scattering objects, by defocusing the image, changes to the phase of the incident wave can be converted to intensity variations. Such phase contrast in the BF image is sensitive to the contrast transfer function (CTF)

which describes the transfer of information by the objective lens for each spatial frequency and is dependent primarily on the low-order aberrations such as defocus, astigmatism and spherical aberration. For weakly scattering objects, such as frozen hydrated specimens in the life sciences, the CTF can be corrected and the true image, or exit wavefunction, recovered. Such an image shows a almost linear relationship with the amount of material in projection. For strongly scattering crystalline objects as normally found in the physical sciences, the interaction of the electron beam with the specimen is dynamical and phase and amplitude components are scrambled by the lens transfer. Diffraction contrast will be present in a BF image and vary dramatically depending on the orientation of the crystal with respect to the incident beam. The variation is extreme and in general has no simple relationship to the structure of the object. Such contrast does not satisfy the projection requirement and will lead to serious artefacts in a tomographic reconstruction. Nevertheless, in some situations as shown above, reconstructions of crystalline specimens from BF TEM micrographs have been performed, and have revealed valid structural details in 3D. However, very careful interpretation of results is required.

6.6.1.2 Experimental Considerations

Crystalline materials should be studied by a tomography technique that minimises diffraction contrast. However, in many cases, due to the lack of the necessary hardware/software or experience in applying them, this is not possible. In such cases there are strategies to minimise the deleterious effects of diffraction contrast. As diffraction contrast is caused by excluding Bragg reflections with an objective aperture, one approach might be to increase the size of that aperture. This will reduce the diffraction contrast but will also reduce the desired amplitude contrast and lead to images of poorer spatial resolution because of the incorporation of off-axis beams. Another approach, that has been demonstrated for EFTEM tomography,[124] is the use of a "hollow cone" illumination mode.[125] In this case, the prespecimen deflectors are used to tilt the beam at a given angle by a few milliradians and then the beam is continuously rotated azimuthally through 360° at high speed. This constantly changes the diffraction conditions resulting in an averaging over all diffraction conditions, thus reducing, but not eliminating, the effects of diffraction contrast. This mode is not without problems, however. Due to the limited accuracy of the microscope deflectors and the range of paths through the objective lens, conical illumination will lead to a reduction in image resolution and for higher magnification studies this approach may be completely unsuitable.

6.6.1.3 Application Example No. 1: Block Copolymers

The earliest applications of electron tomography in materials science were to block copolymer systems,[12] where the native complexity of the microstructure makes this kind of analysis invaluable. There are a range of structures that can form in these systems, depending on the number of individual monomers used

and their mutual immiscibility, ranging from simple lamellar to complex gyroid structures, a summary of the types of structures possible is given in Figure 6.26. This structure is thought to arise from the competition between attaining the lowest possible mean surface curvature that minimises the contact between immiscible phases[126] and the cost of accommodating that curvature on the packing of molecules.[127]

One example of the more complex morphologies studied is the triblock system produced by the combination of styrene-b-isoprene-b-styrene studied by Laurer *et al.*[14] Here, the b refers to a ''midblock'' connection between monomer units, which can be structurally accommodated in two different ways. This is different to a simple covalent link that has to be accommodated at the boundary between the two monomer units. As a consequence, midblock systems tend to more complex structures. A set of 2D BF TEM images of the microstructure of this system is shown in Figure 6.27. The strong contrast in these images is due to an OsO_4 stain that is preferentially absorbed by the isoprene monomer. A BF tilt series was acquired from this specimen using

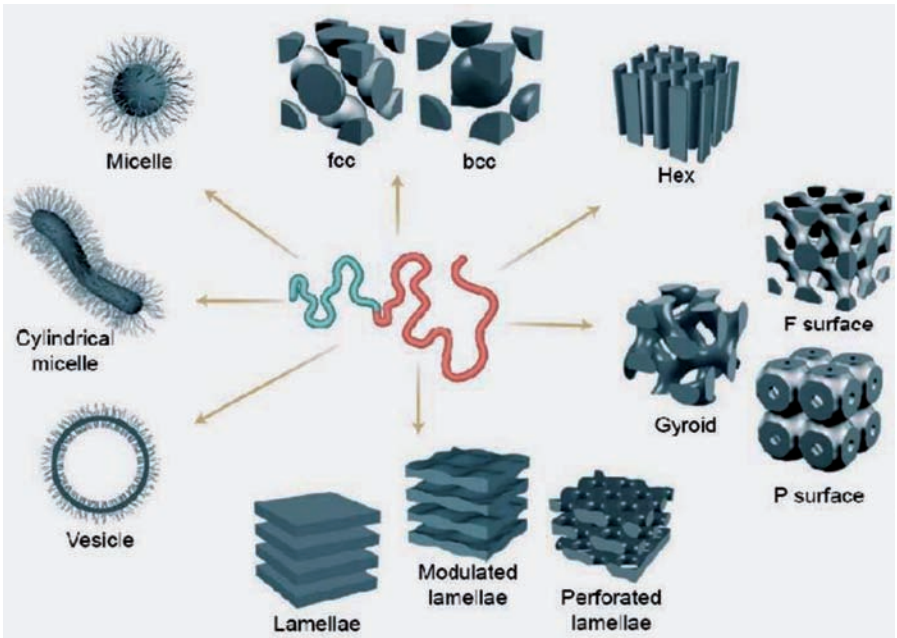


Figure 6.26 The range of structures formed from block copolymer systems. These simple chains of two or more different monomer units can be mixed to form any of the structures shown above. The actual structure formed depends on the monomer chemistry, fractions and chain lengths. From Bucknell and Anderson,[192] adapted from Forster and Plantenberg.[193] Reprinted with full permission from D.G. Bucknall, and H.L. Anderson, *Science*, **302**, 1904–1905 (12 December 2003). Copyright 2003 AAAS.

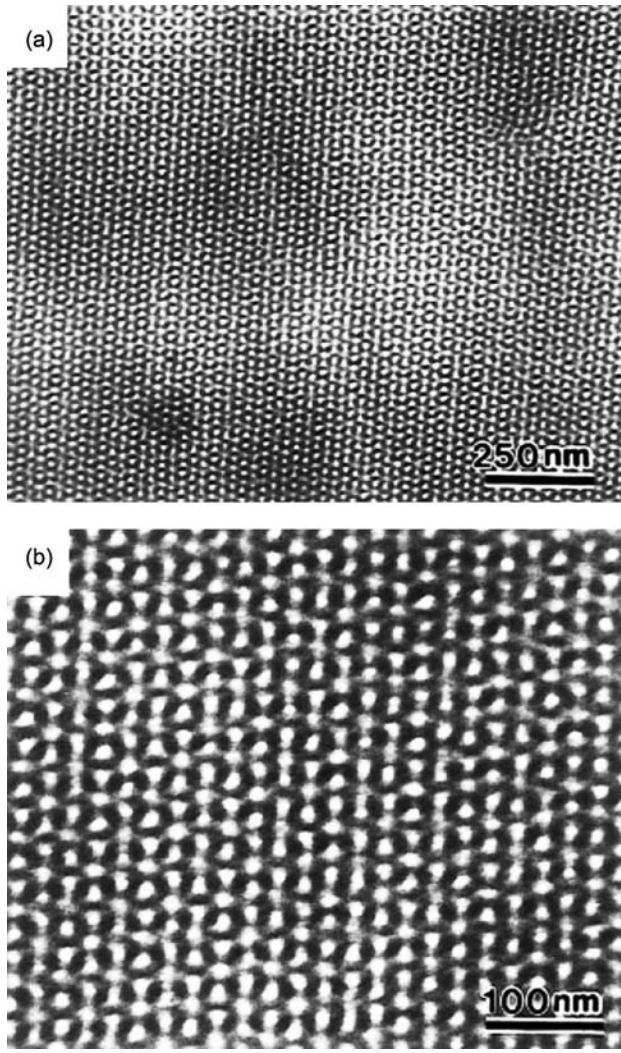


Figure 6.27 BF TEM images of an SIS (styrene-isoprene-styrene) triblock copolymer. The strong contrast arises from osmium tetroxide staining of the isoprene rich, dark, phase. This system appears to form an extremely regular, gyroid like, structure Reprinted with permission from J.H. Laurer, *et al.*, *Macromolecules*, 1997, **30(13)**, 3938. Copyright 1997 American Chemical Society.

automated acquisition, with images recorded over a range of $\pm 60^{\circ}$ every 2.5°. Examples of the reconstruction from this dataset are shown in Figure 6.28 where the styrene components have been volume rendered from different angles. This reconstruction clearly reveals the complex interlocking gyroid type structure that arises from two interlocking, nonconnected, regions of styrene in

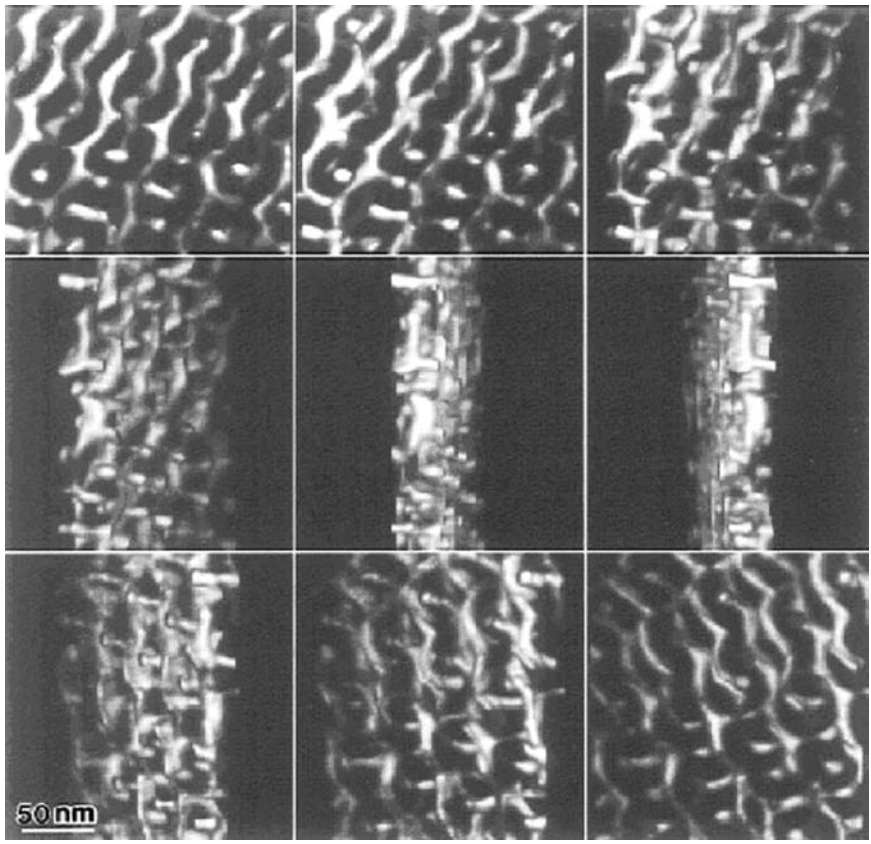


Figure 6.28 Volume renders taken from the tomographic reconstruction of the SIS block-copolymer. The images show a rotation, 20° between images, of the volume through a single axis (vertical). Reprinted with permission from J.H. Laurer, *et al.*, *Macromolecules*, 1997, **30(13)**, 3938. Copyright 1997 American Chemical Society.

the (clear) isoprene matrix. The volumetric nature of the data, and the clear separation between volumetric components allows quantification of various structural parameters. The styrene volume fraction, at 0.35, closely matches the expected value, 0.32, and the significant repeat unit lengths closely match those measured by small-angle X-ray scattering (SAXS).

However, the reconstructions generated from this system can also be used to examine the fundamental forces behind the formation of such structures. If this formation is driven by interfacial energies only, then a constant mean curvature (CMC) will be observed throughout the volume, the value of which is that of the ideal gyroid. However, it has been postulated that deviations from this curvature will arise due to the difficulties in accommodating such curvature by the stacking of individual monomer units and the limited interfacial area for a given amount of material, a factor known as ''packing frustration''. The results of Jinnai *et al.*,[111] following on from those of Laurer, carried out a quantitative

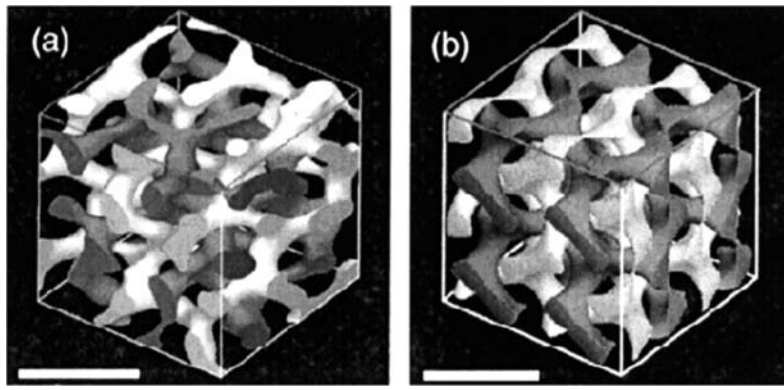


Figure 6.29 Comparison between reconstruction (a) and ideal (b) gyroid structure for the SIS triblock copolymer. While the experimental and theoretical structures match closely, analysis of the small deviations allows an investigation into the physics of its formation. Reprinted with permission from: Hiroshi Jinnai *et al.*, *Phys. Rev. Lett*, 2000, **84**, 521. Copyright 2000 by the American Physical Society.

analysis of the curvature of the styrene network. Figure 6.29 shows (a) the tomographic reconstruction and (b) the ideal gyroid structure, which at first glance appear similar. However, the results of quantitative analysis shows a statistically significant deviation from the value of mean curvature from that which would be expected if this was the only factor. This study allowed the first quantification of the effects of packing frustration on a real system, invaluable information when trying to predict structures and properties from first principles.

6.6.1.4 Application Example No. 2: Heterogeneous Catalysts

The performance of any catalyst is strongly dependent on the surface area available for the reaction being catalysed and the density of sites for that reaction. For reactions involving small molecules, some of the most effective catalysts are based on highly porous "molecular sieve" supports with pore diameters in the range of 1–50 nm. These small pores offer huge surface areas. For heterogeneous catalysts a high number density of active particles, spread evenly throughout the support volume is optimal. With such small length scales and complex pore/particle morphologies, electron tomography is one of the few ways to study directly their local structure.

This was first demonstrated by the application of BF electron tomography to zeolitic structures by Koster *et al.*[15] Zeolite structures are classical examples of "molecular sieves" as a catalytic support material, being low density, chemically inactive and with a complex pore structure. The native sub-nanometre pore structures of a "microporous" system are at present too fine to be resolved by 3D electron tomography. However many chemical reactions require larger

pores to allow access for larger molecules and such ''mesoporous'' materials can be created by chemical treatment, by acid leaching or steaming. Two significant problems that cannot be solved by 2D examination are the partitioning of active particles between exterior and interior sites and the morphology of the pores. In Koster's paper the positions of catalytic Ag particles in a NaY type zeolite were studied, the reconstruction of which is shown in Figure 6.30. Large silver particles are located on the zeolite surface and smaller ones in the interior. The second example is the reconstruction of a mesoporous system, based on acid-leached mordenite, is shown in Figure 6.31. Here, Koster compares a conventional 2D TEM projection, in which the pore position is ambiguous, with a slice through a reconstruction, which clearly shows the position and size of the pores inside the zeolite volume. This is a classic example of the unambiguous information available in a voxel slice of finite thickness compared to a projection.

6.6.1.5 Limitations

Bright-field TEM should have seriously limited application in materials tomography due to its sensitivity to diffraction contrast. Despite this it has been shown, surprisingly, to be able to produce meaningful reconstruction detail in a number of examples.[19,61,123] A study by Friedrich *et al.*[62] has shown that BF electron tomography of crystalline structures can introduce serious artefacts in reconstructions. Diffraction contrast, and to some extent Fresnel fringes, leads to a bright ''ring'' around features, in this study isolated single-crystal nanomagnets, in the reconstruction. In addition, the study has shown that restored intensities of real, and test, objects have significant deviations from expected values. In general, it seems BF tomography of such objects can restore the exterior shapes (topography) of objects with reasonable accuracy, even if the intensities of the interior of the object are questionable. This suggests that the technique will be seriously compromised for buried structures, especially those that are crystalline. The study of Kubel *et al.*[60] compared reconstructions of a semiconductor device using both BF TEM and HAADF STEM. The improvements gained from the latter technique proved to be considerable, even to the point of revealing porosity inside a metallic contact that was invisible in the BF reconstruction. The conclusion seems to be that BF TEM tomography can give 3D information on crystalline objects, but the results should be interpreted very carefully.

Another issue that can be problematic with BF TEM is one of sensitivity/interpretability of very small features. A prime example is the kind of heterogeneous catalyst similar to that studied in the above example by the Koster group. Such catalysts exist over a wide range of length scales down to single atom active sites and subnanometre pores. A recent study of the 2D imaging of catalysts at smaller length scales with 3 nm pores and subnanometre active particles, by Thomas *et al.*[128] shows very clearly the difference between BF TEM and HAADF STEM, (Figure 6.32). The lack of delineation between particles in the BF image, especially compared to the clarity of the HAADF

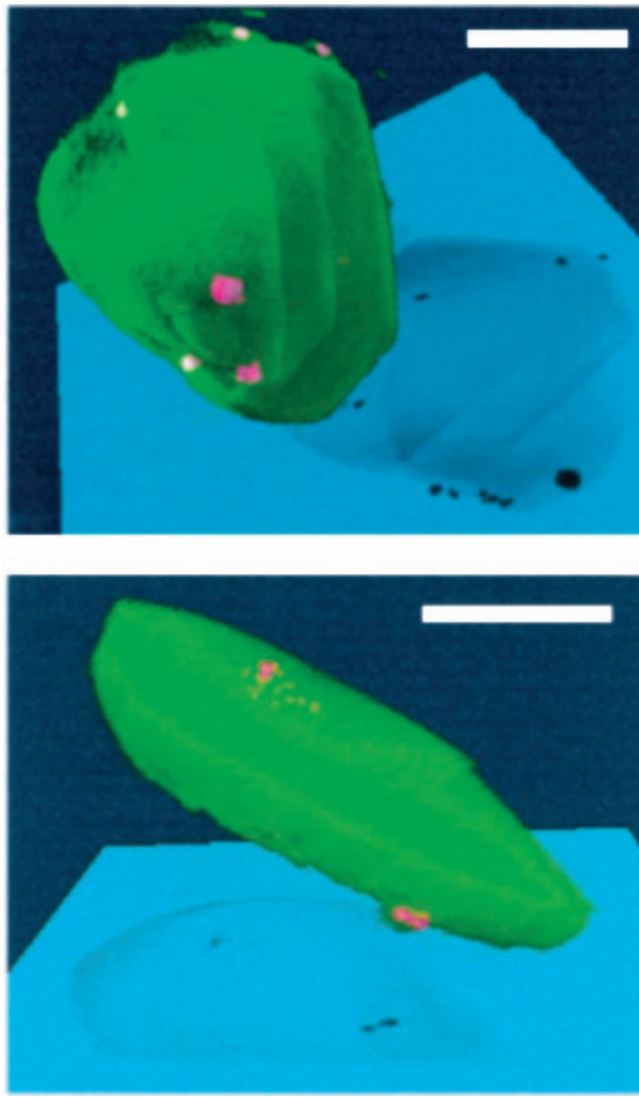


Figure 6.30 Voxel visualisation of the reconstruction of a NaY zeolite substrate (green) embedded with Ag particles (pink). There are two distinct distributions of particles, large particles on the surface and small particles on the interior. Scale bar is 200 nm. Reprinted with permission from A.J. Koster, *et al*., *J. Phys. Chem. B*, 2000, **104**, 9368. Copyright 2000 American Chemical Society.

STEM example is striking. Although the effects on the quality of a BF reconstruction from such an area has yet to be studied, it does suggest for certain materials that there may well be problems of detection for very small features.

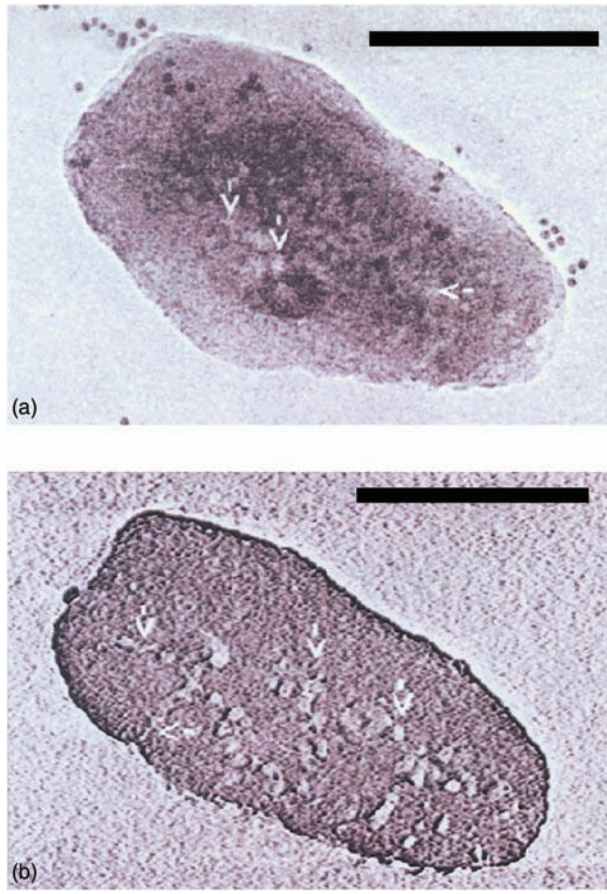


Figure 6.31 Comparison between projection (a), (a conventional TEM micrograph), and a slice through the tomographic reconstruction (b) from the same structure. Note the small dark objects are gold fiducial markers used for alignment, rather than active particles. Scale bar is 100 nm. Reproduced with permission from A.J. Koster, *et al.*, *J. Phys. Chem. B*, 2000, **104**, 9368. Copyright 2000 American Chemical Society.

6.6.2 Dark-Field Tomography

Whilst diffraction contrast does not satisfy the projection requirement, there is a way in which diffraction can be used for tomography. Dark-field (DF) TEM imaging is used to image features by selecting electrons that scatter to a particular Bragg reflection using an objective aperture in the back focal plane of the objective lens. In this situation, the projection requirement can be partially satisfied by ensuring the diffraction condition remains constant throughout the tilt series for the object(s) of interest.

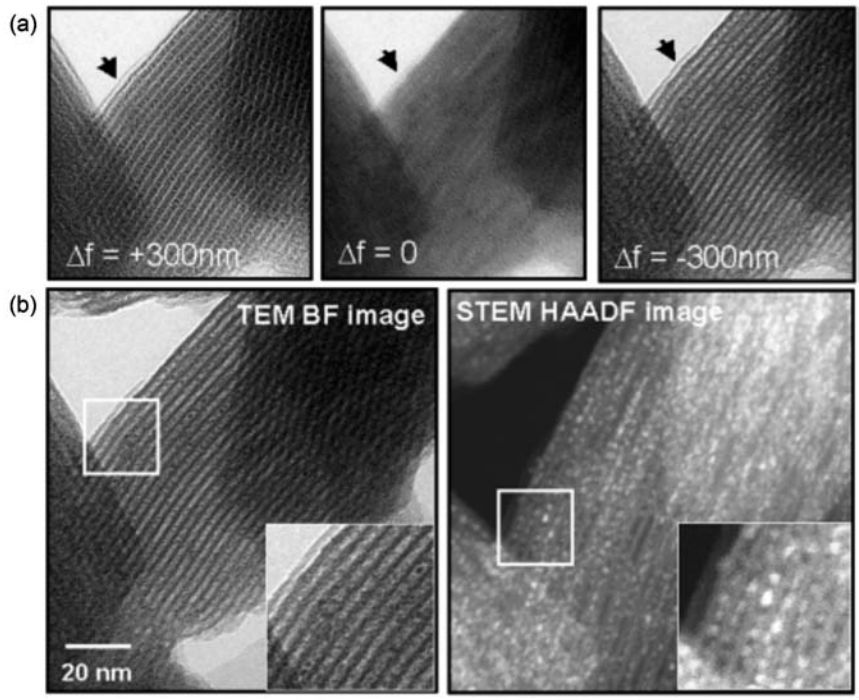


Figure 6.32 Comparison between BF/HAADF STEM for the imaging of nanoscale heterogeneous catalysts. (a) Bright-field TEM images from three different defocal values of nanoparticle loaded mesoporous sili*ca*. (b) A comparison between BF TEM and HAADF STEM imagine of an identical area of catalyst. The magnified view, boxout, shows the clear differences in visibility of the active particles. Reproduced from Thomas *et al.*[128]

To date there have been very few examples of this form of tomography. One is the use of a superlattice reflection to image the block-like structure of γ' precipitates in a Ni-based superalloy.[129] Another is the use of weak-beam dark-field (DF) TEM tomography,[130] to reconstruct the 3D structure of a dislocation network in a GaN epilayer, (Figure 6.33). This was the first time 3D images of dislocations had been reconstructed with electron tomography.

The central limitation to DF TEM tomography is the need to maintain near-constant diffraction conditions throughout the tilt series. The dislocation example works because there is a tightly defined relationship with the host crystal and the dislocations at this level are simple linear structures. In general, this technique is however very difficult to perform and suggests its application may be limited. However, the extreme sensitivity of the technique, allowing differentiation between structures showing only very small, milliradian, deviations in crystalline orientation, may make DF TEM tomography a technique worth exploring further.

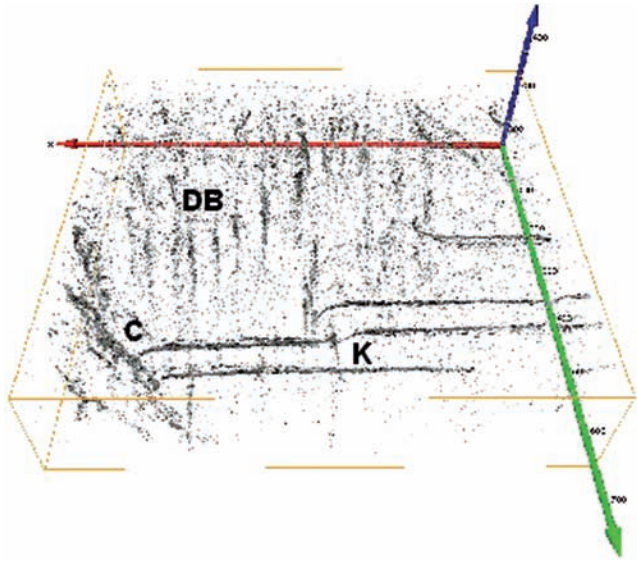


Figure 6.33 Tomographic reconstruction of an array of dislocations in a GaN epilayer. Threading dislocations align to form low-angle domain boundaries (**DB**) while threading dislocations close to a crack (**C**) turn over and glide away leaving an inplane segment that is a pure screw. One of these has interacted directly with another threading dislocation (**K**).

6.6.3 HAADF STEM

Of all the new techniques for electron tomography HAADF STEM shows, arguably, the most potential for application to nanostructured materials. First demonstrated in 2001[131] this technique has already been used to study the 3D structures of a range of materials systems and has also lead to the availability,[81] and application,[63] of automated acquisition software for STEM tomography.

The first application of the technique was to mesporous hetreogeneous catalysts. In this case a study by Midgley *et al.*[131] applied HAADF STEM tomography to an ordered mesoporous silica, MCM-41, with 3 nm pore widths, embedded with nanometre-sized Pd_6Ru_6 particles. The large atomic number difference between the support and the particles led to high image contrast and both the pore geometry and particle position were reconstructed successfully, even from a somewhat limited tilt series. More recent studies on heterogeneous catalysts by HAADF tomography, using more complete tilt series, clearly demonstrate the power of the technique for resolving the structure of heavy active particles embedded in a lighter matrix.[114,128] Other early results include the application to the faceting of biogenic crystallites in magnetotactic bacteria, the shape of which may be of importance in identifying biomarkers in Martian meteorites.[132] The quality of the surface render from these crystals, examined further in a more recent study,[133] suggests the

possibility of using HAADF STEM tomography for 3D metrology of nanoscale systems. One of the areas that metrology seems most relevant is that of semiconductor materials, and the results of Kubel *et al.*[60] show that HAADF tomography may be an ideal technique for such studies. The high resolution attainable by the technique also makes it the ideal tool for examining nanomaterials, and examples studied include quantum dots,[133,134] nanotubes[135] and optical nanomaterials.[136] Indeed, it is the broad applicability of HAADF STEM tomography across a range of nanostructured materials that is, possibly, its most salient feature.

6.6.4 Meeting the Projection Requirement

There are a number of different geometries and detectors used for STEM, a range of which are shown in Figure 6.34. A conventional annular dark-field (ADF) detector collects intensity from a large number of Bragg spots simultaneously, an effect that averages over the dark-field reflections producing an image that is approximately the inverse of the BF image. In addition, electrons are collected on the ADF detector from high-angle elastic (Rutherford scattering) and thermal diffuse scattering (TDS). Early studies using ADF detectors produced images of sufficient resolution to image small heavy-metal clusters, and even single atoms, on light substrates.[137] This is due to the metal clusters being sufficiently small such that to a large degree each atom can be thought of as scattering independently, showing no diffraction effects. As all the contrast is dominated by Rutherford-type scattering, the intensity of the image depends on

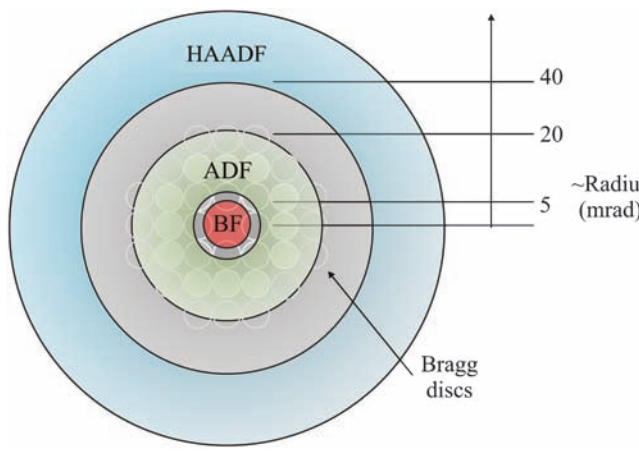


Figure 6.34 STEM detector geometry in the diffraction plane, for 300 kV accelerating voltage. The BF detector collects the primary, unscattered beam. The ADF detector collects electrons that have been both Bragg (coherently) scattered and incoherently scattered (TDS/Rutherford). The HAADF detector radius is large enough that Bragg reflections are excluded, and the collected electrons have been primarily incoherently scattered.

the atomic number, Z. However such "incoherent" imaging using ADF was restricted to small crystals or noncrystalline specimens. It was proposed by Howie[138] that it should be possible to form incoherent images, even from crystalline specimens, by increasing the inner radius of the ADF detector so as to exclude Bragg-scattered beams. The intensity of images collected with this detector is a function of both the atomic number (Z) of the scattering atom, approaching a Z^2 relationship at high detector inner radii and the projected thickness. In essence, this is very similar relationship to mass-thickness contrast and in qualitative terms, the HAADF signal can be thought of as the reciprocal of a mass-thickness BF image. However, at low scattering angles electron screening will serve to reduce the Z-dependence and as a consequence, *high-angle* annular dark-field images show enhanced contrast over equivalent *low-angle* mass-thickness images. This is a distinct advantage in systems where the structures of interest are characterised by only small changes in average Z. The lack of coherence in HAADF images means that such images record a simple projection of the structure in terms of thickness and atomic number. This relationship clearly meets the projection requirement and as such HAADF images are eminently suitable for tomographic reconstruction.

Although STEM HAADF imaging is essentially incoherent in nature the intensity of the image can be modified by the orientation of a crystalline specimen. When atomic columns are aligned closely with the beam direction, such as at a major zone axis, strong Bloch wave channeling tends to redistribute the beam intensity onto the atomic columns as it propagates through the crystal.[139] This increases the relative current density around the atomic cores, increasing the likelihood of core localised scattering events such as Rutherford and thermal diffuse scattering. This has the overall effect of increasing the amount of high-angle scattering and consequentially the intensity in the HA-ADF image.[140] To avoid this effect the crystal can be tilted away from strong channeling conditions, but as a tilt series is being acquired from many different angles, whose increments are tightly controlled, this effect may be unavoidable. However, unlike, for example, BF diffraction contrast, the effect is far more uniform across a crystal, the change in contrast is not as extreme as that of conventional diffraction contrast and is significant only at a small number of orientations and thus in a small number of images within the series. As such, the effect of channeling-enhanced contrast through a tilt series tends to have limited effect on the tomographic reconstruction.

6.6.5 Experimental Considerations

The fundamental differences between TEM and STEM imaging enforce a number of significant changes to tomographic methodology.

The position of the STEM probe on the specimen is determined by two sets of electromagnetic scan coils that are usually mounted to be mutually perpendicular. When the scan coils are calibrated they allow scanning of the beam in a square grid pattern that can be mapped to a square array of pixels. However, if

the scan coils are inaccurately calibrated, there can be significant deviation from this ideal. This can manifest as both angular distortions, where x and y are not perpendicular, and magnitude distortions where x and y steps are not identical in size. Of course, the recorded image will still be square (if set to be so) and so this leads to significant distortions in the resultant 3D reconstruction. Normally, this distortion is corrected by eye with reference to a cross-grating specimen and whilst this is reasonably accurate, the application of a more advanced, preferably automated, procedure will be required for metrology applications.

As has already been mentioned, an ADF detector will suppress coherent contrast when operated with a inner radius large enough to avoid collection of low-angle Bragg scattering. Therefore, for crystalline specimens care should be taken to operate the instrument in a true HAADF mode with a sufficiently high collection angle; $\sim$40 mrad at 200 kV. On modern TEM/STEM instruments this can be adjusted by modifying the camera length of the projection lenses. The exact inner radius can be calibrated and measured with reference to a known diffraction pattern,[3] such as Si $<110>$. However, if studying amorphous objects, or crystals that are small enough to scatter independently,[137] there will be no significant diffraction contribution to the contrast. As the overall scattered intensity falls rapidly with increasing scattering angle it is more dose efficient to use a lower inner angle (*i.e.* greater camera length) in order to maximise the signal collected on the annular detector.

Acquisition times for TEM tilt series have dropped dramatically in recent years by minimising the ''tilt overhead'' for each tilt increment through the use of prediction or precalibrated acquisition methodologies.[77,78] Whilst a similar reduction in STEM acquisition times can be achieved, STEM DF images take significantly more time to record than parallel-beam techniques. This is simply a factor of the time taken for acquiring a nondistorted STEM image with sufficient SNR. Acquisition times for TEM images are on the order of $\sim$1 s but STEM images are rarely acquired in less than 10 s and often considerably more. In order to acquire a STEM image the probe is scanned, point-by-point, in a grid pattern at each point waiting (the dwell time) whilst the signal is recorded by the relevant detector. While the current density in a STEM probe is very high, the total current delivered to the specimen is similar to that for parallel illumination.[141] Therefore, even for an equivalent dose, the STEM image will take longer to acquire than BF TEM because the number of electrons collected on the high-angle detector is only a small fraction of those incident on the specimen. Whilst the contrast achieved by collecting at high angles is considerable, the acquisition time must be sufficiently long to overcome background noise. While acquiring rapid scans is possible, indeed acquisitions of 1–3 s are generally used when traversing the specimen grid, the nature of the scan electronics means that such images show significant image distortion. Thus, unless the electronics are improved, rapidly scanned images, even if pixel-to-pixel noise is low, are unsuitable for tomography. A study of the comparative signal and dose incurred by parallel beam TEM and high angle STEM has been presented by Rez.[141] Although this study concentrates on applications to

biological specimens it is instructional when comparing the differences in acquisition times between the techniques.

Automated acquisition in TEM tomography makes extensive use of autofocusing based on the shift-with-tilt methodology, as discussed in Section 6.4.[76] However, this approach cannot be made practicable for STEM as the convergent nature of the beam introduces significant image distortions when tilting the beam that mask any induced shift. Another consequence of a convergent beam is that STEM is far more susceptible to small changes in defocus than TEM. However, this short focal depth can be exploited for automated focusing, as the sharpness and contrast of a STEM image will change rapidly with defocus. A defocus series should show a trend of increasing/decreasing values for these factors and, if large enough, allow the optimum focus to be estimated. For indistinct or noisy features, a contrast-enhancing filter, similar to that used for cross-correlation, may be used. Autofocusing in this mode requires a series of images, rather than just the two for shift-with-tilt, and it is more time-consuming than its TEM counterpart. Combined with the slow acquisition times in STEM, this results in a significant tilt overhead for STEM autofocusing.

A further consequence of the small depth of field is that there is a limited thickness of specimen that can be focused to give optimal resolution. In addition, as the specimen is tilted to high angles and the projected thickness increases, it is likely that not all of the specimen can be in focus at the same time. Theoretically, this is also a concern in TEM, but the specimen is rarely thick enough to have a significant defocus effect for parallel-beam illumination. However, there is an elegant solution to this problem in STEM. By rotating the scan (essentially changing the balance of the x to y coils) the tilt axis of the specimen can be made perpendicular to the direction of the rapid STEM scan. Given the simple geometric relationship between tilt, specimen height and defocus it is possible, for every scan line, to adjust the beam crossover to match the change in specimen height. A focal ramp is applied across the image, minimising the problems associated with the limited depth of field; this is known as a ''dynamic focusing'' approach, which works best for a slab-like specimen.

When carrying out STEM tomography the effects of carbon deposition, and indeed beam damage, can be minimised by the application of good ''beam etiquette''. This involves minimising the time for which the beam is stopped in one place on the specimen, and if the beam has to be stopped it is well away from the area needed for the tilt series. If a pause for some time is required it is often better to acquire a rapid scan and drop the magnification, spreading the dose over a wide range, than stop the beam altogether.

6.6.5.1 Application Example: Magnetotactic Bacteria

Magnetotactic bacteria, discovered by Blakemore in 1975,[142] are an unusual class of organisms that have internal chains of small magnetic crystals (magnetosomes) within their cell walls. The alignment of these single-domain

nanomagnets allows the bacteria to orient themselves with the Earth's magnetic field, and hence a preferential direction to remain in an ideal habitat. These bacteria are seemingly ubiquitous in bodies of fresh water in which the magnetic particles are magnetite (Fe_3O_4) and in seawater where they are iron sulfide (Fe_3S_4).[143] These nanomagnets are an important example of biomineralisation that may also play a part in the homing instinct of pigeons,[144] have health implications in humans,[145] and for possible use as recording media. International interest in magnetotactic bacteria was stimulated further by a paper[146] that suggested similar magnetite crystals were found in Martian meteorite ALH84001 and may indicate past life on Mars. The key similarity between these crystals is their three-dimensional morphology, as studied by Thomas-Kerpta *et al.*,[147] which they suggested arose only from biogenic sources. These morphological studies were based on simple 2D BF comparisons and there was a great deal of controversy as to the actual shapes of both terrestrial and (supposed) Martian magnetosomes.[132]

One of the first demonstrated applications of HAADF STEM tomography was to examine the exact morphology of magnetosomes of known terrestrial origins. These crystallites are ideal for the STEM technique, see Figure 6.35, because of the strong Z-contrast between cellular material and magnetite. The first tomography data published was acquired over a limited tilt range, $\pm 56^{\circ}$,

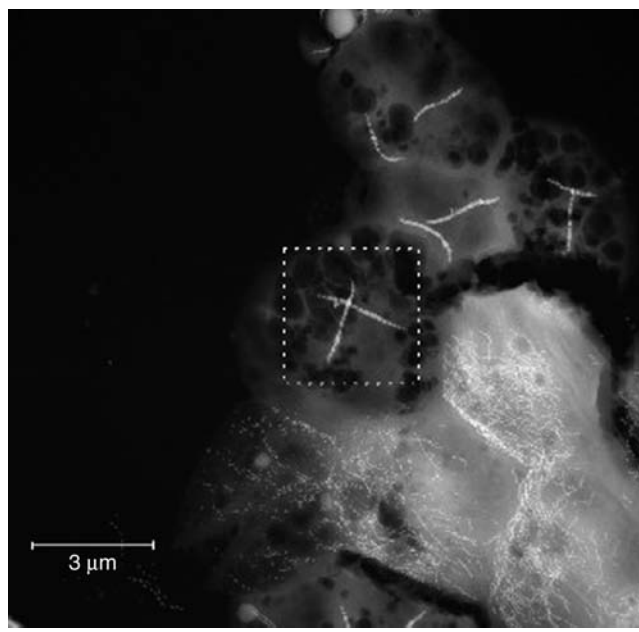


Figure 6.35 Low-magnification STEM HAADF image of a typical area of the magnetotactic bacteria specimen. The high-Z magnetosomes show strong contrast compared to the rest of the low-Z cell remnants. Also clearly visible are at least two different size distributions and chain morphologies.

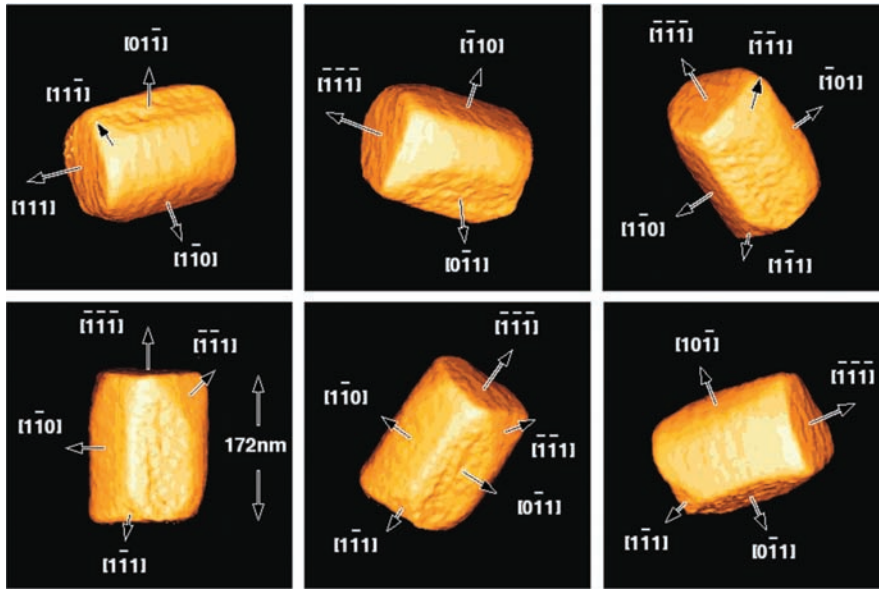


Figure 6.36 Isosurface renders from the reconstruction of a single magnetosome from the horizontal chain indicated in Figure 6.35. The viewing angles have been chosen that highlight the shape and position of the surface facets and demonstrate how the faces can be indexed.

yet the reconstructions, carried out using iterative techniques,[33] reveal the morphology of the magnetosomes unambiguously. Figure 6.36 shows that the major facets can even be indexed. This study demonstrated the applicability of the technique to the Martian meteorite specimens, but perhaps more importantly to nanocrystalline specimens of all kinds.

The particular bacteria to which the Martian magnetite have been compared are the strain of bacteria MV-1 (marine vibrios-1), which show small (<60 nm) magnetite crystals of regular elongated octahedral shape. Elucidating the shapes of these smaller crystals required the use of more sophisticated tomography techniques and, in particular, the first application of an ultrahigh tilt specimen holder designed for narrow polepiece gaps. The resulting tilt series was acquired from a wide tilt range of $\pm 76^{\circ}$ and resulted in a reconstruction of both the magnetite crystals and the outer membrane, (Figure 6.37).[133] The effects of the increased tilt range are shown in Figure 6.38, with the resultant increase in reconstruction fidelity, especially with regard to the surfaces in the lowest resolution direction.

6.6.6 Limitations

The limited depth of field available in STEM appears at first to offer a serious limitation to the applicability of the technique. If the effects of increased object depth at high tilts are tackled using dynamic focusing, as discussed above, the

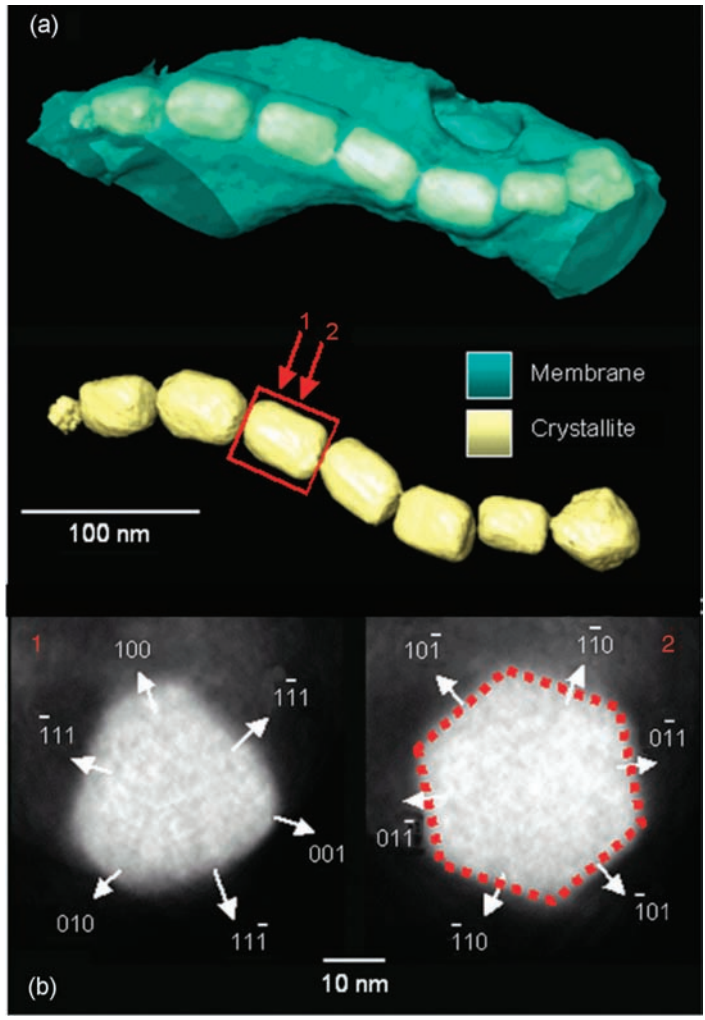


Figure 6.37 (a) Tomographic reconstruction of magnetotactic bacteria strain MV-1. The chain of magnetite crystals is evident. (b) Slices taken through a single magnetite crystal in the chain reveal the perfect crystal faceting of the cubic crystals, as revealed by the superposition of a regular hexagon on slice 2.[133]

limitation then becomes the resolution required in the reconstruction from the analysed volume. As the volume increases the maximum attainable resolution decreases. From a simple geometric optics limit there is approximately 1 nm of spread in 100 nm thickness of material (given a 10 mrad convergence semiangle). The resolution of tomographic reconstruction also scales with the volume under study, following the Crowther formula;[30] a comparison of the two factors is shown in Figure 6.39 for a 141 image tilt series. In this case it becomes clear that

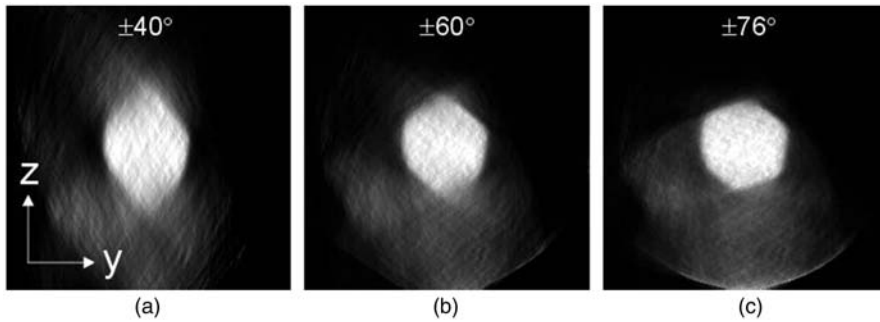


Figure 6.38 The effect of holder-tilt range on the reconstruction of a magnetite crystal from a HAADF STEM tilt series; a single 2D slice (with the tilt range used). (a) Tilt limit ±40° (typical analytical microscope tilt limits). (b) Tilt limit ±60°. (c) Tilt limit ±76° from an ultrahigh-tilt holder.

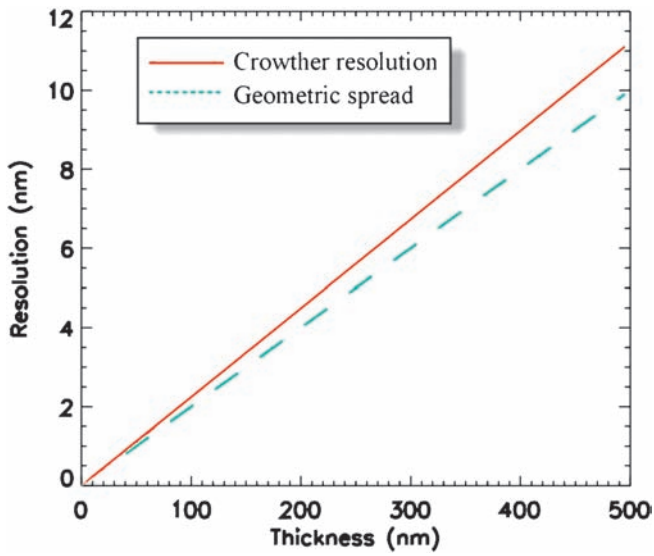


Figure 6.39 Effect of simple geometric beam spreading versus Crowther resolution with specimen thickness assuming the tilt series contains 141 images (±70° tilt range with a 1° increment).

for most tilt series, where the tilt increment is rarely less than one degree, the controlling factor for tomographic resolution will be the limited sampling rather than the depth of field. One effect of the convergent beam, however, will be that different heights inside the specimen will be sampled with different resolution and what effect this has on the overall reconstruction has yet to be studied.

Ultimately, very dense or thick specimens will be a limit to HAADF STEM tomography. The thicker the specimen, or the higher the atomic number, the

larger the average scattering angle. However, as the outer angle of the detector is finite, a significant proportion of the scattering may fall outside the outer edge of the detector and for specimens of high mass-thickness, this can lead to contrast reversals, and strong deviation from monotonic behaviour. This is demonstrated in Figure 6.40 for a focused ion beam (FIB) prepared "finger" specimen of a semiconductor device. At zero tilt the contrast is as expected, with the tungsten contacts appearing brighter than the lower-Z silicon substrate. However, at high tilts, and large projected thickness, the contacts appear less intense as the majority of the signal has fallen beyond the outer radius of the detector.

Another consequence of long acquisition times is that STEM images can be distorted by specimen drift. This is a particular problem when carrying out high-magnification, atomic-resolution, STEM imaging where crystal lattice images can show significant "shear" due to stage drift. However, at the lower magnifications typically used for tomography, drift should not be so severe as to cause problems during acquisition. However, there are occasions, where a strong thermal gradient between holder and microscope/room is present and large drift can occur. These should be easy to identify and the source of the problem removed before acquiring a tilt series.

As already mentioned, the effects of beam damage should be considered on a specimen-by-specimen basis. However, operating in STEM modes makes several considerations about beam damage relevant. Initially, one would consider that the large current density in the STEM probe will result in greater damage. However, in specimens where beam heating is the controlling factor, STEM appears to minimise the damage caused; the total dose to the specimen can be much lower than in conventional TEM imaging and the rastering nature of the

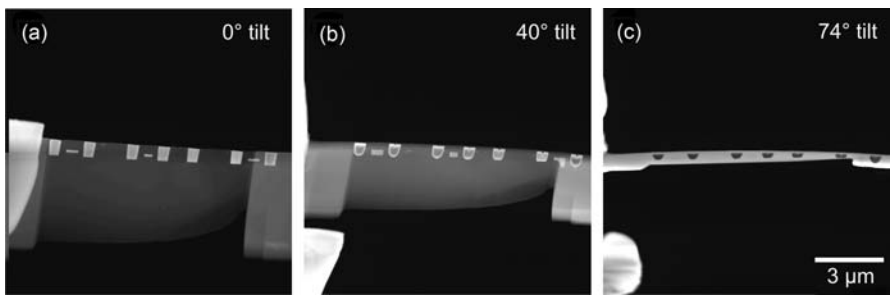


Figure 6.40 The effect of high thickness/atomic number on HAADF STEM contrast. A focused ion beam (FIB) milled pinnacle specimen of a Si based transistor with tungsten (W) contacts). (a) At zero degrees the W contacts appear bright, as they are of higher average Z than silicon; this meets the projection requirement. (b) and (c) As the projected thickness increases the contrast reverses, so that by 74° the contacts appear dark and the silicon bright; this image no longer meets the projection requirement. Rafal Dunin-Borkowski is acknowledged for the specimen and preparation.

scan allows the dissipation of heat (phonons) into surrounding, nonilluminated areas.[70]

6.6.7 Core-Loss (Chemical Mapping) EFTEM

There are several methods of producing element-specific maps in the TEM/STEM and of these perhaps the most promising for tomographic applications is Energy-Filtered Transmission Electron Microscopy (EFTEM). In the biological sciences energy filtering is used routinely for tomography, to generate a ''zero-loss'' image. In this case, the energy filter is being used to exclude all electrons that have undergone significant inelastic scattering, minimising chromatic blurring and improving the image contrast, particularly for thick specimens.[148] However, by making use of characteristic ionisation edges at well-defined energy losses a 2D elemental distribution image can be formed in a relatively short space of time (5–60 s). By using a tilt series of such images it is possible to reconstruct a three-dimensional elemental distribution map. This technique was first demonstrated in 2001 by Mobus and Inkson[18] and, independently, by Weyland and Midgley.[149] The first demonstrated applications in materials science were in engineering ceramics, where interphases form nanoscale precipitates and in stainless steels, to uncover the shape of chromium carbides at grain boundaries in stainless steels. A simpler form of inelastic tomography, making use of improved postedge contrast from a single energy loss rather than generating true elemental maps, has also been demonstrated by Yamauchi *et al.*[150] In this example, different phases in a complex polymer have been reconstructed with the aid of staining and energy-loss filtering. A similar approach is also used in biological TEM by placing the slit just before the carbon edge, known as ''structure-sensitive imaging'', maximising the contrast from species other than carbon.[151] Further applications of EFTEM tomography have been used to study iron/oxygen partitioning in magnetotactic bacteria[17] and to resolve internal chemical features in FeNi nanoparticles.[133] Despite the promising results EFTEM tomography has yet to receive as much attention or research effort as other new tomography techniques such as STEM HAADF tomography. This may be due to the combination of two technically demanding techniques and the significant limitations discussed below.

6.6.7.1 Meeting the Projection Requirement

EFTEM is an extended form of Electron Energy-Loss Spectroscopy (EELS) in which images are formed using electrons of a certain energy loss ''filtered'' by an energy-selecting window. By acquiring several energy-filtered images, elemental distribution maps can be formed. A schematic of the technique is shown in Figure 6.41. The ideal way to form elemental maps is to acquire an extended energy-loss series[152] (image spectra) over the energy-loss edges of interest. By using this approach[153,154] fully quantitative elemental distribution maps can be generated from any spectral feature within the acquired energy range.

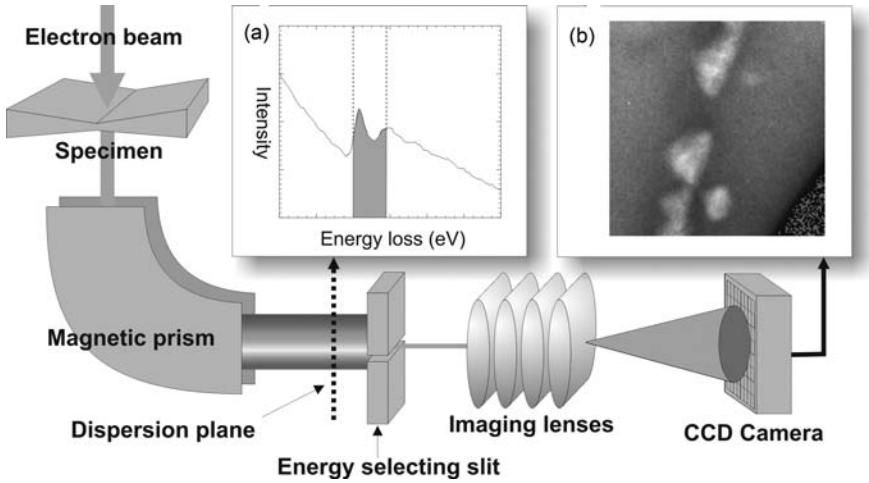


Figure 6.41 Schematic of EFTEM using a post-column energy filter. After passing through the specimen the electron beam is dispersed in energy using a magnetic prism to generate an electron energy-loss spectrum (a). Part of the spectra (shaded) is selected using a slit of finite width and an image reformed using an array of images. This image is recorded on a slow-scan CCD camera to generate an energy-loss image (b).

Tomographic reconstruction from a tilt series of energy loss images would generate a large four dimensional dataset (x, y, z, ΔE). However the acquisition time of a core-loss energy-loss series, typically tens of minutes, means that a tomography experiment would take a prohibitively long time, and result in a very high total dose. Consequently all demonstrated core-loss tomography has been based on simple three-image, elemental map, or two-image, jump-ratio, methods with each map taking only a few minutes to acquire. However, novel 4D volume spectroscopy has recently been demonstrated for low-loss energy losses.[155]

An energy filter will collect only inelastically scattered electrons of a certain energy loss defined by the position and the width of the energy-selecting window. However, the distribution of electrons that scatter inelastically is modified by elastic scattering events occurring within the same specimen area.[156] Therefore, the contrast observed in an energy-loss image is in general derived from a combination of inelastic scattering (through changes in composition and electronic structure) and elastic effects (via crystal thickness and orientation). The compositional information from a single energy-loss image may be isolated by generating either a background-subtracted elemental map (from three or more images) or a jump-ratio map (from two images). Taking due account of inelastic cross sections, both maps will show intensity that is related to the amount of an atomic species at a given pixel and to a first approximation EFTEM maps fulfil the projection requirement for tomographic reconstruction.

However this approximation is somewhat simplistic. For elemental maps, the conventional three-window approach takes little account of diffraction contrast that may effect the validity of the projection assumption, as it would for a BF image. This effect can be clearly seen in Figure 6.42 by the comparing (a) the zero-loss (filtered BF) image of a grain boundary in a stainless steel with b) the Cr and c) the Fe elemental maps. The effect is especially noticeable in the Fe elemental map, which shows diffraction features, particularly in the grain-boundary carbides, comparable to those in the zero-loss (BF) image. Such effects can be partially removed by dividing the elemental map by the zero-loss image[157] but this can introduce further artefacts due to changes in image resolution as a function of energy loss.[158] Jump-ratio images are an ideal way of removing residual diffraction contrast in the energy loss images, as shown in Figure 6.42(d) and (e). However, these images, although showing apparently higher sensitivity than that of an elemental map, have intensity values that are not quantitative. The jump ratio also increases with thickness, in an approximately linear fashion up to a maximum thickness comparable with the overall inelastic mean free path, λ. Beyond this thickness, the jump-ratio falls, primarily because the energy-loss background intensity rises faster than the ionisation-edge signal. Therefore, within the field of view of a single projection, thicker areas may give rise to a lower jump-ratio value. The effects of thickness and changes in composition on the jump-ratio across the image will no longer be monotonic, which invalidates the projection requirement. The degree of deviation from the projection assumption will vary from specimen to specimen, but for some specimens neither an elemental map nor a jump-ratio image will show suitable contrast for tomography.

Tomographic reconstructions carried out using both regular EFTEM and hollow-cone illumination[124] suggest that diffraction contrast is not an

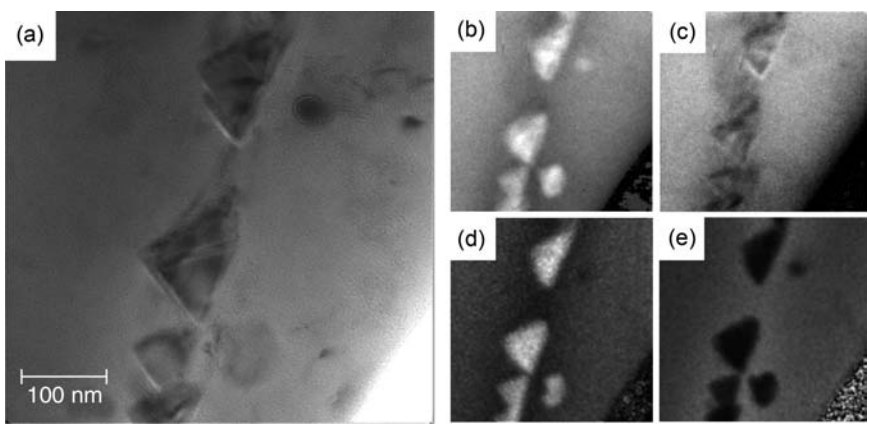


Figure 6.42 Comparison of contrast differences between imaging modes for a grain boundary in a 316 stainless steel specimen. (a) Zero-loss (BF) image, (b) Cr and (c) Fe elemental maps, showing residual diffraction contrast. (d) Cr and (e) Fe jump-ratio images, showing removal of diffraction effects and enhanced contrast.

insurmoutable barrier to EFTEM tomography. Its effects tend to be averaged out over the tilt series and results in an increase in background intensity in the reconstruction rather than introducing new features. However, an increase in projected thickness with tilt may lead to a deviation from monotonic behaviour and as a rule of thumb, the projected thickness across the tilt range should never exceed the inelastic mean free path, λ. For a typical value of λ, (100–150 nm) this implies a slab-like specimen should be prepared with a thickness of 40–50 nm in order to avoid multiple scattering artefacts at high tilts.

6.6.7.2 Experimental Considerations

Filtered images are acquired using an in-column[159] or post-column[160] imaging filter by using an energy-selecting slit, of a given energy width, in the dispersion plane. This slit is aligned on the zero-loss peak, to accurately set the zero energy loss reference, and the effective energy position of the slit is controlled by applying an energy offset, usually to the microscope HT. This offset can drift because of microscope instabilities or environmental factors, and as such its alignment needs to be periodically checked during the tilt series. At each tilt a number of energy-loss images, usually two/three at each energy loss edge, need to be acquired. This can be carried out manually, but it is more convenient to automate it using a simple script. This becomes especially important if a large number of different elements are studied. At each energy loss the positions for the energy windows will closely control the SNR of the resultant maps,[153] and as the quality of the reconstruction will depend on the quality of the projections, these should be chosen with care.

The resolution of EFTEM is controlled by a number of factors: delocalisation, chromatic aberration, spherical aberration and the diffraction limit.[161,162] In most cases the dominant term is the chromatic aberration of the objective lens. This leads to a blurring of the filtered image that can be minimised by reducing the slit width or the collection angle (typically the objective aperture) but always at the expense of total signal. A balance must be struck to achieve the optimum 2D EFTEM image and thus in turn the optimum 3D resolution.

While the optimal focus for BF images can be measured using the shift-with-tilt method this will not have the same value for energy-loss images. As inelastically scattered electrons have lost energy the postfield of the objective lens has a different focusing action than for the unscattered beam. As such there is a fixed offset between zero-loss and energy-loss images. If there is sufficient signal, the correct focus for filtered images is best set by focusing live at the energy loss of interest. If not, a compromise can be made by focusing at an intermediate energy loss.

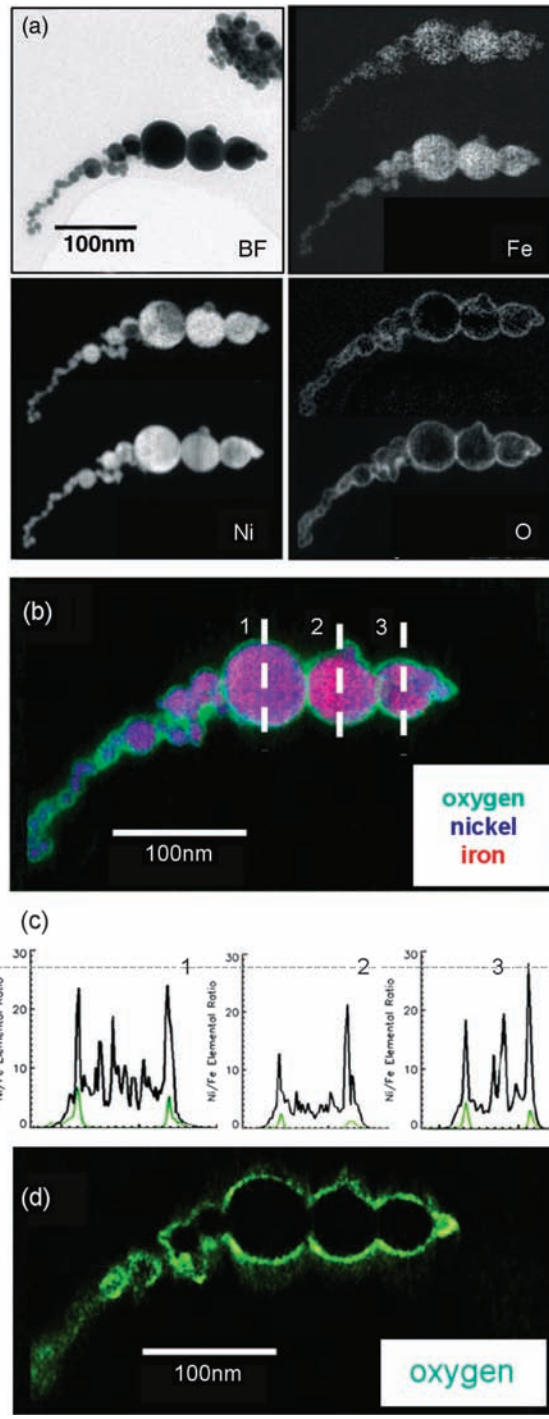
6.6.7.3 Application Example: FeNi Particles

An example of the unique information multiple element analysis in three dimensions can provide is given by the study of cryogenically evaporated iron-nickel nanoparticles.[133] These spherical, individually superparamagnetic,

particles, with diameters ranging from 10–100 nm, can be grown with closely defined stoichiometry and have been characterised previously by conventional 2D EFTEM[163] and electron holography,[164] both techniques revealing an oxide shell surrounding the agglomerated particles. There were two outstanding questions that could not be solved by 2D analysis: what is the dominant metal in the oxide and whether the oxide coating forms before, or after, particle agglomeration.

Tilt series were acquired from a chain of particles, shown in Figure 6.43(a), with three energy-loss images acquired at each of the Fe, Ni and O edges. The tilt series was acquired from -76° to 64° with a 4° increment. The processing of such a large amount of data was made less laborious by applying corrections from the higher-SNR dataset, Ni, to the other two series. Average elemental maps, generated by summing the reconstructions along the *Z*-direction, from the three sampled elements are also shown in Figure 6.43(a). One benefit of tomographic reconstruction is the generation of an improved 2D elemental map compared with a single projection, shown above. This improvement in SNR, known as the *polytropic montage* effect,[8] arises from the repeated sampling of the same volume through the tilt series. The spherical nature of the particles complicates the measurement of accurate 2D profiles through the surface oxide, but with the 3D reconstruction it is a simple matter of selecting out a voxel ''core sample'' through a particle, as plotted in Figure 6.43(b). These clearly show that the Ni/Fe ratio is higher at the oxide shell, clear evidence for a Ni rich oxide. A simple 2D map/projection seems to show that oxide exists between the particles, suggesting oxide growth before agglomeration. However, a 2D slice through the centre of the O reconstruction, Figure 6.43(d), clearly shows that there is no oxide where the nanoparticles touch, therefore they agglomerate and sinter before oxidising. This study clearly demonstrates of the advantage of tomography over simple projection mapping for certain nanostructure geometries.

Figure 6.43 (a) Zero-loss bright-field image and two-dimensional elemental maps of nickel, iron and oxygen recorded from a chain of FeNi nanoparticles chosen for EFTEM tomography acquisition. The lower elemental maps are generated by reprojecting the tomographic reconstructions in the zero-tilt direction. The upper elemental maps are the equivalent raw conventional elemental maps. (b) RGB voxel (volume pixel) projection of the combined reconstructions of the FeNi nanoparticles. (c) Three ''core samples'' have been extracted from the central volume of the three largest particles (marked 1,2 and 3, respectively, in (b)). The ''core samples'' have been averaged in the two shorter dimensions to give average line traces of all three elements. Plotted for all three samples are the nickel/iron ratios (black) and the equivalent oxygen line traces (green). (d) A slice through the oxygen tomographic reconstruction showing how the oxide covers the ensemble of nanoparticles rather than individual nanoparticles.


(a)
100nm
BF
Fe
Ni
O
(b)
1
2
3
oxygen
nickel
iron
100nm
(c)
Ni/Fe Elemental Ratio
30
20
10
0
1
2
3
(d)
100nm
oxygen

6.6.7.4 *Limitations*

Whilst fulfilling the projection requirement is the primary challenge for EFTEM tomography the effects of high dose also offer a serious barrier to many specimens. As any physical change in the specimen whilst acquiring a tilt series will invalidate that series, beam damage is always a concern for electron tomography. Great effort has been expended to develop low-dose techniques for BF electron tomography. However, the rapid decrease in signal with increasing energy loss means that to achieve suitable count rates at core-loss edges within sensible acquisition times, it is usual to use a highly converged beam.[165] Radiation damage can be the fundamental limitation to chemical analysis using EELS/EFTEM even for 2D analysis.[161] This combination of factors means that EFTEM tomography can be difficult to apply for beam-sensitive specimens. Despite this, initial studies have been reported by Leapman *et al.*[166] successfully applying EFTEM tomography to map phosphorous in biological sections. This suggests that it is feasible, with appropriate care, to apply EFTEM tomography to many systems without undue beam damage. The most important step for increasing the application of the technique would appear to be the development of fully automated acquisition, which would allow the use of low-dose, or at least dose-managed, methodologies such as those commonly used with frozen hydrated biological specimens.

6.6.8 Low-Loss EFTEM

There are situations where conventional elemental mapping or STEM HAADF imaging cannot differentiate components of a structure. In such situations other parts of the energy-loss spectrum can be used to identify the components. A recent example of this is the tomographic study of a nanocomposite material based on multiwall carbon nanotubes sheathed in nylon.[155] In this case, each component is carbon-based and the only way to differentiate them was to use a ratio of two plasmon images, recorded at 22 eV (dominated by the bulk plasmon loss of the nylon) and 28 eV (dominated by the nanotube plasmon) energy loss. The ratio images satisfied the projection requirement and could be used as input to a tomographic reconstruction. The final reconstruction showed very clearly the nanotube and nylon but also revealed the presence of catalyst particles and voids, not seen by conventional 2D imaging. By recording an energy-loss series at every tilt angle, a series of energy-dependent tomograms can be reconstructed. This then allows subvolumes to be interrogated and energy-loss spectra to be extracted from individual voxels or groups of voxels. This form of ''volume-spectroscopy'' circumvents the problems of overlap in projection when recording spectra from complex 3D structures.

6.6.9 Energy Dispersive X-Ray (EDX) Mapping

An alternative, more widely used, approach to forming elemental maps is STEM Energy Dispersive X-ray (EDX) mapping. Whilst EELS measures the

primary inelastic scattering events, EDX measures the event indirectly by recording the X-rays that are generated when the electrons fill the core-hole generated by the inelastic scattering. As the transition energy is quantised, the X-ray spectrum shows a unique "fingerprint" for each elemental species. Therefore, for each STEM pixel a spectrum of the emitted X-rays is collected that can then be used to extract elemental maps of each element. As the intensity of each pixel in the elemental map is related directly to the amount of the species in projection EDX would initially seem ideal for electron tomography.

However, the technical challenges for EDX tomography are considerable. The major hurdle is that the position of the EDX detector within the polepiece gap is such that is subtends only a small solid angle for the collection of emitted X-rays. This has two serious consequences. The first is that the collection efficiency of EDX is very low; it can take ~ 1 hour to acquire a single "spectrum image" of only a modest size (256×256 pixels) with a reasonable SNR in the resultant elemental maps. In that time the specimen can drift and the sample will be subject to a large electron dose. For 2D mapping only a single map of the area of interest is required, so damage can often be tolerated, but with tomography requiring multiple images from the same (unchanged) volume, damage of any kind can make meaningful reconstruction difficult if not impossible. The second major hurdle is that the material surrounding the specimen, including the mounting grid and specimen holder, can absorb a large proportion of X-rays before they reach the detector. To minimise this absorption the specimen is usually tilted towards the detector. As tomography requires extensive tilting, including directions away from the detector, much of the acquisition will take place at angles that are far from optimal, and with conventional holder designs may block a large proportion of the emitted X-rays.

Such problems mean that only one study using EDX tomography has been published to date. The work of Mobus[21] attempted to overcome the difficulties of EDX tomography and demonstrated its application to examine the position and morphology of Y_2O_3 particles in a FeAl alloy. The principal innovation was the use of a low-profile specimen holder, in which one side of the holder, facing the EDX detector, was cut away. In addition the specimen was a self-supporting wedge, requiring no extra supporting material in the direction of the EDX detector. However, the study was compromised by the limited tilt available in the microscope, $\pm 50^\circ$, and the large tilt increment of 10°. Each elemental map was acquired in a relatively short time (20 min) and as such the noise is rather high. While a 3D reconstruction was achieved, the number of compromises required in the acquisition and spectral artefacts, such as secondary X-ray generation, led to reconstruction artefacts making results difficult to interpret.

With the current design of EDX systems in TEM the key problems for EDX tomography continue to be signal and geometry. The SNR achievable in EDX mapping is simply too low for rapid mapping and the small solid angle of most EDX detectors is both too small and too directional. The X-ray output can be

increased by working with a larger electron probe current, achievable by either increasing the STEM probe size or by the use of an aberration-corrected STEM. However, an increase in probe current will also lead to an increased likelihood of beam damage. A better solution to both problems would be to use a larger solid angle detector, or multiple detectors. Perhaps another direction has been demonstrated by Koguchi *et al*.[64] who have carried out 2D EDX on a ''needle'' specimen that allows full 360° in-holder tilt without shadowing. While this study specifically mentions the plausibility of 3D-EDX from such a system, no tomography has yet been demonstrated.

6.6.10 Holographic Tomography

Off-axis electron holography[167] can be used to reconstruct the phase of the electron wavefunction to reveal the presence of magnetic fields and electrostatic potentials. In recent years there has been some progress towards combining electron holography with electron tomography to enable 3D magnetic[168] and electric fields[169] to be reconstructed and visualised.[102] The projection requirement for tomography can be satisfied by the phase image if the contribution to any phase change by diffraction contrast is limited. Normally this means tilting away from a major zone axis – as is often the case in conventional electron holography – but in a tilt series, this means some images in the tilt series will have to be discarded. Recent work[170] has shown that the potential distribution in a device containing a p-n junction can be reconstructed in 3D, which revealed quite dramatically the influence on the device's electronic properties through surface modification by FIB-based sample-preparation methods.

6.7 New Techniques

Although there has been a rapid growth in the range of imaging modes demonstrated to be suitable for electron tomography of materials, there are new techniques, that show significant potential for expanding both the resolution and range of information attainable using electron tomography.

6.7.1 Electron Energy-Loss Spectroscopy (EELS) Spectrum Imaging

The use of energy-loss electrons for tomography has been demonstrated for both core-loss[17] and low-loss EFTEM.[155] The distinct disadvantage of EFTEM is that spectral information is sampled coarsely, well below the energy resolution of the instrument/spectrometer. It is recognised that direct recording of EELS yields a more quantitative chemical analysis than EFTEM and has the ability to study spectral fine structure that arises due to changes in electronic/optical properties and the nature of the specimen bonding.[156] The drawback with EELS, however, is that spatial information is slow to acquire. Spectrum

imaging[171] allows the acquisition of spectra on a 2D grid of image points. Unlike X-ray microanalysis, EELS mapping is not limited by the collection angle, as primary inelastic scattering is strongly forward peaked, but by the slow shutter/readout rates of early diode array/CCD cameras:[172] spectrum images can take $\sim$1 h to acquire. The arrival of modern CCD spectrometers, with rapid/multiple readouts, bins and/or continuous readouts, has vastly increased the collection efficiency of EELS, with spatially resolved mapping also being much faster.[173] As such, spatial information can be acquired at speeds that should make EELS tomography feasible. An EELS-based tomographic reconstruction would be 4-dimensional, with a loss spectrum (ΔE) associated with each voxel at position (x, y, z). A further advantage of EELS is that, unlike EFTEM, it is not limited by chromatic aberration, allowing higher-resolution 2D mapping and therefore the potential for higher-resolution 3D reconstructions.

Despite this potential, EELS tomography has yet to be demonstrated, the nearest being the study by Gass *et al.*[155] discussed previously who used an EFTEM image-spectroscopy[174] approach that is still limited in energy resolution by the use of a finite slit width. There are still some technical challenges associated with EELS tomography, the most pertinent being adequate distortion correction of the spectral information prior to reconstruction, but the promise of high spectral sensitivity in 3D suggests it is a case of when, rather than if, it will be applied.

6.7.2 Confocal STEM

One of the central limitations to STEM tomography is the limited depth of field associated with the nonparallel nature of the illumination. This restricts the depth of field for high-resolution ($\sim$1 nm) reconstructions to (100–200 nm^3) volumes. The convergence angle, and hence the depth of field, in a STEM is defined by the diameter of the probe-forming aperture. This can either be the objective aperture, in a dedicated STEM, or a condenser aperture, in a combined TEM-STEM instrument. However, the convergence angle is also a key parameter that defines the size of the probe that is formed on the specimen; and hence also defines the 2D resolution. If a small aperture is used the probe is diffraction limited and if a large aperture is used the spherical-aberration coefficient (C_3) of the objective lens limits the probe size.[175] The typical convergence angles used to determine the volume/resolution limit in tomography, around 10 mrad at 200 kV, are close to optimal for typical values of C_3 ($\sim$1.2 mm) in an analytical STEM.

In order to increase the spatial resolution of STEM images, without decreasing the polepiece gap to an impractical size, many researchers have pursued the design of spherical-aberration correction.[176,177] These correctors have been installed to correct the aberrations of existing instruments, or installed in brand new instruments specifically engineered to take advantage of the corrector optics.[178] The spatial resolution attained by these instruments

is now well below 1 Å.[179] These correctors also allow smaller probes to be formed by increasing the convergence angle to ~20 mrad [179] with current-corrected instruments and a prediction of 50 mrad [178] for a next-generation dedicated instrument. A side effect of this increased convergence angle is a loss of depth of field; this loss is sufficiently high that the resolution in the depth direction becomes significantly smaller than the thickness of the specimen under study. Recent research[180] demonstrates how this limited depth of field can actually be an advantage. As the STEM probe is localised in the specimen both in the depth direction as well as laterally, the image generated is effectively an "optical section" of the specimen, analogous to that used in confocal optical microscopy.[181] Indeed, scanning confocal electron microscope (SCEM), was first suggested, and demonstrated, in 2002 by Frigo *et al.*[182] However, their work concentrated on the ability of such a technique to produce images of extremely thick sections rather than 3D reconstructions, though this possibility was mooted. Van Benthem predicted that their aberration-corrected instrument should show depth resolution on the order of 6 nm. Whilst much worse than the 2D resolution of the instrument (~0.1 nm) this was sufficient to determine the depth positions of individual Hf atoms in a SiO_2 gate dielectric, as illustrated in Figure 6.44. The resolution in the depth direction appears to be better than that predicted, around 2 nm or so, but this is attributed to the loss of the vertical "tails" from the Hf atom in the background scattering intensity from the surrounding structures. This finding points to a potential limitation of the technique; this loss of information in the background intensity may make objects without the high contrast seen between Hf and SiO_2 difficult to detect in a defocus series. However, this blurring will decrease as the localisation is improved and a dedicated aberration-corrected STEM, with a larger convergence angle, may show significantly lower background signal.

The operation of functional nanodevices is often controlled by a small concentration of atoms at interfaces, and knowing their distribution in 3D may be key to describing the device's properties. The resolution attainable by defocus tomography is matched only by that of a three-dimensional atom probe (3D-AP),[183] but as the samples studied by this technique are limited, it may be that when defocus tomography is fully developed it becomes applicable to a much wider range of materials.

6.7.3 Atomistic Tomography

The ultimate resolution achievable with electron tomography should, theoretically, be close to the resolution achievable in 2D TEM/STEM. This is achievable only under the special conditions set for high-resolution lattice imaging; where a precisely oriented crystal may be imaged with atomic resolution (~1 Å). Reaching this resolution in 3D, however, is a significant challenge. Interest in achieving 3D lattice resolution is focused on structures that cannot be made sufficiently large for bulk techniques, such as X-ray diffraction to be applicable. Such specimens can have unknown crystal

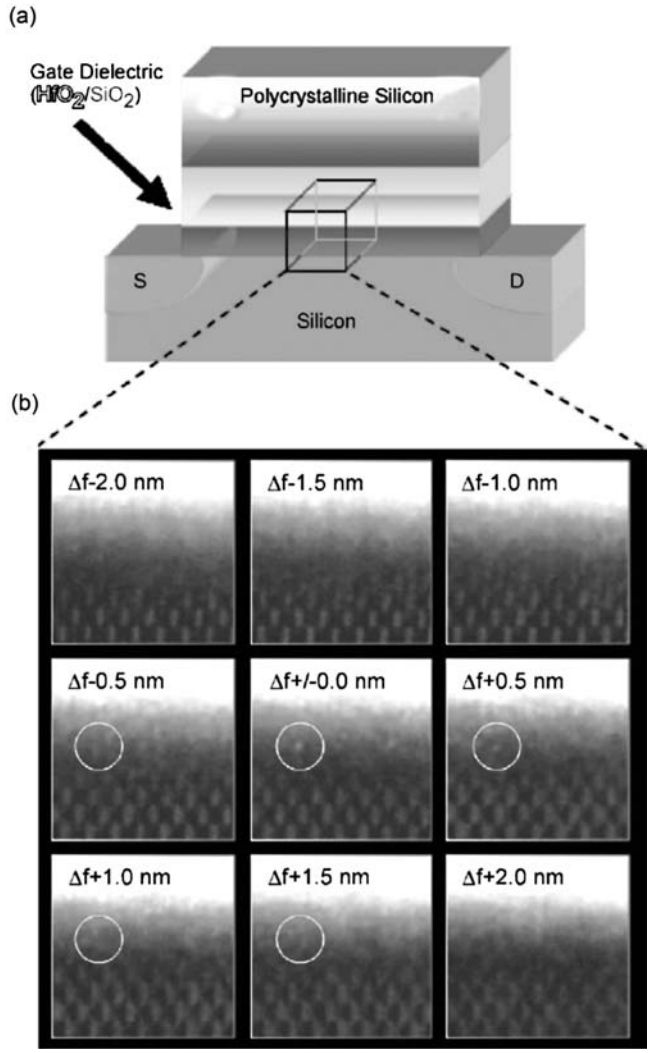


Figure 6.44 Depth localisation of individual Hf atoms in a 3 nm wide SiO_2 dielectric layer. (a) A schematic of the transistor containing the layer, where S is the source and D the drain, respectively. (b) Nine micrographs taken at different defocus levels highlighting the ability to localise a single Hf atom in the dielectric layer due to the limited depth-of-field of the aberration corrected probe. Reused with permission from Klaus van Benthem, *Appl. Phys. Lett.*, 2005, **87**, 034104. Copyright 2005, American Institute of Physics.

structures and morphology, for example buried quantum dots.[184] Whilst a unique structure, such as a pore network, requires information in all spatial dimensions in order to be described adequately, the symmetry of crystalline systems implies a redundancy of information; the only projections required are

those that reveal the directions of lattice repeat units. As such, true atomistic 3D reconstruction should be possible from a discrete number of projections if recorded at high-symmetry orientations. The theory of determining crystal structures in this way,[185] is essentially the same as the Fourier synthesis approach using X-ray diffraction.[27] There are two connected sets of information that can be determined, however, the first is the crystal structure of such nanoscale objects, the second is the spatial position of all the atoms in a unique object. While the former has been achieved, for example for tungsten carbide particles grown in glass by Qin and Fraundorf,[186] the latter has not been achieved to date. However, the information that such a reconstruction might provide, such as the atomic arrangement of a quantum-sized objects or the location of atoms at a structural defect or interface, would be of great value.

Achieving atomistic tomography requires overcoming some significant technical hurdles, the most challenging of which is the tracking of the individual structure while tilting the specimen to a suitable zone axis. This involves observing the diffraction pattern from an extremely small volume whilst tilting the specimen; with existing microscopes an extremely difficult task, particularly for isolated crystals. If, however, the nanocrystal is embedded in a parent crystal matrix, and assuming some kind of crystallographic relationship between the two crystals, the nanocrystal may be tilted onto axis by tracking the tilt of the bulk crystal. An approach to automating such a search, by calibration of the goniometer tilts with reference to the bulk crystal structure has been described by Fraundorf *et al.*[187] The object of interest is now tracked whilst tilting to the required angle without reference to a diffraction pattern during tilting. This study also raises the importance of spatial resolution to such an "autogoniometric" study; higher resolution, such as that offered by aberration correction, increases the number of lattice projections that can be imaged successfully hence decreasing the tilting required. Although this system has yet to be applied, it could offer a standard approach to acquiring tilt data in such cases. Another approach to reaching a larger number of key zone axes is to increase the movement range of the specimen holder/goniometer.[187] The recent availability of a high-tilt ($\pm 70^\circ$) holder also allowing 360° of rotation, designed initially for acquiring dual-axis tilt series, would seem ideal for reaching such axes. The second major limitation is achieving reconstruction from such tilt data. Although standard reconstruction methods, as used for conventional tomography can be used, new algorithms are required to take full advantage of the discrete nature of the data. In theory, constrained reconstruction techniques, such as SIRT, may produce a volume with plausible atomistic features but whether the results would be truly valid is debatable. A more intelligent approach may be to apply the mathematics of discrete systems, as described by Batenburg.[188] Here, atoms are treated as point objects and projections used to simply count the atoms in each column. Such an approach requires a very careful analysis of the number of projected atoms in each column, a significant task in itself.

Whilst the technical hurdles for atomistic reconstruction are significant, it is feasible that the first such reconstructions will appear within the next couple of

years. This is driven by zone-axis-tracking programs, the availability of new specimen holders, novel discrete reconstruction algorithms and the ever-increasing interest in complex nanoscale devices and structures that require 3D analysis for complete characterisation.

6.8 Conclusions

Over the past 5 years, electron tomography in the physical sciences has advanced from a niche offshoot of biological tomography into a burgeoning new field of materials characterisation. This rapid advancement has been made possible by the long history of research and innovation in structural biology combined with the kind of multitechnique approach typical in materials microscopy. There is now a large array of tomography options to determine the 3D structure and composition of materials at the nanoscale and a suitable technique is available for nearly any kind of material, structure and property. In addition, there are indications that alternative modes of tomography may find some application in the biosciences, for example the use of HAADF tomography by Zeise *et al.*[189] as an improved means to locate ultrasmall immuno-gold labels or the 3D mapping of phosphorous in ribosomes using EFTEM tomography.[166] Electron tomography seems set to be a central technique in the future of the biological and physical sciences as well as an area ripe for interdisciplinary research.

References

1. E. Meyer, H.J. Hug and R. Bennewitz, *Scanning Probe Microscopy: The Lab on a Tip*, Springer, Berlin, Hedielberg, New York, 2003.
2. J. Goldstein, ed. *Scanning Electron Microscopy and X-ray Microanalysis*, 3rd ed. 2003, Springer, 689.
3. D. Williams, B. and C.B. Carter, *Transmission electron microscopy: a textbook for materials science*, Plenum Press, New York, London, 1996.
4. Hounsfield, *Computed Medical Imaging*. 1979: Nobel Lecture.
5. A.M. Cormack, *J. Appl. Phys.*, 1963, **34**, 2722.
6. G.N. Hounsfield, *A method and apparatus for examination of a body by radiation such as X or gamma radiation*, in *The Patent Office, London*. 1972: England.
7. D.J. De Rosier and A. Klug, *Nature*, 1968, **217**, 130.
8. R.G. Hart, *Science*, 1968, **159**, 1464.
9. W. Hoppe, R. Langer, G. Knesch and C. Poppe, *Naturwissenschaften*, 1968, **55**, 333.
10. P.N.T. Unwin and R. Henderson, *J. Mol. Biol.*, 1975, **94**, 425.
11. W. Baumeister, R. Grimm and J. Walz, *Trends in Cell Biology*, 1999, **9**, 81.
12. R.J. Spontak, M.C. Williams and D.A. Agard, *Polymer*, 1988, **29**, 387.

13. D.A. Hajduk, P.E. Harper, S.M. Gruner, C.C. Honeker, E.L. Thomas and L.J. Fetters, *Macromolecules*, 1995, **28**, 2570.
14. J.H. Laurer, D.A. Hajduk, J.C. Fung, J.W. Sedat, S.D. Smith, S.M. Gruner, D.A. Agard and R.J. Spontak, *Macromolecules*, 1997, **30**, 3938.
15. A.J. Koster, U. Ziese, A.J. Verkleij, A.H. Janssen and K.P. de Jong, *J. Phys. Chem. B*, 2000, **104**, 9368.
16. R.J. Spontak, J.C. Fung, M.B. Braunfeld, J.W. Sedat, D.A. Agard, A. Ashraf and S.D. Smith, *Macromolecules*, 1996, **29**, 2850.
17. M. Weyland and P.A. Midgley, *Microscopy and Microanalysis*, 2003, **9**, 542.
18. G. Mobus and B.J. Inkson, *Appl. Phys. Lett.*, 2001, **79**, 1369.
19. H. Stegmann, H.H. Engelmann and E. Zschech, *Microelectron. Eng.*, 2003, **65**, 171.
20. P.A. Midgley and M. Weyland, *Ultramicroscopy*, 2003, **96**, 413.
21. G. Mobus, R.C. Doole and B.J. Inkson, *Ultramicroscopy*, 2003, **96**, 433.
22. J. Radon, *Ber. Verh. K. Sachs. Ges. Wiss. Leipzig, Math.-Phys. Kl.*, 1917, **69**, 262.
23. S.R. Deans, *The Radon transform and some of its applications*, Wiley, New York, Chichester, 1983.
24. G.T. Herman, *Image Reconstruction From Projections, The Fundamentals of Computerised Tomography.*, Academic Press, New York, 1980.
25. H. Cramer and H. Wold, *J. London Math. Soc.*, 1936, **11**, 290.
26. R.N. Bracewell, *Aust. J. Phys.*, 1956, **9**, 297.
27. M.F. Perutz, M.G. Rossmann, A.F. Cullis, H. Muirhead and A.C.T. North, *Nature*, 1960, **185**, 416.
28. R.M. Glaeser, *Annual Review of Physical Chemistry*, 1985, **36**, 243.
29. P.F.C. Gilbert, *Proc. R. Soc. Lond. B.*, 1972, **182**, 89.
30. R.A. Crowther, D.J. de Rosier and A. Klug, *Proc. Roy. Soc. Lond. A.*, 1970, **317**, 319.
31. B.K. Vainshtein, *Soviet Physics-Crystallography*, 1970, **15**, 781.
32. J. Carazo, in *Electron tomography: three-dimensional imaging with the transmission electron microscope*, ed. J. Frank, Plenum Press, New York, London, 1992.
33. P. Gilbert, *J. Theor. Biol.*, 1972, **36**, 105.
34. J. Frank, *Three-Dimensional Electron Microscopy of Macromolecular Assemblies*, Academic Press, San Diego, 1996.
35. U. Skoglund and L. Ofverstedt, *Journal of Structural Biology*, 1996, **117**, 173.
36. R. Gordon, R. Bender and G.T. Herman, *J. Theor. Biol.*, 1970, **29**, 471.
37. R.A. Crowther and A. Klug, *J. Theor. Biol.*, 1971, **32**, 199.
38. S.H. Bellman, R. Bender, R. Gordon and J.E. Rowe, *J. Theor. Biol.*, 1971, **32**, 205.
39. R. Marabini, E. Rietzel, R. Schroeder, G.T. Herman and J.M. Carazo, *Journal of Structural Biology*, 1997, **120**, 363.
40. M.I. Sezan, *Ultramicroscopy*, 1992, **40**, 55.

41. M. Radermacher, T. Wagenknecht, A. Verschoor and J. Frank, *Journal of Microscopy*, 1987, **146**, 112.
42. C.R. Smith and W.T. Grandy Jr., *Maximum-Entropy and Bayesian methods in inverse problems*, Dordrecht, Holland, 1985.
43. J.J. Hu and F.H. Li, *Ultramicroscopy*, 1991, **35**, 339.
44. M.C. Lawrence, M.A. Jaffer and B.T. Sewell, *Ultramicroscopy*, 1989, **31**, 285.
45. C.E. Shannon, *Proceedings of The Institute of Radio Engineers*, 1949, **37**, 10.
46. M. Radermacher and W. Hopee. proceedings of *7th European Congr. Electron Microscopy*. Den Haag. 1980.
47. M. Radermacher, in *Electron tomography: three-dimensional imaging with the transmission electron microscope*, ed. J. Frank, Plenum Press, New York, London, 1992, 91.
48. J.W. Strutt (Rayleigh), *Philosophical Magazine*, 1879, **7**, 261.
49. B.F. McEwen, M. Marko, C.E. Hsieh and C. Mannella, *Journal of Structural Biology*, 2002, **138**, 47.
50. M. Van Heel and G. Harauz, *Optik*, 1986, **73**, 119.
51. J. Frank, A. Verschoor and M. Boublik, *Science*, 1981, **214**, 1353.
52. M. Unser, B.L. Trus and A.C. Steven, *Ultramicroscopy*, 1987, **23**, 39.
53. R. Henderson, J.M. Baldwin, T.A. Ceska, E. Beckman, F. Zemlin and K. Downing, *J. Mol. Biol.*, 1990, **213**, 899.
54. P.A. Penczek, *Journal of Structural Biology*, 2002, **138**, 34.
55. B. Bottcher, S.A. Wynne and R.A. Crowther, *Nature*, 1997, **386**, 88.
56. P.W. Hawkes, in *Electron tomography: three-dimensional imaging with the transmission electron microscope*, ed. J. Frank, Plenum Press, New York, London, 1992, 17.
57. B. Sander, M.M. Golas and H. Stark, *Journal of Structural Biology*, 2003, **142**, 392.
58. J.J. Fernandez, S. Li and R.A. Crowther, *Ultramicroscopy*, 2006, **106**, 587.
59. J.A. Velazquez-Muriel, C.O.S. Sorzano, J.J. Fernandez and J.M. Carazo, *Ultramicroscopy*, 2003, **96**, 17.
60. C. Kubel, A. Voigt, R. Schoenmakers, M. Otten, D. Su, T.C. Lee, A. Carlsson and J. Bradley, *Microscopy and Microanalysis*, 2005, **11**, 378.
61. J. Mardinly, *proceedings of Microscopy and Microanalysis*, Long Beach, California, 2001.
62. H. Friedrich, M.R. McCartney and P.R. Buseck, *Ultramicroscopy*, 2005, **106**, 18.
63. M. Weyland, L. Laffont and P.A. Midgley, *Inst. Phys. Conf. Ser.*, 2004, **179(EMAG 2003)**, 349.
64. M. Koguchi, H. Kakibayashi, R. Tsuneta, M. Yamaoka, T. Niino, N. Tanaka, K. Kase and M. Iwaki, *Journal of Electron Microscopy*, 2001, **50**, 235.
65. J. Szot, R. Hornsey, T. Ohnishi and S. Minagawa, *Journal of Vacuum Science & Technology B*, 1992, **10**, 575.

66. T. Ishitani and T. Yaguchi, *Microscopy Research and Technique*, 1996, **35**, 320.
67. C. Burkhardt and W. Nisch, *Praktische Metallographie-Practical Metallography*, 2005, **42**, 161.
68. R. Hegerl and W. Hoppe, *Zeitschrift Fur Naturforschung Section A-A Journal of Physical Sciences*, 1976, **31**, 1717.
69. R. Grimm, H. Singh, R. Rachel, D. Typke, W. Zillig and W. Baumeister, *Biophys. J.*, 1998, **74**, 1031.
70. R.F. Egerton, P. Li and M. Malac, *Micron*, 2004, **35**, 399.
71. L.W. Hobbs, in *Introduction to Analytical Electron Microscopy*, ed. J.J. Hren, J.I. Goldstein and D.C. Joy, Plenum, New York, 1979.
72. T.C. Isabell, P.E. Fischione, C. O'Keefe, M.U. Guruz and V.P. Dravid, *Microscopy and Microanalysis*, 1999, **5**, 126.
73. A.J. Koster, R. Grimm, D. Typke, R. Hegrel, A. Stoschek, J. Walz and W. Baumeister, *Journal of Structural Biology*, 1997, **120**, 276.
74. K. Dierksen, D. Typke, R. Hegerl and W. Baumeister, *Ultramicroscopy*, 1993, **49**, 109.
75. K. Dierksen, D. Typke, R. Hegerl, A.J. Koster and W. Baumeister, *Ultramicroscopy*, 1992, **40**, 71.
76. A.J. Koster, A. Van Den Bos and K.D. Van Der Mast, *Ultramicroscopy*, 1987, **21**, 209.
77. U. Ziese, A.H. Janssen, J.L. Murk, W.J.C. Geerts, T. Van der Krift, A.J. Verkleij and A.J. Koster, *Journal of Microscopy*, 2002, **205**, 187.
78. D.N. Mastronarde, *Journal of Structural Biology*, 2005, **152**, 36.
79. Q.X.S. Zheng, M.B. Braunfeld, J.W. Sedat and D.A. Agard, *Journal of Structural Biology*, 2004, **147**, 91.
80. U. Ziese, A.H. Janssen, J.L. Murk, W.J.C. Geerts, T. Van der Krift, A.J. Verkleij and A.J. Koster, *Journal of Microscopy-Oxford*, 2002, **205**, 187.
81. Fei, *Xplore3D*. 2001: Eindhoven (The Netherlands).
82. M.C. Lawrence, in *Electron tomography: three-dimensional imaging with the transmission electron microscope*, ed. J. Frank, Plenum Press, New York, London, 1992, 197.
83. J. Frank and B.F. McEwen, in *Electron tomography: three-dimensional imaging with the transmission electron microscope*, ed. J. Frank, Plenum Press, New York, London, 1992, 205.
84. D.E. Olins, A.L. Olins, H.A. Levy, R.C. Durfee, S.M. Margle, E.P. Tinnel and S.D. Dover, *Science*, 1983, **220**, 498.
85. M.C. Lawrence, *Proceedings of the Electron Microscopy Society of South Africa*, 1983, **13**, 19.
86. J. Berriman, R.K. Bryan, R. Freeman and K.R. Leonard, *Ultramicroscopy*, 1984, **13**, 351.
87. S. Nickell, F. Forster, A. Linaroudis, W. Del Net, F. Beek, R. Hegerl, W. Baumeister and J.M. Plitzko, *Journal Of Structural Biology*, 2005, **149**, 227.
88. D. Ress, M.L. Harlow, M. Schwarz, R.M. Marshall and U.J. McMahan, *Journal of Electron Microscopy*, 1999, **48**, 277.

89. J.C. Fung, W.P. Liu, W.J. deRuijter, H. Chen, C.K. Abbey, J.W. Sedat and D.A. Agard, *Journal of Structural Biology*, 1996, **116**, 181.
90. S. Brandt, J. Heikkonen and P. Engelhardt, *Journal of Structural Biology*, 2001, **133**, 10.
91. J.C. Russ, *The Image Processing Handbook*, CRC Press, London, 1995.
92. R. Guckenberger, *Ultramicroscopy*, 1982, **9**, 167.
93. W. Hoppe, H.J. Schramm, M. Sturm, N. Hunsmann and J. Gassman, *Zeitschrift Fur Naturforschung Section A-A Journal of Physical Sciences*, 1976, **31**, 645.
94. J. Frank, in *Computer processing of electron microscope images*, ed. P.W. Hawkes, Springer-Verlag, Berlin, New York, 1981, 187.
95. R.A. Crowther, *Philos. Trans. R. Soc., London B.*, 1971, **261**, 221.
96. Y. Liu, P.A. Penczek, B.F. McEwen and J. Frank, *Ultramicroscopy*, 1995, **58**, 393.
97. C. Renken and B. McEwen, *Microscopy and Microanalysis*, 2003, **9**, 1170.
98. J. Dengler, *Ultramicroscopy*, 1989, **30**, 337.
99. H. Winkler and K.A. Taylor, *Ultramicroscopy*, 2006, **106**, 240.
100. S. Brandt, J. Heikkonen and P. Engelhardt, *Journal of Structural Biology*, 2001, **136**, 201.
101. J.C. Russ, *The Image Processing Handbook*, CRC Press, London, 2002.
102. V. Stolojan, R.E. Dunin-Borkowski, M. Weyland and P.A. Midgley, proceedings of *Electron Microscopy and Analysis 2001*, 2001.
103. A.S. Frangakis and R. Hegerl, *Journal of Structural Biology*, 2002, **138**, 105.
104. N. Volkmann, *Journal of Structural Biology*, 2002, **138**, 123.
105. C. Bajaj, Z.Y. Yu and M. Auer, *Journal of Structural Biology*, 2003, **144**, 132.
106. C.Y. Xu and J.L. Prince, *Ieee Transactions on Image Processing*, 1998, **7**, 359.
107. A.S. Frangakis and F. Forster, *Current Opinion in Structural Biology*, 2004, **14**, 325.
108. V. Lucic, T. Yang, G. Schweikert, F. Forster and W. Baumeister, *Structure*, 2005, **13**, 423.
109. B.K. Rath, R. Hegerl, A. Leith, T.R. Shaikh, T. Wagenknecht and J. Frank, *Journal of Structural Biology*, 2003, **144**, 95.
110. A.S. Frangakis, J. Bohm, F. Forster, S. Nickell, D. Nicastro, D. Typke, R. Hegerl and W. Baumeister, *Proc. Natl. Acad. Sci. USA.*, 2002, **99**, 14153.
111. H. Jinnai, Y. Nishikawa, R.J. Spontak, S.D. Smith, D.A. Agard and T. Hashimoto, *Phys. Rev. Lett.*, 2000, **84**, 518.
112. Y. Ikeda, A. Katoh, J. Shimanuki and S. Kohjiya, *Macromol. Rapid Commun.*, 2004, **25**, 1186.
113. S. Kohjiya, A. Katoh, J. Shimanuki, T. Hasegawa and Y. Ikeda, *Journal of Materials Science*, 2005, **40**, 2553.
114. P.A. Midgley, J.M. Thomas, L. Laffont, M. Weyland, R. Raja, B.F.G. Johnson and T. Khimyak, *J. Phys. Chem. B*, 2004, **108**, 4590.

115. A. Al-Amoudi, L.P.O. Norlen and J. Dubochet, *Journal of Structural Biology*, 2004, **148**, 131.
116. J. Frank, *Electron Tomography: Three-dimensional imaging with the Transmission Electron Microscope*, Plenum Press, New York, London, 1992.
117. V. Lucic, F. Forster and W. Baumeister, *Annual Review of Biochemistry*, 2005, **74**, 833.
118. L.H. Radzilowski, B.O. Carragher and S.I. Stupp, *Macromolecules*, 1997, **30**, 2110.
119. H. Sugimori, T. Nishi and H. Jinnai, *Macromolecules*, 2005, **38**, 10226.
120. U. Ziese, K.P. de Jong and A.J. Koster, *Appl. Catal., A.*, 2004, **260**, 71.
121. T. Kizuka, T. Sumi, S. Bansho and T. Saito, *Japanese Journal of Applied Physics Part 2-Letters*, 2001, **40**, L56.
122. M. Numata, M. Asai, K. Kaneko, A.H. Bae, T. Hasegawa, K. Sakurai and S. Shinkai, *Journal of The American Chemical Society*, 2005, **127**, 5875.
123. K. Inoke, K. Kaneko, M. Weyland, P.A. Midgley, K. Higashida and Z. Horita, *Acta Materialia*, 2006, **54**, 2957.
124. M. Weyland and P.A. Midgley, *Microscopy and Microanalysis*, 2003, **9**, 148.
125. W.O. Saxton, W.K. Jenkins, L.A. Freeman and D.J. Smith, *Optik*, 1978, **49**, 505.
126. E.L. Thomas, D.M. Anderson, C.S. Henkee and D. Hoffman, *Nature*, 1988, **334**, 598.
127. S.M. Gruner, *J. Phys. Chem.*, 1989, **93**, 7562.
128. J.M. Thomas, P.A. Midgley, T.J.V. Yates, J.S. Barnard, R. Raja, I. Arslan and M. Weyland, *Angew. Chem., Int. Ed.*, 2004, **43**, 6745.
129. K. Kimura, S. Hata, S. Matsumura and T. Horiuchi, *Journal of Electron Microscopy*, 2005, **54**, 373.
130. J.S. Barnard, J. Sharp, J.R. Tong and P.A. Midgley, *Science*, 2006, **313**, 319.
131. P.A. Midgley, M. Weyland, J.M. Thomas and B.F.G. Johnson, *Chem. Commun.*, 2001, 907.
132. P.R. Buseck, R.E. Dunin-Borkowski, B. Devouard, R.B. Frankel, M.R. McCartney, P.A. Midgley, M. Posfai and M. Weyland, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 13490.
133. M. Weyland, T.J.V. Yates, R.E. Dunin-Borkowski, L. Laffont and P.A. Midgley, *Scripta Materialia*, 2006, **55**, 29.
134. I. Arslan, T.J.V. Yates, N.D. Browning and P.A. Midgley, *Science*, 2005, **309**, 2195.
135. T.J.V. Yates, *The development of electron tomography for nanoscale materials science applications*, PhD Thesis, *Department of Materials Science and Metallurgy*, 2005, University of Cambridge.
136. J.P. Barnes, N. Beer, A.K. Petford-Long, A. Suarez-Garcia, R. Serna, D. Hole, M. Weyland and P.A. Midgley, *Nanotechnology*, 2005, **16**, 718.

137. M. Isaacson, M. Ohtsiki and M. Utlaut, in *Introduction to Analytical Electron Microscopy*, ed. J.J. Hren, J. Goldstein and D.C. Joy, Plenum Press, 1979, 343.
138. A. Howie, *Journal of Microscopy*, 1979, **117**, 11.
139. P.B. Hirsch, A. Howie, R.B. Nicholson, D.W. Pashley and M.J. Whelan, *Electron Microscopy of Thin Crystals*, Krieger, New York, 1977.
140. S.J. Pennycook and P.D. Nellist, in *Impact of Electron and Scanning Probe Microscopy on Materials Research*, ed. D.G. Rickerby, Kluwer, New York, 1999, 161.
141. P. Rez, *Ultramicroscopy*, 2003, **96**, 117.
142. R. Blakemore, *Science*, 1975, **190**, 377.
143. S. Mann, N.H.C. Sparks, R.B. Frankel, D.A. Bazylinski and H.W. Jannasch, *Nature*, 1990, **343**, 258.
144. V.P. Shcherbakov and M. Winklhofer, *Eur. Biophys. J.*, 1999, **28**, 380.
145. J. Dobson and P. Schultheiss-Grassi, *J. Biochem. Mol. Biol. & Biophys.*, 1998, **1**, 247.
146. D.S. McKay, E.K. Gibson, K.L. Thomas-Keptra, H. Vali, C.S. Romanek, S.J. Clemett, X.D.F. Chillier, C.R. Maechling and R.N. Zare, *Science*, 1996, **273**, 924.
147. K.L. Thomas-Kerpta, D.A. Bazylinski, J.L. Kirschvink, S.J. Clemett, D.S. McKay, S.J. Wentworth, H. Vali, E.K. Gibson and C.S. Romanek, *Geochim. Cosmochim. Acta.*, 2000, **64**, 4049.
148. R. Grimm, D. Typke and W. Baumeister, *Journal of Microscopy*, 1998, **190**, 339.
149. M. Weyland and P.A. Midgley, *Inst. Phys. Conf. Ser.*, 2001, **161(EMAG 2001)**, 239.
150. K. Yamauchi, K. Takahashi, H. Hasegawa, H. Iatrou, N. Hadjichristidis, T. Kaneko, Y. Nishikawa, H. Jinnai, T. Matsui, H. Nishioka, M. Shimizu and H. Fukukawa, *Macromolecules*, 2003, **36**, 6962.
151. L. Reimer and M. Rossmessemer, *Journal Of Microscopy-Oxford*, 1990, **159**, 143.
152. J.L. Lavergne, J.M. Martin and M. Belin, *Microscopy Microanalysis Microstructures*, 1992, **3**, 517.
153. P.J. Thomas and P.A. Midgley, *Ultramicroscopy*, 2001, **88**, 179.
154. P.J. Thomas and P.A. Midgley, *Ultramicroscopy*, 2001, **88**, 187.
155. M.H. Gass, K.K.K. Koziol, A.H. Windle and P.A. Midgley, *Nano Letters*, 2006, **6**, 376.
156. R.F. Egerton, *Electron energy-loss spectroscopy in the electron microscope*, Plenum Press, New York, London, 1996.
157. P.A. Crozier, *Ultramicroscopy*, 1995, **58**, 157.
158. F. Hofer, W. Grogger, G. Kothleitner and P. Warbichler, *Ultramicroscopy*, 1997, **67**, 83.
159. J. Mayer. Proceedings of *50th Annual Meeting of the Electron Microscopy Society of America*, 1992.

160. O.L. Krivanek, A.J. Gubbens, M.K. Kundmannn and G.C. Carpenter, *proceedings of Proc. 51st Annual Meeting of the Microscopy Society of America*, 1993.
161. R.F. Egerton, *Journal of Electron Microscopy*, 1999, **48**, 711.
162. O.L. Krivanek, M.K. Kundmann and K. Kimto, *Journal of Microscopy*, 1995, **180**, 277.
163. R.K.K. Chong, R.E. Dunin-Borkowski, T. Kasama, M.J. Hytch and M.R. McCartney, *Proceedings of Electron Microscopy and Analysis 2003*, 2003.
164. M.J. Hytch, R.E. Dunin-Borkowski, M.R. Scheinfein, J. Moulin, C. Duhamel, F. Mazaleyrat and Y. Champion, *Phys. Rev. Lett.*, 2003, **91**, 257207.
165. L. Reimer, in *Energy-Filtering Transmission Electron Microscopy*, ed. L. Reimer, Springer-Verlag, Berlin, 1995.
166. R.D. Leapman, E. Kocsis, G. Zhang, T.L. Talbot and P. Laquerriere, *Ultramicroscopy*, 2004, **100**, 115.
167. P.A. Midgley, *Micron*, 2001, **32**, 167.
168. G.M. Lai, T. Hirayama, K. Ishizuka, T. Tanji and A. Tonomura, *Applied Optics*, 1994, **33**, 829.
169. G.M. Lai, T. Hirayama, A. Fukuhara, K. Ishizuka, T. Tanji and A. Tonomura, *J. Appl. Phys.*, 1994, **75**, 4593.
170. A. Twitchett, R.E. Dunin-Borkowski, T.J.V. Yates and P.A. Midgley, *unpublished.*
171. R.D. Leapman, C.E. Fiori, K.E. Gorlen, C.C. Gibson and C.R. Swyt, *Ultramicroscopy*, 1984, **12**, 281.
172. R.F. Egerton, Y.Y. Yang and S.C. Cheng, *Ultramicroscopy*, 1993, **48**, 239.
173. Z. Tang, R. Ho, Z. Xu, Z. Shao and A.P. Somlyo, *Journal Of Microscopy-Oxford*, 1994, **175**, 100.
174. J.L. Lavergne, C. Foa, P. Bongrand, D. Seux and J.M. Martin, *Journal of Microscopy-Oxford*, 1994, **174**, 195.
175. J.R. Michael and D.B. Williams, *Journal of Microscopy-Oxford*, 1987, **147**, 289.
176. M. Haider, H. Rose, S. Uhlemann, E. Schwan, B. Kabius and K. Urban, *Ultramicroscopy*, 1998, **75**, 53.
177. O. Krivanek, N. Dellby and A. Lupini, *Ultramicroscopy*, 1999, **78**, 1.
178. O.L. Krivanek, P.D. Nellist, N. Dellby, M.F. Murfitt and Z. Szilagyi, *Ultramicroscopy*, 2003, **96**, 229.
179. P.E. Batson, *Ultramicroscopy*, 2003, **96**, 239.
180. K. van Benthem, A.R. Lupini, M. Kim, H.S. Baik, S. Doh, J.H. Lee, M.P. Oxley, S.D. Findlay, L.J. Allen, J.T. Luck and S.J. Pennycook, *Appl. Phys. Lett.*, 2005, **87**.
181. P.C. Cheng, ed., *Multidimensional microscopy*, 1994, Springer-Verlag, New York. 387.
182. S.P. Frigo, Z.H. Levine and N.J. Zaluzec, *Appl. Phys. Lett.*, 2002, **81**, 2112.

183. D. Blavette, A. Bostel, J.M. Sarrau, B. Deconihout and A. Menand, *Nature*, 1993, **363**, 432.
184. A.P. Alivisatos, *J. Phys. Chem.*, 1996, **100**, 13226.
185. P. Fraundorf, *Ultramicroscopy*, 1987, **22**, 225.
186. W. Qin and P. Fraundorf, *Ultramicroscopy*, 2003, **94**, 245.
187. P. Fraundorf, W.T. Qin, P. Moeck and E. Mandell, *J. Appl. Phys.*, 2005, **98**.
188. K.J. Batenburg, *Electronic Notes in Discrete Mathematics*, 2005, **20**, 247.
189. U. Ziese, C. Kubel, A.J. Verkleij and A.J. Koster, *Journal Of Structural Biology*, 2002, **138**, 58.
190. L.J. Wu, P. Lo, X.K. Yu, J.K. Stoops, B. Forghani and Z.H. Zhou, *J. Virol.*, 2000, **74**, 9646.
191. O. Medalia, I. Weber, A.S. Frangakis, D. Nicastro, G. Gerisch and W. Baumeister, *Science*, 2002, **298**, 1209.
192. D.G. Bucknall and H.L. Anderson, *Science*, 2003, **302**, 1904.
193. S. Forster and T. Plantenberg, *Angewandte Chemie-International Edition*, 2002, **41**, 689.

CHAPTER 7

In-situ Environmental Transmission Electron Microscopy

P. L. GAI[1,2,3]

[1] *Department of Chemistry and Department of Physics, The York–JEOL Nanocentre, University of York, Heslington, York, YO10 5DD, UK*
[2] *Central Research and Development Laboratories, DuPont, Experimental Station, Wilmington DE 19880-0356, USA*
[3] *University of Delaware, Department of Materials Science and Engineering, Newark, DE 19716, USA*

7.1 Introduction

The dynamic surface atomic structure of solid-state catalysts under operating conditions plays a crucial role in heterogeneous catalysis.[1–4] Electron microscopy is a very powerful research tool for investigating the local structure and chemistry of complex heterogeneous catalytic materials directly, from the macroscopic to the atomic scale. Conventional High-Resolution Transmission Electron Microscopy (HRTEM) operates in high vacuum at ambient temperature and uses thin samples including as-synthesised catalyst powders. Two or more Bragg reflections are used for imaging. In HRTEM, very thin samples can often be treated as weak-phase objects where the image intensity can be correlated to the projected electrostatic potential, leading to the atomic-structural information. In addition, the detection of electron-stimulated X-ray emission in the Microscope (Energy Dispersive X-Ray Spectroscopy, EDX) and Electron Energy Loss Spectroscopy (EELS) permit the simultaneous determination of chemical compositions of catalysts at the subnanometer level.

However, heterogeneous catalysis is a dynamic process with reactions taking place in gas (or liquid) high temperature environments at the catalyst surface at the atomic level. The direct *in-situ* observation of the catalyst nanostructural

evolution under dynamic reaction environments is therefore a very powerful scientific tool in the chemical sciences. *In-situ* Environmental (Scanning) Transmission Electron Microscopy (E(S)TEM, of gas–solid catalyst reactions under controlled reaction environments provides real-time dynamic information about reaction processes on the solid surface, which can not be obtained readily by other methods. *In-situ* ETEM under controlled operating conditions has been demonstrated to be of vital importance in studies of heterogeneous catalysts used in, or proposed for, commercial reactions, which may include environmental control, and the structure–property relationships associated with them.[1–4] Probing catalysts directly under reaction environments is key to gaining insights into catalytic reaction mechanisms, active sites, the nature of defect structural evolution and the associated chemical bonding, which are crucial to the development of new systems and processes. The unique contribution of ETEM lies in its ability to characterise dynamic catalyst local structures, which may be aperioidic and determine the nature, composition and adjacency of particles of different types. Systematic associations or repulsions of catalyst particles of different size and shape, as well as the response of nanosystems such as carbon nanotubes (CNTs) to reaction environments can also be examined.

7.2 Background

A number of important *in-situ* experiments reported in the literature have relied upon modifications to conventional TEM operation. The main electron optical functions of the TEM, especially the electron gun, depend on a high vacuum environment, which is of the order of 10^{-6}–10^{-7} mbar. With a gas reaction cell or environmental cell (ECELL) as a microreactor inside the microscope column, controlled, chemically reducing/oxidising atmospheres or solvent-rich media can be maintained and a wide range of gases and vapours can be used. These developments are described in the following sections.

Some of the previous *in-situ* experiments carried out in high vacuum in the TEM produced important results especially in probing dislocation dynamics.[5] The value of using a hot stage to study *in-situ* crystallisation[6] and the growth of clusters and ceramics at very high temperatures ($\sim$1500 °C)[7] have been demonstrated in a high-vacuum environment of a high-resolution TEM. Incorporation of a genuine ultrahigh vacuum (UHV) range suitable for surface science studies[8] has also been developed into an important scientific tool.[9]

However, in heterogeneous catalysis the role of environment on a sample is critical.[1–4,10] The high-vacuum environment of a conventional microscope is not suitable for studying many catalytic reactions. Post-reaction studies of static catalysts (*i.e.* catalysts cooled to room temperature and removed from the reaction environment) are often not representative of the dynamic catalyst. Early *in-situ* ETEM experiments are well documented in the literature.[11–13] Window cells have been used to contain gases, solvent vapours and hydrated samples,[14,15] with the sample sandwiched between electron transparent

windows to contain the environment for the reaction. However, window cells present problems in reliably sustaining a large pressure difference across a window thin enough to permit electron transmission. Window cells are generally not compatible with heating systems and furthermore, have limited image contrast. Recent advances in the investigation of liquid–solid reactions in the ETEM at the subnanometer level are described in the following sections.

The complications and potential for failure of windows can be overcome by substituting small apertures above and below the sample to restrict the diffusion of gas molecules, and to allow the transmission of the electron beam. This is accomplished by adding pairs of apertures above and below the sample with differential pumping lines between them. Most previous ECELL systems have used older,[13,16] or larger[12,17,18] electron microscopes. Furthermore, many of these ECELL systems were made to be interchangeable, necessitating the frequent rebuilding of the microscope to effect the changeover in functionality. To maintain reliability, ECELL systems were designed that could be inserted between the polepieces of the image-forming objective lens of the electron microscope. In the early *in-situ* experimentation, ECELL systems could be inserted inside the EM column vacuum between the objective lens polepieces.[12] The gas reaction chamber and hermetically sealed objective aperture assembly occupied the gap between the upper and lower objective polepieces, leading to a gas reservoir around the sample. Such ECELL systems were a major step forward in scientific capability and important work was carried out with them.[1,12,13,16,19–33] Early studies with limited spatial resolution have included reactions of metal particle catalysts at low gas pressures of a few mbar,[16] as realistic gas pressures and temperatures were not possible.

The development of *in-situ* ETEM has played a key role in the studies of dynamic catalysis.[1] Gai *et al.*[1,9–31] developed *in-situ* High-Voltage EM (HVEM) operating at 1 MV to meet the demand of realistic high gas pressures and temperatures (up to 1000 °C) for catalysis, and performed the first *in-situ* studies of selective hydrocarbon oxidation reactions over metal oxides under high gas pressures ($\sim$1 atm) at operating temperatures.[23–27] These studies provided insights into the fundamental role of defects in selective oxidation catalysis.[1] With this system, image resolutions of $\sim$1–2 nm at high gas pressures ($>$100 mbar) and 0.5 nm at lower gas pressures of $<$30 mbar, were obtained at catalyst operating temperatures. HVEM development thus laid the basis for the future development of atomic-resolution *in-situ* ETEM.[34–39]

7.3 Recent Advances in Atomic-Resolution *In-Situ* ETEM

The quest to probe gas molecule–solid catalyst reactions directly at the atomic level has resulted in the development of an atomic-resolution *in-situ* ETEM at

DuPont.[34,35–38] This demonstrated for the first time that probing dynamic gas–solid (catalyst) reactions at the atomic level was possible under controlled reaction environments consisting of gas pressure of a few mbar and elevated temperatures.[40] In this development a new approach was taken to design an ETEM instrument that is dedicated to ECELL operations and the ECELL was permanently mounted and integrated with HRTEM. Firstly, it was based on a modern Philips CM30T (S)TEM system with high resolution crystal lattice imaging performance. Secondly, the whole microscope column, including the region surrounding the sample, was redesigned for ECELL functionality. Thirdly, a set of polepieces incorporating radial holes for the first stage of differential pumping, with no deleterious effect on imaging, was designed for the instrument (Figures 7.1(a) and (b)).

In this instrument,[35,37] the ECELL, HRTEM, STEM, hot-stage and parallel EELS/Gatan imaging filter (GIF) functionalities have been combined in a single instrument. The combination is required to aid the simultaneous determination of dynamic nanostructure and composition of the reactor contents. The E(S)TEM is used as a *nanolaboratory* with multiprobe measurements and the design of novel reactions and nanosynthesis is thus possible. The structure and chemistry of dynamic catalysts are revealed by atomic imaging, electron diffraction and chemical analysis while the sample is immersed in controlled gas atmospheres and is at the operating temperature. For chemical microanalyses, a commercial Gatan PEELS/GIF system is fitted to the ETEM, to provide elemental analysis during *in-situ* experiments with low gas pressures[41] and fast minimally invasive high-resolution chemical mapping with filtered TEM images. The analysis of oxidation state in intermediate phases of the reaction and in principle, EXELFS studies are possible. In many applications, the size and subsurface location of particles require the use of the dynamic ESTEM system integrated with ETEM,[35] with complementary methods for chemical and crystallographic analyses.

The basic geometry of the novel atomic-resolution ETEM design[35–38] is a four-aperture system, located in pairs above and below the sample (indicated in Figure 7.1, but the apertures are now mounted inside the bores of the objective-lens polepieces, rather than between them as in previous designs. The controlled environment ECELL (microreactor) volume is the regular sample chamber of the microscope. Differential pumping systems connected between the apertures are achieved using molecular drag pumps and turbomolecular pumps. This permits relatively high gas pressures in the sample region while maintaining high vacuum in the rest of the ETEM. A conventional reactor-type gas-manifold system enables the inlet of flowing gases into the specimen region, and a hot stage allows samples to be heated. A mass spectrometer is also included for gas analysis. For dynamic atomic resolution a few millibars of gas pressure are used in the ECELL. Higher gas pressures (up to a substantial fraction of 1 atm) are possible but this compromises the resolution (due to multiple scattering effects of the electron beam through thicker gas layers). A video system facilitates digital image processing and real-time recording of dynamic events. Because of the small amounts of solid reactant in the

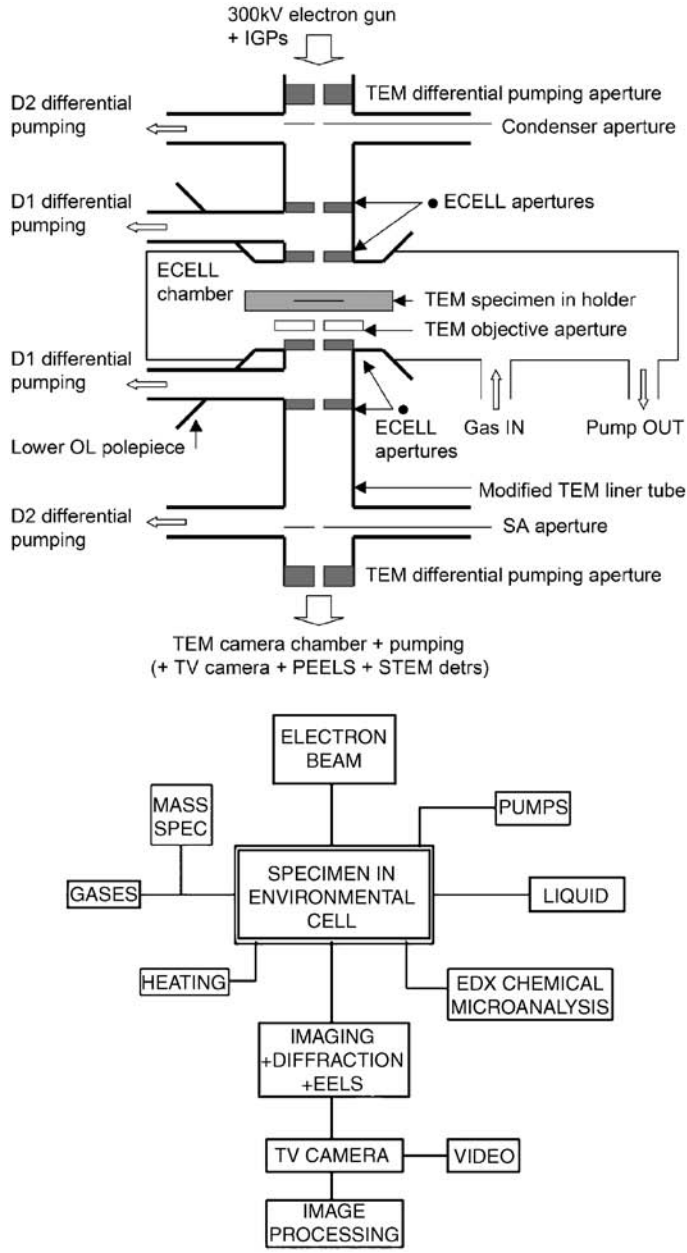


Figure 7.1 Schematic representations of the basic ETEM geometry showing radial holes *through* objective lens (OL) polepieces and the aperture system.[34–38] The OL polepieces are above and below the specimen holder and the lower OL polepiece is indicated. Gas inlet, the first stage differential pumping lines (D1) between the environmental cell (ECELL) apertures, condenser aperture, second stage of pumping (D2) at the condenser lens, selected area (SA) diffraction aperture and TEM camera vacuum are also indicated.

microscope sample, parallel measurements of reaction products are performed on larger samples in an *ex-situ* microreactor operating under similar working conditions and are used for the nanostructure–surface reactivity correlation.

Electronic image shift and drift compensation help to stabilise high-resolution images for data recording on film or with real-time digitally processed video; and minimally invasive electron-beam techniques and calibration studies are used throughout in a contamination-free environment. The impact of the charged-particle electron beam on the chemical processes must always be minimised or eliminated by the use of appropriate very low electron dose techniques (with doses well below the threshold for damage). These procedures also reduce any beam-induced local temperature rise. The effects of the conditions are checked in a parallel "blank" calibration experiment, with the beam switched off for the *in-situ* reaction and the sample exposed to the beam only to record the reaction endpoint.[1] Real-time *in-situ* studies are then confirmed by comparison with data from calibration experiments. The aim is completely non invasive characterisation under benign conditions. Under carefully simulated conditions close to those in practical reactors, data from *in-situ* ETEM studies can be directly correlated to nanostructure–activity relationships in technological processes, resulting in considerable cost and time savings. In addition, several conditions are required for successful studies with ETEM; thin, electron-transparent samples are necessary and most catalyst powders meet this requirement; ultrahigh-purity heater materials and sample grids capable of withstanding elevated temperature and gases are also required.

In the atomic-resolution *in-situ* ETEM the alignment and atomic resolution (0.2 nm) performance of the microscope have been maintained with the ECELL facilities, with sample temperatures above 700 °C, and with modest (mbar) amounts of gas flowing through the ECELL.[34–39] The relatively large apertures in the cell provide useful angles of electron diffraction and some Convergent-Beam Electron Diffraction pattern (CBED) analysis capability with a dynamic STEM probe. Smaller objective apertures can be used inside the ECELL for diffraction-contrast experiments to determine the nature of defects critical to catalysis.

7.4 Impact of Atomic-Resolution *In-Situ* ETEM and Applications

The design of the atomic-resolution *in-situ* ETEM providing controlled reaction conditions[34–37] has been adopted by commercial TEM manufacturers (such as FEI) for production. Later versions of this *in-situ* ETEM instrument have been installed in laboratories around the world and atomic-scale ETEM studies[34–39] are being performed and replicated by other researchers. Studies carried out with the commercial instruments include reactions of promoted Ru

catalysts in different gas environments[42,43] and Ziegler–Natta catalysts, supported nickel catalysts and redox processes in ceria and ceria-zirconia.[44–48]

7.5 Applications of Atomic-Resolution *In-situ* ETEM to Studies of Gas–Catalyst and Liquid–Catalyst Reactions

7.5.1 Liquid-Phase Hydrogenation and Polymerisation Reactions

Technological processes for the hydrogenation and polymerisation of complex organic molecules are primarily derived from solutions at low temperatures and the associated chemical reactions occur on the nanoscale. Recent advances in biological catalysis and molecular electronics also require studies in liquid environments. Probing reactions in liquids is therefore of great importance in the development of advanced catalytic and nanotechnologies.[49–51] I review the progress of ETEM in wet environments (*i.e.* wet ETEM) in the following sections.

As described in the previous sections, ''window cells'' where a sample is sandwiched between electron transparent windows, have been used to study hydrated or biological samples.[14,15] Several research groups in Japan have been active in the area of wet ETEM. Nagata and Ishikawa[52] used liquid injection directly into an ECELL to study biological samples. Fukushima *et al.*[53] have developed an elegant flow route to inject liquid from outside the electron microscope column into the ECELL of ETEM. These authors use top-entry ECELL systems and employ a liquid line with flowmeters and pressure gauges to inject liquids directly into the ECELL to study liquid–solid reactions. Daulton *et al.*[54] have used a commercial JEOL wet ETEM with window cells for studies of biological samples. However, in all these developments, liquid–solid reactions have been studied with limited resolution at room temperature, with nonheating and nontilting sample stages and the results tend to be averaged over large areas. Recently, Gai[55] has developed a heating and tilting liquid sample stage to study low vapour pressure liquid–gas–solid catalyst, as well as liquid–solid reactions at the sub-nm level in wet ETEM under controlled conditions.

Alkylated amines are key organic intermediates in the manufacture of linear polyamides for fibre, plastics and mechanical industries. Liquid-phase hydrogenation of aliphatic dinitriles at low temperatures ($\sim 100\,^{\circ}C$) is used in chemical technology to produce the corresponding diamines. These are subsequently reacted with adipic acid in the liquid phase and polymerised to produce linear polyamides. Of particular commercial importance is the liquid-phase hydrogenation of adiponitrile (ADN, $NC(CH_2)_4CN$) to produce hexamethylene diamine (HMD, $H_2N(CH_2)_6NH_2$), an important intermediate in the manufacture of the polyamide.[56] Conventional hydrogenation is carried out with the dinitrile in a solvent such as methanol, under gaseous hydrogen over

Raney nickel (or cobalt) complexes formed by an Ni-Al alloy from which Al is dissolved in alkaline solutions leaving a hydrogenated surface.[56] However, the recovery of the fragile catalysts and the disposal of the alkaline solution waste can be difficult and residual amounts of Al can have a deleterious effect on the catalyst selectivity.

7.5.2 Development of Nanocatalysts for Novel Hydrogenation Chemistry and Dynamic Imaging of Desorbed Organic Products in Liquid-Phase Reactions

Using wet ETEM, Gai *et al.*[55] have developed novel hydrogenation chemistry using a novel concept of dispersing promoted ruthenium nanocatalysts on *nanosupports* (instead of conventional bulk supports). The nanocatalysts derived from this concept are catalytically stable and active. Desorbed organic products in the dynamic liquid-phase hydrogenation have been imaged on the subnanometer scale for the first time.[57]

Cobalt-promoted Ru nanocatalysts on nanotitania supports show catalyst nanoparticles of ~ 1 nm dispersed in size on well-ordered nanotitania particles of 10–20 nm.[57] Nanoprobe EDX shows the nanoparticles to be distinct. Nitrogen pore-size distributions (Figures 7.2(a) and (b)) show that pores are generally mesoporous with an average pore diameter of about 5–6 nm. Nano-Co-Ru/nanotitania mesoporous nanosystems ((Figure 7.3(a), at room temperature) are immersed in ADN in methanol liquid (and 0.75 wt% NaOH solvent) using the liquid holder[55] for the liquid-phase hydrogenation. Nano/microliters of the liquid are injected (using procedures similar to those employed in chromatography methods) over the catalysts, and the catalyst is immersed in flowing liquids. Flowing hydrogen gas is passed over the samples simultaneously and the sample is heated to $\sim 100\,^{\circ}$C. *In-situ* wet ETEM studies reveal the formation of HMD (Figure 7.3(b)). Rapid growth of the reaction product (~ 0.25 nm s^{-1}) and its clean desorption observed at $100\,^{\circ}$C indicate a highly selective catalyst.[57] Atomic-resolution imaging after the hydrogenation of the sample shows anion vacancy defects in nanotitania (Figure 7.3(c)) following anion loss and the formation of defects. The glide shear mechanism preserves anion vacancy sites associated with Lewis acid sites.[34,37] Mass spectrometry of the desorbed product confirms the presence of HMD (Figure 7.4) with the primary peak at ~ 117 (HMD + H). Parallel reactivity tests (inset in Figure 7.4) have confirmed the ETEM experiments (Figure 7.3) and show very high hydrogenation activity for the nano-Co-Ru/nanotitania catalyst system. The tests also show that the promoted system is more selective to HMD than the unpromoted nanosystem.

Simultaneously, *in-situ* wet imaging at the subnanometer scale and mass spectrometery have revealed the presence of bis-hexamethylene triamene ($NH_2(CH_2)_6NH$ $(CH_2)_6NH_2$, or BHMT) on the catalyst surface during the liquid-phase hydrogenation.[57] BHMT is an important organic product used in

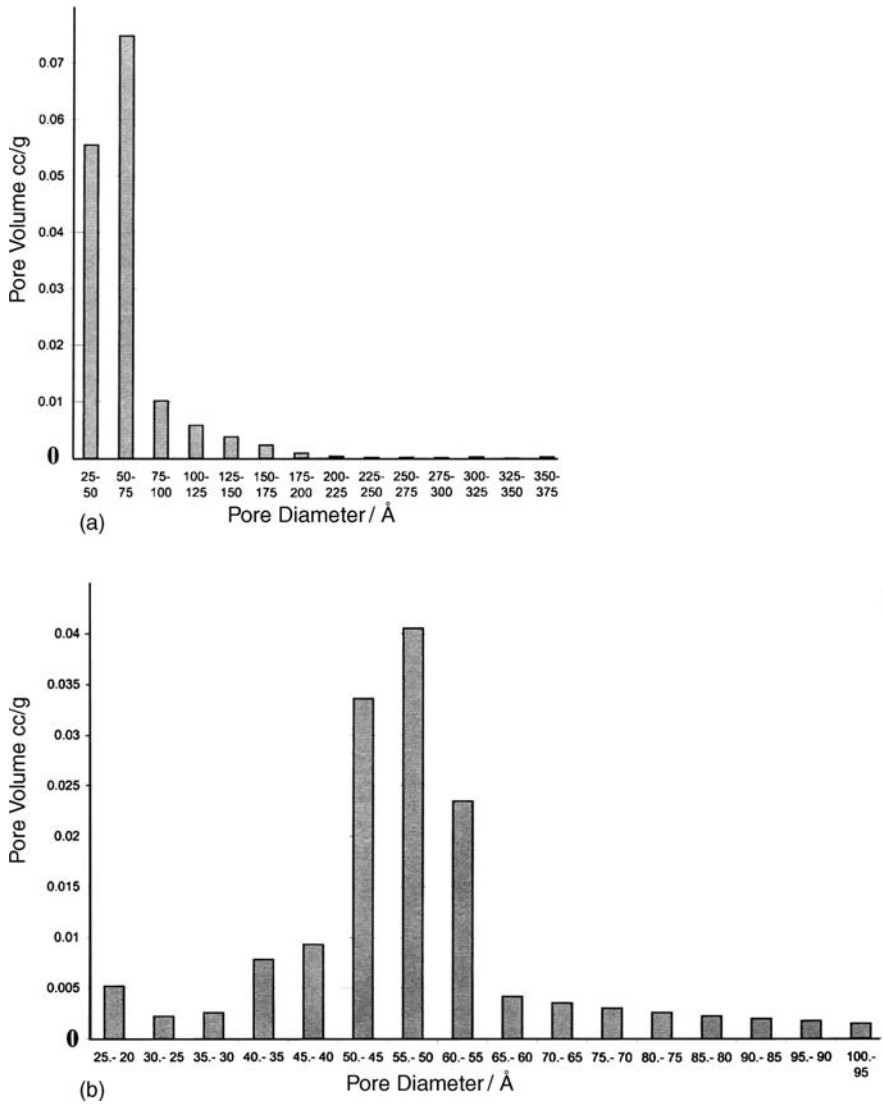


Figure 7.2 N_2 pore-size distribution of a ($Co_{0.1}Ru_{0.05}$) nanocatalyst on nanotitania: (a) Pore volume ($cm^3\ g^{-1}$), V, *vs.* pore diameter (A) in Å. The horizontal scale extends from 25–50 Å to 350–375 Å. (b) Enlargement of (a), with average pore diameters of ~5–6 nm. The horizontal scale extends from 25–20 Å to 100–95 Å.

coatings, sealants, epoxies and polyamides. Figures 7.5(a) and (b) illustrate images of the wet catalyst at room temperature and at ~100 °C, respectively. Figure 7.5(c) shows an enlargement of the BHMT nanolayers (at H) shown in (b). They indicate the hitherto unknown lattice structure of BHMT with a periodicity of ~0.56 nm. *In-situ* studies have been important in elucidating the

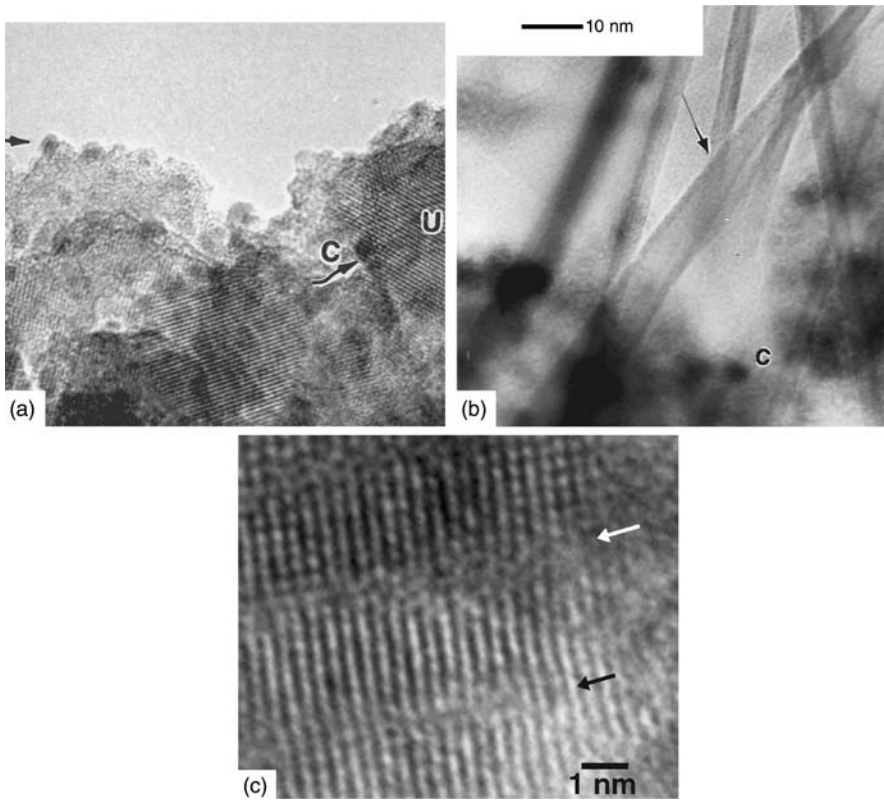


Figure 7.3 *In-situ* imaging of gas–liquid–catalyst reactions and dynamic imaging of desorbed organic products (a) Dry nano-Co-Ru/nanotitania in vacuum at room temperature. The nanocatalysts are indicated by C and the arrow, and the well-ordered nanosubstrate lattice, by U. (b) Wet sample: *In-situ* hydrogenation of adiponitrile (ADN) in the liquid medium over the catalyst surface at nominal 100 °C showing desorption of nanolayer product (arrowed) (temperature gradient may be present in the liquid). Long fibres are observed after 2 h. (c) Atomic-resolution image of anion vacancy nanodefects in nanotitania (arrowed) in dry sample following hydrogenation.

nucleation of BHMT and its lattice, as upon exposure to air it can interact with CO_2 and form compounds such as carbamates.

These direct *in-situ* studies demonstrate high hydrogenation activity in the presence of anion vacancy defects associated with Lewis acid sites at the nanosupport surface and an electronic and synergistic contribution to the promoter mechanism. Based on the ETEM data, it is proposed that the combined synergistic effect between the two nanometals (Co and Ru) and the interaction with the reduced nanosupport containing anion vacancies leading to an electronic modification yield highly reactive site for the hydrogenation catalysis.[57] These key wet-ETEM findings have led to

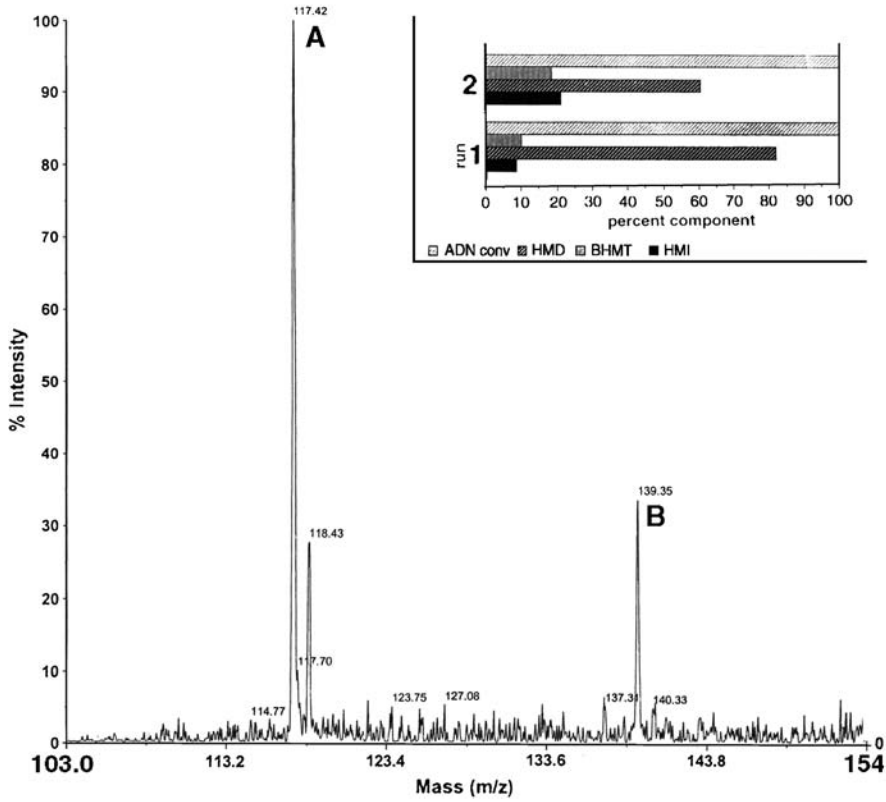


Figure 7.4 Mass spectrum (% intensity *vs.* mass (m/z)) of the desorbed product, confirming hexamethylene diamine (HMD): primary HMD peak at $\sim$117 (HMD + H) at A. Inset shows parallel reactivity tests showing high hydrogenation activity for Co-Ru nanocatalysts over nanotitania (bar 1) confirming the ETEM experiments (Figure 7.3). The promoted catalysts are also more selective to than unpromoted Ru nanocatalysts (bar 2).

low-temperature heterogeneous routes for the hydrogenation of ADN and the polymerisation of HMD in the liquid phase and indicate that the mesoporous systems are capable of high hydrogenation activity.

7.5.3 Butane Oxidation Technology

The selective catalysation of n-butane to maleic anhydride (MA) over vanadium phosphorus oxides is a very important commercial process.[58] MA is subsequently used in catalytic processes to make polyurethane intermediates, fibres and agricultural products. The active phase in the selective 14-electron oxidation of butane has been identified as vanadyl pyrophosphate $(VO)_2P_2O_7$ (VPO.)[58]

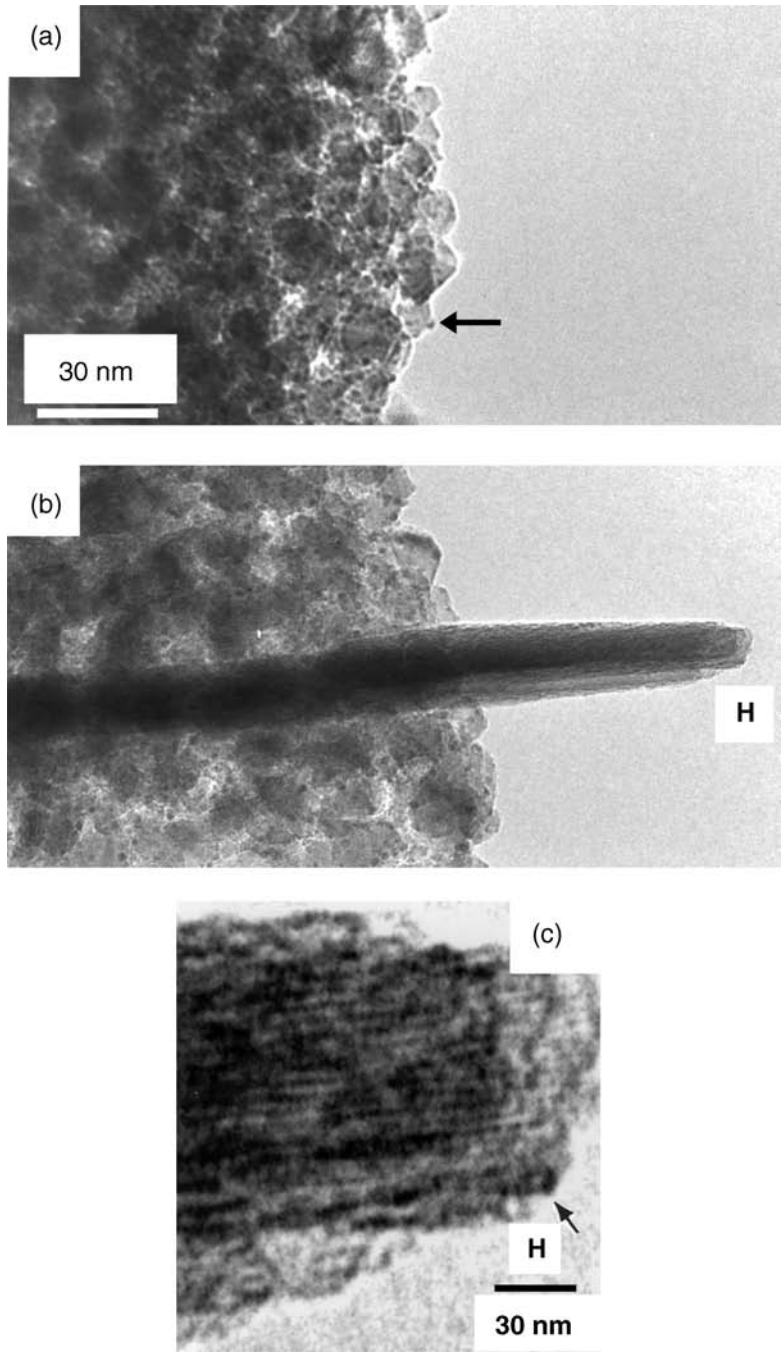


Figure 7.5 (a) Wet nano-Co-Ru/nanotitania in ADN liquid and H_2 gas at room temperature. (b) *in-situ* growth of BHMT lattice (*e.g.* at H) at the wet catalyst surface at $\sim 100\,^{\circ}C$. (area from (a)). Particles remain stable. (c) Enlargement of wet sample in (b) indicating ~ 0.56 nm periodicity.

The 3-dimensional structure of orthorhombic VPO consisting of vanadyl octahedra and phosphate tetrahedra, is shown in Figure 7.6, with a = 1.6594 nm, b = 0.776 nm and c = 0.958 nm.[58]

In-situ ETEM has achieved the formidable task of unraveling the atomic structures of active sites and the mechanism for the release of catalyst structural oxygen and the accommodation of anion deficiency in butane oxidation technology.[34,37] The *in-situ* results of calcined and activated (010) VPO catalyst in 20% butane /He (5 mbar) are summarised in Figure 7.7. The structure image of VPO in n-butane /He at room temperature is shown in Figure 7.7(a) with the corresponding electron diffraction (ED) pattern inset. In butane reduction,

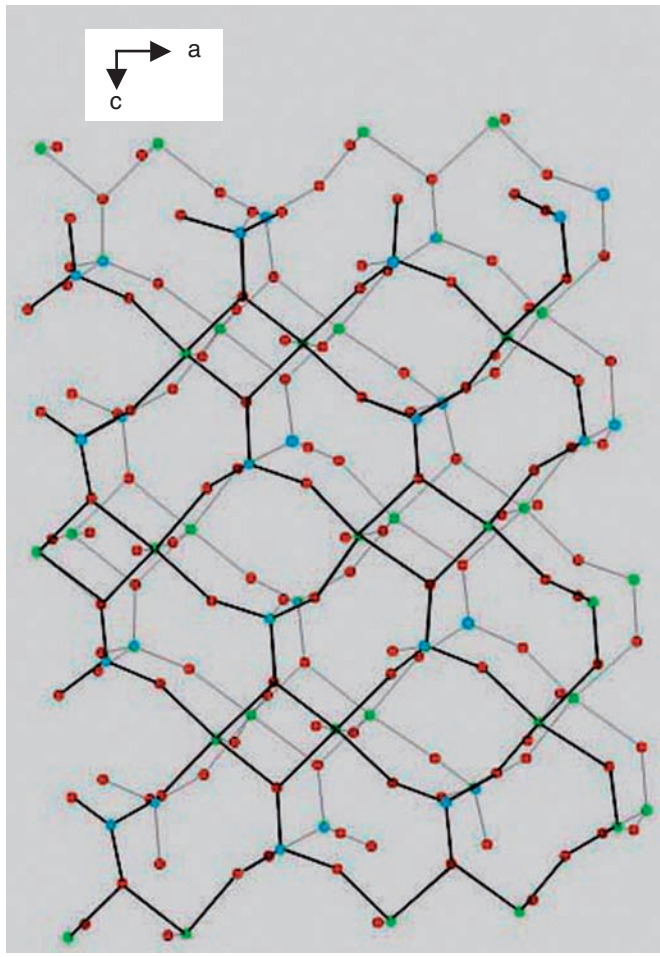


Figure 7.6 Structure of complex $(VO)_2P_2O_7$, viewed down the *b*-axis. Vanadium octahedra and phosphate tetrahedra link together forming a 3D network. Front and back layers are shown.

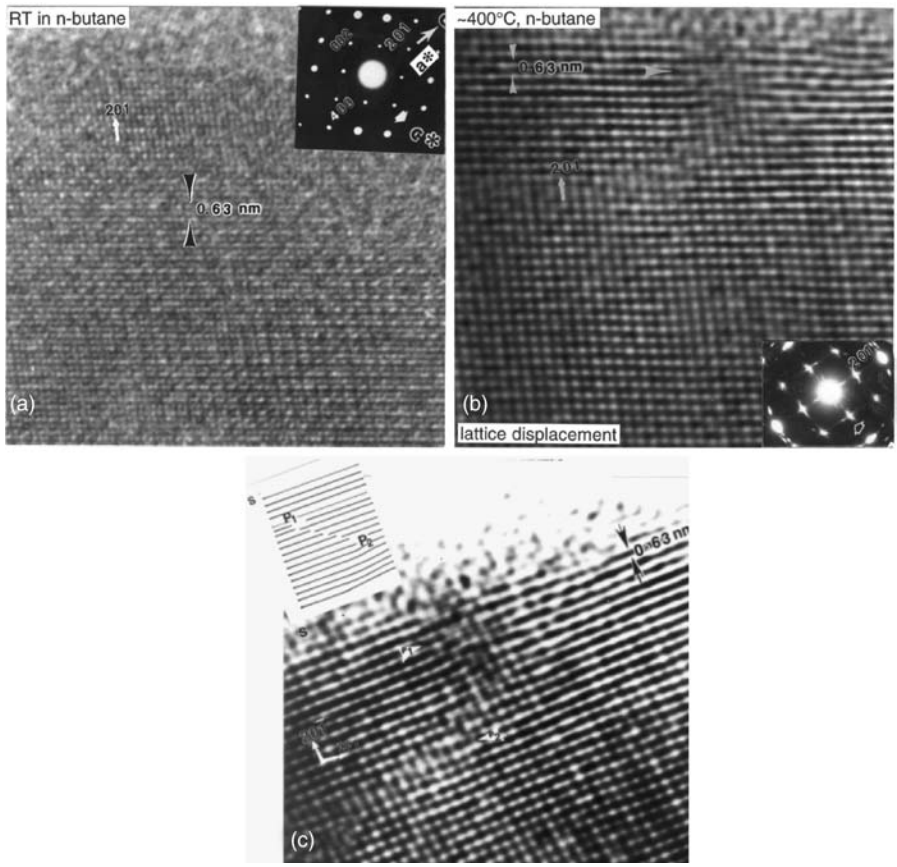


Figure 7.7 (a) *In-situ* atomic-resolution ETEM image of (010) VPO in n-butane at room temperature with electron diffraction pattern inset. (201) reflection is arrowed.[34,61] (b) *In-situ* direct imaging of dynamic atomic motion of reacting VPO in n-butane at $\sim 400\,^{\circ}C$. (c) Enlarged image of (b). The (201) lattice displacements (disturbing the periodicity) due to the reaction are close to the surface S. The resulting defects P1 and P2 are formed by novel glide shear and the lattice is not collapsed. The corresponding electron diffraction pattern (inset in (b)) shows diffuse streaks along a $<201>$ direction (arrowed).

the surface structural development due to the catalyst anion loss shows the formation of extended defects along $<201>$ at an operating temperature of $\sim 400\,^{\circ}C$ (Figure 7.7(b)). The dynamic electron diffraction (inset) shows streaking along $<201>$. The image in Figure 7.7(b) is enlarged in (c). The image reveals a dislocated lattice with terminating lattice planes and the presence of partial dislocations (defects) on (201) lattice planes. The two partial dislocations P1 and P2 (arrowed) are close to the catalyst surface (shown in

profile at S, with the projection of the structure along the electron beam direction), bounding an associated stacking fault. The streaking in the electron diffraction pattern illustrates structural disorder due to the defects on (201) planes. This means that anions in (201) planes, between the vanadyl octahedra and the phosphate tetrahedra are involved in the alkane oxidation reaction. The disorder due to the catalyst anion loss is revealed only in (201) lattice planes, thus excluding all other planes in the crystal structure. These findings, coupled with detailed diffraction contrast experiments[37] show that the defects are formed by glide shear and the lattice is not collapsed (Figures 7.8(a)–(c)). Dynamic atomic studies show that only a few monolayers of the catalyst are involved in butane catalysis.

The disorder along $<201>$ demonstrates that in the catalyst–adsorbate interaction, lattice oxygen loss leads to the formation of coplanar anion vacancies between the vanadyl octahedra and phosphate tetrahedra. Extended defects are introduced along $<201>$. They elucidate that the release of structural oxygen in the oxidation catalysis is accompanied by a novel glide shear mechanism in which a few surface layers of the oxide undergo a structural transformation by glide shear to accommodate the misfit at the interface between the reduced surface layer containing anion vacancies formed during the reaction and the underlying matrix. This important mechanism explains the release of structural oxygen and the preservation of active Lewis acid sites at the surface without changing the overall bulk structure of the catalyst. The positively charged vacancy defect sites are readily available for the electrons of the butane molecule, leading to the alkane activation, and for gas–oxygen exchange. The novel glide shear mechanism revealed by ETEM is of fundamental importance in the understanding of the solid-state heterogeneous catalytic process.

Using atomic-resolution ETEM, it is possible to correlate crystal glide shear defects with catalytic activity.[59,60] Earlier *in-situ* EM studies correlated with reaction chemistry[1,31] have shown that crystallographic shear plane defects (CS) produced by the well-known CS mechanism, which eliminate supersaturation of anion vacancies in reducing oxides by shear and lattice collapse, are secondary to catalysis. That is, CS planes are consequences of oxide reactions and not the origins of catalytic activity. The novel glide shear mechanism leading to glide shear defects revealed by ETEM and correlations of the defects with catalytic reactivity[37,39] show that the glide mechanism is key to efficient butane oxidation catalysis and more generally to hydrocarbon oxidation processes.[37] Catalysts can accommodate anion deficiency without collapsing the crystal lattice and continue to operate, lengthening the catalyst life under optimised ratios of butane/air.[61] The work has led to the development of improved catalysts for the butane oxidation process, *via* substitution of double cation promoters (such as, Fe,Sb or Bi,Mo) in the vanadium site to induce selective glide transformations and optimising anion vacancy defect concentrations.[59,60] An example of improved (Fe,Sb) VPO catalyst performance in butane only environment is shown in Figure 7.9.

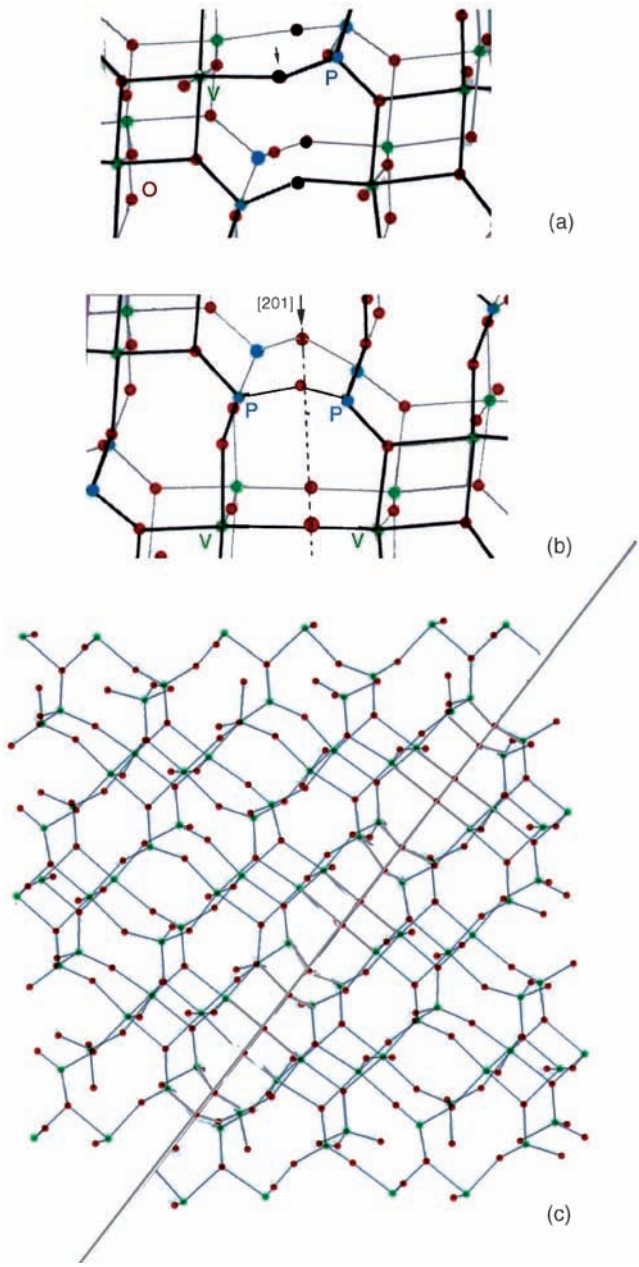


Figure 7.8 (a) Top: Projection of (010) VPO and generation of anion vacancies along <201> (thick lines denote the front layers and thin lines, back layers, respectively) in n-butane. V and P are denoted. (b) Model of glide shear mechanism for butane catalysis (idealised structure). The atom arrowed in (a) (*e.g.* front layer) moves to the vacant site O and so on. The nature of the glide shear is deduced from Figure 7.7 and the defect displacement experiments.[34,37] (c) Schematic of the glide shear defect in 3D. It does not collapse the lattice and the regeneration of the catalyst is possible.

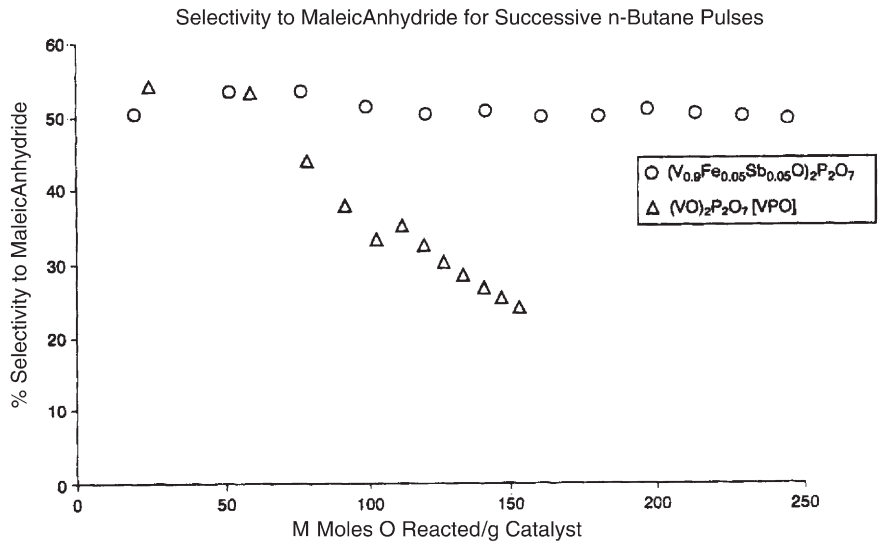


Figure 7.9 Improved selectivity data to successive pulses of butane only for (Fe,Sb) VPO catalysts. Fe and Sb are substituted in the V-lattice to induce selective glide shear transformations to optimise anion vacancy defect concentrations.

7.5.4 *In-Situ* Observations of Carbon Nanotubes in Chemical and Thermal Environments

The discovery of carbon nanotubes (CNTs)[62] has sparked extensive research to explore their remarkable electronic and mechanical properties.[63] They are of interest in applications including, catalyst supports, gas sensors, energy storage and nanoelectronics. CNTs can be thought of as rolled-up sheets of graphene capped with half a fullerene molecule.[64] Single-walled CNTs (SWCNTs) and multi-walled CNTs (MWCNTs) can be synthesised under different process conditions.[63–66] SWCNTs are normally produced in the presence of small quantities of transition metal catalysts either by laser vaporisation of a carbon target,[63,64,67,68] or by the carbon arc method,[64] followed by cleaning, but some residual nanocatalysts may remain. SWCNTs consist of a single layer of carbon atoms with diameters <1 nm and possess superior properties.[62–64] Ordered arrays of SWCNTs form bundles or ropes. A major challenge for technological applications of CNTs is the precise control of the nanostructure and stability. While the basis of the synthesis and structural properties of SWCNTs have been extensively reported, the reaction of SWCNTs with chemical (reducing and oxidising gases) and thermal environments and the effect on their nanostructure, which are critical to the performance of nanotubes are not well understood. Direct observations of the dynamic behaviour of SWCNTs under oxidising, reducing and thermal environments using atomic-resolution *in-situ* ETEM[69–71] have implications in understanding the nanostructural behaviour

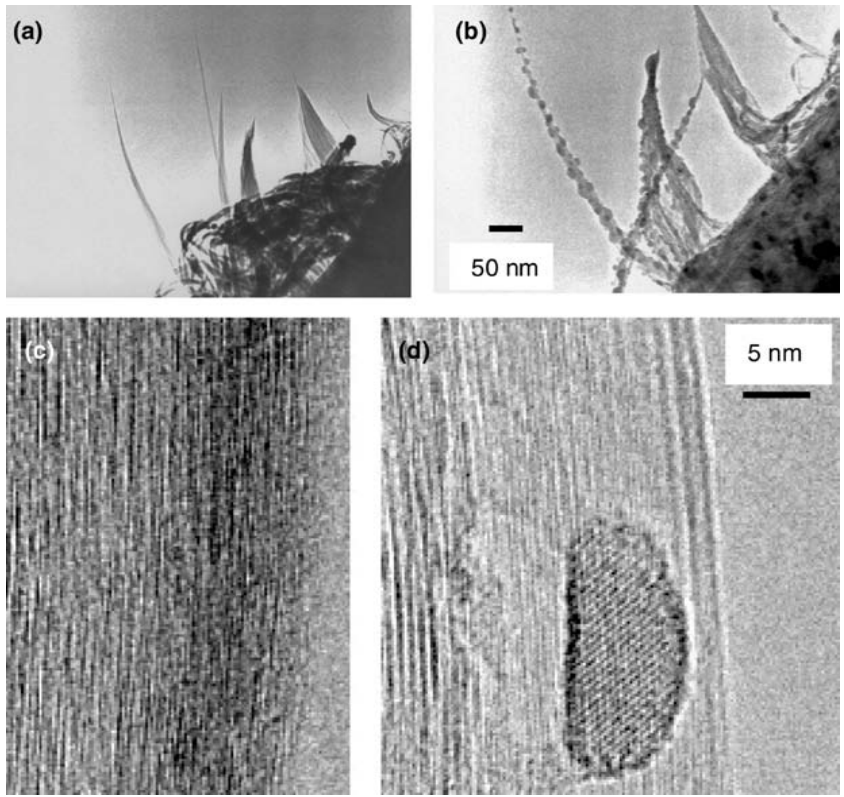


Figure 7.10 *In-situ* ETEM of SWCNT reactions from the *same area:* (a) SWCNT ropes in air at room temperature; (b) air treated, at 575 °C showing the motion of the nanocatalysts; (c) room temperature and (d) 575 °C show the corresponding atomic resolution images (with the nanocatalyst particle at atomic resolution), respectively.

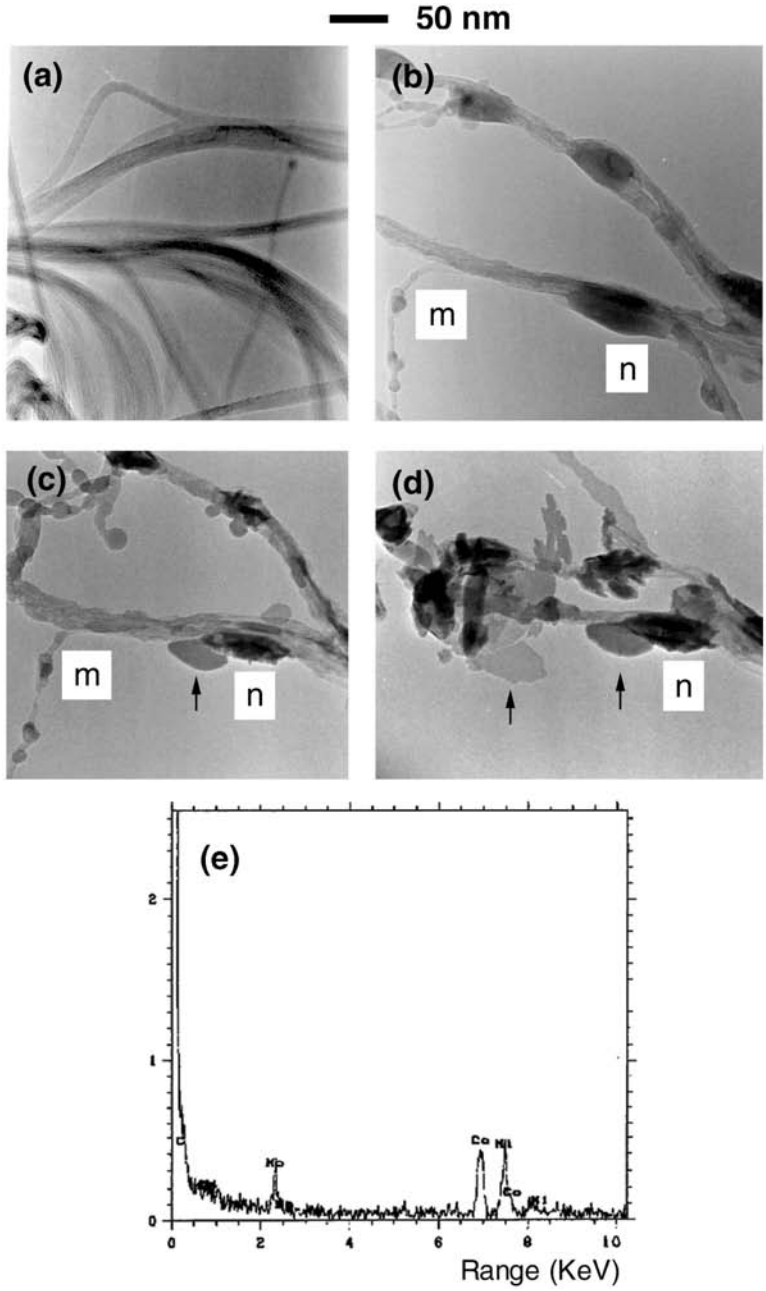


Figure 7.11 *In-situ* ETEM of SWCNT reactions: (a) SWCNT ropes in air at room temperature; (b) air-treated SWCNTs at 575 °C (catalysts at m,n) and cooled to room temperature; (c) the same area (around m,n): in H_2/He at 520 °C, with amorphous carbon evolution at catalysts (*e.g.* arrowed at n) (d) the amorphous growth at ~580 °C. (e) EDX spectrum of the chemical composition (Ni,Co) of the particles. (Mo is the sample support grid).

of CNTs and optimising conditions for their operation in the presence of residual catalysts.

SWCNTs, prepared by laser vapourisation with Ni-Co catalysts[63,64,67,68] have been studied using *in-situ* electron microscopy. SWCNTs show uniformly thick ropes as well as ''ribbon-like'' ropes (where the number varies) in the samples. *In-situ* ETEM in oxidising and reducing environments at a few mbar gas pressures from room temperature to $\sim 800\,^{\circ}C$, for different time periods are summarised in Figures 7.10 and 7.11. Figure 7.10(a) shows an ETEM example of a clean SWCNT rope in air at room temperature, which appears to have very little residual nanocatalyst. However, Figure 7.10(b) shows a dynamic image of the same area in air (3 mbar) at $575\,^{\circ}C$, revealing Ni-Co catalyst particles (confirmed by EDX) and the modified nanostructure of SWCNTs. Figures 7.10(c) and (d) show corresponding atomic-resolution images at room temperature and $575\,^{\circ}C$, respectively. Figures 7.11(a) and (b) show temperature-resolved sequences of SWCNTs from the same area, as follows: (a) in air at room temperature; (b) air-treated at $575\,^{\circ}C$. Sintered Ni-Co rich particles (*e.g.* at m and n), confirmed by EDX are found in and along the SWCNTs. The nanoparticles are also extracted from the tube surface for further analysis to confirm the chemical composition. The ETEM microreactor is then outgassed and cycled with reducing hydrogen gas balanced by helium ($20\%H_2$/He) and reactions are studied in the flowing gas from room temperature to $\sim 580\,^{\circ}C$. Figures 7.11(c) and (d) show dynamic snapshots of SWCNTs (*e.g.* at n) in the H_2/He atmosphere (5 mbar) at $520\,^{\circ}C$ and $580\,^{\circ}C$, respectively. They reveal the evolution and growth of amorphous carbon (arrowed) in the presence of the catalysts (n), confirmed by electron diffraction and atomic imaging. An EDX chemical composition spectrum of the catalyst nanoparticles (n) is shown in Figure 7.11(e). The studies show that the SWCNT nanostructure is significantly altered in the reducing reaction environments. SWCNTs prepared in the presence of catalysts from other methods show similar behaviour.[69–71] Plasma-treated MWCNTs are reported to show surface modifications of carbon nanotubes with graphite formation.[72]

It is proposed[69–71] that the driving force for the particle motion and agglomeration in air is due to contributions from a resultant volume change of the particle in oxidation and a reduction in the surface energy between the particle and the gas. In a hydrogen-containing environment, the reaction mechanism for the amorphisation of SWCNTs at elevated temperatures appears to involve the dissolution of the metal atoms in carbons, followed by the reaction and precipitation of the amorphous carbon.

7.6 Conclusions

This chapter describes the origin and development of atomic-resolution ETEM[35–37] that preserves atomic-scale imaging, electron-diffraction and analytical capabilities under gas environments and at elevated temperatures. The design of this ETEM development has been adopted by commercial TEM

manufacturers and is replicated in laboratories worldwide. Applications of the atomic-resolution ETEM to catalysis and nanosystems have led to the development of novel catalysts and reactions.

More recently, Gai *et al.*[73,74] have proposed that Aberration-Corrected ETEM (AC-ETEM) would offer superior resolution under dynamic reaction conditions. AC-ETEM would be particularly beneficial for supported nanoparticle systems for catalytic, energy and fuel-cell technologies.

Acknowledgements

It is a pleasure to thank L.G. Hanna, E.D. Boyes, K. Kourtakis and S. Ziemecki.

References

1. P.L. Gai, *Catal. Rev. Sci. Eng.*, 1992, **34**, 1.
2. P.L. Gai, *Curr. Opin. Solid State Mater. Sci.*, 1999, **4**, 63.
3. P.L. Gai, *Top. Catal.*, 2002, **21**, 168.
4. P.L. Gai and E.D. Boyes, *Electron Microscopy in Heterogeneous Catalysis*, Institute of Physics Publishing, Bristol, UK, Philadelphia, USA, 2003.
5. P.B. Hirsch, A. Howie, R. Nicholson, D.W. Pashley and M.J. Whelan, *Electron Microscopy of Thin Crystals*, Butterworths, London, 1965.
6. R. Sinclair, T. Yamashita and F. Ponce, *Nature*, 1981, **290**, 386.
7. H. Saka and K. Kamino, *In situ Electron Microscopy in Materials Research*, ed. P.L. Gai, Kluwer Academic Publishers, London, Boston, 1997.
8. D.W. Pashley, M.J. Stowell, M.H. Jacobs and T.J. Law, *Philos. Mag.*, 1964, **10**, 127.
9. K. Yagi, K. Kobayashi, Y. Tanishiro and K. Takayanagi, *Thin Solid Films*, 1985, **126**, 95.
10. J.M. Thomas and W.J. Thomas, *Principles and Practices of Heterogeneous Catalysis*, VCH, Weinheim, 1997.
11. H. Hashimoto, *Jpn. J. Appl. Phys.*, 1968, **7**, 946.
12. P. R. Swann and N. Tighe, *Jernkont. Ann.*, 1971, **155**, 251.
13. E. P. Butler and K. F. Hale, *Dynamic Experiments in Electron Microscopy*, North Holland, Amsterdam, 1981.
14. D. Double, A. Hellawell and S. Perry, *Proc. Roy. Soc.*, 1978, **A359**, 435.
15. D. F. Parsons, *Science*, 1974, **186**, 407.
16. R.T.K. Baker, *Catal. Rev. Sci. Eng.*, 1979, **19**, 161.
17. H. Fujita, ed., *In Situ Experiments in High Voltage Electron Microscopy*, Osaka University Press, Osaka,1985.
18. R.C. Doole, G. Parkinson and J.M. Stead, *Inst. Phys. Conf. Ser.*, 1991, **119**, 161.

19. P.L. Gai and P.B. Hirsch, *Proceedings on the Chemistry and Uses of Molybdenum*, of the Climax Mo Co. Ltd and Chemical Society (Dalton Division) University of Oxford, UK, 1976.
20. P.L. Gai, W. Thoni and P.B. Hirsch, *J. Less-Common Met.*, 1977, **54**, 263.
21. P.L. Gai and M.J. Goringe, *Krist. Tech.*, 1979, **14**, 1385.
22. P.L. Gai, C.J. Humphreys, A.E. Webb, D.R. Pyke and J.C.J. Bart, *Inst. Phys. Conf. Ser.*, 1980, **52**, 317.
23. P.L. Gai, *Philos. Mag.*, 1981, **43**, 841.
24. P.L. Gai and M.J. Goringe, *Proc. 39th El. Micr. Soc. of America*, San Francisco Press, San Francisco, 1981, p. 68.
25. P.L. Gai, E.D. Boyes and J.C.J. Bart, *Philos. Mag.*, 1982, **A45**, 531.
26. P.L. Gai, *J. Solid State Chem.*, 1983, **49**, 25.
27. P.L. Gai, *Philos. Mag.*, 1983, **48**, 359.
28. P.L. Gai and P.A. Labun, *J. Catal.*, 1985, **94**, 79.
29. P.L. Gai, B.C. Smith and G. Owen, *Nature*, 1990, **348**, 430.
30. P.L. Gai and B.C. Smith, *Ultramicroscopy*, 1990, **34**, 17.
31. P.L. Gai, *J. Solid State Chem.*, 1993, **104**, 119.
32. T.C. Lee, D. Dewald, J. Eades, I.M. Roberetson and H.K. Birnbaum, *Rev. Sci. Instrum.*, 1991, **62**, 1438.
33. P. Crozier, R. Sharma and A. Datye, *Proc. of Micr. Soc. of America*, 1998, **4**, 228.
34. P. L. Gai and K. Kourtakis, *Science*, 1995, **267**, 661.
35. E.D. Boyes and P.L. Gai, *Ultramicroscopy*, 1997, **67**, 219.
36. P.L. GaiE.D. Boyes, *In situ Microscopy in Materials Research*, Kluwer Academic Publishers, Boston, London, 1997.
37. P.L. Gai, *Acta. Crystallogr.*, 1997, **B53**, 346.
38. P.L. Gai, *Adv. Mater.*, 1998, **10**, 1259.
39. P.L. Gai, *Top. Catal.*, 1999, **8**, 97.
40. J. Haggin, *Chem. Eng. News* (American Chemical Society) 1995, **73** (No. 30), 39.
41. E.D. Boyes and P.L. Gai, *Electron Microscopy*. ICEM 14; ed. H.A. Claderon Benavides, M.J. Yacaman, Institute of Physics Publishers, UK, 1998, **3**, p.511.
42. T.W Hansen, J. Wagner and P.L. Hansen, *Proc of 12th Euro. Congr. on EM* (Czech EM Soc.), 2000, **2**, 537.
43. T. Hansen, J. Wagner, P.L. Hansen, H.T. Topsoe and S. Dahl, *Science*, 2001, **294**, 1508.
44. V. Oleshko, P. Crozier, R. Cantrell and A. Westwood, *J. Electron Microsc.*, 2002, **51**, S27.
45. R. Wang, P. Crozier, R. Sharma and J. Adams, *Microsc. Microanal.*, 2005, **11**(Suppl. 2), 216.
46. R. Sharma and P. Crozier, *Microsc. Microanal.*, 2003, **9**, CD314.
47. R. Sharma, P. Crozier, L. Eyring and F.C. Kang, *Philos. Mag.*, 2004, **84**, 2731.
48. P.J. Li, J. Liu, N. Nag and P. Crozier, *Microsc. Microanal.*, 2005, **11**(suppl. 2), 1538CD.

49. P.L. Gai, *Inst. Phys. Conf. Ser.*, 2001, **168**, 401.
50. P.L. Gai, K. Kourtakis and S. Ziemecki, *Microsc. Microanal.*, 2000, **6**, 335.
51. P.L. Gai, *Curr. Opin. Solid State Mater. Sci.*, 2001, **5**, 371.
52. F. Nagata and I. Ishikawa, *Jpn. J. Appl. Phys.*, 1972, **11**, 1293.
53. K. Fukushima, A. Ishikawa and A. Fukami, *J. Electron Microsc.*, 1985 **34**, 47.
54. T. Daulton, B. Little, K. Lowe and J. Jones Meehan, *Proc. MSA*, ed. G. Bailey, 2001, 134.
55. P.L. Gai, *Microsc. Microanal.*, 2002, **8**, 21.
56. C. De Bellefon and P. Fouilloux, *Catal. Rev. Sci. Eng.*, 1994, **36**, 459.
57. P.L. Gai, K. Kourtakis and E.D. Boyes, *Catal. Lett.*, 2005, **102**, 1.
58. G. Centi (ed.) *Catal. Today*, 1993, **16**, 1.
59. P.L. Gai, K. Kourtakis, D.R. Coulson and G.C. Sonnichsen, *J. Phys. Chem.*, 1997, **101**, 9916.
60. K. Kourtakis and P.L. Gai, *J. Molec. Catal.*, 2004, **220**, 93.
61. J.M. Thomas and P.L. Gai, *Adv. Catal.*, 2004, **48**, 171.
62. S. Iijima, *Nature*, 1991, **354**, 56.
63. A. Thess, P.G. Collins, A. Fettle and R.E. Sunaccey, *Science*, 1997, **273**, 219.
64. M.S. Dresselhaus, S. Dresselhaus and P.C. Eklund, *Fullerenes, Carbon Nanotubes*, Academic Press, New York, 1996.
65. M.J. Goringe, A. Rawcliffe, A. Burden, J.L. Hutchison and R.C. Doole, *Faraday Discuss.*, 1996, **105**, 102.
66. S. Iijima, *Microsc. Microanal.*, 2005, **11**(suppl.2), 1544CD.
67. P.L. Gai, J.G. Lavin and E.D. Boyes, *Electron Microsc.*, 1998, **3**, 510.
68. P.L. Gai, O. Stephan, K. McGuire, A.M. Rao, M.S. Dresselhaus, G. Dresselhaus and C. Colliex, *J. Mater. Chem.*, 2004, **14**, 669.
69. E.D. Boyes and P.L. Gai, *Microsc. Today*, 2004, July 07 issue.
70. P.L. Gai and E.D. Boyes, *Mater. Res. Soc. Sym. Proc.*, 2005, **12**, 876E.
71. P.L. Gai, *Microsc. Microanal.*, 2006, **12**, 48.
72. S. Trasobares, J. Bivvel, X. Xiao, C. Ewels, O. Stephan, J. Carlisle and P.M. Ajayan, *Adv. Mater.*, 2004, **16**, 610.
73. P.L. Gai and E.D. Boyes, (a) *Frontiers of Electron Microscopy Proceedings*, 2003, Conference Abstracts Published by Frontiers of Electron Microscopy Berkeley, CA, USA (b) *Adv. Catal.*, 2004, **48**, 171.
74. P.L. Gai, L.G. Hanna and E.D. Boyes, *Microsc. Microanal.*, 2005, **11** (suppl. 2), 1526CD.

Subject Index

Page references in *italic* refer to Figures not included in the page ranges.