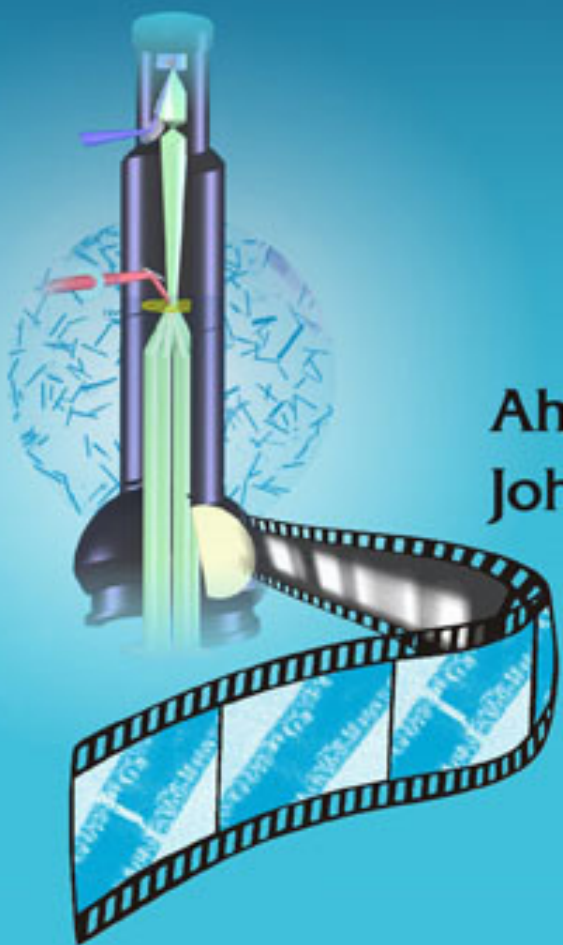
4D Electron Microscopy

Imaging in Space and Time



Ahmed H. Zewail
John M. Thomas

Imperial College Press

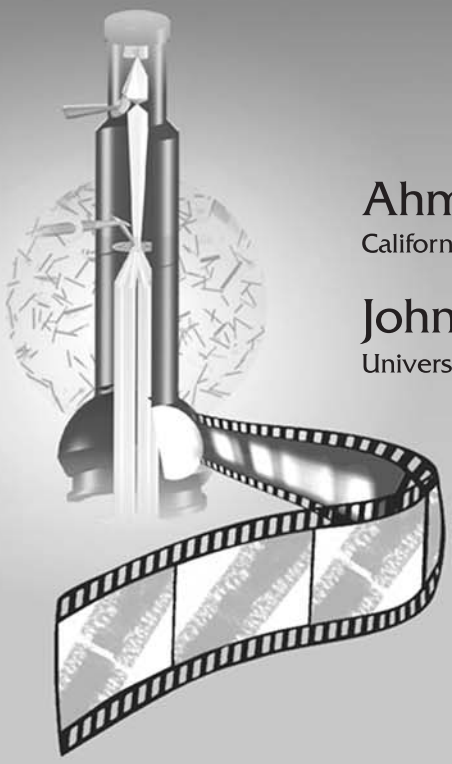
4D Electron Microscopy

Imaging in Space and Time

This page intentionally left blank

4D Electron Microscopy

Imaging in Space and Time



Ahmed H. Zewail
California Institute of Technology, USA

John M. Thomas
University of Cambridge, UK

ICP

Imperial College Press

Published by

Imperial College Press
57 Shelton Street
Covent Garden
London WC2H 9HE

Distributed by

World Scientific Publishing Co. Pte. Ltd.

5 Toh Tuck Link, Singapore 596224

USA office: 27 Warren Street, Suite 401-402, Hackensack, NJ 07601

UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

Library of Congress Cataloging-in-Publication Data
Zewail, Ahmed H.
4D electron microscopy : imaging in space and time / Ahmed H Zewail , John M Thomas.
p. cm.
Includes bibliographical references and index.
ISBN-13: 978-1-84816-390-4 (hardcover : alk. paper)
ISBN-10: 1-84816-390-8 (hardcover : alk. paper)
ISBN-13: 978-1-84816-400-0 (pbk. : alk. paper)
ISBN-10: 1-84816-400-9 (pbk. : alk. paper)
1. Electron microscopy. 2. Materials science. 3. Biological sciences. 4. Imaging.
I. Thomas, J. M. (John Meurig) II. Title. III. Title: Four dimensional electron microscopy.
QH212.E4Z49 2009
570.28'25--dc22

2009037249

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

Printed in Singapore.

Acknowledgements

The science and technology discussed in this book are the result of dedicated efforts by our students and associates, and by colleagues around the globe. The writing and production of the final manuscript, however, required the devotion and interest of a few special individuals. We owe a great deal to Dr. Dmitry Shorokhov of Caltech for his untiring commitment to the production of a consistent and illustrative text, and for his drive for completion of the work on time. We have benefited from numerous discussions with present members of the Physical Biology Center (Caltech) and the Department of Materials Science and Metallurgy (Cambridge), especially P. A. Midgely with whom we bounced around some of our concepts and ideas. Special thanks go to C. J. Humphreys of Cambridge University and M. Chergui of EPFL in Lausanne for their penetrating and helpful reviews of the manuscript.

John Spence, a world expert in the field of electron microscopy, reviewed the book in its entirety and made constructive suggestions and detailed comments on an early draft of the manuscript. We very much appreciate John's invaluable insights and the "on time" e-mails! Throughout the writing of the text we had stimulating discussions (and correspondence) with many colleagues: W. Baumeister, E. D. Boyes, D. J. Buttrey, W. Chiu, R. A. Crowther, C. Ducati, R. E. Dunin-Borkowski, R. F. Egerton, M. A. A. Franco, P. L. Gai, R. M. Glaeser, J. M. Gonzalez-Calbet, S. Helveg, R. Henderson, J.-C. Hernandez-Garrido, A. Howie, A. B. Hungría, M. José-Yacamán, G. J. Jensen, C. J. Kiely, A. I. Kirkland, A. Klug, R. D. Leapman, B. V. McKoy, S. Mejia-Rosales, J. M. Plitzko, O. Terasaki, M. M. J. Treacy, P. A. Wright, E. Zeitler, and W. Zhou. AHZ would like to thank all members, past and present, of Caltech group, who contributed significantly to the success of the story told here.

Gratefully acknowledged are the financial support from the Gordon and Betty Moore Foundation, the National Science Foundation, and the Air Force Office of Scientific Research that made the research advances at Caltech possible, and the administrative assistance of De Ann Lewis and Maggie Sabanpan at Caltech (especially during the "high-maintenance" visits of JMT to Pasadena), as well as the enthusiastic commitment of Dr. K. K. Phua and Ms. Sook Cheng Lim of Imperial College Press to publish the book on time.

Last but not least, we wish to acknowledge the unqualified and loving support of our families.

This page intentionally left blank

Preface

Our intention in writing this book is threefold: to distill, to impart exciting new knowledge, and to predict. We have been impelled to delineate and recall the many remarkable achievements already accomplished using electron microscopy, and to focus on what we perceive to be key opportunities for the future exploration of the unique attributes that this powerful technique possesses — and this we do in an unorthodox manner so as to link the fundamentals of microscopy to emerging new concepts and developments.

First, we distill the basic principles involved in imaging and diffraction and in so doing illuminate the central role occupied by notions of coherence and interference. One of us first began using the electron microscope over 40 years ago, principally as an aid to elucidating certain enigmas in his main fields of study: solid-state chemistry, catalysis and surface and materials science. We devote two chapters to the diverse applications of static 2D and 3D imaging covering both organic and inorganic materials as well as biological extended structures, including proteins, viruses, and molecular machines.

The principal focus of the book, however, is on the new development at Caltech of 4D electron microscopy. Because more than a million million frames per second have to be taken in order to produce the picture of atomic motion, the time resolution achieved by stop-motion recording in "real time" far exceeds that of the human brain (milliseconds). Each of us passionately believes that the addition of the time domain to "static" imaging by conventional electron microscopy in its various modes of operation opens up an almost bewildering variety of exciting and important opportunities, especially in probing the intricacies of fundamental dynamic processes. It is well known that the electron microscope (EM) yields information in three distinct ways: in real space, in reciprocal space and, when used in a spectroscopic mode, in energy space. By introducing time resolution, we achieve a fourth domain. Here, our use of the term 4D EM also emphasizes the four fundamental dimensions of matter — three spatial and one temporal. These two notions of 4D EM are, however, inextricably interwoven as will become apparent in the ensuing text.

Time resolution in EM now exceeds by ten orders of magnitude that of conventional video cameras, making it feasible to follow the unfolding of various processes in their real length-time coordinates, from micrometer-subsecond down to angstrom-femtosecond. With such resolutions in both space and time, using "tabletop instruments," we believe that the modern 4D electron microscope in its numerous variants is unique (when compared, for example, with sources of synchrotron radiation) in the exploration of new phenomena and in the retrieval

of new knowledge. For this reason, we devote several chapters to the principles of the 4D approach, emphasizing the concept and the merits of timed single-electron imaging in real space, Fourier space, and energy space. Besides the time resolution, there are consequential advantages which are discussed, with emphasis on the opportunities resulting from the spatial and temporal coherence, the brightness, and the degeneracy of (fermionic) electrons when generated as ultrafast packets and in the absence of space–charge repulsion. The applications span physical, mechanical, and some biological systems. We also discuss recent advances made possible in electron-energy-loss spectroscopy (EELS), in tomography and holography. We conclude with an outlook on the future.

This book is not a *vade mecum* — numerous other texts are available for the practitioner of 3D microscopy for that purpose. It is, instead, an in-depth exposé of new paradigms, new concepts and developed techniques, including the significant advances that can now be executed to retrieve new knowledge from the corpus of physical and biological sciences. We presuppose that the reader is acquainted with some introductory aspects of laser science and with the rudiments of basic knowledge of Fourier transformation and other crystallographic procedures, although we have made an effort didactically to present self-sufficient accounts in each chapter.

The first draft of the manuscript was written in Pasadena in August–September of 2008. During that period, we worked with frenetic zeal (up to 12 hours a day!) so as to do justice to the existing rich literature and numerous recent developments. Since then, and before sending the manuscript to the publisher, a variety of new and exciting results have emerged, the most important of which are incorporated herein.

We hope that the reader will share our enthusiasm for the book, and that we have succeeded in its execution to aim at a broad readership within the various disciplines of science, engineering, and medicine. Above all, we hope that we shall induce many young (and old!) readers to enter this field.

Ahmed H. Zewail
Pasadena

John M. Thomas
Cambridge

Contents

Chapter 1

Historical Perspectives

From Camera Obscura to 4D Imaging

Over their time on Earth, the ever-increasing progress made by humans in making visible and tangible the very small and the very large is remarkable. Although the acuity of the human eye is not limited by diffraction, its spatial and temporal resolutions are limited to about 100 μm and a fraction of a second, respectively, as discussed in Chap. 2. Today, we are aided by tools that can visualize objects that are below a nanometer in size and that move in femtoseconds or attoseconds (see Figs. 1.1 and 1.2, and also Chap. 8).

How did it all begin? First came the power of light for observation, which has been with humans since their creation, stretching back over many millennia (Fig. 1.3). For the imaging of small objects, major concepts were developed by the Arab polymath Alhazen (Ibn al-Haytham; 965–1040). His brilliant work was published in the *Book of Optics* or *Kitab al-Manazir* (in Arabic); see Fig. 1.4 and references therein. He is recognized for his quantitative experimentation and thoughts on reflection and refraction, and is also credited with the correct explanation of the mechanism of vision, prior to the contributions of Kepler, Descartes, da Vinci, Snell, and Newton. But of considerable significance to our topic is his conceptual analysis of the *camera obscura*, the "dark room," which aroused the photographic interests of Lord Rayleigh in the 1890s.[1] Alhazen's idea that light must travel along straight lines and that the object is inverted in the image plane is no different from the modern picture shown in Fig. 2.8!

In the 14th and 15th centuries, the art of grinding lenses was perfected in Europe, and the idea of optical microscopy was developed. In 1665, Robert Hooke (the man who coined the word "cell") published his studies in *Micrographia* (Fig. 1.5), and among them was a description of plants and feathers, as well as of cork and its ability to float in water. Contemporaneously, Anton van Leeuwenhoek used a simple, one-lens microscope to examine blood, insects, and other objects, and was the first to visualize bacteria, among other microscopic objects. More than a hundred years later, an experiment by the physicist, physician, and egyptologist Thomas Young demonstrated the interference of light, an achievement that revolutionized our views on the nature of light. His double-slit experiment of 1801,

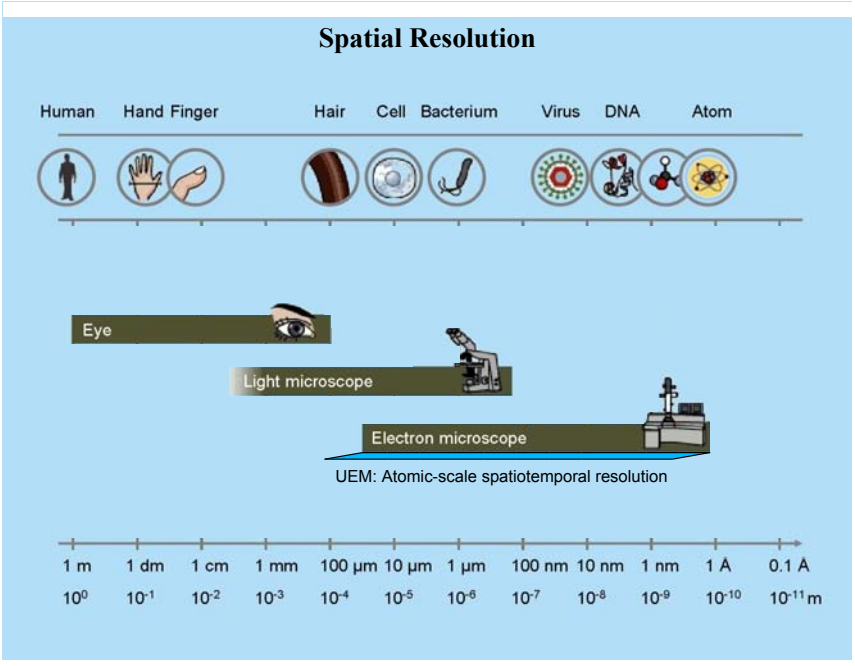


Figure 1.1 The spatial resolution, from the eye to modern electron microscopes (adapted from *www.nobelprize.org*), together with the now improved temporal resolution in UEM (10 orders of magnitude better than that of a human eye).

performed at the Royal Institution of Great Britain, led to the demise of Newton's corpuscular theory of light. Of relevance here is the phenomenon of diffraction due to interference of waves (coherence), and much later such diffraction was found to yield the (microscopic) interatomic distances characteristic of molecular and crystal structures, as discovered in 1912 by von Laue and elucidated later that year by W. L. Bragg (see Chap. 2).

Resolution in microscopic imaging was brought to a whole new level by two major developments in optical microscopy. In 1878, Ernst Abbe formulated a mathematical theory correlating resolution to the wavelength of light (beyond what we now designate the empirical Rayleigh criterion for incoherent sources; Chap. 2), and hence the optimum parameters for achieving higher resolution. At the beginning of the 20th century, Richard Zsigmondy, by extending the work of Faraday and Tyndall,[2] developed the "ultramicroscope" to study colloidal particles. Then came the penetrating developments by Frits Zernike in the 1930s, who introduced the phase-contrast concept in optical microscopy. The spatial resolution of optical microscopes was understood to be limited by the wavelength of visible light.

Just before the dawn of the 20th century, the *corpuscle* (electron) was discovered by J. J. Thomson in 1897. With the subsequent pervasive influence

of quantum mechanics in the physical and adjacent sciences in the first three decades of the 20th century, Louis de Broglie formulated the concept of particle–wave duality in 1924, to the extent that the wave character of an electron, which is quantified in the relationship $\lambda_{\text{de Broglie}} = h/p$, became essential for our understanding of diffraction and imaging. The first evidence of the wave character of the electron was established in 1927 by Davisson and Germer (diffraction from a nickel surface) and, independently, by G. P. Thomson (the son of J. J.), who, with Reid, observed diffraction of electrons penetrating a thin foil; Hans Bethe, then a student with Sommerfeld in 1928, discussed in his thesis multiple scattering and the role of the inner potential in relation to the Davisson–Germer experiment. Even earlier (in 1912), the 22-year-old Lawrence Bragg had made an imaginative leap by assuming that X-rays, then universally regarded as corpuscles, were waves, thereby leading to his famous $n\lambda = 2d\sin\Theta$ equation. In 1923, Dirac postulated "single-particle interference."

In 1933, Max Knoll and Ernst Ruska invented the electron microscope and improved the resolution to the (sub)micrometer scale, in the transmission mode. Boersch introduced the diffraction lens in transmission electron microscopy (TEM) in 1936, and a few years later (1940) he found so-called Fresnel fringes as "diffraction at edges" in the microscope. Concurrently, Walther Kossel and

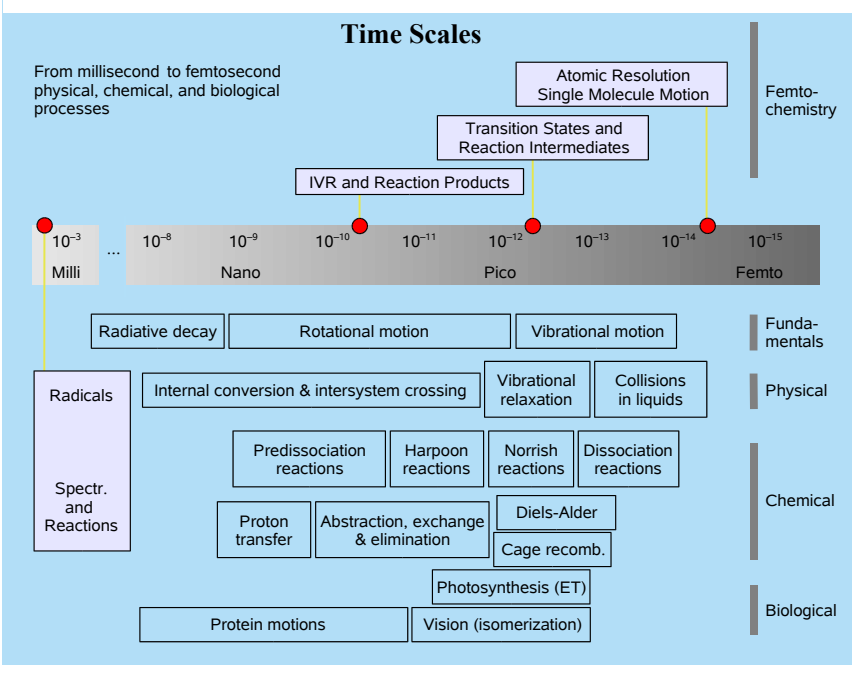
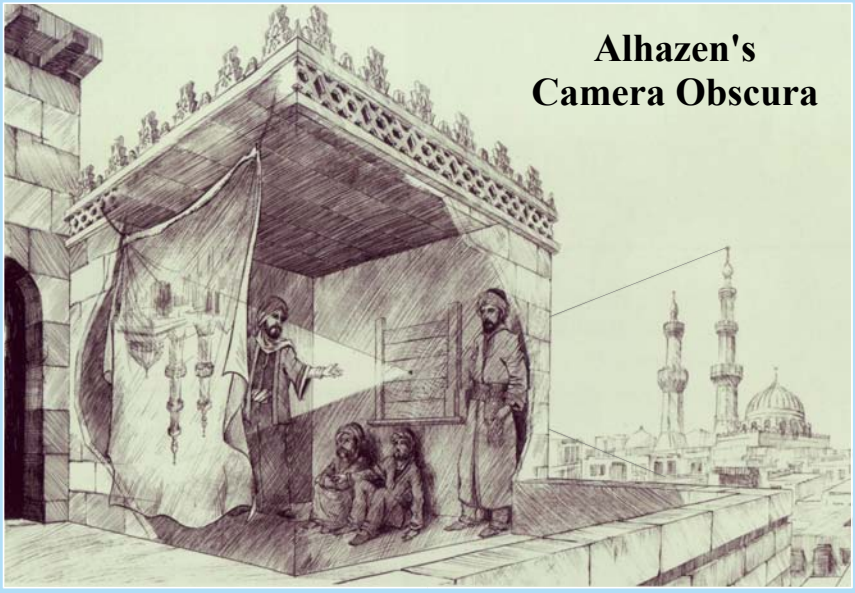


Figure 1.2 Time scales of relevance to physical, chemical, and biological changes. The fundamental limit of the vibrational motion defines the regime for femtochemistry. Examples are given for each change and scale.[21]

Gottfried Möllenstedt in 1939 combined in their electron microscope the ability to record projected 2D images and electron diffraction patterns, which contain information on the structure, the repeating lattice distances, and other aspects pertaining to crystallographic symmetry. These and other related developments in microscopy led to electron interferometry and holography. The original proposal of electron holography by Denis Gabor in 1948 and the birth of electron biprism interference by Möllenstedt in 1953 laid the foundation[3,4] for the advances made by Tonomura and others[5] in the years to follow.

Figure 1.3 The significance of light–life interaction as perceived more than three millennia ago, since Akhenaton and Nefertiti. Note the light's "ray diagram" from a spherical (point) source, the Sun.

The first images of individual atoms were obtained in 1951 by Erwin Müller,[6–8] who introduced the technique of field-ion microscopy to visualize them at fine tips of metals and alloys, and to detect vacancies and atomic steps and kinks at their surfaces. With the invention of field-emission sources and scanning transmission electron microscopy, pioneered by A. V. Crewe, isolated heavy atoms became readily visible in 1970.[9] (The scanning-tunneling microscope was developed in the 1980s and made possible atomic-scale images of conducting surfaces.) Today, with aberration-corrected microscopes, imaging has reached a resolution of less than an angstrom.[10] This history would be incomplete if we did not mention that the totality of technical developments and applications in the



In his celebrated book "*Kitab al-Manazir*", the "*Book of Optics*", al-Hasan Ibn al-Haytham (Latinized as *Alhacen* or *Alhazen*) coined the Arabic expression "al-bayt al-muzlim", which literally means "dark room" (or *camera obscura* in Latin). Sabra of Harvard gives a detailed account and translation of the original work in which the discovery is documented in a series of volumes.[22] In Book I, Chapter 3 (I/3), Alhazen describes a series of experiments with a sunlight beam entering a darkened room through a hole in the door; he repeats the experiments with a row of candles (or lamps) and notes (in I/6) that the image formed on the wall opposite to the door is inverted. Alhazen's interest was in the understanding of the behavior of light, and his explanation of light traveling along straight lines from the source (object) to the detector plane indeed constitutes the basis for many phenomena in optics and imaging.[22-25] Introducing such methods of experimental observation earned Alhazen the title of the First Scientist,[23] and the place alongside Galileo (Figure 1.5) who came into the world of observational science centuries after Alhazen. Besides this and related work, Alhazen made revolutionary contributions to the science of vision. Combining investigations of the behavior of light with inventive geometrical proofs, and exploration of the physiology of visual perception, he formed a coherent alternative to the Euclidean and Ptolemaic theories of "visual rays" originating from the eye. This work had a real impact on the science pursued later in 17-th century Europe. The Latin translation of Alhazen's work influenced scientists and philosophers such as (Roger) Bacon and da Vinci, and formed the foundation for the work by mathematicians like Kepler, Descartes and Huygens, as acknowledged in the literature on the history of science.[24] Rashed of CNRS/Université Paris, in his book "*Geometry and Dioptrics in Classical Islam*", gives an exposé of the influence of Alhazen in this context.[25] René Descartes who, in 1637, published three essays—Dioptrics, Meteorology, and Geometry—expanded Alhazen's work on refraction, the rainbow, and an alytical geometry. Ibn al-Haytham, like the other notable contemporaries, such as Ibn Sina (Avicenna) and al-Biruni, was a polymath concerned with the subject of faith and reason. He wrote on philosophy, the rational thinking that he recognized in Aristotelian natural philosophy, and on rigor of mathematics which he enjoyed immensely. He is quoted to have said:[22] "*All matters, secular and religious, are the fruits of the philosophical sciences*" which, in his time, comprised all of mathematics, the natural sciences, and theology or metaphysics. The totality of his scientific contributions places him among the giants, as evidenced by the immortal recognition by historians, but equally impressive is his philosophical heritage. In "*Doubts Concerning Ptolemy*" he wrote:[23] "*Truth is sought for itself, but the truths are immersed in uncertainties, and scientific authorities are not immune from error, nor is human nature itself.*" Ibn al-Haytham, who was born in Basra (Iraq), died *ca*. 1040 after a scholarly rich and productive life in both Iraq and Egypt.

Figure 1.4 The concept of *camera obscura* as perceived a millennium ago by Alhazen (Ibn al-Haytham), who coined the phrase (see text). Note the formation of the inverted image through a ray diagram. Graphics piece adapted from Ref. 26.

Figure 1.5 Microscopy timeline, from *camera obscura* to 3D electron microscopes. 4D ultrafast electron microscopy and diffraction were developed a decade ago (see text). The top inset shows the frontispiece to Hooke's *Micrographia* published by the Royal Society of London.[27] In the frontispiece to Hevelius's *Selenographia* (bottom inset), Ibn al-Haytham represents *Ratione* (the use of reason) with his geometrical proof and Galileo represents *Sensu* (the use of the senses) with his telescope. The two scientists hold the book's title page between them, suggesting a harmony between the methods.[22b,23]

investigations of inorganic and organic materials have benefited enormously from the contributions of many other scientists, and for more details we refer the reader to books by Cowley,[11] Spence,[12] Humphreys,[13a] and recent reviews by Hawkes and Howie.[13b]

Biological EM has been transformed by several major advances including electron crystallography, single-particle tomography, and cryo-microscopy, aided by large-scale computational processing. Beginning with the electron crystallographic work of DeRosier and Klug[14] in 1968, 3D density maps from electron microscope images became retrievable. Landmark experiments revealing the high-resolution (atomic-scale) structure from 2D crystals, single-particle 3D cryo-EM images of different but identical particles (6 Å resolution), and 3D cryo-EM images of the same particle (tomography with 6 Å resolution) represent the tremendous progress made. With these methods the first membrane protein structure was determined; the first high-resolution density maps for the protein shell of an icosahedral virus were obtained; and the imaging of whole cells was accomplished. Minimizing radiation damage by embedding the biological macromolecules and machines in vitreous ice affords a noninvasive, high-resolution imaging technique of visualizing the 3D organization of eukaryotic cells, with their dynamic organelles, cytoskeletal structure, and molecular machines in an unperturbed context, with a resolution of 6 Å to 2 nm, being limited by radiation damage. We refer the reader to articles by Henderson,[15] Crowther,[16] Baumeister and Glaeser,[17] and the books by Glaeser *et al.*[18] and by Frank.[19]

Shortly after the first electron microscope was built by Knoll and Ruska, it was realized that the resolution of the instrument, working under ideal conditions, would far exceed that attainable using a light microscope. It is known from the Rayleigh criterion (Chap. 2) that, with a wavelength of λ for the probing beam, the smallest distance that can be resolved is given by *ca.* 0.5 λ. Thus, in conventional optical microscopy, green light cannot resolve distances smaller than *ca.* 3000 Å (300 nm). When, however, an electron is accelerated in a 100 kV microscope, its wavelength is *ca.* 4×10^{-2} Å, reaching the picometer scale, a distance far shorter than that separating atoms in a solid or molecule. Equally important, electron imaging provides structural information (shapes), whereas light microscopes, with the best resolution, and using fluorophores, provide only positions (coordinates). Special variants of optical microscopy can nowadays resolve small objects of several tens of nanometers, below the diffraction limit.[20] For a variety of reasons, a principal one being the inherent imperfection of electron-optical lenses, it is not yet possible to achieve the theoretical limit of resolution set by the wavelength of the electron. But steady advances have been made, especially in recent years, in achieving resolutions of less than 1 Å, thanks to the arrival of so-called aberration-corrected electron lenses (Chap. 8).

Of the three kinds of primary beams (neutrons, X-rays and electrons) suitable for structural imaging, the most powerful are coherent electrons, which are readily produced from field-emission guns. The source brightness, as well as the temporal and spatial coherence of such electrons, significantly exceeds achievable values for neutrons and X-rays. Moreover, their minimum probe diameter is as small as

1 Å, and their elastic mean-free path is *ca.* 100 Å (for carbon), much less than for neutrons and X-rays (Chaps. 5 and 7).

As a result of these developments and inventions, new fields of research have now emerged. First, by combining energy-filtered electron imaging with electron tomography, chemical compositions of sub-attogram ($< 10^{-18}$ g) quantities located at the interior of microscopic or mesoscopic objects may be retrieved non-destructively. Second, transmission electron microscopes fitted with field-emission guns to provide coherent electron waves can be readily adapted for electron holography to record the magnetic fields within and surrounding nanoparticles or metal clusters, thereby yielding the lines of force of, typically, a nanoferromagnet encapsulated within a multi-walled carbon nanotube. Third, advances in design of aberration-corrected high-resolution electron microscopes have greatly enhanced the quality of structural information pertaining to nanoparticle metals, binary semiconductors, ceramics and complex oxides. Moreover, electron tomography sheds light on the shape, size and composition of bimetallic catalysts attached to nanoporous supports (Chaps. 3 and 4).

Whereas in all of the above methods the process of imaging, diffraction, and chemical analyses has been conducted in a static (time-integrated) manner, it has now become possible to unite the time domain with the spatial one, thereby creating four-dimensional (4D) electron microscopy which has opened up new vistas and applications as discussed here. Using timed, coherent single-electron pulses or electron packets, which are liberated from a photocathode with femtosecond laser pulses, structural phase transitions, mechanical deformations, and the embryonic stages of nucleation and crystallization are examples of phenomena that can be charted in unprecedented structural detail at a rate which is 10 orders of magnitude as fast as that previously attainable. Furthermore, because electrons are focusable and can be pulsed at femtosecond rates, and have appreciable inelastic cross-sections, the electron microscope yields information in four distinct ways: in real space, in reciprocal space, in energy space, and in the time domain. Thus, besides structural imaging, the energy landscapes of macromolecules may be explored, and, under optimal conditions, elemental compositions, valence-states bonding and 3D information (from tomography) may also be retrieved. Lastly, convergent-beam and energy-filtering variants of the technique provide nanoscale diffraction and near-field imaging.

Figure 1.6 depicts the space–time dimensions of TEM and ultrafast electron microscopy (UEM). The boundaries of time resolution are representative of the transition from the utilized video speed (millisecond) imaging in TEM, to fast or high-speed (nanosecond-to-microsecond) imaging, to the ultrafast (picosecond-to-femtosecond) imaging regime, as discussed in Chap. 5. The spatial resolution in the high-speed, nanosecond domain indicated in the figure is limited by electron–electron (space–charge) repulsion in nanosecond pulses of electrons. The UEM

landscape is that of single-electron imaging which, owing to the absence of interelectron repulsion, reaches the spatial resolution of TEM. Figure 1.7 displays a collection of images from examples of time-integrated electron microscopic studies. For UEM (discussed in Chaps. 5 and 6), the emphasis is on stroboscopic recording of structural, morphological, and electron dynamics, and examples are given in Fig. 1.8. A universal layout of the UEM microscope for operation in single-electron and single-pulse modes is provided in Fig. 1.9.

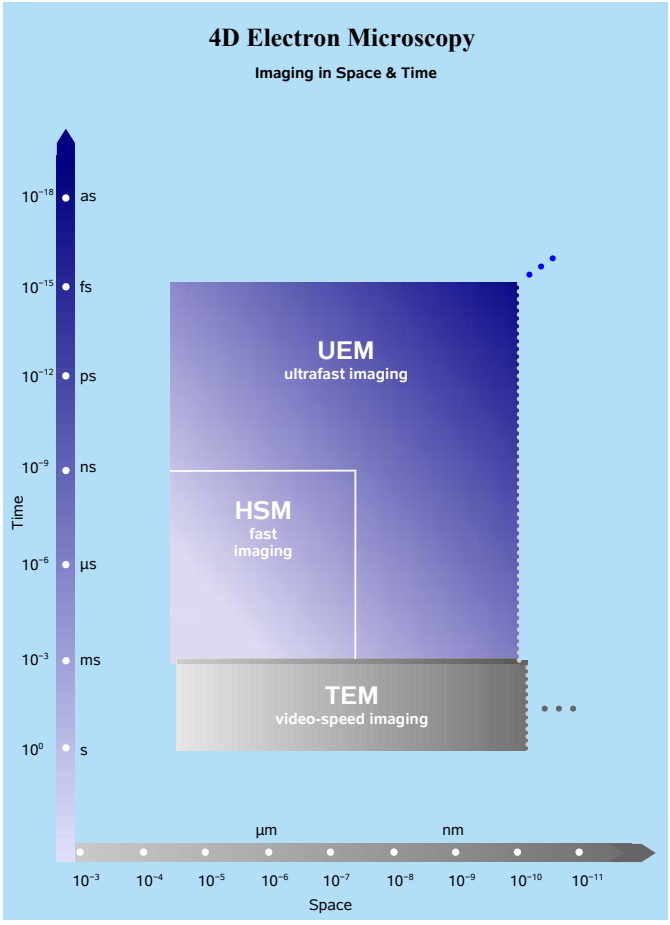


Figure 1.6 Resolutions in space and time achieved in electron microscopy. The focus here is on comparisons of UEM and TEM, but other variants of the techniques (scanning EM, tomography and holography, as well as electron spectroscopy) can similarly be considered. The horizontal dimension represents the spatial resolution achieved from the early years of EM to the era of aberration-corrected instruments. The vertical axis depicts the temporal resolution achieved up to the present and the projected extensions in the near future. The regions of fast and ultrafast time resolution are discussed in the text. Vertical dotted lines separate the spatial resolutions characteristic of real-space (microscopy) imaging from the spatial resolution that is obtainable using the reciprocal-space (diffraction) techniques that reach picometers. White solid lines define the region of high-speed microscopy (HSM) at the boundaries of 1 ns and 50 nm, as discussed in Chap. 5.

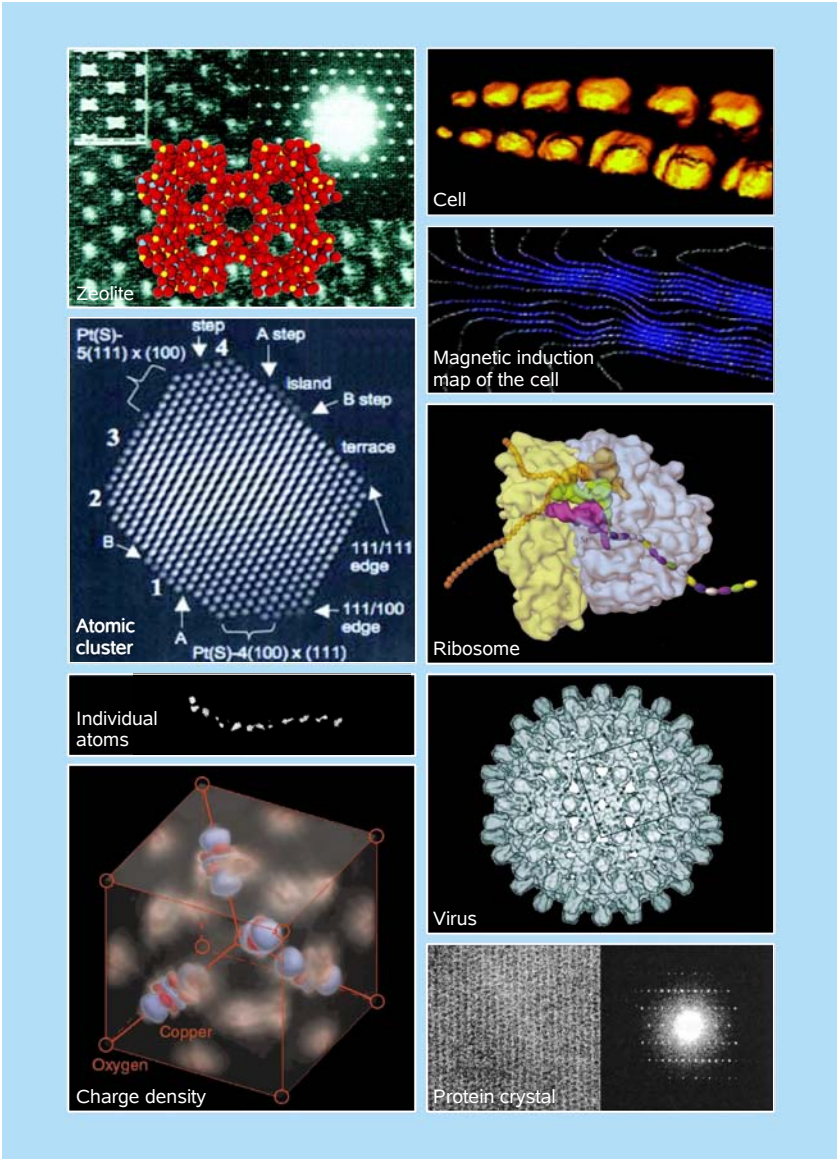


Figure 1.7 Examples of TEM images of materials and biological assemblies. Shown on the left are atomic-scale reconstruction of a nanoporous catalyst, CoALPO-36 (*Zeolite*; see Fig. 4.4 and Ref. 10 in Chap. 4); atomic surfaces of a Pt nanocrystal (*Atomic cluster*; see Fig. 4.2 and Ref. 6 in Chap. 4); separations between individual thorium atoms, which were observed for the first time by Crewe *et al*. (*Individual atoms*; see Ref. 9 in Chap. 8); and charge density maps in cuprite (Cu_2O) showing the difference between the total charge density measured by convergent beam electron diffraction and that of a collection of spherical ions obtained by calculations (*Charge density*; see Ref. 43 in Chap. 8). Shown on the right are tomographic reconstruction and magnetic induction map of a magnetosome chain (*Cell*; see Fig. 3.23 and Ref. 29 in Chap. 3); structural model of *Escherichia coli* ribosome based on cryo-electron microscopy data as obtained by Frank at about 1 nm resolution (*Ribosome*; see Ref. 44 in Chap. 8); 3D map of the hepatitis B core protein (*Virus*; see Fig. 4.20 and Ref. 69 in Chap. 4); and high-magnification image, and diffraction, of a catalase crystal as obtained by Glaeser and Taylor (*Protein crystal*; see Ref. 45 in Chap. 8).

No monograph known to us attempts to cover the revolutionary dimensions that EM in its various modes of operation nowadays makes possible. Our objective in this text is therefore to chart these developments and also to compare the merits of coherent electron waves with those of synchrotron radiation. In the next chapters,

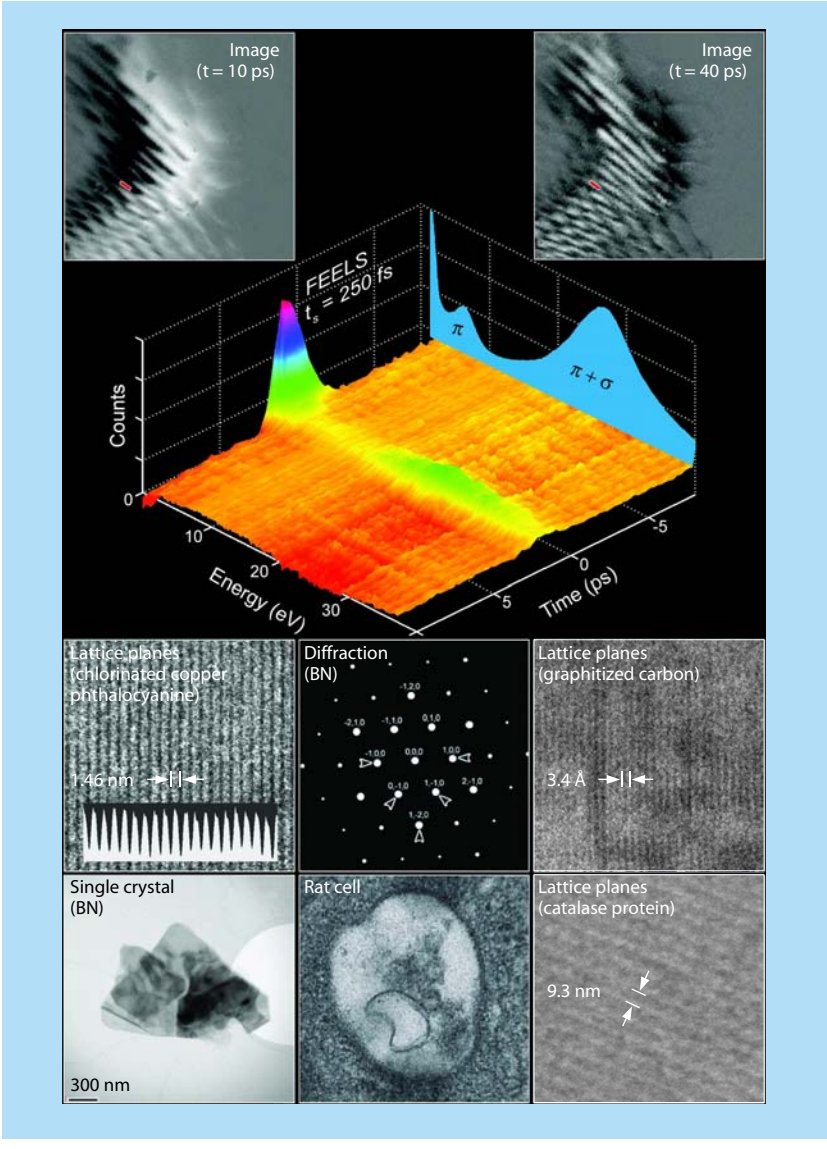


Figure 1.8 Examples of UEM images, diffraction patterns, and electron-energy-loss spectrum (EELS). Temporal evolution of images taken following atomic motions in graphite is illustrated at the top. A femtosecond-resolved EELS (FEELS) is depicted in the center. The spatial resolution characteristic of the chlorinated copper phthalocyanine, graphitized carbon, and catalase images suffices to observe the lattice atomic planes. Also shown are real-space images of a BN single crystal and a rat cell, and a typical diffraction pattern. Details are given in Chaps. 5 and 6.

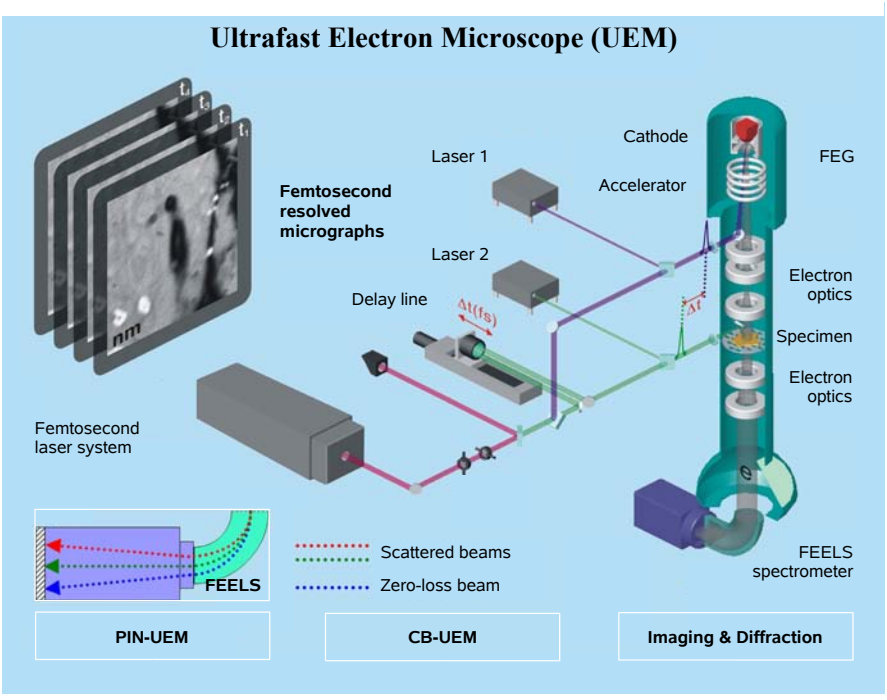


Figure 1.9 Universal layout for the ultrafast electron microscope (UEM). Shown are the components of the EM column, together with the femtosecond and nanosecond (1 and 2) laser systems and different optical components, which are integral parts of UEM. The probing electrons, which are generated by femtosecond laser pulses illuminating the photocathode, are further accelerated in a field-emission-gun-type arrangement. Optical delay assembly is the critical component for determining the temporal evolution by delaying the clocking laser pulse on the specimen (green beam) with respect to the probing electron pulse also on the specimen (gray beam); an optical path difference of 1 μm in the optical delay line provides a time delay (and hence a temporal resolution) of 3.3 fs. Micrographs are recorded with atomic-scale spatiotemporal resolution, whereas femtosecond-resolved electron-energy-loss spectra are obtained as 3D plots. Variant techniques are given: CB-UEM for the convergent-beam method, and PIN-UEM for the photon induced near-field technique, both discussed in Chap. 6.

we judge it prudent to recall some important basic procedural and theoretical concepts of imaging and diffraction using TEM and its scanning variant, STEM, so that we may better comprehend the significance of the new dimensions and the myriad applications now available. We then adumbrate these developments and the applications made possible in physical and biological domains. We deem it outside the scope of this monograph to discuss field-ion or scanning-tunneling microscopy, or the numerous recent advances and manifestations of optical microscopy.

Our aim, and hope, is that, by introducing the fourth dimension (4D) into imaging by electron microscopy, this book will stimulate research in many areas of science and technology. 4D imaging, together with the arsenals of different methods, such as tomography, holography, and electron-energy-loss spectroscopy (EELS), defines the state of progress over 70 years of developments (Fig. 1.5).

References

1. J. W. Strutt (later known as Lord Rayleigh), *Phil. Mag.*, **31**, 87 (1891).
2. R. A. Zsigmondy, *Nobel Lecture* (December 11, 1926).
3. H. Lichte, *Phil. Trans. R. Soc. Lond. A*, **360**, 897 (2002), and references therein.
4. (a) J. C. H. Spence, in "*Compendium of Quantum Physics: Concepts, Experiments, History and Philosophy*," Eds. D. Greenberger, K. Hentschel and F. Weinert, Springer, Berlin–Heidelberg (2009).
   (b) M. P. Silverman, W. Strange and J. C. H. Spence, *Am. J. Phys.*, **63**, 800 (1995).
5. (a) A. Tonomura, "*Electron Holography*," 2nd ed., Springer, Berlin–Heidelberg (1999).
   (b) A. Tonomura, "*The Quantum World Unveiled by Electron Waves*," World Scientific, Singapore (1998).
6. E. W. Müller, *Z. Physik*, **131**, 136 (1951).
7. T. T. Tsong, *Phys. Today*, **59**, 31 (2006).
8. J. M. Thomas, in "*Physical Biology: From Atoms to Medicine*," Ed. A. H. Zewail, Imperial College Press, London, p. 51 (2008).
9. (a) A. V. Crewe, J. Wall and J. Langmore, *Science*, **168**, 1338 (1970).
   (b) J. M. Thomas, *Nature*, **281**, 523 (1979).
10. P. D. Nellist, M. F. Chisholm, N. Dellby, O. L. Krivanek, M. F. Murfitt, Z. S. Szilagyi, A. R. Lupini, A. Borisevich, W. H. Sides, Jr. and S. J. Pennycook, *Science*, **305**, 1741 (2004).
11. J. M. Cowley, "*Diffraction Physics*," 3rd ed., Elsevier, Amsterdam (1995).
12. J. C. H. Spence, "*High-Resolution Electron Microscopy*," 3rd ed., Oxford University Press, New York (2003).
13. (a) "*Understanding Materials: A Festschrift for Sir Peter Hirsch*", Ed. C. J. Humphreys, Maney Publishing, London (2002).
    (b) Lucid reviews by P. W. Hawkes and A. Howie were recently published in a Special Issue on new possibilities with aberration-corrected electron microscopy. See: P. W. Hawkes, *Phil. Trans. R. Soc. A*, **367**, 3637 (2009); A. Howie, *Phil. Trans. R. Soc. A*, **367**, 3859 (2009).
14. A. Klug, *Nobel Lecture* (December 8, 1982).
15. R. Henderson, *Q. Rev. Biophys.*, **28**, 171 (1995).
16. R. A. Crowther, *Phil. Trans. R. Soc. B*, **363**, 2441 (2008).
17. (a) A. Sali, R. M. Glaeser, T. Earnest and W. Baumeister, *Nature*, **422**, 216 (2003).
    (b) R. M. Glaeser, *Proc. Natl. Acad. Sci. USA*, **105**, 1779 (2008).
18. R. M. Glaeser, K. Downing, D. DeRosier, W. Chiu and J. Frank, "*Electron Crystallography of Biological Macromolecules*," Oxford University Press, New York (2007).
19. J. Frank, "*Three-Dimensional Electron Microscopy of Macromolecular Assemblies: Visualization of Biological Molecules in Their Native State*," Oxford University Press, New York (2006).
20. (a) T. A. Klar, S. Jakobs, M. Dyba, A. Egner and S. W. Hell, *Proc. Natl. Acad. Sci. USA*, **97**, 8206 (2000).

(b) C. W. Freudiger, W. Min, B. G. Saar, S. Lu, G. R. Holtom, C. He, J. C. Tsai, J. X. Kang and X. S. Xie, *Science*, **322**, 1857 (2008).

21. A. H. Zewail, in "*Les Prix Nobel: The Nobel Prizes 1999*," Ed. T. Frängsmyr, Almqvist & Wiksell, Stockholm, p. 110 (2000), and references therein.
22. (a) A. I. Sabra in "*Inside the Camera Obscura: Optics and Art under the Spell of the Projected Image*," Ed. W. Lefèvre, Max-Planck-Institut für Wissenschaftsgeschichte, Berlin, p. 53 (2007).
    (b) A. I. Sabra, *Harvard Mag.*, **IX–X**, 54 (2003).
    (c) Ibn al-Haytham, "*Optics*" (Books I–III on Direct Vision translated with introduction and commentary by A. I. Sabra of Harvard University), Part 2, The Warburg Institute, London, p. lxxvi (1989).
23. B. Steffens, "*Ibn al-Haytham: First Scientist*," Morgan Reynolds Publishing, Greensboro (2006).
24. (a) D. C. Lindberg, "*Theories of Vision: From al-Kindi to Kepler*," University of Chicago Press, Chicago (1976).
    (b) A. I. Sabra, "*Theories of Light, from Descartes to Newton*," 2nd ed., Cambridge University Press, New York (1981).
    (c) J. R. Voelkel, "*Johannes Kepler and the New Astronomy*," Oxford University Press, Oxford (1999).
    (d) M. H. Morgan, "*Lost History: The Enduring Legacy of Muslim Scientists, Thinkers, and Artists*," National Geographic Press, Washington (2007).
25. R. Rashed, "*Geometry and Dioptrics in Classical Islam*," Al-Furqan Islamic Heritage Foundation, London (2005).
26. "*1001 Inventions: Muslim Heritage in Our World*," Eds. S. T. S. al-Hassani, E. Woodcock and R. Saoud, FSTC, Manchester (2006).
27. R. Hooke, "*Micrographia: Or Some Physiological Descriptions of Minute Bodies Made by Magnifying Glasses with Observations and Inquiries thereupon*", Royal Society, London (1665).

Chapter 2

Concepts of Coherence

Optics, Diffraction, and Imaging

The coherence of a wave field underpins its ability to produce interference effects and is a fundamental property associated with many phenomena — physical, chemical, and biological. From interference of light to the motion of atoms and to the function of biological machines, coherence is a key to our understanding of the underlying behavior. In many cases, its presence is readily apparent, but in numerous other cases experiments must be designed carefully in order to project coherence out of the underlying incoherent processes. Diffraction, holography, and lasers are examples of phenomena and tools that exist as a result of coherence. In imaging, the coherence property of specimen spatial waves is manipulated to enhance phase contrast, forming the basis of atomic-scale high-resolution electron microscopy. In addition, there can be no proper temporal imaging without consideration of coherence. It is therefore essential to address the origin of coherence and how to exploit its unique features for the benefit of optimum resolutions in structure and dynamics. Here, we shall limit ourselves to three relevant cases, those of (i) optical coherence in quantum systems (molecules and materials), (ii) diffraction and crystallography, and (iii) electron imaging. But first we provide a simple description to highlight the concepts and definitions.

2.1 Coherence — A Simplified Prelude

Coherence is a measure of interference in systems exhibiting wave or wave-like behavior. Two waves can interfere constructively or destructively depending on the relative phase between them, and the degree to which this interference is "visible" is a measure of the degree of coherence; in a more formal sense, it is the amplitude correlations in time and space of the waves involved.

Consider the simplest case of two waves with amplitudes ε_1 and ε_2, and only with time dependence:

$$\begin{aligned} E_1(t) &= \varepsilon_1 \exp(-i\varphi_1 t), \\ E_2(t) &= \varepsilon_2 \exp(-i\varphi_2 t). \end{aligned} \tag{2.1}$$

If wave one is detected, the intensity would be $I_1 = |E_1(t)|^2 = |\varepsilon_1|^2$, and similarly for wave two, $I_2 = |E_2(t)|^2 = |\varepsilon_2|^2$. However, if the two waves are simultaneously detected, then the total intensity would be

$$I(t) = |E_1(t) + E_2(t)|^2 = |\varepsilon_1|^2 + |\varepsilon_2|^2 + \varepsilon_1\varepsilon_2^* \exp(-i[\varphi_1 - \varphi_2]t) + \text{c.c.}, \tag{2.2}$$

where the complex conjugate (c.c.) is similar to the third term except for the exponential sign. Taking $\varepsilon_1\varepsilon_2^*$ to equal $\varepsilon_1^*\varepsilon_2$, we now have a simple expression,

$$I(t) = |\varepsilon_1|^2 + |\varepsilon_2|^2 + 2|\varepsilon_1\varepsilon_2| \cos([\varphi_1 - \varphi_2]t) \tag{2.3}$$

It follows that in addition to the expected first two terms, there is an additional term which describes the interference of the two waves and the presence of which is critical to the degree of coherence: when $\varphi_1 - \varphi_2 = 0$, the total intensity becomes larger than the sum of each wave independently (for $\varepsilon_1 = \varepsilon_2 = \varepsilon, I = 4|\varepsilon|^2$, four times as large as for either wave); when $\varphi_1 - \varphi_2 = 90°$, the total intensity is the sum of the individual two waves. In the former case, the component waves are in phase (constructive interference) while for the latter they are out of phase (destructive interference). The degree of coherence tells us the extent of this correlation or perfection in interference and will depend on so-called beating between different frequencies, the recording time and other factors that are discussed later.

The correlation in time is characterized as *temporal coherence* whereas in space it is that of *spatial coherence*.[1] We can assign in the simplest case a time constant for these correlations. For the temporal coherence, the average correlation function for an observable (Ω), say, intensity, at different times t and t' is defined by

$$\langle \Omega(t)\Omega(t') \rangle \sim \exp(-t/\tau_c), \tag{2.4}$$

and the time constant τ_c is the coherence time, that is, the time ($t = \tau_c$) over which the phase or amplitude changes significantly, from the value at $t = 0$ and where the waves were in phase (Fig. 2.1). The coherence length (l_c) is simply defined as the distance traveled in the time τ_c; at a speed υ,

$$l_c = \upsilon\tau_c, \text{ or } c\tau_c \text{ (for light)}. \tag{2.5}$$

Several noteworthy points now arise. For a monochromatic wave, τ_c, in principle, is "infinitely" long, while for pulses (a group of waves), τ_c is finite. Because the frequency bandwidth ($\Delta\nu$) of a pulse is a measure of the extent of group-frequency differences, the larger the frequency bandwidth, the shorter τ_c. From Fourier analysis, it can be shown that

$$\Delta\nu\tau_c \sim 1, \tag{2.6}$$

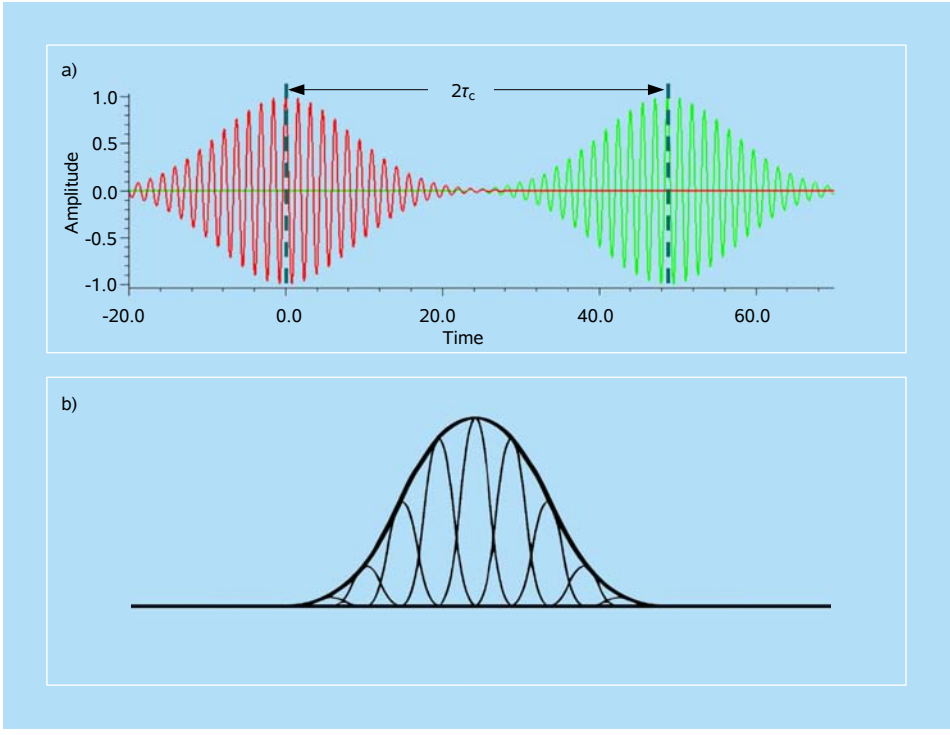


Figure 2.1 Wave packets and spatial distributions. (a) A wave packet whose amplitude changes significantly in time τ_c (red) and a copy of the same wave delayed by $2\tau_c$ (green). The correlation in time is indicated by $2\tau_c$ as noted by amplitude and phase changes. Adapted from *wikipedia.org*. (b) Homogeneous and inhomogeneous broadenings such as those existing in gases (Doppler effect) and solids (due to defects and imperfections).[2]

and consequently

$$\left(\frac{\Delta\lambda}{\lambda^2}\right) l_c \sim 1. \tag{2.7}$$

A continuous wave (CW) laser has l_c on the meter scale, whereas the coherence length of a 10 fs light pulse is 3 μm. If the pulse duration in time is the same as τ_c, the pulse is said to be (uncertainty) *transform-limited* or *coherent*, and all frequency components are intrinsically homogeneous; but when the pulse is inhomogeneously-broadened, τ_c is less than the temporal width (Δt_p) of the pulse (Fig. 2.1). For transform-limited pulses with well-defined shapes, the widths in frequency ($\Delta\nu$) and time (Δt_p) domains are quantitatively expressed as

$$\begin{aligned} \Delta\nu\Delta t_p &= 0.441 \text{ (Gaussian)}, \\ \Delta\nu\Delta t_p &= 0.315 \text{ (sech}^2\text{)}. \end{aligned} \tag{2.8}$$

White light has therefore very short τ_c because of the large spread of frequencies involved. For electrons, because of their finite de Broglie wavelength, their wave-like property defines coherence, temporal and spatial, as we shall discuss below.

The coherence features of waves discussed above are relevant when considering coherence in matter, as defined in the following section for two distinct states of matter and in Chap. 5 for atomic-scale resolution of motion; further discussion of coherence in optics can be found in Ref. 1.

2.2 Optical Coherence and Decoherence

In this section, we are concerned with interference that can be created in matter using optical excitation by, for example, lasers, a process common to many methods of coherence creation and probing. The case discussed is that of a two-level system.[2] When a laser interacts with an ensemble of atoms or molecules, under certain conditions, one can create a coherent linear combination of the two levels, say ground (*s*-like) and excited (*p*-like) wave functions (Fig. 2.2). This coherent state, or what is sometimes called a pure state, possesses a definite phase

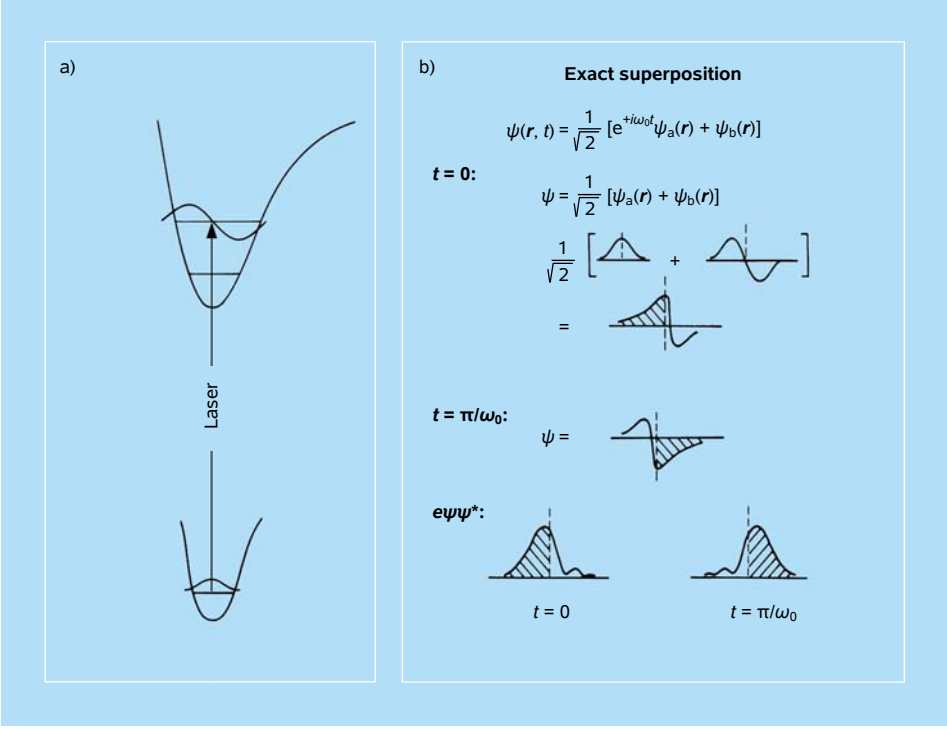


Figure 2.2 A schematic description of optical coherence in a two-state system. An *s*-type state function that is coherently coupled by a laser to a *p*-type excited state is shown. The superposition is displayed on the right as additions of the functions and densities involved.[3]

relationship between the wave function of the ground and that of the excited state. Due to interactions of the optically excited atoms or molecules among themselves or with the environment, the coherent state can become incoherent. In other words, the well-defined superposition of states created by the laser becomes random or statistical in nature. Randomization of phase coherence in time is termed *dephasing* or *decoherence*, and the processes responsible for the phenomenon are the dephasing or decoherence mechanisms.

For molecules, measurements of optical coherence[3] require the coherent preparation of a set of molecules in a given vibronic or rovibronic state, and the probing of the prepared coherent state with the passage of time and as the molecules dephase by, for example, collisions with other molecules (say solvent) or energy flow to other states. We consider, for simplicity, the semiclassical approach[2]: the molecule is treated quantum mechanically, and the laser field, classically. Semiclassically, we describe the phenomenon as follows: the laser field E interacts with an ensemble of molecules to produce a time-dependent polarization, $P(t)$, which in turn changes as the molecules dephase. So, our task now is to find how the polarization is related to coherence and what coherence means on the molecular level.

The two vibronic states are the ground state, $\psi_a(\boldsymbol{r})$, and the excited state, $\psi_b(\boldsymbol{r})$, and the laser field is simply a wave (propagation direction, z) of the form:

$$E(z,t) = \varepsilon \cos(\omega t - kz) = \frac{1}{2}\varepsilon\left[\exp(i[\omega t - kz]) + \exp(-i[\omega t - kz])\right], \tag{2.9}$$

where ε is the amplitude and ω is the frequency of the radiation. The state of the molecule (driven by the laser) at time t may be represented as

$$\psi(\boldsymbol{r},t) = a(t)\exp(-i\omega_a t)\psi_a(\boldsymbol{r}) + b(t)\exp(-i\omega_b t)\,\psi_b(\boldsymbol{r}), \tag{2.10}$$

the analogue of the two-wave superposition that we described earlier. We can now calculate the time-dependent molecular polarization:

$$\begin{aligned} P_m(t) &= \langle \psi(\boldsymbol{r},t)|\bar{\mu}|\psi(\boldsymbol{r},t)\rangle \\ &= ab^*\mu_{ba}\exp(-i[\omega_a - \omega_b]t) + a^*b\mu_{ab}\exp(-i[\omega_b - \omega_a]t), \end{aligned} \tag{2.11}$$

where $\bar{\mu}$ is the dipole-moment operator and μ_{ba} and μ_{ab} are the transition moment matrix elements. Taking these matrix elements to be equal ($\equiv\mu$) and setting $\omega_b - \omega_a = \omega_0$, the transition frequency, we obtain

$$P_m(t) = \mu[ab^*\exp(+i\omega_0 t) + a^*b\exp(-i\omega_0 t)]. \tag{2.12}$$

Again, as discussed above, the cross terms define the coherent state and the induced polarization. Hence, the polarization, which is related to the radiation power, is zero if there is *no* coherent superposition or, in other words, if the molecule is certain to be in the state $|a\rangle$ or $|b\rangle$. As shown in Fig. 2.2, as a result of the superposition of

states the charge in a volume element $d\upsilon$, $e\psi\psi^* d\upsilon$, oscillates coherently between two distributions and the system is nonstationary.

From Eq. (2.12), the total polarization for N molecules in the sample (assuming equal contribution and ignoring propagation effects) is therefore

$$P(t) = \frac{1}{2}[\bar{P}\exp(+i\omega_0 t) + \bar{P}^*\exp(-i\omega_0 t)] = N\mu\,[\rho_{\mathrm{ab}} + \rho_{\mathrm{ba}}], \tag{2.13}$$

where $P(t) \equiv NP_{\mathrm{m}}(t)$ and $\bar{P}$ is its complex amplitude, i.e., $\bar{P} = \bar{P}_{\mathrm{real}} + i\bar{P}_{\mathrm{imag}}$. We choose the notation ρ_{ab} and ρ_{ba} for the cross terms $ab^*\exp(+i\omega_0 t)$ and $a^*b\exp(-i\omega_0 t)$ because they are indeed the off-diagonal elements of the ensemble density matrix ρ, as shown below; for an introduction to density matrices, see, for example, Ref. 2.

So far, all that we have shown is that to create an optical coherence or polarization we need a nonvanishing interference term or equivalently the off-diagonal elements of ρ must be nonzero in the zero-order basis set of $|a\rangle$ or $|b\rangle$. Quantum mechanically, one can calculate these coherence terms and, for the most experimental descriptions, one can perform a rotating coordinate frame analysis for the components of P. Now that we have obtained the polarization, the resultant coherent field in the sample can be found using the following self-consistent prescription:

$$\text{laser field} + \text{molecules} \rightarrow \text{polarization, } P \rightarrow \text{sample field}. \tag{2.14}$$

Basically, P can be calculated by using Eq. (2.13), and Maxwell's equations can be used with $\bar{P}$ being the source term to calculate the resultant sample field amplitude $\bar{\varepsilon}$. It is through $\bar{P}$ (or $\bar{\varepsilon}$) that we can monitor the nonlinear optical behavior of the sample and hence the changes in the rate of optical dephasing or decoherence.

From Eq. (2.10) we see that the probability of finding the system in the excited (ground) state is simply $|b|^2$ ($|a|^2$). These probabilities decay by time constants, say $T_{1\mathrm{b}}$ and $T_{1\mathrm{a}}$, respectively. Such phenomenological decay is the result of the Wigner–Weisskopf approximation, i.e., an exponentional decay of the amplitudes a and b; a or $b \sim \exp(-t/2T_1)$. The cross terms of Eq. (2.10) also decay, possibly by a different rate from the diagonal terms.[4] Hence, the ensemble density matrix can now be written as

$$\begin{gathered} |a\rangle = \psi_{\mathrm{a}}(\boldsymbol{r}), \qquad |b\rangle = \psi_{\mathrm{b}}(\boldsymbol{r}) \\ \begin{matrix} \langle a| \\ \langle b| \end{matrix} \begin{bmatrix} |a_0|^2 e^{-t/T_{1\mathrm{a}}} & \left(a_0 b_0^* e^{+i\omega_0 t}\right) e^{-t/T_2} \\ \mathrm{cc} & |b_0|^2 e^{-t/T_{1\mathrm{b}}} \end{bmatrix} \end{gathered} \tag{2.15}$$

where

$$\frac{1}{T_2} = \frac{1}{T_2'} + \frac{1}{2}\left(\frac{1}{T_{1\mathrm{a}}} + \frac{1}{T_{1\mathrm{b}}}\right). \tag{2.16}$$

The T_1 terms in Eq. (2.16) come from the diagonal elements and represent an average rate for the loss of population in the *a*, *b* levels. Physically, the T_2' term represents the additional decay caused by phase changes in the cross terms.[4] In other words, the random and rapid variation in $\omega_0(t)$ causes the off-diagonal elements to decay faster than the diagonal ones. The total dephasing rate is therefore $1/T_2$: it contains $(T_2')^{-1}$, the rate for phase coherence loss (pure dephasing), and $(T_1)^{-1}$, the rate for irreversible loss of population in the two levels. The phenomenology described here is the optical analogue of magnetic resonance T_1 and T_2 from Bloch's equations,[5] but the physics is different.[4]

If we monitor the population, say from the *b* state by emission, it is clear now that all we can measure is T_{1b}. Furthermore, if the ensemble is homogeneous (i.e., consists of only those molecules that follow the uncertainty relationship $\Delta\nu T_2 = \pi^{-1}$ for a Lorentzian line shape), then an absorption experiment will give T_2 and an emission experiment will give T_1, and one can deduce the value of T_2'. In general, however, optical transitions are inhomogeneously broadened (IB), implying that the homogeneous ensemble is a subensemble of a grand ensemble as, for example, in the case of Doppler broadening in gases or atoms and molecules in materials. In frequency-domain experiments, if the homogeneous broadening (HB) is very small, it is difficult to measure it through absorption methods. However, in time-resolved experiments the narrower the resonance, the longer the decay time, and it is relatively easy to measure T_2. It is this ability to separate T_2 (T_1 and T_2') from IB effects that makes coherent spectroscopy a useful technique for unraveling dynamical (optical) processes in matter.

The coherent state must be created on a time scale that is shorter than that of decoherence. Under such condition it is useful to express and calculate the amplitudes responsible for coherence. For the case at hand, one can quite easily obtain *a* and *b* in Eq. (2.10) using the time-dependent Schrödinger equation. On resonance, i.e., when $\omega = \omega_0$, and in the absence of relaxation (decoherences):

$$a(t) = \cos\left[\frac{1}{2}\left(\frac{\mu\varepsilon}{\hbar}\right)t\right] \equiv \cos\frac{\omega_R t}{2}, \tag{2.17}$$

$$b(t) = -i\sin\left[\frac{1}{2}\left(\frac{\mu\varepsilon}{\hbar}\right)t\right] \equiv -i\sin\frac{\omega_R t}{2}, \tag{2.18}$$

where ω_R is the well-known Rabi frequency ($\mu\varepsilon/\hbar$). Note that the total Hamiltonian now contains the molecule–radiation interaction Hamiltonian ($\boldsymbol{\mu}\cdot\boldsymbol{E}$). From Eqs. (2.17), (2.18), and (2.13), the polarization is maximum when $\omega_R t = \pi/2$, and that is why, in practice, $\pi/2$ optical pulses are very desirable. At time $t = \pi/(2\omega_R)$, P_m becomes simple: $P_m = -\mu\sin\omega_0 t$. This polarization is created out-of-phase with the applied field, $\varepsilon\cos\omega t$, as it should. We now suppose that we are starting with N_b molecules per unit volume in state $|b\rangle$ and N_a in state $|a\rangle$. If the

exciting field is turned off at $t = \pi/(2\omega_R)$, then the net macroscopic polarization, or radiating dipole moment per unit volume, is

$$P = -(N_a - N_b)\mu \sin \omega_0 t \tag{2.19}$$

Hence, the net radiated power per unit volume (time averaged) becomes

$$\left\langle E_c(t)\frac{dP}{dt}\right\rangle = \frac{1}{2}(N_b - N_a)\omega_0(\mu\bar{\varepsilon}), \tag{2.20}$$

where $\bar{\varepsilon}$ is the field amplitude created by the radiating molecules since the external field is turned off.

For the coherent coupling to take place between the molecules and the applied field, the inequality $\omega_R > 1/T_2$ must hold. Near the coherent limit we can therefore estimate the laser field amplitude needed, provided we know the transition width or T_2. For a transition with a moment of 3 debyes and a width of $1\ \text{cm}^{-1}$, $\varepsilon \approx 10^4$ V/cm or $\approx 10^5\ \text{W/cm}^2$, which nowadays can easily be achieved. If the radiated field is $10^{-3}\varepsilon$, then the radiated power $\approx 13\ \text{mW/cm}^2$, a flux readily handled by conventional detectors. The example given shows that optical coherence can be created, detected, and used to obtain the decoherence rates involved.

2.3 Coherence in Diffraction

2.3.1 Rayleigh criterion and resolution

Diffraction is usually discussed by making the analogy with the renowned double-slit experiment of Thomas Young. Many texts deal with this issue. Here, the objective is to state key points relating to diffraction and to emphasize the role of phase coherence and limits of resolution.

For a single slit of width w, the diffraction pattern, in the Fraunhofer limit (when the distance to the detector, L, is much larger than w), is described by a sinc function $[\text{sinc}(\beta) = \sin(\beta)/(\beta)]$; see Fig. 2.3. The intensity becomes (see also Sec. 2.5)

$$I(x) = I_0 \,\text{sinc}^2(\pi x w/\lambda L). \tag{2.21}$$

The function reaches its maximum at $x = 0$ $[\sin(0)/0 = 1]$, and I becomes I_0. The first minimum occurs at $x = \lambda L/w$, with $\sin(\pi)/\pi = 0$ (Fig. 2.3). For multiple slits with width w and separation between slits d, the region between $+\lambda L/w$ and $-\lambda L/w$ further resolves into a pattern of peaks, where the minima are separated by $\lambda L/d$, when $w < d$. (This is why a diffraction grating is capable of separating colors with high resolution.) The above conditions for $I(x)$ can be obtained either from simple geometric optics or using Fourier analysis. For the former, one considers the path lengths involved in the interferences, destructive

and constructive, of wavelets originating from the upper and lower halves ($w/2$) of the slit; Fig. 2.3. In Fourier analysis, the light would be considered to be passing through the slit as a single square pulse.

When a two-dimensional circular aperture is considered, a similar treatment can be implemented. Unlike the one-dimensional slit case, the circular aperture gives rise to a central bright spot and surrounding rings, the Airy disc. The expression, which was derived by the astronomer George Airy in 1835, can be derived from the Fourier transform of a circular aperture. It is more complicated than that of a square one-dimensional slit, and the final intensity is given by $2J_1(x)/x$, where J_1 is the Bessel function of the first kind. The central maximum of this function extends to the first minimum located at $x = 1.22\pi$ instead of $x = \pi$ in the case of the sinc function.

In many applications involving diffraction there are a few concepts that are essential to the understanding of resolution and explain remarkably well how we see through lenses, with telescopes and microscopes, and with our complex eyes. For this reason, we shall first discuss here a simple geometric optics consideration for a single slit. In Fig. 2.3, the path difference at the first destructive interference, i.e., $\lambda/2$, is displayed; we note that the slit is divided into the upper and lower halves so that interferences of wavelets are consistently throughout the w length. It follows that for the first minimum

$$l_2 - l_1 = (w/2)\sin\Theta \equiv \lambda/2, \tag{2.22}$$

or

$$\sin\Theta = \lambda/w. \tag{2.23}$$

This simple equation is central to many applications. It gives the "angular resolution" due to a slit of width w. The angle Θ is a "constant of the system" — depending on the distance L of observation, x increases accordingly. Moreover, the condition for diffraction is now clear: a large slit ($w \gg \lambda$) will almost provide the geometrical shadow with no appreciable diffraction; when $w \approx \lambda$, diffraction (fringes) is observed; and when $w \ll \lambda$, diffraction disappears but the intensity in the full domain will fall off at large angles.[6] Another expression that is worth mentioning is the one derived from Eq. (2.23) for the limit of small Θ, i.e., when $\sin\Theta \approx \tan\Theta \approx \Theta$ (Fig. 2.3):

$$x = \lambda L/w \quad \text{(small } \Theta \text{ limit)}. \tag{2.24}$$

Equations (2.23) and (2.24) will be used throughout and have analogy to Bragg's diffraction [Eq. (2.23)] and to the lateral coherence length [Eq. (2.24)], both of which are discussed below.

A commonly used criterion for resolution is that of Lord Rayleigh. In connection with lenses, one is interested in their resolving power, which is ultimately limited by diffraction (the spread of the Airy disc). The lens aperture

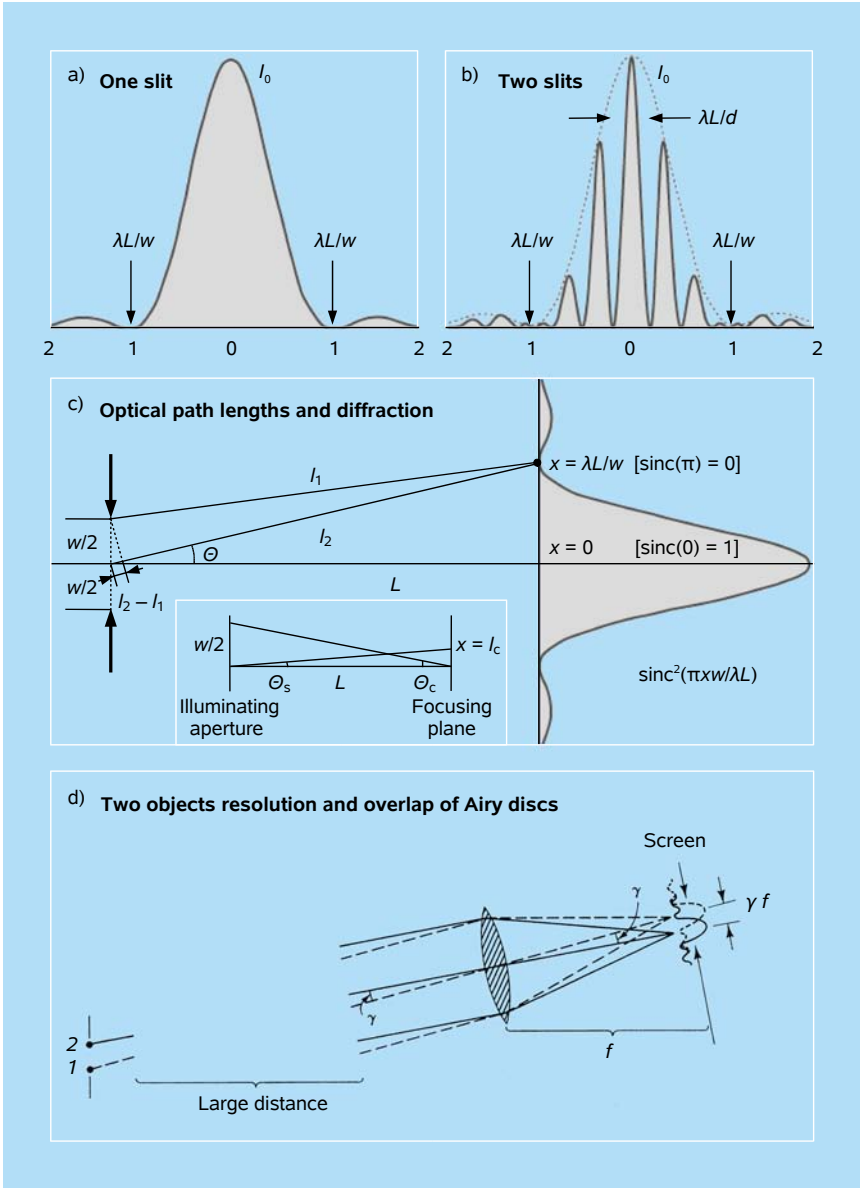


Figure 2.3 Concepts pertinent to diffraction. Panels (a,b) describe one- and two-slit diffraction. Ray diagram shown in panel (c) displays the difference in optical pathways and the broadening (and interferences) expressed by the Airy function (when 2D diffraction is considered). Panel (d) describes the resolution which is determined by the overlap of two Airy functions,[6a] when objects 1 and 2 are observed from far away distances. We note that, from the lens to the screen, the resolution is determined by the lens aperture and the distance to the screen which, in this case, is the focal length. Note also that γ is the angular separation of the two objects and this makes possible the determination of the separation between objects 1 and 2.

is analogous to the circular orifice equation, and, because of wavelet interference, diffraction blurs the images. Then the image of even an ideal point object of infinitesimal width is broadened as shown in Fig. 2.3c, and this distribution cannot be made smaller without using a larger lens. Rayleigh's criterion states that two overlapping Airy-disc images are just resolvable when the central maximum of one coincides with the first minimum of the other, giving

$$\sin\Theta = 1.22\lambda/D \qquad \text{(Rayleigh criterion)}, \tag{2.25}$$

where D is the diameter of the lens aperture. As discussed later, Rayleigh obtained this result only for the case where the intensities of images of point-like objects (such as those of a binary star) were added together at the detector plane. (His original paper also treats the coherent imaging case.) Sometimes, the aperture radius is used (the analogue of the upper and lower halves of the slit), and the numerical factor in Rayleigh criterion becomes 0.61. Also, the refractive index n should be included for the path difference, i.e., replacing λ by λ/n.

We now consider several cases of interest; for these, the outcome is determined by the limit of diffraction, the result of the wave nature of the particles involved. First, for an ideal lens (aperture) with focal length f, the minimum spatial resolution $[\Delta x = f\tan\Theta \approx f\sin\Theta]$ becomes $\Delta x = 1.22 f\lambda/D$. For a laser beam of $D = 1$ mm (the laser beam diameter on the surface of the lens), $f = 10$ cm, and $\lambda = 500$ nm, the diffraction-limited $\Delta x \approx 60$ μm. We note that better limits can be obtained by decreasing f, expanding the beam diameter (D), and shortening wavelength. For the lens, we assume that the phase fronts are parallel. If it were not for diffraction, the beam would be forced to a point, but because of diffraction there is a spread Δx, and similarly Δy — the focus is said to be diffraction-limited.

In other applications, it is relevant to consider parallel light from two sources, with a lens to focus the light from the two sources on a screen. A telescope is a prototype example, as it is a question of how to resolve the images of two closely spaced objects that are located far away using a lens (mirror) system; the detector is a camera (or our eyes). The small angle γ between the two (nearly parallel) sources results in two spots in the diffraction focal plane behind the lens.[6] Using the Rayleigh criterion, the angular resolution γ [rad] becomes $1.22\lambda/D$. From this simple consideration we can understand why telescopes must possess large lenses (mirrors). The Mount Palomar mirror is 5.08 m (200 inches) in diameter, and for light of $\lambda = 500$ nm, the angular resolution required to distinguish two objects becomes $\gamma = 1.2\times10^{-7}$ rad (0.69×10^{-5} degrees; 0.025 s); the resolution is sometimes expressed in arcseconds, which is equivalent to 4.8 μrad. For a distance of 10,000 km, two objects 1.2 m apart can be resolved. [Sirius, the brightest star in the night sky, is 8.6 light years away, and with a 10 m (Keck) telescope, millions-of-kilometer separations may be resolved.] With the new 30 m (Caltech – UC) telescope, the resolution will improve by a factor of three and, as importantly, the

sensitivity will increase by an order of magnitude because of the larger area of collection. (Another calculation of interest is that of the coherence length of rays coming from the Sun to the Earth; see Sec. 2.5.2.)

On the human scale, the Rayleigh criterion can predict the diffraction limit of our vision. The pupil, the opening governing the diffraction, may change its diameter from bright light (~1 mm) to darkness (~8 mm). For a diameter of 8 mm at $\lambda = 500$ nm, the diffraction-limited angular resolution is therefore 0.6×10^{-4} rad taking the refractive index of the eye's fluid to be 1.33. The letter "E" on the Snellen chart is 88 mm at 20/200 (the 20/20 line would have letters of 8.8 mm height), thus giving at 6 m away (without diffraction) $\tan\Theta \approx \sin\Theta \approx \Theta \approx 1.47\times10^{-3}$ rad. A remarkable result: 20/20 is about 20 times the Rayleigh diffraction limit. Our vision is not limited by diffraction — the most acute vision is characterized by an angular resolution of nearly 2×10^{-4} rad.

Finally, we discuss the microscope for which we wish to know the size of the smallest resolvable object, and hence its diffraction-limited resolution. The most significant element in determining the resolution is the objective lens, but in this case the objective lens is close to the object, and the Fraunhofer diffraction criterion, although used, may not be strictly applicable. What is important is to consider differences in path lengths[6,7] when the two objects to be resolved are separated by a distance $d_{\min}$. In terms of the Rayleigh criterion, when considering differences in path lengths from the two objects to the lens limits (hence the aperture), one can relate $d_{\min}$ to the aperture angle of the lens α (determined by the radius of the lens' aperture and distance from the lens to the object plane). We note that, in the microscope, there is no further path differences introduced at or beyond the lens, as the lens focuses the rays in the focal plane. Accordingly, the minimum resolvable separation of incoherently illuminated objects (points) becomes

$$d_{\min} = \frac{0.61}{n}\frac{\lambda}{\sin\alpha} \tag{2.26}$$

For a high-quality lens, $\alpha \approx 70°$, and for $\lambda = 450$ nm and $n = 1.56$, $d_{\min}$ reaches ~200 nm. The term $n\sin\alpha$ is called the numerical aperture (NA)

It is important to realize that the Rayleigh criterion of resolution is only applicable to *incoherent* sources of illumination. In other words, it is assumed that the source emission is from different points (and of different energies) with no phase correlation. For each source point and each energy value, the detected interferences, which represent the incoherent sum of intensities, display patterns of certain lateral position and period, as discussed below. Coherent sources are characterized by a different resolution criterion as pointed out by Goodman (see also Fig. 6.19 in Ref. 6).[6] Thus, the Rayleigh criterion, which was introduced for distinguishing between images of two incoherent point sources, is not relevant to

the phase-contrast imaging we will be dealing with in bright-field high-resolution transmission electron microscopy (HRTEM). It may be useful, however, for dark-field scanning transmission electron microscopy (STEM), which is also capable of atomic resolution.

2.3.2 Diffraction from atoms and molecules

In diffraction from matter, the atoms (or atomic planes) are the "slits," and the interference is what gives rise to the observed patterns. It is instructive to consider the origin of interferences and address features of the elastic scatterings involved. For electrons, the subject of focus here, an incident plane wave of constant-phase front, propagating along the z-direction, has the form

$$\psi_i = A\exp(i\boldsymbol{k}_0\cdot\boldsymbol{z}) = A\exp(ik_0 z), \tag{2.27}$$

$$k_0 = |\boldsymbol{k}_0| = 2\pi/\lambda_e = [2m_e c^2 E_k + E_k^2]^{1/2}/\hbar c, \tag{2.28}$$

where k_0, the absolute value of the wave vector $\boldsymbol{k}_0$, is given by the relativistic expression for the de Broglie wavelength (λ_e) at kinetic energy (E_k); c is the speed of light, and A is the wave amplitude (the intensity is $|A|^2$). The scattered (modified spherical) wave in the direction $\boldsymbol{R}$, which is at an angle Θ with respect to z, is given by

$$\psi_s = \frac{A}{R} f(\Theta)\exp(ikR). \tag{2.29}$$

We note that $|\boldsymbol{k}| = |\boldsymbol{k}_0|$ because we are interested only in elastic scattering, for which no change in the magnitude of k occurs, but the direction of the wave vector does change ($\Theta \geq 0$). The scattering amplitude at Θ is $f(\Theta)$ whose square, $|f(\Theta)|^2$, is the differential cross-section $d\sigma$ for a unit solid angle $d\Omega$. The expression is valid at long large values of R, and the $1/R$ term in Eq. (2.29) ensures the attenuation of electron density as $R \to \infty$. The differential scattered intensity thus becomes

$$I = \left(\frac{I_0}{R^2}\right) f^2(\Theta). \tag{2.30}$$

The task is to find $f(\Theta)$, and this is done in many texts by assuming certain forms of the potential and solving the Schrödinger equation, typically using a Green's function and invoking the first Born approximation, assuming the scattered wave amplitude to be very small when compared with that of the incident wave (single-scattering events).[8,9]

The solution gives, after some mathematical manipulation,

$$\psi_s(\boldsymbol{R}) = \frac{A}{R}\exp(ikR) f(s)\exp(i\boldsymbol{s}\cdot\boldsymbol{r}_0), \tag{2.31}$$

where $\boldsymbol{s}$ is simply the momentum transfer vector,

$$\boldsymbol{s} = \boldsymbol{k}_0 - \boldsymbol{k} \tag{2.32}$$

such that

$$|\boldsymbol{s}| = s = 2k_0 \sin\left(\frac{\Theta}{2}\right) = \frac{4\pi}{\lambda_e}\sin\left(\frac{\Theta}{2}\right), \tag{2.33}$$

and the scattering amplitude (in units of length) in s-space is given by

$$f(s) = \frac{2m_e e^2}{\hbar^2}\left(\frac{Z - f_x(s)}{s^2}\right); \tag{2.34}$$

$$f_x(s) = 4\pi \int_0^\infty r\rho(r)\frac{\sin(sr)}{s}dr\,, \tag{2.35}$$

where $\rho(r)$ is the electron density of the atom. Thus, for a single atom interacting with electrons (plane waves), and displaced from the origin by $\boldsymbol{r}_0$, the elastic scattering amplitude depends on Z and decreases with scattering angle by s^2. The (dimensionless) X-ray atomic scattering factors, $f_x(s)$, are tabulated for different atoms.

Coherence comes in when we consider more than one atom in molecules or atomic planes in crystals. Taking these atoms or planes to be independent, it is straightforward to write

$$\psi_s(\boldsymbol{R}) = \frac{A}{R}\exp(ikR)\sum_{i=1}^{N} f_i(s)\exp(i\boldsymbol{s}\cdot\boldsymbol{r}_i) \tag{2.36}$$

with the sum being over all atoms. The intensity is thus the product of two series that can be reindexed as a double summation:

$$\begin{aligned} I(s) &= \frac{I_0}{R^2}\sum_{i=1}^{N}\sum_{j=1}^{N} f_i(s)f_j(s)\exp(i\boldsymbol{s}\cdot[\boldsymbol{r}_i - \boldsymbol{r}_j]) \\ &\equiv \frac{I_0}{R^2}\left[\sum_{i=1}^{N} f_i^2(s) + \sum_{i=1}^{N}\sum_{j=1}^{N} f_i(s)f_j(s)\exp(i\boldsymbol{s}\cdot\boldsymbol{r}_{ij})\right] \qquad (i \neq j)\,. \end{aligned} \tag{2.37}$$

Equation (2.37) displays important features of scattering and diffraction, and it is worth emphasizing the following points. First, the analogy with "diagonal" and "off-diagonal" terms, or with "incoherent" and "coherent" contributions, discussed in the first sections of this chapter, is clear. The first term in Eq. (2.37), f_i^2, is interference-free while the second term is the coherent one and in fact has all the information about the positions of atoms (r_{ij}) and hence the structure; it also contains both the amplitudes and the phases. Second, because f_i^2 is an "incoherent" contribution, it is usually subtracted along with any inelastic contribution. Third,

in order to connect with observables, the coherence term must be related to the material phase in which the experiment is carried out. For example, in isolated gas-phase electron diffraction from an isotropically oriented collection of molecules, the $\exp(i\boldsymbol{s}\cdot\boldsymbol{r}_{ij})$ is averaged over all orientations with an isotropic distribution $P(\alpha)d\alpha$, resulting in

$$\left\langle \exp(i\boldsymbol{s}\cdot\boldsymbol{r}_{ij})\right\rangle_{\text{spatial}} = \frac{\sin sr_{ij}}{sr_{ij}}. \tag{2.38}$$

Similarly, if the atoms are vibrating relative to one another, an average (over their motion) can be made, resulting in broadening (the analogue of T_2 damping), and lastly, if atoms have large Z, the atom-dependent phase shift (second Born approximation or method of partial waves) has to be included, especially for scattering in the gas phase.

Many books and reviews deal with such issues, but the objective here is to clarify the physics of coherence and interference and the possibilities and ideas that could emerge, beyond what is usually assumed. For example, recognizing that randomly-oriented samples can be made to "orient" (or more accurately align) by a laser field led to "oriented diffraction patterns" which can be utilized to enhance interferences of certain bonds or directions (Fig. 2.4a); see Chap. 8.[10,11] Thus, the isotropic averaging is replaced by field-induced distributions. Moreover, the averaging over vibrational amplitudes assumes a well-defined structural geometry with the vibrations being typically harmonic about the equilibrium geometry. In molecules with many conformations, or macro in structure, this picture breaks down, and coherence is directly connected to a "dynamical landscape model" of populations on a potential energy surface of multiple minima (see Fig. 2.4b and Chap. 8).[12]

2.4 Coherence and Diffraction in Crystallography

From the above discussion, it follows that, for electrons, the scattering factors $f(\Theta)$, or $f(s)$, give the probability that an electron scatters into a specific direction determined by the angle between the incident beam and the scattering direction. In Sec. 2.3 this angle between the incident and scattered beams was denoted Θ, and $s = 2k_0\sin(\Theta/2)$, but in other contexts, such as crystallography, the angle of incidence is labeled Θ and the angle between the incident and scattered beams becomes 2Θ; for the latter case, $s = 2k_0\sin\Theta$. Also, Eq. (2.34) can be rewritten as

$$f(\Theta) = \text{const.}\times\left(\frac{\lambda}{\sin\Theta}\right)^2\left[Z - f_x(\Theta)\right], \tag{2.39}$$

which shows the decay of $f(\Theta)$ when plotted against $\sin(\Theta)/\lambda$ for higher scattering angles. These atomic scattering factors are listed in the International Tables of

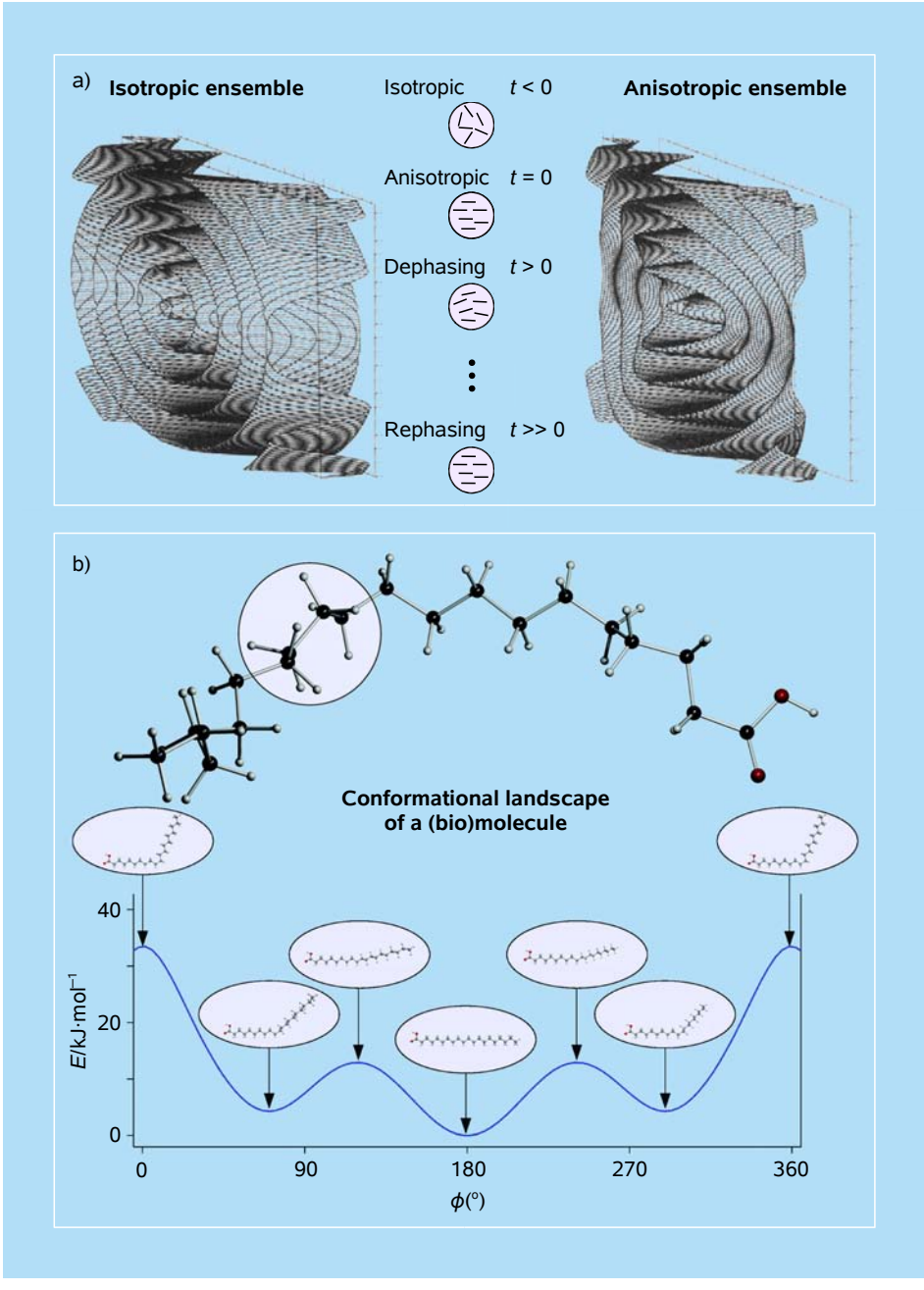


Figure 2.4 Orientations and conformations. (a) Isotropic and anisotropic gas electron diffraction patterns; temporal evolution of a molecular ensemble oriented with a laser field is shown in the center.[10,11] (b) Conformational landscape of a (bio)molecule reduced to a single dimension for simplicity[12]; see also Chap. 8.

Crystallography, and the larger the atomic number Z, the larger the amplitude. Similarly, Eq. (2.28) can be rewritten to obtain λ_e by replacing E_k by eV, the energy of acceleration (potential difference V in volts):

$$\lambda_e = h[2m_e eV + (eV)^2/c^2]^{-1/2}, \tag{2.40}$$

giving the values of de Broglie wavelengths: 0.0370 Å (for 100 kV) and 0.0251 Å (for 200 kV).

One of the most important and general quantities in crystallography is the structure factor $F(hkl)$, which determines the probability that an electron is scattered by the crystal (unit cell) in a given direction:

$$F(\boldsymbol{u}) = \sum_{i=1}^{N} f_i(u) \exp(2\pi i(\boldsymbol{u} \cdot \boldsymbol{r}_i)) \tag{2.41}$$

or

$$F(hkl) = \sum_{i=1}^{N} f_i(hkl) \exp(2\pi i[hx_i + ky_i + lz_i]), \tag{2.42}$$

where, as before, the scattering amplitude f_i is that of the ith atom in a unit cell with N atoms. The position of an atom is designated by (x_i, y_i, z_i) and the reflections (Bragg spots) are indexed by Miller indices (h, k, l). The phases (due to differences in path lengths from lattice planes to the detector) are defined by the dot product $\boldsymbol{u} \cdot \boldsymbol{r}_i$, with

$$\boldsymbol{u} = h\hat{a}^* + k\hat{b}^* + l\hat{c}^*, \tag{2.43}$$

$$\boldsymbol{r}_i = x_i\hat{a} + y_i\hat{b} + z_i\hat{c}. \tag{2.44}$$

The stars denote the unit vectors in reciprocal space. Thus, the intensity of scattering is determined by scattering amplitude for the type of atom f_i, its position in the unit cell (x_i, y_i, z_i), and the involved planes (h, k, l) that make up a crystal structure. The intensity contains both the amplitude and phase from the indicated sums, and in order to obtain structures from the intensities of diffraction reflections, the "phase problem" must be solved.

For electron scattering, because k_0 is much larger than in the case of X-rays (for electrons, $\lambda \approx 0.02$ Å; for X-rays, $\lambda \approx 1$ Å), the angles involved are uniquely small; the length of k_0 is the radius of the Ewald sphere, as discussed below. Some simple relationships are useful to recall. First, from Bragg's relationship,

$$n\lambda_e = 2d \sin\Theta, \tag{2.45}$$

where n is the order and d is the separation between the planes, and with $d/n \equiv d_{hkl}$, one obtains

$$\sin\Theta = \lambda_e/2d_{hkl}. \tag{2.46}$$

For 30 kV electrons and $d_{hkl} \approx 1$ Å, $\Theta < 2°$. Second, the relationship between s and d can be obtained by rearranging Eq. (2.46) and defining s (see the geometrical representation in Fig. 2.5):

$$\begin{aligned} sd_{hkl} &= 2\pi \quad (k_0 = 2\pi/\lambda); \\ sd_{hkl} &= 1 \quad\;\; (k_0 = 1/\lambda). \end{aligned} \tag{2.47}$$

From Eq. (2.47) it follows that:

$$\Delta s/s = -\Delta d/d. \tag{2.48}$$

We note the confusion which arises in the definition of Θ and k_0, but the important point is that the values of s in reciprocal space yield the separations in real space. (For comparison with the X-ray scattering literature, where the scattering vector is invariably labeled $\boldsymbol{q}$, it is important to note that $\boldsymbol{q} = 2\pi\boldsymbol{s}$.) Thus, if a Bragg spot is measured at $s = 0.5$ Å^{-1} away from the (000) position of the direct beam, d_{hkl} is measured to be 2 Å. In Fig. 2.5, the relationships between s and d are depicted using the Ewald sphere in reciprocal space; $1/\lambda$ defines the radius of the sphere whereas the momentum-transfer vector $\boldsymbol{s}$ is perpendicular to the diffracting planes in the case of elastic scattering and when satisfying the Bragg diffraction condition for (h, k, l).

At the detector, the measured distance x from the (000) position is related to d_{hkl} (and s of the Ewald sphere) by a simple relationship

$$x_{hkl}d_{hkl} = \lambda_e L, \tag{2.49}$$

where L is the camera length, and $\lambda_e L$ is the same for all reflections. This is the basis of indexing of diffraction in electron crystallography (see Fig. 2.6); for details, the reader is referred to cited articles and books on the subject.[13–16,22] Equations (2.46) and (2.49) above are reminiscent of Eqs. (2.23) and (2.24) obtained for a destructive interference (minimum); for constructive interferences (maxima), similar conditions may be obtained. Also, the much used Scherrer formula, $D\Delta s \sim 1$, for the estimation of crystallite size (D) from the diffraction peak width (Δs), is now understandable.

Clearly, the atoms are not stationary as dictated by the uncertainty principle and the temperature at which the experiment is carried out. For this reason, the well-known Debye–Waller factor, $W(T)$, is introduced to relate the vibrational amplitude to the intensity of diffraction. This temperature effect on diffraction (loss of interference by the incoherent movements of atoms) results in loss of intensity, which is expressed by the following relationship:

$$\ln(I/I_0) \equiv -2W(T) = -\boldsymbol{s}^2\langle u^2\rangle/3, \tag{2.50}$$

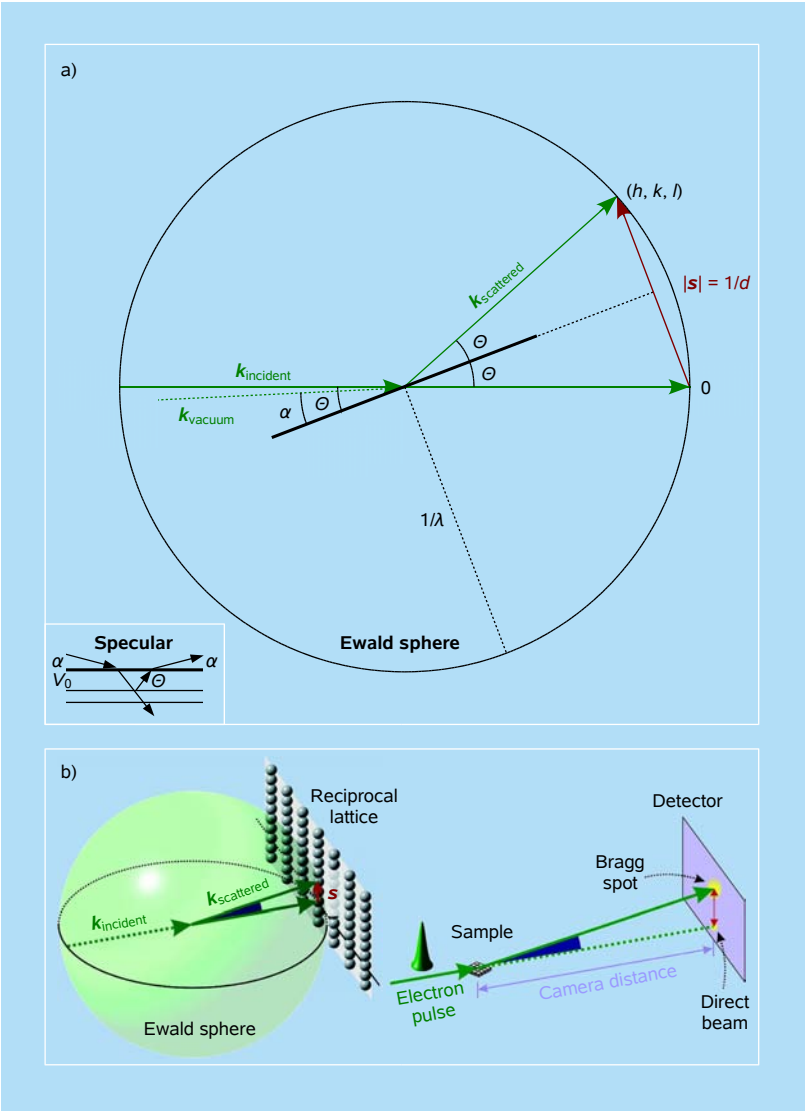


Figure 2.5 Electron scattering and Ewald sphere. (a) Ewald construction defined in a space with dimensions of reciprocal distance. Note that the scattering vector $\boldsymbol{s}$ of the length $|\boldsymbol{s}| = 1/d$ is perpendicular to the crystal plane that lies at an angle Θ with respect to the incident beam. (b) Reciprocal lattice and Bragg spot formation.[16] Note that the wave vectors $\boldsymbol{k}_{\text{incident}}$ and $\boldsymbol{k}_{\text{scattered}}$ shown here are those excited *inside* the crystal, and are related to the electron beam wave vector in vacuum, $\boldsymbol{k}_{\text{vacuum}}$, by $|\boldsymbol{k}_{\text{incident}}|^2 = |\boldsymbol{k}_{\text{vacuum}}|^2 + 2m|e|V_0/h^2$. Inset: The effect of the mean inner potential V_0. Because the refractive index for high-energy electrons $n \sim 1 + V/2V_0$ (where the accelerating potential is denoted V and a typical value of V_0 is 10 eV) is greater than unity, electron rays entering a solid are refracted toward the surface normal, according to Snell's law, as shown. If the incident beam arriving at a glancing angle α to the surface happens to be at the Bragg angle Θ inside the crystal after refraction, a reflected beam is generated, which is also refracted as it exits the surface. The result is a specular reflection. For $\alpha = 0$, a critical angle Θ can be defined below which no beam can escape from the crystal. As a result, the lower part of the diffraction pattern (for the smallest Bragg angles) is inaccessible; courtesy of J. C. H. Spence.

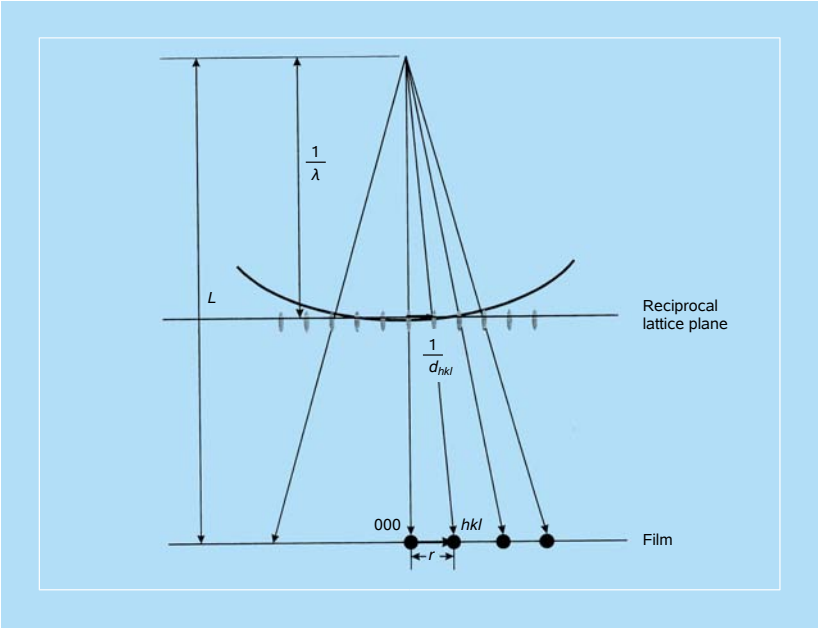


Figure 2.6 Schematic drawing illustrating the geometry of electron diffraction on a crystal lattice.[14,15]

where I (I_0) denotes the diffraction intensity from a lattice with (without) thermal vibrations, $\boldsymbol{s}$ the scattering vector for a particular Bragg spot, and $\langle u^2 \rangle$ the mean-square harmonic displacement of the constituent atoms in a unit cell. A fundamental assumption here is the thermal equilibrium among all modes of the lattice at a given well-defined temperature. However, this assumption, as discussed later, is not valid when the time scale is ultrashort and/or coherent motions of atoms are involved.

This Gaussian-type description can be modified to include the temporal dependence by expressing the last term of Eq. (2.50) as $\left[-\boldsymbol{s}^2\langle u(t)^2\rangle/3\right]$, with $\langle u(t)^2\rangle$ being the time-dependent mean-square displacement. If the crystal is excited and this "disorder" causes displacements (Fig. 2.7), then the time averaging in Eq. (2.50) may be replaced by a spatial averaging over all excited atoms (or molecules). If the displacement is ballistic, $\langle u(t)^2\rangle^{1/2}$, for example, becomes $\upsilon_{\mathrm{RMS}}t$, and $I(t)$ becomes Gaussian in both s and t; see Chap. 6.[17]

2.5 Coherence in Imaging

2.5.1 Basic concepts

It is perhaps useful to reflect on the transforming power of a lens and to recall the fundamental concepts developed by Abbe, Zernike, and others, and which are

pertinent to phase and amplitude contrast, spatial frequencies, etc. Both geometric optics (ray diagrams) and Fourier optics can be used to illustrate the process of imaging.[6,7] However, hidden in each of these approaches are details that need to be understood in order to acquire a clearer picture of the relationships of diffraction and imaging to the actual structure of the object. Here again, coherence plays a central role.

In Fig. 2.8a, an object, objective lens, and the two well-defined planes — back focal plane and image plane — are shown together in a ray diagram. An electron beam of parallel phase fronts is represented by the arrows, but the same could be said if, say, a laser beam were behind the object. We note several points. The two undiffracted (zero-order) or, more correctly, forward scattered rays focus at a point in the back focal plane of the lens (solid lines). The rays originating from the head and the tail of the object (diffraction by 2Θ) focus in the back focal plane at another "point" (dotted lines). All the rays scattered at different angles put together yield a *diffraction pattern* in the back focal plane of the lens. The location of this plane is given by the focal length of the lens (f), and it is related to the image distance simply by $f^{-1} = X^{-1} + Y^{-1}$, where X and Y are distances from the object to the lens and from the lens to the image plane, respectively. Thus, in the diffraction plane, the coherence is the result of interferences between scattered waves originating from a specific distance within the object, and the

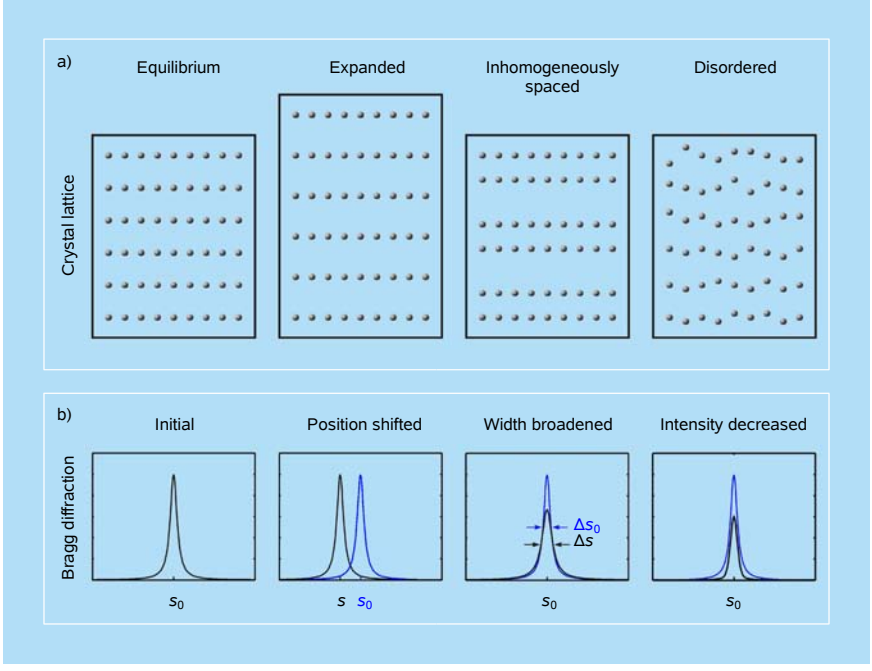


Figure 2.7 Influence of crystal lattice changes on observed Bragg diffraction. Adapted from Ref. 27.

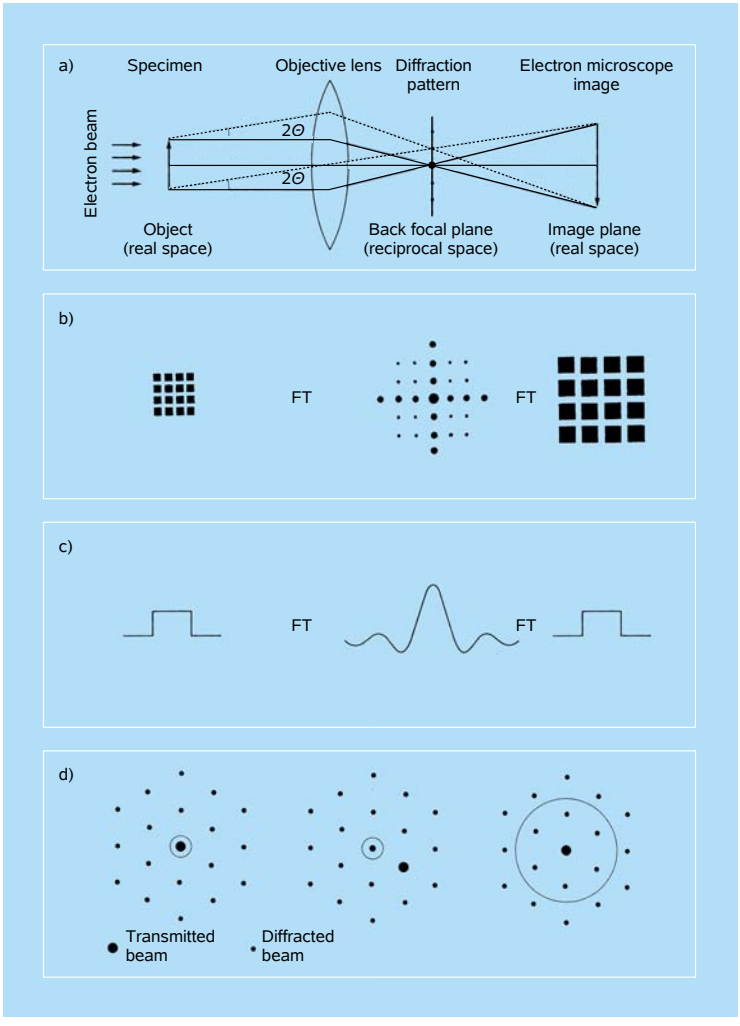


Figure 2.8 Imaging and diffraction by a ray diagram and through Fourier transformation. (a) The ray diagram depicting interferences between scattered rays (diffraction) and between the unscattered and scattered rays (image). (b) The Fourier transformation from an object (a grid of squares) to a diffraction pattern, and then to an image in real space. (c) The same transformation for a slit (meander).[6a] (d) Modes of detection in EM: bright-field (BF, left), dark-field (DF, center), and phase-contrast (high-resolution) imaging (right).[23]

formation of constructive or destructive interferences depends on the differences in optical path lengths, as we saw in Fig. 2.3. Quantum mechanically, the interferences are expressed in the phases [$\exp(i\boldsymbol{s}\cdot\boldsymbol{r})$] given in Eq. (2.36), and the distances are defined as in Eq. (2.37) for the case of a molecule.

We note that because the term $\exp(i\boldsymbol{s}\cdot\boldsymbol{r})$ is complex, with real $(\cos\boldsymbol{s}\cdot\boldsymbol{r})$ and imaginary $(\sin\boldsymbol{s}\cdot\boldsymbol{r})$ parts, the amplitude, $(Re^2 + Im^2)^{1/2}$, and the phase, (Im/Re), are determined by all the interferences originating from the entire set of distances

within the object, or the *structure* of the object. A measurement of the diffracted intensity provides the amplitudes, $(I_{\text{diff}})^{1/2}$, but the phases remain undetermined. Thus, the structure cannot be solved by simple Fourier transformation of these measured (real) amplitudes (the *phase problem*). For small systems, where the approximate structures are already known, the phases can be guessed close to their actual values, and by refining the structure (and thus the phases) iteratively, it is possible to achieve consistency between the measured and calculated intensities. For large systems, where even the approximate structures are unknown, more advanced methods have to be employed. In the 1950s, Max Perutz discovered that attaching a heavy metal (mercury) to hemoglobin changed the relative intensities of the diffraction spots measurably — and this, in principle, meant that the phase problem could be solved — and the isomorphous replacement method for determining crystal structures had been invented and was used by Perutz's colleague, John Kendrew, to determine the structure of myoglobin. This method has since been used to determine the structures of thousands of proteins.

If the object were a simple slit, then the positions of maxima and minima would indeed provide the slit width (as detailed earlier in this chapter), and the scattered intensity [Eq. (2.21)] for this simple case is readily understandable. Considering Huygens spherical wavelets, the amplitude of diffraction is the Fourier transform of the object:

$$A_{\text{diff}} = f_{\text{s}} \int_{-w/2}^{+w/2} e^{-iky \sin\theta} dy, \tag{2.51}$$

where f_{s} is a constant of the slit and the integration is over the slit dimension, from $-w/2$ to $+w/2$, with the origin being located in the center of the slit (see Fig. 2.3). The scattered intensity is simply $I_{\text{diff}} = AA^*$, and in fact, when the integration is made, one obtains Eq. (2.21), the sinc function. The analogy with Eqs. (2.36) and (2.37) is now clear.

In imaging, the situation is entirely different. First, without thinking about the diffraction focal plane, the diagram in Fig. 2.8 reveals that, for the head and the tail of the image, the interferences are between the incident (zero-order) and the scattered (diffracted) rays in the image plane. Because we measure intensities, we can now write the analogue of Eq. (2.37), but in the following simple way:

$$I = \left| f_0 \exp i(kz + \varphi_0) + f_{\text{s}} \exp i(kz + \varphi_{\text{s}}) \right|^2 \equiv \left| f_0 \right|^2 + \left| f_{\text{s}} \right|^2 + 2 f_0 f_{\text{s}} \cos\varphi_{\text{s}}, \tag{2.52}$$

where φ_{s} is a phase shift acquired by the scattered wave during the interaction with the specimen. To illustrate the point, we take f_{s} to be real, and φ_{s} is now expressed relative to φ_0. It is apparent that the coherent interference of the incident (0) and scattered (s) waves gives rise to a modulation on top of the "background

scattering", $|f_b|^2 = |f_0|^2 + |f_s|^2$, and this modulation is responsible for the contrast variation in the image plane. Notably, the image is reflecting the structure of the object through interferences with the "reference" beam (which is also the case in holography), while in diffraction the interferences are due to the differences in the optical paths associated with different distances within the object. Thus, a single atom can be imaged as a position point without consideration of the phase problem of the reciprocal space. In practice, however, the issues of resolution and sensitivity have to be considered.

The relation between imaging and diffraction is given by a Fourtier transformation (FT). As we have already seen, the FT of an object potential yields a diffraction pattern. A further FT gives a magnified and inverted image (see Figs. 2.8b,c) of that potential. These relationships make possible the progress achieved in electron crystallography of complex systems. From the FT of the observed intensity distribution on a selected-area, real-space grid in the image, the phases of the diffraction spots (h, k) can be obtained:

$$T(h,k) = \mathrm{FT}[I_{\mathrm{image}}(x,y)] = \int_{x_1}^{x_2}\int_{y_1}^{y_2} I_{\mathrm{grid}}(x,y)e^{2\pi i(hx+ky)}dxdy$$

$$\equiv \sum_{x_1}^{x_2}\sum_{y_1}^{y_2} I_{\mathrm{grid}}(x,y)[\cos 2\pi(hx+ky) + i\sin 2\pi(hx+ky)]. \tag{2.53}$$

Thus, by pixelizing the area of the grid in x and y directions and measuring the intensity of each pixel, it is possible to obtain the real and imaginary parts for a given (h, k). For example, if $T(6, 10) = 14 + i0$, then for this diffraction spot ($h = 6$, $k = 10$) the amplitude (the length of the vector) is 14, and the phase is 0. When the process is extended to the third dimension, the amplitudes and phases can be obtained in 3D Fourier space (h, k, l).

A few final comments are worthy of emphasis. First, the use of filters in diffraction results in image modifications, which should be clear from the above discussion. Known examples include masks on gratings, which lead to the suppression of certain wavelengths. Accordingly, a diffraction pattern may be interpreted as a superposition of "spatial waves," of well-defined frequencies, originating from different distances within the object: the higher the spatial frequency, the smaller the distance and the better the resolution. This can be illustrated by the superposition of three spatial waves of the form of $\sin[2\pi(x/\lambda)]$, associating +1 with black, −1 with white, and 0 with gray (Fig. 2.9). It follows that contributions from high-frequency spatial waves are directed to large scattering angles [see Eq. (2.23)]. Second, the concept of optical filtering, which is used in improving the appearance of photographs, is based on the selection of certain spatial waves; blocking, for example, high-frequency ones with a filter installed in

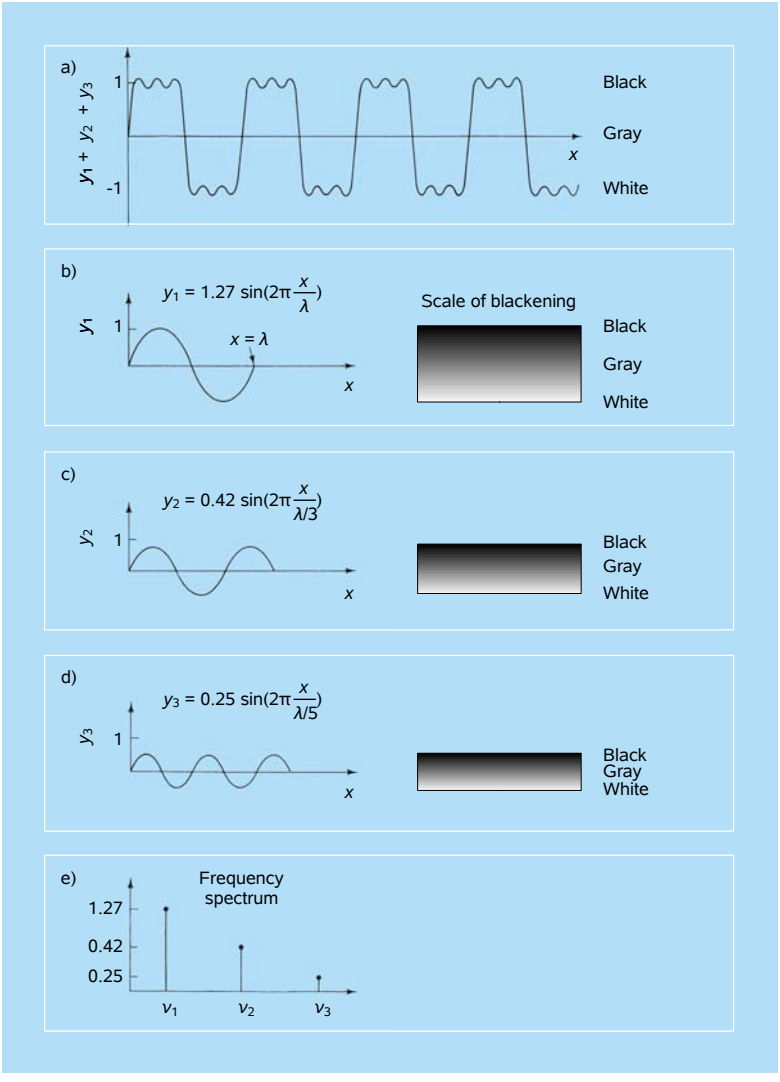


Figure 2.9 Spatial waves decomposition and contribution to contrast. The high-frequency waves, which scatter at larger angles, have in this case smaller amplitudes.[6a]

the focal plane of the lens suppresses structural details at small distances, and thus smoothes the picture, as explained with the aid of optical transforms in Chap. 3 (see Fig. 3.2).

Historically, the theory of image formation as it relates to the diffraction pattern of the object, and the effect of filtering, dates back to the work of Abbe in 1870s. However, it was not until the 1930s that the method of phase contrast was developed by Zernike for imaging when the change in refractive index and thickness of specimen is too small to produce a pronounced contrast variation.

This is often the case for biological imaging of a thin "carbon material" in vitreous water. Staining, though used to enhance contrast, is not the method of general choice if only because of its intrusive nature. The enhancement of phase contrast through zone plates (phase shift) and other schemes is a subject well-developed in optical microscopy, and here and in Chap. 5 we consider it specifically for the case of electron microscopy.

Before we proceed, it is instructive, however, to outline the general methods of contrast generation in electron microscopy (Fig. 2.8d): amplitude contrast and phase contrast. The latter method is used to obtain high resolution images of the specimen and can be carried out in either the bright field mode (direct beam reaches the image plane) or the dark field mode (direct beam is blocked). Amplitude contrast does not rely on the phases of the scattered wave, rather it passes only rays that scatter at small angles, while intersecting rays are scattered at higher angles by a limiting aperture in the back focal plane of the objective lens. Consequently, regions of the specimen that scatter electrons more strongly (heavy atoms or high density) will appear dark in the image because fewer electrons were able to pass these specimen regions undeflected. Thus, amplitude contrast measures the "scattering power" or the mass-thickness of certain region in the specimen. Phase contrast, on the other hand, relies on coherent interferences of the scattered rays with the unscattered ray (bright field) or with themselves (dark field). Since phase contrast imaging relies on the phase information of the scattered rays, a larger back focal plane aperture passes more scattered rays (spatial frequencies), resulting in more information content in the image and, consequently, better resolution. The term "amplitude contrast" has different connotations depending on the discipline, materials science or biology.

2.5.2 Coherence of the source, lateral and temporal

In the above sections of this chapter we discussed wave interferences and optics, both of which are pertinent to the description of coherence and image formation. Here, we focus the discussion on properties of the source that are relevant for evaluating the extent of lateral and temporal coherence in imaging. First, we consider the ray diagram in Fig. 2.3 (middle) but define two angles, Θ_s, the semi-angle subtended by the focused spot at the illuminating aperture, and Θ_c, the semi-angle subtended by the source at the plane of focus (*see inset*). In analogy with the results of Eq. (2.24), a lateral coherence length can be defined:

$$w\Theta_s = \lambda; \quad l_c = \left(\frac{L}{w}\right)\lambda. \tag{2.54}$$

The right-hand side of Eq. (2.54) can be rewritten to give (*see inset*):

$$2l_c\Theta_c = \lambda. \tag{2.55}$$

We note that the criterion here is that "incoherency" occurs at $x = \lambda L/w$ (first minimum, destructive interference) and beyond; one could have used some other criteria, such as the value of $x = 1$ corresponding to about 15 percent drop of intensity, with a factor of π difference [see Eqs. (4.16) and (4.17) in Ref. 18]. By the van Cittert–Zernike theorem (discussed in Born and Wolf[6c]; see also Ref. 18), the results are still valid for an aperture being filled with an extended quasi-monochromatic (incoherent) source; for a completely filled aperture with spherical waves a coherent source is defined. Thus, knowing the size of the source and the distance L, one can obtain the coherence length. A prime example is that of the coherence length of the sunrays falling on the Earth. The angular radius of the Sun, the ratio of the Sun's radius to the distance to the Earth, is 0.0047 rad. Thus, for $\lambda = 0.55$ μm, the coherence length is on the order of 100 μm.

We now turn to the description of the waves involved and possible decoherences due to the differences in Θ-distribution of the source and k-distribution of the emitted waves. The superposition is of the type given in Eq. (2.10) and dephasing must be considered; for electron interference we follow the terminology of Lichte.[19] In the 1950s, Möllenstedt and colleagues found the two-beam electron interference corresponding to the known Fresnel biprism fringes in light optics; the analogy in double-slit electron interference will be discussed later. Taking the form of the two waves to be

$$\psi_i = a_i \exp(i\boldsymbol{k}_i \cdot \boldsymbol{r} + i\varphi_i), \qquad i = 1, 2, \tag{2.56}$$

one propagating along the y-axis and the other at a slight angle β, the interference of the two waves gives rise to the following intensity in the y-direction and at any point (x) perpendicular to it:

$$I(x, y) = a_1^2 + a_2^2 + 2a_1 a_2 \cos[(\boldsymbol{k}_1 - \boldsymbol{k}_2) \cdot \boldsymbol{r} + (\varphi_1 - \varphi_2)]. \tag{2.57}$$

Because β is small and $|\boldsymbol{k}| = |\boldsymbol{k}_1| = |\boldsymbol{k}_2|$, $(\boldsymbol{k}_1 - \boldsymbol{k}_2) \cdot \boldsymbol{r} \approx |\boldsymbol{k}| \beta x$. The degree of coherence can now be introduced (see Sec. 2.2) through the function

$$\mu_{\text{coh}} = \mu_{\text{sc}}(\Theta)\mu_{\text{tc}}(x), \tag{2.58}$$

where μ_{sc} is the degree of the spatial (angular) coherence and μ_{tc} is the degree of the temporal (longitudinal) coherence. The assumption in this separation is that the emission of each source point is characterized by the same energy distribution. Both μ_{sc} and μ_{tc} represent distributions centered at mean values and over angles and distances, respectively, and can be written as FTs of the intensity and wavenumber distributions, respectively.[19] These coherences contribute to changes in the total amplitude and phase, and thus, for convenience, one writes

$$\mu_{\text{coh}} = |\mu_{\text{coh}}| \exp(i\rho_{\text{coh}}), \tag{2.59}$$

and Eq. (2.57) becomes

$$I(x,y) = a_1^2 + a_2^2 + 2a_1a_2|\mu_{\text{coh}}|\cos\left(|\boldsymbol{k}|\beta x + \Delta\varphi + \rho_{\text{coh}}\right). \tag{2.60}$$

Using these relations, spatial and temporal coherences have been experimentally measured for electron beams emitted by different sources. Inelastic interactions involving one beam (by ΔE energy change) can be accounted for as a further phase shift of $(\Delta E/\hbar)t$, which represents the difference in the number of cycles of the waves involved.

2.5.3 Imaging in electron microscopy

Building on the concepts outlined above, for EM it is important to consider the concept of specimen potential in imaging; for details, see the books by Spence, Cowley, Reimer, and others,[18,20–26] and Ref. 14. Within a sample it is the electrostatic potential difference, by both the electrons and nuclei, that is responsible for the scattering of incident electrons. Here, the potential is denoted by $\Phi(\boldsymbol{r})$, where $\boldsymbol{r}(x, y, z)$ is a set of atomic coordinates. For crystals, the potential (in volts) is given by the sum

$$\Phi(\boldsymbol{r}) = \sum_{i=1}^{N} \Phi_i(\boldsymbol{r} - \boldsymbol{r}_i), \tag{2.61}$$

with

$$\boldsymbol{r}_i(x_i, y_i, z_i) = x_i\hat{a} + y_i\hat{b} + z_i\hat{c}, \tag{2.62}$$

and the sum, as before, is over the atoms in the unit cell. The FT of the potential (in real space) is directly related to the structure factor of the crystal:

$$\Phi(\boldsymbol{u}) \equiv \text{FT}[\Phi(\boldsymbol{r})] = \frac{\lambda_{\text{e}}}{\sigma V_{\text{c}}} F(\boldsymbol{u}), \tag{2.63}$$

where σ is the interaction constant ($\sigma = 2\pi m_{\text{e}} e\lambda/h^2$) and V_{c} is the volume of the unit cell. It follows that

$$\Phi(\boldsymbol{r}) = \frac{\lambda_{\text{e}}}{\sigma V_{\text{c}}} \sum F(\boldsymbol{u}) \exp(-2\pi i[\boldsymbol{u}\cdot\boldsymbol{r}]) \tag{2.64}$$

is the inverse Fourier transformation. In general, the potential $\Phi(\boldsymbol{r})$ is continuous in real 3D space as it is a map of the atomic structure of the specimen. Considering the case in Fig. 2.10, for the so-called "phase object approximation," where the scattering angle is very small and no change in amplitude takes place, the wave function of scattered waves in a plane at a distance R from the specimen is

given by

$$\psi = \text{const.} \times \iint q(x, y) \frac{\exp(ikR)}{R} dxdy, \tag{2.65}$$

which indicates that the change is in $q(x, y)$ and the scattering is that of spherical waves. It can be shown that in the Fraunhofer diffraction limit (or in the back focal plane of the objective lens), the scattering amplitude is given by the FT of $q(x, y)$; the scattering wave at the exit surface is simply determined by $q(x, y)$.

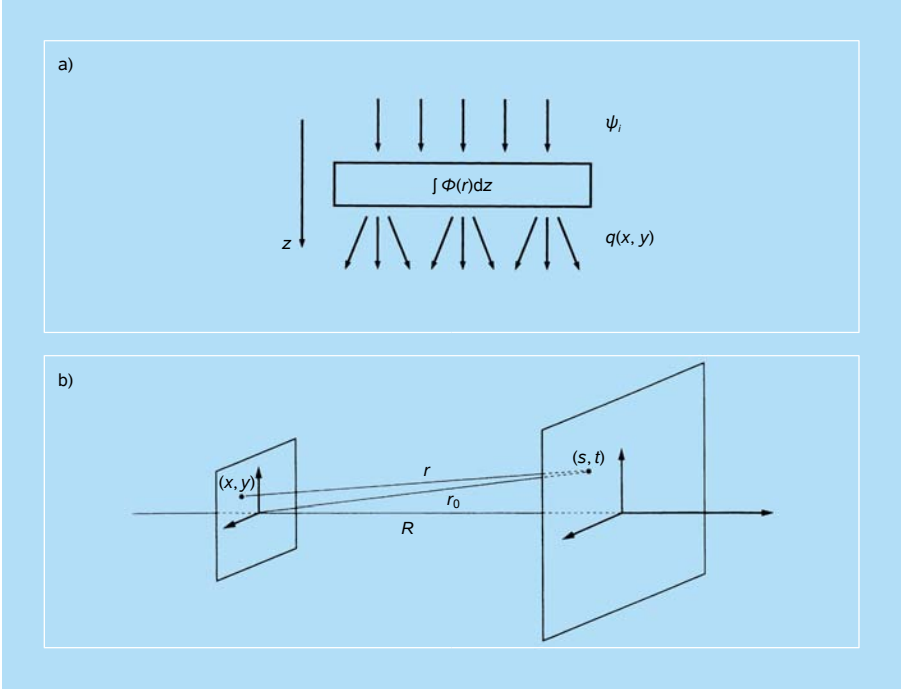


Figure 2.10 Electron scattering by a specimen potential and a ray diagram. (a) The effect of the potential on exit (scattered) waves.[14] (b) Schematic illustration showing the electron scattering from a specimen to a plane at a distance R; in this plane u and v are defined: $u = s/\lambda r_0$, $v = t/\lambda r_0$.[23]

In this regime of the phase object approximation, our task is to find the scattering phase shift resulting from the propagation of the electron in a specimen of thickness Δz. Physically, there is an analogy with light. In vacuum the wavelength of the electron is λ, whereas in the specimen it is λ_s due to the potential $\Phi(\boldsymbol{r})$, and λ_s without relativistic correction is given by

$$\lambda_s = h[2m_e e(V + \Phi(r))]^{-1/2}, \tag{2.66}$$

or

$$\frac{1}{\lambda_s} = \frac{1}{\lambda}\left[1 + \frac{\Phi(r)}{V}\right]^{1/2} \approx \frac{1}{\lambda}\left[1 + \frac{\Phi(r)}{2V}\right]. \tag{2.67}$$

The phase shift becomes

$$2\pi \int_0^z \left(\frac{1}{\lambda_s} - \frac{1}{\lambda}\right) dz \equiv \sigma \Phi(x, y) \Delta z, \tag{2.68}$$

where σ is the interaction constant in $\mathrm{V}^{-1}\mathrm{\AA}^{-1}$ and equals $2\pi\left[V\lambda(1 + (1 - \alpha^2)^{1/2})\right]^{-1}$; it is $\sim \pi/V\lambda$ when α, the velocity of the electron relative to that of light, is ~ 0.3. The relevant transmission function becomes

$$q(x, y) = \exp[i\sigma\Phi(x, y)\Delta z], \tag{2.69}$$

and this exit electron wave is no longer a single plane wave, rather it is a rippled function describing the potential distribution projected along z. As such, the intensity, if measured, would be one with no contrast. The transmission function $q(x, y)$ is best defined when used in conjunction with the first Born approximation, as the ratio of the transmitted complex wavefield and the incident complex wavefield.

For thin specimens (a few nanometers), the phase factor in the exponent takes on a value much smaller than one (weak-phase object approximation), i.e., $\sigma\Phi(x, y)\Delta z \ll 1$, and

$$q(x, y) \approx 1 + i\sigma\Phi(x, y)\Delta z. \tag{2.70}$$

The presence of an objective lens modifies such phases due to the spherical aberrations and the aperture of the lens. As mentioned above, the scattering amplitude in a plane at a distance R from the specimen, or in the back focal plane of the lens, is the FT of the transmission function (ignoring some constants):

$$\psi(u, v) = \mathrm{FT}[q(x, y)], \tag{2.71}$$

where u and v are the coordinates in that plane (Fig. 2.10). Besides its focusing power, the effect of a real lens is to multiply by the "contrast transfer function" (discussed also in Chap. 5), which describes the changes of phases due to aberrations of the lens, i.e.,

$$\psi(u, v) = \mathrm{FT}[q(x, y)] \exp(i\chi(u, v)), \tag{2.72}$$

with

$$\chi(u,v) = \pi\left[\Delta f\lambda(u^2+v^2) - \frac{1}{2}C_s\lambda^3(u^2+v^2)^2\right], \tag{2.73}$$

defining Δf and C_s, respectively, as the defocus (sometimes designated *defect of focus*) and the spherical aberration coefficient (symmetric about the incident beam axis) of the objective lens (see Fig. 2.11). The wave amplitude at the image plane is then obtained (ignoring some constants) by further Fourier transformation of the diffraction amplitudes at the back focal plane of the lens:

$$\psi(x,y) = \mathrm{FT}[C(u,v)\psi(u,v)], \tag{2.74}$$

where C describes the effect of the objective aperture, and the intensity

$$I(x,y) = \psi(x,y)\psi^*(x,y) = \left|1 + i\,\mathrm{FT}\left[\mathrm{FT}[\sigma\Phi(x,\ y)\Delta z]\exp(i\chi(u,v))\right]\right|^2, \tag{2.75}$$

for which we ignored the aperture effect putting $C(u,v) = 1$.[23] For the simple case of $\exp(i\chi(u,v)) = \pm i$, $(u,v) \neq 0$, the intensity becomes $\left|1 \mp \sigma\Phi(-x,\ -y)\Delta z\right|^2$, recovering the change in the specimen potential in the image (see Figs. 2.8 and 2.12).

For simplicity, we now consider the case of only "1D" spatial frequencies, that of u. Then, the phase changes, employing the terminology of Ref. 22 for use in Chaps. 3 and 4, can be expressed by the following (approximate) contrast transfer function (CTF):

$$\mathrm{CTF}(u) = A(u)E(u)\sin[\chi(u)]. \tag{2.76}$$

In this "simple" expression, $\chi(u)$ is a function of the aberration of the lens system, $A(u)$ is the aperture function cutting off beams scattered above a certain angle, and $E(u)$ is the envelope function that describes the attenuation of high spatial frequencies. The sinusoidal term determines the magnitude and sign for the contribution of the spatial frequency (u) amplitude to the contrast of the image. From Eq. (2.73) it follows that

$$\chi(u) = (\pi/2)C_s\lambda^3u^4 - (\pi)\Delta f\lambda u^2, \tag{2.77}$$

where C_s is the spherical aberration coefficient. If the defocus Δf is zero, the first term will be entirely determined by the u^4 dependence. If, however, Δf is finite, then the λu^2 term counters the first one, and it can, in principle (and in practice), flatten $\chi(u)$, thus allowing low spatial frequencies (and hence larger spacings within the object) to be transferred to the image (see Fig. 2.11). The oscillations reflect interferences as rays go through the lens, with the contrast transfer at high frequencies alternating rapidly between +1 and −1 (Fig. 2.11).

For the ideal case of CTF taking on a constant value (the entire range of spatial frequencies), the high-resolution images would reflect the true crystal

potential and the resolution is optimum. In reality, there are resolution limits imposed on the omission of the highest spatial frequencies. The known Scherzer defocus depends on microscope properties (C_s and λ) and is given by (realizing that $\sin \chi = 0$, when $\chi = n\pi$, and optimizing by taking the derivative)[22]:

$$\Delta f = -1.2(C_s\lambda)^{1/2}. \tag{2.78}$$

The limit for the highest spatial frequencies being involved with the correct sign is critical to resolution, and this point resolution is defined as the spatial frequency where the CTF crosses the abscissa for the first time. At Scherzer defocus, the value of u becomes

$$u_{\text{Scherzer}} = 1.5\lambda^{-3/4}C_s^{-1/4}, \tag{2.79}$$

and the resolution at Scherzer defocus is simply the inverse, giving

$$d_{\text{Scherzer}} = 0.66\lambda^{3/4}C_s^{1/4}. \tag{2.80}$$

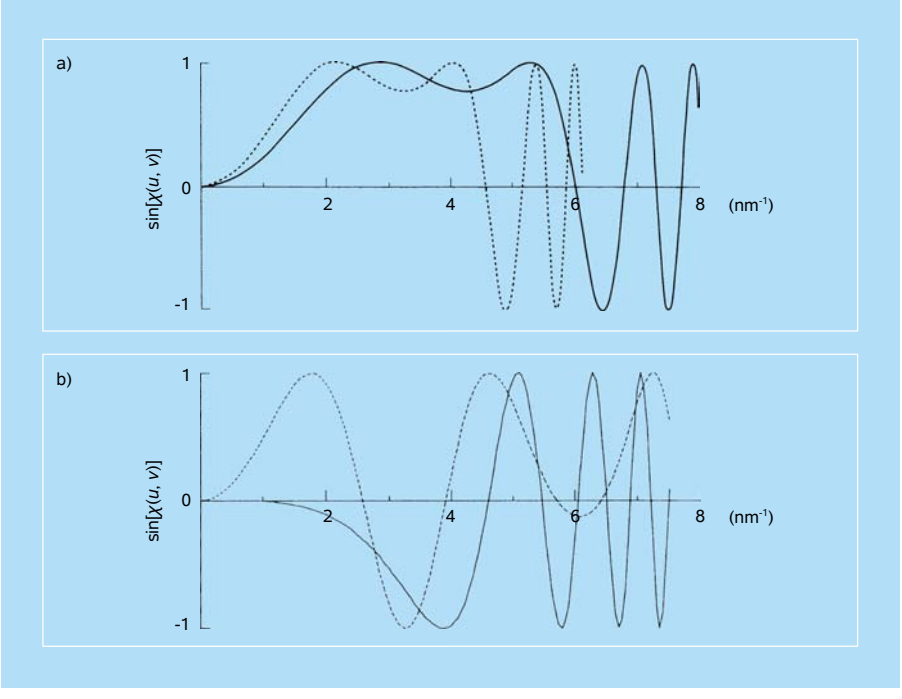


Figure 2.11 Contrast transfer function of the objective lens, imaginary part. (a) Under Scherzer focus [see Eq. (2.79)] for a 400 kV electron microscope (solid line) and a 200 kV electron microscope (dotted line). (b) For a 400 kV electron microscope with Δf = 0 nm (solid line) and Δf = 100 nm (dotted line).[23]

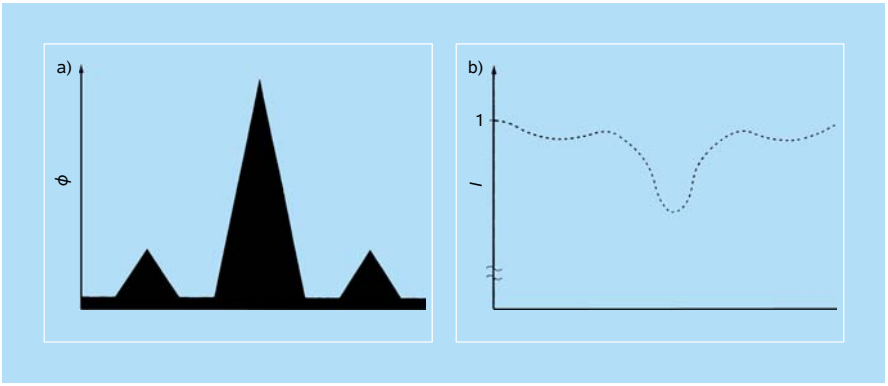


Figure 2.12 Schematic illustration showing the correspondence of (a) the crystal potential and (b) the intensity distribution of a high-resolution image.[23]

The CTF of the objective lens under the Scherzer defocus condition is seen in Fig. 2.11 to change with λ, as expected. The function takes values near 1 in the range 1.7–4.3 nm^{-1} for 200 kV and 2.1–5.7 nm^{-1} for 400 kV for a specific (C_s value) microscope. The CTF changes with defocus (Fig. 2.11b), enhancing the imaging of lower frequencies (larger-scale objects) at larger defocus, which is of particular significance to biological imaging when the entire object needs to be seen, but hopefully with CTF having a good coherence property to allow also for high resolution of structures within the object. Decreasing λ and/or C_s clearly improves the resolution.

The envelope function $E(u)$ dampens the CTF and thus acts as a "virtual aperture," cutting off high spatial frequency components. It is a product of several terms: the chromatic aberration of the lens, angular spread of the source, specimen drift and vibration, and the detector. As discussed earlier in this chapter, and in the coming chapters, the angular spread of the source limits the spatial (lateral) coherence whereas the chromatic dispersion mainly influences the temporal coherence (see Sec. 2.5.2). The chromatic contribution, $E_c(u)$, to the total envelope function can be expressed as a Gaussian distribution:

$$E_c(u) = \exp\left[-\frac{1}{2}(\pi\lambda\delta)^2 u^4\right], \tag{2.81}$$

independent of defocusing Δf. The spread δ, which is due to chromatic aberration, is given by the sum of dephasing terms:

$$\delta = C_c[4(\Delta I_{obj}/I_{obj})^2 + (\Delta V/V)^2 + (\Delta E/V)^2]^{1/2} \tag{2.82}$$

where the first two terms are due to changes in the current through the objective lens and voltage fluctuations in the electron source, and the last term is due to the energy spread of electrons relative to the accelerating voltage energy; C_c is the chromatic aberration coefficient of the lens. The coherence envelope due to the angular spread, unlike that due to the chromatic spread, depends on defocus. The spatial envelope function can be written as:

$$E_s(u) = \exp[-(\pi\alpha/\lambda)^2(C_s\lambda^3u^3 + \Delta f\lambda u)^2], \tag{2.83}$$

where α is the divergence angle describing the distribution. The damping can be minimized by optimizing Δf ("Lichte defocus"; the *Lichte focus* is the defocus setting at which the contrast transfer function has the optimum contrast for electron holography).[19] Consideration of spatial and temporal coherence is important in the unique concept and design of single-electron imaging, as discussed in Chaps. 5 and 6.

Several other points are noteworthy. First, for the electron passing through the specimen, we can define a "refractive index." Because the refractive index is the ratio of $(k/\omega)/(k/\omega)_0$, equivalent to $[V + \Phi(r)]^{1/2}/V^{1/2}$, it may be approximated by $\sim 1 + (\Phi/2V)$, a value mentioned earlier as responsible for the phase shift caused by the potential Φ of the specimen. Second, for thicker samples, where inelastic scattering and adsorption have to be considered, $q(x, y)$ is multiplied by a phenomenological decay function $\exp[-\mu(x, y)]$, the analogue of what was done for the diagonal terms in optical coherence. The intensity then becomes proportional to $\exp[-2\mu(x, y)]$, and this is referred to sometimes as "amplitude contrast"; it is an "absorption" effect due to inelastic processes, but it should not be confused with phase contrast obtained from the coherent interference of elastically scattered waves with the unscattered wave (bright field) or with themselves (dark field). Third, when multiple scattering takes place, dynamical scattering methods, such as the "multi-slice" approach of Cowley and Moodie, have to be involved.[20] Finally, the resolution, as influenced by the decay envelope, can operationally be checked by taking the FT of the image — when displaying "Thon rings," which correspond to the oscillations on the CTF abscissa, the extent of coherence can be assessed by quantifying the attenuation of the rings.

2.6 Instrumental Factors Limiting Coherence

As discussed in the previous section, for proper image formation in the electron microscope we must take into consideration instrumental implementations of lens aberrations and electron trajectories and energies through the contrast transfer function and its damping envelopes. Similarly, in diffraction and crystallography, such "instrumental factors" are unavoidable, and they have to be optimized in

order to project as much coherence out of the system as possible. Because each electron interferes only with itself, the image is composed of the independent superposition of single-electron interferences and, thus, interference fringe blurring occurs in the presence of an uncertainty in the electrons' trajectory, location and their wavelength at the moment of scattering. The magnitude of these instrumental uncertainties determines the degree to which the visibility of the coherent interferences (100 percent in the ideal case) is attenuated.

Common to all diffraction techniques are Eqs. (2.32) and (2.33) for the momentum transfer during the scattering process. The error in the observable momentum transfer at a given radial position s in the diffraction plane can then be evaluated by the convolution of the above-mentioned contributions:

$$\delta s = \sqrt{\left(\frac{\partial s}{\partial \Theta}\right)^2 \delta\Theta^2 + \left(\frac{\partial s}{\partial \lambda}\right)^2 \delta\lambda^2 + \left(\frac{\partial s}{\partial x}\right)^2 \delta x^2 + \left(\frac{\partial s}{\partial y}\right)^2 \delta y^2 + \left(\frac{\partial s}{\partial z}\right)^2 \delta z^2 + \left(\frac{\partial s}{\partial b}\right)^2 \delta b^2}. \tag{2.84}$$

The $\delta\Theta$-term represents the uncertainty in the scattering angle, originating from the incident angular spread at each position in the interaction volume (lateral coherence length of the beam). The $\delta\lambda$-term represents the longitudinal energy spread (longitudinal coherence length of the beam), and the δx-, δy- and δz-terms are reflective of the 3D size of the interaction region defined by the electron and the specimen. The δb-term is an uncertainty in measuring the momentum transfer due to the electron detection process, for example, through blurring and binning in the detector (signal converter, amplifier, and digitizer). In the case of gas-phase electron diffraction discussed earlier, each distinct internuclear distance produces sinusoidal interference fringes [Eq. (2.38)], while in crystallography, repeated long-range order of scattering centers produces a series of Bragg spots with a corresponding frequency in spacing. The uncertainty in s will reduce the visibility of the interference fringes or the intensity of the Bragg spots, such that they become unobservable and/or indistinguishable. The instrumental coherence length can then be defined as

$$R_c = \frac{1}{\delta s} \equiv \frac{1}{2\sigma_s}. \tag{2.85}$$

This value gives an estimate of the maximum distance that can be resolved with high confidence in an electron diffraction experiment.

Finally, we should mention that when electrons are pulsed at high rates, not in the continuous-wave mode, the coherence properties outlined above must be discussed in relation to the duration of the electron pulse and the number of electrons per pulse. Unlike photons (bosons), electrons (fermions) follow the exclusion principle. At high densities, the interactions between them (such as

the Boersch effect) lead to an increase of the effective source size (decrease of lateral coherence) and to a limited temporal resolution. These and other issues will be addressed in Chap. 5, where we discuss ultrafast electron microscopy and diffraction.

References

1. (a) L. Mandel and E. Wolf, "*Optical Coherence and Quantum Optics*," Cambridge University Press, Cambridge (1995).
   (b) R. Loudon, "*The Quantum Theory of Light*," 3rd ed., Oxford University Press, Oxford (2000).
   (c) H. Haken, "*Light: Waves, Photons, Atoms*," Vol. 1, North Holland, Amsterdam (1981).
2. M. Sargent III, M. O. Scully and W. E. Lamb, Jr., "*Laser Physics*," Addison-Wesley, Reading (1974).
3. A. H. Zewail, *Acc. Chem. Res*., **13**, 360 (1980).
4. (a) K. E. Jones and A. H. Zewail, in "*Springer Series in Chemical Physics*, *Vol. 3: Advances in Laser Chemistry*," Ed. A. H. Zewail, Springer, Berlin–Heidelberg, p. 196 (1978).
   (b) K. E. Jones, A. H. Zewail and D. J. Diestler, in "*Springer Series in Chemical Physics*, *Vol. 3: Advances in Laser Chemistry*," Ed. A. H. Zewail, Springer, Berlin–Heidelberg, p. 258 (1978).
5. A. Abragam, "*Principles of Nuclear Magnetism*," Oxford University Press, Oxford (1983).
6. (a) K. D. Möller, "*Optics*," University Science Books, Mill Valey (1988).
   (b) J. W. Goodman, "*Introduction to Fourier Optics*," 3rd ed., Roberts & Co., Greenwood Village (2005).
   (c) M. Born and E. Wolf, "*Principles of Optics*," 7th ed., Cambridge Unversity Press, Cambridge (2002).
7. E. M. Slayter and H. S. Slayter, "*Light and Electron Microscopy*," Cambridge University Press, Cambridge (1992).
8. (a) L. D. Landau and E. M. Lifshitz, "*Quantum Mechanics: Nonrelativistic Theory*," 3rd ed., Butterworth-Heinemann, Oxford (2003).
   (b) M. I. Davis, "*Electron Diffraction in Gases*," Marcel Dekker, New York (1971).
   (c) "*Stereochemical Applications of Gas-Phase Electron Diffraction, Part. A: The Electron Diffraction Technique*," Eds. I. Hargittai and M. Hargittai, VCH, New York (1988).
9. J. C. Williamson, *Ph.D. Thesis*, California Institute of Technology, Pasadena, CA, USA (1998).
10. J. C. Williamson and A. H. Zewail, *J. Phys. Chem*., **98**, 2766 (1994).
11. (a) J. S. Baskin and A. H. Zewail, *Chem. Phys. Chem*., **6**, 2261 (2005).
    (b) J. S. Baskin and A. H. Zewail, *Chem. Phys. Chem*., **7**, 1562 (2006).
12. (a) D. Shorokhov and A. H. Zewail, *Phys. Chem. Chem. Phys*., **10**, 2879 (2008).
    (b) M. M. Lin, D. Shorokhov and A. H. Zewail, *Chem. Phys. Lett*., **420**, 1 (2006).

(c) M. M. Lin, D. Shorokhov and A. H. Zewail, *J. Phys. Chem. A*, **113**, 4075 (2009).
13. (a) L.-M. Peng, S. L. Dudarev and M. J. Whelan, "*High-Energy Electron Diffraction and Microscopy*," Oxford University Press, Oxford (2004).
(b) J. C. H. Spence and J. M. Zuo, "*Electron Microdiffraction*," Springer, Berlin–Heidelberg (1992).
14. X. Zou and S. Hovmöller, *Acta Cryst. A*, **64**, 149 (2008); "*Notes on Electron Crystallography*" (private communication).
15. "*NATO Science Series E, Vol. 347: Electron Crystallography*," Eds. D. L. Dorset, S. Hovmöller and X. Zou, Springer, Berlin–Heidelberg (2004).
16. D.-S. Yang, N. Gedik and A. H. Zewail, *J. Phys. Chem. C*., **111**, 4889 (2007).
17. A. M. Lindenberg, J. Larsson, K. Sokolowski-Tinten, K. J. Gaffney, C. Blome, O. Synnergren, J. Sheppard, C. Caleman, A. G. MacPhee, D. Weinstein, D. P. Lowney, T. K. Allison, T. Matthews, R. W. Falcone, A. L. Cavalieri, D. M. Fritz, S. H. Lee, P. H. Bucksbaum, D. A. Reis, J. Rudati, P. H. Fuoss, C. C. Kao, D. P. Siddons, R. Pahl, J. Als-Nielsen, S. Duesterer, R. Ischebeck, H. Schlarb, H. Schulte-Schrepping, T. Tschentscher, J. Schneider, D. von der Linde, O. Hignette, F. Sette, H. N. Chapman, R. W. Lee, T. N. Hansen, S. Techert, J. S. Wark, M. Bergh, G. Huldt, D. van der Spoel, N. Timneanu, J. Hajdu, R. A. Akre, E. Bong, P. Krejcik, J. Arthur, S. Brennan, K. Luening, J. B. Hastings, *Science*, **308**, 392 (2005).
18. J. C. H. Spence, "*High Resolution Electron Microscopy*," 3rd ed., Oxford University Press, New York (2003), and references therein.
19. (a) H. Lichte, *Phil. Trans. R. Soc. Lond. A*, **360**, 897 (2002).
(b) H. Lichte, *Ultramicroscopy*, **38**, 13 (1991).
20. (a) J. M. Cowley, "*Diffraction Physics*," 3rd ed., Elsevier, Amsterdam (1995).
(b) J. M. Cowley and A. F. Moodie, *J. Phys. Soc. Japan*, **17** (Suppl. B2), 86 (1962).
21. L. Reimer, "*Transmission Electron Microscopy: Physics of Image Formation and Microanalysis*," 2nd ed., Springer, Berlin–Heidelberg (1989).
22. D. B. Williams and C. B. Carter, "*Transmission Electron Microscopy: A Textbook for Materials Science*," Vol. I–IV, Plenum Press, New York (1996).
23. D. Shindo and K. Hiraga, "*High-Resolution Electron Microscopy for Materials Science*," Springer, Tokyo (1998).
24. Z. L. Wang, "*Elastic and Inelastic Scattering in Electron Diffraction and Imaging*," Plenum Press, New York (1995).
25. P. B. Hirsch, A. Howie, R. B. Nicholson, D. W. Pashley and M. J. Whelan, "*Electron Microscopy of Thin Crystals*", Butterworths, London (1965).
26. B. Fultz and J. M. Howe, "*Transmission Electron Microscopy and Diffractometry of Materials*," 3rd ed., Springer, Berlin–Heidelberg (2008).
27. A. Rousse, C. Rischel and J.-C. Gauthier, *Rev. Mod. Phys*., **73**, 17 (2001).

This page intentionally left blank

Chapter 3

From 2D to 3D Structural Imaging

Salient Concepts

Irrespective of whether we use a transmission electron microscope (TEM) or a scanning transmission one (STEM), the uniform electron intensity of the incident beam is in each case transformed into a nonuniform intensity after scattering by the specimen. To deduce the desired structural information, we need to recognize the *various modes of contrast* that can arise in the image. Two main modes are phase contrast, which, as discussed in Chap. 2 and below, is of great importance in structural imaging, and amplitude contrast in the sense defined in Chap. 2. There are other sources of contrast that also become relevant, including diffraction contrast and the phenomenon known as *Z*-contrast, which is of paramount importance in high-angle annular dark field (HAADF) STEM imaging and which has great relevance in the context of chemical analysis and in electron tomography.

Diffraction contrast arises typically when, because of some local deformation or bending within the specimen, the Bragg diffraction condition ($n\lambda = 2d \cdot \sin\theta$) is obeyed, thereby leading to a loss in the transmitted beam intensity at that location. *Z*-contrast is the consequence of the incoherent scattering of the incident electrons through high angles, a feature of so-called Rutherford scattering. The intensity of such scattering is proportional to the atomic number (Z) at the scattering nucleus raised (very nearly) to the power of two. *Z*-contrast can, therefore, yield the local chemical nature of the specimen, and is especially useful when heavy atoms are distributed over a light support (e.g., Pd or Pt on carbon or silica). These different contrast schemes can be elucidated with the simple ray diagrams shown in Figs. 2.8a and 3.1, where the objective lens produces a magnified image of the specimen in the so-called Gaussian image plane, and where the diffraction pattern is formed at the back focal plane of the lens.[1]

Historically, 2D and 3D imaging developed along separate lines for inorganic materials (like metals, oxides, ceramics and a few zeolites where beam damage is usually minimal), on the one hand, compared with the imaging of biological materials (like proteins and viruses which are severely damaged under electron beam radiation), on the other. And, for heuristic purposes, it is instructive to treat them here as distinct but related procedures. There are, however, numerous

common features concerning imaging that apply to both. For example, with each kind of specimen it is necessary, in order to achieve the highest image resolution, to contain within the objective aperture as many diffracted beams as possible: the larger the distance to which the diffraction spots extend in reciprocal space, the greater, in principle, will be the resolution in the magnified image.

Because an object (or its magnified image) viewed under a microscope is related to its diffraction pattern by the process of Fourier transformation — the diffraction pattern is the Fourier transform of the object — it is instructive to use optical transforms (as first suggested by Lawrence Bragg in 1938) to illustrate visually this reciprocity. Figure 3.2a, taken from the *Atlas of Optical Transforms*,[2] shows a series of diffraction patterns in which some of the diffraction spots have been eliminated by use of an appropriate aperture. Figure 3.2b, also taken from Ref. 2, shows the corresponding images that have been reconstructed (by Fourier transformation) of the respective diffraction patterns. It is immediately apparent that, to achieve high resolution, one needs to include as many diffraction spots as possible.

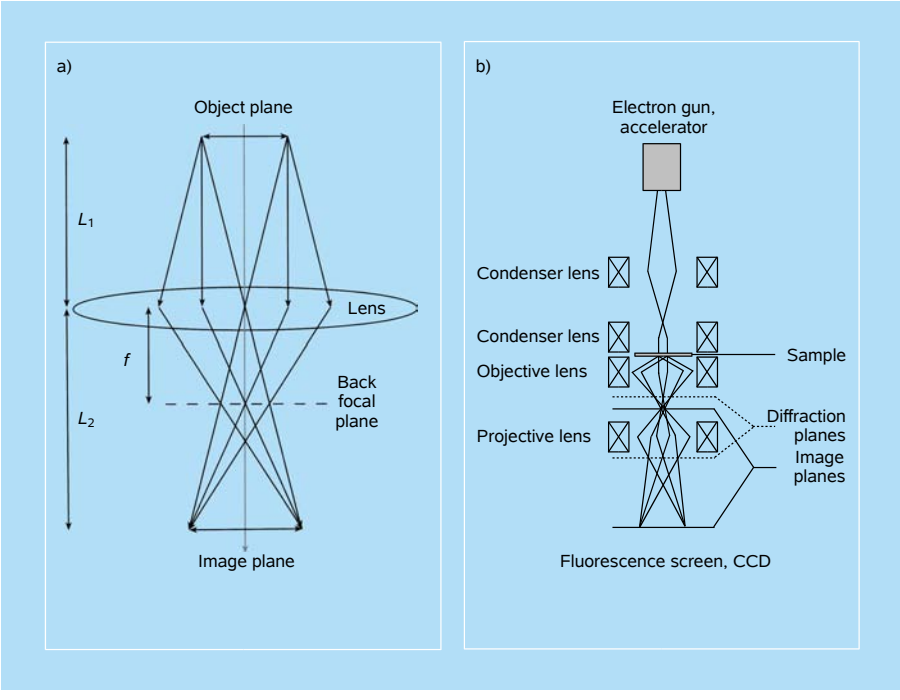


Figure 3.1 Image formation. (a) A ray diagram for a finite object, symmetrically positioned around the optical axis. All rays emerging from a point in the object (distance L_1 from the lens) that are gathered by the lens converge to a point in the image (distance L_2 from the lens), and all rays are focused in the focal plane (distance f from the lens).[1] (b) Complete ray diagram in a typical TEM; adapted from *commons.wikimedia.org*.

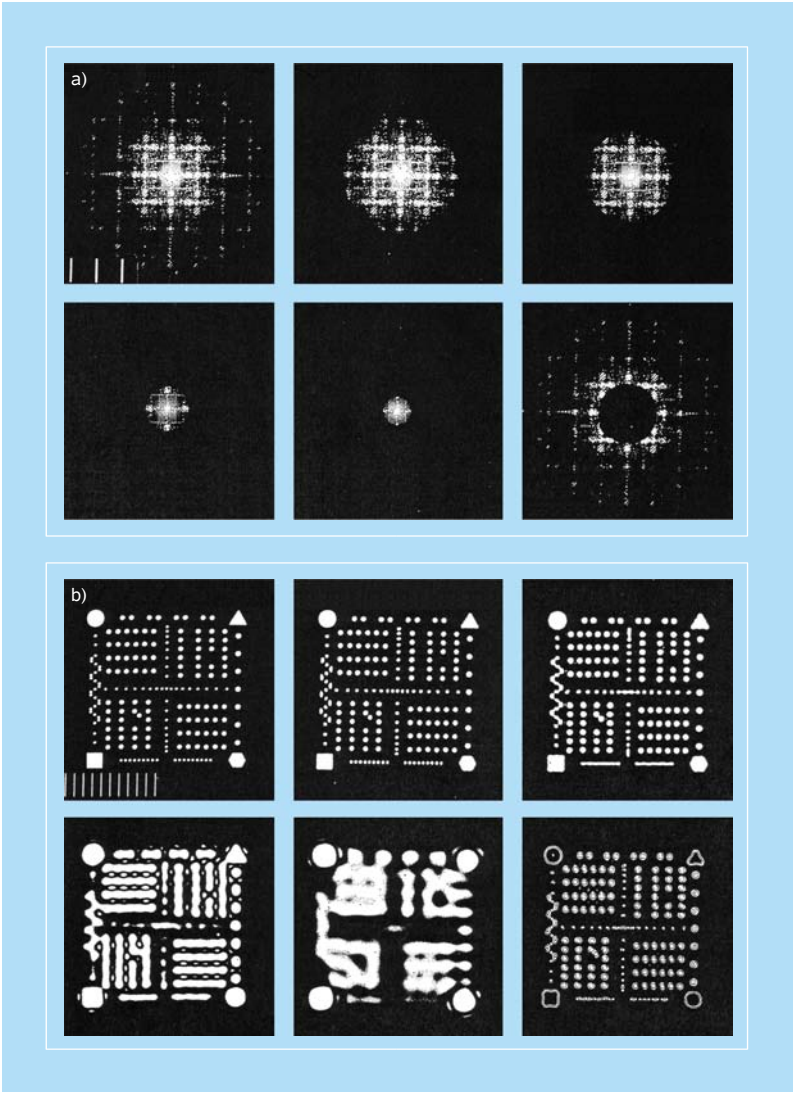


Figure 3.2 Reciprocal-space and real-space construction. (a) Fourier-space patterns limited to the indicated areas. (b) Corresponding real-space images reconstructed by Fourier synthesis.[2]

Lens imperfections may limit the number of diffracted beams that should contribute to the image. For both types of specimen (inorganic and biological) imperfections in the electron-optical lens are a key factor, as detailed in Chap. 2.

3.1 2D and 3D Imaging

Successful 2D imaging in high-resolution (HR) TEM occurs when the specimen under investigation is thin. This is because of the *weak phase object approximation*,

discussed in Chap. 2, and which is of cardinal importance in the context of deriving authentic 2D projected images in HRTEM instruments. The actual value of the "thinness" depends on the precise composition of the material. With heavy atoms (e.g., Nb, W) it may be as small as *ca.* 10 Å; with lighter elements (as in zeolites or organic materials) it may be in excess of 100 Å or so.

Because the objective lens is a "Fourier transformer," in that at the back focal plane a Fourier transform of the specimen function is generated, the diffraction pattern is further Fourier transformed to yield the image. If the lens functioned in a perfect manner, and if all the other conditions (pertaining to monochromaticity and coherence of the electron beam) were perfect, the enlarged image of the specimen would be a genuine picture of the 2D projection of the structure. As noted before, in practice, all electron-optical lenses are imperfect to a degree. It therefore becomes necessary to talk of a "*contrast transfer function*" (CTF; Chap. 2), which in effect is a quantitative measure of the reliability of the imaging that a particular lens and set of conditions will generate.

In Chap. 2, we already expressed CTF as a function of u, where u is inversely proportional to spacings in the object. The ideal form of CTF would be a constant value as u increases, as shown in Fig. 3.3. Note that, when CTF is negative, positive phase contrast results, and therefore atoms will appear dark against a bright background. When CTF is positive, negative phase contrast results, and atoms will appear bright against a dark background. When CTF is zero, there is no detail in the image for that value of u. Note that CTF becomes zero at $u = u_1$. The ideal HRTEM we would like to have is one where u is as large as possible. When CTF crosses the u axis, the sign of the transfer function reverses. It is seen, therefore, that u defines the limit at which the image may be directly interpreted by visual inspection.

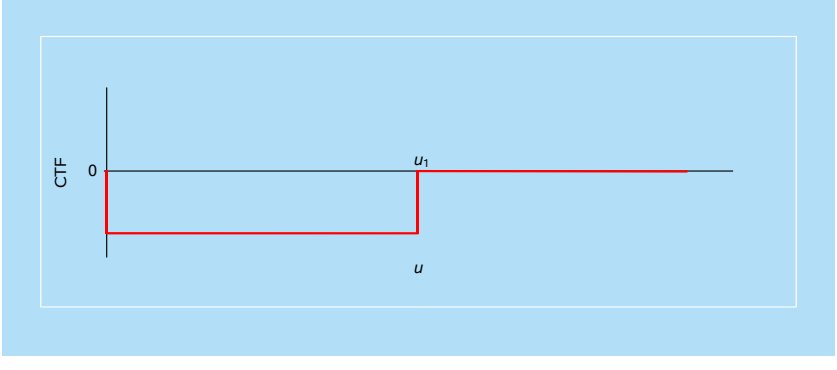


Figure 3.3 The ideal form of the contrast transfer function. In this example, CTF is large and negative between $u = 0$ and $u = u_1$.[1]

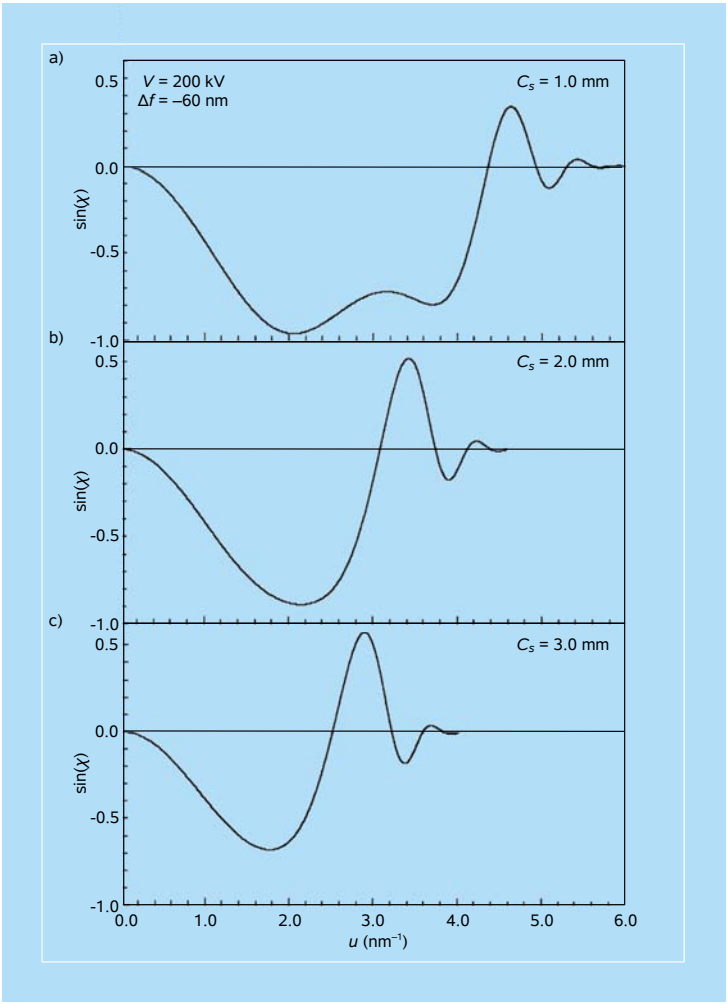


Figure 3.4 A series of sin(χ) curves calculated for different values of C_s.[1]

For a given microscope with a fixed value of spherical aberration coefficient C_s, the curve of CTF versus u, which is expressed as $\sin[\chi(u)]$, depends on the electron wavelength, λ, and also on the extent of defocusing Δf. In Chap. 2 we provided the expression [see Eqs. (2.76) and (2.77)], and it is useful here to consider the dependence on C_s. Typical values of $C_s = 1.0$, 2.0, and 3.0 mm, at the accelerating voltage of 200 kV and $\Delta f = -60$ nm, give the CTF curves shown in Fig. 3.4.[3] The resolution deteriorates as C_s increases.

In Fig. 3.5, which summarizes the essence of, and differences between, TEM and STEM, typical profiles of contrast transfer functions are shown, the one on the right being for the new generation of microscopes that are *aberration-corrected* such that C_s values are very small (0.1 mm). Note that the plots shown in Fig. 3.5

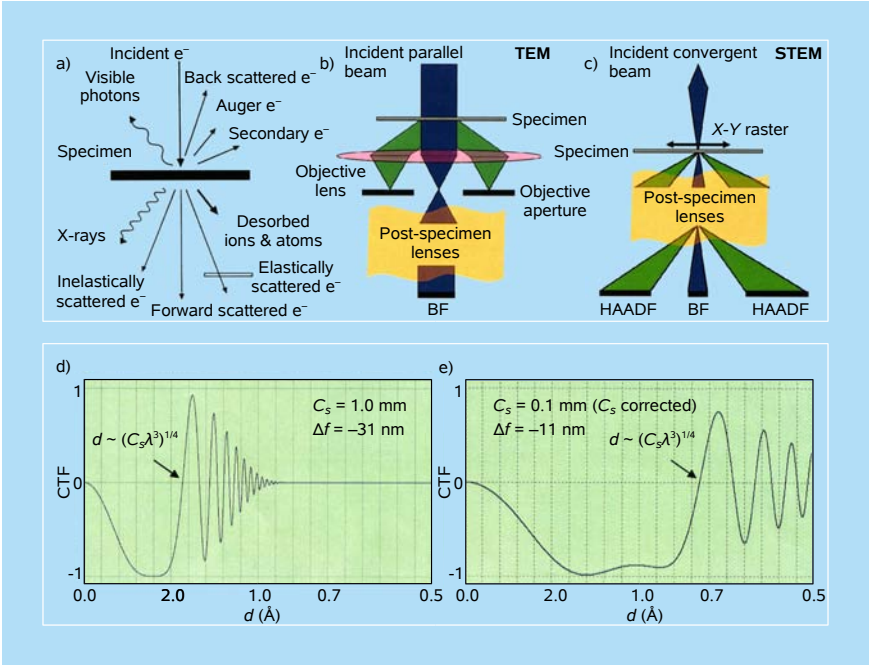


Figure 3.5 TEM versus STEM. (a–c) Schematics of the various arrangements of lenses and detectors in transmission electron microscopy (TEM) and scanning transmission electron microscopy (STEM); and (d,e) the nature of the contrast transfer function, plotted against corresponding values of separation distances (see Chap. 5), for lenses of two different values of the coefficient of spherical aberration (C_s). See text for more details.[23b]

have an envelope damping feature: in other words the contrast transfer function falls off with spatial frequency. This is due, in part, to the spatial coherence of the electron source and also to chromatic aberration, as discussed in Chap. 2. The position of the first crossover of these curves is the value of the *interpretable resolution*, d, which is given by Eq. (2.80). It is clear that as the wavelength decreases (i.e., as the accelerating voltage increases and also as C_s is diminished), the resolution is improved. For the new generation of aberration-corrected microscopes now in commercial production, the first crossover is down to *ca.* 0.8 Å, so that all spacings larger than this value are faithfully imaged.

A summary of the mathematical terms (see Chap. 2) relating to what has been described hitherto in regard to imaging using TEM is given in Fig. 3.6. Here, a distortion function is written with u and v being the dimensions in reciprocal space. In other words the *phase-distortion function* has the form of a phase shift expressed as $2\pi/\lambda$ times the path difference traveled by those waves affected by spherical aberration (C_s), defocus (Δf), and astigmatism. What this means in qualitative terms is that the beams that traverse the lens far away from its optic axis are distorted in phase more than those that are closer to that axis. The transfer

function could be optimized by balancing the effect of spherical aberration against a particular negative value of Δf, the value of *Scherzer defocus* given by Eq. (2.78). At the Scherzer defocus, all the beams will have nearly constant phase out to the "first crossover" of the zero axis of the transfer function. This crossover point is defined as the *instrumental resolution limit*. It is the best performance that can be expected from a microscope unless one uses sophisticated image processing schemes to extract more information. This limit is, in effect, the limit where we can use intuitive arguments to interpret what is seen in the 2D projected image.

In the traditional approach to HRTEM, images are recorded down two or more different zone axes. A sequence of such 2D images is readily converted to that of 3D ones. This is how the structure of the well-known zeolite called ZSM-5 (which is isostructural with silicalite and now designated by the International Zeolite Association as MFI) was determined.[4] Interestingly, this structure determination was done even before it was obtained from X-ray single crystal diffraction (see Fig. 3.7). This approach to 3D imaging has benefited enormously from the arrival of aberration-corrected TEM. HRTEM images recorded with such

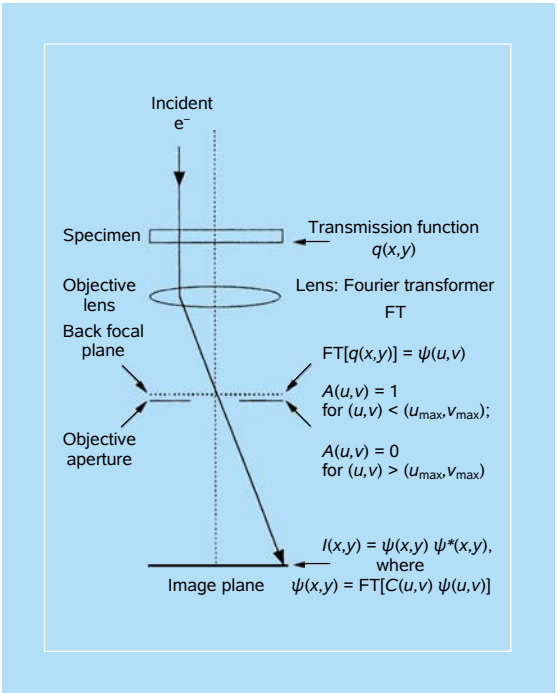


Figure 3.6 The objective lens of a microscope serves as a Fourier transformer.[23b] The diffraction pattern formed at the back focal plane is further Fourier transformed to yield the image. There are phase changes in the electron waves for (u, v). The phase contrast imaging performance of an HREM is governed by the contrast transfer function, which contains the spherical aberration coefficient of the lens and the extent of its defocus. The intensity $I(x, y)$ at the image plane is proportional to the projected electrostatic potential density in the "weak-phase" approximation (see Chap. 2).

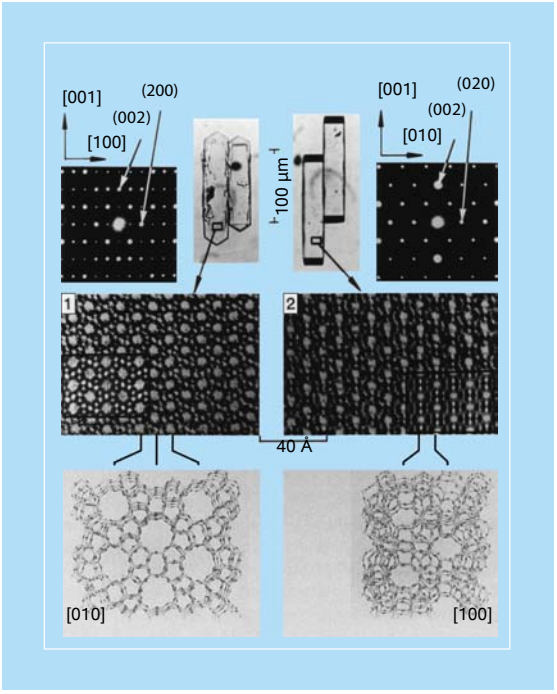


Figure 3.7 High-resolution micrographs (middle) together with selected-area diffraction patterns of ZSM-5 looking down the [010] direction (top-left) and the [100] direction (top-right). Corresponding views of the structural model are shown at the bottom. Based on the work of G. R. Millward and J. M. Thomas.[4a]

instruments now allow quantitative atomic scale analysis of crystals, and of crystal interfaces and of crystal (point) defects, where oxygen atoms and oxygen vacancies can be visualized directly; see examples cited by Thomas and Zhou[4b] and also in Refs. 5–9. The repercussions of such advances in retrieving 3D images of inorganic solids have recently been discussed by Thomas and Zhou[4b] and by Zou and Hovmöller.[8b]

Since atomic scattering amplitudes for electrons are approximately 10^4 to 10^5 times as large as they are for X-rays and neutrons, it follows that, with electrons as probes, structural information may be retrieved from single crystals of almost miniscule dimension. To illustrate this point, recall that even with the best attainable X-ray performance (with synchrotron radiation), crystal dimensions of *ca.* $2\times2\times2$ μm^3 are required (i.e., some 10^9 or more unit cells for a typical specimen). With the arrival of the so-called free-electron laser, smaller crystal dimensions and hence fewer unit cells may be studied by synchrotron radiation. Because of the strong interaction of electrons with atoms, only some 10^3 unit cells of the specimen, signifying masses of a sample of as little as 10^{-18} g, are needed to yield meaningful HREM images and diffraction patterns. When regions of the

specimen so minute can be readily assessed by electron microscopy — so that elemental composition as well as perfect and defective structures can be retrieved by EELS, X-ray emission and real-space crystallography from such miniscule volumes — one realizes the supremacy of the technique in this regard over other powerful probes such as synchrotron radiation (see also Chap. 7).

There are, however, both fundamental and practical difficulties that require caution in the interpretation of such projected images. First, there is a tendency for the 2D projections to be influenced by overlap of atoms within the structure, a situation that becomes serious when there is a compact packing of the atoms in that structure. Then there is the strong interaction between electrons and atoms that leads to multiple scattering (i.e., so-called dynamical scattering phenomena), which are quite complex, though computable.[5] In addition, there are the "corrupting" influences of the lens (contrast transfer function) alluded to earlier and also the chromatic aberration created by energy losses in the sample. Standard programs are available[6] using variants of the Cowley–Moodie multi-slice procedures,[5a,5b] to compute a matrix of images in which the influence of both defect of focus (Δf) and specimen thickness is taken into account.[7b,10]

Quite apart from the method of direct comparison of observed with calculated images for solving inorganic structures, there is also available the method of electron crystallography,[7,8] which we summarize in the next section.

3.2 Electron Crystallography: Combining Diffraction and Imaging

The approach of electron crystallography entails solving the structure of very small crystals in three dimensions using both the diffracted beams and the image in an electron microscope. (In the determination of the 2D and 3D structures of biological and organic materials, electron crystallography is of great importance; see Sec. 3.4 below.) The incoming beam is focused onto a small part of the specimen and then a "selected-area electron diffraction" (SAED) pattern is recorded using the aperture at the image plane.

One of the great advantages of solving 3D structures by electron crystallography is that the beams that form the diffraction pattern may be refocused to form an image (the electron micrograph). This electron micrograph is in essence a hologram constructed from the diffracted beams and contains all the information about the relative phases of those beams [see Eq. (2.75)]. That is, the micrograph is a result of both the reconstructed amplitude and relative phases of the diffracted beam. The process is depicted in Fig. 3.8. Since the crystallography is carried out in the electron microscope, the 3D structural determination can be concerned with regions of the crystal of differing elemental composition (a feature that is of growing importance in the study of composite materials).

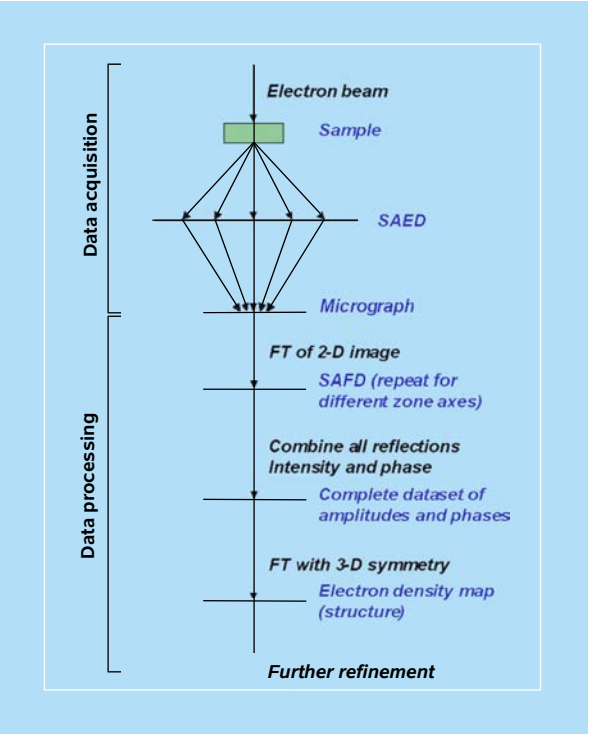


Figure 3.8 Schematics of the electron crystallography method.[7a]

The method operates by using suitable regions of the electron micrograph rather than the electron diffraction pattern itself. This is known rather inelegantly as a selected-area Fourier diffractogram (SAFD). The micrograph in essence becomes the crystal from which the structure is solved. In effect a micrograph is recorded along suitable zone axes. Then only a section of this micrograph corresponding to a particularly good part of the specimen is chosen. This section should be as thin as possible to avoid contributions from multiple scattering, and where the resolution in the image is at its highest. From this section of the micrograph a SAFD pattern may be constructed by straightforward Fourier transformation in two dimensions. Now, however, the diffraction pattern that is created contains not only the intensity of the diffracted peaks but also their relative phases, i.e., all that is required to determine the structure of the material. A difficulty arises because there is usually not enough information from one micrograph along one zone axis to solve the 3D structure. For example, a micrograph along the [100] zone axis of a cubic crystal will contain information only about crystal planes parallel to this direction (such as [011] or [002]). Consequently, in order to collect enough information required to solve the structure, micrographs are also recorded along several zone axes.

When Wenk *et al.*[8a] solved by electron crystallography the structure of the close-packed iron silicate mineral, staurolite (ideal composition $HFe_2Al_9Si_4O_{24}$), where Al ions were thought to occupy octahedral interstices and Si and Fe ions occupy tetrahedral interstices of the close-packed oxygen assembly, they used five distinct orientations of the HRTEM images (containing 59 independent reflections to a resolution of 1.38 Å) to construct a 3D electron potential map which clearly resolved all cations (Al, Si, Fe), including those with partial occupancy and all the O atoms. Numerous other examples of 3D inorganic structures determined in this fashion are given by Anderson *et al.*,[7a] Thomas *et al.*,[7b] and by Zou and Hovmöller.[8b] This technique is particularly useful in solving the structures of mesoporous solids typified by MCM-48,[7a,10] where, because their unit cell dimensions are so large and the walls of the "vacancy lattice" (composed of large channels and cages) are amorphous, the structures are well nigh impossible to solve by X-ray crystallography.

3.3 High-Resolution Scanning Transmission Electron Microscopy

Images with this technique (see Figs. 3.5 and 3.9a) are formed by scanning a finely focused probe (1–2 Å diameter) over a thin specimen. High-resolution scanning transmission electron microscopy (HRSTEM) images have slightly diminished resolutions compared with analogous HRTEM ones because of the finite source size and the reduced detection efficiency imposed by the geometry of the collection. However, an HRSTEM image has at least two important advantages: first, it has ultrahigh resolution micro-analytical (chemical composition) capabilities of X-ray emission and EELS; second, all these signals may be collected simultaneously, together with backscattered electrons. Since HRSTEM readily yields HAADF images, as well as both bright field images, formed from the transmitted electrons, and ordinary dark-field images formed from Bragg scattered beams, there are extra benefits to be gained in locating nanoparticles composed of as little as two- or three-atom clusters, and, under favorable circumstances, single atoms. As we shall see in Sec. 4.2.5 (see also Fig. 4.10), HAADF-STEM imaging is now so powerful, owing to the arrival of aberration-corrected instruments, that one may retrieve the spatial coordinates as well as the site occupancies of heavy atoms (large Z values) in complex oxide structures. We shall also see in Sec. 4.6 the spectacular results that have been obtained recently, also using a specially constructed aberration-corrected system,[10b] which, *inter alia*, enables an extremely high flux of electrons in a 1 Å primary beam (and hence boosts considerably the signal-to-noise ratio) to be focused on a single-atom column. Such a system achieves atomic-scale chemical imaging of composition and reveals the bonding of beam-resistant specimens

such as the technologically important perovskites, in particular, multilayers of $La_{0.7}Sr_{0.3}MnO_3/SrTiO_3$ (see Fig. 4.25).

One useful practical feature that STEM possesses, but which TEM does not, concerns the question of minimizing beam-damage because, with STEM, the energy that the beam endows to the specimen is more readily dissipated, as schematized in Fig. 3.9b. Another important aspect of HRSTEM is *Z*-contrast imaging (or atomic number imaging), which exploits the fact that high-angle scattering intensity of electrons from their specimens follows (very nearly) the Z^2 dependence of Rutherford's law, where *Z* is the atomic number. Thus, one Pt atom scatters in the HAADF mode as strongly as about 100 oxygen or 32 silicon atoms.

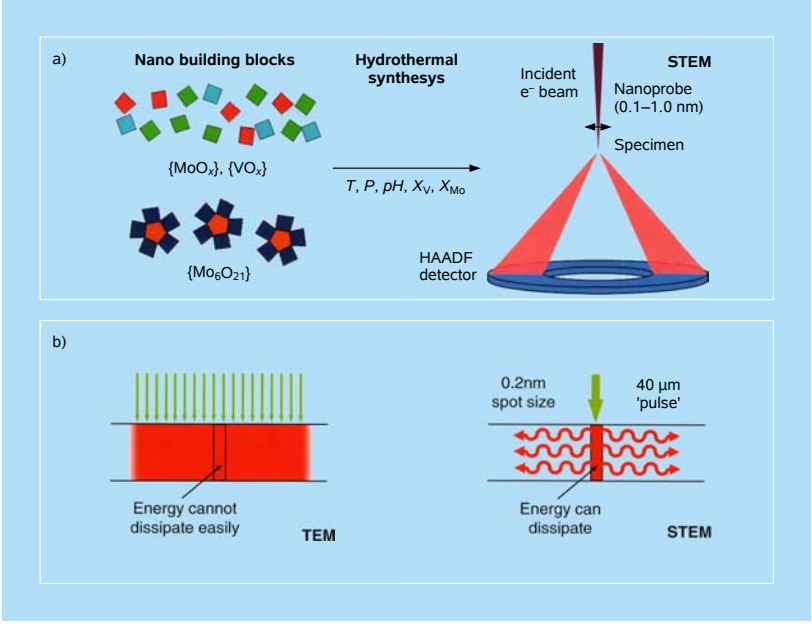


Figure 3.9 High-resolution STEM. (a) Schematic illustration of the HAADF-STEM mode of imaging complex mixed oxides and other solids containing heavy atoms (courtesy of T. Vogt and D. J. Buttrey). (b) Because of its mode of operation, a STEM instrument inflicts less beam damage than its TEM counterpart.[23b]

A particularly elegant example of HRSTEM imaging, coupled with the use of HAADF, is seen in the recent work of Li *et al.*,[11a] who recorded the 3D atomic-scale structure of size-selected gold nanoclusters consisting of 309 atoms (Fig. 3.10). They used an aberration-corrected STEM and utilized simple image simulation to discover the ino-decahedral, cuboctahedral and icosahedral geometries that they report. Significantly, they also carried out a series of integrated HAADF intensity measurements on size-selected Au_n clusters in the range $n = 55$ to 1,500 atoms. This enabled them to establish (Fig. 3.11) that there is a linear relationship between

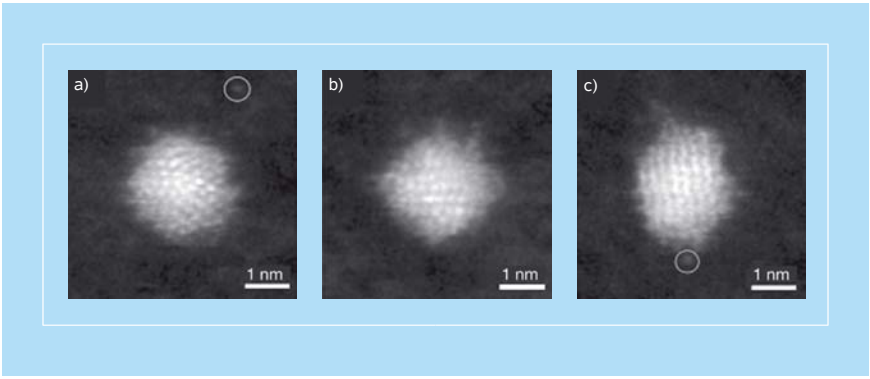


Figure 3.10 HAADF-STEM imaging of Au_{309} clusters on a carbon film. Typical images show various outline shapes, that is, cluster projections: (a) pentagon, (b) square and (c) hexagon. The intensity variation within the clusters clearly demonstrates atomic column resolution. Single atoms can be seen in the vicinity of the clusters, as indicated by the circle in (c), and occasionally some distance away, as indicated by the circle in (a). The resolution of the mass selector is ±2%.[11a] Courtesy of Z. Y. Li and R. E. Palmer.

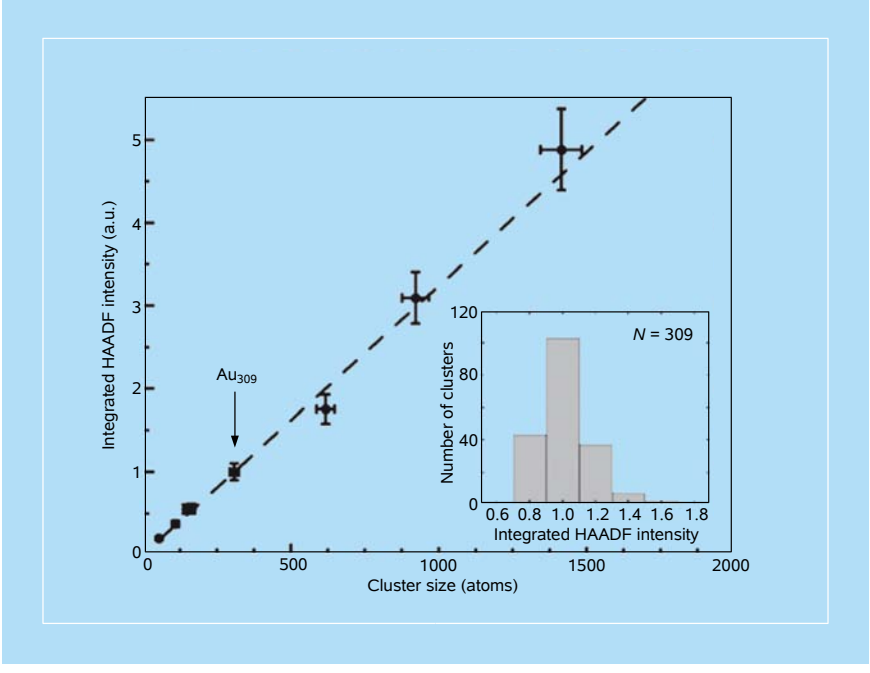


Figure 3.11 Integrated HAADF intensity of size-selected Au clusters on amorphous carbon film plotted as a function of the number of atoms they contain, showing a linear relationship. The line is drawn as a guide to the eye. Each data point is obtained from a statistical intensity distribution analysis over a large number of clusters with a given number of atoms. The standard deviation is used for estimating the error bars. An example of such a distribution for Au_{309} is shown in the inset.[11a] Courtesy of Z. Y. Li and R. E. Palmer.

integrated HAADF intensity and the size of the Au cluster. The linearity suggests that the HAADF intensities for the individual atomic columns in Fig. 3.10 can be directly associated with the number of atoms in each column.

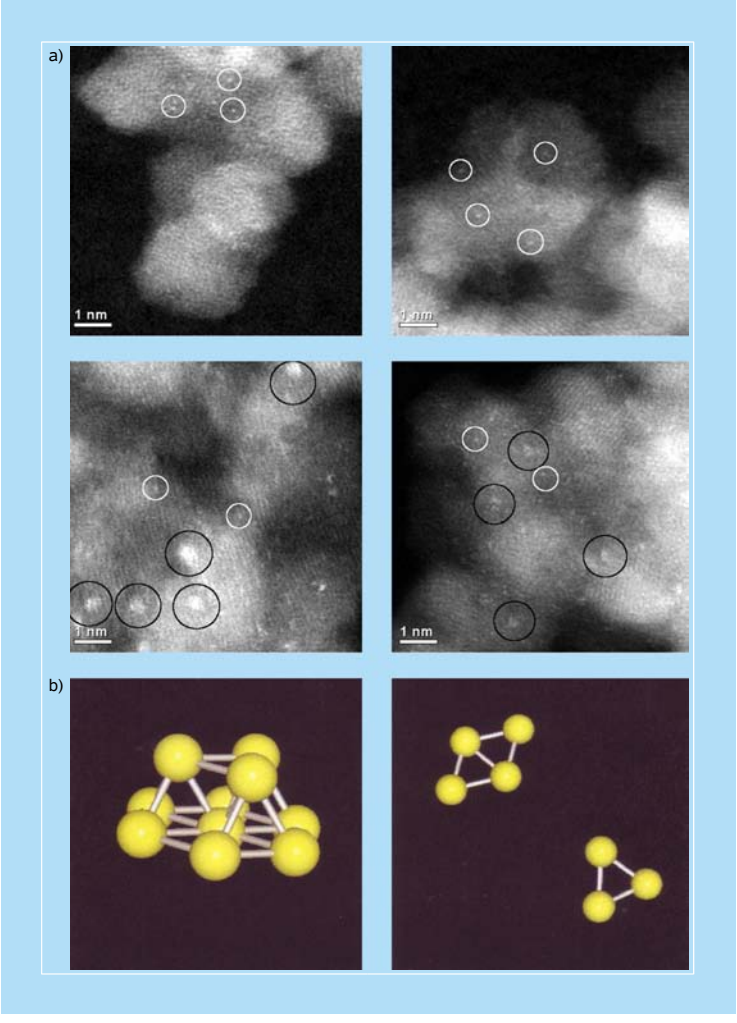


Figure 3.12 Images of catalytic clusters. (a) High-magnification aberration-corrected HAADF-STEM images of (upper) the inactive (sample 1) and (lower) the active (sample 2) Au/FeO_x catalysts. The white circles indicate the presence of individual Au atoms, whereas the black circles indicate sub-nanometer Au clusters consisting of only a few atoms. Note the presence and image intensity difference of two distinct cluster-types: In the lower left there are 0.5 nm higher-contrast clusters, whereas in the lower-right 0.2–0.3 nm low-contrast clusters dominate. This difference indicates that bilayer and monolayer sub-nanometer Au clusters are present in the active catalyst.[11b] (b) Schematic diagrams of potential clusters of (left) 0.5 nm bilayer containing 10 atoms and (right) 0.2–0.3 nm planar monolayer structures consisting of only three or four atoms. Models proposed by Kiely *et al.*[11b] for the structure of the catalytic gold clusters, as ascertained by HRSTEM used in conjunction with HAADF.

The technique of *Z*-contrast is ideally suited for detecting catalytically active nanoclusters of metals or bimetals supported on both nanoporous silicas or zeolites or on close-packed, but light, materials such as carbon or iron oxide. The work of Kiely[11b] and colleagues on the enigmatic catalytic performance of nanoclusters of gold (on iron oxide supports) has benefited enormously from the application of aberration-corrected HRSTEM in the HAADF mode, as may be assessed from Fig. 3.12a. Models of the structures of the nanoclusters of gold, as proposed by Kiely *et al.*[11b] are shown in Fig. 3.12b.

3.3.1 Use of STEM for electron tomography of inorganic materials

Like other well-known forms of tomography (X-ray or positron-emission), electron tomography (ET) enables a 3D picture to be computed (using proven mathematical procedures) from a series of projected 2D images taken over a wide range of angular orientations. The greater the number of distinct 2D images accumulated, the higher the resolution of the resulting tomogram. (The relationship between resolution, d, of a tomographic reconstruction of a cylindrical object of diameter D from N equally spaced projections covering the full angular range of 180° is given by the Crowther relationship:[12] $d \approx \pi D/N$.) Typically, the specimen to be investigated is examined in STEM mode over a range of angular settings. From such a series the tomogram is reconstructed by suitable so-called "back projection" as described by De Rosier, Klug, Crowther and others.[12] Figure 3.13 illustrates the principles involved.

The question arises as to whether the ET analysis should be performed under "bright field" (using the direct, forward-scattered beam; BF) or "dark field" (DF),

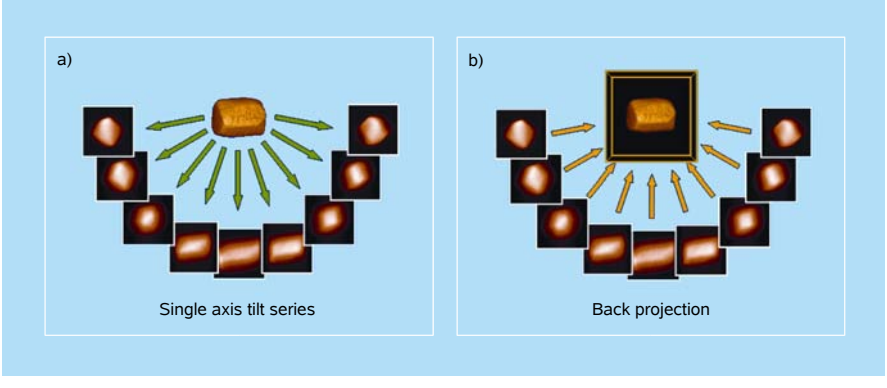


Figure 3.13 A schematic diagram of tomographic reconstruction using the back-projection method. In (a) a series of images are recorded at successive tilts. These images are back projected in (b) along their original tilt directions into a 3D object space. The overlap of all of the back-projections will define the reconstructed object.[23b] Courtesy of W. Baumeister.

where the diffracted beams are employed. We shall see below that BF imaging is the customary procedure for biological specimens, for example, in the cryotomographic work of Baumeister (see Sec. 3.4.1) at resolutions of some 40–90 Å, which are described in the subsequent section. But when attempts are made to extend the resolution of BF electron tomography, certain insurmountable electron-optical drawbacks are encountered. The coherence of the BF signal leads to strong contrast (e.g., Fresnel contrast) whenever the electron wave encounters a change in refractive index; or diffraction contrast (which produces bend contours and thickness fringes) seen in crystalline specimens, and these, as well as other complications,[13] make the BF method less attractive than the HAADF approach, which is readily carried out in STEM instruments.

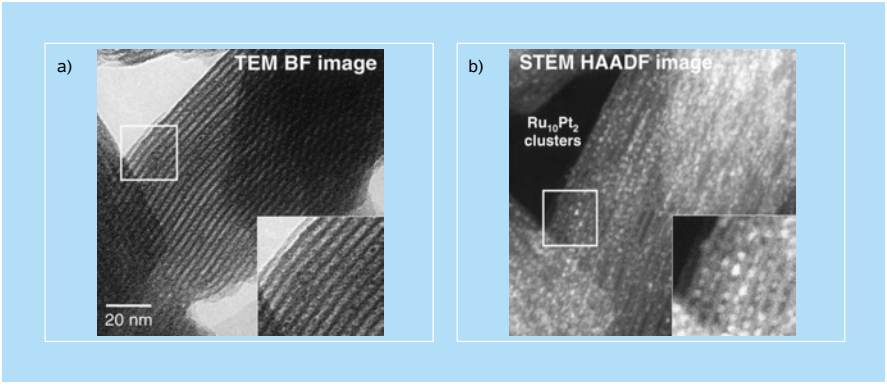


Figure 3.14 The superiority of HAADF-STEM over BF-TEM for detecting nanoparticles and clusters of metals or bimetals is illustrated here (see text).[13] The $Ru_{10}Pt_2$ nanoparticles are barely visible in the BF images, whereas they stand out clearly in the HAADF images.

The electron beams scattered by the specimen under this condition (high angles) gives *Z*-contrast and the process is incoherent, so that signals accumulated by a HAADF detector (see Fig. 3.5) are free from the complications outlined above (with BF), which are sensitive to the contrast transfer function of the lens. An authentic image of the internal structure of the specimen is thus recorded, and there are no "missing" atomic spacings (as in BF coherent imaging) arising from purely electron-optical effects. A further advantage with STEM imaging is that specimens suffer far less beam damage during investigation (which is rather long because several images, typically 70, over an angular span of $+70°$ to $-70°$ of specimen tilt, have to be recorded); see also Fig. 3.9b. The superiority of the HAADF-STEM approach over that of BF-TEM is seen clearly in Fig. 3.14, where nanocluster catalysts of $Ru_{10}Pt_2$ dispersed within the interior walls of a mesoporous silica support are clearly imaged.

3.4 Biological and Other Organic Materials

The first 3D structure of biological complexes determined by TEM were carried out by De Rosier and Klug in 1968,[12a] followed shortly thereafter by Crowther.[12b] These researchers established the methodology for solving the structure of the helical tail of bacteriophage T4 and spherical viruses, respectively. Their method entailed measuring the amplitudes from the intensities of the electron diffraction spots and the extraction of phases of the corresponding reflections from the high-resolution image. (As is well known, phases are more important in solving the structure, while amplitudes are more important for refining it.) For membrane proteins, a major breakthrough in obtaining the 3D near-atomic structure came in the work of Henderson and Unwin,[14] who solved by electron crystallography the 3D structure of bacteriorhodopsin to 7 Å, sufficient to reveal the seven transmembrane α-helices.

Because of the large depth of field of the electron microscope as compared with the size of many of the objects under investigation (notably viruses), all the features at different heights in the object are simultaneously in focus and the recorded 2D image is a projection of the 3D scattering matter in the object. In other words, features at the top and bottom and at all levels in between become superimposed on the image and are therefore difficult to unravel. To get at the 3D structure of the object, one has to look at the specimen from different directions; this can be done by tilting the object inside the microscope, or (especially for viruses) by using individual particles sitting in different orientations on the grid. De Rosier and Klug solved the problem of unraveling the projected data by exploiting Fourier transforms, which split the density in the image into all their 2D spatial frequency components.

Only relatively few 2D crystals of biological molecules have so far yielded to electron microscopic structural elucidation, aquaporin being one particular example; see Ref. 15a for a recent comprehensive list. Kornberg and Uzgiris[15b] showed the ease of forming 2D crystals of antibodies and their suitability for image analysis by electron microscopy on lipid monolayers. Kornberg *et al.*[16] grew crystals of yeast RNA polymerase preserved in vitreous ice and determined the structure by both electron crystallography and by the so-called "single-particle" method, which we shall outline next.

The single-particle method of arriving at the 3D structures of viruses, phages and other biological materials by electron microscopy essentially does crystallography without crystals.[12d] This is because individual virus particles are each treated separately, as described in Fig. 3.15, rather than imaging the whole virus crystal. Because of the extreme vulnerability of biological specimens to radiation damage in the electron microscope, it is necessary to "preserve" the specimen in vitreous ice and to conduct the examination and data collection under

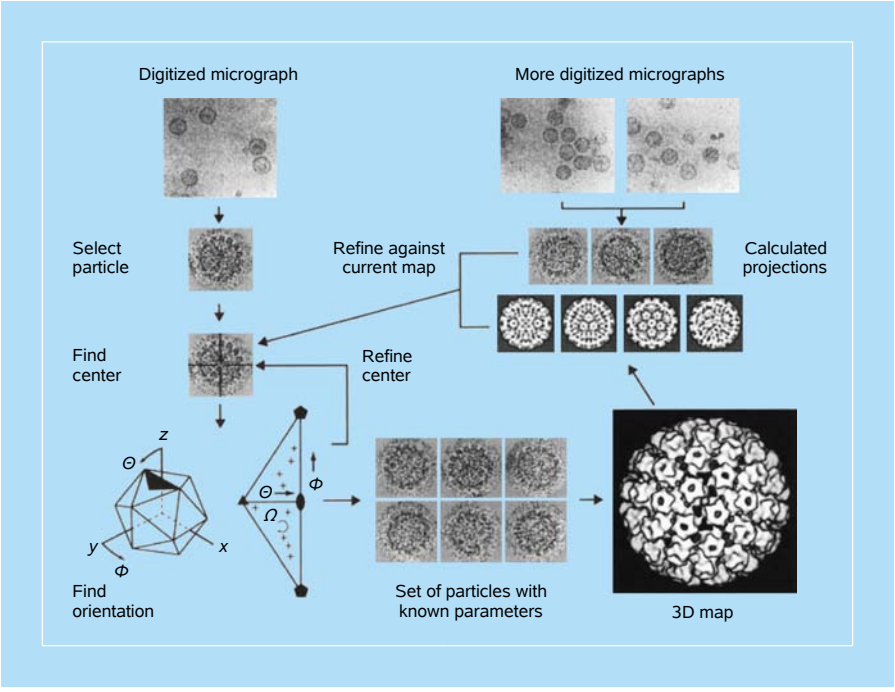


Figure 3.15 Outline of the scheme for processing the images of virus particles. The micrograph has first to be digitized on a film scanner to convert the blackness of the film into an array of numbers suitable for processing in a computer. Particles are selected and boxed out, and approximate centers and orientations relative to the icosahedral symmetry axes are found by self-common lines.[12b] A preliminary three-dimensional map is calculated from an initial set of images. From the map, a set of projections can be calculated, effectively simulating particle images but with a greatly reduced noise level. This set of projections provides a basis for determining by cross-common lines the centers and orientations of additional particles selected from further micrographs (Crowther *et al.* 1994). The whole process is iterated, increasing the number of particles included, refining their parameters more accurately and increasing the resolution of the 3D map.[35]

cryoscopic conditions.[17] The sequence of steps required to achieve a final 3D picture of the structure is shown in Fig. 3.16.

Advances in electron cryo-electron microscopy (cryo-EM) have made possible the structural determination of large biological machines at a resolution of 6–9 Å. The detailed computations and visualization methodologies necessary for the so-called "structural mining" of the computed cryo-EM maps of these sub-cellular machines have been described by Chiu,[18] a leading exponent of the technique.

3.4.1 Macromolecular architecture visualized by cryo-electron tomography

As mentioned earlier in connection with rather beam stable inorganic specimens, cryo-electron tomography (cryo-ET) acquires 2D projections of the biological

material as viewed from different directions and then merges them computationally into a 3D reconstruction, the tomogram. Again, as with inorganic specimens, ET is accomplished by tilting the specimen holder incrementally around an axis perpendicular to the electron beam.

Until recently, formidable experimental problems stood in the way of practical applications (in molecular biology) for several decades. Basically, there are two conflicting requirements, which are difficult to reconcile. On the one hand, the resolution and quality of a tomogram are directly dependent on the spacing of the 2D projections and on the angular range covered — data must be gathered over as wide a tilt range as possible, and with increments as small as possible, as mentioned earlier. On the other hand, the exposure to the electron beam must be minimized to prevent radiation damage.[19]

The development of elaborate automated data-acquisition procedures (by Baumeister *et al.*[19–21]) helped to reduce beam exposure dramatically and to take advantage of the principle of so-called *dose fractionation*. With these revolutionary improvements, ET could be applied to radiation-sensitive specimens, such as most biological materials embedded in vitreous ice. Baumeister's approach has enabled his group to characterize not only whole cells but also different organelles and their components, which include the architecture of eukaryotic cells and nuclear pore

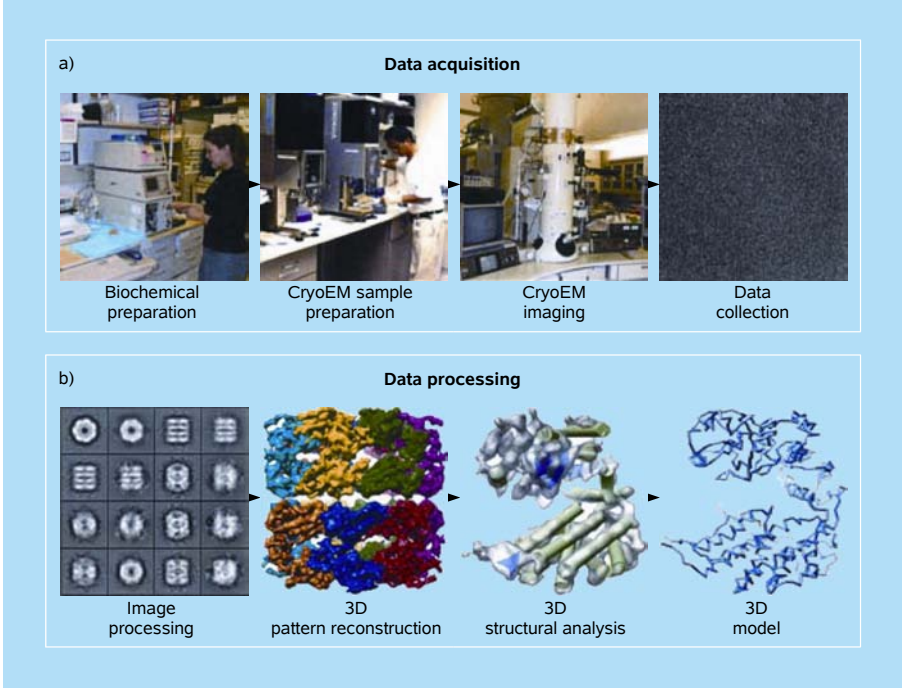


Figure 3.16 The "pipeline" in biological cryo-EM (see text). After W. Chiu; see Refs. 18 and 23b.

complexes (NPC) from *Dictyostelium discoideum* (Fig. 3.17). Another interesting application of cryo-ET (illustrated fully in Sec. 4.3 and Fig. 4.12) concerns the characterization of the magnetosome chains of magnetotactic bacteria studied by Jensen[22a] and Schüler[22b] and their co-workers (see also Fig. 3.23 below).

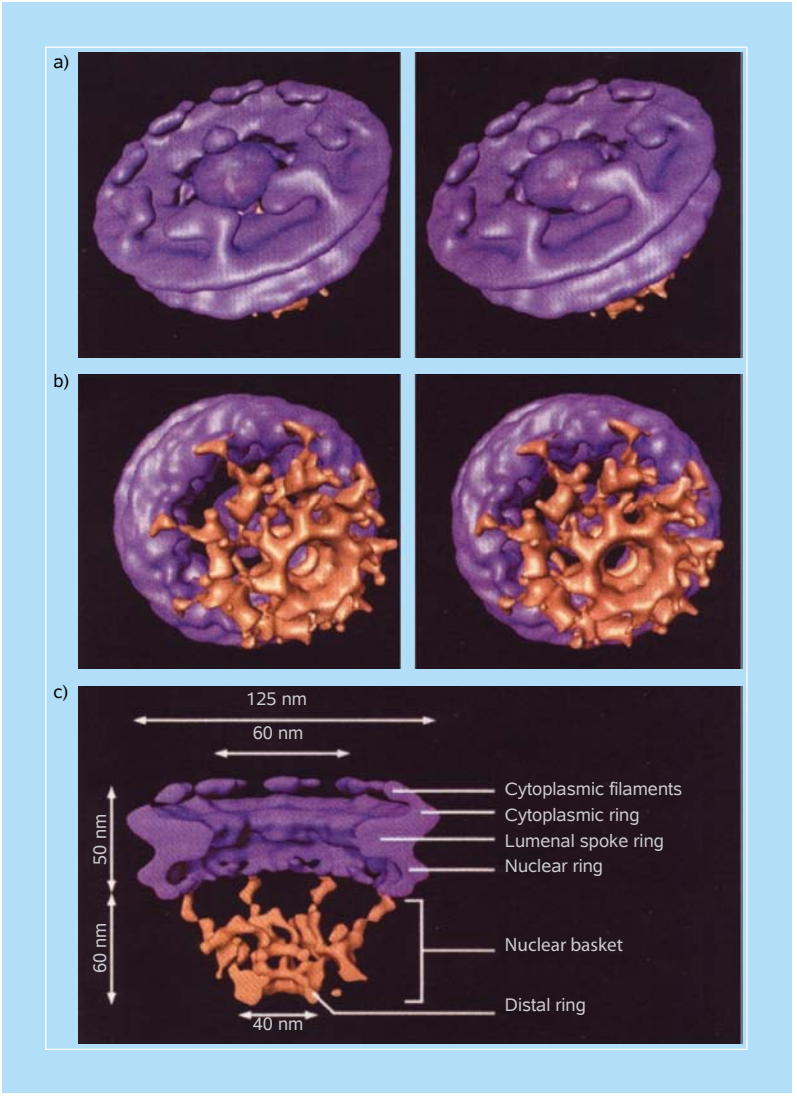


Figure 3.17 Structure of the *Dictyostelium* NPC. (a) Cytoplasmic face of the NPC in stereo view. The cytoplasmic filaments are arranged around the central channel; they are kinked and point toward the CP/T. (b) Nuclear face of the NPC in stereo view. The distal ring of the basket is connected to the nuclear ring by the nuclear filaments. (c) Cutaway view of the NPC with the CP/T removed. The dimensions of the main features are indicated. All views are surface-rendered (nuclear basket in brown). See Ref. 20.

3.5 Electron-Energy-Loss Spectroscopy and Imaging by Energy-Filtered TEM

EELS imaging employs a STEM instrument to form an image using only those electrons which have lost energy while traversing the sample to a particular excitation. If this excitation is an inner-shell process, the energy loss will be characteristic of that atomic species. Hence, by scanning the sub-nanometer probe, a map of that species will be formed using those "energy-loss" electrons. However, there are limitations to this technique arising from multiple scattering, including those due to multiple inelastic scattering in addition to elastic scattering. More importantly, for all but the very thinnest samples, we must remember that it is not possible to prevent the elastic scattering of inelastically scattered electrons. Hence, if we set our spectrometer to collect all those electrons which have lost, say, 285 eV, corresponding to the K-ionization energy of carbon, we will also collect all those electrons which have first ionized a carbon atom and then been elastically scattered by, say, a stacking fault. Then the image we form will show both the carbon atoms and the stacking faults. The example of the stacking faults can be replaced by any other source of diffraction contrast, including bending effects, and these will show up in the image.[23a]

The great merit of EELS and electron-loss imaging in STEM and energy-filtered TEM (EFTEM) is the wealth of compositional and structural information they yield,[23] as is apparent in Figs. 3.18 and 3.19. Valence states, bond distances and composition may all, under ideal conditions, be retrieved from the kind of spectra shown in Fig. 3.18. Even plasmon-loss spectroscopy[24] and the Compton scattering envelope[25] can access information that is well-nigh inaccessible by other means, such as, respectively, the composition of alkali-metal alloys and the electronic bonding in amorphous carbon. It is now readily feasible[25a] to extend the analysis of EELS spectrum imaging data from elemental to coordination fingerprinting[25b] to bond mapping in complex nanostructures (especially those containing boron).

By ingeniously combining the EELS approach with that of ET (as described above), Midgley has devised a new technique labeled *volume spectroscopy*, whereby the composition of a minute volume in the interior of a specimen may be nondestructively and noninvasively determined down to the sub-attogram (10^{-18} g) level.[26] This new method of quantitative analysis combines EFTEM measurements at every tilt angle that is involved in the concomitant energy-loss tomography; see Fig. 3.20a,b. Advantage has been taken of this approach in the nanochemical characterization of composites consisting of nylon and multi-walled carbon nanostructures (Fig. 3.20c,d). Although this chapter deals with the rudiments of various aspects of (static) 2D to 3D imaging, we draw the reader's attention to recent advances in the femtosecond time-resolved EELS made at Caltech (see Chap. 6).

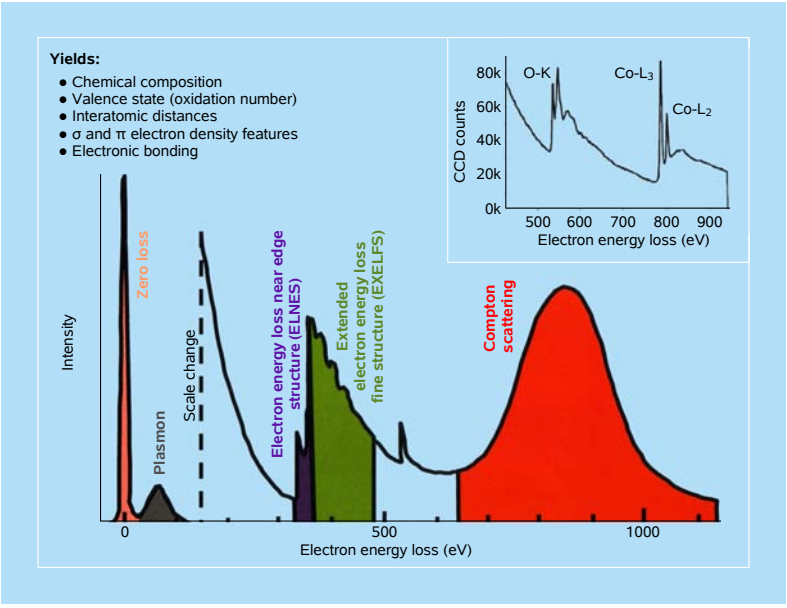


Figure 3.18 EELS yields a wealth of structural and electronic information, including chemical composition, valence state (or oxidation number, obtained from ELNES), interatomic distances (obtained from EXELFS), information pertaining to σ- and π-electrons (obtained from plasmon peaks) and momentum density of bonding electrons (obtained from Compton scattering).[23,27,36] The inset was adapted from Ref. 37.

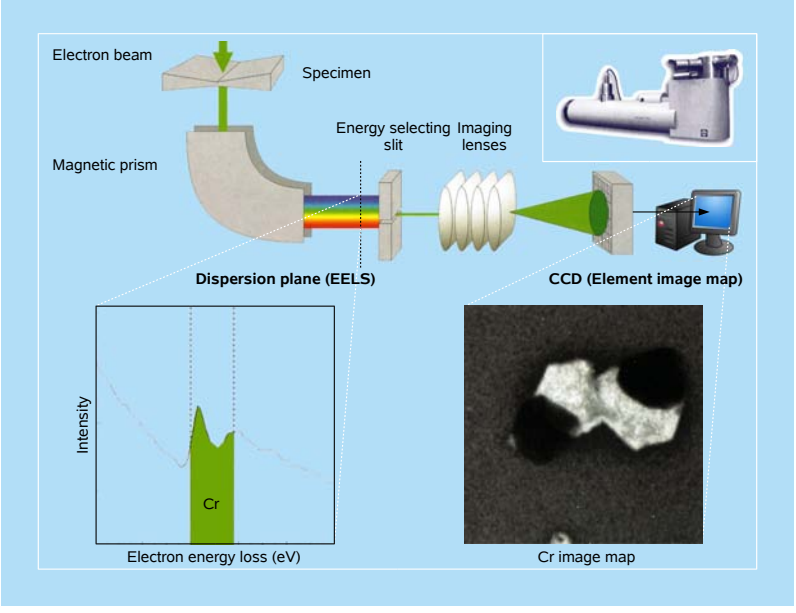


Figure 3.19 Energy-filtered TEM is carried out using a so-called Gatan Imaging Filter, schematized here.[23b] Courtesy of M. Weyland and P. A. Midgley.

3.5.1 Combined EELS and ET in cellular biology

Although the electron-beam susceptibility of biological macromolecules will not allow the degree of spatial resolution achievable with inorganic systems, EELS and EFTEM as well as ET have an important role to play in cellular biology, as may be gleaned from the recent work of Leapman *et al.*,[27] who mapped phosphorus-containing ribosomes within a cell (see Fig. 3.21). There is no discernible phosphorus EELS ($L_{2,3}$) signal for the cytoplasm, but there is for the ribosome. Even though the attainable resolutions are modest, the location of individual ribosomes at different heights within sections of the *Caenorhabditis elegans* cell may be unmistakably identified. A combination of EFTEM and ET could yield further insights into cellular structures and compositions, as outlined elsewhere,[23] for example:

- the total distribution of biomolecules, including proteins and nucleic acids, from nitrogen EELS signals;
- the proteins that contain high levels of the amino acids cysteine and methionine from sulfur signals; and
- the distribution of nucleic acids, phosphorylated proteins and phospholipids from phosphorous signals.

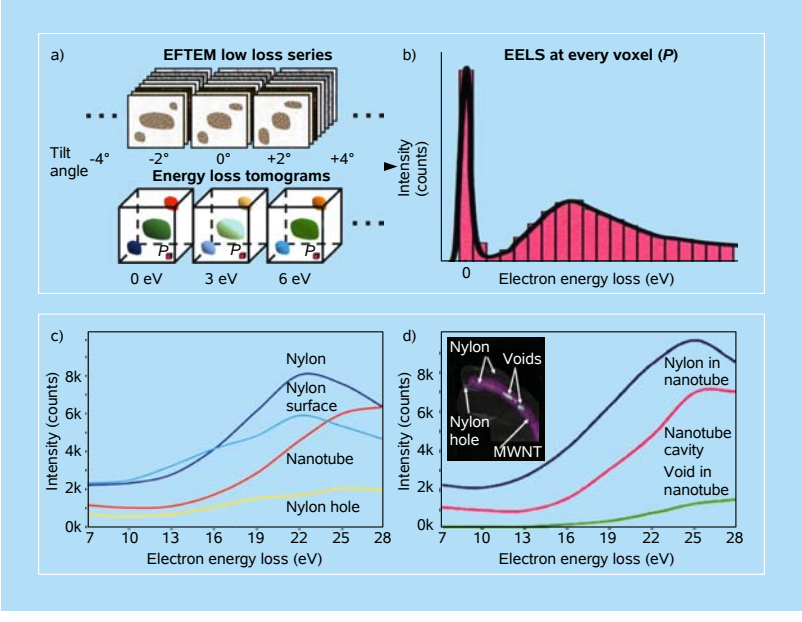


Figure 3.20 By recording both a series of frames for tomographic analysis, and an EELS spectrum at each angle of tilt, Midgley *et al.* have introduced a method of chemical analysis known as volume spectroscopy. This enables the composition of sub-attogram quantities of the specimen to be determined by nondestructive methods. See Refs. 26b and 23b. Courtesy of P. A. Midgley and A. H. Windle.

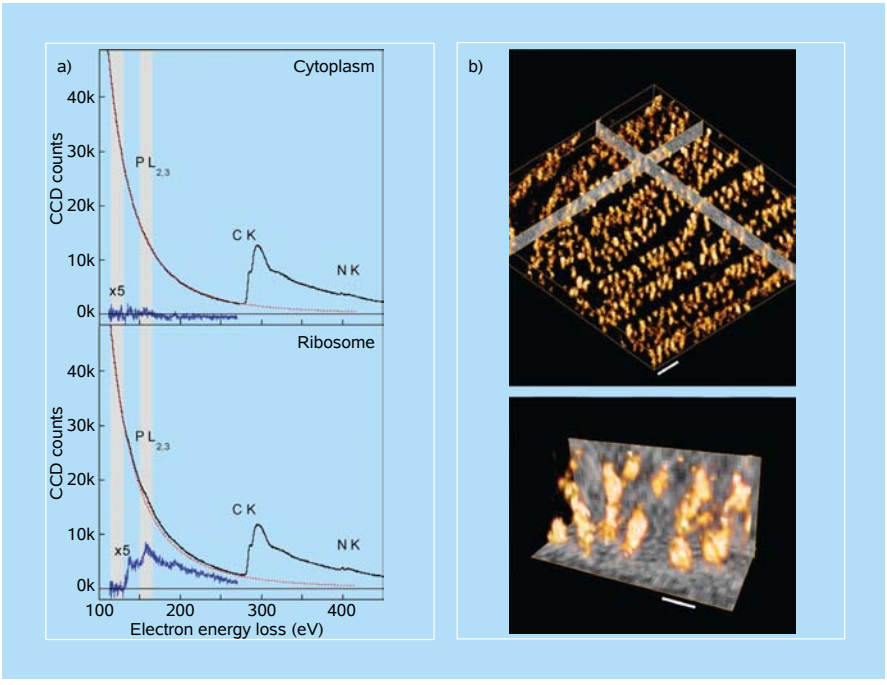


Figure 3.21 Leapman *et al.* have combined EELS and volume-rendered tomography to reveal rows of ribosomes in endoplasmic reticulum membranes. (a) EELS from 10 nm diameter region of cytoplasm from cell in *C. elegans*, obtained from STEM spectrum-image (top). EELS from 10 nm diameter region of single ribosome. No significant phosphorus is detected in the cytoplasm, whereas a clear phosphorus signal is evident in the spectrum from the ribosome with a P $L_{2,3}$ signal-to-background fraction of approximately 0.1 at an energy loss of 160 eV. The background is fitted from 110 to 130 eV and extrapolated above the phosphorus edge (dashed curve); subtraction of the fitted background produces the characteristic P $L_{2,3}$ shape in the spectrum (bottom). Energy windows used for EFTEM ratio mapping are indicated on each spectrum. (b) Volume-rendered, tomographic reconstruction of phosphorus in a section of *C. elegans* cell: rows of ribosomes are evident along stacks of endoplasmic reticulum membranes. Top: Slices through the reconstruction of the *x*–*z* and *y*–*z* planes are also shown. Bar = 100 nm. Bottom: Higher magnification of volume-rendered phosphorus distribution showing individual ribosomes located at different heights within the section. Bar = 20 nm. Adapted from Ref. 38.

This methodology could be further extended to the detailed 3D mapping of certain key inorganic elements within living cells. The importance of potassium can hardly be overemphasized, and the growing realization that RNA folding is strongly dependent on the concentration of Mg^{2+} ions makes this an important determinant.[23,27] Although EELS in STEM microscopes provides high sensitivity for measuring the important element Ca^{2+} in biological systems, the technique has been difficult to apply routinely because of the required long accumulation times. Leapman has recently refined the complementary analytical technique of EFTEM that enables rapid imaging of large cellular regions and measurement of Ca^{2+} concentrations approaching physiological levels.[27b]

3.6 Electron Holography

Any electron microscope that uses a source of coherent electron waves is, in principle, capable of yielding holographic information of the kind shown in Fig. 3.22, where the lines of force of, for example, an incarcerated nanoferromagnet are imaged. The technique of electron holography (EH) can be used to demonstrate the nature of magnetic vector fields in nanomaterials.[28,29] In particular, unprecedented insights into the magnetic nature of magnetotactic bacteria,[22,29,30] nanoparticle (magnetic) chains and chiral bracelets,[31] and geochemically relevant-phenomena involving exsolution (the unmixing of two mineral phases) in a magnetic solid[32] have been achieved.

The off-axis TEM mode of recording electron holograms involves the examination of an electron-transparent specimen using defocused illumination from a highly coherent electron source (such as a field-emission gun). The region of interest is positioned so that it covers approximately half the field of view (see Fig. 3.22). The application of a voltage to an electron biprism results in overlap of a "reference" electron wave that has passed through vacuum with another part of the

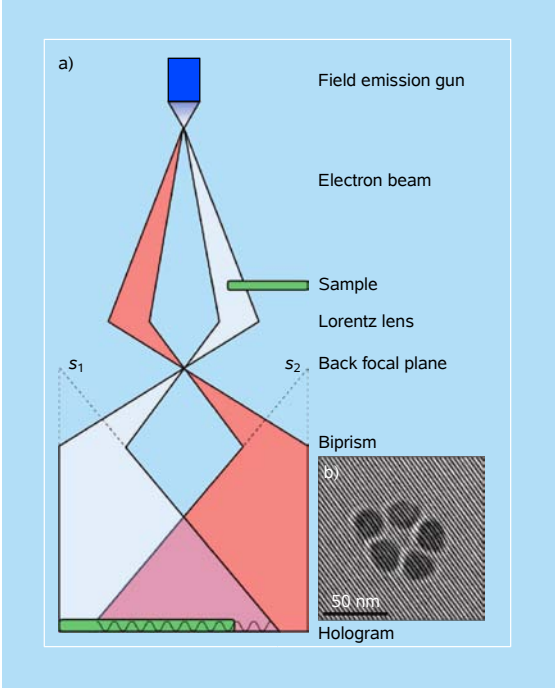


Figure 3.22 Electron holography. (a) Schematic diagram illustrating the application of a voltage to an electron biprism located close to a conjugate image plane in the microscope, in order to overlap a "reference" electron wave that has passed through vacuum with the electron wave that has passed through the specimen, to form an off-axis electron hologram. (b) Representative off-axis electron hologram of five 20−30 nm cobalt nanoparticles.[29]

same electron wave that has passed through the specimen. Since the electron source is then coherent, in addition to an image of the specimen, an interference fringe pattern is formed in the overlap region, as shown in Fig. 3.22b. The amplitude and the phase shift of the electron wave that leaves the specimen are recorded in the intensity and the position, respectively, of the holographic interference 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 specimen located in magnetic-field-free conditions.

The measured phase shift is sensitive both to the in-plane component of the magnetic induction and to the electrostatic potential in the specimen. If it is assumed that the latter is weakly diffracting (i.e., ignoring dynamical scattering contributions to the contrast), then the phase shift as expressed in the notation of Ref. 33 is

$$\psi(x) = C_E \int V(x,z)dz - \frac{e}{\hbar}\iint B_\perp(x,z)dxdz, \tag{3.1}$$

where 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 x and z, V is the electrostatic potential, and C_E is a constant value of 6.53×10^6 rad V^{-1} m^{-1} for the acceleration voltage of 300 kV. If neither V nor $B_\perp$ varies along the direction of the electron beam in a specimen of thickness t, and in the absence of electrostatic or magnetic fringing fields outside the specimen, Eq. (3.1) may be simplified as

$$\psi(x) = C_E V(x)t(x) - \frac{e}{\hbar}\int B_\perp(x)t(x)dx. \tag{3.2}$$

Using Eqs. (3.1) and (3.2), information about both V and $B_\perp$ may be recovered from a measured phase change. Of particular interest for magnetic materials is the (digital) determination of the gradient of the phase image. The phase gradient is directly proportional to the in-plane component of the magnetic induction in the specimen, and a graphical representation of the strength and direction of the local projected in-plane magnetic induction may be obtained simply by adding contours to the recorded magnetic contributions to the phase image. While this method recovers the total field B, for the study of magnetic materials one is normally more interested in obtaining a map of the material's property M, the magnetization or dipole moment density. To obtain this from B, it is necessary to map out the fringing fields due to H.

The magnetotactic bacteria[22,29,30] that contain single or multiple chains of crystals of magnetite (Fe_3O_4) that are between 350 to 1200 Å in length are shown in various electron microscopic (BF, tomographic and holographic) modes in Fig. 3.23. Many other such examples of magnetic vector fields are given in the recent review by Thomas *et al.*[29] The extremely rapid development of combined ET and EH, recently chronicled by Midgley and Dunin-Borkowski[34]

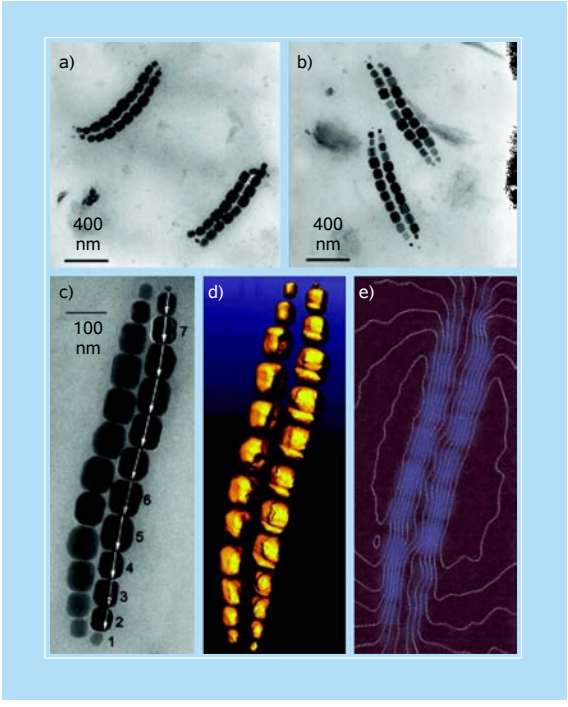


Figure 3.23 Magnetotactic bacteria imaging. (a,b) Low-magnification bright-field images of magnetotactic bacteria that were air-dried onto carbon TEM grids. Each cell contains two pairs of double chains of magnetite (Fe_3O_4) crystals. (c) BF TEM image of one double chain of magnetite crystals from a cell similar to those shown in panels (a) and (b). The white arrows correspond to directions in the crystals that were identified as [111]. (d) Isosurface visualization of a HAADF tomographic reconstruction of the double magnetosome chain shown in the previous panel. (e) Magnetic induction map recorded at room temperature using off-axis electron holography from the same double chain of magnetite crystals. See Ref. 29, and references therein.

with specific reference to inorganic materials, in which they describe retrieval of both electrostatic and magnetic potentials in three dimensions, coupled with the diversity of imaging modes available in EM, highlights afresh the great scope that now exists for the quantitative measurements of nanomaterials utilizing *in situ* environmental cells and the ability to examine working devices under applied bias or during mechanical deformation.

References

1. D. B. Williams and C. B. Carter, "*Transmission Electron Microscopy: A Textbook for Materials Science*," Vol. 3, Plenum Press, New York, p. 461 (1996).
2. G. Harburn, C. A. Taylor and T. R. Welberry, "*Atlas of Optical Transforms*," Bell and Hyman, London (1975).
3. D. B. Williams and C. B. Carter, "*Transmission Electron Microscopy: A Textbook for Materials Science*," Vol. 3, Plenum Press, New York, p. 464 (1996).

4. (a) J. M. Thomas, G. R. Millward and L. A. Bursill, *Phil. Trans. R. Soc. Lond. A*, **300**, 43 (1981).
   (b) J. M. Thomas and W. Zhou, *Chem. Phys. Chem*., **4**, 927 (2003).
5. (a) J. C. H. Spence, "*High Resolution Electron Microscopy*," 3rd ed., Oxford University Press, New York (2003).
   (b) J. M. Cowley, "*Diffraction Physics*," 3rd ed., Elsevier, Amsterdam (1995).
6. P. A. Stadelmann, *Ultramicroscopy*, **21**, 131 (1987).
7. (a) M. W. Anderson, T. Ohsuna, Y. Sakamoto. Z. Liu, A. Carlsson and O. Terasaki, *Chem. Commun*., 907 (2004).
   (b) J. M. Thomas, O. Terasaki, P. L. Gai, W. Zhou and J. M. Gonzalez-Calbet, *Acc. Chem. Res*., **34**, 583 (2001).
8. (a) H. R. Wenk, K. H. Downing, H. Meisheng and M. A. O'Keefe, *Acta. Cryst. A*, **48**, 700 (1992).
   (b) X. Zou and S. Hovmöller, *Acta. Cryst. A*, **64**, 149 (2008).
9. (a) C. L. Jia, M. Lentzen and K. Urban, *Science*, **299**, 870 (2003).
   (b) C. L. Jia and K. Urban, *Science*, **303**, 2001 (2004).
10. (a) M. Kaneda, T. Tsubakiyama, A. Carlsson, Y. Sakamoto, T. Ohsuna, O. Terasaki, S. H. Joo and R. Ryoo, *J. Phys. Chem. B*, **106**, 1256 (2002).
    (b) D. A. Muller, L. Fitting Kourkoutis, M. Murfitt, J. H. Song, H. Y. Hwang, J. Silcox, N. Dellby and O. L. Krivanek, *Science*, **319**, 1073 (2008).
11. (a) Z. Y. Li, N. P. Young, M. Di Vece, S. Palomba, R. E. Palmer, A. L. Bleloch, B. C. Curley, R. L. Johnston, J. Jiang and J. Yuan, *Nature*, **451**, 46 (2008).
    (b) A. A. Herzing, C. J. Kiely, A. F. Carley, P. Landon and G. J. Hutchings, *Science*, **321**, 1331 (2008).
12. (a) D. J. De Rosier and A. Klug, *Nature*, **217**, 130 (1968).
    (b) R. A. Crowther, *Phil. Trans. R. Soc. Lond. B*, **261**, 221 (1971).
    (c) R. Henderson, *Quart. Rev. Biophys.*, **37**, 3 (2004).
    (d) R. M. Glaeser, *Proc. Natl. Acad. Sci. USA*, **105**, 1779 (2008).
13. J. M. Thomas, P. A. Midgley, T. J. V. Yates, J. S. Barnard, R. Raja, I. Arslan and M. Weyland, *Angew. Chem*., *Intl. Ed.*, **43**, 6745 (2004).
14. P. N. T. Unwin and R. Henderson, *J. Mol. Biol.*, **94**, 425 (1975).
15. (a) D. L. Dorset, *Ultramicroscopy*, **107**, 453 (2007).
    (b) E. E. Uzgiris and R. D. Kornberg, *Nature*, **301**, 125 (1983).
16. F. J. Asturias, W. Chang, Y. Li and R. D. Kornberg, *Ultramicroscopy*, **70**, 133 (1998).
17. (a) M. Adrian, J. Dubochet, J. Lepault and A. W. McDowall, *Nature*, **308**, 32 (1984).
    (b) R. M. Glaeser, K. Downing, D. De Rosier, W. Chiu and J. Frank, "*Electron Crystallography of Biological Macromolecules*," Oxford University Press, New York (2007).
    (c) R. M. Glaeser, *Phys. Today*, **61**, 48 (2008).
18. W. Chiu, M. L. Baker, W. Jiang, M. Dougherty and M. F. Schmid, *Structure*, **13**, 363 (2005).
19. O. Medalia, I. Weber, A. S. Frangakis, D. Nicastro, G. Gerisch and W. Baumeister, *Science*, **298**, 1209 (2002).
20. M. Beck, F. Förster, M. Ecke, J. M. Plitzko, F. Melchior, G. Gerisch, W. Baumeister and O. Medalia, *Science*, **306**, 1387 (2004).

21. M. Cyrklaff, C. Risco, J. J. Fernández, M. V. Jiménez, M. Estéban, W. Baumeister and J. L. Carrascosa, *Proc. Natl. Acad. Sci. USA*, **102**, 2772 (2005).
22. (a) A. Komeili, Z. Li, D. K. Newman and G. J. Jensen, *Science*, **311**, 242 (2006).
(b) A. Scheffel, M. Gruska, D. Faivre, A. Linaroudis, J. M. Plitzko and D. Schüler, *Nature*, **440**, 110 (2006).
23. (a) J. C. H. Spence and J. Lynch, *Ultramicroscopy*, **9**, 267 (1982).
(b) J. M. Thomas, in "*Physical Biology: From Atoms to Medicine*," Ed. A. H. Zewail, Imperial College Press, London, p. 51 (2008).
24. J. M. Thomas, B. G. Williams and T. G. Sparrow, *Acc. Chem. Res.*, **18**, 324 (1985).
25. (a) R. Arenal, F. de la Peña, O. Stéphan, M. Walls, M. Tencé, A. Loiseau and C. Colliex, *Ultramicroscopy*, **109**, 32 (2008).
(b) R. Brydson, H. Sauer, W. Engel, J. M. Thomas and E. Zeitler, *J. Chem. Soc., Chem. Commun*., 1010 (1989).
(c) B. G. Williams and J. M. Thomas, *Intl. Rev. Phys. Chem.*, **3**, 39 (1983).
26. (a) P. A. Midgley, E. P. W. Ward, A. B. Hungría and J. M. Thomas, *Chem. Soc. Rev.*, **36**, 1477 (2007).
(b) M. H. Gass, K. K. K. Koziol, A. H. Windle and P. A. Midgley, *Nano Lett.*, **6**, 376 (2006).
27. (a) R. D. Leapman, *Curr. Opin. Neurobiol.*, **14**, 591 (2004).
(b) M. A. Aronova, Y. C. Kim, N. B. Pivovarova, S. B. Andrews and R. D. Leapman, *Ultramicroscopy*, **109**, 201 (2009).
28. A. Tonomura, "*The Quantum World Unveiled by Electron Waves*," World Scientific, Singapore (1998).
29. J. M. Thomas, E. T. Simpson, T. Kasama and R. E. Dunin-Borkowski, *Acc. Chem. Res.*, **41**, 665 (2008).
30. R. E. Dunin-Borkowski, M. R. McCartney, R. B. Frankel, D. A. Bazylinski, M. Pósfai and P. R. Buseck, *Science*, **282**, 1868 (1998).
31. S. L. Tripp, R. E. Dunin-Borkowski and A. Wei, *Angew. Chem., Intl. Ed.*, **42**, 5591 (2003).
32. G. D. Price, *Am. Mineral.*, **66**, 751 (1981).
33. R. E. Dunin-Borkowski, T. Kasama and R. J. Harrison, in "*Nanocharacterisation*," Eds. A. I. Kirkland and J. L. Hutchison, Royal Society of Chemistry, Cambridge (2007).
34. P. A. Midgley and R. E. Dunin-Borkowski, *Nature Mater.*, **8**, 271 (2009).
35. R. A. Crowther, *Phil. Trans. R. Soc. B*, **363**, 2441 (2008).
36. R. F. Egerton, "*Electron Energy Loss Spectroscopy in the Electron Microscope*," 2nd ed., Plenum Press, New York (1996).
37. Z. L. Wang, J. Bentley and N. D. Evans, *Micron*, **31**, 355 (2000).
38. R. D. Leapman, E. Kocsis, G. Zhang, T. L. Talbot and P. Laquerriere, *Ultramicroscopy*, **100**, 115 (2004).

This page intentionally left blank

Chapter 4

Applications of 2D and 3D Imaging and Related Techniques

4.1 Introduction

Ever since Menter, Hirsch, Whelan, Howie, Amelinckx, Allpress, Anderson, Cosslett, Cowley, Moodie, Iijima, Thomas and others from the mid-1950s onwards popularized the power and importance of the examination of thin films of materials by electron microscopy, chemists, materials scientists, applied physicists and earth scientists have found it almost indispensable to deploy this technique, often in conjunction with others, to elucidate the structure and properties of condensed matter. Equally, on the biological front, the breakthroughs by Klug, Henderson, Unwin, Crowther, Frank, Baumeister, Glaeser, Chiu and others from the 1960s onwards have spawned an enormous and growing body of investigations that has led to many advances in our knowledge of the macromolecules of life.

While we shall dwell in this chapter more on examples of 2D and 3D structural imaging, we shall also deal with associated chemical analytical procedures and outline the value of less sophisticated techniques such as electron diffraction, especially in the investigation of superstructures formed by many inorganic materials. Throughout, we shall build upon the key principles presented in Chaps. 2 and 3, but we will also refer to a few technicalities of other procedures not hitherto discussed in this monograph.

4.2 Real-Space Crystallography via HRTEM and HRSTEM

Under ideal imaging conditions (described in Chap. 3), remarkable insights into the nature of nanoscopic materials by both TEM and STEM instruments have been gained. We now proceed to describe specific examples.

4.2.1 Encapsulated nanocrystalline structures

Kirkland and Sloan[1] examined the structure taken up by potassium iodide (KI) when it is encapsulated in a single-wall (carbon) nanotube (SWNT). A typical

HRTEM image is shown in Fig. 4.1, from which it is seen that the binary ionic halide is tightly accommodated in a manner that can be quantitatively imaged. To improve the resolution and the interpretability of such structural data, Kirkland and others[2,3] employed various methods to reconstruct the specimen exit plane wave. Such methods involve inclusion of corrections of lens aberrations and improvements in the signal-to-noise ratio of the recorded image.[3]

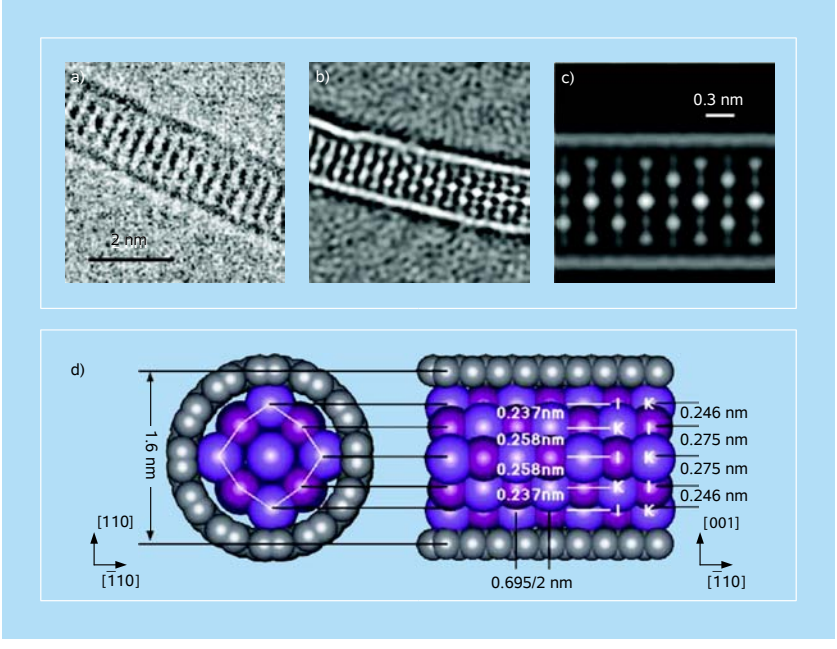


Figure 4.1 High-resolution imaging of KI. (a) Conventional HRTEM image (at Scherzer defocus) of a KI nanocrystal formed within a 1.6 nm diameter SWNT in a [110] projection. (b) Reconstructed phase of the same crystal as in (a). Note that the contrast in the reconstructed phase is reversed. (c) Enlargement of region from (b) after averaging along the tube axis. (d) Schematical model derived from (b).[1]

4.2.2 Nanocrystalline catalyst particles of platinum

It has long been suspected that monatomic steps and kinks at the surfaces of (minute nanocrystals of) Pt and certain other metals are loci for catalytic processes, such as those employed in auto-exhaust systems (where both CO and unburnt hydrocarbons are converted to CO_2 in air).[4] Using aberration-corrected imaging and the procedure of exit-wave reconstruction,[2,3,5] Gontard *et al.*[6] revealed detailed surface structures (including steps, vacancies and partially complete monolayers) on the surface of a typical (*ca.* 60 Å diameter) platinum catalyst supported on carbon (Fig. 4.2).

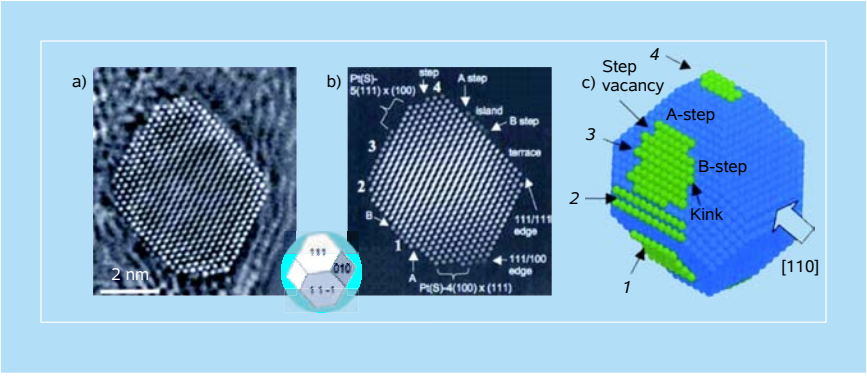


Figure 4.2 Exit-wave restoration methods in electron microscopy (worked out by Kirkland *et al.*)[1–3,5] are able to image atomic locations and distributions at the exterior surface of nanocrystalline particles of platinum.[6]

Aberration-corrected electron microscopy has also uncovered an unexpected compositional feature in Pd–Au bimetallic nanoparticles, where a shell of three layers of Pd surrounds another shell of three layers of Au, which in turn envelopes a four-layer core of Pd. This seemingly bizarre compositional arrangement[7,8] is quite unexpected on the basis of the bulk properties of the two metals. The surface energy of Au is significantly smaller than that of Pd, so one would expect Au to be at the outermost shell. For completeness, an HRTEM image of a gold nanoparticle is shown in Fig. 4.3.

Figure 4.3 Columns of Au atoms, visible as white spots, in this nanoparticle imaged by José-Yacamán *et al.* are separated from one another by *ca.* 2 Å. Adapted from Ref. 11a.

4.2.3 Microporous catalysts and molecular sieves

The structural elucidation of these materials has been transformed by real-space crystallography carried out by HRTEM.[9–11] As described in Chap. 3, the structure of ZSM-5, which is isostructural with silicalite, was first determined by taking high-resolution, real-space images down mutually perpendicular zone axes (see Fig. 3.7). Several other open-structure solids have had their 3D atomic structure solved via HRTEM, often supplemented by the use of atomistic calculations such as in the case of the solid (Brønsted) acid catalyst CoALPO-36[12] (Fig. 4.4), in which a small fraction of the Al^{III} ions in the structure has been isomorphously replaced by Co^{II}. Electron and powder X-ray diffraction data also played a key role, in addition to the real-space projected images, in the solution of this structure.[12,13] Other noteworthy achievements of HRTEM registered in this field include the elucidation of many intergrowth structures in which one zeolitic structure type is recurrently arranged in coherent fashion with another different one, but where they share a common interface. An example of this is shown in Fig. 4.5, where the industrial catalyst known as Exxon Central Research (ECR-1) is seen to be a regular intergrowth structure of the naturally occurring minerals mazzite and mordenite.

Both regular intergrowths of two distinct structures, as in ECR-1 and ZSM-23, where the zeolitic structure known as theta-1 is recurrently intergrown with its twin structure, and the numerous members of the vast family of so-called ABC-6 structures are now known thanks to real-space imaging to be composed of unit-cell scale recurrent intergrowths.[14] The "local" crystallography of zeolitic structures as

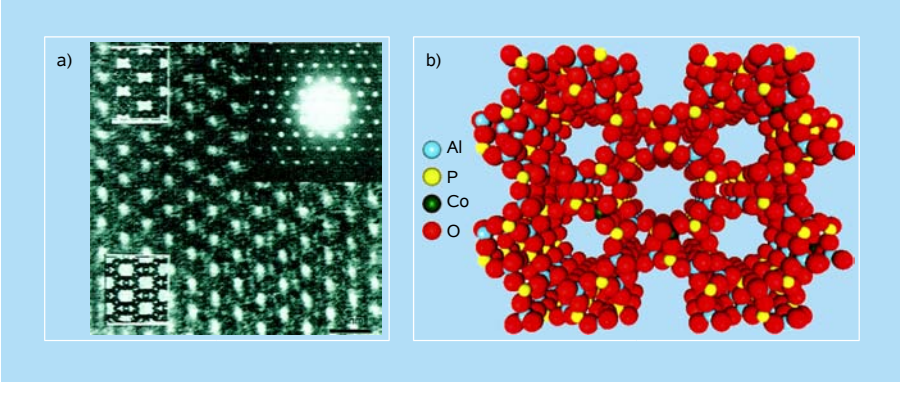


Figure 4.4 Structure determination of CoAlPO-36. (a) HRTEM showing well-defined cage apertures, and with the computed image inset, revealing the outline of 12-membered rings (top and bottom are with and without taking beam damage into account). (b) Combining data from (a) with other data, including atomistic calculations, an atomic-scale structure of the ALPO-36 framework with random distribution of Co^{II} ions is obtained.[10]

well as the chemical composition of supported nanoclusters are readily amenable to real-space HRTEM imaging, as shown in Fig. 4.6. The change in stacking order close to a structural defect is seen in the large-pore zeolite beta (see Ref. 11b and references therein) imaged in Fig. 4.6 and within nephrite jade;[15] and the 5:1 ratio of Ru:Pt in a zeptogram (10^{-21} g) cluster of $Ru_{10}Pt_2$ supported on mesoporous silica is seen in Fig. 4.14 (see also Sec. 3.1).

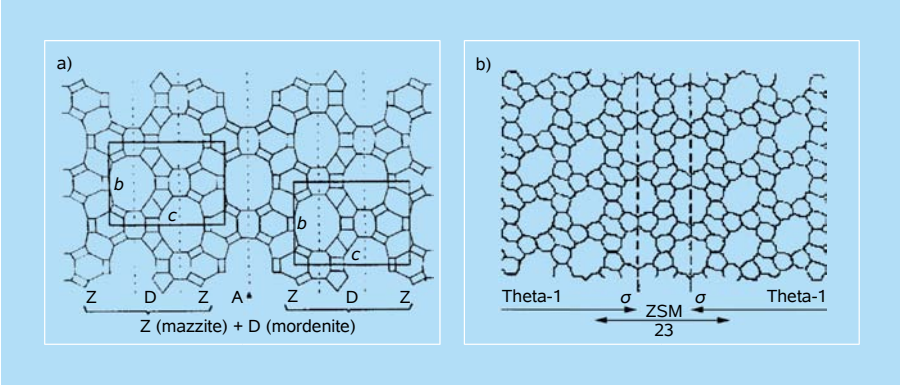


Figure 4.5 High-resolution imaging of zeolites. (a) HRTEM reveals the zeolite ECR-1 to be a sub-unit-cell intergrowth (recurrent) of mordenite and mazzite layers. (b) Schematic of ZSM-23 structure, revealed by HRTEM to be a recurrently twinned version of the zeolite theta-1 (TON).[10] See also Ref. 67.

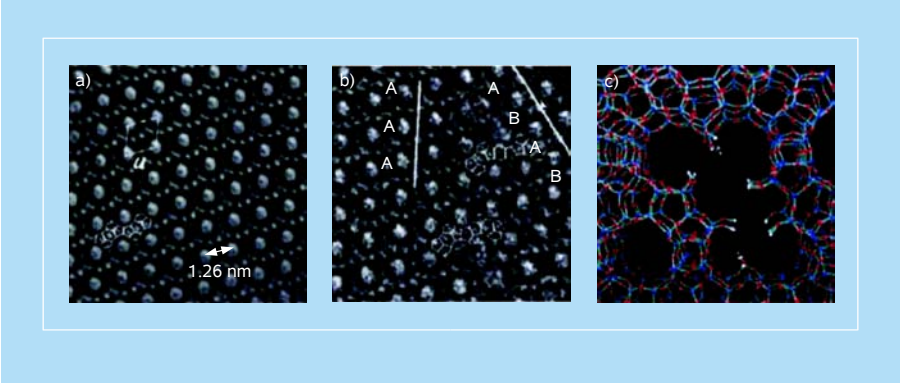


Figure 4.6 HRTEM can readily identify local structure and the coexistence of distinct crystallographic phases as in this case, with the microporous catalyst known as zeolite beta.[11] Similarly, HRTEM shows the presence of triple, quadruple, and hextuple chains of kinked SiO_4 tetrahedra occurring as defects within nephrite jade, an amphibole made up of double chains.[15]

4.2.4 Other zeolite structures

Despite substantial advances in crystal structure determination methodology for polycrystalline materials, some solids have proved particularly recalcitrant. One of them is the zeolite IM-5, first identified in 1998, which is a good catalyst for the cracking of hydrocarbons. It has an unusually large unit cell (*C*-centered orthorhombic with $a = 14.229$ Å, $b = 57.413$ Å and $c = 20.143$ Å) with a volume almost treble that of ZSM-5. Baerlocher *et al.*[16] used the recently introduced, very successful "charge-flipping" structure-solution algorithm (devised by Oszlányi and Sütő[17] for single crystals) adapted to accommodate powder diffractogram data,[18] and by combining this approach with that of HRTEM imaging they solved the structure. Very recently the structure of a novel, large-pore chiral zeolite, known as ITQ-37, has been solved by an extension of this method.[18b,18c] However, Sun *et al.*[19] have also solved the remarkably complicated 3D structure of IM-5 — it has 24 unique Si and 47 unique O atoms and 864 atoms in the unit cell — from electron crystallography alone, taken along three different projections (see Fig. 4.7). Compare Fig. 3.7 pertaining to ZSM-5.

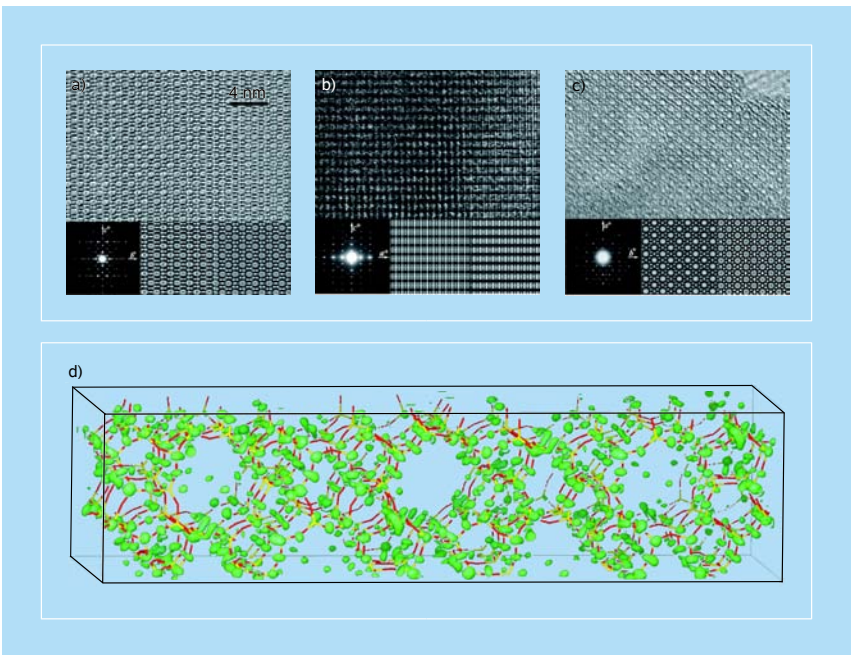


Figure 4.7 HRTEM images taken along different zone axes of the zeolite IM-5: (a) [100], (b) [010], and (c) [001]. Shown as insets are corresponding electron diffraction patterns (left) and the contrast-transfer-function-correlated and symmetry-averaged image (middle), and the computer simulation (multi-slice) from the structural model (right). (d) The final 3D density map (in green) with a stick model of the final structure of IM-5 (Si atoms, yellow; O atoms, red) superimposed for comparison. After Baerlocher *et al.*[16]

4.2.5 Structures of complex catalytic oxides solved by HRSTEM

A major breakthrough was made recently by Pyrz *et al.*[20] using an aberration-corrected STEM in that they solved the basic structural elements of the selective oxidation catalyst composed of Mo, V, Nb, Te and oxygen. Their manner of doing so will be applicable to the solution of other complex inorganic solids that play a central role in modern technology from phosphors through ceramic membranes to (solid) sources of hydrogen and oxygen, and in heterogeneous catalysis. Most of these solids are complex oxides consisting of four or more different cationic species and are extremely difficult to prepare as single crystals amenable to structural determination by X-ray crystallography.

To appreciate the significance of what Pyrz *et al.* have achieved, it is relevant to recall that only after the arrival of Rietveld's[21] method of refinement of X-ray powder diffractograms did it become possible to determine the atomic positions within these complex oxides, a procedure which gained in reliability with the availability of variable wavelength, highly monochromatized synchrotron X-ray sources.[22] When the results of such Rietveld-refined, high-resolution powder diffractograms came to fruition[23] it was generally agreed amongst crystallographers that powder X-ray structural analysis rivaled that of the traditional X-ray single-crystal approach for mixed oxides as well as quite complicated organic materials (such as oligopeptides).[24] Indeed, when allied to parallel results from X-ray absorption spectroscopy (also using synchrotron sources), great steps forward were taken in resolving inorganic cations and anionic oxygen species in complex titanosilicates typified by the synthetic catalysts JDFL-1, where Ti^{4+} ions were identified[22] in their rare five-fold (square pyramidal) coordination. Synchrotron radiation proved invaluable in the determination of the structure of JDFL-1.

When, however, the fiendishly complicated selective oxidation catalysts within the systems Mo–V–Nb–Sb–O and Mo–V–Nb–Te–O were investigated by powder diffractometry, sensible progress was achieved only after a laborious two-year effort that combined electron, X-ray and neutron diffractometry to effect the Rietveld refinement.[23,25] The Mo–V–Nb–Te–O system is a particularly important one for it catalyzes the conversion of propane (in oxygen and ammonia) directly to acrylonitrile, a major advance in that it enables paraffin-based reactants to replace the more expensive, currently used olefin-based ones for the production of acrylate polymers. This catalyst, it transpires, is itself made up of two phases, designated M1 and M2.[25a,26]

De Santo *et al.*'s[25a] work on the M1 phase, which has the composition $[TeO]_{[6.94]}(Mo_{7.8}V_{1.2}Nb)O_{28}$, was based on a simultaneous Rietveld refinement of high-resolution X-ray (XRD) and neutron powder diffraction (NPD) data.

"Conventional" electron microscopy and selected area electron diffraction (SAED) were used to derive symmetry, space group and unit cell parameters. The [TeO] entities were found to be accommodated in the channels formed by rings of corner-shared transition-metal octahedra. (The extent of the [TeO] entities accommodated in the channels is sample dependent, and it is still rather unclear whether they are restricted to the hexagonal channels or are also present in the heptagonal ones.)[26,27]

Figure 4.8 Schematic representation of the structure (projected along [001]) of the M1 complex oxide phase of the oxidation catalyst composed of Mo–V–Nb–Te–O (red: Mo^{6+}/V^{5+}; green: Mo^{5+}/V^{4+}; purple: Mo^{6+}; orange: Nb^{5+}; teal: Mo^{5+}; pink: Mo^{6+}/Mo^{5+}; yellow: Te^{4+}).[20]

As shown in Fig. 4.8, the hexagonal and heptagonal channels are formed by rings of corner-shared, transition-metal octahedra. Structural refinement of the M1 phase is exacerbated by the fact that the complex oxides of which it is composed require models that have close to 200 adjustable parameters, so that extremely high quality diffraction data as well as adroitly chosen starting models (with imposed constraints) are required to ensure a stable, converging refinement. Detailed bright-field (BF) transmission electron microscopy (TEM) in the hands of Buttrey[25a] and Ueda[26] and their co-workers yielded plausible models of the projected structure of the M1 phase and, in particular, led to the picture, shown in Fig. 4.8, of

an open structure in which hexagonal and heptagonal channels run along the [001] direction. In addition, the BF TEM images, coupled with SAED data, provided firm evidence for the location of the Nb-centered pentagonal biprisms (e.g., Nb9 in Fig. 4.8). The location of Nb cannot be directly determined in any straightforward way with TEM, XRD, or NPD. Studies of Ta substitution provided support for the contention that the Nb would preferentially occupy the pentagonal biprismatic center site.

The profound significance of the papers by Pyrz, Vogt, Buttrey *et al.*[20] is that, using an aberration-corrected HAADF-STEM, they have independently derived a model for the M1 structure of the oxidation catalyst. Moreover, this derived model agrees very well with that obtained by the more laborious method of multiple Rietveld refinement; see Figs. 4.9 and 4.10. As mentioned above, the approach of Pyrz *et al.*[20] should, with relative ease, be applicable to the atomic resolution of many other complicated structures where the (heavy) rare-earth elements figure eminently as constituents thereby taking advantage of *Z*-contrast imaging. With adequate precautions — such as ensuring by ion-beam thinning and testing image contrast using the multislice method of Moodie and Cowley[a] beforehand — the precise location of the heavy (high *Z*) elements in projected images down to appropriate zone axes should be readily determined. The use of Kirkland's exit-wave restoration methods[5] should also help in further elucidating the structures of such complicated oxidic materials, especially with regard to the location of lighter elements.

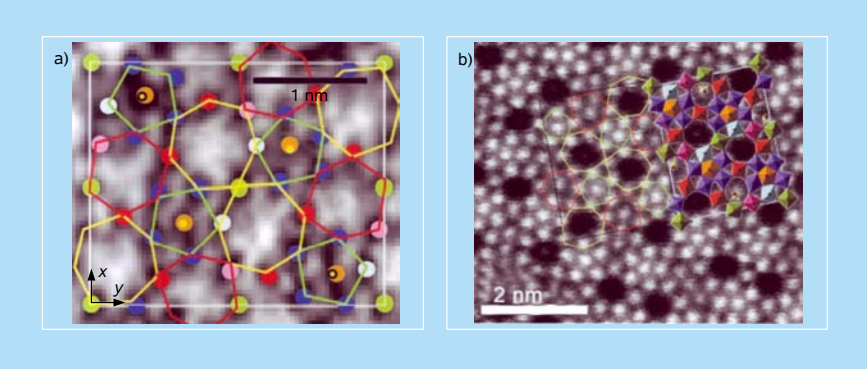


Figure 4.9 Comparison of the projected electron microscope image of a thin section of the M1 phase, projected along [001] taken with (a) a conventional HRTEM and (b) an aberration-corrected STEM.[20] Courtesy of W. D. Pyrz and D. J. Buttrey.

[a]For a concise account of the Moodie and Cowley "multislice" method, see Ref. 28.

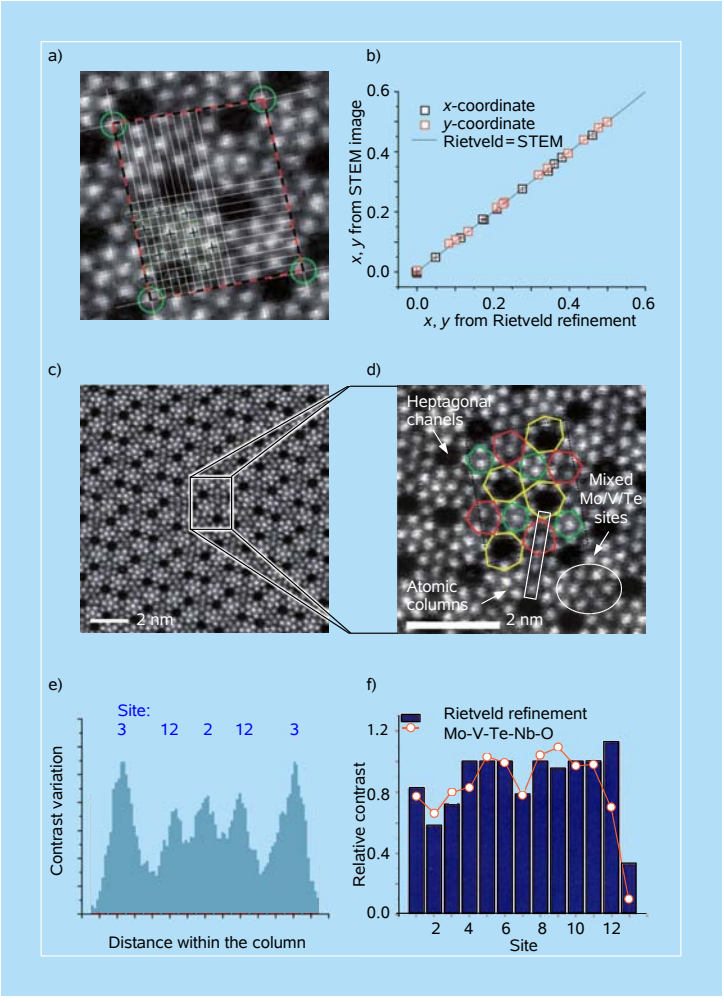


Figure 4.10 HRSTEM imaging of the catalyst of Fig. 4.8. (a, b) Direct analysis of atomic coordinates. C_s-corrected HAADF micrographs in the center (c, d) clearly reveal the pentagonal biprisms linked to octahedra, as well as the heptagonal and hexagonal channels. Because of the nature of the *Z*-contrast of HAADF-STEM imaging (see Sec. 3.3), the spatial coordinates of the heavy atoms may be directly read off from the real-space images (to within 0.1–0.2 Å) and their site occupancies to within 15 percent. Alternatively, the site occupancies can be obtained from the atomic column contrast variations (e, f).[20]

4.2.6 The value of electron diffraction in solving 3D structures

The role that SAED patterns play in facilitating the solution of 3D structures of complex solids must never be underestimated. This fact is eloquently illustrated in the work of Buttrey *et al.*[29] dealing with the 16 distinct crystallographic phases that have been discovered (so far) in the Bi_2O_3–MoO_3 mixed oxide system.

Arranged from Mo-rich to Bi-rich compositions, the sequence of distinct phases from α–$Bi_2Mo_3O_{12}$ to δ(H)–$Bi_{26}Mo_{10}O_{69}$ and on to H–$Bi_{14}MoO_{24}$ is impressive. Many of these phases are good selective oxidation catalysts and their structures are capable of releasing oxygen from the solid to the hydrocarbon in the presence of gaseous oxygen.

By judicious use of SAEDs it was discovered early on[30] that very many of these seemingly distinct phases are, in fact, superlattice variants of the structure of fluorite (CaF_2) in a defective state. To be sure, they are derived from a defective $Bi_2O_3\Delta$ structure, which may be thought of as an analogue of fluorite with an oxygen vacancy, symbolized by Δ. Structural views of some of these fluorite-related bismuth molybdates are shown in Fig. 4.11.

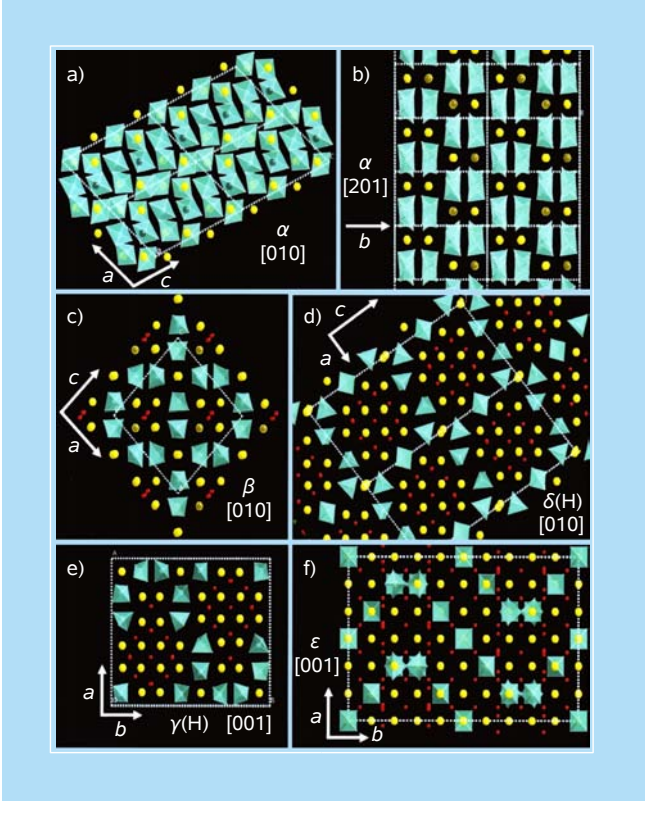


Figure 4.11 Structural views of some of the fluorite-related bismuth molybdates: (a) α-$Bi_2Mo_3O_{12}$ [010] "scheelite projection," (b) α–$Bi_2Mo_3O_{12}$ [201] principle fluorite projection, (c) β–$Bi_2Mo_2O_9$ [010] projection, (d) δ(H)–$Bi_{26}Mo_{10}O_{69}$ [010], (e) γ(H)–Bi_2MoO_6 [001] "Latin cross" projection, and (f) ε–$Bi_{38}Mo_7O_{78}$ proposed cation arrangement in the [001] projection. Note that in each case, the fluorite subcell axes are oriented along the chosen true-cell zone axis as well as vertically and horizontally in the projections so as to produce the approximately square pattern of metal sites expected in projection from the fluorite substructures. Bi (yellow), Mo polyhedra (blue), O outside polyhedra (red), metal vacancies (gray).[29] Courtesy of D. J. Buttrey.

With the aid of SAED patterns, Jefferson *et al.*[31] were able to solve by HRTEM imaging down various zone axes the structure of the selective oxidation catalyst $Bi_2(Mo,W)_{12}O_{35}$. It transpires that many solid-solution systems derived from Bi_2O_3 and a heavy metal oxide also display a wide diversity of fluorite-based superlattices, the $Bi_2O_3 \cdot Nb_2O_5$ one being a very rich example.[32] Again, SAED patterns facilitate the derivation of the 3D structures of such materials.

4.3 Electron Tomography

Here we consider first a particularly striking example of how cryo-electron tomography (cryo-ET) sheds light on magnetotactic bacteria, followed by an example of ET involving HAADF-STEM imaging of nanocluster heterogenous catalysts.

Whereas the images (BF, tomographic and holographic) that we showed in Fig. 3.23 referred to double chains of magnetite nanocrystals within the bacteria, here we consider the single chain variety, and the so-called MamK gene thought to be responsible for its proper organization, and studied by Komeili *et al.*[33] and Scheffel *et al.*[34] These magnetosome chains contain 15 to 20 magnetite crystals of *ca.* 500 Å size which, because of their superparamagnetism, orient the bacteria such that they swim parallel to Earth's magnetic field, so as to locate their food.

The 3D reconstruction of the individual 2D projections of each cell was in this instance achieved by the use of dual-axis tilt cryo-ET. This showed that each magnetite crystal was surrounded by a layer that was an invagination of the cell membrane, along with the presence of some filaments parallel to the chain.

Figure 4.12 shows the cryo-ET after the identification of the gene MamK. The 3D reconstruction of a wild cell (Fig. 4.12a) can be compared with visualization of a MamK-deficient mutant cell (Fig. 4.12b), in which instead of long, aligned chains of magnetitic nanocrystals, only small clusters of them are found. The alignment of the magnetitic crystals and the formation of the filaments can be reversed by means of the mutant cells being complemented with MamK-GFP [green fluorescent protein (GFP) fused to the C-terminus of MamK] as shown in Fig. 4.12c.

Turning to a recent example, insofar as exploring the location, distribution and shape of nanocluster bimetallic catalysts supported inside nanoporous silica is concerned, the HAADF-STEM imaging procedure is the preferred method (also elaborated in Sec. 3.3.1). The predominantly coherent nature of low-angle scattered electrons means that conventional BF and DF images are prone to contrast reversals from changes of specimen thickness, orientation or defocus. But high angle scattering is predominantly incoherent, and STEM images formed using a scanned focused probe yield in addition the *Z*-contrast that arises from the ensuing Rutherford scattering. This sensitivity to composition, together with the incoherence of the scattering process, renders small heavy particles highly visible

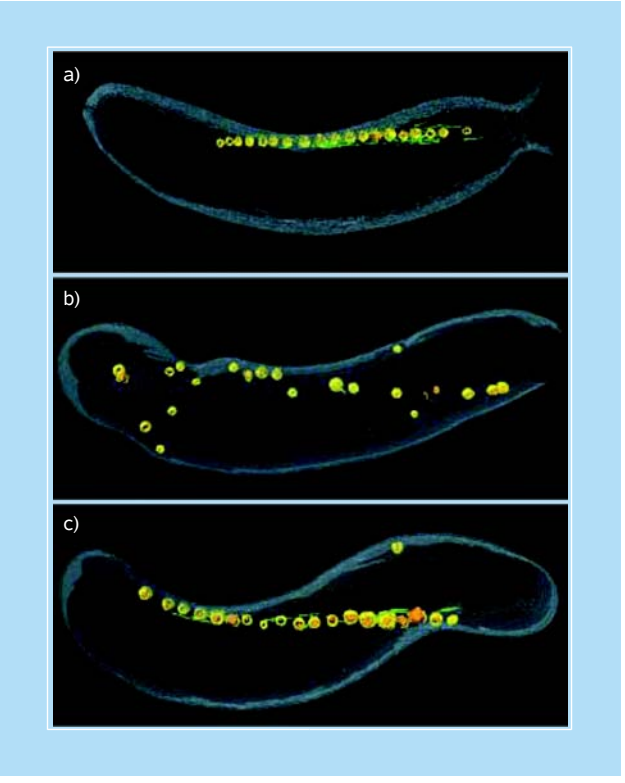


Figure 4.12 The MamK gene is required for the proper organization of the magnetosome chain in magnetotactic bacteria, as illustrated by this tomographic study. (a) 3D reconstruction of a wild-type AMB-1 cell. The cell membrane (gray), magnetosome membrane (yellow), magnetite (orange), and magnetosome-associated filaments (green) are rendered. (b) ΔmamK mutant, where magnetosomes appear disordered and no filaments are found in their vicinity. (c) ΔmamK cell expressing mamK-GFP on a plasmid showing full reversal of the mutant phenotype. See Refs. 33 and 34 (adapted from Ref. 33).

within a matrix or support. Figure 4.13 illustrates how readily the nanoclusters of $Ru_{10}Pt_2$ are imaged within the 3D nanoporous structure of the siliceous support (see also Fig. 3.14). The simultaneous retrieval of the elemental composition (from electron-induced X-ray emission spectra) of the individual catalytic nanoparticles, each of which weighs around a zeptogram (10^{-21} g), is also achieved (Fig. 4.14).

4.4 Electron Holography

The principles presented in Chap. 3 are readily applied to examine the nature of the magnetic lines of force emanating from incarcerated ferromagnetic nanomaterials. Figure 4.15c, with its sharply defined contours, shows the lines of force generated by a minute single crystal of iron (*ca.* 360 Å in diameter) shrouded by a multiwalled carbon nanotube approximately 1800 Å in diameter. The contours,

which were generated from the phase image, as described in Sec. 3.6, Fig. 3.23e, are overlaid onto a BF TEM image of the crystal. For comparison, an original picture of Michael Faraday's illustration of lines of magnetic force, first conceived by him in the 1840s, is also shown in Fig. 4.15a.

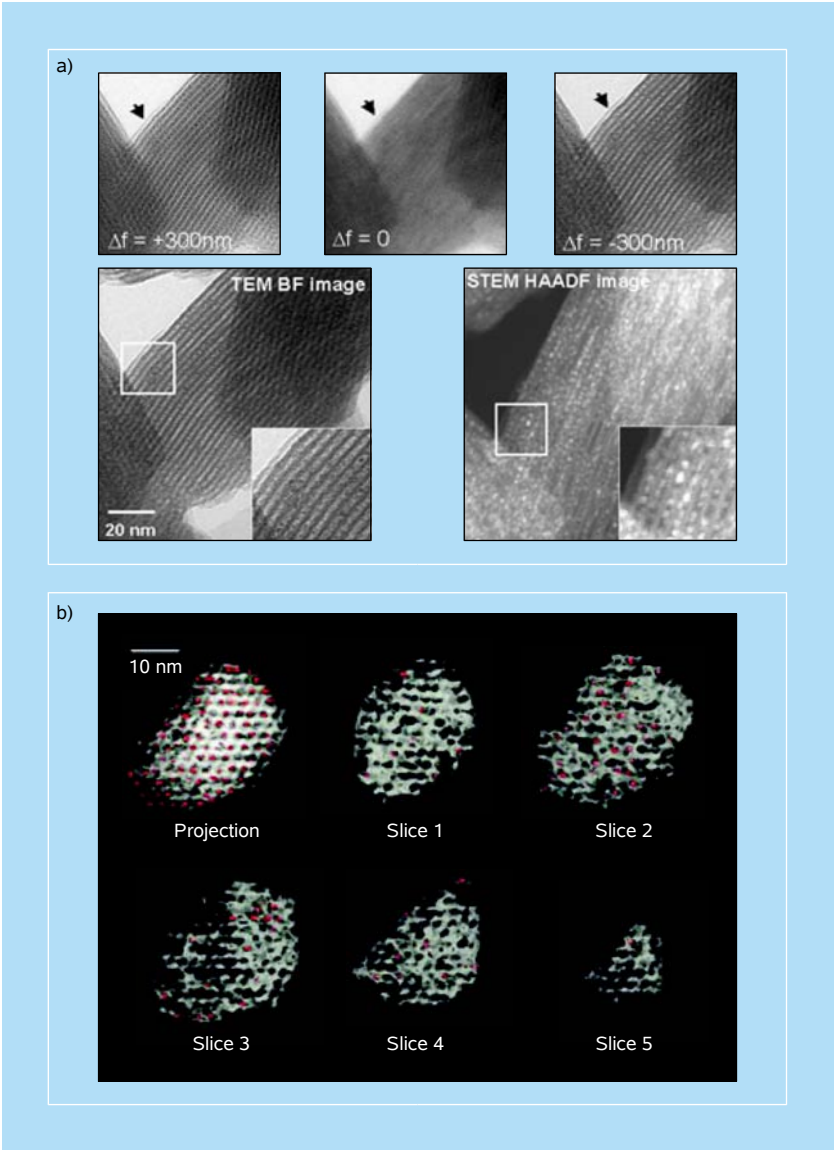


Figure 4.13 Tomographic slices and STEM of bimetallic clusters. (a) Bimetallic clusters of $Ru_{10}Pt_2$ (hydrogenation catalyst) supported on mesoporous silica are visible in dark-field STEM tomographs, but not in BF TEM. (b) An axial projection of a 30 nm thick specimen and five successive 3 nm thick slices through a tomogram of silica-supported $Ru_{10}Pt_2$ nanoclusters (red) on microporous silica (white).[68]

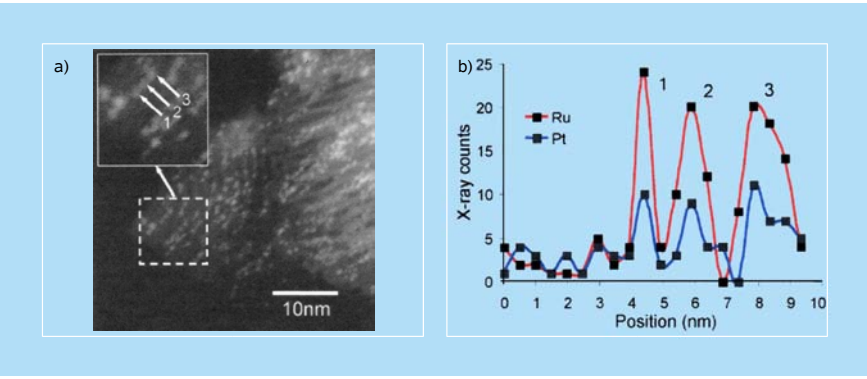


Figure 4.14 Metallic composition of nanoparticle catalysts. (a) HAADF-STEM image of the Ru–Pt-based catalyst. The inset shows the particles studied by X-ray microanalysis. X-ray spectra were recorded approximately every 0.5 nm and the resulting compositional profile across the three particles is shown in (b). From the area under the curve in the region of the three nanoparticles, the ratio of Ru:Pt for each particle is determined as 5:1, in line with the composition expected from stripping the carbonyl groups from the precursor $[Ru_{10}Pt_2(CO)_{28}]$ dianions.[69]

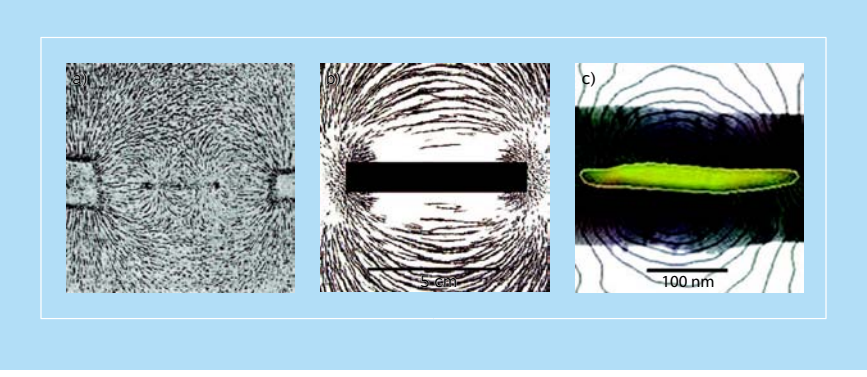


Figure 4.15 Faraday's magnetic lines of force and electron holography. (a) Faraday's image of "magnetic lines of force," formed using magnets and iron filings. This is a copy of the actual illustration made by Michael Faraday, who "fixed" the lines of force (made of iron filings) by using adhesive paper. (b) A contemporary image similar to that shown in panel (a) formed using a single bar magnet and iron filings. (c) Magnetic phase contours recorded using off-axis electron holography from a multiwalled carbon nanotube, approximately 180 nm in diameter, containing a 36 nm-diameter encapsulated iron crystal (in yellow). The contours, which were generated from the phase image, were overlaid onto a BF TEM image of the crystal.[36]

Many other instances have been reported where minute nanoferromagnetic solids are encapsulated in carbon nanotubes, the work of Jourdain *et al.*[35] in this field being of interest as it shows how, *inter alia*, Fe–Co–P nanocrystals are periodically laid down along the axis of a growing multiwalled carbon nanotube (Fig. 4.16). Electron holography has been especially useful in the study of

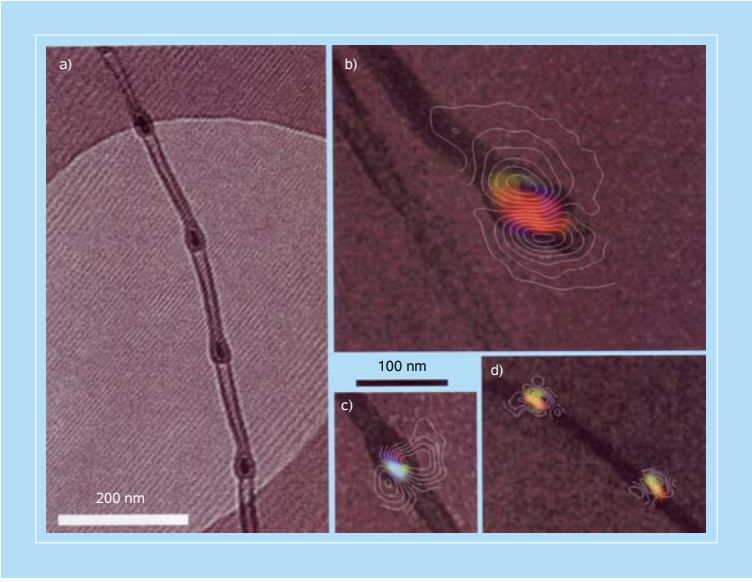


Figure 4.16 Electron holography of carbon nanotubes. (a) Off-axis electron hologram recorded from a multiwalled carbon nanotube containing four periodically spaced Fe–Co–P crystals. (b–d) Magnetic induction maps recorded using electron holography from crystals similar to those shown in panel (a) and overlaid onto amplitude images: (b) 52 nm × 98 nm particle, (c) 40 nm × 16 nm particle, and (d) two particles in a nanotube.[35]

intergrowths of magnetite and another mineral known as ulvöspinel (Fe_2TiO_4), which occurs in Sweden. Although the magnetite (Fe_3O_4)–ulvöspinel system forms a complete solid solution[36] at temperatures above *ca.* 450°C, intermediate bulk compositions can exsolve (i.e., un-mix) during slow cooling to yield an intergrowth of single domains or pseudo-single-domain magnetite-rich blocks separated by nonmagnetic ulvöspinel-rich lamellae.

Harrison *et al.*[37] used off-axis electron holography to elucidate the magnetic microstructure of a natural finely exsolved intergrowth of submicrometer magnetite blocks in an ulvöspinel matrix. Some of their key observations are reproduced in Fig. 4.17. Figure 4.17a shows a chemical map (derived from energy-filtered imaging) of a representative area of the sample. Ulvöspinel-rich exsolution lamellae (red) subdivide the original titaniferous grain into an array of magnetite-rich blocks (blue). The profiles in Fig. 4.17b, which were obtained from the line marked with a short white arrow in Fig. 4.17a, show that little Ti is present in the blocks as they are essentially pure magnetite. Harrison *et al.*[37] showed that the individual blocks of magnetite contain primarily single-domain states. They also revealed the magnetostatic interaction fields between them, as illustrated in Fig. 4.17c. These images reveal the complexity of the magnetic structure of this system: it is dominated by the shapes of the blocks and by magnetostatic interactions. Magnetic

superstates, in which clusters of magnetite blocks act collectively to form vortex and multidomain states that have zero net magnetization, were also observed.

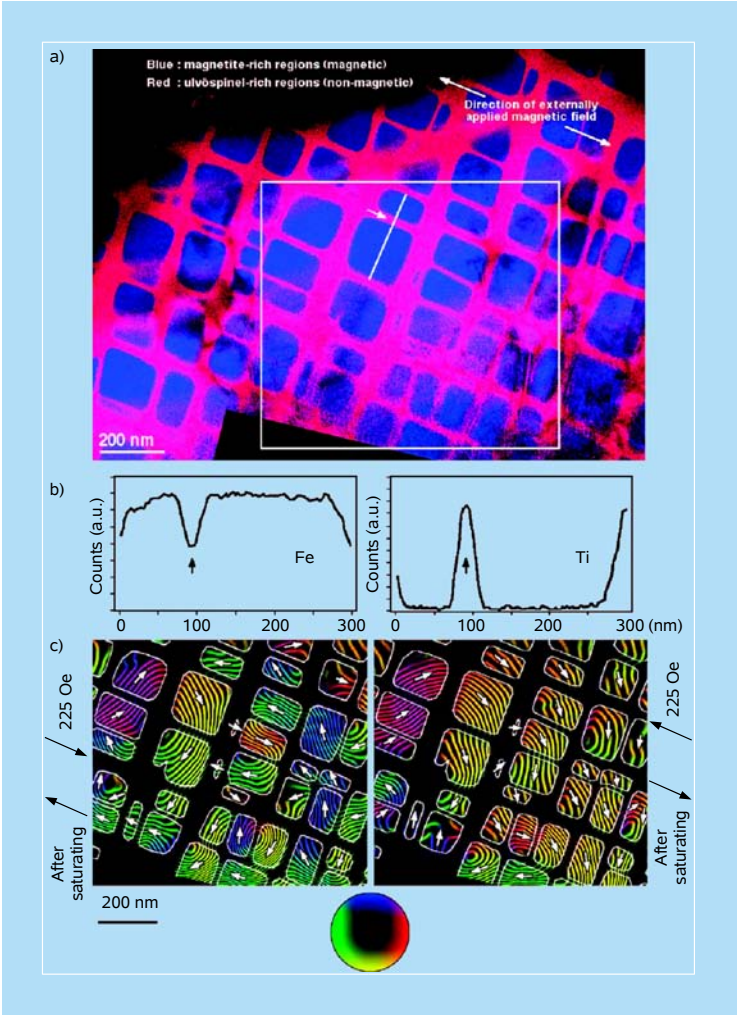


Figure 4.17 Chemical maps and magnetic microstructures. (a) Chemical map of an exsolved titanomagnetite (Fe_3O_4–Fe_2TiO_4) sample obtained using elemental mapping at the Fe $L_{2,3}$ and Ti $L_{2,3}$ edges in a Gatan imaging filter (GIF). The GIF separates electrons that have lost energy in the sample due to inelastic scattering from elastically scattered electrons and refocuses them to form an image of the sample. The blue regions are magnetic and are rich in magnetite (Fe_3O_4), while the red regions are nonmagnetic and rich in ulvöspinel (Fe_2TiO_4). The box corresponds to the region examined in detail using off-axis electron holography. Panel (b) shows line profiles obtained from the Fe and Ti chemical maps, along the solid line marked in panel (a). The short arrows mark the same point in the three pictures. (c) Magnetic microstructure of the boxed region in panel (a) measured using off-axis electron holography. The two images correspond to different magnetic remanent states, acquired with the sample in magnetic-field-free conditions.[36,37]

It has long been known that the microstructures of minerals are petrogenetic indicators. The remanent magnetization is used by geophysicists to map the motions of continents and ocean beds resulting from the dynamics of plate tectonics. Electron holography has considerable potential in the field of measuring both remanent magnetizations and magnetization reversal mechanisms in rocks, and for understanding mineral magnetism at the nanometer scale in general.

4.5 Electron Crystallography

Earlier (see Sec. 3.2) we discussed how the electron microscope is nowadays used to solve the 3D structures of inorganic materials, especially those that are microporous and mesoporous. Some much-used, ordered mesoporous siliceous solids are designated MCM-48, SBA-6, SBA-16, and AMS-*n*, the 3D structures of which have all been solved from the HR image in an electron microscope by Terasaki *et al.*[38–41] Figure 4.18 gives both the raw EM data and the solved structure of MCM-48.

The structure of the very complex intermetallic compound *ν*-AlCrFe with $a = 40.687$ Å, $c = 12.546$ Å, space group $P6_3/m$, was solved[42] (see also Ref. 18 and references therein) by a combination of HRTEM images and SAED patterns from

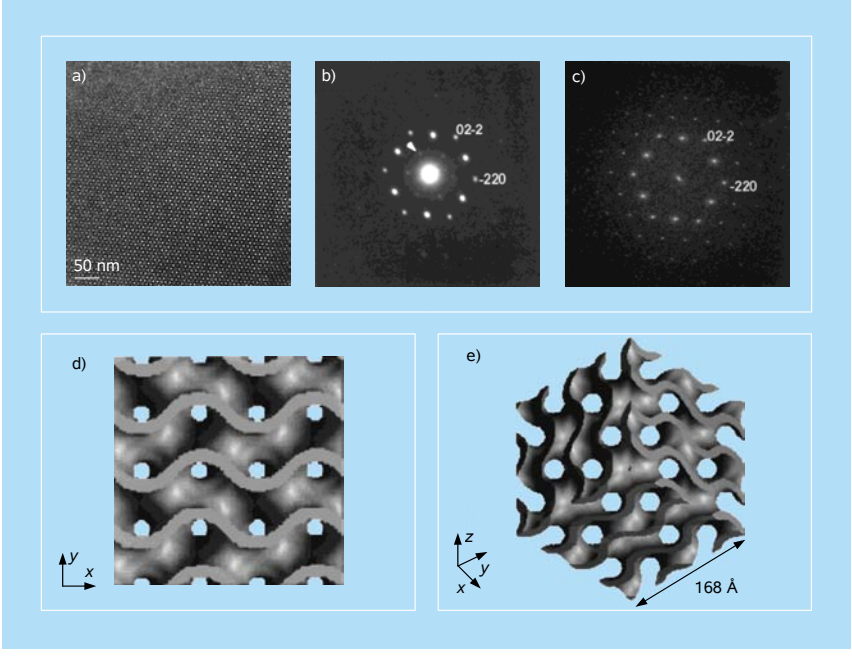


Figure 4.18 Images and diffraction of MCM-48. (a) HRTEM of MCM-48 along [111] with (b) electron diffraction and (c) Fourier diffractogram of (a). (d,e) Structure of MCM-48, (d) along [100]; (e) is a 3D representation along [111].[10]

13 zone axes. It turns out that 124 of the 129 unique atoms — there are 1,176 in the unit cell — were found in the clear 3D potential maps obtained by electron crystallography (Fig. 4.19). This structure, it seems, is the most complicated one ever solved to atomic resolution by electron crystallography to date.[19] It is a reminder that ostensibly there seems to be no limit to determining complex structures of inorganic materials by electron crystallography.

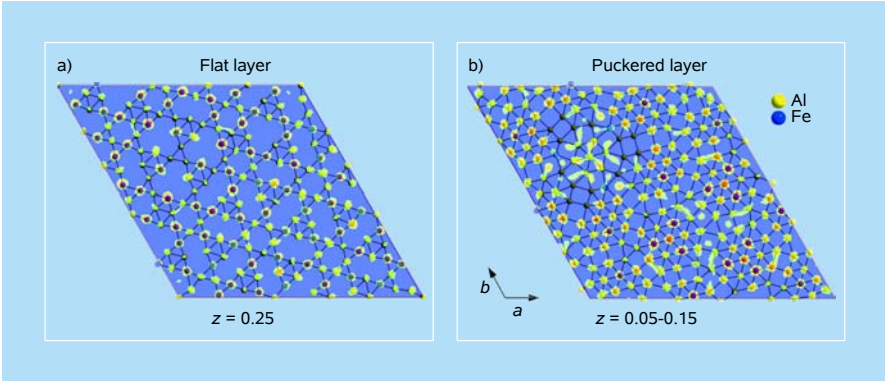


Figure 4.19 Two sections of the 3D potential map of *v*-AlCrFe obtained by electron crystallography with the structure superimposed. All atoms are clearly resolved and it is possible to distinguish transition metals (Fe, Cr, strong peaks) from Al (weak peaks). Adapted from Zou *et al*.[42a]

4.5.1 Other complex inorganic structures

Although electron crystallography as outlined above and in Sec. 3.2 — see, in particular, Fig. 3.8, with its use of SAED patterns — is often effective, it is not universally so. This is because electrons interact very strongly with matter — much more so than X-rays (see Table 4.1). This strong interaction results in complicated multiple scattering (i.e., dynamic) effects, which cause the reflection intensities to deviate severely from the so-called kinematic assumption that I_{hkl} is proportional to $|F_{hkl}|^2$ (see Ref. 28). This dynamical scattering is the main reason why electron diffraction (ED) data are not generally used to solve structures. Nevertheless, recording ED patterns in an electron microscope has many advantages, for example, the ease and rapidity of doing so; the size of the investigated crystal may be extremely small (and inaccessible to X-ray studies) and individual crystallographic phases can be readily targeted; and the resolution is high, typically thousandths of an Å. This is where precession electron diffraction (PED) comes into its own: it is an ingenious experimental method of recording ED intensities effectively under kinematic conditions. Introduced by Vincent and Midgley[42b] in 1994, it is now gaining rapidly in popularity (see Table 4.2).

Table 4.1 The Complementarity of X-ray Powder and Electron Diffraction[42c]

	X-ray Powder Diffraction	Electron Diffraction
Crystallite Size	μm	nm
Data Set	Complete	Incomplete
Resolution	Low	High
Data Precision	High	Low
Intensities	Kinematical: easy to interpret	Dynamical: difficult to interpret (multiple scattering)
Multiple Scattering	No	Yes
Overlap	Yes	No

Table 4.2 A Selection of Key Illustrative Examples of Structure Solution Using the Vincent–Midgley[42b] (Precession Electron Diffraction) Approach to Electron Crystallography*

1.	K. Boulahya, L. Ruiz-González, M. Parras, J. M. González-Calbet, M. S. Nickolsky and S. Nicolopoulos, "*Ab Initio Determination of Heavy Oxide Perovskite Related Structures from PED Data*"	445 *et seq.*
2.	D. L. Dorset, "*Electron Crystallography of Organic Materials*"	453 *et seq.*
3.	D. L. Dorset, C. J. Gilmore, J. L. Jorda and S. Nicolopoulos, "*Direct Electron Crystallographic Determination of Zeolite Zonal Structures*"	462 *et seq.*
4.	Z. B. He, X. D. Zou, S. Hovmöller, P. Oleynikov and K. H. Kuo, "*Structure Determination of the Hexagonal Quasicrystal Approximant μ'–(Al, Si)$_4$Cr by the Strong Reflections Approach*"	495 *et seq.*

*Taken from *Ultramicroscopy*, **107**, (2007), issues 6 and 7 ("*Proceedings of the Electron Crystallography School 2005*," Eds. S. Nicolopoulos and T. E. Weirich). The page numbers of the respective articles are indicated in the numerical entries given above.

The PED technique is equivalent to the Buerger precession method used in single-crystal X-ray diffraction, where the crystal is precessed around the incident X-ray beam. In the PED case, however, the electron beam is deflected and precessed around a stationary crystal. With this tilted illumination only a few reflections are excited simultaneously, so the possibilities for multiple scattering are greatly reduced and the resulting diffraction intensities are more kinematic in nature. Nowadays, commercial PED attachments for existing TEMs are available. As may be seen in Table 4.2, a wide range of inorganic structures which were inaccessible to the earlier electron crystallographic methods have been solved via PED.

Even though PED reflection intensities are more reliable than those used from SAED patterns (Fig. 3.8), they are still not ideal. But it has been demonstrated

that PED patterns can be used to advantage in combination with X-ray powder diffraction data. In essence, the single-crystal so-called "charge-flipping" structure solution algorithm of Oszlányi and Sütő (see Ref. 17 and references therein) is applied to the 2D data for selected zones to obtain phases for the contributing reflections. These phases are then included in the starting phase sets for the application of the powder charge-flipping (PCF) algorithm[42d] to the X-ray powder diffraction data. This approach is similar in principle to that using a combination of X-ray powder diffraction data with phases derived from PED.[42c]

4.5.2 Complex biological structures

Here, the materials to be solved structurally are almost invariably extremely vulnerable to radiation damage in the electron microscope. It is therefore necessity to use elaborate procedures of specimen preparation, preservation in vitreous ice, image recording at liquid helium temperatures, and image processing, as outlined in Chap. 3. Here, however, we illustrate how electron crystallography has led to spectacular advances that have, in particular, transformed our knowledge of the structure and mode of operation of viruses and phages. The procedural details outlined by Crowther, a leading figure in virus structural investigations, have been summarized in Fig. 3.15, where it is seen that the single-particle approach is used to solve the structure.

The power and success of the single-particle approach adopted by Crowther *et al.*[43] on the hepatitis B virus, and by Chiu *et al.*[44] on the infectious ε15 virus capsid are examples that we focus upon. Hepatitis B virus is a major human pathogen; it is estimated that there are at least 350 million carriers of the virus worldwide and that it causes in excess of one million deaths a year from liver disease.

Figure 4.20 shows a map of the recently determined 3D structure of the hepatitis B core protein, which was completed from *ca.* 6,300 particle images from 34 electron micrographs. This is a sufficiently good map for the fold at the protein, which is largely α-helical, to be deduced (Fig. 4.21). Moreover, it proves possible, using biochemical and immunological data, to propose an amino acid numbering scheme for the protein chain. Subsequent to this electron microscope 3D structure determination, the structure of the core shell was solved to higher resolution by X-ray crystallography,[45] thereby confirming and extending the results from electron microscopy. The core shells consist of 240 copies of the core protein arranged as 120 dimers. Each dimer makes a spike that protrudes from the shell and consists of two α-helical hairpins, one from each molecule in the dimer. The spike thus consists of a four-helix bundle. The shells that were analyzed in Crowther's studies were made from the core protein with the basic C-terminal region removed (Fig. 4.21).

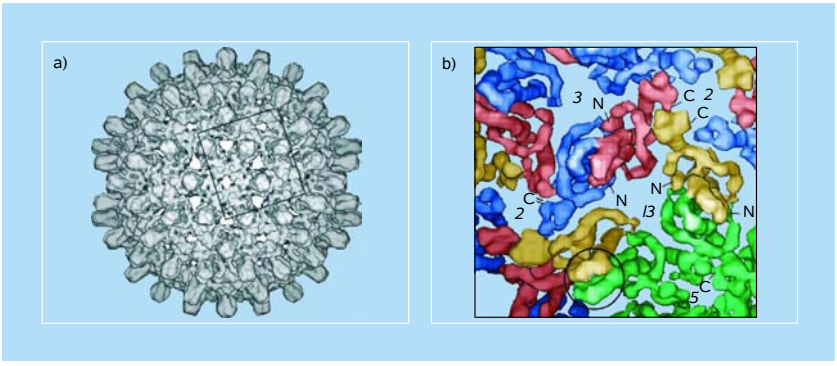


Figure 4.20 Map of the hepatitis B core protein. (a) View of the whole shell. Map computed from approximately 6,300 particle images from 34 micrographs. (b) Enlarged view of the boxed area indicated in (a). This form of the shell is made from 240 copies of the core protein, so after icosahedral averaging there are four computationally independent subunits, which are here colored blue, red, yellow and green. The red–blue and yellow–green molecules form closely associated dimers. The positions of the icosahedral twofold, threefold, fivefold and local threefold axes are indicated (2, 3, 5, *I*3), as are the positions of the N- and C-termini of the protein. The fold of the four subunits is identical, and the thicker tubular regions correspond to *α*-helices. Bundles of four helices (an example is encircled) form the spikes protruding from the surface.[70]

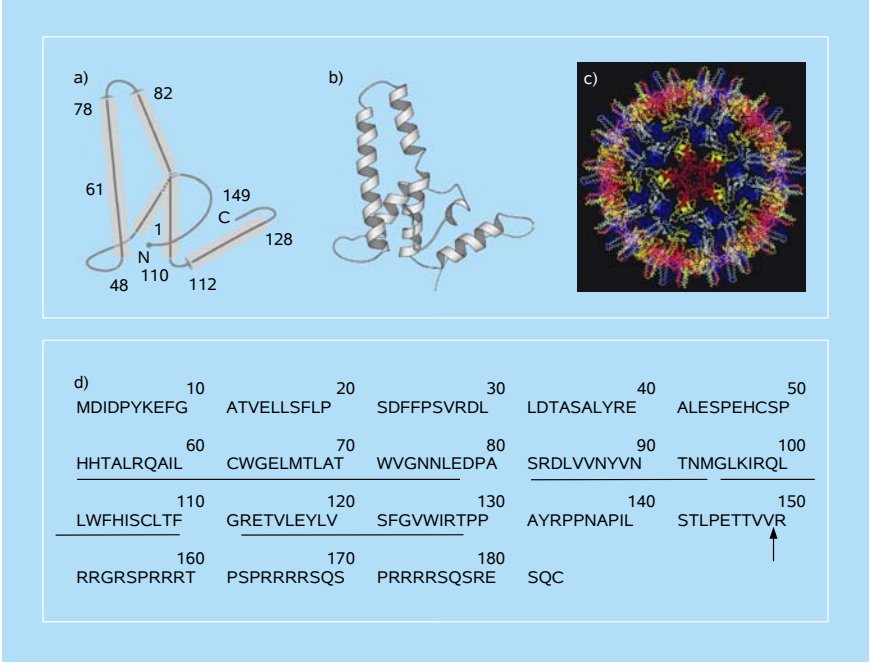


Figure 4.21 Fold of the hepatitis B core protein. (a) Fold deduced from the cryomicroscopy map. (b) Fold determined by X-ray crystallography. (c) The crystallographic structure of the core shell viewed down an icosahedral fivefold axis, with the four independent subunits colored red, yellow, blue and gray. (d) The amino acid sequence of the core protein, indicating where the protein was truncated after residue 149 to remove the very basic C-terminal region.[70]

In Chiu *et al.*'s[44] study of the ε15 virus capsid, some 20,000 individual particle images were analyzed, so that the electron density map could be reconstructed to a resolution of 4.5 Å. This resolution is the result of advances and accumulated experience in instrumentation, the collection of large data sets, and image processing, and also in the availability of large-scale distributed computing. The resulting electron density map (Fig. 4.22) clearly shows protruding densities around all the icosahedral and local two-fold positions.

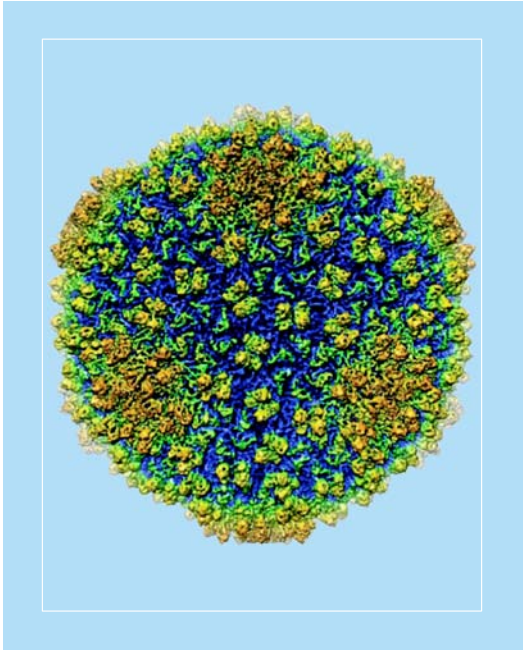


Figure 4.22 The 4.5 Å resolution 3D reconstruction of ε15 phage, about 700 Å in diameter, radially colored. After Jiang *et al.*[44]

Cryo-EM is an indispensable method for capturing the nature of the molecular components of phages, a fact well represented by Fig. 4.23 for the ε15 bacteriophage. We may summarize the kind of information that cryo-EM yields as follows:

- reveals all molecular components of a spherical virus;
- traces the backbone of a virus capsid protein;
- reveals conformational changes of capsid particles in a maturation process;
- reveals that double-stranded DNA viruses have similar but distinct structural details and are probably critical for respective host infection.

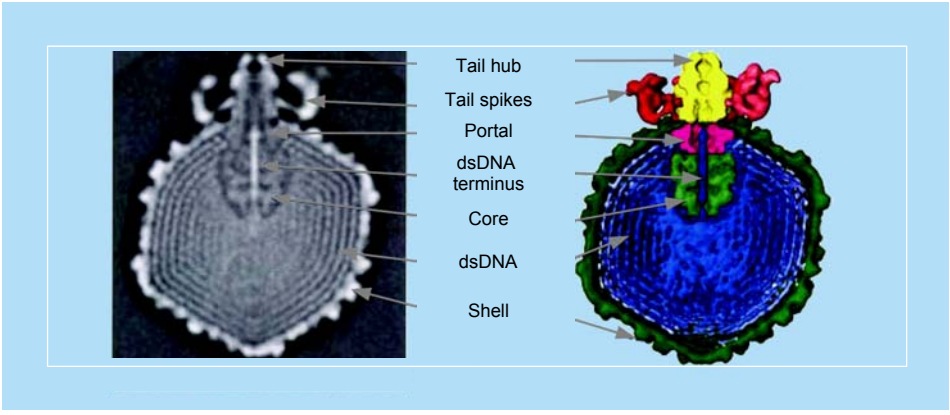


Figure 4.23 The molecular components of the ε15 phage structure determined from cryo-EM. After Jiang *et al.*[71]

As mentioned in Chap. 3, only a few 2D crystals of biological molecules have so far yielded electron microscopic structural elucidation by the classic Henderson and Unwin approach worked out for bacteriorhodopsin.[46a] It is noteworthy that, in this way, Kühlbrandt and Wang determined the 3D structure of plant-light-harvesting complex.[47] A recent UK–US collaboration,[46b] using low-dose aberration-corrected microscopy, holds considerable promise for such improved, high-resolution EM images of organic and biological specimens. This work demonstrates that, with low electron-dose illumination of the specimen at liquid N_2 temperatures and low defocus values, sub-nanometer image resolution ought to be possible.

The work of Unwin[48] on the acetylcholine receptor channel imaged in the open state, by the Klug method of electron crystallography, constitutes a notable landmark. The open-channel form of the receptor was determined from electron images of Torpedo ray postsynaptic membranes activated by brief (< 5 ms) mixing with droplets containing acetylcholine. Comparison with the closed-channel form showed that acetylcholine initiates small rotations of the subunits in the extracellular domain, which trigger a change in configuration of α-helices lining the membrane-spanning pore. All this work is of central molecular biological importance as it is known that the brain functions by sending electrical signals across synapses connecting cells of the nervous system. Neurotransmitter-gated ion channels of the acetylcholine and glutamate receptor families are the protein molecules in the postsynaptic membranes that control the rapid propagation of these signals between most cells.

It is instructive to summarize the situation concerning the quality of structural information retrievable from the various cryo-electron microscopic approaches now available.[48b]

- Where 2D crystals are studied, the resolution in relative terms is high, approaching atomic values in special cases.
- Single-particle imaging requires many identical molecules, which are imaged in random orientations in a thin film of vitreous ice. Here, the resolution is about 6 Å at least, this being limited by several factors including a range of molecular conformations and the effects of radiation damage.
- Where tomography is utilized, because of the inevitable damage inflicted by the electron beam during data collection, the resolution is limited to about 15 Å.

4.6 Electron-Energy-Loss Spectroscopy and Imaging

We recall that, by adding an electron spectrometer to a TEM, electron-energy-loss (EELS) spectra and energy-filtered (EFTEM) images can be obtained, and from those, the specimen thickness and a good deal of chemical information can be extracted. In addition to EELS as a means of deriving chemical compositions, by placing a solid-state X-ray detector within a few millimeters of the specimen and analyzing the intensities of the X-ray peaks, which are characteristic of each element, a chemical map of the specimen may be constructed (from the energy dispersive X-rays, EDX, or wavelength dispersive spectroscopy, WDS), with a spatial resolution determined largely by the diameter of the focused electron probe (i.e., *ca.* 10 Å). EELS offers superior sensitivity (compared with EDX), especially to the very light elements ($Z < 10$).[49–51]

A typical EELS spectrum, shown in Fig. 4.24 (see also Fig. 3.18), contains a sharp and intense zero-loss peak, representing electrons that have passed through the specimen without losing energy through inelastic scattering.[49–51] The width of this peak depends on the instrumentation; it is typically of the order of 1 eV. The low-loss region (up to 50 eV energy loss) is dominated by one or more broader peaks which arise from inelastic scattering of the transmitted electrons by valence or conduction electrons in the specimen. These peaks are often referred to as plasmon peaks, since the inelastic scattering process involves the interaction of many valence electrons, which behave like a solid-state plasma. The strength of the plasmon peaks depends on specimen thickness: in fact, their integrated intensity I_p (relative to the intensity I_0 of the zero-loss peak) can be used to measure the local thickness $d = \lambda(\beta)\ln(I_p/I_0)$, where $\lambda(\beta)$ is the mean free path for inelastic scattering for a given collision angle β, which depends on the chemical composition of the specimen and on the electron accelerating voltage (see Sec. 6.2).

At higher energy loss (Fig. 4.24b), the electron intensity is much weaker, requiring a more sensitive detector or longer recording time. These features are analogous to the absorption edges found in X-ray absorption spectra (as we discuss

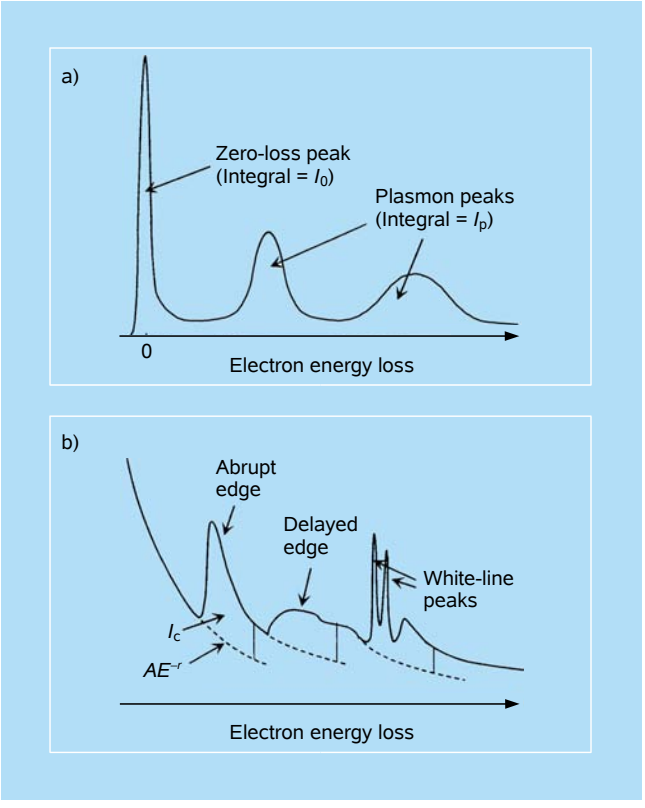


Figure 4.24 Typical regions of electron-energy-loss spectra. (a) General form of the low-loss region of an EELS spectrum, typically up to E = 50 eV. (b) Core-loss region, recorded at higher detector sensitivity and illustrating the typical shapes of abrupt, delayed and white-line edges, each superposed on a spectral background (dashed curves).[51]

comparatively in Chap. 7): they arise from scattering of the transmitted electrons by inner-shell (core) electrons of each element present in the specimen. The integrated intensity I_c below each ionization edge (in excess of the extrapolated background) is proportional to the amount of each element (N atoms per unit area): $I_c = N(I_0 + I_p)\sigma_c$. Here, σ_c is the core-loss scattering cross-section for the appropriate element: its value can be calculated or obtained experimentally. Although the expression for I_c allows the possibility of determining an absolute aerial density of N of each element, it is usually sufficient to measure concentration ratios of two or more elements.

Elsewhere,[52,53] the extent of detailed chemical information that may be gleaned from other features of Figs. 4.24b and 3.18 have been fully described. Valence states, the number of *d*-electrons in a transition metal ion, coordination fingerprints,[52b] bond mapping[52c] as well as bond distances may all, under favorable circumstances, be retrieved from EELS spectra. However, since we are concerned

here with 3D imaging in a chemically sensitive manner, we concentrate now on the kind of information that EFTEM (see Fig. 3.19) yields.

Owing to the very recent availability of high-intensity, sharply focused electron beams operating within an aberration-corrected microscope, truly atomic-scale chemical imaging is now possible as shown[52d] in Fig. 4.25. This example of spectroscopic imaging in the electron microscope represents the ultimate logical extension of the paths taken towards this end by Crewe (with his individual uranium atom images in 1970), by Suenaga on gadolinium atoms inside a single chain of $Gd@C_{82}$ metallofullerenes that were encapsulated by single-walled carbon nanotubes (SWNT), and by Gai and Colliex on boron-dopped SWNT in the early 2000s. Note that Fig. 4.25 shows clear atomic-scale intermixing of Mn and Ti ions on the B-sublattice of the ABO_3 perovskite structure.

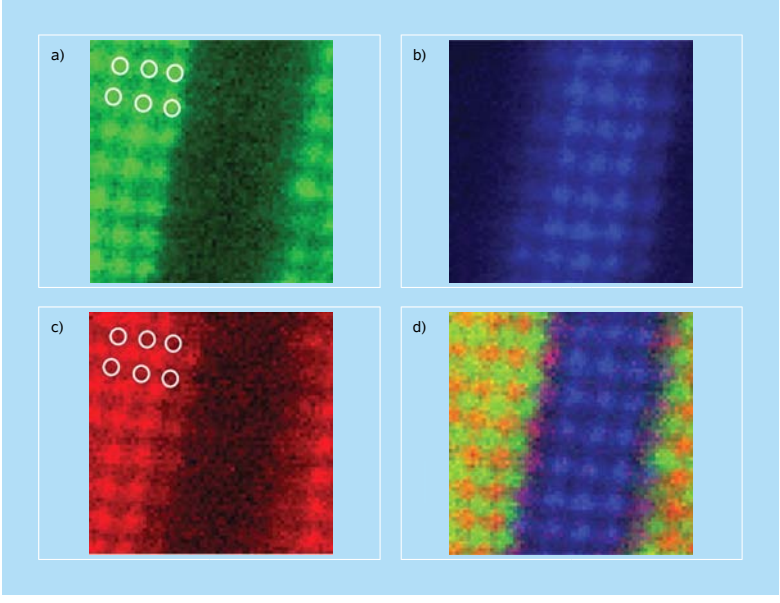


Figure 4.25 Spectroscopic imaging of an $La_{0.7}Sr_{0.3}MnO_3/SrTiO_3$ multilayer, showing the different chemical sublattices in a 64 × 64 pixel spectrum image extracted from 650 eV-wide electron-energy-loss spectra recorded at each pixel: (a) La M edge, (b) Ti L edge, (c) Mn L edge, and (d) red-green-blue false-color image obtained by combining the rescaled Mn, La, and Ti images. Each of the primary color maps is rescaled to include all data points within two standard deviations of the image mean. Note the lines of purple at the interface in (d), which indicate Mn–Ti intermixing on the B-site sublattice. The white circles indicate the position of the La columns, showing that the Mn lattice is offset. Live acquisition time for the 64 × 64 spectrum image was ~30 s; field of view: 3.1 nm.[52d]

A striking example of the analytical power of EFTEM is shown in Fig. 4.26, where a sequence of EELS spectra is recorded along a line (shown dashed at top right), where the zero-energy loss intensity was also measured. This example

involves a minute cluster that has nucleated inside a specimen of "316" stainless steel. From the resulting energy landscape depicted in Fig. 4.26, where six distinct elemental components are identified, the precise composition of (and the spatial variation within) the minute precipitate may be determined.[11a,54]

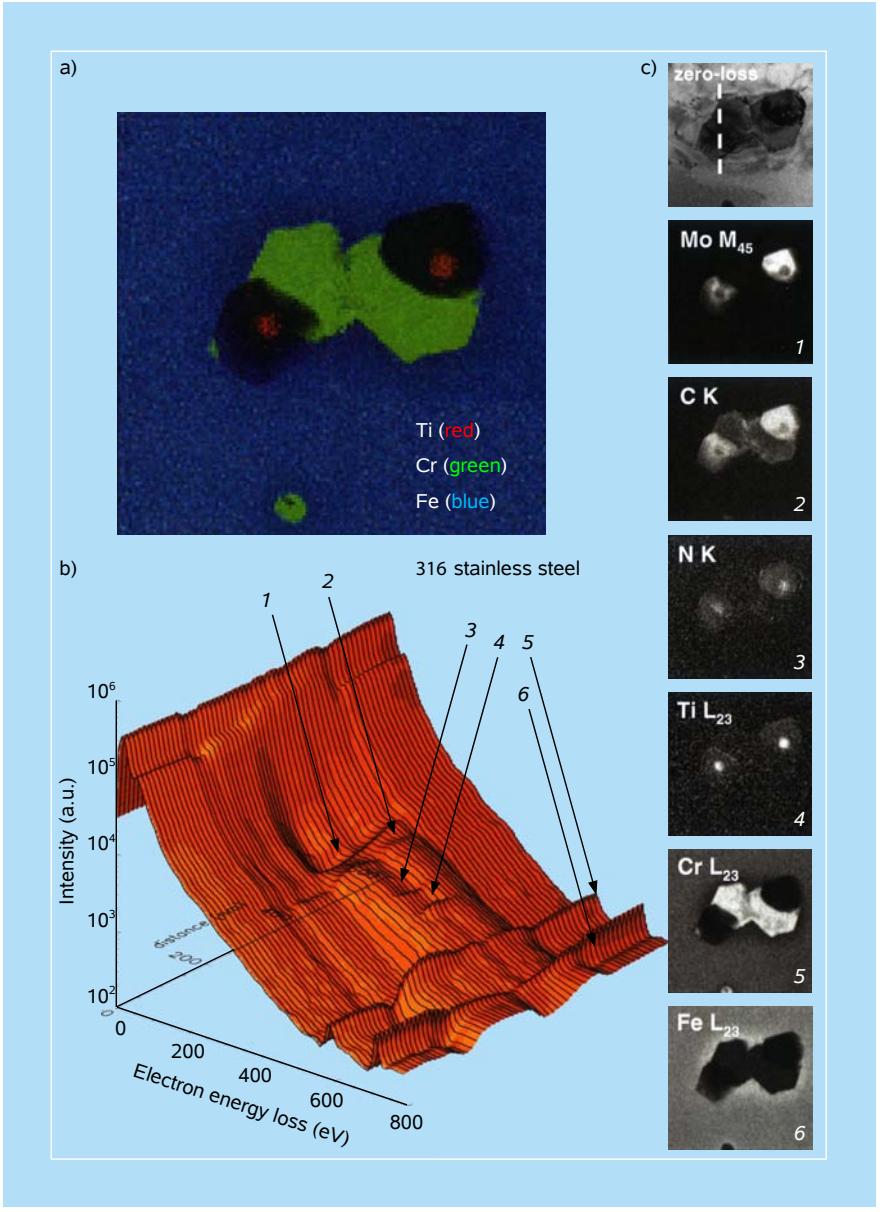


Figure 4.26 A line trace through a minute precipitate in a stainless steel specimen is the starting point in constructing the "landscape" which is composed of EELS spectra at discrete points long the (zero-loss) line. From the EELS spectra, traces of Mo, C, N, Ti, Cr, and Fe are detected.[54]

4.7 Atomic Resolution in an Environmental TEM

Notwithstanding that it is extremely difficult (but see Sec. 4.7.1 below) to operate TEM and STEM instruments in such a manner as to allow routine gas–solid and solid–liquid interactions to be studied under normal environmental conditions—simply because the mean free path of the electron is infinitesimal under these conditions — it is very encouraging to observe the major progress made by Gai, Helveg, Henry, Crozier and others in designing TEM systems that permit high-resolution electron microscopic images to be recorded of objects undergoing dynamic changes under modest (reduced) gas pressures.

Work done several decades ago[b] on the catalytic oxidation of single crystals of graphite by nanoparticles of a range of metals involved an electron microscope fitted with a hot stage that "tolerated" an ambient pressure of O_2 up to 3 mbar.[55] The stage in which the specimen was mounted and the dynamics of the catalysis investigated involved a differential pumping system, which was a primitive form of the methods pioneered by Gai.[56–60]

Gai[58–59] has perfected her *in situ* environmental TEM in such a fashion as to achieve atomic resolution in the image of nanoparticles participating in catalytic turnover. The set-up devised by her and Boyes and adapted by others is schematized in Fig. 4.27; and a typical result, illustrating how time-resolved and temperature-resolved heterogeneously catalyzed reactions may be followed by electron-microscopic imaging, is shown in Fig. 4.28. The highest degree of time resolution so far achieved by Gai *et al.*[60] falls in the range of milliseconds. Gai *et al.*[60] have also followed, *in situ*, the twinning transformation of tungsten carbide at elevated temperatures (Fig. 4.29). Again, the time resolution fell in the millisecond range. With faster camera sweep, submillisecond resolution may be obtained.

In a recent, timely review the full potential of environmental TEM (ETEM) has been discussed and illustrated, with special reference to its use in the study of model heterogeneous catalysts.[60] One especially important aspect of catalysis, that concerned with *nanoparticle shape dynamics*, is of considerable interest. The size, shape, and detailed surface structure of supported metal nanoparticles play a key role in their catalytic performance. Although nanoparticles with predefined shapes can be synthesized, there is no guarantee that such materials remain stable when exposed to reaction conditions. The gas constituents can induce changes in surface and interface energies and hence drive restructuring of the nanoparticles. Inspection of the ETEM images provides direct insight into the nature of gas-induced shape changes and helps to elucidate the driving forces for such transformations.

[b] *In situ* cinematographic study by electron microscopy of the catalytic oxidation of graphite, made by J. M. Thomas, was shown at the Royal Institution Christmas Lectures, December 1987, broadcast by the BBC, January 1988.

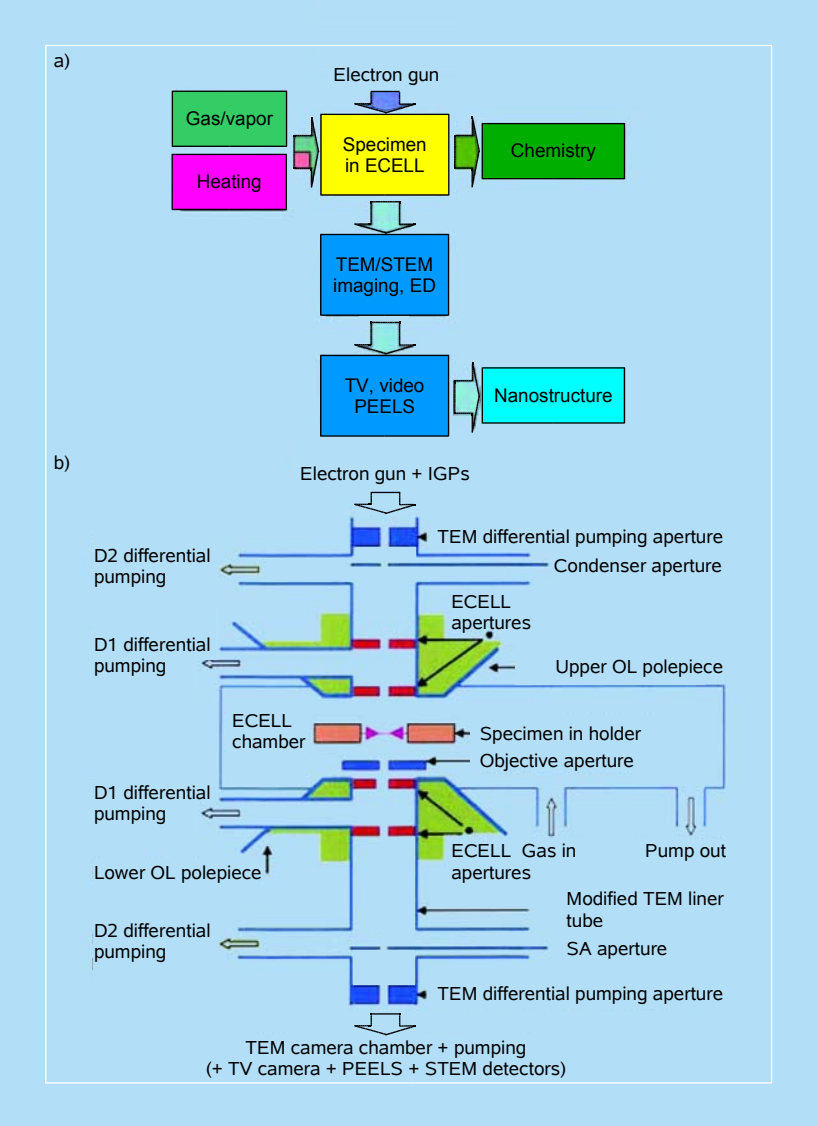


Figure 4.27 Environmental TEM. (a) Block diagram illustrating the Gai-Boyes set-up used for environmental TEM (ETEM). (b) Details of the ETEM pioneered by Gai. PEELS stands for parallel recording of EELS spectra.[60]

One prototype (model) system for studying complex gas-induced structural dynamics consists of copper nanoparticles dispersed on zinc oxide crystallites representing industrial methanol synthesis catalysts. It has been shown that the Cu particles can undergo shape changes in response to changes in the gas environment and that the shape changes affect the catalytic activities of the Cu nanoparticles.[61,62] Figure 4.30 shows atomically-resolved ETEM images reflecting the equilibrated

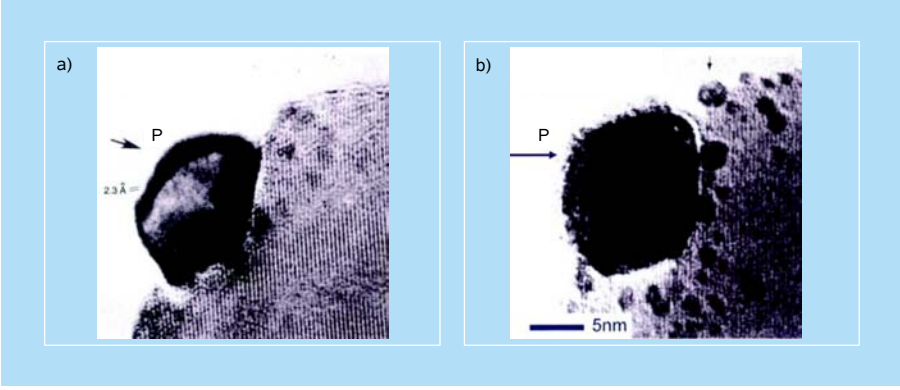


Figure 4.28 The first atomically resolved *in situ* images of a Pt nanoparticle catalyst supported on TiO_2 in an atmosphere (3 mbar) of H_2 at high temperatures:[72] (a) $T = 300°C$ and (b) $T = 450°C$; the horizontal arrow shows the passivation of TiO_x coating on the same Pt particle (P). Direct evidence of strong metal support interactions (SMSI) and particle shape changes was obtained. Reproduced from Gai's work.[56–59]

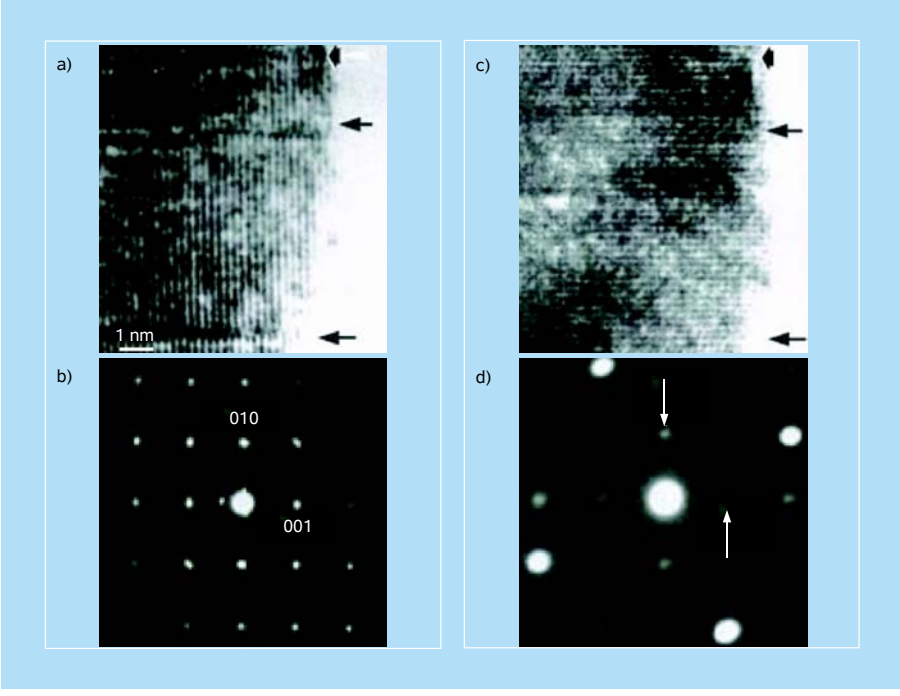


Figure 4.29 Atomically resolved *in situ* ETEM images of tungsten carbide exposed to H_2/He gas mixture (a,c) and electron diffraction patterns (b,d) recorded at room temperature (a,b) and at 450°C (c,d). The elimination of the thin boundaries could be followed in real time (at 30 frames per second).[73]

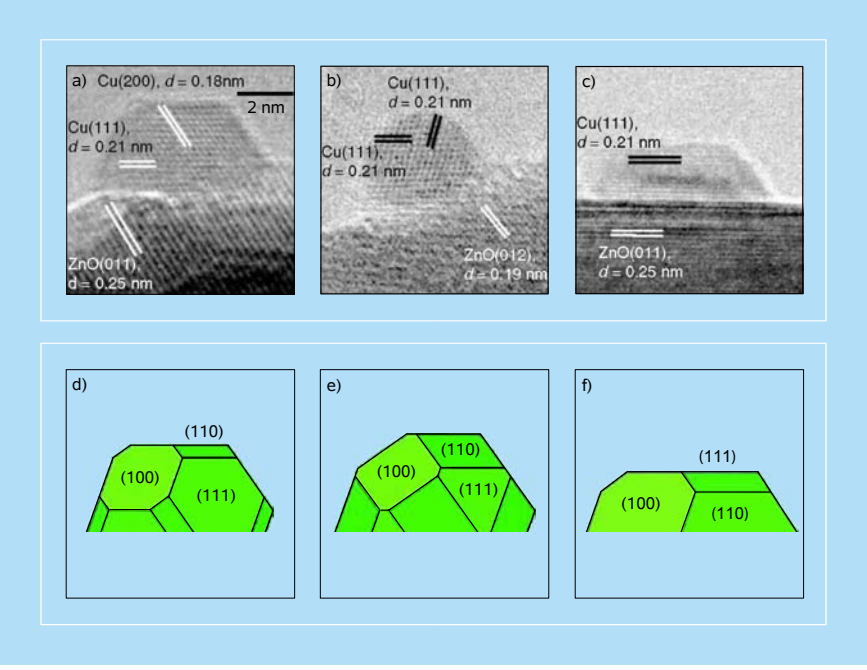


Figure 4.30 High-resolution imaging of exposed catalysts. (a–c) Atomically-resolved ETEM images of a Cu/ZnO catalyst in various gas environments. (d–f) The corresponding Wulff constructions of the Cu nanocrystals. (a) Image recorded at a pressure of 1.5 mbar H_2 at 220°C. (b) Image obtained with the Cu nanocrystals exposed to a gas mixture of H_2 and H_2O, H_2/H_2O = 3:1, at a total pressure of 1.5 mbar at 220°C. (c) Image obtained with the Cu nanocrystal exposed to a gas mixture of H_2 (95%) and CO (5%) at a total pressure of 5 mbar at 220°C.[62]

shapes of Cu nanocrystals on ZnO during exposure to different gas environments. In hydrogen gas, faceted metallic copper nanocrystals are observed (Fig. 4.30a). Upon addition of water vapor (Fig. 4.30b) to the hydrogen gas, the Cu nanocrystals transform into a more spherical shape, and this reshaping is attributed to water adsorption on the exposed Cu facets, driving surface reconstructions. Upon addition of carbon monoxide (Fig. 4.30c), the Cu nanocrystals transform into more disc-like structures (Fig. 4.30f). It was proposed that this wetting effect is due to a CO-induced change in the oxygen content of the ZnO surface and, thereby, the Cu/ZnO interface energy.[61,62] Insight into gas-dependent metal support interactions and confirmation for this interpretation can be obtained by combining the ETEM images with *in situ* EELS measurements.[63] A sufficient increase in the reduction potential of the gas, caused by CO, induces changes in the energy-loss near-edge structure (ELNES) of the Cu L_3 ionization edge, which can be attributed to a Cu–Zn alloy formation either at the Cu–Zn interface or on the Cu surface.[63]

Furthermore, on the basis of the lattice-resolved ETEM images, the crystallographic identity of the exposed facets of the projected Cu nanocrystals and the epitaxial relationship between the Cu/ZnO interface are revealed. The atomic-resolution information allows Wulff reconstructins (Figs. 4.30d–f) for the nanocrystals and quantitative determinations of the relative metal surface free energies and the interface free energies in the different gas environments to be obtained. A determination of surface and interface properties *in situ* represents an important perspective as it provides information about the concentrations of the different types of surface sites (facets, corners, steps, etc.) under the various environmental conditions. The finding that the nanoclusters can undergo dynamic shape changes in response to changes in the gas environment has the general implication that the distribution of catalytic active sites, and hence the catalytic activity, can be self-regulated by the reaction conditions.

Another aspect of structure–activity relationships in the Au nanoparticle catalysts is also discussed by Gai *et al.*[60] This is the study by Henry and co-workers[64] (see Fig. 4.31) of reversible change in the shape of Au nanoparticles depending upon the gas to which they are exposed. Atomically-resolved images were produced of Au nanoparticles, deposited at room temperature on carbon in 2 mbar of H_2 (Fig. 4.31a), then 2 mbar of O_2 (Fig. 4.31b) and again 2 mbar of H_2 (Fig. 4.31c). These resolved images show a reversible change from a faceted shape in H_2 to a whole spherical shape during O_2 exposure to the final state in which the faceted shape is recovered. Because this behavior of Au particles was observed on both TiO_2 and amorphous carbon, it indicates that the shape changes are driven by the process of adsorption on the metal surface.

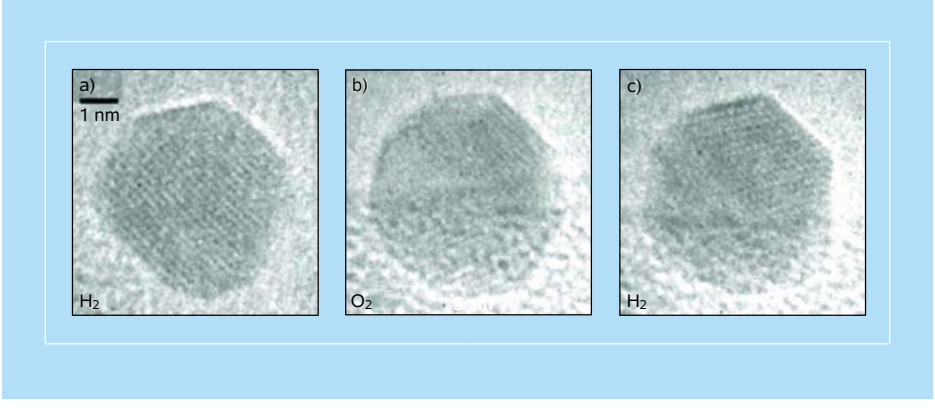


Figure 4.31 High-resolution imaging of exposed clusters. Lattice-resolved ETEM images of an Au cluster during exposure to a cycle of (a) H_2, (b) O_2, and (c) H_2 at room temperature.[64]

4.7.1 Atomic-scale electron microscopy at ambient pressure by exploiting the technology of microelectromechanical systems

A highly significant advance in environmental TEM has been achieved very recently in a joint Dutch–Danish collaboration reported by Creemer *et al.*[65] The strategy used by these workers (in order to limit the number of gas atoms in the specimen stage of the microscope) was to build a novel nanoreactor that is functionalized with micrometer-sized gas-flow channels, electron-transparent windows (made of amorphous SiN_x) and a special heating device, as depicted in Fig. 4.32. The microelectromechanical systems (MEMS) technology approach enables the integration of a heater and a thermal sensor for variable-temperature experiments at a sub-second time scale. In a proof-of-principle experiment Creemer *et al.*[65] investigated the Cu/ZnO system, the catalyst of choice in methanol synthesis from CO and H_2 mixtures. Figure 4.33 illustrates the results that they obtained.

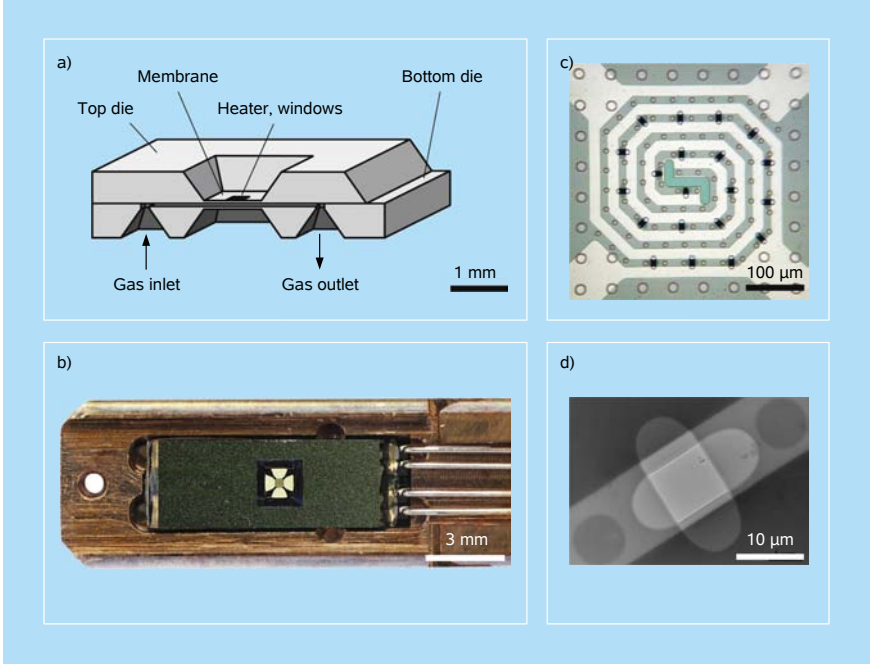


Figure 4.32 Illustration of the nanoreactor device. (a) Schematic cross-section of the nanoreactor. (b) Optical image of the TEM holder with the integrated nanoreactor and the four electrical probe contacts. (c) Optical close-up of the nanoreactor membrane. The bright spiral is the Pt heater. The small ovaloids are the electron-transparent windows. The circles are the SiO_2 spacers that define the minimum height of the gas channel. (d) A low-magnification TEM image of a pair of superimposed 10 nm thick windows. Their alignment creates a highly electron-transparent (bright) square through which high-resolution TEM imaging can be performed.[65]

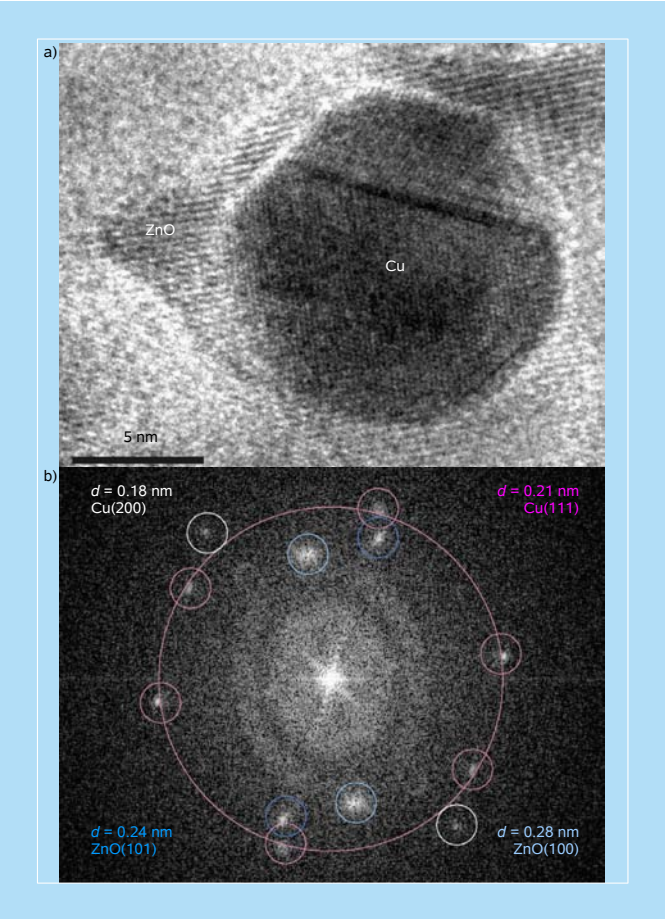


Figure 4.33 A representative HRTEM image of the Cu/ZnO catalyst during exposure to 1.2 bar hydrogen at 500°C. (a) The image displays lattice fringes of a twinned Cu nanocrystal and of the ZnO support. (b) A Fourier transform of (a). The bright dots represent sets of lattice fringes. Their lattice spacing corresponds to the distance to the origin and reveals the crystallographic identity. The large red circle corresponds a spacing of 0.21 nm. The smallest resolved lattice spacing is 0.18 nm.[65a]

This work is well suited for the *in situ* study of a variety of other academically and commercially important problems at very high spatial resolution and good temporal resolution. The concept of the nanoreactor devised by Creemer *et al.*[65a] and extended by Allard *et al.*[65b] seems versatile enough to be extended to other environments, including higher pressures, heavier constituents, and possibly even liquids. Moreover, parallel measurements may be made with the nanoreactor by utilizing mass spectrometry, optical and X-ray techniques. This advance promises[66] much in the future investigation of functional nanomaterials and the manner in which they interact with various ambient environments, both in the physical and biological sciences and in engineering (see also Sec. 7.1).

References

1. A. I. Kirkland and J. Sloan, *Top. Catal.*, **21**, 139 (2002).
2. W. O. Saxton in "*Image and Signal Processing in Electron Microscopy*," Eds. P. W. Hawkes, F. P Ottensmeyer, W. O. Saxton and A. Rosenfeld, Scanning Microscopy International, Chicago (1988).
3. R. R. Meyer, A. I. Kirkland and W. O. Saxton, *Ultramicroscopy*, **92**, 89 (2002).
4. P. Gambardella, Ž. Šljivančanin, B. Hammer, M. Blanc, K. Kuhnke and K. Kern, *Phys. Rev. Lett.*, **87**, 056103 (2001).
5. (a) A. I. Kirkland and R. R. Meyer, *Microsc. Microanal.*, **10**, 401 (2004).
   (b) C. L. Jia and A. Thust, *Phys. Rev. Lett.*, **82**, 5052 (1999).
   (c) C. L. Jia, M. Lentzen and K. Urban, *Science*, **299**, 870 (2003).
   (d) C. L. Jia and K. Urban, *Science*, **303**, 2001 (2004).
6. L. C. Gontard, L.-Y. Chang, C. J. D. Hetherington, A. I. Kirkland, D. Ozkaya and R. E. Dunin-Borkowski, *Angew. Chem., Intl. Ed.*, **46**, 3683 (2007).
7. D. Ferrer, A. Torres-Castro, X. Gao, S. Sepúlveda-Guzmán, U. Ortiz-Méndez and M. José-Yacamán, *Nano Lett.*, **7**, 1701 (2007).
8. A. A. Herzing, M. Watanabe, J. K. Edwards, M. Conte, Z.-R. Tang, G. J. Hutchings and C. J. Kiely, *Faraday Discuss.*, **138**, 337 (2008).
9. L. A. Bursill, E. A. Lodge and J. M. Thomas, *Nature*, **286**, 111 (1980).
10. J. M. Thomas, O. Terasaki, P. L. Gai, W. Zhou and J. Gonzalez-Calbet, *Acc. Chem. Res.*, **34**, 583 (2001).
11. (a) J. M. Thomas in "*Physical Biology: From Atoms to Medicine*," Ed. A. H. Zewail, Imperial College Press, London, p. 51 (2008).
    (b) P. A. Wright, W. Zhou, J. Pérez-Pariente and M. Arranz, *J. Am. Chem. Soc.*, **127**, 494 (2005).
12. (a) P. A. Wright, S. Natarajan, J. M. Thomas, R. G. Bell, P. L. Gai-Boyes, R. H. Jones and J. Chen, *Angew. Chem., Intl. Ed.*, **31**, 1472 (1992).
    (b) J. Chen and J. M. Thomas, *J. Chem. Soc., Chem. Commun.*, 603 (1994).
13. P. L. Gai-Boyes, J. M. Thomas, P. A. Wright, R. H. Jones, S. Natarajan, J. Chen, and R. Xu, *J. Phys. Chem.*, **96**, 8206 (1992).
14. G. R. Millward, S. Ramdas and J. M. Thomas, *Proc. R. Soc. Lond. A*, **399**, 57 (1985).
15. L. G. Mallinson, D. A. Jefferson, J. M. Thomas and J. L. Hutchison, *Phil. Trans. R. Soc. Lond. A*, **295**, 537 (1980).
16. C. Baerlocher, F. Gramm, L. Massüger, L. B. McCusker, Z. He, S. Hovmöller and X. Zou, *Science*, **315**, 1113 (2007).
17. G. Oszlányi and A. Sütő, *Acta Cryst. A.*, **61**, 147 (2005).
18. (a) C. Baerlocher, L. B. McCusker and L. Palatinus, *Z. Kristallogr.*, **222**, 47 (2007).
    (b) J. Sun, C. Bonneau, Á. Cantín, A. Corma, M. J. Díaz-Cabañas, M. Moliner, D. Zhang, M. Li and X. Zou, *Nature*, **458**, 1154 (2009).
    (c) K. D. M. Harris and J. M. Thomas, *Chem. Cat. Chem.*, **1**, 223 (2009).
19. J.-L. Sun, Z. B. He, S. Hovmöller, X. D. Zou, F. Gramm, C. Baerlocher and L. B. McCusker, *Z. Kristallogr.*, in press.
20. (a) W. D. Pyrz, D. A. Blom, T. Vogt, D. J. Buttrey, *Angew. Chem., Int. Ed.*, **47**, 2788 (2008).

(b) J. M. Thomas, *Chem. Phys. Chem*., **9**, 1363 (2008).
(c) T. Vogt, D. J. Buttrey, W. D. Pyrz, D. A. Blom and J. M. Thomas, manuscript submitted.
21. E. Prince, in "*The Rietveld Method*," Ed. R. A. Young, Oxford Universtiy Press, New York, p. 43 (1995).
22. M. A. Roberts, G. Sankar, J. M. Thomas, R. H. Jones, H. Du, J. Chen, W. Pang and R. Xu, *Nature*, **381**, 401 (1996).
23. T. Vogt, in "*Encyclopedia of Inorganic Chemistry*," 2nd ed., Ed. R. B. King, Wiley, New York (2005).
24. F. Guo and K. D. M. Harris, *J. Am. Chem. Soc*., **127**, 7314 (2005).
25. (a) P. DeSanto, Jr., D. J. Buttrey, R. K. Grasseli, C. G. Lugmair, A. F. Volpe, Jr., B. H. Toby and T. Vogt, *Z. Kristallogr*., **219**, 152 (2004).
(b) T. Uchikubo, K. Oshima, A. Kayou, T. Umezawa, K. Kiyono and I. Sawaki, European Patent 529,853 (1993) assigned to the Mitsubishi Chemical Corporation, Tokyo, Japan.
26. M. Sadakane and W. Ueda, in "*Turning Points in Solid-State, Materials and Surface Science*," Eds. K. D. M. Harris and P. P. Edwards, RSC Publishing, Cambridge, p. 507 (2008).
27. (a) P. DeSanto, Jr., D. J. Buttrey, R. K. Grasselli, C. G. Lugmair, A. F. Volpe, B. H. Toby and T. Vogt, *Top. Catal*., **23**, 23 (2003).
(b) R. K. Grasselli, C. G. Lugmair and A. F. Volpe, Jr., *Top. Catal.*, **50**, 66 (2008).
28. J. C. H. Spence, "*High-Resolution Electron Microscopy*," 3rd ed., Oxford University Press, New York, p. 125 (2003).
29. (a) D. J. Buttrey, in "*Turning Points in Solid-State, Materials and Surface Science*," Ed. K. D. M. Harris and P. P. Edwards, RSC Publishing, Cambridge, p. 754 (2008).
(b) M. Sadakane, K. Yamagata, K. Kodato, K. Endo, K. Toriumi, Y. Ozawa, T. Ozeki, T. Nagai, Y. Matsui, N. Sakaguchi, W. D. Pyrz, D. J. Buttrey, D. A. Blom, T. Vogt and W. Ueda, *Angew. Chem*., *Intl. Ed*., **48**, 3782 (2009).
30. D. J. Buttrey, D. A. Jefferson and J. M. Thomas, *Phil. Mag. A*, **53**, 897 (1986).
31. D. A. Jefferson, J. M. Thomas, M. K. Uppal and R. K. Grasselli, *J. Chem. Soc*., *Chem. Commun*., 594 (1983).
32. W. Zhou, D. A. Jefferson, M. Alario-Franco and J. M. Thomas, *J. Phys. Chem*, **91**, 512 (1987).
33. A. Komeili, Z. Li, D. K. Newman and G. J. Jensen, *Science*, **311**, 242 (2006).
34. A. Scheffel, M. Gruska, D. Faivre, A. Linaroudis, J. M. Plitzko and D. Schüler, *Nature*, **440**, 110 (2006).
35. V. Jourdain, E. T. Simpson, M. Paillet, T. Kasama, R. E. Dunin-Borkowski, P. Poncharal, A. Zahab, A. Loiseau, J. Robertson and P. Bernier, *J. Phys. Chem. B*, **110**, 9759 (2006).
36. J. M. Thomas, E. T. Simpson, T. Kasama and R. E. Dunin-Borkowski, *Acc. Chem. Res*., **41**, 665 (2008).
37. R. J. Harrison, R. E. Dunin-Borkowski and A. Putnis, *Proc. Natl. Acad. Sci. USA*, **99**, 16556 (2002).
38. Y. Sakamoto, T.-W. Kim, R. Ryoo and O. Terasaki, *Angew. Chem*., *Intl. Ed*., **43**, 5231 (2004).

39. Y. Sakamoto, I. Díaz, O. Terasaki, D. Zhao, J. Pérez-Pariente, J. M. Kim and G. D. Stucky, *J. Phys. Chem. B*, **106**, 3118 (2002).
40. A. E. Garcia-Bennett, O. Terasaki, S. Che and T. Tatsumi, *Chem. Mater.*, **16**, 813 (2004).
41. M. Kaneda, T. Tsubakiyama, A. Carlsson, Y. Sakamoto, T. Ohsuna, O. Terasaki, S. H. Joo and R. Ryoo, *J. Phys. Chem. B*, **106**, 1256 (2002).
42. (a) X. D. Zou, Z. M. Mo, S. Hovmöller, X. Z. Li and K. H. Kuo, *Acta Cryst. A*, **59**, 526 (2003).
    (b) R. Vincent and P. A. Midgley, *Ultramicroscopy*, **53**, 271 (1994).
    (c) D. Xie, C. Baerlocher and L. B. McCusker, *J. Appl. Cryst.*, **41**, 1115 (2008).
    (d) J. Wu, K. Leinenweber, J. C. H. Spence and M. O'Keeffe, *Nature Mater.*, **5**, 647 (2006).
43. B. Böttcher, S. A. Wynne and R. A. Crowther, *Nature*, **386**, 88 (1997).
44. W. Jiang, M. L. Baker, J. Jakana, P. R. Weigele, J. King and W. Chiu, *Nature*, **451**, 1130 (2008).
45. S. A. Wynne, R. A. Crowther and A. G. W. Leslie, *Mol. Cell*, **3**, 771 (1999).
46. (a) P. N. T. Unwin and R. Henderson, *J. Mol. Biol.*, **94**, 425 (1975).
    (b) J. E. Evans, C. Hetherington, A. Kirkland, L.-Y. Chang, H. Stahlberg and N. Browning, *Ultramicroscopy*, **108**, 1636 (2008).
47. W. Kühlbrandt and D. N. Wang, *Nature*, **350**, 130 (1991).
48. (a) N. Unwin, *Nature*, **373**, 37 (1995).
    (b) J. C. H. Spence, private communication (May 2009).
49. R. F. Egerton, "*Electron Energy Loss Spectroscopy in the Electron Microscope*," 2nd ed., Plenum Press, New York (1996).
50. (a) D. A. Jefferson, J. M. Thomas and R. F. Egerton, *Chem. Brit.*, **17**, 514 (1981).
    (b) M. Tanaka, M. Takeguchi and K. Furuya, *Ultramicroscopy*, **108**, 1427 (2008).
51. R. F. Egerton, *Top. Catal.*, **21**, 185 (2002).
52. (a) J. M. Thomas, B. G. Williams and T. G. Sparrow, *Acc. Chem. Res.*, **18**, 324 (1985).
    (b) R. Brydson, H. Sauer, W. Engel, J. M. Thomas and E. Zeitler, *J. Chem. Soc., Chem. Commun.*, 1010 (1989).
    (c) R. Arenal, F. de la Peña, O. Stéphan, M. Walls, M. Tencé, A. Loiseau and C. Colliex, *Ultramicroscopy*, **109**, 32 (2008).
    (d) D. A. Muller, L. Fitting Kourkoutis, M. Murfitt, J. H. Song, H. Y. Hwang, J. Silcox, N. Dellby and O. L. Krivanek, *Science*, **319**, 1073 (2008).
53. (a) J. C. H. Spence and M. R. Howells, *Ultramicroscopy*, **93**, 213 (2002).
    (b) J. M. Thomas and P. A. Midgley, *Chem. Commun.*, 1253 (2004).
54. P. J. Thomas and P. A. Midgley, *Top. Catal.*, **21**, 109 (2002).
55. (a) J. M. Thomas, in "*Chemistry and Physics of Carbon*," Vol. 1., Ed. P. L. Walker, Marcel Dekker, New York, p. 122 (1965).
    (b) G. M. Parkinson, *Catal. Lett.*, **2**, 303 (1989).
56. P. L. Gai-Boyes, *Catal. Rev.*, **34**, 1 (1992).
57. P. L. Gai and K. Kourtakis, *Science*, **267**, 661 (1995).
58. J. M. Thomas and P. L. Gai, in "*Advances in Catalysis*," Vol. 48., Eds. H. Knözinger and B. C. Gates, Academic Press, London, p. 171 (2004).
59. P. L. Gai, in "*Nanocharacterisation*," Eds. A. I. Kirkland and J. L. Hutchison, Royal Society of Chemistry, Cambridge, p. 268 (2007).

60. P. L. Gai, E. D. Boyes, S. Helveg, P. L. Hansen, S. Giorgio and C. R. Henry, *MRS Bull.*, **32**, 1044 (2007).
61. J.-D. Grunwaldt, A. M. Molenbroek, N.-Y. Topsøe, H. Topsøe and B. S. Clausen, *J. Catal.*, **194**, 452 (2000).
62. P. L. Hansen, J. B. Wagner, S. Helveg, J. R. Rostrup-Nielsen, B. S. Clausen and H. Topsøe, *Science*, **295**, 2053 (2002).
63. J. B. Wagner, P. L. Hansen, A. M. Molenbroek, H. Topsøe, B. S. Clausen and S. Helveg, *J. Phys. Chem. B*, **107**, 7753 (2003).
64. S. Giorgio, S. Sao Joao, S. Nitsche, D. Chaudanson, G. Sitja and C. R. Henry, *Ultramicroscopy*, **106**, 503 (2006).
65. (a) J. F. Creemer, S. Helveg, G. H. Hoveling, S. Ullmann, A. M. Molenbroek, P. M. Sarro and H. W. Zandbergen, *Ultramicroscopy*, **108**, 993 (2008).
(b) L. F. Allard, W. C. Bigelow, M. José-Yacamán, D. P. Nackashi, J. Damiano and S. E. Mick, *Microsc. Res. Tech.*, **72**, 208 (2009).
66. J. M. Thomas and J.-C. Hernandez-Garrido, *Angew. Chem., Intl. Ed.*, **48**, 3904 (2009).
67. M. E. Leonowicz and D. E. W. Vaughan, *Nature*, **329**, 819 (1987)
68. J. M. Thomas, P. A. Midgley, T. J. V. Yates, J. S. Barnard, R. Raja, I. Arslan and M. Weyland, *Angew. Chem.*, *Intl. Ed.*, **43**, 6745 (2004).
69. P. A. Midgley, J. M. Thomas, L. Laffont, M. Weyland, R. Raja, B. F. G. Johnson and T. Khimyak, *J. Phys. Chem. B*, **108**, 4590 (2004).
70. R. A. Crowther, *Phil. Trans. R. Soc. B*, **363**, 2441 (2008).
71. W. Jiang, J. Chang, J. Jakana, P. Weigele, J. King and W. Chiu, *Nature*, **439**, 612 (2006).
72. (a) E. D. Boyes and P. L. Gai, *Ultramicroscopy*, **67**, 219 (1997).
(b) P. L. Gai, *Adv. Mater.*, **10**, 1259 (1998).
73. P. L. Gai, C. C. Torardi and E. D. Boyes, in "*Turning Points in Solid-State, Materials and Surface Science*," Eds. K. D. M. Harris and P. P. Edwards, RSC Publishing, Cambridge, p. 745 (2008).

This page intentionally left blank

Chapter 5

4D Electron Imaging in Space and Time

Principles

In this and subsequent chapters, the focus is on the principles of electron imaging in real space, Fourier space, and energy space, all with time being the fourth dimension. Before discussing the applications in Chap. 6, here we address three fundamental questions of 4D imaging with electrons:

- How can atomic-scale time resolution be reached?
- How can timed single electrons be used for real-space imaging?
- How can ultrafast dynamics be resolved with slow-response detectors?

5.1 Atomic-Scale Resolution in Time

5.1.1 Matter particle–wave duality

One of the profound findings of quantum mechanics is the particle–wave duality of matter expressed by the remarkable relationship of de Broglie in 1924: $\lambda = h/p$, where λ is the de Broglie matter wavelength, h is the Planck constant, and p is the momentum of the particle/wave, $p = m\upsilon$. In wave mechanics, the state of any material system is totally defined by a spatially varying wave function from which it is possible to determine only *probabilities* of the system's having a certain spatial configuration or having a certain momentum, rather than specific values for those quantities. For example, the probability that a particle the wave function of which, $\Psi(r,t)$, will be found at a position r in a small volume dv, $|\Psi(r,t)|^2\, dv$, is proportional to the squared modulus of the wave function.

The form of wave functions representing atomic motion is determined by the energy landscape of the system, called the potential energy surface, which in a molecule, for example, depends on the specific arrangement of electrons around the nuclei. The total energy of an isolated system is constant, and the difference between the total energy and the potential energy is energy of motion, or kinetic

energy (E_{kin}), of the constituent particles. The classically allowed region of the potential energy landscape of a system consists of all regions in which the total energy is greater than the potential energy; that is, where E_{kin} is nonnegative. According to classical mechanics, motion of the system is only possible in this region. If the classically allowed region is limited in extent (i.e., the system is confined or bound), only certain specific values of the energy are possible.

When the energy of a system is exactly defined (energy uncertainty $\Delta E = 0$), its wave function spreads over (and even somewhat beyond) the entire classically allowed region of space, and the position probability density (PPD) is independent of time. Mathematically, the spatial and temporal dependences of the wave function are separable, and $\Psi(r,t)$ may be written as

$$\Psi(r,t) = \psi(r)\exp(-iEt/\hbar), \tag{5.1}$$

where the position-independent temporal phase factor, $\exp(-iEt/\hbar)$, is a complex number of unity modulus. Thus,

$$|\Psi(r,t)|^2 = \psi(r)\exp(-iEt/\hbar)\psi^*(r)\exp(-iEt/\hbar) = |\psi(r)|^2, \tag{5.2}$$

with no time dependence.

To illustrate these concepts, we present in Fig. 5.1 the simple case of a diatomic molecule vibrating in a parabolic, or harmonic, potential energy well, which is an approximate representation of the vibrating iodine molecule, one of the first studied in real time (herein we speak of molecules but the concepts are applicable to more complex systems such as crystals or biological assemblies). The horizontal axis represents the separation, r, of atomic nuclei. The possible energy values of the system are represented by the horizontal lines labeled by the quantum number n. Examples of wave functions $\psi_n(r)$ corresponding to $n = 0$ and 16–20 are plotted, with the zero of the wave function amplitude axis set equal in each case to the corresponding energy value. The sign of the wave function alternates from positive to negative with a spatial period determined by the kinetic energy, which is zero at the walls of the well and has its maximum value at the center.

When the physical state of a molecule is described by one of these wave functions of definite energy, a measurement of atomic position will reflect only the stationary probability distribution, with no nuclear motion. Taking $n = 20$ of Fig. 5.1, for example, there are 21 peaks in the PPD (corresponding to both positive and negative peaks in the wave function), distributed from one wall of the potential to the other. As another example, the PPD and the momentum probability distribution for a stationary state of a more realistic potential, one modeled on the quasi-bound excited state of sodium iodide (NaI) (see below), are shown in Fig. 5.2. The PPD fills the classically allowed region of the potential well, from 2.75 Å to 10.19 Å, and the momentum probability has peaks that correspond to the kinetic energy ($\sim$5300 cm^{-1}; 1 $\text{cm}^{-1} = 1.98648\times10^{-16}$ ergs) near the center of the well.

Motion is observed, on time scales shorter than the decoherence time, when the system is prepared in a *coherent* (*superposition*) *state*, making the state function a coherent sum of stationary wave functions for different energies E_i:

$$\Psi(r,t) = \sum_i c_i \psi_i(r) \exp(-iE_i t/\hbar). \tag{5.3}$$

The coefficients c_i determine the relative contributions of the wave functions to the superposition. The sum is said to be coherent when well-defined phase relations exist between the constituent wave functions, similar to the wave and optical coherence we discussed in Chap. 2. The PPD in this case is simply

$$\Psi(r,t)\Psi^*(r,t) = \left[\sum_i c_i \psi_i(r) \exp(-iE_i t/\hbar)\right]\left[\sum_j c_j^* \psi_j^*(r) \exp(+iE_j t/\hbar)\right]. \tag{5.4}$$

One can see that cross-terms $(i \neq i)$ in this product are *not* time independent (the phase factors do not cancel), but oscillate at the frequencies determined by the energy differences $(E_i - E_j)$, so the expected outcome of a position measurement changes with time — *the system is in motion*!

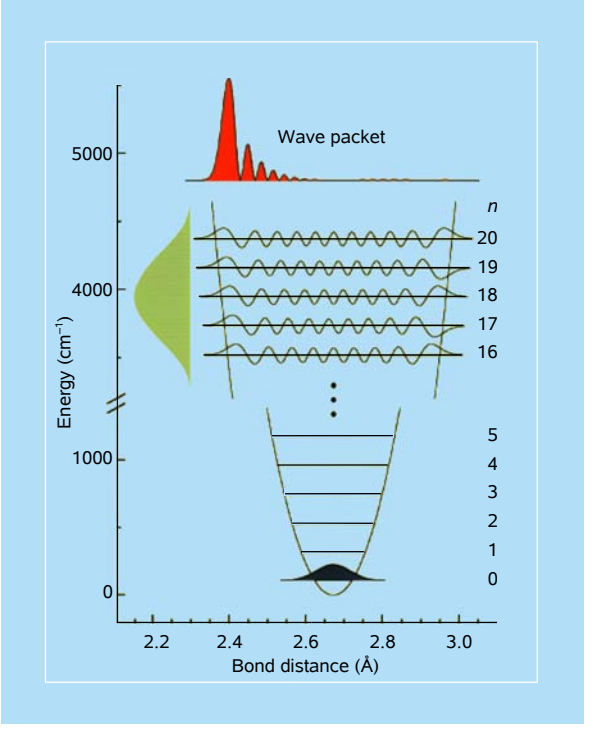


Figure 5.1 Diatomic molecule in a harmonic oscillator potential: stationary wave functions and formation of a localized wave packet. Figures 5.1–5.13 are adapted from Ref. 4 in Further Reading.

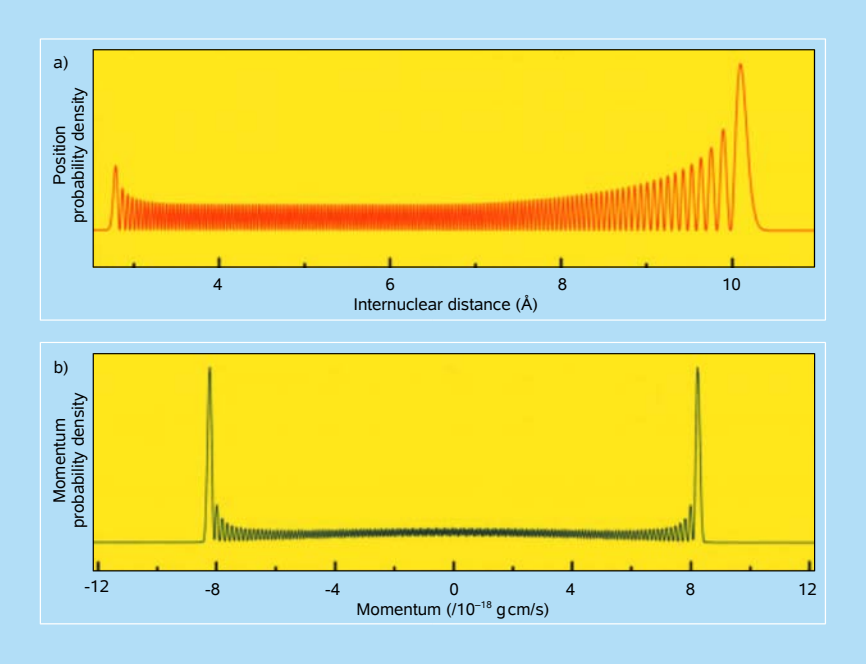


Figure 5.2 Stationary state of the theoretical bound potential of sodium iodide (NaI, $E = 30191.0\ \text{cm}^{-1}$, $n = 150$), modeled on the real quasi-bound potential that is covalent in nature at short distances and ionic at long distances — position and momentum probability distributions.

5.1.2 Analogy with light

The above is an example of wave superposition and *interference*, which has an analogy with the more familiar Young double-slit light interference experiment. As the wave function is a probability wave, so light is a wave of electric and magnetic fields whose amplitudes change at a regular interval, the wavelength, from positive to negative to positive along the direction of propagation. Like the PPD, the intensity of light at a point in space is given by the square of the field amplitude. When light from two or more sources overlaps in space, the instantaneous field amplitudes (not intensities) from each source must be added together to produce the resultant light field, as discussed in Chap. 2.

In Thomas Young's two-slit experiment of 1801, light from a single source passes through two parallel slits cut in an opaque screen to produce, in the space beyond, two *phase-coherent* fields of equal wavelength and amplitude. At points for which the distances to the two slits differ by $n + ½$ wavelengths (for integer n), the amplitudes of the two superposed fields are opposite in sign and add to zero at all times; no light is detected. Elsewhere, the amplitudes do not cancel. Thus, a stationary pattern of light and dark interference fringes are produced. If, instead,

incoherent (without a well defined phase relation) fields of the same amplitude were superposed, the intensities observed from each slit individually would simply add together and no interference fringes would be seen. Waves are coherent if they originate from a common point on the source. The addition of many such source points leads to partial coherence and reduced fringe visibility. An extended source similarity reduces contrast in phase–contrast images.

5.1.3 Classical atoms: Wave packets

In a manner analogous to the interference of light waves in the Young experiment, the constituent matter wave functions of a coherent superposition state interfere constructively and destructively, and the resultant total wave function may have a large amplitude in only a limited area of the classical region at any given time. This constitutes a *wave packet*. At the top of Fig. 5.1 is shown a specific example of the PPD of a wave packet formed from a superposition of the plotted wave functions $n = 16$–20 with weighting according to the distribution curve at the left of the figure. As time advances [$t > 0$ in $\exp(-iE_i t/\hbar)$], the phase relationships between wave functions change, causing the wave packet to move. As long as the PPD remains sufficiently localized on the scale of the total classical region, as it does in the figure, a discussion in terms of the classical concepts of particle position and momentum is entirely appropriate. Such is the case for the wave packet of Fig. 5.1, which will oscillate back and forth across the potential well indefinitely as would a classical particle at the same energy in that potential.

In order for a laser pulse to prepare a system in a localized wave packet, the deposited energy cannot be precisely defined, as wave functions at a broad range of energies E_i must be included in the superposition state. Given that the sizes of the energy quanta of a radiation field are determined by the constituent wavelengths, this translates into a requirement for a broad range of wavelengths in the field, which is also consistent with the nature of ultrafast pulsed laser sources. (Ultrashort light pulses are produced much as wave packets are formed, by coherently combining a large number of different wavelengths of light, each corresponding to a characteristic mode of the laser cavity; see below.) Thus, localization in time and in space are simultaneously achievable by virtue of the energy uncertainty!

When the extent of the classically allowed region is unbounded, the energy may take on all values rather than being restricted to quantized levels, but the same principle of superposition of wave functions applies to building wave packets; the sum in Eq. (5.3) becomes an integral. One example of unbound motion is motion at constant potential energy — that is, for a *free particle*. It is possible in this case to create a wave packet in one dimension with a very simple bell-shaped spatial profile defined by the Gaussian equation

$$\Psi(r,t) = N(t)\exp\left(-\frac{\left(x - x_0 - \langle v\rangle\ t\right)^2}{2\left[\Delta x(t)\right]^2}\right), \tag{5.5}$$

where $\Delta x(t)$ can be shown to be the root mean square (RMS) deviation in the PPD, $|\Psi(x,t)|^2$, and $N(t)$ is a normalization constant. The RMS deviation of a probability distribution $P(x)$ is given by

$$\left[\int_{-\infty}^{+\infty}(x-\bar{x})^2 P(x)dx\right]^{1/2}, \tag{5.6}$$

where $\bar{x}$ is the mean value of $\int xP(x)dx$ integrated from minus to plus infinity. (The time dependence of Δx is considered below.) The maximum of the wave function or PPD is at $x_0 + \langle\upsilon\rangle t$ and thus moves at velocity $\langle\upsilon\rangle$. Note that the full width at half maximum (FWHM) of a Gaussian function is $2(2\ln 2)^{1/2} = 2.3548$ times its RMS deviation, so the full width, $\Delta x^{\mathrm{FW}}(t)$, of the PPD is equal to $2.3548\times\Delta x(t)$.

The well-known Heisenberg uncertainty principle states that the degree of localization in space of a quantum system is inversely related to the minimum uncertainty in the system's momentum. That is, the more limited the extent of the PPD $|\Psi(x,t)|^2$, the larger the uncertainty in the momentum, and thus the less well-defined the future trajectory is. When the uncertainties Δx and Δp are defined as RMS deviations of the probability distributions, then the uncertainty relation is $\Delta x\Delta p \geq \hbar/2$, where $\hbar = h/2\pi$.

For a Gaussian free-particle wave packet having the minimum value of the uncertainty product, $\Delta x\Delta p = \hbar/2$, the momentum distribution is also Gaussian, and the contribution of the momentum uncertainty to the widening of the packet as it moves is expressed in the relationship

$$\Delta x(t) = \left[\Delta x^2(t=0) + (\Delta\upsilon\times t)^2\right]^{1/2} \tag{5.7}$$

where $t = 0$ is the time of minimum uncertainty and $\Delta\upsilon = \Delta p/m$. Because Δp is inversely related to $\Delta x(0)$, it is only possible to slow the spreading by an increase in the initial Δx, thereby allowing a smaller $\Delta\upsilon$. Note that spreading is inevitable for the minimum-uncertainty, free-particle wave packet, but not for wave packets in general, which can be shaped by their potentials. Equation (5.7) can be used to give a feeling of the time needed for an appreciable spreading to take place — for $\Delta x(t)$ to increase by a factor of $2^{1/2}$, the product $\Delta\upsilon\times t$ must be equal to $\Delta x(t=0)$. From this equality and the uncertainty relation for RMS deviations, we can express t_s, the time for spreading by ~40 percent, as

$$t_s = \Delta x(0)/\Delta\upsilon = m\Delta x(0)/\Delta p = 2m\Delta x^2(0)/\hbar. \tag{5.8}$$

When a free-particle wave packet is initiated by an ultrashort light pulse so that it has the minimum value of the $\Delta x \Delta p$ uncertainty product, one can calculate the relation between the pulse duration and the wave-packet spatial width. It turns out to be quite simple: $\Delta x = \langle \upsilon \rangle \Delta t$.

The size of $\hbar$, 1.05457×10^{-27} erg s, means that the fuzziness required by the uncertainty principle is imperceptible on the normal scales of size and momentum, but becomes important at atomic scales. For example, if the position of a stationary 100 g apple is initially determined to a small fraction of a wavelength of light, say $\Delta x = 10$ nm, a momentum uncertainty greater than

$$\frac{\hbar/2}{\Delta x} = \frac{0.5 \times 10^{-27} \text{ erg s}}{10^{-6} \text{ cm}} = 5 \times 10^{-22} \text{ g cm/s} \tag{5.9}$$

will be imposed, or $\Delta \upsilon \geq 5 \times 10^{-24}$ cm/s. Under the minimum uncertainty assumption, Eq. (5.8) can be used to calculate that the apple's position uncertainty will spread by ~ 40 percent only after 2×10^{17} s, or six billion years (on the scale the age of the Universe)! On the other hand, an electron with a mass 29 orders of magnitude smaller would spread by 40 percent from an initial 1 Å localization after only 0.2 fs. Clearly, the much greater masses of atomic nuclei make their wave packets spread orders of magnitude slower than the wave packet of the electron (see below).

Atomic-scale spatial localization is critical for motion in real time: not only localization of the wave packets formed in each molecule but of the total spread in position among wave packets formed in the large number of molecules on which the measurement is performed. When both types of spread are small with respect to the relevant extent of the classical region, the conditions for observation of a classical "*single-molecule trajectory*" have been met. The key to achieving localization of the molecular ensemble is generally provided by the naturally well-defined initial equilibrium configuration of the studied molecules before excitation by the initiating (clocking) pulse, as discussed below. For example, most molecules (e.g., iodine) at room temperature or below are in their $n = 0$ vibrational state, the wave function of which, shown in Fig. 5.1, yields a width of the PPD of only 0.08 Å FWHM. This spatial confinement establishes the phase relations between the excited wave functions that produce and define the localized wave packets, so the entire sample of molecules is launched from essentially the same starting point — a *single-molecule trajectory* is observed. The femtosecond time resolution is critical to exploit this localized launch configuration before the dispersion or propagation of the wave packet.

5.1.4 Paradigm case study: Two atoms

Femtosecond spectroscopic observation of the motion of atoms as classical particles, or wave packets, was realized in experiments conducted at Caltech in the 1980s. A prime example is represented by experiments on the wave packet motion of sodium iodine. Figure 5.3 depicts the potential energy for the motion and the experimental observations. The quasi-bound upper potential curve is composed of a covalent inner part and an ionic long-range part created by the crossing of the covalent and ionic curves, as shown. As discussed below, the measurements

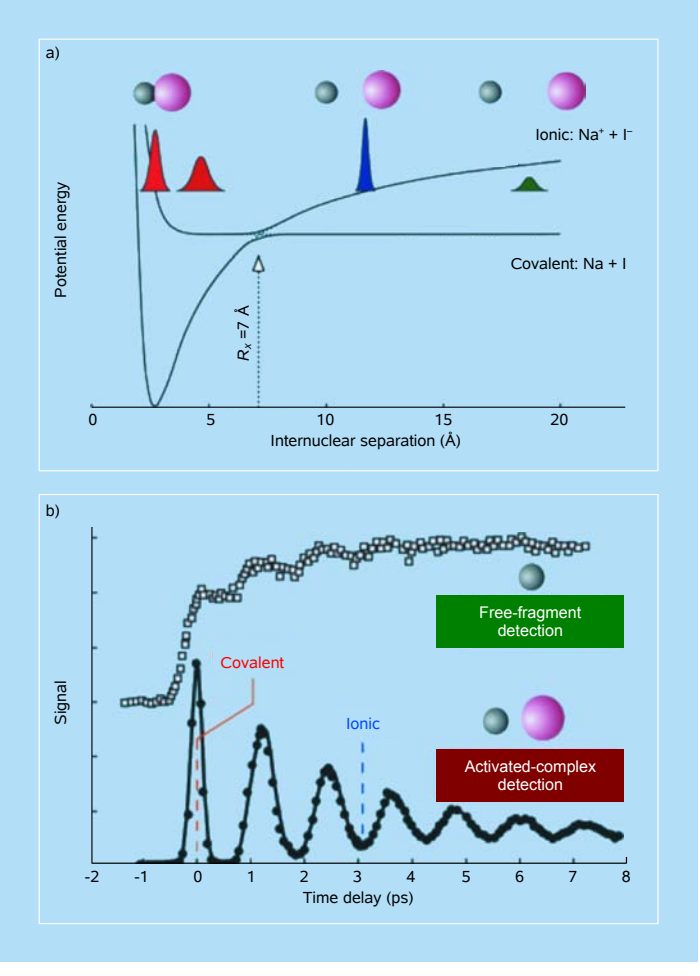


Figure 5.3 Femtosecond spectroscopy of the bond breakage and bond reformation of sodium iodide. Top: The two potential energy functions, the two channels of wave packet evolution (bound and free), and schematic of wave packet motion. Bottom: Time sequence of measured populations of free fragments, formed Na atoms (open squares), and of quasi-bound complex transition-state $(Na^{\cdot\cdot}I)^{\ddagger}$ configurations. The resonance and steps indicate that the motion exhibited is a single-molecule dynamical trajectory, not a kinetic behavior!

show a large fraction of the excited wave packet oscillating in this well, between covalent and ionic bonding, while with each stretch a smaller fraction remains on the covalent surface and separates as free atoms. The role of coherence is critical in preparing the wave packet, and a single-molecule trajectory of the ensemble is assured when an initial pulse synchronizes all molecules near the 2.67 Å bond length of the ground state minimum at time zero.

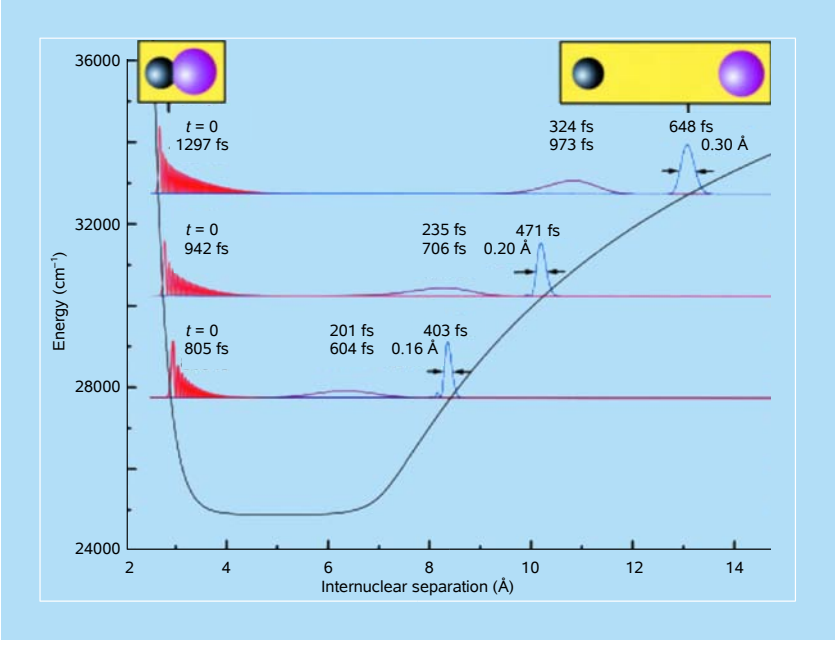


Figure 5.4 Evolution of a wave packet (200 cm^{-1} FWHM bandwidth), corresponding to pulses of 73.6 fs (~75 fs) FWHM, on the theoretical bound potential of sodium iodide (see legend to Fig. 5.2) at three excess energies: 2862, 5362, and 7862 cm^{-1}. The spatial and temporal coordinates are indicated for different snapshots.

This system of two atoms has been studied in great detail, highlighting these and other concepts, and it represents a paradigm case of molecular dynamics with atomic-scale resolution. It is natural therefore to illustrate the properties and time evolution of wave packets as developed above with calculations based on the NaI system. First, we give some examples of wave packet propagation on the theoretical bound state potential, which is modeled on the quasi-bound NaI state. This potential is shown in Fig. 5.4, with an energy axis referenced to the NaI ground-state potential minimum. Also shown are five snapshots, representing one full vibrational cycle, of the PPDs of wave packets formed at three different center energies by coherent superpositions with 200 cm^{-1} FWHM Gaussian population bandwidths. (This bandwidth corresponds to ~75 fs FWHM laser pulses.) Note that

the wave function whose PPD is shown in Fig. 5.2 is one of the $\psi_i(r)$ contributing to the sum in Eq. (5.3) for the superposition at the middle energy, 30,250 cm^{-1} or 5362 cm^{-1} excess vibrational energy. At time zero, the constituent wave functions are all taken as real and positive at the inner turning point of the potential. As time advances, the phase factors $\exp(-iE_i t/\hbar)$ cause the wave packets to move, in each case broadening in the center of the well and sharpening at each turning point.

These wave packets are seen to remain well localized in comparison to the extent of the classically allowed region throughout their first vibrational periods. In fact, at this initial Δx, 10 or more vibrational cycles can be completed without the progressive spreading of the wave packet destroying the usefulness of a classical description of the motion. The femtochemistry experiments referred to above (Fig. 5.3), with pulses of similar duration, have shown this long-time particle-like behavior. In contrast, a calculation for a 1000 cm^{-1} FWHM bandwidth superposition, or ~15 fs pulse, at 5362 cm^{-1} excess energy, shown in Fig. 5.5, is initially much more localized [smaller $\Delta x(0)$] but spreads much more quickly (larger Δp). Now the spreading of the wave packet in even one vibrational period is substantial.

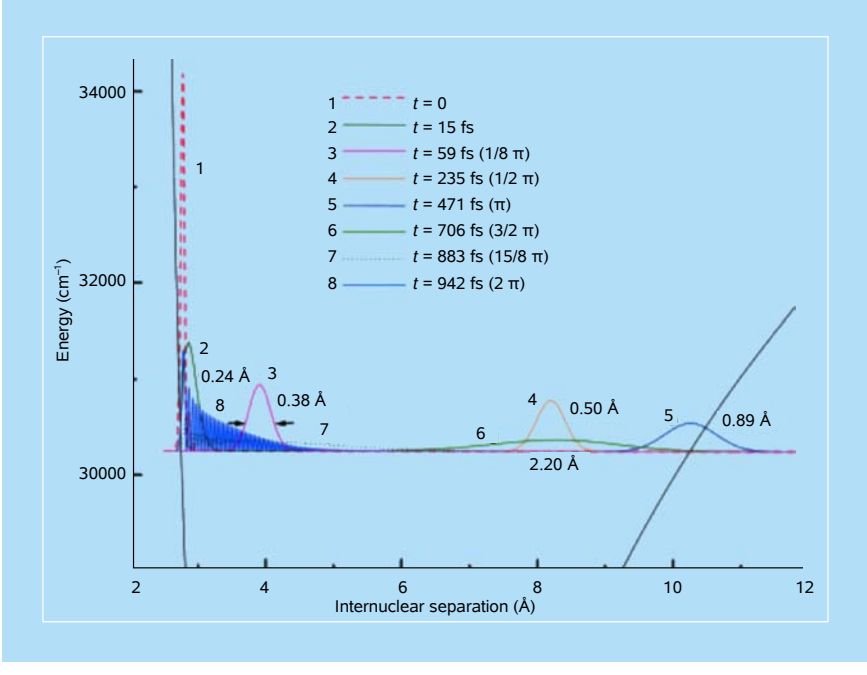


Figure 5.5 Evolution of a wave packet (1000 cm^{-1} FWHM bandwidth), corresponding to pulses of 14.7 fs (~15 fs) FWHM, on the theoretical bound potential of sodium iodide (see legend to Fig. 5.2) at 5362 cm^{-1} excess energy. Vibrational phase angles ($2\pi t/\tau_v$, where τ_v is the vibrational period) are given in parentheses and the FWHM for wave packets 2–6 are indicated.

As shown in Fig. 5.3, the real NaI system involves two crossing potentials, one of which allows the Na and I to separate totally as free atoms. If we use the relative motion of Na and I at a constant potential as an example of free particle motion, for an average (kinetic) energy of 5362 cm^{-1}, the mean velocity of separation $\langle \upsilon \rangle$ is 25.7 Å/ps or 2.57×10^5 cm/s ($E_{kin} = \frac{1}{2}\mu \langle \upsilon \rangle^2$, where $\mu = 19.46$ amu $= 3.232 \times 10^{-23}$ g, the reduced mass of Na and I). A 1000 cm^{-1} FWHM Gaussian energy bandwidth corresponds to a 14.72 fs FWHM Gaussian pulse. Such a pulse will create a minimum uncertainty wave packet of 0.378 Å FWHM. Restating the minimum uncertainty relation in terms of the FWHM as $\Delta x^{FW} \times \Delta p^{FW} \geq 8 \ln 2 \times \hbar/2 = 2.7726\hbar$, we can calculate a momentum spread Δp^{FW} of 7.73×10^{-19} g cm/s or $\Delta \upsilon^{FW} = \Delta p^{FW}/\mu = 2.39$ Å/ps.

The growth with time of the wave-packet spatial width can then be calculated according to Eq. (5.7). Using FWHM values for the present case leads to

$$\Delta x^{FW}(t) = \left[(0.378 \text{ Å})^2 + (2.39 \text{ Å/ps} \times t)^2 \right]^{1/2}. \tag{5.10}$$

This gives a width of 2.4 Å after 1 ps. If the initial wave packet is twice as broad (0.756 Å), $\Delta \upsilon^{FW}$ will be half as large and the wave packet will spread more slowly, to only 1.4 Å after 1 ps. Twice the initial wave packet width will be produced for the same $\langle \upsilon \rangle$ by a pulse twice as long, or for the same pulse width by an average energy four times as high. Similarly, Eq. (5.8) can be used to calculate the time for the packet to spread by 40 percent: if $\Delta x^{FW}(0) = 1$ Å, that is, $\Delta x(0) = (1/2.3548)$ Å, then $t_s = 1.1$ ps; the wave packet in that time travels over 28 Å. If $\Delta x(0)$ doubles, the wave packet takes 4.4 ps, four times as long, to spread by the same factor.

Figure 5.6 shows the evolution of an unbound NaI wave packet with energy spread corresponding to a 14.72 fs laser pulse. The wave packet is formed on the steep repulsive potential wall of the bound well but evolves after the initial repulsion on the unbound, constant-energy surface. The initial packet is quite narrow and virtually identical to the bound packet of Fig. 5.5; that is, its form depends only on the local potential surface. While the potential is changing sharply with distance ($t < 30$ fs), the above calculations concerning width and spreading of the wave packet for a constant potential energy are not applicable. However, after the packet enters the force-free region, it reproduces quite well the behavior predicted above for a free particle wave packet of minimum uncertainty at $t = 30$ fs.

These calculations for NaI show how, despite the important role of quantum uncertainty and wave mechanical behavior, it is entirely appropriate to treat the nuclear motion of reacting molecular systems as "particle-like," and to achieve spatial localization on the femtosecond time scale, the result of the two uncertainty relationships:

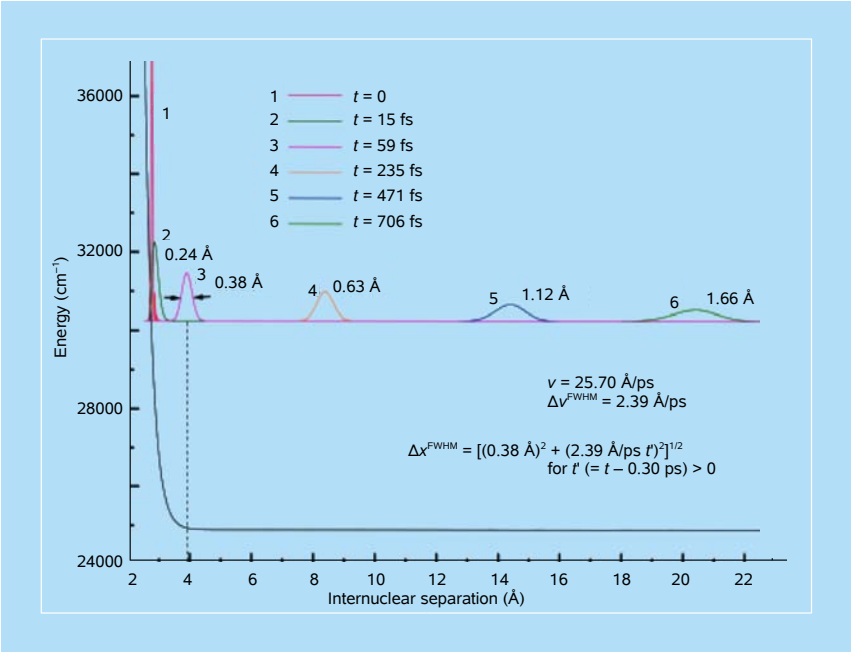


Figure 5.6 Evolution of a wave packet (1000 cm^{-1} FWHM bandwidth), corresponding to pulses of 14.7 fs (~15 fs) FWHM, for unbound sodium iodide at 5362 cm^{-1} excess energy.

$$\Delta x \Delta p \geq \frac{\hbar}{2} \quad \text{and} \quad \Delta t \Delta E \geq \frac{\hbar}{2}, \tag{5.11}$$

which imply that the shorter the time, the better the localization (without quantum violation), provided coherence is properly introduced experimentally and understood conceptually. Historically, these issues and concerns regarding the uncertainty principle were raised and addressed in the context of dynamics of matter and similarly in the development of lasers (see Refs. 1 and 7 in Further Reading).

5.2 From Stop-Motion Photography to Ultrafast Imaging

5.2.1 High-speed shutters

As a starting point, we may consider the historic experiments in stop-motion photography performed more than a hundred years ago by photographer Eadweard Muybridge. Some of those experiments, carried out under the patronage of Leland Stanford at Stanford's Palo Alto farm, are commemorated by a plaque on the campus of Stanford University (Fig. 5.7). Muybridge's work, which began in 1872, was inspired by a debate over the question of whether all four hooves of a trotting

Figure 5.7 Tablet erected on the Stanford University campus in 1929, commemorating "motion picture research" of Eadweard Muybridge at the Palo Alto farm of Leland Stanford in 1878 and 1879.

horse are simultaneously out of contact with the ground at any point in its stride. Muybridge believed he could answer this question by taking photographs with a camera shutter and film capable of capturing reasonably sharp images of a horse trotting at high speed. (The trot is a mode of locomotion, or gait, in which diagonal legs move together in pairs — it is a smooth, steady motion, in contrast to the more energetic bounding motion of the gallop, which is the gait used to attain ultimate speed for short distances. The details of each gait were eventually elucidated by

Muybridge, but his written account clearly describes the seminal role of the trot in inciting his interest.)

We can estimate the duration of the shutter opening, Δt, that Muybridge needed for his camera from consideration of the necessary spatial resolution and the velocity υ of the horse. For a clearly defined image of a horse's legs, a resolution (Δx) of 1 cm is reasonable; that is, 1 cm is small compared to the relevant dimensions of the problem (the dimensions of the leg and its displacement during the course of a stride). Taking $\upsilon \approx 10$ m/s for the velocity of a horse (the legs will, in fact, at times be moving several times faster), and using the relation $\Delta x = \upsilon \Delta t$, leads to

$$\Delta t \approx \frac{1 \text{ cm}}{10 \text{ m/s}} = 10^{-3} \text{ s} = 1 \text{ ms}. \tag{5.12}$$

Indeed, Muybridge was able to achieve the necessary exposure times to capture an image of a trotting horse with all four feet in the air. Subsequent to this initial success, he devoted many years to the photographic study of animals and humans in motion, for some time in Palo Alto and later at the University of Pennsylvania.

In these studies, Muybridge sought not only to provide the isolated images required to answer questions such as that which first attracted his interest, but also to document the entire sequence of an animal's leg motions during its stride. To establish the required absolute timing of the photographs, he initially set up a row of equally spaced cameras along a track at the Palo Alto farm. The shutter of each camera was activated by a thread stretched across the track in front of the camera. Thus, a horse running down the track at speed υ recorded a series of photographs, and the point in time associated with the ith photo could be calculated as $\sim d_i/\upsilon$, where d_i was the distance from the starting gate to the ith camera. The separation in time between frames was $\tau = \Delta d/\upsilon$, where $\Delta d = d_{i+1} - d_i$, so the number of frames per second was $\upsilon/\Delta d$. Although the absolute timing of the frames of the series was imperfect, tied as it was to the velocity of the horse from camera to camera, the images nevertheless permitted detailed analysis of the motion. The imprecision of the chronology of images obtained in this manner was subject to some criticism, and in his later studies, Muybridge used cameras with shutters triggered sequentially by a clockwork mechanism to obtain photographs regularly spaced in time, such as those shown in Fig. 5.8.

Etienne-Jules Marey, professor at the Collège de France and a contemporary of Muybridge, invented "chronophotography," a reference to the regular timing of a sequence of images recorded by a *single* camera using a rotating slotted-disk shutter (the analogue of shutter-stroboscopy, the concept of which was developed as early as 1832; the related flash-stroboscopy is discussed in what follows). The recording was made on either a single photographic plate or a film strip, the precursor to modern cinematography. Marey, like Muybridge, focused on investigations of

humans and animals in motion, such as that involving "cat righting reflex," the innate ability of a cat to orient itself as it falls and land on its feet, often uninjured.

Muybridge also developed an apparatus to project sequences of images obtained in his experiments to give "the appearance of motion" in demonstrations representing an early stage in the development of motion picture technology. His device, which he called a zoopraxiscope, made use of a projecting lamp and counter-rotating image and shutter disks to produce a rapid succession of apparently stationary bright images and periods of intervening darkness. With image frames appearing at a rate greater than about 20 per second, the viewer's perception of each bright image persisted across the dark intervals, merging with its successor to form the desired continuous, but moving, picture. A simpler device for viewing the sequences, which was well known in Muybridge's time, is the zoetrope (from the Greek *zoe*, "life," and *tropos*, "turning"). It consists of a cylinder spinning on its axis with a series of slits equally spaced around its circumference. Through the slits, photos on the opposite inner surface are briefly glimpsed in rapid sequence, giving an impression of continuous motion.

The authors of this book attest to the reality of Muybridge's device. One of us (AHZ) when giving (in March 1991) the Faraday Discourse at the Royal Institution

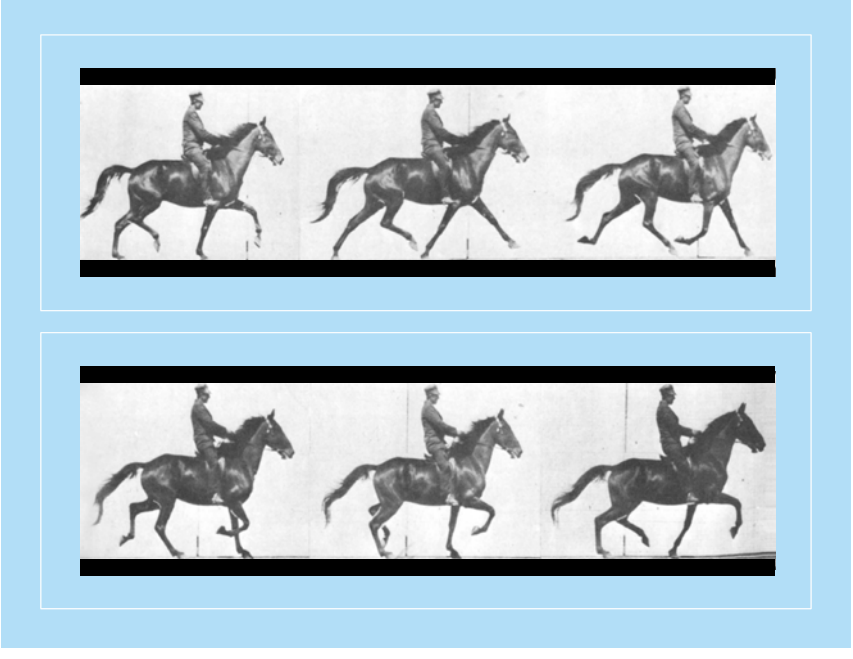


Figure 5.8 A series of photographs by Eadweard Muybridge of a trotting horse, taken during studies at the University of Pennsylvania (1884–1885). The time interval between photos is regular, 0.052 s. Note that in the third photograph (top right), all four hooves are simultaneously off the ground in the trot motion.

(directed at the time by JMT) demonstrated the concept of freezing motion by displaying "horses in motion" using Muybridge's device, which was in the archives of the Royal Institution, and then compared its time scale (20 frames per second; slowing down the motion by 50 times) with the vastly different femtosecond time resolution (slowing down the motion by nearly 10^{14}) needed to record "atoms in motion," the subject of the Discourse.[a]

5.2.2 Stroboscopy

An alternative approach to the study of rapid motion, which has also proved capable of reaching much shorter time scales than possible with fast shutters, is the use of short light flashes, which make an object moving in the dark visible to a detector (observer's eye or photographic plate, for example) only during the light pulse. Thus, the pulse duration Δt plays the same role as the opening of a camera shutter and can be thought of in just the same way. An instrument that provides a series of short light pulses is a stroboscope (*strobos* from the Greek word for "whirling" and *scope* from the Greek for "look at," for its original use in viewing rotating objects). Combined with a camera with an open shutter and with an appropriately chosen Δt for the light pulses, a stroboscope can produce a well-resolved image of an object as fast as a moving bullet.

In the mid-19th century, spark photography had been demonstrated for "stopping" rapid motions. The development of stroboscopic photography in the mid-20th century was greatly advanced by Harold Edgerton, professor at MIT and cofounder of EG&G electronics, through the development of electronic flash equipment capable of producing reliable, repetitive, and microsecond-short flashes of light. Edgerton and EG&G also developed camera shutters based on optical principles with no moving parts that are far faster than any conventional mechanical shutter.

An example of the use of a stroboscope is shown in Fig. 5.9, a precisely timed sequence of images of a falling apple. With the apple's velocity limited to $v \leq 5$ m/s and a value of Δx of ~1 mm for a sharp image, Δt of the flash must be $\sim (1\text{ mm})/(5\text{ m/s}) = 2\times10^{-4}$ s or less, well within the stroboscope's range. An absolute time axis is established here by electronic timing of the flashes. The picture shows the acceleration due to gravity, which can be quantified by analyzing the successive positions in which the light illuminates the apple. According to the law of uniformly accelerated motion $x = \frac{1}{2}at^2$, for the position x of an object at rest

[a]On 13 March, 1882 Muybridge was invited to give a Discourse at the Royal Institution in London to illustrate his zoopraxiscope, which was then donated to the Institution and was the one used by AHZ in March 1991. In Muybridge's audience were the Prince of Wales, the Prime Minister (W. E. Gladstone), T. H. Huxley, Sir James Dewar, Lord Tennyson, and John Tyndall, Faraday's successor — would political personalities of today attend such a scientific discourse?

at $t = 0$ and subjected to acceleration a. Therefore, flashes equally spaced in time record images of uniformly increasing separation (up until impact and rebound from the table); the slope of the plot of image separation versus time is equal to $g\tau$, where g is the acceleration of gravity (~ 9.8 m/s^2) and τ is the spacing of the flashes. When an apple height of 10 cm is assumed in Fig. 5.9, one finds that τ is ~ 0.04 s; that is, the strobe flash rate was ~ 25 per second. [This is the modern equivalent of Galileo's experiments carried out in 1604 and beyond. He used water clocks (or pendulums) to record the distances traveled by balls running down a ramp in equal time intervals in order to derive his famous projectile equations. His original apparatus is presented in a museum in Florence.]

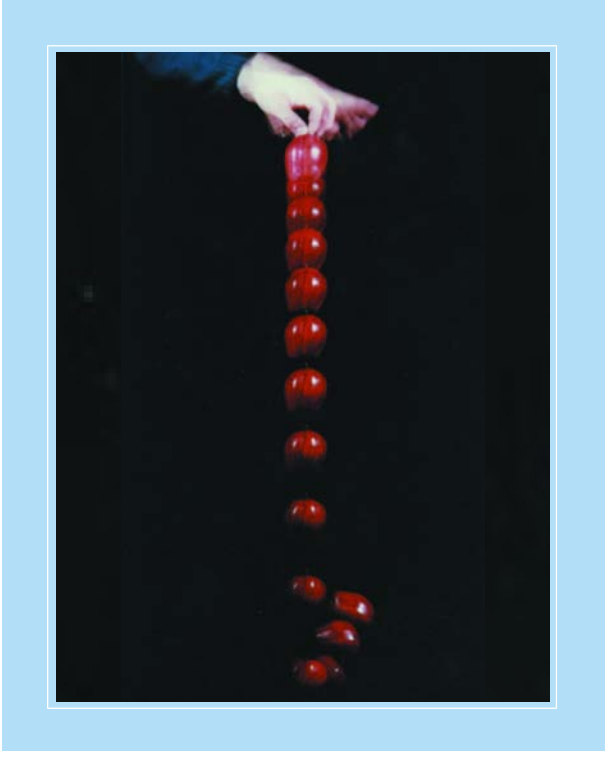


Figure 5.9 A falling apple photographed by stroboscopic illumination at intervals of ~1/25 s. The acceleration due to gravity is clear.

5.2.3 Ultrafast techniques

If the above ideas can be carried over in a straightforward manner to the study of atoms in motion, then the requirements for ultrafast recording are easily determined. For a molecular structure in which atomic motions of a few angstroms

typically characterize the change, a detailed mapping of the process will require a spatial resolution Δx of less than 1 Å, more than eight orders of magnitude smaller than that needed by Muybridge. Therefore, the Δt required to observe with high definition molecular transformations in which atoms move at speeds of the order of 1000 m/s is (0.1 Å)/(1000 m/s) = 10^{-14} s = 10 fs. While this time scale has been recognized theoretically as the time scale for molecular transformations since the 1930s, it became possible to study directly the transition states for the first time in the 1980s with the development of femtochemistry, as discussed above. Such minute times and distances mean that molecular-scale phenomena are governed by quantum mechanical principles.

Flashing a molecule (or specimen, in general) with a femtosecond laser pulse can be compared to the effect of a stroboscope flash or the opening of a camera shutter. Thus, a pulse from a femtosecond laser, combined with an appropriate detector, can produce a well-resolved "image" of a molecule (or specimen) as it passes through a specific configuration in a process of rearrangement, as Muybridge caught the horse with all four feet in the air. However, because the detection step in femtochemistry utilizes optical spectroscopy, mass spectrometry, or photoelectron spectroscopy, one analyzes the measured spectroscopic signal in order to obtain information about the positional change of atoms. In real-space imaging and diffraction, the subject of this book, the actual structure can be obtained because the strobing pulses are composed of ultrafast electrons which, unlike light, have a wavelength smaller than interatomic distances. In general, the pulse (light or electron) that produces the spectrum or the image is called a *probe* pulse because it is used to probe the system just as a shutter opening and a stroboscope flash probed the positions of the horse and apple in Figs. 5.8 and 5.9, respectively. The use of such pulses to "stop the motion" of atoms and obtain instantaneous molecular or materials structures is the cornerstone of ultrafast imaging, and it is entirely different in principle from conventional fast techniques, as discussed below. Structures determined at different stages of a process can be treated like the frames of a motion picture, allowing the motion of the atoms to be visualized; the number of frames in a molecular movie (Fig. 5.10) could then be as high as 10^{14} per second, which implies 10-femtosecond intervals (femto, the prefix meaning 10^{-15}, is from *femton*, the Scandinavian word for "fifteen"). Imaging with electron probing pulses is detailed in Chap. 6.

Probing is not the whole story. For the entire course of any motion to be recorded, that motion must be initiated so that it takes place in the time span accessible to a sequence of probe snapshots. In photographing the horse and apple, the processes are initiated by opening a starting gate for the horse and by releasing the apple, and the respective probing sequences are arranged to coincide closely in time with those actions. For ultrafast recording, the analogous operation is realized by launching the molecule (or material) on its path using a

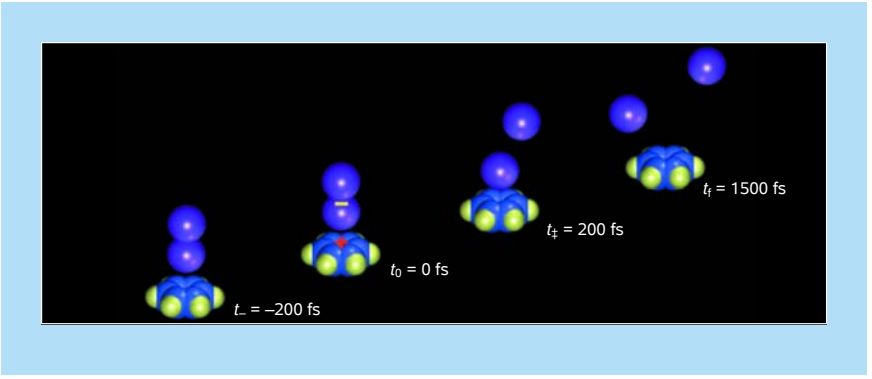


Figure 5.10 Atomic motions for a reaction in progress involving two molecules (bimolecular). The diatomic iodine molecule (I_2, top) is split by exchange of an electron with the ring molecule benzene (C_6H_6).

femtosecond initiation pulse (the clocking or *pump* pulse) passing through the sample. This establishes a temporal reference point (time zero) for the changes that occur in the motion. The timing relative to the probe pulses is accomplished by generating the clocking and probe pulses from a common source and sending either the clocking or the probe along an adjustable optical path to the sample, as shown in Fig. 5.11 for clocking and probing light pulses. The difference between clocking and probe path lengths divided by the constant speed of light of about 300,000 km/s (actually 299,792 km/s) precisely fixes each probe spectrum on the time axis established by the clocking pulse. When electron and light pulses are introduced in imaging, we must consider the difference in speed and the fermionic/bosonic character (see Chap. 6 and below).

Such use of optical path differences to measure transient phenomena dates back at least to 1899, when Abraham and Lemoine[1] reported their measurement in France of the Kerr response of carbon disulfide (Fig. 5.12). They used the breakdown of air at a spark gap to create *simultaneously* ($t = 0$) a probe light pulse (spark) and discharge a parallel plate capacitor immersed in carbon disulfide. Before the discharge of the capacitor, the carbon disulfide molecules between the plates retained a net alignment with the applied electric field, resulting in optical birefringence, a difference in refractive index for light polarized parallel or perpendicular to the plates (Kerr effect). Upon discharge, the spark-produced probe light monitored the Kerr effect as a function of time by traversing a variable-length optical delay before passing between the plates of the capacitor. Polarization analysis showed that the substantial Kerr effect to which the probe light was subjected at the shortest measurable delays was progressively reduced and ultimately disappeared as the delay path length was increased. From this observation, a half-life of 2.5 ns, corresponding to a path length of 75 cm, was

determined for the combined electrical and Kerr (alignment) response time of the system. We now know that the Kerr response due to reorientation of the molecules occurs in ~ 2 ps ($1\ \text{ps} = 10^{-12}$ s), with a femtosecond component associated with electronic polarization effects.

A fundamental difference between ultrafast techniques and the horse and apple analogies is that, in typical experiments, one probes millions to trillions of atoms or molecules for each initiation pulse, or one repeats an experiment many times, to provide a signal strong enough for adequate analysis. A comparable situation would arise in stroboscopy of the apple if capture of a distinct photographic image could only be accomplished by using many different apples or repeated exposures. It is clear that success in such a case would require (i) precise *synchronization* of the strobe (probe) pulse sequence with the release of the apple to less than or about the strobe pulse duration, for optimum resolution, and (ii) a precisely defined launching configuration of each apple, to a fraction of an apple diameter.

By the same reasoning, to synchronize the motion of many independent *atoms or molecules* so that all have reached a similar point in the course of their structural evolution when the probe pulse arrives to capture the desired impression,

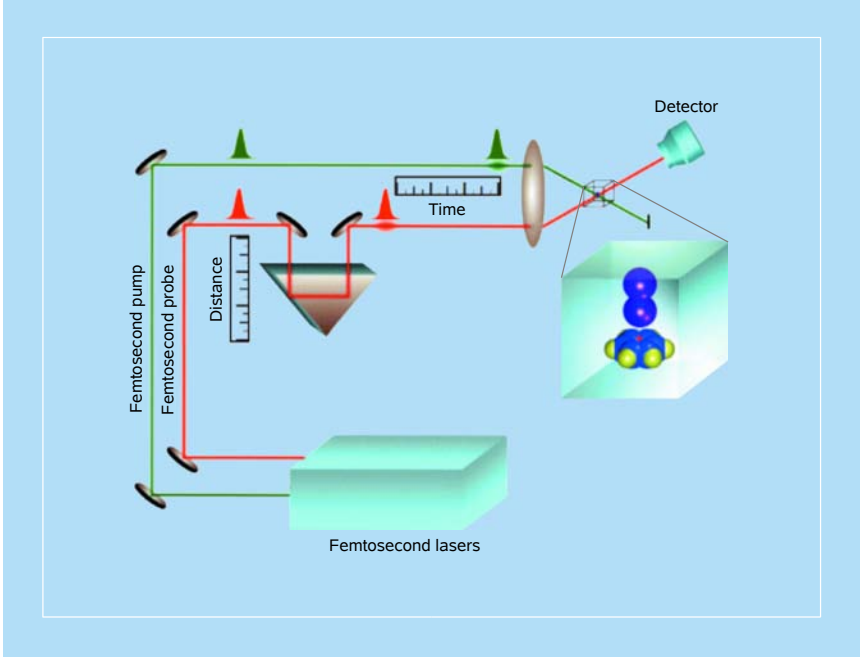


Figure 5.11 The concept of femtosecond all-optical (pump-probe) experiments. After the probe pulse has been delayed by diversion through a variable length optical path, femtosecond pump and probe pulses are focused into a volume containing the specimen to be studied. The detector responds to the probe pulse by any of a variety of schemes.

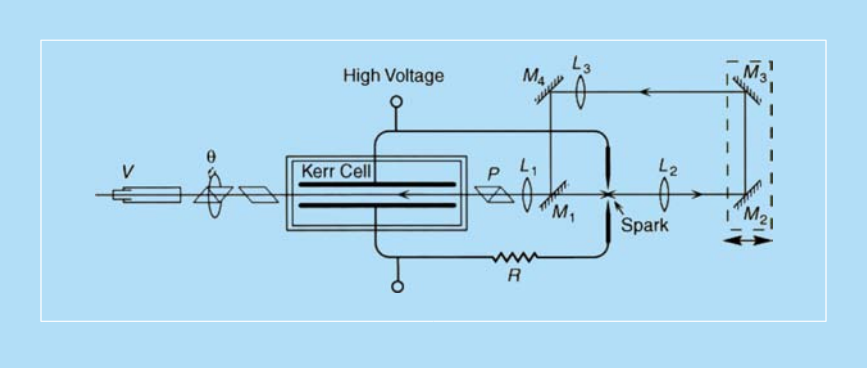


Figure 5.12 Kerr-cell response measurement of Abraham and Lemoine (1899). The high voltage applied across the plates of the capacitor induces a Kerr effect in the carbon disulfide. The breakdown of air at the spark gap discharges the capacitor and creates a pulse of light, which follows the path indicated (reflecting from mirrors M_2, M_3, M_4, and M_1; L_2, L_3, and L_1 are collimating lenses) to probe the Kerr effect. The time dependence of the effect is determined by moving mirrors M_2 and M_3 as indicated, to vary the time delay for passage of the probe pulse through the Kerr cell. The light is polarized before the cell by the polarizer P, and the Kerr effect is quantified by adjusting the rotation angle θ of an analyzing polarizer to measure the polarization properties of the light after passage through the Kerr cell, as observed at the viewing telescope *V*.

the relative timing of clocking and probe pulses must be of femtosecond precision, and the launch configuration must be defined to sub-angstrom resolution. It is only by means of such synchronization that the signals from many atoms or molecules may be added together without hopelessly blurring the molecular structure derived from the measurement. The use of pulses generated from the same source and an optical path delay as described above provides the required high degree of timing precision. As mentioned before, an optical path accuracy of 1 μm corresponds to absolute timing of the molecular snapshots of 3.3 fs (see Fig. 5.11).

Of equal importance is the required definition of the launch configuration. This definition is naturally realized because the femtosecond clocking pulse snaps all members of the atomic or molecular ensemble from their ground states, which have a "single" well-defined structure. Moreover, on the femtosecond time scale, moving atoms are coherent or particle-like in their trajectories because of localization in each atom or molecule and in the ensemble, as discussed above.

5.2.4 Ultrafast lasers

The laser, light amplification by stimulated emission of radiation, operates on two concepts: amplification and feedback. Thus, one must have a gain medium (e.g., Ti:sapphire crystal) for amplification and a cavity (e.g., two mirrors) for the feedback. The excited Ti:sapphire crystal amplifies the light, and with the feedback of the cavity, the bouncing of light back and forth, an intense laser beam is

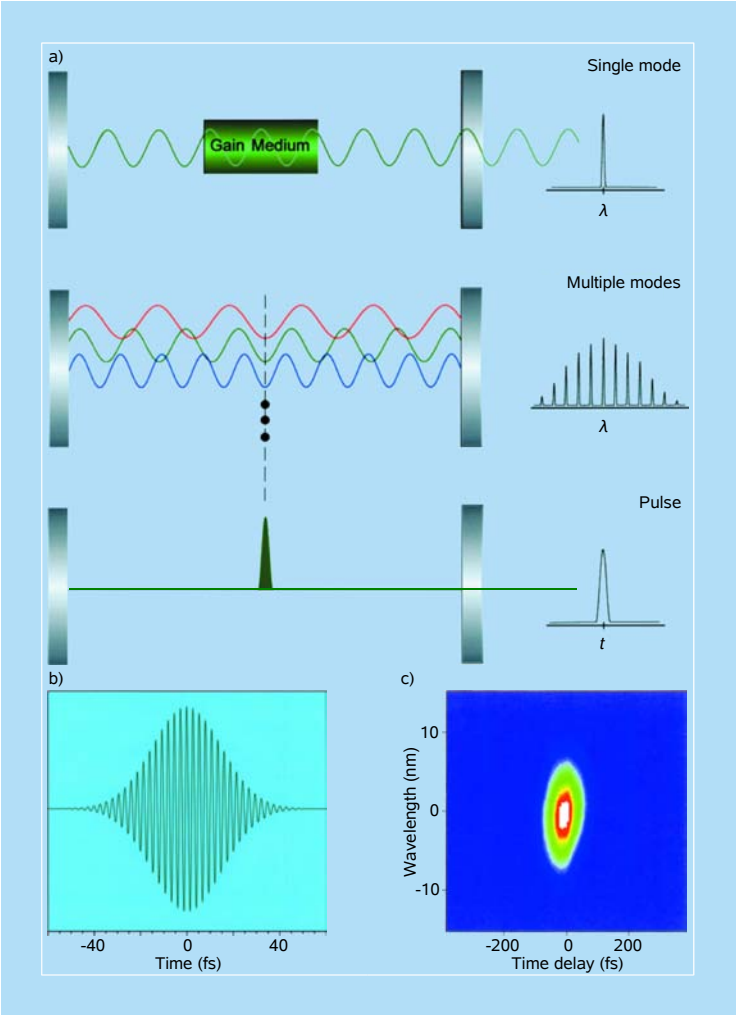


Figure 5.13 The basic concept of ultrashort laser pulse generation by the process of mode-locking. The upper panel shows (top) continuous-wave lasing in single-mode operation; (middle) the superposition of several modes, which, if locked together in large numbers as indicated by the three vertical dots, gives (bottom) pulsed operation such as that depicted here. The lower panels show (left) the field of a femtosecond pulse at 800 nm [Eq. (5.13), $\sigma = 15$ fs], together with (right) an experimental display of a pulse in the wavelength (nm)–time (fs) domain.

produced from one of the mirrors, which is partially transmitting (Fig. 5.13). The cavity (of length L) has modes, or light of certain wavelengths, which are separated in frequency by $c/(2L)$. If one mode is isolated, then lasing occurs in a narrow-band frequency in what is referred to as CW or continuous-wave operation. For generation of an ultrashort pulse, light of a large number (possibly millions) of modes (wavelengths) is emitted together coherently; the modes are said to be locked

together. Mode-locking is a superposition of light waves in phase (Fig. 5.13), and as such it is similar to the creation of wave packets of matter waves. In Fig. 5.13 we also display the shape of a femtosecond pulse at 800 nm; the variation of the electric field at a point may be expressed as

$$E(t) = E_0 \exp\left(-\frac{t^2}{2\sigma^2}\right) Re\left[\exp(i\omega_0 t)\right], \tag{5.13}$$

a packet with a carrier frequency $\omega_0/2\pi$ and RMS width σ of the field amplitude envelope.

Three essential elements are contained in the cavity of a femtosecond laser: a gain medium, a mode-locking element, and an element that compensates for group velocity dispersion. The gain medium has a broad spectrum to ensure a large number of modes sufficient for femtosecond pulse generation; a 10 fs pulse requires a spectral width of about 100 nm at an 800 nm center wavelength. The group-velocity-dispersion element usually consists of a combination of prisms arranged so that the broad range of spectral components in the pulse can all stay in step while circulating in the cavity. There are two forms of mode-locking: active and passive; the idea is to provide a time-dependent modulation of loss and/or gain in the cavity. Active mode-locking uses an externally driven modulator, whereas in passive mode-locking an element that lets the pulse itself vary the gain/loss is included in the cavity. One method of passive mode-locking uses an organic molecule as a "saturable absorber" to enhance the propagation of a pulse over that of continuous lasing in the cavity. When the light is so intense that the saturable absorber's rate of absorption exceeds its rate of ground-state population recovery, it becomes temporarily transparent — it is saturated. This transparency occurs at the peak intensity levels of pulsed operation, allowing light to pass and circulate in the cavity to ultimately produce ultrashort pulses; the lower light intensities achievable in continuous lasing are blocked.

For Ti:sapphire lasers, the pulsed operation is achieved using self-mode-locking. The lasing material has a refractive index that changes depending on the light intensity. Because the laser beam in the cavity has a transverse spatial distribution with the largest intensity in the middle (Gaussian profile), the speed of light propagation for the intense part decreases relative to that of the lower intensity part, resulting in a lens-type focusing behavior (optical Kerr effect). This effect is pronounced only under the high intensities possible with short pulses. With a narrow aperture in the cavity where the pulsed beam focuses, much of the beam from any lower intensity mode of operation, which does not focus, can be blocked, while the pulses pass unimpeded. Thus, the favored pulsed lasing will be the natural mode of operation.

As discussed earlier, to carry out femtosecond experiments, we need both an initial (clocking) and a probe pulse. It is useful here to consider first experiments using optical pulses; in Chap. 6, we shall focus on those involving ultrafast electron pulses. The pulses are usually created from the output of a single femtosecond laser by a variety of nonlinear optical mixing processes chosen to produce the specific wavelengths of light appropriate for the molecule or system under study. The generated femtosecond pump and probe pulses are characterized by measuring the intensity, duration, spectral bandwidth, and polarization of the fields. The clocking pulse sets the $t = 0$ by initiating the process under study, and the probe pulse is usually absorbed; the absorption varies as a function of the probe delay time in response to the spectral changes of atoms or molecules with time. The amount of probe absorption is reflected in a signal acquired by one of a variety of detection schemes. Two of these are laser-induced fluorescence and mass spectrometry.

In laser-induced fluorescence, the absorption of probe light energy creates an electronically excited species which then emits light characteristic of the species; that is, it fluoresces. The fluorescence intensity is measured at each delay time to determine the amount of absorption. For example, for NaI, the absorption spectra of various configurations of the transition-state complex Na···I are very different from those of free Na and I atoms. Thus, by tuning the probe to the wavelength absorbed by Na···I with Na and I close together, we can observe a periodic fall and rise of the fluorescence signal with delay as the molecules vibrate back and forth in the quasi-bound well, passing repeatedly through the strongly absorbing configuration. Since some of the molecules break apart with every bounce, the peak of every return is also smaller than the last. On the other hand, when the probe is tuned to a wavelength absorbed by free Na atoms, the fluorescence increases in stepwise fashion as, at each bounce, a new batch of free atoms is added to the existing pool; for the entire process all molecules move coherently, as evidenced by the steps and resonance shown in Fig. 5.3.

In mass spectrometric detection, the probe creates a population of ions which can be sorted by mass. Thus, changes in mass as well as changes in absorption spectra can be monitored as bonds are broken and formed. There are various methods of detection such as photoelectron spectroscopy, fragment kinetic energy detection, and nonlinear coherent anti-Stokes Raman scattering (CARS), as well as others. Applications of these spectroscopic methods have been made in different phases of matter (gases, liquids, and solids) and for small as well as complex molecules including biological systems. Methods for ultrafast imaging will be discussed in Chap. 6.

Without the technological advances made since the invention of the laser in 1960, the generation of ultrafast laser pulses would not have been possible. The seminal work by C. Townes, N. Basov, and A. Prokhorov on masers and C. Townes, A. Schawlow, N. Bloembergen, T. Maiman, and others on lasers led to

the development of Q-switching (R. Hellwarth), mode locking (A. De Maria), dye lasers (P. Sorokin and F. Schäfer), femtosecond dye lasers (C. Shank, E. Ippen, and R. Fork), and femtosecond solid-state lasers and amplifiers (W. Sibbett and G. Mourou). These became cornerstones for achieving pulses of a few optical cycles long. Together with other advances, including the development of white-light continuum, sensitive detectors and computerized data handling, the pulses have provided the required sources in many areas of study (see Ref. 1 in Further Reading and Ref. 2 in References section).

5.3 Single-Electron Imaging

At this point, we distinguish two general time-resolved imaging modes: (i) stroboscopy of reversible or repetitive processes; and (ii) single-shots of irreversible processes. While the formation of a real-space image clearly requires at least about one electron per image pixel, we have seen that by the use of synchronously triggered repetitive processes, an image may be built up using even a single electron per exposure. We call this "single-electron imaging". Before considering some key principles of imaging with single electrons, it is instructive to emphasize the issues involved. In ultrafast imaging we are dealing with timed, single-electron packets which possess well-defined coherence time and length, and these characteristics are distinct from those of continuous electron beams used in time-integrated diffraction and imaging. Moreover, because electrons are fermions, we have to consider consequences of the Pauli exclusion principle, and behavior in the single-shot regime. Finally, imaging with ultrashort pulses is an entirely different methodology from those of video recording (millisecond or longer) or fast (μs to ns) electronic methods. We address these issues in what follows.

5.3.1 Coherence of ultrafast packets

For light pulses of frequency width $\Delta\nu$, the temporal and spatial coherence mentioned in Chap. 2 are expressed by the simple relation:

$$\tau_c \sim \Delta\nu^{-1} \quad \text{and} \quad l_c = c\tau_c \sim c\Delta\nu^{-1} \ (l_c \equiv \lambda^2/\Delta\lambda), \tag{5.14}$$

where l_c is the corresponding longitudinal coherence and λ is the mean wavelength of the pulse. [For electrons, the relations are only slightly different: $\tau_c = l_c/\upsilon$ and $l_c = \lambda(E/2\Delta E)$.] In order to measure the temporal coherence of a pulse, we simply divide the beam into two in a Michelson interferometer, delay one of the resulting two beams relative to the other, and then recombine both to observe "fringes" of their interference (see Fig. 5.14). If the path difference in the interferometer is sufficiently small, fringes are formed, and they are manifestations of the temporal coherence. In other words, the time delay Δt_d gives the time over which the fields

of the waves are correlated, and this correlation is over $\Delta\nu$ of the pulse, the degree of monochromaticity. Similarly, the longitudinal coherence length tells us the distance over which the fringes and correlations are possible. It follows that the larger $\Delta\nu$, the shorter τ_c and l_c. For a laser radiation with $\Delta\nu \sim 1$ kHz, $\tau_c \sim 10^{-3}$ s and $l_c \sim 300$ km. For a 100 fs (transform-limited) pulse, τ_c is the same as the pulse width and $l_c \sim 30$ μm. For nanosecond pulses, τ_c is much shorter than the pulse width, as discussed below. Thus, as mentioned before, the incoherent light from the Sun (very broad spectral range) has a very short τ_c of fs or less depending on the $\Delta\nu$ filtered.[3–5]

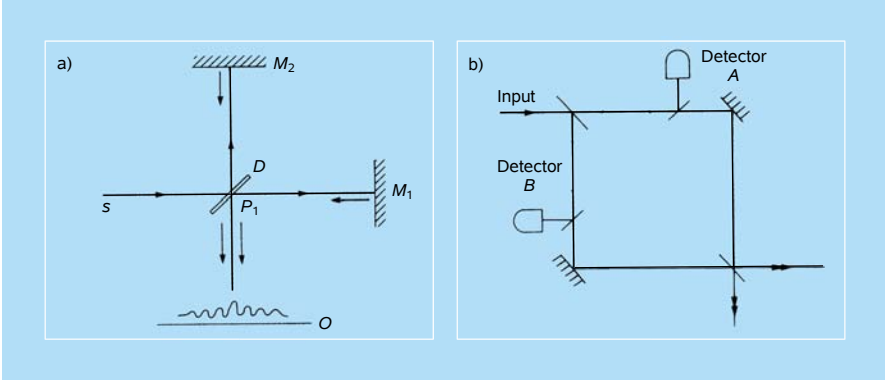


Figure 5.14 Interferometry for temporal-coherence and path-length measurements. (a) Temporal coherence illustrated by means of an interference experiment using a Michelson interferometer; s = source, D = beam divider, $M_{1,2}$ = mirrors, O = plane of observation. For simplicity, a compensating plate and a collimating lens system are not shown. (b) Mach–Zehnder interferometer allows the path chosen by an incident photon to be identified, as detector A or detector B registers the output. Adapted from Ref. 3.

This temporal and longitudinal coherence length of the pulse should not be confused with the transverse spatial coherence area and length (or width; see Chap. 2). The latter features are related to the extent of maintaining the phase fronts over a distance of travel, and observing fringes in a Young's type arrangement (Fig. 2.3). Here, the lateral extent of the source for generating the pulses is a key to the visibility of the fringes. The emitting area of the source, when considered to consist of a set of statistically independent atomic elements, is termed "ideally incoherent". The degree of coherence of the particles it produces increases with distance from the source. For optical pulses, the issue becomes whether or not they are generated from the same laser or different ones, and the transverse-mode characteristics of the beams. For electrons, as discussed below, we use femtosecond optical pulses to generate the electron packets, using the photoelectric effect. The effective size of the disc on the photocathode (for electron packet generation) is

therefore one parameter we must consider; the others are the distance over which the packet travels, the number of electrons in each pulse, and the coherence of the packet and pulse.

The divergence of waves from a source of finite size (say, diameter w; see Fig. 2.3) with angle Θ results in change in the phase fronts of the involved waves. At a large distance L from the source, the interference fringes will form only if

$$w\Theta \leq \lambda, \tag{5.15}$$

where λ is the mean wavelength [see Eq. (2.23)]. The coherence area is given by $(L\tan\Theta)^2$, and for small Θ, the transverse coherence length becomes

$$l_c^T \approx L\Theta \approx (L/w)\lambda \tag{5.16}$$

and coherence area becomes

$$A_c \approx (L\Theta)^2 \approx \left[(L/w)\lambda\right]^2, \tag{5.17}$$

which will be used below in considering coherent imaging. For light of $\lambda = 800$ nm, at $L = 2$ m and for a 1 mm incoherent source, the coherence length and area are about 1.6 mm and 2.6 mm^2, respectively. (The sunlight at the surface of the Earth, at the known angular radius of its disc, gives a coherence area of 10^{-2}–10^{-3} mm^2, and similar calculations can be made for other stars.) For the electron, λ_e is very small and, as considered below, the transverse coherence length in our typical arrangement is on the order of 100 nm.

It may be helpful now to re-emphasize that there are differences in the meaning of longitudinal and transverse coherences, and between single-electron packets and many-electron pulses. The de Broglie wavelength is on the pm scale, shorter than the spacing between the atoms, but it is important that the transverse coherence length is on the order of the length scale of the specimen under study of its coherent scatterings (see Chap. 2). For ultrashort pulses, and trains thereof, coherence can be classified into three categories: (i) mutual coherence between different "particles" in a pulse, (ii) intrinsic coherence of the single particles that a pulse is made of, and (iii) the coherence, or well-defined phase relation, which may exist between subsequent pulses within a pulse train. One extreme case is that of a flashlight—the photons are incoherent with respect to each other, but each single photon (burst) has a finite coherence length, and certainly there is no coherence of significance between individual light bursts.

The electron packets used for ultrafast electron imaging are generated by photoionization or field emission triggered by a femtosecond laser, and subsequent acceleration by a static electric field. The electron packets contain at least one and up to 10^5 electrons per femtosecond or picosecond pulse. The coherence between the waves of individual electrons within a single pulse is determined by the initial

generation process, and by the propagation dynamics that follow. For typical conventional operation by heating a filament (without lasers), the electrons from the cathode come from a large number of initial states which are thermally distributed and highly incoherent similar to those of an incandescent lamp; in contrast, field emission sources are coherent emitters. Accordingly, coherence between individual electrons is not expected. Moreover, even if multiple electrons within a single pulse had an initial coherence, this would suffer strongly from Coulomb repulsion during propagation, resulting in a rapid dephasing of macroscopic coherence. For ultrashort pulses of electrons, the excitation is more well defined (laser bandwidth), and the excess energy can be tuned. The temporal coherence, say for a 0.1 eV energy spread, as shown below, is comparable to the pulse width, or shorter.

Ideally, a single-electron packet has a minimum energy uncertainty that is related to the energy width of the laser used to create the packet. For an fs pulse, this represents tens of meV, and the packet is essentially transform-limited with an envelope [see Eq. (5.13)] of the pulse width of the optical pulse but with λ_e in pm; for a 100 fs pulse, $\lambda_e \sim 3$ pm and $\upsilon \sim 0.7c$ — the number of wavelengths of the packet is near 10^7. Below the space–charge limit, for a pulse with many electrons each created with energy E_e, giving rise to a distribution of energy (ΔE), the coherence time becomes shorter and is determined by ΔE ($\equiv h\Delta\nu$). Thus, for $\Delta E = 1$ eV, $\tau_c = 4$ fs is significantly shorter than the pulse width of 100 fs. As a result of such multichromaticity (and phase shift), the number of wavelengths decreases to $\sim 10^5$.

When the coherence time becomes much shorter than the pulse duration, the electrons of the pulse become uncorrelated and can be treated as the "sum" of individual single-electron packets. On the other hand, if $\tau_c = \Delta t_p$, then all packets are correlated. Even in the former case, diffraction and imaging is possible because of the relatively large coherence lengths (transverse and longitudinal), or more appropriately, the coherence volume of single-electron packets, as discussed below. We conclude this part by noting that for ultrafast imaging with single-electron packets, there is one electron in the microscope column at a given time. Because the time between packets (repetition rate) is nanoseconds to milliseconds, the single electron can be made to travel as long as 2 m to 2 km before the second one comes in! The likelihood of crowding by more than one electron will be addressed in Sec. 5.3.4.

There are several other noteworthy points. First, the coherence between optical pulses in a train, or between two pulses from different sources, can be determined using interferometry methods, and usually has the characteristic of being very short. However, there are techniques to phase-stabilize pulses. Second, in a regular train of femtosecond pulses with a locked carrier-envelope phase, all pulses become identical, and in the frequency domain such trains, when averaged for infinitely long times, have very narrow and equally spaced spectral lines (frequency comb), the

Fourier transform of the pulse train, giving rise to $1{:}10^{15}$ precision in the frequency domain. Third, in a chirped laser pulse, the different frequencies have a different group delay (i.e., the frequency increases or decreases with time), leading to a temporally stretched pulse with different wavelengths arriving at different times. A chirp is conveniently described by a spectral phase, including the phase delay of individual frequencies relative to one another. However, coherence is generally independent of chirp or pulse length. The temporal coherence measured in a Michelson interferometer as described above will not depend on the amount of chirp within the initial pulse. As a consequence, for measuring the real pulse length, a nonlinear interferometer is required, for example, with second harmonic generation.

Lastly, when electron pulses of a few femtoseconds or possibly attoseconds are generated, the temporal coherence will ultimately become that of the pulse duration (or exceed it for a frequency comb), opening up new routes to phase-sensitive investigations with ultrashort electron waves. For future diffraction and imaging studies, and along the electron's way to the attosecond domain, a thorough consideration of coherence will be essential. As discussed in Chap. 8, some of the emerging developments include possibilities for attosecond electron pulse generation, reversal of chirp, and the use of energy filtering. In Sec. 5.3.4, we consider the issue of velocity mismatch in ultrafast imaging.

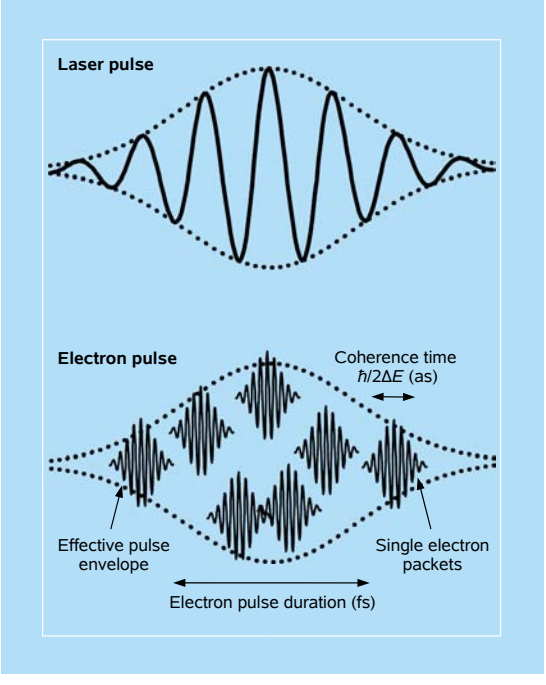


Figure 5.15 An optical pulse envelope (top) and electron pulse (bottom) with coherence time shorter than the pulse width. Adapted from Ref. 17g in Chap. 8.

5.3.2 The double-slit experiment revisited

Historically, the double-slit experiment is one of the most fundamental contributions to science, not only in displaying interference but because of its intrinsic meaning. In Chap. 2, the relevance to coherent diffraction was discussed. Here, we consider one of the "mysteries" of the single-photon (electron) limit of diffraction from two slits, and in the following section we discuss the relevance to our single-electron microscopy imaging. Young's 1801 experiment with light[6–8] stimulated J. J. Thomson to suggest that reducing the light intensity would modify ordinary phenomena of diffraction because the thinking at the time was that its origin was due to an "average effect." In 1909, G. I. Taylor demonstrated[9] that interference (Fig. 5.16) can be obtained with feeble light — after weeks of exposure time. Importantly, he did not appreciate the significance of the finding in relation to Einstein's 1905 paper describing light as particles (photons) and Planck's earlier work on quantization.[8]

Taylor gave the energy (5×10^{-6} ergs/sec) of light falling on 1 cm^2 of the photographic plate he used (Fig. 5.16 caption) but he did not translate this figure

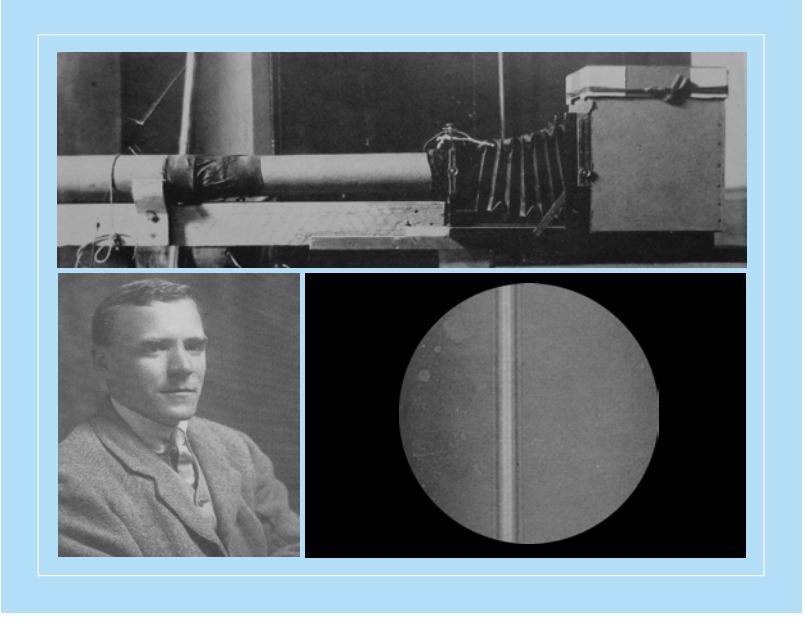


Figure 5.16 During the longest exposures of three months, G. I. Taylor sailed his yacht to Spitzbergen and back! In these cases he estimated that the light energy falling on the plate was about 5×10^{-6} erg s^{-1} cm^{-2}. This would indicate a separation between successive photons considerably greater than the likely beam path he employed. He was evidently deeply puzzled by the result. With hindsight it is a pity that he did not study the underexposed plates more carefully. Their spottiness would not have surprised his supervisor J. J. Thomson, but seeing how the fringes arose with increasing exposure might have led them to connect probability to wave intensity. Adapted from Ref. 9; courtesy of A. Howie (2007).

into number of photons; had he been aware of Planck's and Einstein's work he would have done so, and realized important concepts for photons and interference. Assuming that he had used 2 eV photon energy and a plate of $5 \times 5\ \mathrm{cm}^2$, the photons must have been 10 m apart, so there was only one of them at any time in the apparatus — it was only one photon interfering with itself!

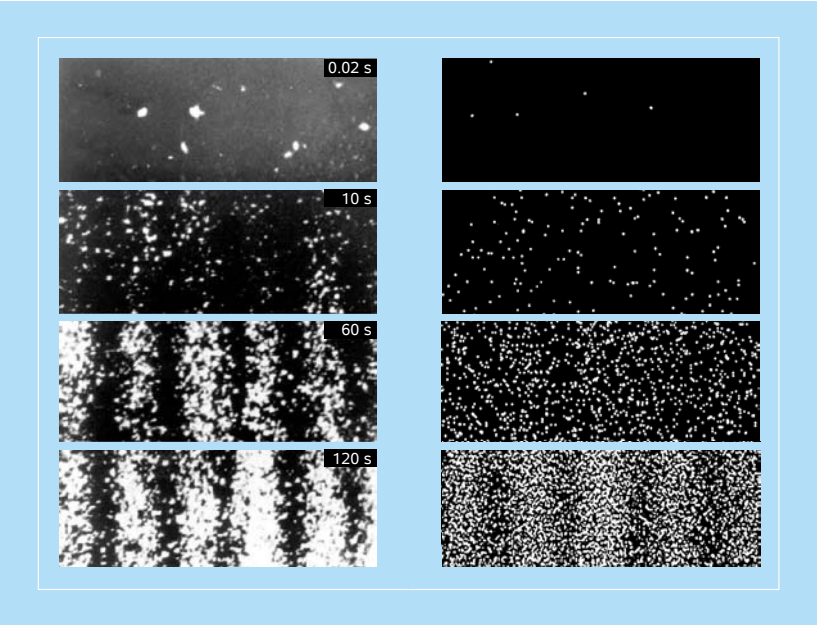


Figure 5.17 Young's double-slit fringes (biprism interference patterns) formed using electrons of very low intensity, recorded as a function of increasing exposure time. There is only one electron in the interferometer at any instant, yet an interference pattern develops with time (the time of flight being much shorter than the time between impacts—1 μs versus 1 ms, respectively; data recorded at the Institut für Angewandte Physik Universität Tübingen in Germany).[10] The work of Tonomura and colleagues is displayed on the right (for details, see Ref. 13).

Interference from a double-slit experiment with electrons was not reported until 1961 using a relatively large number of electrons. Striking results were obtained when the experiments[10–14] were carried out with "one electron at a time" — it was still possible to observe interference (Fig. 5.17)! The results ushered in numerous discussions about the physical meaning of these observations of double-slit experiments with photons and electrons, and many variant thought experiments (*Gedankenexperiment*) were proposed. Richard Feynman was so intrigued by the double-slit finding that in public lectures he would always quiz: "Which slit did the one photon (electron) go through?" — of course building up the drama as the uncertainty and duality of quantum mechanics approached. From the results shown in Fig. 5.17, and the build-up of a diffraction pattern as the number of electrons

increases, it is clear that the spatial distribution for uncorrelated electrons gives the fringes on the detector.

The events have characteristic distributions in time. The arrival times of electrons in the double-slit experiment or in conventional electron microscopes, as measured by video recording, are in general of random intervals. For a continuous wave beam, the statistics becomes that of a Poisson-like distribution, and there is a finite, yet small, probability that two, and three, electrons be a correlated event in a window of time. Such correlations can be zero (random), positive (bunched arrival time) or negative (antibunched arrival time); see Fig. 5.18. Much has been done to study these correlations for photons (with optical interferometry) and electrons (with biprisms). One characteristic of 4D single-electron microscopy is the control over time of each packet, as illustrated in Fig. 5.18a.

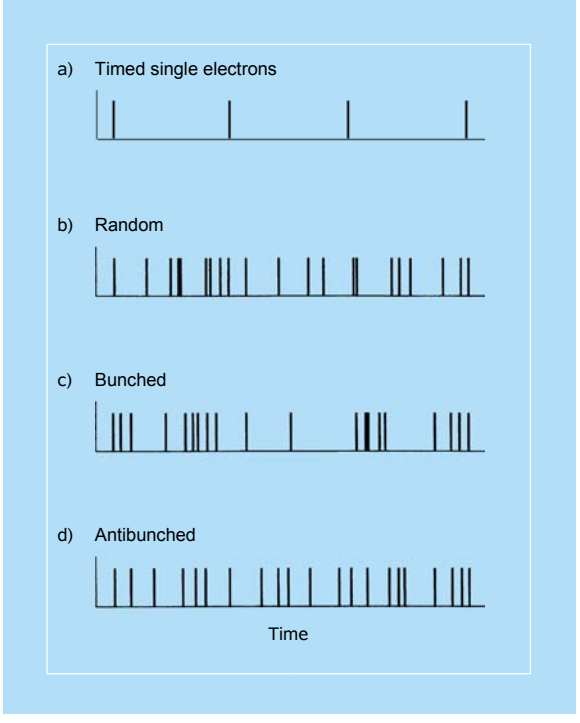


Figure 5.18 Timed versus random (noisy) electron pulses: (a) arrival times for timed, single electrons, (b) random arrival times (zero correlation), (c) bunched arrival times (positive correlation), and (d) antibunched arrival times (negative correlation). The bottom three are adapted from Ref. 4.

5.3.3 Ultrafast versus fast imaging

Here, it is instructive to consider the fundamental differences between imaging on fast (ns or slower) and ultrafast (ps and shorter) time scales. One difference is

clear. For the former, the time resolution will not permit any study of elementary processes which characterize nuclear motions, phase transitions, transient molecular structures, etc. Instead, one measures the average changes, on longer time scales, of rates of physical and reactive transformations, not the dynamics on the atomic and molecular scale; the study of catalysts by environmental TEM (see Chaps. 4 and 7) provides, e.g., such rates on the ms-to-s time scale. This difference is reminiscent of the transition from "flash-photolysis" (with the ms-to-μs kinetics) to fs dynamics of vibrational and rotational motions (see Further Reading, Refs. 1 and 7).

Another important difference lies in the ability to reach both the ultrafast time resolution and atomic-scale spatial resolution using timed single-electron packets. Clearly, because these packets are used stroboscopically (see Sec. 5.2.2), the structures involved have to be recovered during the time between pulses, or the specimen be scanned for each strobe (Chaps. 6 and 8). In experiments using single pulses, the electrons are packed in a short-time envelope and they are space–charge limiting both the temporal and spatial resolutions (ns scale and sub-μm scale, respectively). The fast-pulse methodology introduced by Bostanjoglo, and improved by Browning and others,[15–17] is not our primary focus here, but we will return to this and other methods in Chap. 8, when discussing emerging developments; for a recent review of fast-pulse recording, the reader is referred to the article of Ref. 15.

The difference in the coherence properties of fast and ultrafast imaging, and the spatiotemporal resolution associated with each, are two more issues that are of concern in examining structural changes. We first deal with the coherence. In the above discussion of temporal coherence we invoked simple concepts, but for a more rigorous treatment we must use the well-known correlation-function methods. In what follows we show, both theoretically and experimentally, how the correlation (coherence) time is investigated for laser pulses. Typically, the two pulsed beams in the interferometer are recombined in a nonlinear crystal to measure the resultant second-harmonic-generated light (see Figs. 5.14 and 5.19); a pulse of wavelength λ becomes $\lambda/2$ when focused through a second-harmonic crystal.

The signal measured in background-free second-harmonic generation for a time delay τ between the pulses is proportional to the autocorrelation function of the pulse intensity $I(t)$, given in normalized form by

$$G(\tau) = \langle I(t+\tau)I(t)\rangle / \langle I^2(t)\rangle \tag{5.18}$$

where the brackets denote an average over a sufficiently long period of time. $G(\tau)$ has a maximum value of 1 at zero delay, and decays to zero at longer delay if the intensity $I(t)$ is a single isolated pulse. $G(\tau)$ is always symmetric about $\tau = 0$ and so contains no information on possible asymmetry in the pulse $I(t)$. If $I(t)$ is a smooth unstructured pulse, then $G(\tau)$ is also a smooth function, with an FWHM equal to the pulse duration multiplied by a numerical constant.

An important consideration which is often overlooked in autocorrelation analysis is the effect of pulse substructure on the autocorrelation functions. This is important because, in practice, most lasers are not perfectly mode-locked but experience some fluctuations of the cavity mode phases resulting in substructure

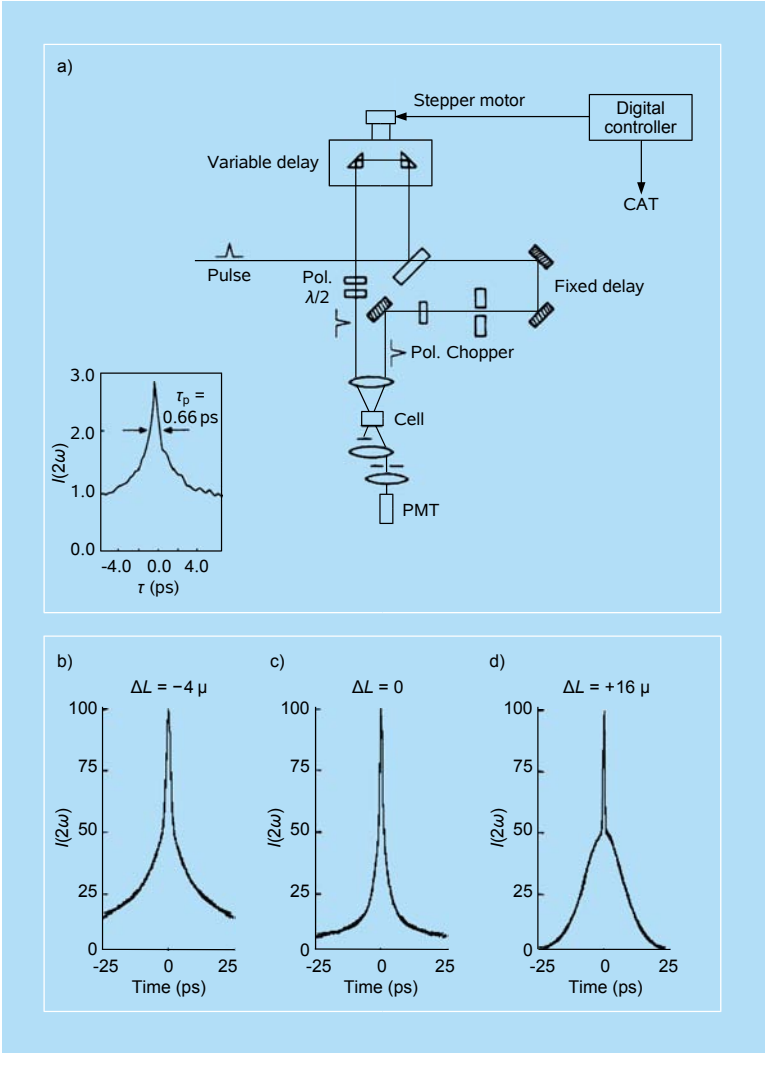


Figure 5.19 Pulse coherence and width as measured by interfereometry. (a) Experimental arrangement for pump-probe spectroscopy. The inset shows the pulse intensity autocorrelation function, obtained by collinear second-harmonic generation. The pulse duration τ_p = 0.66 ps is obtained for a hyperbolic secant pulse. The contrast ratio (peak-background) indicates that the pulse train is well mode-locked. Adapted from Ref. 18a. Bottom: Autocorrelation traces of the output of the synchronously mode-locked dye laser as a function of cavity length detuning: (b) $\Delta L = -4$ μm, (c) $\Delta L = 0$ (exact synchrony), and (d) $\Delta L = +16$ μm. The points (experimental data) have all been normalized to the same maximum intensity and show the change in coherence time relative to the pulse width (see text). Adapted from Ref. 18b.

in the pulse intensity envelope. The pulse intensity function becomes the product of an envelope function and a Gaussian-distributed random variable describing the thermal noise. The pulse autocorrelation function is then found to be

$$G(\tau) = G_{\mathrm{p}}(\tau)\left[1 + G_N(\tau)\right], \tag{5.19}$$

where $G_{\mathrm{p}}(\tau)$ is the autocorrelation function of the pulse envelope, and the term in brackets is the autocorrelation of the Gaussian noise substructure. If the width $\Delta\tau_N$ of $G_N(\tau)$ is much smaller than the width $\Delta\tau_{\mathrm{p}}$ of $G_{\mathrm{p}}(\tau)$, then the autocorrelation described by Eq. (5.19) has the form of a narrow "coherence spike" superimposed on a broad base, with equal relative amplitudes at $\tau = 0$. An example of this type of autocorrelation function is shown in Fig. 5.19, which is obtained from the synchronously mode-locked dye laser when the cavity is too long for optimum mode-locking. Physically, the broad base results from overlap of the pulse envelope, while the "coherence spike" results from overlap of the substructure. The width of the broad base is related to the pulse envelope width (pulse duration) whereas the width of the coherence spike is related to the coherence time of the laser.[18]

When the pulse duration and coherence time are similar in magnitude, the autocorrelation may appear to be quite smooth with no obvious evidence of a coherence spike. The FWHM of the autocorrelation trace will, however, be considerably narrower than the true envelope autocorrelation width. Simply estimating the pulse duration from the observed FWHM will lead to serious underestimates in this case. Ultrashort pulses[19–21] can be made with the pulse width being the same as the coherence width, but fast pulses have incoherent substructures that do not permit such characteristics. Next, we consider the spatiotemporal resolution.

5.3.4 The velocity mismatch and attosecond regime

Besides the pulse duration, there is another temporal spread that imposes an often more severe limitation on time resolution. When performing experiments with pulses of different speed, such as the case here for electrons and photons (used to initiate the change), one has to consider the different group velocities involved, especially for experimental geometries involving tilting.[22] For electrons accelerated at 30–200 keV, this difference results in a large smearing of the time resolution during propagation in the sample (dispersion of time zero); see Fig. 5.20. This mismatch becomes even more significant when a collinear arrangement is not possible, such as in ultrafast electron crystallography (UEC), during which the electrons interrogate the sample at near grazing incidence and the initiating pulse is nearly perpendicular to the surface (see Chap. 6); for a collinear geometry, the only dispersion is that determined by the path length in the specimen divided by the (axial) difference in velocity (for a 100 nm thickness, this dispersion is ~1 fs).

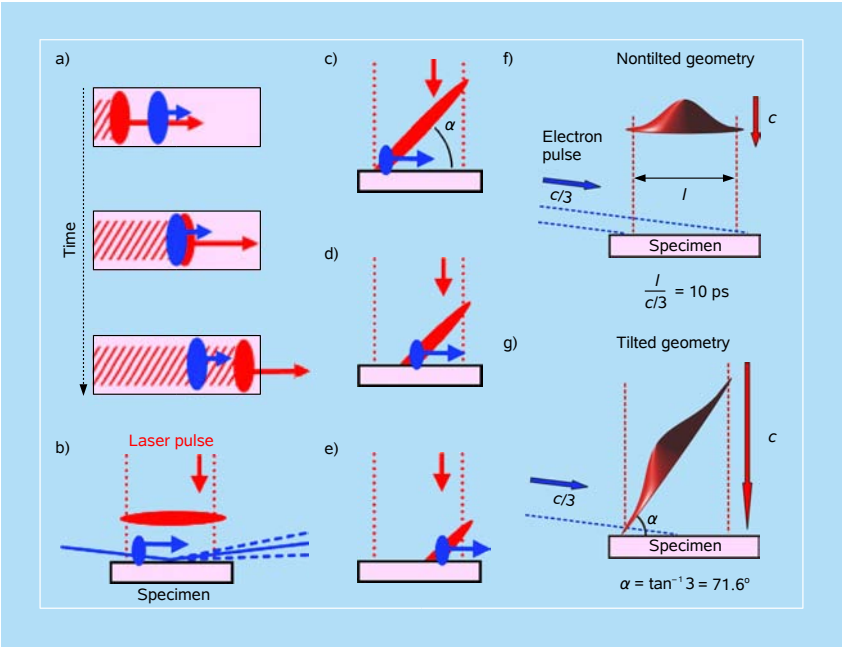


Figure 5.20 Tilted pulse geometry. (a) Concept of group velocity mismatch (GVM). The difference in speed of electron (blue) and laser (red) pulses leads to a spread (dispersion) of time zero throughout the sample and spoils the time resolution. (b) Basic methodology of UEC. Electron pulses (blue) are diffracted from the sample surface after excitation with a laser pulse (red). Different group velocities and the geometry lead to a large timing mismatch on the sample surface. (c–e) The velocity mismatch between the relatively slower electron pulses (33 percent of *c* for UEC) and the laser excitation can be overcome by tilting the optical pulses. Depicted are the (c) early, (d) intermediate, and (e) late time during electron and laser propagation. (f,g) The GVM is represented by the geometries which describe the distances, times, and angles involved (see Ref. 49 in Chap. 6). Figures 5.20–5.25 are adapted from Ref. 23.

Some time ago, in a publication[22] from Caltech, this problem was considered for ultrafast electron diffraction (UED), and conditions were established for minimizing the velocity mismatch, basically by reducing the sample size and/or the initiating laser beam diameter and by changing the angle between the two beams. The optimum spread was found to be subpicosecond, but no solution was offered to reduce it beyond this limit; other geometries, such as UEC, were not considered.

Recently, it was shown how to overcome the effect of group velocity mismatch (GVM) between electrons and light by tilting the optical pulses (Fig. 5.21) such that both the electron and the initiating pulses precisely coincide at every point in the sample and throughout the propagation time.[23] This method can be generally applied to UED, UEC, and UEM, for studies in the gas and condensed phases, and for biological specimens. The method was demonstrated for UEC-type experiments, the worst of all cases. Although the necessary pulse front tilt turns out to be as large as 72°, it was possible to achieve this tilting with ease, and the results showed

a more than 25-fold reduction in time spread. In the same paper,[23] we reported the results of electron-pulse autocorrelation measurements and discussed the advantage of tilting the electron packets to overcome space charge problems that are especially important in single-shot experiments. A method for obtaining the shortest pulse duration was described.

The optical pulses must have an angle between the propagation direction and the plane of highest intensity. Basically, the optical pulses can be viewed as tilted discs propagating in a given direction. Such pulses still have their phase front perpendicular to the propagation direction, but the point of constructive interference between different spectral components is distributed over the spatial beam profile so that constructive interference occurs earlier on one side and later on the other. Tilted pulses have an angular dispersion, and different colors propagate in different directions.

For the arrangement indicated by Fig. 5.21, the necessary pulse front tilt α is determined by the incidence angle γ between the laser and sample, the speed υ_e of electrons, and by the small sample tilt β. From geometrical considerations, the

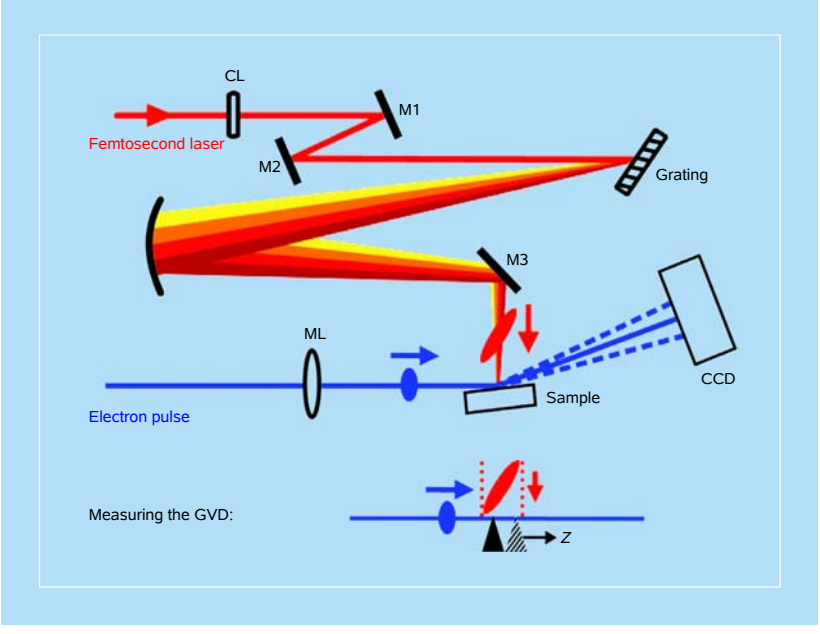


Figure 5.21 Experimental arrangement (relevant part only) for tilted pulse excitation and footprint match. Top: The optical pulses (red) are focused by a cylindrical lens (CL) and fall on a grating. The tilted optical pulses are generated and imaged on the sample with a spherical mirror; M1–M3, steering mirrors 1–3; ML, magnetic lens system. The rainbow colors depict different wavelengths within the 800 nm laser spectrum. Bottom: For measuring the effect of the pulse front tilt, a needle (black) is used. A time scan yields time zero for that particular needle position, which is then varied along the electron propagation direction (Z).

following expression was obtained[23]

$$\alpha = \frac{\pi}{2} - \arctan\left(\frac{\upsilon_e \cos(\beta)^{-1} \sin(\gamma - \beta)}{c - \upsilon_e \cos(\beta)^{-1} \cos(\gamma - \beta)} \right). \tag{5.20}$$

Similar expressions can be derived for other arrangements. For $\gamma = 90°$ and $\upsilon_e = 0.33c$, a pulse front tilt of $\alpha = 72°$ is obtained. Small angles of β at $< 5°$ change the tilt value only by $< 0.5°$. A single tilt angle is therefore sufficient for recording data at different sample angles, allowing for a broad range of experiments involving rocking curves.

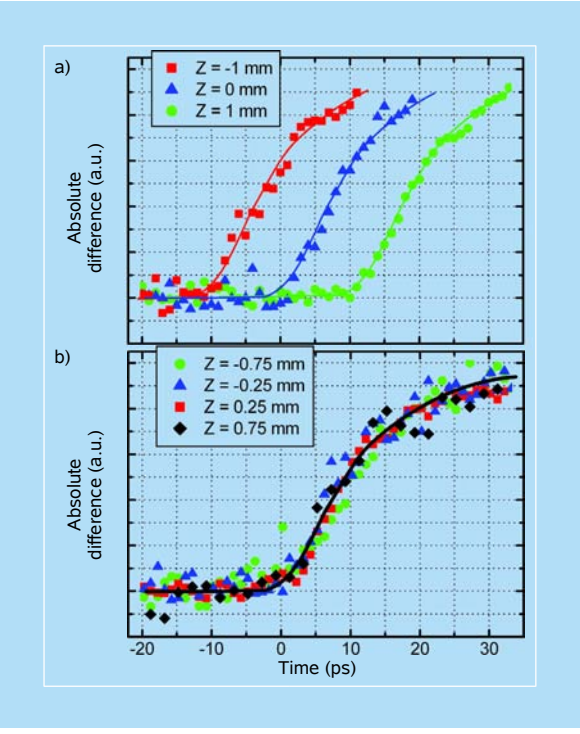


Figure 5.22 Experimental results for measurements of time zero for the nontilted and tilted geometry. (a) Without tilted pulses, the traces are all similar but shifted in time when moving the point of measurement in space. (b) With tilted pulses, all traces coincide regardless of the position *Z*.

The experimental results are shown in Fig. 5.22. The spatial variation of time zero for a nontilted excitation pulse was first measured. For different *Z* positions along the interaction region, very similar but temporally shifted traces were obtained. This behavior clearly shows that time zero shifts (spreads) by the expected dispersion of 10 ps/mm because of the effect of velocity mismatch

between the electron and laser pulses. When the prescribed tilted pulse excitation scheme with the optical parameters set as calculated was used, the results in Fig. 5.22 were obtained. The resulting time traces for different positions are shown in the figure. Independent of the position, all traces coincide, showing that time zero is the same regardless of where in space the interaction region is probed. For more accurate analysis, time zero was determined from all transients by linearly fitting the initial slope and establishing the crossing with the base line. The results are plotted in Fig. 5.23 for the nontilted excitation; one can see the expected 10 ps/mm dispersion.

The same measurement after properly tilting the optical pulses is shown in the lower panel of Fig. 5.23. In contrast to the untilted geometry, over a range of nearly 2 mm, no significant variation in time zero was observed within the measurement accuracy; the error bars are the result of the uncertainty in the characterization of time zero (~1 ps) by the ionization method and not from the tilting scheme. Over the full spatial range, better than 1 ps control of the resolution was observed, and extension to the femtosecond resolution was made (see Chap. 6). This tilting

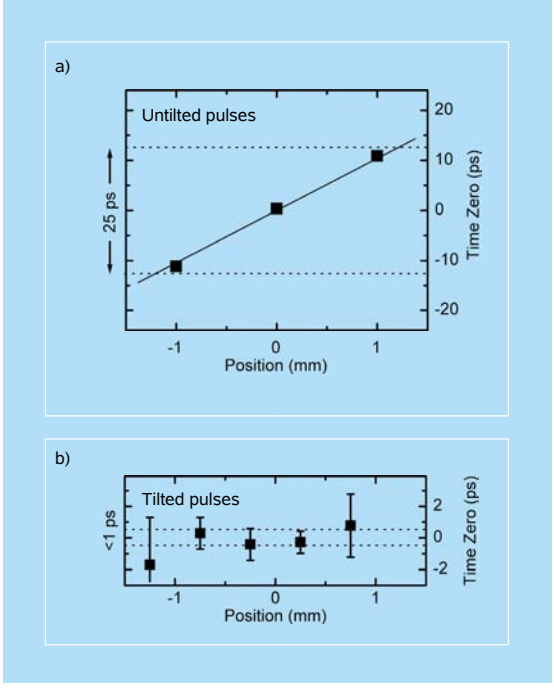


Figure 5.23 Change of time zero with spatial position. (a) Without tilted pulses, time zero (closed square) changes by the dispersion of 10 ps/mm, reducing the overall time resolution by the range indicated (25 ps for a 2 mm sample). The solid line is a fit including data points outside the displayed range. (b) With tilted pulses, time zero does not change significantly across the 2 mm range. The difference between (a) and (b) in position range is due to the footprint used. The dotted lines give a better than 1 ps velocity match, and femtosecond resolution was also possible (see text).

approach eliminates one major hurdle in UEC and other related imaging studies, namely that of velocity mismatch. Moreover, it also offers the advantage of a better match between the excited and the probed (elliptical) regions on the sample surface, and hence, more sensitivity in the experiments because of the absence of contributions from spatial regions that are not initially excited. Applications of the tilting method can be found in Refs. 24–27.

The same tilting scheme was considered for reaching subpicosecond duration in a single-shot recording (Fig. 5.24). Similarly, a ponderomotive deflection scheme was introduced for possible generation of few-femtosecond pulses for single-electron imaging (Fig. 5.25), and a method was proposed[28] for reaching the electron attosecond time domain. The concept is centered on the compression of femtosecond electron packets to trains of 15-attosecond pulses by the use of the ponderomotive force in synthesized gratings of optical fields (Figs. 5.26 and 5.27); chirped electron pulses and energy filtering (Fig. 5.28) were also considered (see also Chap. 8). Such attosecond electron pulses are significantly shorter than those achievable with extreme UV light sources near 25 nm (~50 eV) and have the potential for applications in the visualization of ultrafast electron dynamics, especially of atomic structures, clusters of atoms, and some materials. For imaging, these electron pulses would be the analogue of photon-based attosecond spectroscopy, pioneered and developed over the past decade by Krausz, Corkum, and others.[2,29–37]

5.4 4D Microscopy: Brightness, Coherence and Degeneracy

The essential concept in UEM is based on the premise that trajectories of coherent and timed single-electron packets can provide the equivalent image obtained using *N*-electrons in conventional microscopes (Fig. 5.29). This goal of obtaining micrographs with timed, single-electron packets was achieved, as detailed in Chap. 6, using a femtosecond optical system integrated with a redesigned electron microscope operating at 120 keV or 200 keV. By directly illuminating the photocathode above its work function with feeble femtosecond pulses, both real-space images and diffraction patterns can be obtained. A high-frequency train of pulses separated by nanoseconds or longer intervals allows for recording of the micrographs in seconds.

The classical Coulomb interaction between electrons in a focused beam causes them to avoid one another. Lateral avoidance results in increased beam divergence angle, which blurs images, while longitudinal avoidance causes an increase in energy spread (Boersch effect). This effect is exactly analogous to the complicating Coulomb interaction between electrons in a crystal in band theory. Single-electron imaging circumvents these “crossovers,” or regions of high current

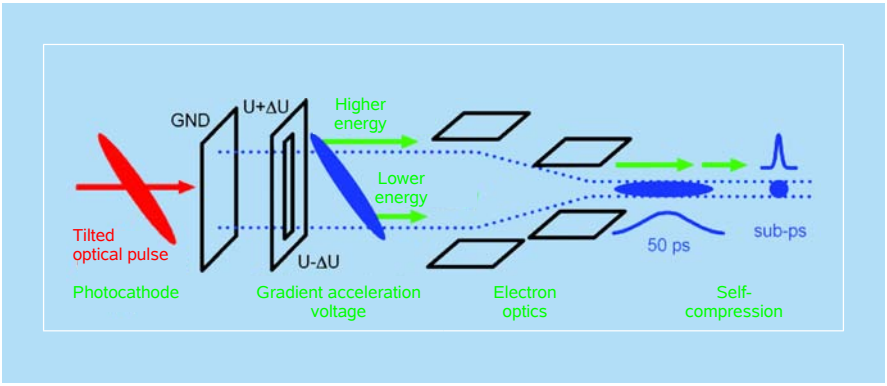


Figure 5.24 Generation of self-compressed femtosecond electron pulses. The concept involves the use of tilted laser pulses to trigger a slit-type photo-activation, generating electrons at different times. An acceleration element with a linear voltage gradient, together with suitable electron optics, then generates an electron pulse with the slow electrons leading the fast ones. The pulse self-compresses in free space to an ultrafast electron density.

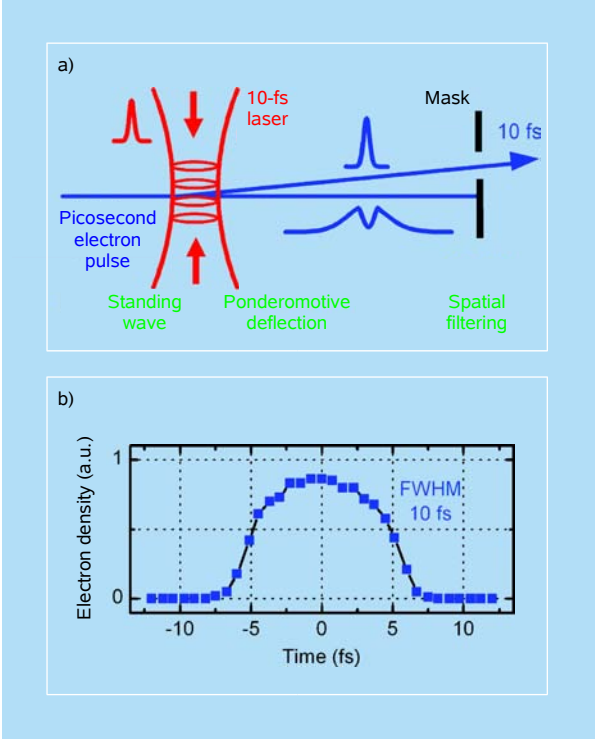


Figure 5.25 Schematic of few-femtosecond electron pulses that may be generated by ponderomotive deflection. (a) Electrons are deflected by the ponderomotive force of a laser lasting a few femtoseconds in a standing-wave geometry. (b) Simulations obtained by integrating the equation of motion of electrons in the ponderomotive force showing that 10 fs electron pulses are possible.

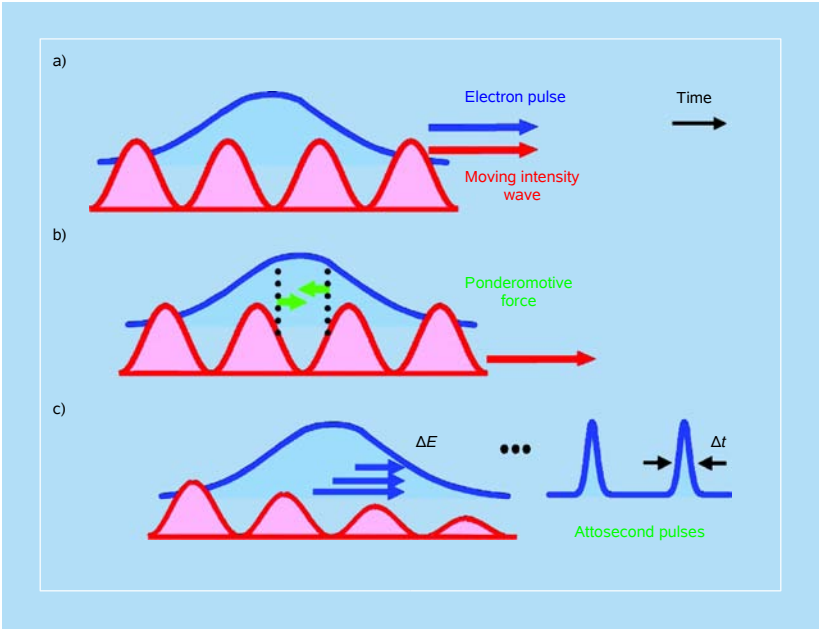


Figure 5.26 Concept of optical compression in a moving intensity grating. (a) An electron packet (blue) copropagates at a velocity of less than the speed of light with the maxima and minima of a moving intensity wave (red), synthesized from two counterpropagating laser pulses of different frequencies. (b) The ponderomotive forces (green arrows) push electrons out of regions with high intensity. Within every optical cycle, the parts of the electron packets that are located at a falling slope are accelerated (left-hand green arrow) and the parts located at a rising slope are retarded (right-hand green arrow). (c) The resultant momentum and velocity distribution leads into self-compression to the final attosecond duration (Δt) within every optical cycle.[28]

density in the so-called space–charge regime, and provides the necessary electron flux for processing of micrographs. The temporal resolution is determined by the pulse width, whereas the spatial and temporal coherence of the image built up from many one-electron pulses remains determined by the same size and energy spread [Eqs. (5.14) and (5.16)]. At atomic-scale resolutions, we have to consider the length and time scales of electron coherence in UEM; namely, the energy–time and space–time relationships. Of particular importance are the fundamental limits of fermionic electrons in determining the coherence volume in imaging and diffraction. The energy–time uncertainty was discussed in a previous section, and here we consider three relevant concepts of space–time imaging.

5.4.1 Coherence volume and degeneracy

Unlike for (bosonic) photons, which can occupy the same space over the duration of a coherent pulse, the Pauli exclusion principle restricts the volume in phase space for electrons. In general, the phase space can be thought of as divided into

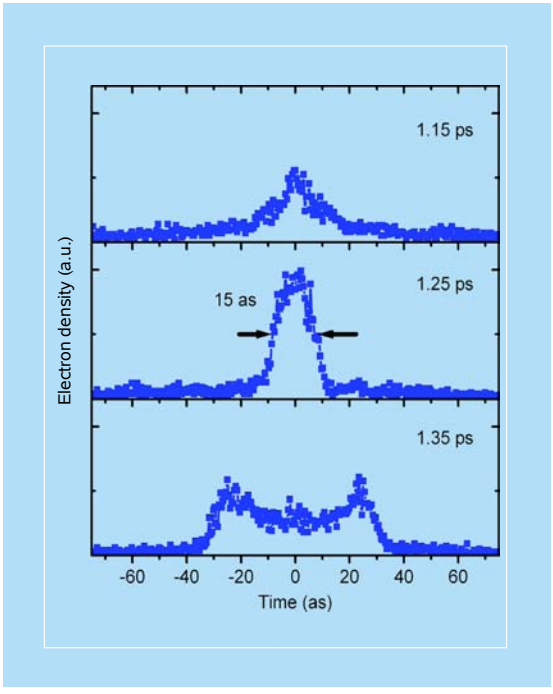


Figure 5.27 Compression dynamics for attosecond electron pulses. After initiating self-compression by an optical intensity wave, the electron density (blue) begins the accumulation (1.15 ps), peaks up to a 15 attosecond duration (1.25 ps), and disperses again (1.35 ps). The times are given relative to the laser pulse overlap.[28]

cells of the size

$$[(\Delta x \Delta y \Delta z) \cdot (\Delta p_x \Delta p_y \Delta p_z)]_{\text{cell}} \sim h^3. \tag{5.21}$$

The volume given by Eq. (5.21) is the *minimum* one determined by the uncertainty principle. An electron belonging to this volume is in a quantum state, and if the probability is approaching one in the volume, statistics has to be invoked in order to deal with populations and correlations of electrons. On the other hand, if the microscope operates in the "dilute regime," i.e., when the number of electrons per cell is much less than one, then the "degeneracy parameter" δ is very small and the problem can be considered without yielding to such statistics.

In general, the degeneracy δ is defined as the mean number of particles (photons or electrons) per cell of phase space. Thus, one can relate δ to the number of particles:

$$n = \delta \cdot \eta, \tag{5.22}$$

where η is the number of occupied cells. Because each cell volume is h^3 [Eq. (5.21)], η is given by the volume of phase space ($2\Delta x \Delta y \Delta z \cdot \Delta p_x \Delta p_y \Delta p_z$; the

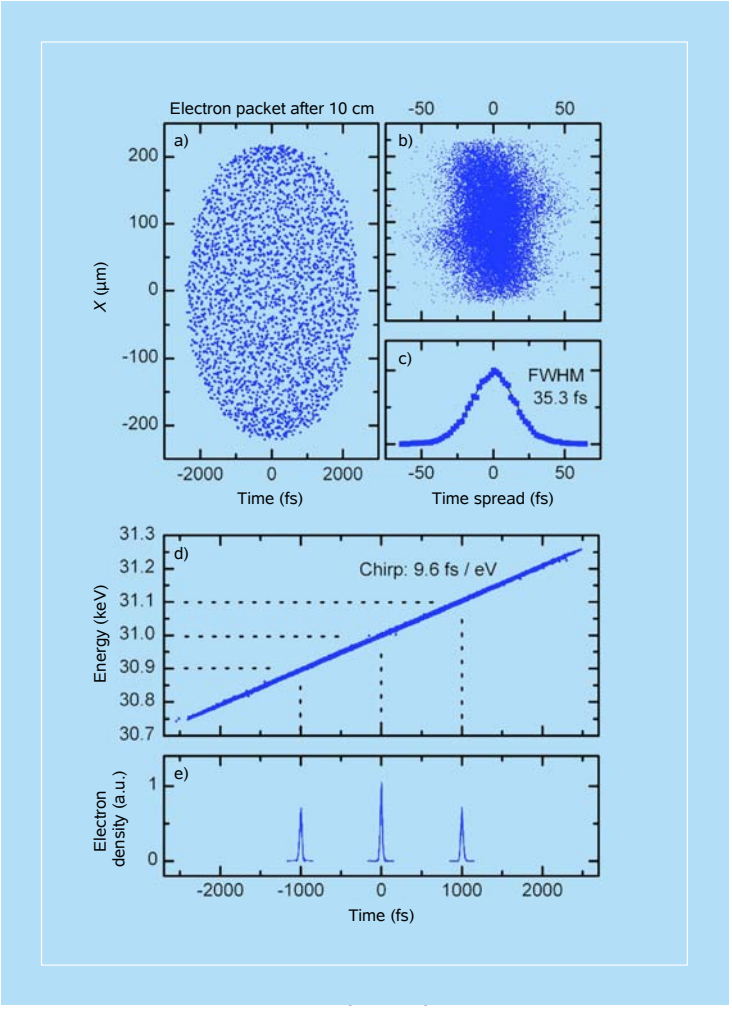


Figure 5.28 Propagation of a chirped electron packet of 20,000 electrons. (a) Spatial packet profile after ~1 ns of propagation; in the panel, *X* is perpendicular to the propagation direction (*Z*), and displayed is the section at $Y = 0$. (b) Residual spread of the correlation trace after subtraction of a linear chirp. (c) Total residual time spread. (d) Momentum–time correlation (chirp). (e) Temporal sections at 30.9 keV, 31.0 keV, and 31.1 keV.[28b]

factor of two is for electron spin) divided by the volume of each cell in phase space (h^3):

$$\eta = \frac{2(\Delta x \Delta y \Delta z) \cdot (\Delta p_x \Delta p_y \Delta p_z)}{h^3}. \tag{5.23}$$

The quantum coherence volume can be related to the classical value by the following relation:

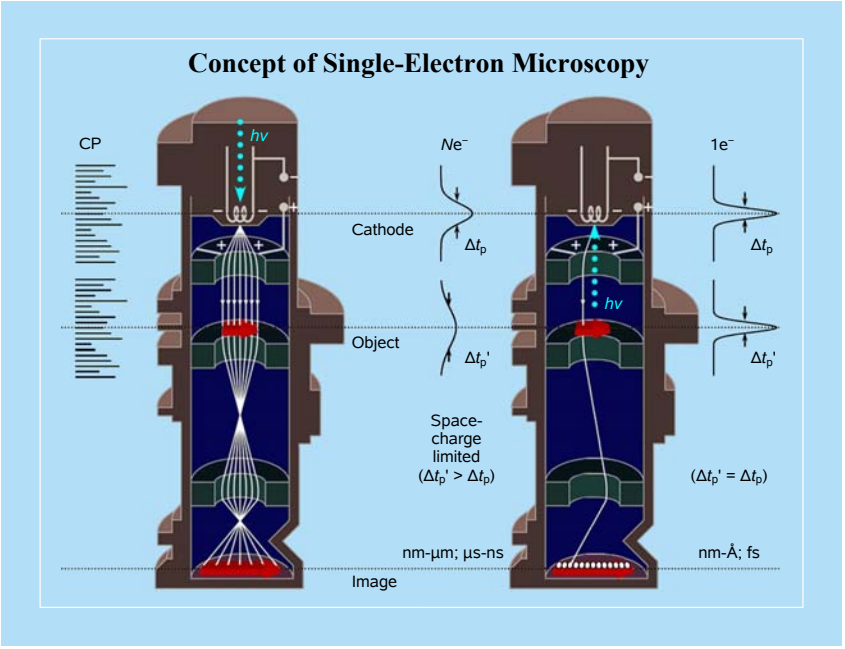


Figure 5.29 Concept of single-electron UEM (right), in comparison with conventional TEM (left). Noted is the difference between conventional (continuous-particle, CP) TEM operation, and single-electron and single-pulse modes of operation in UEM. The influence of the space–charge effect on the spatial and temporal resolutions is indicated in the figure. Adapted from Ref. 7 in Further Reading; see also Ref. 44.

$$\Delta x \Delta y \Delta z = \ell_{\mathrm{tc}}(temporal) \cdot \ell_{\mathrm{sc}}^{2}(spatial) \equiv V_{\mathrm{c}}(cell). \tag{5.24}$$

The temporal coherence is the longitudinal coherence determined by the speed (or wavelength λ) at which the accelerated electrons pass the specimen and by the spread in electron energies (ΔE), or in wavelengths ($\Delta\lambda$), of the packet:

$$\ell_{\mathrm{tc}}(temporal) = \upsilon_{\mathrm{e}} \cdot \frac{h}{\Delta E} \equiv \frac{\lambda^{2}}{\Delta\lambda}. \tag{5.25}$$

The spatial (transverse or lateral) coherence length, in a simple geometry, is the product of λ and (L/w), where L is, in this case, the distance from the specimen to the source, and w is the dimension of the source [see Eq. (2.24)]; in a microscope, the geometry is complicated by the combination of magnetic lenses. It follows that

$$V_{\mathrm{c}} = \left(\lambda \frac{L}{w}\right)^{2} \upsilon_{\mathrm{e}} \cdot \frac{h}{\Delta E}. \tag{5.26}$$

Using typical values for a 120 kV acceleration potential ($\lambda = 3.348$ pm), V_{c} is obtained to be about 5×10^{5} nm^{3}. For our second microscope operating at

200 kV, V_c is *ca.* 10^7 nm^3 (10^{-14} cm^3) at λ of 2.507 pm, assuming the above simple geometry and not including the actual lens system.

5.4.2 Brightness and degeneracy

In classical terms, the brightness B of the source is given by the current density j divided by the solid angle Ω:

$$B = \frac{\hat{j}}{\Omega} \tag{5.27}$$

where $\hat{j}$ is the number of particles per unit area per second (for electrons, $\hat{j} = j/e$). It follows that B is the number of particles Δn received during the time period Δt within a solid angle $\Delta\Omega$ and through a detecting surface ΔA. The solid angle is defined by a cone of apex angle θ ($\sim \pi\theta^2$ for $\theta \ll 1$). From Eqs. (5.22) and (5.23),

$$n = \delta[2(p^2 \Delta p \Delta\Omega) \cdot (\upsilon \Delta t \Delta A)]h^{-3}, \tag{5.28}$$

with B given by[38,39]

$$B = \delta[2(p^2 \Delta p) \cdot \upsilon]h^{-3}, \tag{5.29}$$

as defined above. The product uncertainty in momentum, $\Delta p_x \Delta p_y \Delta p_z$, is written as $p^2 \Delta p \Delta\Omega$ from the definition of solid-angle distances by noting that $\Delta\Omega \approx \Delta p_x \Delta p_y / p^2$ where Δp_x and Δp_y are given by the lateral distance uncertainty in time t, and p, with the momentum being in the z-direction, is given by the total distance traveled in the same time (see the inset of Fig. 5.30 defining these terms).

For photons, δ can be much larger than one (as discussed below), but for electrons the maximum value is one because of the Pauli exclusion principle — one electron in a state or in a cell. Thus, for fermions the maximum brightness is when $\delta = 1$ [in Eq. (5.29)], giving

$$B = \delta B_{\max}. \tag{5.30}$$

One interesting consequence of this result is that, unlike the classical definition of B (continuous value) in Eq. (5.27), the quantum uncertainty sets a limit on B.

From the above relation, Eq. (5.28), another useful expression can be obtained by replacing h^3, the volume of a cell, by the dimensions of coherence expressed in Eq. (5.24). We can then define the degeneracy as

$$\delta = n \frac{t_{tc}}{\Delta t} \frac{A_{sc}}{\Delta A} \equiv \hat{j} t_{tc} A_{sc}, \tag{5.31}$$

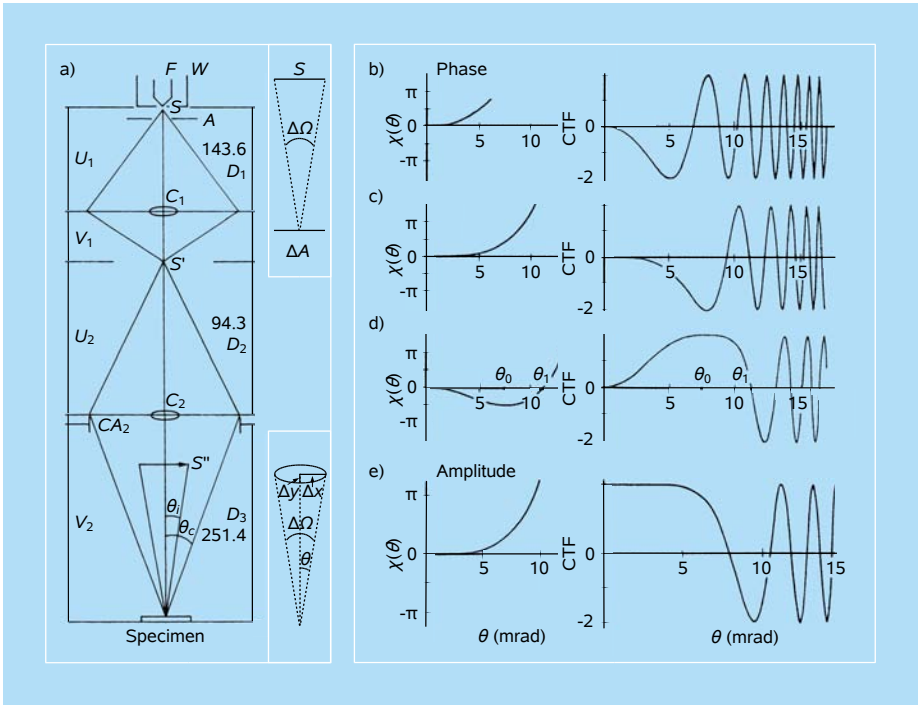


Figure 5.30 Electron-microscope layout and associated contrast-transfer function. (a) The illumination system of a modern electron microscope. There are two condenser lenses: C_1 (a strong lens) and C_2 (a weak lens). F is a filament electron source; W is a Wehnelt cylinder; S is a virtual electron source, with images S' and S''; CA_2 is the second condenser aperture. The distances U_1, U_2, V_1, V_2 are electron optical parameters, while distances D_1, D_2, D_3 are easily measured from the microscope column. Adapted from Ref. 42, and values of the distances are given for a particular microscope in Table 2.1 therein. (b–d) Wave-front aberration and phase contrast transfer function (C_s = 1 mm, λ = 3.7 pm) for (b) $\Delta f = -(C_s\lambda)^{1/2}$, (c) $\Delta f = 0$, and (d) $\Delta f = (C_s\lambda)^{1/2}$. Wave-front aberration and amplitude transfer function for $\Delta f = 0$ (e). Adapted from Ref. 14. The two insets in panel (a) are drawn to indicate some geometrical considerations discussed in this chapter.

with A_{sc} being the spatial coherence area and t_{tc} being simply the temporal coherence length divided by υ. In fact, Eq. (5.31) is for 2δ in order to take the spin into account (two states per cell in the phase space). Degeneracy, therefore, is interpreted as the mean number of particles per coherence time traversing a coherence area normal to the beam; one electron crosses the coherence area during the coherence time when $j = 2e/(A_{sc}t_{tc})$. Several points can now be emphasized. First, although the above definition of B is given in terms of the number of particles that are detected in an area ΔA and in time Δt, it can be shown by simple mathematical manipulation that B is the *total* number of electrons emitted by the source in time Δt (total current) per unit area of the source. Second, in a classical description [Eq. (5.30)], both the number of particles and the number of cells are, in

principle, unlimited; for photons, the number of particles per cell is unlimited but the number of cells (states) is uncertainty-limited; and for electrons, the maximum number of particles per cell is one and the number of cells is uncertainty-limited. Finally, it can be shown[38] that, for example, for fermionic emission sources, $\delta \approx \exp(-\phi\beta)$, where ϕ is the work function and $\beta = (kT)^{-1}$, again implying that there is a maximum of one particle with the probability being diminished by the exponential term.

For electrons, the measurement of δ for a point-like field emitter can be made from the observation of Fresnel edge fringes (the total width of which also gives a simple estimate of spatial coherence and source brightness), typically on the order of 10^{-4} or lower.[39] It is this quantity δ which controls the performance of electron interferometers and the magnitude of quantum interference effects involving electron correlations and antibunching. As discussed in Chap. 2, the complex degree of coherence μ can be expressed as a product of factors describing spatial and temporal coherence, and these factors are proportional to the Fourier transform of the source intensity distribution (spatial coherence) and the distribution of wavenumbers (temporal coherence). The temporal coherence measurement requires an interferometer (biprism)[10] in which a variable time (distant) delay is introduced. The lateral coherence, the width of which is evident in the degree of interferences observed in the diffraction/image, is measured from the width of all the Fresnel fringes seen; the effective source size is about equal to the width of the finest Fresnel fringe, provided that other factors are not limiting.[39–42]

Electron interferences have been used to demonstrate the electron antibunching effect.[40] Unlike the bunching of photons observed in the Hanbury Brown and Twiss experiment, the Pauli exclusion principle for electrons prevents overlapping wave trains due to antisymmetrization of the wave functions.[41] This results in a reduced probability (compared with classical particles) of detecting two electrons within the coherence time. The antibunching effect requires crowding of electrons in phase space, yet the degeneracy of the field emitter is only *ca.* 10^{-4} electrons per cell in phase space. In the high-electron density regime, we must consider Coulomb interactions. The simplest directly observable many-body effect is the Boersch effect,[42] in which Coulomb interactions along the direction of electron travel broaden the energy distribution. Lateral Coulomb repulsion causes an angular divergence, which degrades the spatial resolution.

The brightness B of the source varies depending on the microscope. For a heated field-emission source (ZrO/W) with a virtual source diameter of 100 nm, the measured B value is 10^7–10^8 A cm^{-2} sr^{-1} at 100 kV, whereas for an LaB_6 source of 5–10 μm diameter, $B = 7\times10^6$ at 75 kV. An RT Field emission with a 2 nm virtual source diameter is characterized by a B value of 2×10^9 A cm^{-2} sr^{-1}.[42] The important point is that high brightness allows for the achievement of high coherence (small A), but at the same time gives us the required number of electrons.

Table 5.1 Some Characteristics of a High-Voltage Field-Emission Electron Microscope[38]

Energy	10^5 eV
Energy dispersion	0.3 eV
Wavelength	4×10^{-10} cm
Coherence time	1.4×10^{-14} s
Current density	3.3×10^{13} e cm^{-2} s^{-1}
Angular divergence	1.3×10^{-7} rad
Transverse coherence length	1.5×10^{-3} cm
Coherence area	2.2×10^{-6} cm^2
Brightness	10^8 A cm^{-2} sr^{-1} 6×10^{26} e cm^{-2} s^{-1} sr^{-1}
Degeneracy	10^{-6}

Field-emission electron sources are the brightest particle sources available, considerably brighter than either synchrotron/wiggler systems or the Sun. Consider a field-emission source with the properties given in Table 5.1: the brightness $B = 10^8$ A cm^{-2} sr^{-1}, and the degeneracy $\delta = 10^{-6}$.[38] Taking a typical microscope geometry, Fig. 5.30,[42] with a 400 μm aperture (after the source) and a 25 cm distance between the aperture and specimen (i.e., angle $= 8 \times 10^{-4}$ rad), the current density at the specimen area is near 200 A/cm^2, or near 10^{21} electrons/(s cm^2), reaching 10^5 electrons per second per angstrom squared. The total solar irradiance at the Earth is ~1 kW/m^2. Knowing the radius of the Sun ($\sim 7 \times 10^8$ m), and the distance to the Earth ($\sim 1.5 \times 10^{11}$ m), that is, an angle of 5×10^{-3} rad, B is deduced to be $\sim 10^{10}$ erg s^{-1} cm^{-2} sr^{-1}, which at 500 nm solar radiation gives 10^{21} photons s^{-1} cm^{-2} sr^{-1}. This value is smaller by nearly a million than that of a field-emission source (10^{27} electrons s^{-1} cm^{-2} sr^{-1}), considering both to produce particles. We note that these sources are coherent emitters and ideal for phase-contrast imaging. As discussed in Chap. 2, there is a difference between the lateral coherence at the specimen and that at the source.

We now consider the fundamental difference beween a continuous electron beam and a pulsed one operating in the single-electron and single-pulse modes. The degeneracy changes by orders of magnitude and one must consider consequences of the coherence width of the pulse. For a laser source, the difference is clear, recalling that δ can be significantly larger than 1. From Eq. (5.31), for a 1 mW CW laser with a bandwidth of 0.2 nm and $\lambda \sim 600$ nm, $\delta = 10^4$. For a pulsed laser, the degeneracy increases. For a 1 mJ Ti:sapphire laser producing ~800 nm light and a 100 fs pulse (bandwidth ~10^{13} s^{-1}), δ has a value of more than 10^{16} photons/cell.

For the single-electron packet, when the pulse width reaches the femtosecond domain, each pulse can increase the degeneracy by orders of magnitude, from a value of 10^{-4}–10^{-6} for CW beams with an average energy, reaching values close to

unity when the number of electrons is 100–1000 below (space–charge limit) and/or the pulse duration is reduced to the sub-femtosecond time scale. It is interesting to note that as we reach this time scale, the temporal coherence gets reduced and the cell changes from a "barrel" to a "pancake." For relatively high δ-values, electron correlations become significant. For a 1 fs pulse, the electron packet is 0.1 μm in length (whereas for a 1 ps pulse it is 100 μm), but for 10 as the packet is only 10 Å in length. New possibilities for studies in quantum electron optics should become feasible as the pulse width and ΔE approach the transform limit we discussed earlier. We note that, for relativistic reasons, the Coulomb interaction problem goes away at very high energies.

5.4.3 Coherence and Contrast

We have seen in Sec. 2.5.1 that diffraction in the back focal plane is the consequence of interferences of scattered waves, which mirrors the different distance separations (say, between atoms) in the specimen (a Fourier transform). In the image plane, the interferences are between scattered and unscattered rays (Fig. 2.8), the equivalent of an additional (inverse) Fourier transform for recovering the actual location of atoms. As shown in Sec. 2.5.3, the potential of the specimen $\Phi(x,y)$ causes a phase shift [Eq. (2.68)] and the scattered waves undergo a difference in path length through the lens (another phase shift) before they are focused in the image plane. The scattering gives rise to real and imaginary components, with the former being in the direction of the incident beam and the latter being a phase shift (for each scattering event!). In the weak-phase object approximation, Eq. (2.70) is valid.

It is useful to summarize the critical factors influencing spatial coherence and contrast transfer function. As mentioned in Chap. 2, in the weak-phase approximation, $q = \exp(i\overline{\Phi}) \approx 1 + i\overline{\Phi}$, with the first term representing the "transmitted part" of the beam and $\overline{\Phi}$ representing the scattered part; $\overline{\Phi} = \sigma\Phi(x,y)\Delta z$ [Eq. (2.70)]. In this limit, the intensity is unity and no contrast is expected. However, if we introduce a phase shift of $\pi/2$ between the transmitted and scattered components, then

$$q \approx 1 + i\overline{\Phi}\exp(-i\pi/2) \equiv 1 + \overline{\Phi} \tag{5.32}$$

and

$$|q|^2 \approx 1 + 2\overline{\Phi}(x,y), \tag{5.33}$$

directly establishing a relationship with the specimen phase (x,y) changes.

The directions of scattering to the lens depend on spacings in the specimen as given by $\theta = \lambda/d$, with θ being the resulting diffraction angle; the smaller the

d, the larger the angle θ; d is the inverse of the spatial frequency u. After the lens, the phase shift can be expressed, similar to Eq. (2.77), as

$$\chi(\theta) = \frac{2\pi}{\lambda}\left(\frac{1}{4}C_s\theta^4 - \frac{1}{2}\Delta f\theta^2\right) \tag{5.34}$$

reflecting deviations in path length [$\equiv (2\pi/\lambda)\Delta$ path lengths, in rad]; note that $\chi(0) = 0$. Taking the derivative one defines θ_0 at a minimum (maximum), and by requiring this to correspond to a phase shift of $-\pi/2$ in Eq. (5.34), the following relations can be obtained:

$$\theta_0 = (\lambda/C_s)^{1/4} \quad \text{and} \quad \Delta f = (\lambda C_s)^{1/2}. \tag{5.35}$$

It follows that $\chi(\theta)$ will begin at zero and reach a minimum at $\theta = \theta_0$, and then intersect the θ-axis at θ_1 when $\chi(\theta_1) = 0$ (Fig. 5.30). The value of θ_1 is given by

$$\theta_1 = (4\lambda/C_s)^{1/4}. \tag{5.36}$$

In general, the effect of the lens can be expressed, ignoring $\overline{\Phi}^2$-terms, as

$$q \approx 1 + i\overline{\Phi}(x, y)\exp[i\chi(\theta)] \tag{5.37}$$

and

$$|q|^2 \approx 1 - 2\overline{\Phi}(x, y)\sin\chi(\theta). \tag{5.38}$$

Comparing Eqs. (5.33) and (5.38) indicates the effect of the lens defect, a phase contrast function given by

$$\text{CTF} \equiv -2\sin\chi(\theta), \tag{5.39}$$

as opposed to simply 2 when there is a maximum phase shift of $-\pi/2$ [$\equiv \chi(\theta_0)$]. A plot of CTF versus θ will exhibit a rise from zero to the maximum of +2, then a crossing at θ_1 ($\lambda/\theta_1 = d_1$) followed by oscillations in ± directions of the CTF (Fig. 5.30).

A resolution can now be defined (for coherent illumination):

$$d_{\text{res}} = 0.8\lambda/\theta_1 \tag{5.40}$$

and, from Eq. (5.36), we obtain

$$d_{\text{res}} = 0.6\lambda^{3/4}C_s^{1/4}, \tag{5.41}$$

which is similar to Eq. (2.80). The resolution decreases with decreasing C_s; when $C_s \to 0$ and $\Delta f \to 0$, CTF $\to 0$. Scherzer used $\Delta f = [3/(2\pi)^{1/2}](\lambda C_s)^{1/2}$ as an optimum value for the CTF.

In the above discussion we did not consider "amplitude contrast." If there is a change of amplitude a, then q becomes

$$q = [1 + a(\theta)] \exp(i\overline{\Phi}) \approx 1 + i\overline{\Phi} + a(\theta), \tag{5.42}$$

and with the lens we have

$$q \approx 1 + [i\overline{\Phi} + a(\theta)] \exp[i\chi(\theta)] \tag{5.43}$$

and

$$|q|^2 \approx 1 - 2\overline{\Phi} \sin\chi(\theta) + 2a(\theta)\cos\chi(\theta), \tag{5.44}$$

with the last term describing the amplitude contribution to intensity. The amplitude contribution to CTF can be expressed as

$$\text{CTF(amplitude)} = 2\cos\chi(\theta). \tag{5.45}$$

The function at $\theta = 0$ is a maximum of 2, and oscillates at larger θ-values. Note that for $\theta = 10$ mrad, the spacing d (for an λ of 3 pm) is ~3 Å. We also note that the decay of the CTF by the envelope function is unique: the chromatic contribution [Eq. (2.81)] to total envelope decay is independent of Δf, whereas the spatial envelope function [Eq. (2.83)] does depend on Δf. Thus, close to the focus, the spatial coherence contributes to the envelope function only through the first exponential term of Eq. (2.83) ($\alpha^2 u^6$, where α is the divergence angle). For biological imaging, Δf is usually set at negative values to enhance contrast, especially for contributions of large distances (small u-values), and it is important that lateral coherence be large enough as the envelope function now approaches the exponential of $\alpha^2 \Delta f^2$.

5.4.4 Contrast, dose, and resolution

We conclude this section with a discussion on the relationship between resolution, contrast, and dose. The relationship of pulse duration and spatial resolution is of special interest.

For contrast in an image, the number of electrons for a resolution of an area $(\Delta x)^2$ is simply $N = \overline{n}(\Delta x)^2$, where $\overline{n}$ is the electron density per unit area; the average current density $j = e\overline{n}/\Delta t$. The trade-off between dose and resolution is usually expressed using a modified form of the Rose equation, originally developed for television tubes (see the book by Rose[43] and a concise summary by Spence[42]). Basically, the minimum dose n needed for contrast C can be written as

$$\overline{n} > \frac{k^2}{(fC^2\Delta x^2)}, \tag{5.46}$$

where f is the fraction contributing to background (e.g., $f = 0.95$ for bright field, and 0.001 for dark field). In obtaining this equation, the idea is that the signal-to-noise ratio between adjacent resolution elements for a given C should exceed about $k = 5$ (Rose condition). It follows that

$$j(\Delta t)(\Delta x)^2 > \frac{\text{const.}}{C^2}. \tag{5.47}$$

The expression may appear as an "uncertainty relationship," but because the current density is per Δt, we can write

$$(\Delta x)^2 > \frac{\text{const.}}{\bar{n}C^2}, \tag{5.48}$$

meaning that, for a given contrast, the resolution depends on the electron density per unit area, and equivalently, the dose is inversely proportional to the square of the resolution (in 2D). Only if experiments are designed to vary Δt for a fixed current density would Eq. (5.47) appear as an "uncertainty relationship" with a significant role for the temporal duration.

In general, one has to consider the actual geometry of the experiment. From the experimental point of view, there are several factors that affect the resolution, such as lens aberrations, mechanical stability and specimen damage. As discussed in the present chapter, there is a direct relationship between the dose and the resolution achieved, and the brightness of the source is an important quantity. It is worth noting that from Eqs. (2.54) and (2.55) and the definition of brightness, one can relate $(\Delta x)^2$ to the total beam spot area D by a constant: $(\Delta x)^2 = \text{const.} \times D$, where the constant is simply given by the value of the wavelength, current and brightness B.[39b]

In closing this chapter, it is worth mentioning one additional unique feature of ultrafast techniques besides the time resolution and coherence, namely, the number of photons per atom or molecule that can be excited. Consider a 1 W CW laser beam versus a 1 μJ, 1 ps laser pulse. If used to excite say a gold specimen with a density of about 10^{28} atoms/m^3 in a volume of ~1 μm^3, we can calculate the number of photons per gold atom. The number of photons of energy 2 eV is for the pulse (10^{-6} J) ~ 10^{12} photons, whereas the number of atoms in the volume mentioned is ~10^{11} atoms. This means that with the 1 ps pulse, 10 photons, in principle, are available for each atom at total absorption; even for 10 percent absorption, we could have a photon/atom ratio of one conduction electron (carrier) per atom. In contrast, for the CW beam (1 W) in 1 ps, the energy is ~10^{-12} J, the number of photons is ~10^6, and only 10^{-6} photons/atom are possible at 10 percent absorption and in 1 ps. This consideration becomes critical especially when the atoms relaxation or dephasing time is on the ultrashort time scale.

Further Reading

1. A. H. Zewail, in "*Les Prix Nobel, The Nobel Prizes 1999*," Ed. T. Frängsmyr, Almqvist & Wiksell, Stockholm, p. 103 (2000).
2. A. H. Zewail, *Sci. Am*., **263**, 76 (1990).
3. J. S. Baskin and A. H. Zewail, *Sci. Spectra*, **14**, 62 (1998).
4. J. S. Baskin and A. H. Zewail, *J. Chem. Educ*., **78**, 737 (2001).
5. V. K. Jain, *The World & I*, **X**, 156 (1995).
6. H. C. von Baeyer, "*Taming the Atom: The Emergence of the Visible Microworld*," Random House, New York (1992).
7. A. H. Zewail, in "*Visions of Discovery: New Light on Physics, Cosmology, and Consciousness*," Eds. R. Y. Chiao, M. L. Cohen, A. J. Leggett, W. D. Phillips and C. L. Harper, Jr., Cambridge University Press, Cambridge, in press.
8. (a) E. Muybridge, "*Animals in Motion*," Dover, New York (1957).
   (b) R. B. Haas, "*Muybridge: Man in Motion*," University of California Press, Berkeley (1976); see also Ref. 11.
9. (a) E. Jussim and G. Kayafas, "*Stopping Time: The Photographs of Harold Edgerton*," H. N. Abrams, New York (1987).
   (b) M. Frizot, "*La Chronophotographie*," Association des Amis de Marey, Beaune (1984).
10. (a) J.-M. Hopkins and W. Sibbett, *Sci. Am*., **283**, 73 (2000), and references therein.
    (b) Refs. 18 and 19 in Ref. 1 give more details.
11. J. M. Thomas, "*Michael Faraday and the Royal Institution: The Genius of Man and Place*," Taylor and Francis, New York (1991).

References

1. H. Abraham and J. Lemoine, *C. R. Hebd. Seances Acad. Sci*., **129**, 206 (1899).
2. (a) N. Bloembergen, *Rev. Mod. Phys*., **71**, S283 (1999).
   (b) N. Bloembergen, in "*Progress in Optics*," Vol. 50, Ed. E. Wolf, Elsevier, Amsterdam, p. 1 (2007).
   (c) F. Krausz and M. Ivanov, *Rev. Mod. Phys*., **81**, 163 (2009), and references therein.
   (d) C. Altucci and D. Paparo, in "*Developments in Surface Contamination and Cleaning: Fundamentals and Applied Aspects*," Eds. R. Kohli and K. L. Mittal, William Andrew, Norwich, p. 25 (2008).
3. L. Mandel and E. Wolf, "*Optical Coherence and Quantum Optics*," Cambridge University Press, New York (1995).
4. R. Loudon, "The Quantum Theory of Light," 2nd ed., Oxford University Press, New York (1983), p. 227.
5. S. Svanberg, "*Atomic and Molecular Spectroscopy: Basic Aspects and Practical Applications*," 4th ed., Springer, Berlin–Heidelberg (2004).
6. T. Young, *Phil. Trans. R. Soc. Lond*., **92**, 12 (1802).
7. T. Young, *Phil. Trans. R. Soc. Lond.*, **94**, 1 (1804).
8. (a) A. Howie and J. E. Ffowcs Williams, *Phil. Trans. R. Soc. Lond. A*, **360**, 805 (2002).
   (b) A. Howie, private communication.

9. G. I. Taylor, *Proc. Cambridge Phil. Soc.*, **15**, 114 (1909).
10. (a) H. Lichte, *Phil. Trans. R. Soc. Lond. A*, **360**, 897 (2002).
    (b) J. C. H. Spence, in "*Compendium of Quantum Physics: Concepts, Experiments, History and Philosophy*," Eds. D. Greenberger, K. Hentschel and F. Weinert, Springer, Berlin–Heidelberg (2009).
11. (a) C. Jönsson, *Am. J. Phys.*, **42**, 4 (1974).
    (b) C. Jönsson, *Z. Physik*, **161**, 454 (1961).
12. P. G. Merli, G. F. Missiroli and G. Pozzi, *Am. J. Phys.*, **44**, 306 (1976).
13. A. Tonomura, J. Endo, T. Matsuda, T. Kawasaki and H. Ezawa, *Am. J. Phys.*, **57**, 117 (1989).
14. A. Tonomura, "*Electron Holography*," 2nd ed., Springer, Berlin–Heidelberg (1999).
15. O. Bostanjoglo, *Adv. Imaging Elecron Phys.*, **121**, 1 (2002).
16. H. Kleinschmidt, A. Ziegler, G. H. Campbell, J. D. Colvin and O. Bostanjoglo, *J. Appl. Phys.*, **98**, 054313 (2005).
17. (a) J. S. Kim, T. LaGrange, B. W. Reed, M. L. Taheri, M. R. Armstrong, W. E. King, N. D. Browning and G. H. Campbell, *Science*, **321**, 1472 (2008).
    (b) T. LaGrange, G. H. Campbell, P. E. A. Turchi and W. E. King, *Acta Mater.*, **55**, 5211 (2007).
    (c) T. LaGrange, G. H. Campbell, B. W. Reed, M. Taheri, J. B. Pesavento, J. S. Kim and N. D. Browning, *Ultramicroscopy*, **108**, 1441 (2008).
18. (a) D. P. Millar, R. Shah and A. H. Zewail, *Chem. Phys. Lett.*, **66**, 435 (1979).
    (b) D. P. Millar and A. H. Zewail, *Chem. Phys.*, **72**, 381 (1982).
19. J.-C. Diels and W. Rudolph, "*Ultrafast Laser Pulse Phenomena: Fundamentals, Techniques, and Applications on a Femtosecond Time Scale*," Academic Press, San Diego (1996).
20. E. N. Glezer, in "*Spectroscopy and Dynamics of Collective Excitations in Solids*," Ed. B. Di Bartolo, Plenum Press, New York (1997).
21. A. E. Siegman, "*Lasers*," University Science Books, Sausalito (1986).
22. J. C. Williamson and A. H. Zewail, *Chem. Phys. Lett.*, **209**, 10 (1993).
23. P. Baum and A. H. Zewail, *Proc. Natl. Acad. Sci. USA*, **103**, 16105 (2006).
24. N. Gedik, D.-S. Yang, G. Logvenov, I. Bozovic and A. H. Zewail, *Science*, **316**, 425 (2007).
25. P. Baum, D.-S. Yang and A. H. Zewail, *Science*, **318**, 788 (2007).
26. F. Carbone, P. Baum, P. Rudolf and A. H. Zewail, *Phys. Rev. Lett.*, **100**, 035501 (2008).
27. D.-S. Yang, C. Lao and A. H. Zewail, *Science*, **321**, 1660 (2008).
28. (a) P. Baum and A. H. Zewail, *Proc. Natl. Acad. Sci. USA*, **104**, 18409 (2007).
    (b) P. Baum and A. H. Zewail, *Chem. Phys. Lett.*, **462**, 14 (2008).
29. P. B. Corkum and F. Krausz, *Nature Phys.*, **3**, 381 (2007).
30. P. Agostini and L. F. DiMauro, *Rep. Prog. Phys.*, **67**, 813 (2004).
31. D. G. Lappas and A. L'Huillier, *Phys. Rev. A*, **58**, 4140 (1998).
32. I. P. Christov, M. M. Murnane and H. C. Kapteyn, *Phys. Rev. Lett.*, **78**, 1251 (1997).
33. G. Sansone, E. Benedetti, J.-P. Caumes, S. Stagira, C. Vozzi, S. De Silvestri and M. Nisoli, *Phys. Rev. A*, **73**, 053408 (2006).

34. R. López-Martens, K. Varjú, P. Johnsson, J. Mauritsson, Y. Mairesse, P. Salières, M. B. Gaarde, K. J. Schafer, A. Persson, S. Svanberg, C.-G. Wahlström and A. L'Huillier, *Phys. Rev. Lett*., **94**, 033001 (2005).
35. M. Drescher, M. Hentschel, R. Kienberger, M. Uiberacker, V. Yakovlev, A. Scrinzi, T. Westerwalbesloh, U. Kleineberg, U. Heinzmann and F. Krausz, *Nature*, **419**, 803 (2002).
36. M. Uiberacker, T. Uphues, M. Schultze, A. J. Verhoef, V. Yakovlev, M. F. Kling, J. Rauschenberger, N. M. Kabachnik, H. Schröder, M. Lezius, K. L. Kompa, H.-G. Muller, M. J. J. Vrakking, S. Hendel, U. Kleineberg, U. Heinzmann, M. Drescher and F. Krausz, *Nature*, **446**, 627 (2007).
37. J. Itatani, J. Levesque, D. Zeidler, H. Niikura, H. Pépin, J. C. Kieffer, P. B. Corkum and D. M. Villeneuve, *Nature*, **432**, 867 (2004).
38. M. P. Silverman, "*Quantum Superposition: Counterintuitive Consequences of Coherence, Entanglement and Interference*," Springer, Berlin–Heidelberg (2008).
39. (a) J. C. H. Spence, W. Qian and M. P. Silverman, *J. Vac. Sci. Technol. A*, **12**, 542 (1994), and references therein.
    (b) W. Chiu and R. M. Glaeser, *Ultramicroscopy*, **2**, 207 (1977).
    (c) C. J. Humphreys and J. C. H. Spence, *Optik*, **58**, 125 (1981)
40. (a) H. Kiesel, A. Renz and F. Hasselbach, *Nature*, **418**, 392 (2002).
    (b) J. C. H. Spence, *Nature*, **418**, 377 (2002).
41. (a) R. Hanbury Brown and R. Q. Twiss, *Nature*, **178**, 1046 (1956).
    (b) M. P. Silverman, *Phys. Lett. A*, **120**, 442 (1987).
42. J. C. H. Spence, "*High-Resolution Electron Microscopy*," 3rd ed., Oxford University Press, New York (2003).
43. A. Rose, "*Vision: Human and Electronic*," Plenum Press, New York (1973).
44. "*Physical Biology: From Atoms to Medicine*," Ed. A. H. Zewail, Imperial College Press, London (2008).

Chapter 6

4D Ultrafast Electron Imaging

Developments and Applications

In this chapter our focus is on the methodologies involved in imaging and diffraction studies and on the applications of ultrafast electron microscopy (UEM). To begin with, we describe the core instruments which were built at Caltech. They involve the integration of ultrafast lasers with modified electron microscopes to enable the recording of real-space images, diffraction patterns, scanning probes, and electron-energy-loss spectra (EELS). Besides two generations of these instruments (UEM-1 and UEM-2), we also briefly describe the associated methods (of central importance here) of ultrafast electron diffraction (UED) and ultrafast electron crystallography (UEC). We highlight different applications using examples of studies of structural, morphological, and mechanical (functional) dynamics.[a]

6.1 Developments at Caltech — A Brief History

As discussed in Chap. 5, the motion of atoms is probed in femtochemistry using spectroscopic methods, which include the use of optical detection, photoelectron detection, and mass spectrometry. In 1991 we suggested the replacement of spectroscopic probes with an ultrafast electron pulse, which is capable of capturing transient structures.[1] This was followed by theoretical study,[2] commentary,[3] and demonstration,[4] but it took three generations of instruments (UED-3)[5] to be able to go beyond demonstration and to investigate real complex systems, covering studies of transient structures of chemical reactions, excited-state structures, dark nonradiative pathways, conformational interconversions, and orientational ordering in diffraction.[6,7] The approach of ultrafast electron probing was further developed during the course of building the UEC apparatus[8] for the study of crystals, interfaces, and surfaces.[9] The applications vary in scope and have included semiconductor surfaces,[10] water at interfaces,[11] high-T_c materials,[12] phase transitions,[13] and nanorods.[14] In retrospect, we began the journey at the

[a]Throughout this chapter, the use of the word "we" appropriately refers to the Caltech authors referenced in the text.

most difficult frontier, that of isolated structures in molecular beams. For such structures, the coherence is limited to a single molecule, not a crystal.

With the success in UED and UEC, we decided to extend the methodology to microscopy[15] but learned that we needed a new way of thinking. The key concept of single-electron (Chap. 5) imaging — to circumvent the problem of Coulomb repulsion between electrons — was the paradigm shift that made possible the temporal and spatial resolutions at the atomic scale. Of relevance also is coherence and the "*uncertainty relationship*" between time and energy and the apparent "uncertainty" that may limit the spatial resolution, which have already been discussed in Chaps. 2 and 5. Armed with these advances and challenges, we constructed UEM-1 and, later, UEM-2; the former operates at 120 keV and the latter at 200 keV. We first demonstrated single-electron imaging in real space, and then augmented the instrument, in UEM-2, with electron-energy resolution and scanning capabilities. We also operated the microscope in two modes, that of single-electron packets and single-pulse imaging, as detailed below. Because imaging became possible in the four dimensions of space and time, our interest branched into various applications.[16–26] The potential of UEM was recognized and highlighted in the commentary by Thomas.[27]

Since the first reports from this laboratory, the technical and theoretical machinery had to be further developed, and currently at Caltech there are five

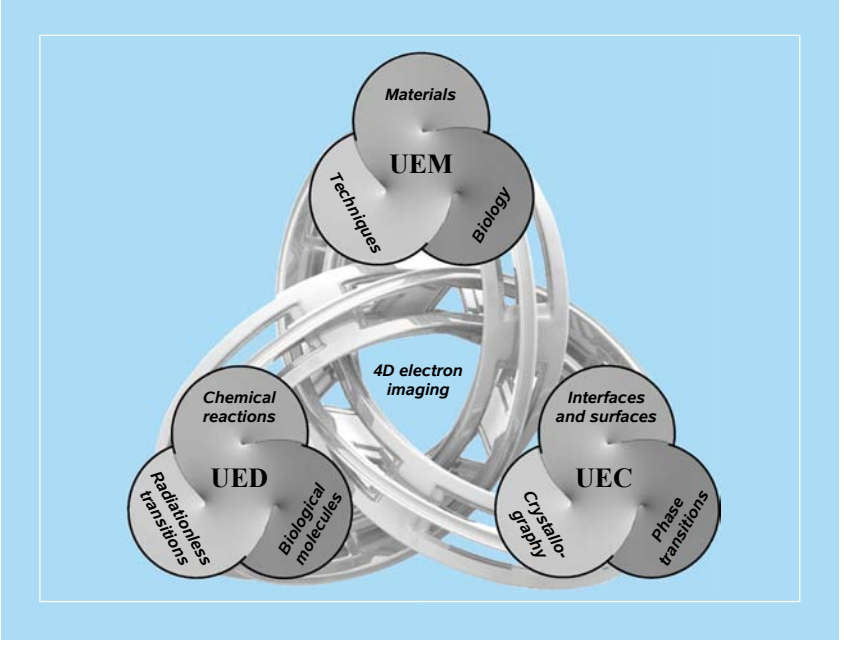


Figure 6.1 The "trefoil knot" of 4D ultrafast electron microscopy (UEM), crystallography (UEC), and diffraction (UED).

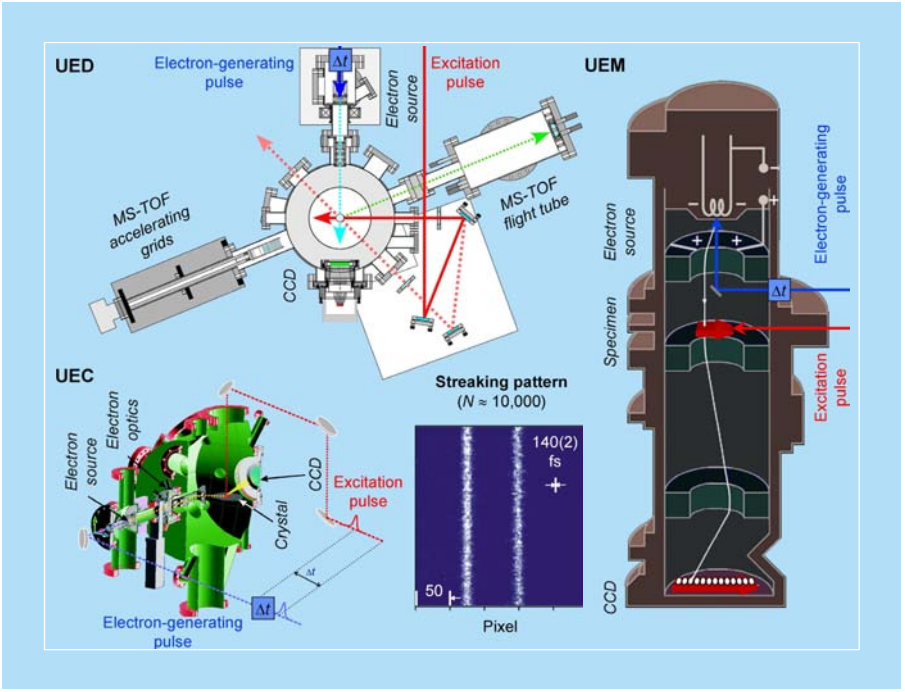


Figure 6.2 4D imaging apparatuses for ultrafast electron diffraction (UED), crystallography (UEC) and microscopy (UEM). For UEM, the single-electron trajectory is schematized. Inset: Streaking pattern for pulses of about 10,000 electrons in UEC.[29]

table-top instruments for studies of gases, condensed matter, and biological systems. The totality of these interconnected studies is artistically given by the trefoil knot designed by D. Shorokhov (see Fig. 6.1). Examples of the experimental setups for UED, UEC, and UEM are displayed in Fig. 6.2, and recent review articles[28,29] summarize the pertinent contributions from Caltech, which are the focus of discussion on applications in what follows.

6.2 Instruments and Techniques

The conceptual framework of the approach is as follows. Upon the initiation of the structural change by either heating of the sample, or through electronic excitation induced by ultrashort laser pulses, a series of electron pulses is used to probe the specimen with a well-defined time delay. An electron diffraction pattern or a microscopic image is then obtained. A series of images recorded at a number of delay times forms a movie which provides an insight into the temporal evolution of the structure.

Shown in Fig. 6.3 is a picture of the second-generation UEM-2 built for operation in the single-electron and single-pulse modes. The integration of two

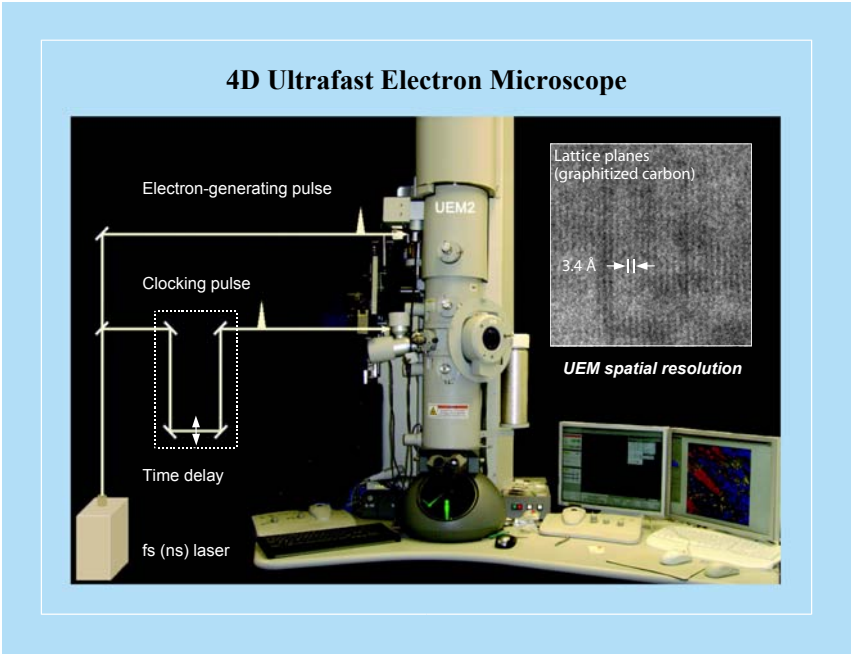


Figure 6.3 Photograph of the UEM-2 microscope at Caltech, together with a high-resolution image. Displayed is the interface of laser systems (fs and ns) to a modified, hybrid 200 kV transmission electron microscope designed with two ports for optical access. The optical pulses are directed to the photocathode to generate electron packets as well as to the specimen to initiate (clock) the change in images with a well-defined delay time (details are given in the text). By simply flipping two mirrors, the operation can vary from single-electron to single-pulse imaging. The lattice image of a graphitized carbon with 3.4 Å separations shows the spatial resolution obtained in UEM.[24]

laser systems to a modified electron microscope is indicated in the figure, together with a high resolution (3.4 Å) image obtained in the UEM mode; this spatial resolution was achieved without the field-emission-gun (FEG) arrangement of a transmission electron microscope (TEM). In the single-electron mode of operation, which affords studies of reversible or repetitive processes, a train of up to 10^8 electron pulses is used to build up the image or diffraction pattern stroboscopically (see Chap. 5). By contrast, in the single-pulse mode, each recorded frame is created with only one pulse of 10^5–10^7 electrons; reversibility of the process is not a requisite in this mode. Conceptually, the experimental design is the same for either imaging mode, with the only difference being the fluence and pulse width involved. As discussed below, by simply flipping two mirrors, the freedom to operate the UEM apparatus for either single electron or single pulse is achieved.

The first atomic-scale (angstrom) resolution in imaging was achieved in the second-generation UEM-2. As mentioned above and shown below, with UEM-2, which operates at 200 keV ($\lambda_{\text{de Broglie}} = 2.5079$ pm), it was also possible to obtain energy-space (electron-energy-loss spectroscopy, EELS) and Fourier-space

(diffraction) patterns of nanostructured materials. The apparatus can operate in the scanning transmission electron microscope (STEM) mode and is designed to explore the vast parameter space bridging the gap between the two ideal operating modes of single-electron and single-pulse imaging. With these features, and with convergent-beam capability, UEM studies provide new limits of resolution, image mapping, and elemental analysis.

Figure 6.4 displays the conceptual design of UEM-2, which, as with the first generation (UEM-1), involves the interface of the electron microscope to the modified laser/optics arrangement—here with the pump-probe configuration outlined. The fs-laser in this case is a diode-pumped Yb-doped fiber oscillator/amplifier, which produces ultrashort pulses of up to 10 μJ at 1030 nm with variable pulse width of 200 fs to 10 ps, and repetition rate of 200 kHz to 25 MHz (in UEM-1, the pulse width is 100 fs and a pulse picker is part of the setup). The

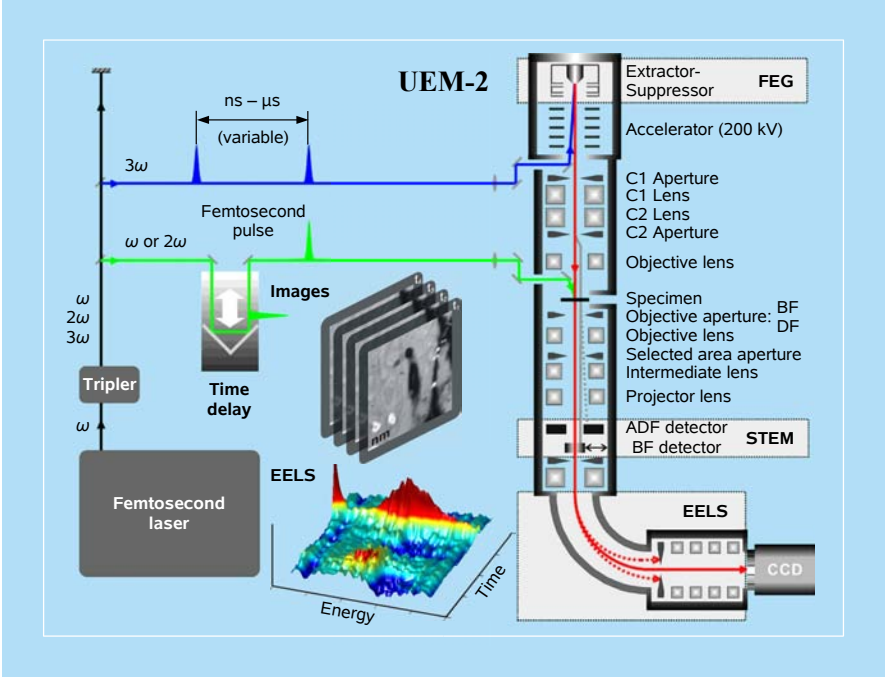


Figure 6.4 Conceptual design of Caltech's UEM-2. Schematic representations of optical, electric, and magnetic components are displayed. The optical pulse train generated from the laser, in this case having a variable pulse width and a variable repetition rate, is divided into two parts after harmonic generation and guided toward the entries of the designed hybrid electron microscope. The frequency-tripled optical pulses are converted to the corresponding probe electron pulses at the photocathode in the hybrid FEG, whereas the other optical pump beam excites (*T*-jump or electronic excitation) the specimen with a well-defined time delay with respect to the probe electron beam. The probe electron beam through the specimen can be recorded as an image (normal or filtered, EFTEM), a diffraction pattern, or an EEL spectrum. The STEM BF detector is retractable when it is not in use. Adapted with modifications from Ref. 18.

output pulses pass through two successive nonlinear crystals to be frequency-doubled (515 nm) and -tripled (343 nm). The harmonics are separated from the residual infrared radiation (IR) beam by dichroic mirrors, and the frequency-tripled pulses are introduced to the photocathode of the microscope to generate the electron pulse train. The residual IR fundamental and frequency-doubled beams remain available to heat or excite samples and clock the process through a computer-controlled optical delay line for time-resolved applications.

The electron microscope column is that of a designed hybrid 200 kV TEM integrated with two ports for optical access, one leading to the photocathode and the other to the specimen. The FEG in the electron-generation assembly adapts a lanthanum hexaboride (LaB_6) filament as the cathode, terminating in a conical electron source truncated to leave a flat tip area with a μm-scale diameter. The tip is located in a field environment controlled by suppressor and extractor electrodes. The gun can be operated as either a thermal-emission or a photo-emission source.

The optical pulses are guided to the photocathode as well as to the specimen by a computer-controlled, fine-steering mirror in an externally mounted and X-ray-shielded periscope assembly. Each laser beam can be focused to a spot size of $< 30\ \mu$m full width at half-maximum (FWHM) at its respective target when the beam is expanded to utilize the available acceptance angle of the optical path. Various pulse-energy, pulse-length, and focusing regimes have been used. The cathode was heated to a level below that needed to produce detectable thermal emission, and images were obtained using both the TEM and the UEM modes of operation.

In applications involving EELS and energy-filtered transmission electron microscopy (EFTEM), we used the Gatan imaging filter (GIF) Tridiem, of the so-called postcolumn type, attached below the camera chamber. The GIF accepts electrons passing through an entrance aperture in the center of the projection chamber. The electron beam passes through a 90 sector magnet, as shown in Fig. 6.4, which bends the primary beam through a 10 cm bending radius and thereby separates the electrons according to their energy into an energy spectrum. An energy resolution of 0.87 eV was measured for the EELS zero-loss peak in thermal mode operation of the TEM, and typically 1–2 eV in UEM, depending on the number of electrons. A retractable slit is located after the magnet followed by a series of lenses. The lenses restore the image or diffraction pattern at the entrance aperture, and finally it can be recorded on a charge-coupled device (CCD) camera at the end of the GIF with DigitalMicrograph software. The digital camera uses a 2048×2048 CCD chip with $14\ \mu$m square pixels. Readout of the CCD is done as four independent quadrants via four separate digitizing signal chains. This four-port readout camera combines single-electron sensitivity and 16-bit pixel depth with high-speed sensor readout (4 Mpix/s).

Additionally, for STEM, the UEM is equipped with a bright-field (BF) detector with a diameter of 7 mm and an annular dark-field (ADF) detector with an inner diameter of 7 mm and an outer diameter of 20 mm. Both detectors are located in the near-axis position underneath the projection chamber. The BF detector usually collects the same signal as the TEM BF image, i.e., the transmitted electrons, while the ADF detector collects an annulus at higher angles where only scattered electrons are detected. The STEM images are recorded with imaging and analysis software.

To observe diffraction patterns, i.e., the scatterings at the back focal plane of the objective lens, a selected-area aperture is inserted into the image plane of the objective lens, thus creating a virtual aperture in the plane of the specimen. The result is a selected-area diffraction (SAD) pattern of the region of interest only. Adjustment of the intermediate and projector lenses determines the camera length. Diffraction patterns are then processed and analyzed for structural determination. Convergent-beam diffraction is achieved by focusing the probe beam, as detailed at the end of this chapter.

Several key features realized in the new UEM-2 machine are worth emphasizing. First, the high-repetition-rate amplified laser source allows for the illumination of the cathode with 343 nm pulses of energies above 500 nJ compared with typical values of 3 nJ near 380 nm for UEM-1. Thus, the system delivers a level of average optical power for electron generation comparable to that of UEM-1 operating at 80 MHz, but at much lower repetition rates. The pulse energy available in the visible and IR beams is also at least two orders of magnitude greater than for UEM-1, allowing for exploration of a much greater range in the choice of sample excitation conditions.

Second, the hybrid 200 kV FEG, incorporating an extractor/suppressor assembly providing an extractor potential of up to 4 kV, permits higher resolving power and greater flexibility and control of the conditions of electron generation. Third, with simple variation of optical pulse width, the temporal and spatial resolutions can be controlled depending on the requirements of each experiment. Fourth, with variation of spacing between optical pulses without loss of pulse energy, a wide range of samples can be explored, allowing them fully to relax their energy after each excitation pulse and rewind the clock precisely; with enough electrons, below or near the space-charge limit, single-pulse recording is possible. Finally, by the integration of the EELS spectrometer, UEM is empowered with energy resolution in addition to the time and space resolutions.

The proof of principle for UEM imaging can be seen in the results shown in Fig. 6.5 for gold particles and gold islands on carbon films. Panels (a) and (d) of Fig. 6.5 show gold particles of uniform size dispersed on a carbon film. From the higher magnification image of Fig. 6.5d, corresponding to the area indicated by the black arrow in Fig. 6.5a, it is found that the gold particles have a size of

15 nm, and the minimum particle separation seen in the image is 3 nm. Panels (b) and (e) were recorded under identical conditions to panels (a) and (d), respectively, but without cathode irradiation by the femtosecond laser pulses. No images were observed, demonstrating that nonoptically generated electrons from the warm cathode were negligible. Similar background images with the light pulses blocked are routinely recorded and checked for all cathode conditions used. The waffle (cross line) spacing of the cross-grating replica (gold islands) seen in Fig. 6.5c is known to be 463 nm. The gold islands are observed in Fig. 6.5f, where the bright regions correspond to the amorphous carbon support film and the dark regions to the nanocrystalline gold islands.

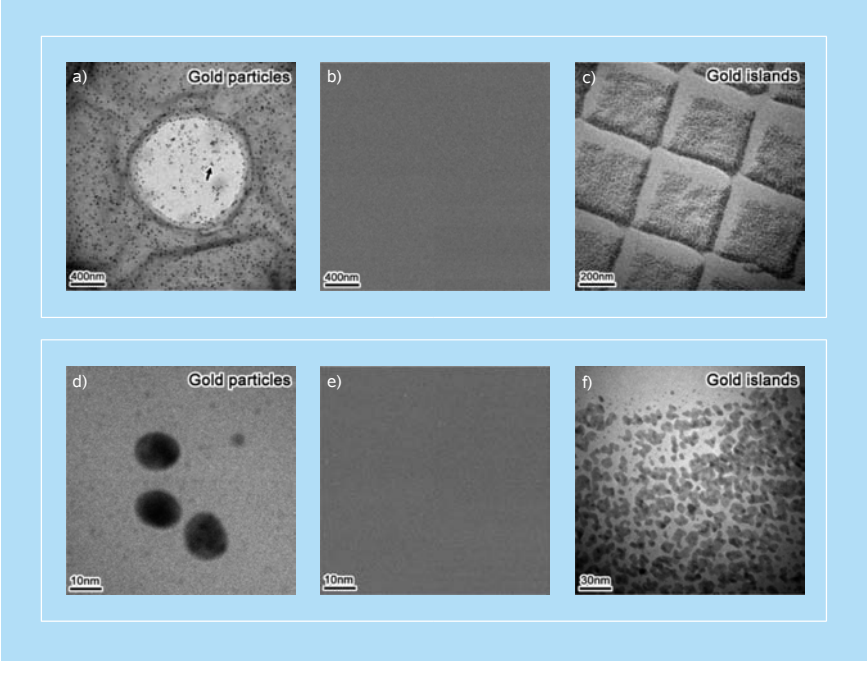


Figure 6.5 UEM images obtained with ultrafast electron pulses. Shown are (a,d) gold particles and (c,f) gold islands on carbon films and (b,e) UEM background images obtained by blocking the photoelectron-extracting femtosecond laser pulses. For the UEM images of gold particles, the objective (contrast) aperture of 40 μm was used to eliminate diffracted beams, while no objective aperture was used for the gold-island images. Figures 6.5 to 6.9 are adapted from Ref. 18.

To test the high-resolution capability of UEM by means of *phase contrast imaging* (Chap. 2), an organometallic compound, chlorinated copper phthalocyanine (hexadecachlorophthalocyanine, $C_{32}Cl_{16}CuN_8$) was investigated. The major spacings of lattice fringes of copper of this molecule in projection along the *c*-axis are known to be 0.88, 1.30, and 1.46 nm, with atomic spacings of 1.57 and 1.76 nm. Figure 6.6a exhibits the lattice fringes observed by UEM,

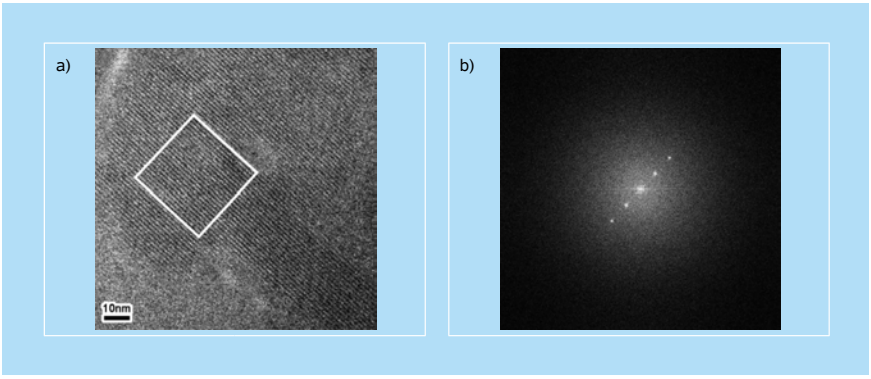


Figure 6.6 High-resolution, phase-contrast UEM images. Shown are (a) the image and (b) digital diffractogram of an organometallic crystal of chlorinated copper phthalocyanine. The diffractogram was obtained by the Fourier transform of the image in (a). The high-resolution image was taken near the Scherzer focus for optimum contrast, which was calculated to be 90.36 nm for a spherical aberration coefficient C_s of the objective lens of 2.26 mm. The objective aperture was not used.

where the black lines correspond to copper layers parallel to the *c*-axis. The Fourier transform of Fig. 6.6a is shown in Fig. 6.6b, discussed below, and the clear reciprocity (without satellite peaks in the FT) indicates the high degree of order in crystal structure. Figure 6.7a is an enlargement of the area outlined in Fig. 6.6a, clearly showing the lattice fringe spacing of 1.46 nm, corresponding to the copper planes highlighted in gray in Fig. 6.7b, in which a unit cell is shown in projection along the *c*-axis. It is known that, in high-resolution images, the lattice fringes produced by the interference of two waves passing through the back focal plane,

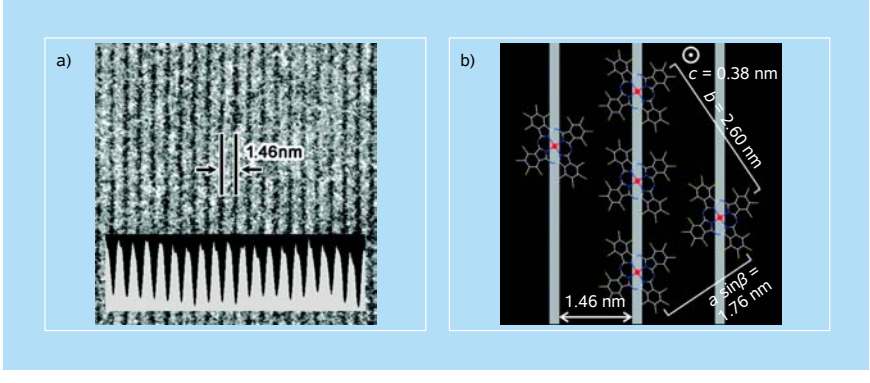


Figure 6.7 High-resolution, phase-contrast UEM image and structure of chlorinated copper phthalocyanine. The high-resolution image (a) is a magnified view of the area outlined in Fig. 6.6a. The representation of the crystal structure (b) is shown in projection along the *c*-axis, and the assignment of the copper planes observed in (a) is indicated by the gray lines. The red spheres are the copper atoms.

i.e., the transmitted and diffracted beams, are observed only in crystals where the lattice spacing is larger than the resolution of the TEM. In the profile inset of Fig. 6.7a, it should be noted that the FWHM was measured to be less than 7 Å, directly indicating that UEM has the capability of subnanometer resolution, and consistent with the 3.4 Å separations resolved for graphitized carbon as indicated in Fig. 6.3.

The digital diffractogram obtained by the Fourier transform of the observed high-resolution image of Fig. 6.6a is shown in Fig. 6.6b. In the digital diffractogram, the peaks represent the fundamental spatial frequency of the copper layers (0.69 nm^{-1}) and higher harmonics thereof. A more direct means of obtaining reciprocal-space information such as this is the recording of electron diffraction, also available in UEM. Figure 6.8 depicts measured and calculated electron diffraction patterns of gold islands and boron nitride (BN) on carbon films, along with the corresponding real-space images of each specimen, all recorded by UEM. In Fig. 6.8b, the electron diffraction patterns exhibit Debye–Scherrer rings formed by numerous diffraction spots from a large number of the face-centered gold

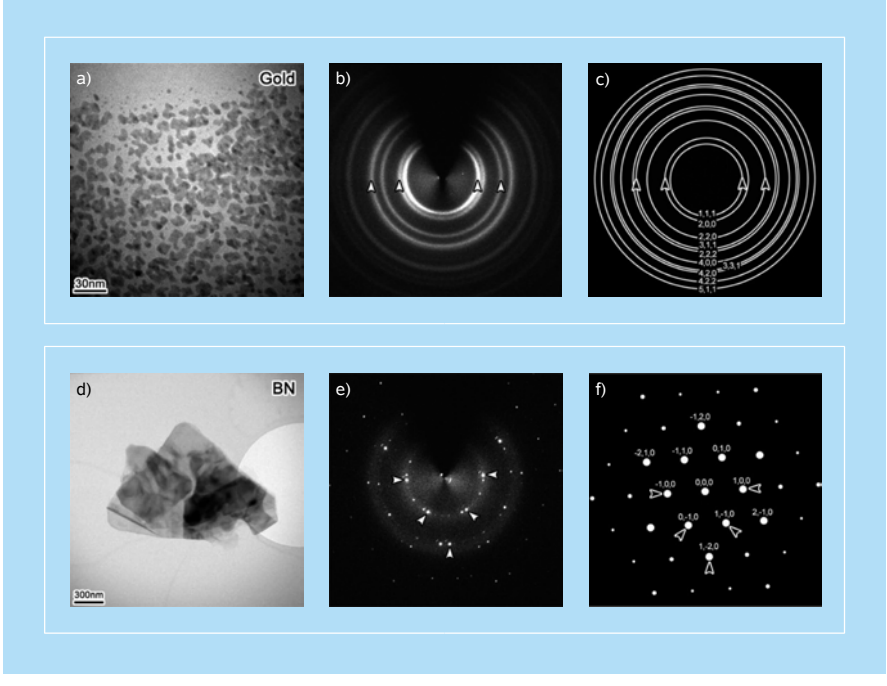


Figure 6.8 UEM images and electron diffraction patterns. Shown are images and measured and calculated electron diffraction patterns of (a–c) gold islands and (d–f) boron nitride (BN) on carbon films. The incident electron beam is parallel to the [001] direction of the BN. All diffraction patterns were obtained by using the selected-area diffraction (SAD) aperture, which selected an area 6 μm in diameter on the specimen. Representative diffraction spots and rings were indexed as indicated by the arrowheads.

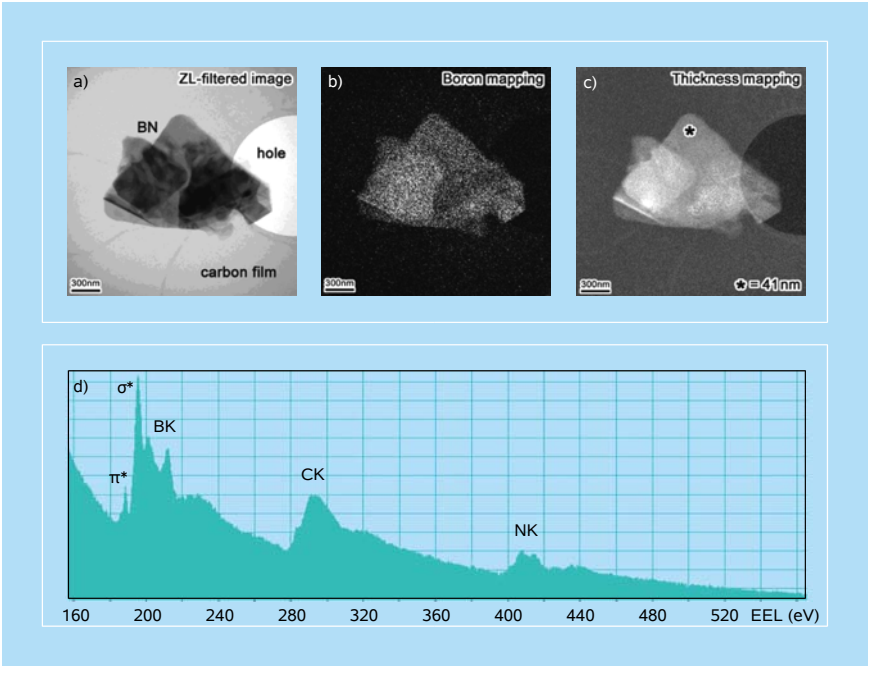


Figure 6.9 Energy-filtered UEM images and spectrum: (a) A zero-loss filtered image, (b) boron K-edge mapping image, (c) thickness mapping image, and (d) EEL spectrum of the BN sample. The 5.0 and 1.0 mm entrance apertures were used for mapping images and EEL spectrum, respectively. The thickness at the point indicated by the asterisk in (c) is estimated as 41 nm. ZL stands for zero-loss.

nanocrystals with random orientations. The rings can be indexed as indicated by the white arrowheads. The diffraction pattern of BN in Fig. 6.8e is indexed by the hexagonal structure projected along the [001] axis, as shown in Fig. 6.8f. It can be seen that there are several BN crystals with different crystal orientations besides that responsible for the main diffraction spots indicated by the white arrowheads.

To explore the energy resolution of UEM, we investigated the BN specimen in detail by EELS and EFTEM. Figure 6.9 displays (a) a zero-loss filtered image, (b) boron K-edge mapping image, (c) thickness mapping image, and (d) the corresponding EEL spectrum. The boron map was obtained by the so-called three-window method. In the boron map of Fig. 6.9b, image intensity is directly related to areal intensity of boron. In the thickness map of Fig. 6.9c, the brightness increases with increasing thickness:

$$d = \bar{\lambda}(\beta)\ln(I_t/I_0), \tag{6.1}$$

where $\bar{\lambda}$ is the mean free path for inelastic scattering for a given collection angle β, I_0 is the zero-loss (ZL) peak intensity, and I_t is the total intensity. The thickness in the region indicated by the asterisk in Fig. 6.9c was estimated to be

41 nm (see below). In the EEL spectrum of Fig. 6.9d, the boron K-edge, carbon K-edge, and nitrogen K-edge are observed at the energies of 188, 284, and 401 eV, respectively (see also Sec. 4.6 and Fig. 4.24). In the boron K-edge spectrum, sharp π^* and σ^* peaks are clearly visible. The carbon K-edge spectrum is considered to result from the amorphous carbon film due to the existence of small and broad peaks at the position π^* and σ^* — quite different from the spectra of diamond and graphite.

The temporal resolution has been characterized in imaging, diffraction, and EELS studies, as discussed below, but first we make a few general remarks regarding the important controlling factors. The optical pulses used to generate the electron packets in UEM-1 and UEM-2 were 100 and 200 fs, respectively. These widths (Δt_{hv}) are not limiting as shorter optical pulses are now available, down to a few femtosecond and even attosecond durations (Chaps. 6 and 8). Below the space–charge limit, for an electron pulse of tens of electrons, and certainly for the single-electron packet, the width is determined by the energy dispersion the electron carries following liberation from the photocathode. But because the electron is accelerated at 200 kV energy and the energy dispersion is near 1 eV, the temporal broadening reaches a similar magnitude to the optical pulse width. The total resolution in the absence of the space–charge effect can be estimated from the relationship[30]:

$$\Delta t = \left[\Delta t_{\mathrm{hv}}^2 + \Delta t_{\mathrm{KE}}^2\right]^{1/2}. \tag{6.2}$$

The results were confirmed by electron dynamics simulations. The "longitudinal broadening" is expressed by

$$\Delta t_{\mathrm{KE}} = \frac{d}{eV}\left(\frac{m_0}{2}\right)^{1/2} \frac{\Delta E_i}{(E_i)^{1/2}}, \tag{6.3}$$

where e is the electron charge and m_0 is the electron rest mass. It is solely determined by the energy spread, ΔE_i, relative to the square root of the mean energy, $(E_i)^{1/2}$, and the potential V across the distance d between cathode and anode (see Fig. 6.10). Given the parameters of UEM-2, Δt_{KE} is of a similar magnitude to the optical pulse, which is convenient for ultrafast imaging. In single-pulse imaging with millions of electrons, the dominant broadening arises from space charge. With modified optical excitation, cathode design, and modified extraction geometry, there are different schemes for the possibility of reaching the the ultrashort single-pulse probing, as outlined in Chap. 5 and which will be discussed further in Chap. 8. We note that transverse broadening is related to spatial coherence, as discussed in the previous chapter, and theoretical

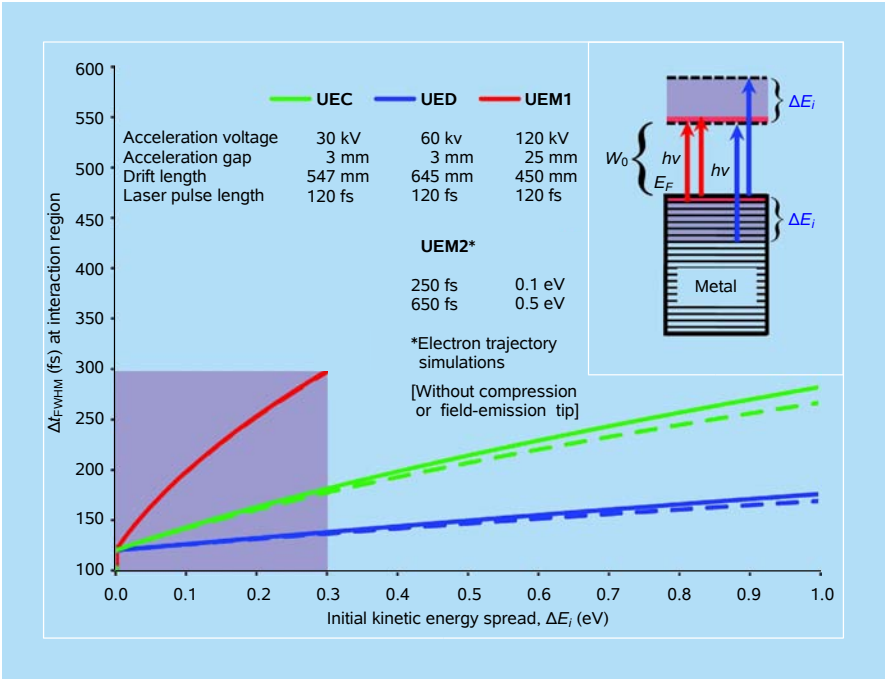


Figure 6.10 Theoretical estimates of the effect of initial kinetic energy spread on temporal broadening due to excess energy of the electrons above the photocathode work function for the UED-4 (blue line), UEC (green line) and UEM-1 (red line) instruments. The pulse length at the specimen interaction region is computed in the absence of space charge using the instruments' specific parameters. In the inset, the red arrow indicates the excitation from near the Fermi surface (E_F) to just above the vacuum, while the blue arrow is for excitation which carries an excess of photon energy.[30] For UEM-2, recent theoretical work with the actual geometry, including lens optics, gives the spread indicated in the table without invoking the compression schemes or using the field-emission tip, as discussed in the text. The experimental femtosecond resolution is obtainable from EELS studies (Secs. 6.3.4.6 and 6.6).

consideration of the effect requires knowledge of the specific microscope design parameters (path length, lenses, and voltages involved).

It is important to understand the critical role of *frame referencing*,[6,28] not as simple subtraction of one image or pattern from another, but as an approach to follow the evolution of structures in time. The image or diffraction frame at time t can be referenced to another one at t', where t' could be a specific (arbitrarily chosen) time point in the evolution of change, including frames taken at negative times (the initial structure). This can perhaps best be illustrated by considering one of the first cases studied in UED, UEC, and UEM. Figure 6.11 gives the three structures involved in a chemical reaction. With frame referencing at two different time points, before and after the breakage of the first chemical bond, the disappearance of the second bond contribution to the diffraction (Fig. 6.11, lower panel) was clearly observed. Mathematically, one can therefore make these

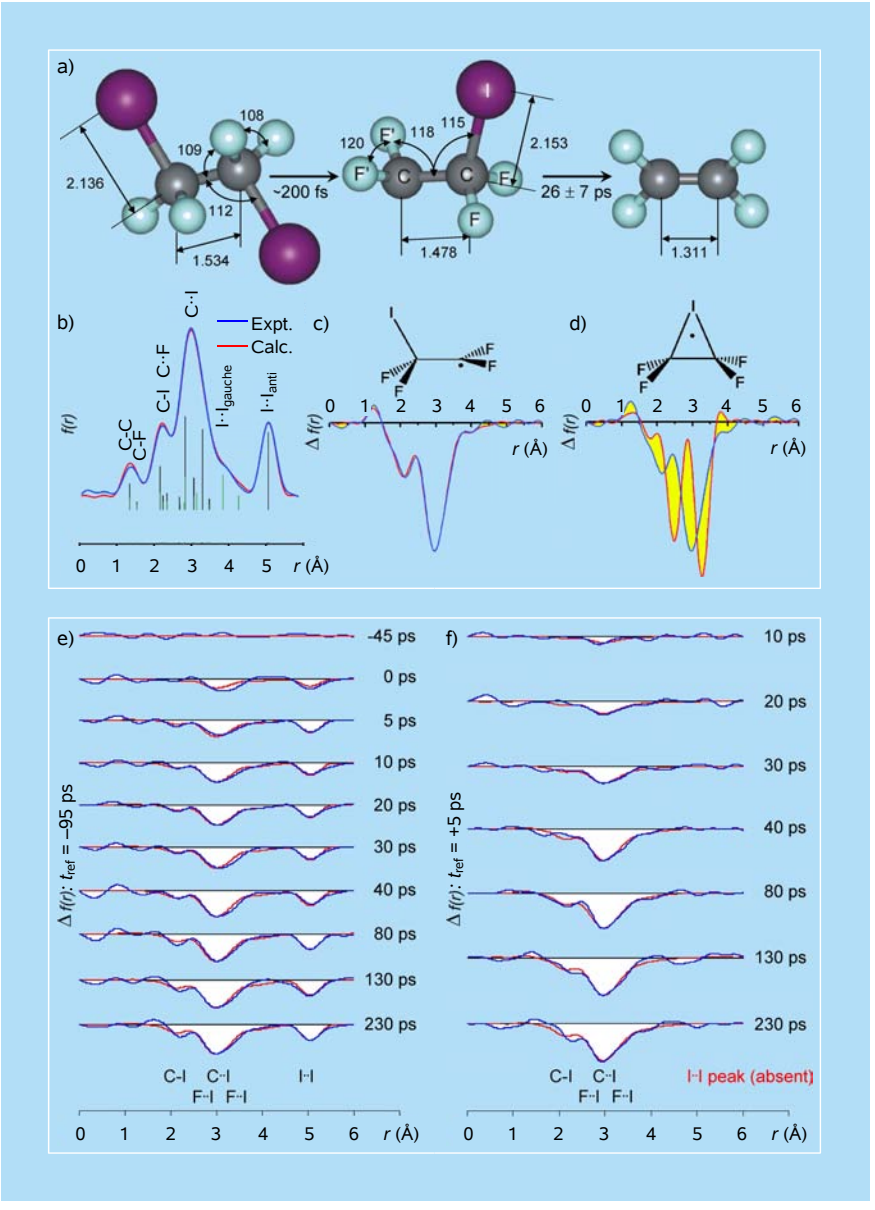


Figure 6.11 UED structural dynamics and *frame-referencing* methodology. Top: (a) Structural dynamics of the elimination reaction of $C_2F_4I_2$, with the reactant, intermediate, and product structures determined under collisionless condition. (b) UED patterns of the *anti-gauche* conformation mixture characteristic of the reactant as obtained for the ground state; internuclear distances for the *anti* (black) and *gauche* (green) isomers are indicated by vertical bars at the bottom of the panel. The structure of the intermediate is determined to be classical (c), not bridged (d), as evidenced by the agreement between diffraction theory and experiment (see Chap. 7 for studies in solution); the discrepancy between theory and experiment is indicated in yellow (c,d). Bottom: Frame-referencing methodology applied to the determination of structures in the elimination reaction of $C_2F_4I_2$. Temporal changes in diffraction with respect to the two different reference points, before (e) and after (f) the arrival of the exciting laser pulse, are shown. Note the absence of the I··I peak in (f).[29]

reference points meaningful to obtain transient structures and their percent contribution to the total (observed) diffraction.[6] The same methodology is applied in imaging, as shown in the following sections.

For materials, the ability to select, if sensitivity permits, a diffraction peak (spot) to study dynamics is of great value, not only because one can examine lattice expansions (spot position), structural inhomogeneity (spot width), and temperature (spot intensity), but also because one can selectively study motions along different directions of the lattice (structure factors), as discussed in Chaps. 2 and 5. Figure 6.12 gives a picture of how the changes of Bragg spots can be observed in the two dimensions of scattering (s) and time (t),[10] e.g., in UEC (Fig. 6.13), and Fig. 6.14 shows how these changes, in the case of the spot intensity, scale correctly with the order of diffraction, the s^2-dependence in the Debye–Waller factor (see Chap. 2).

Finally, selected-area image dynamics (SAID) provides a real advantage when seeking to unravel the intricacies of spatially inhomogeneous specimens. With selected-area imaging it is possible to study morphological, structural and mechanical dynamics of single sites (particles) with sensitivity — in seconds of recording — reaching at least one part in 10^6 when comparing with bulk

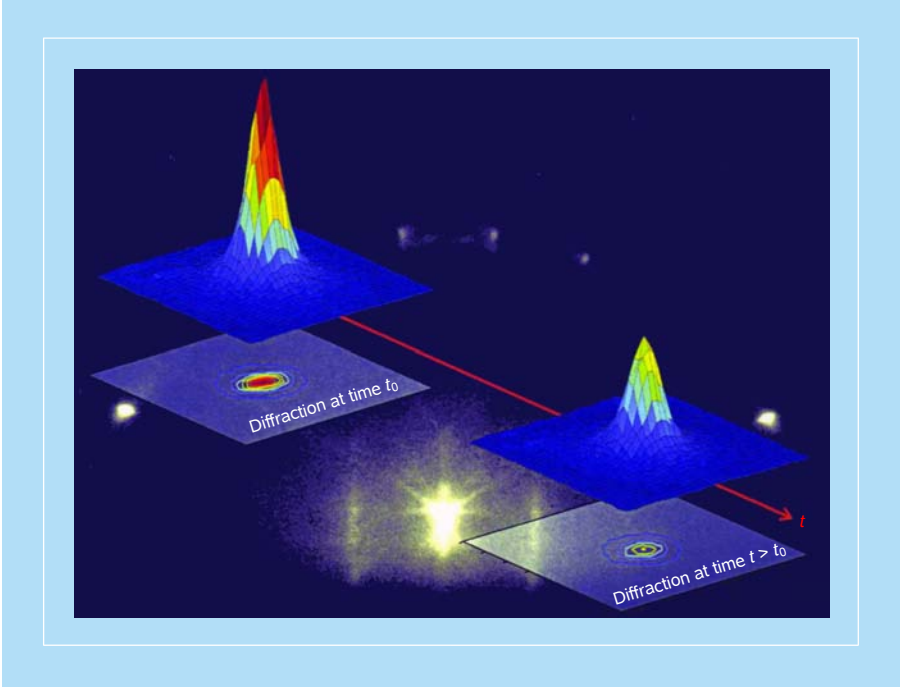


Figure 6.12 A Bragg spot change observed in the two dimensions of scattering (s) and time (t).[10a]

diffraction. The high spatiotemporal resolution and the dissipation of specimen heat between pulses are key elements for the success of this approach. As shown in the following section, the time scales of (ultrafast) structural heating and (slower) morphological deformations (and restructuring) are vastly different, suggesting that for nanoscopic materials or surface reactions the actual rates may not only be determined by the local temperature but also by allotropic changes.

The level of sensitivity reached in UEM, UEC, and UED, and SAID in UEM, together with the ability to control the temporal cooling of specimens, while maintaining the current spatiotemporal resolutions, have benefited greatly from technological advances in detection (CCD cameras), large-scale computation, ultrafast lasers, and electron optics of microscopy. Equally important is the conceptual framework realized — based on the power of quantum physics (uncertainty and coherence) — for developing single-electron imaging.

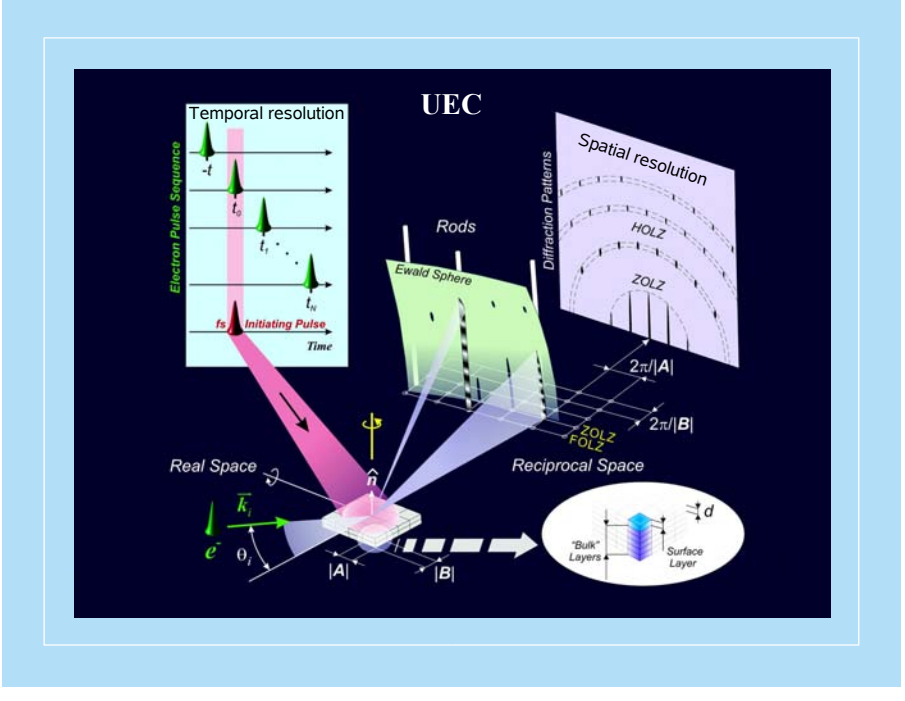


Figure 6.13 Schematic of the methodology of UEC in the reflection geometry. The relationships between real and reciprocal lattices and Ewald construction for the formation of diffraction patterns are shown (see Chap. 2). The temporal and spatial resolutions of UEC are achieved by varying the delay time between the electron probe and initiating heating pulses (upper-left panel), and the recording of the diffraction patterns at different times; transmission patterns can also be obtained in the same apparatus.[10a]

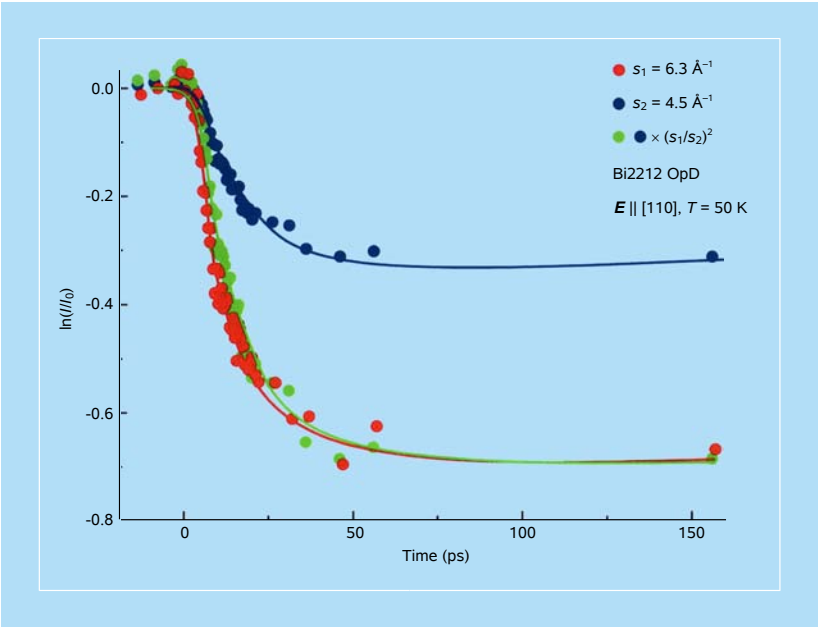


Figure 6.14 The intensity changes for two different Bragg spots, recorded for the same pattern, at different scattering vectors. In a time-resolved diffraction experiment, different Bragg spots at a given time should exhibit intensity changes in proportion to the scattering vector $\boldsymbol{s}$. Therefore, two distinct Bragg diffraction features appearing at $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$ should obey the scaling relation for a Debye–Waller change: $\ln[I(\boldsymbol{s}_1)/I_0]/\ln[I(\boldsymbol{s}_2)/I_0] = (\boldsymbol{s}_1/\boldsymbol{s}_2)^2$; see Eq. (2.50). These studies were made on cuprates,[12b] and in Sec. 6.4.4.4 similar results are shown for graphite. Adapted from Ref. 49.

6.3 Structure, Morphology, and Mechanics

6.3.1 Selected-area image (diffraction) dynamics

In the present section, we specifically discuss applications of UEM to the study of structural, morphological, and mechanical dynamics and their different time and length scales. Unlike measurements obtained from the whole of the specimen, microscopy images from a 2K×2K CCD camera provide up to four million simultaneous observations, each attributable to a single, independent location on the specimen (subject to point spread function limitations). As a result, the spatiotemporal changes, frame by frame, can be observed for all specimen points at once, the analog of parallel processing, without the spatial averaging associated with selected-area diffraction, for example. Moreover, structural and morphological visualizations are concurrently possible on the ultrashort time scale, using single-electron packets, and on arbitrarily-long time scales, using single pulses of ~10^5–10^7 electrons, with sufficient time between pulses to allow studies of both repetitive and irreversible processes on longer time scale. Both SAID and selected-

area-diffraction dynamics (SADD) can be obtained — for the former in real space (from contrast) and for the latter in Fourier space (from changes of the Bragg peak separations, amplitudes, and widths).

In order to quantify image changes, the following method of cross correlation is invoked. The normalized cross correlation of an image at time t with respect to that at time t' is expressed as[24]

$$\gamma(t';t) = \frac{\sum_{x,y} C_{x,y}(t) C_{x,y}(t')}{\sqrt{\sum_{x,y} C_{x,y}(t)^2 \sum_{x,y} C_{x,y}(t')^2}}, \tag{6.4}$$

where the contrast is defined as $C_{x,y}(t) = [I_{x,y}(t) - \bar{I}(t)]/\bar{I}(t)$; $I_{x,y}(t)$ and $I_{x,y}(t')$ are the intensities of pixels at the position (x, y) at times t and t', and $\bar{I}(t)$ and $\bar{I}(t')$ are the means of $I_{x,y}(t)$ and $I_{x,y}(t')$, respectively. The correlation coefficient $\gamma(t'; t)$ is a measure of the temporal change in "relief pattern" between the two images being compared, which can be used as a guide to image dynamics as a function of time. Two types of cross-correlation plots (frame referencing) are typically made: those referenced to a fixed image frame before $t = 0$ and others that show correlation between adjacent time points.

6.3.2 Dynamical morphology: Time-dependent warping

When a crystal is excited by absorbing energy from a heating light pulse, the lattice must undergo changes to accommodate the deposited energy. In the absence of a phase change in the material, the lattice will typically expand. The primary influence of this expansion on the level of the individual lattice unit cell, and the effect on diffraction, are illustrated in Fig. 6.15. Unconstrained thermal expansion of a high-symmetry lattice, such as a face-centered cubic (fcc) gold crystal, is isotropic and will not affect the orientation of any crystal zone axis. A different situation applies when a crystal's elastic properties are anisotropic, as in the case of the layered structure of graphite. Then, all diagonal zone axes, such as [011], will change directions, while others, such as [001], will not. These different geometric characteristics must be considered in the analysis of diffraction and of imaging, to the extent that image contrast depends on diffraction efficiency.

In addition to the above primary effect of expansion, change in zone axis alignment can also be induced in both isotropic and anisotropic crystals when boundary constraints cause warping or bulging of the film as a secondary response to the lattice expansion (Fig. 6.15). Because very small zone-axis misalignment is sufficient to produce pronounced asymmetry in a diffraction pattern, zone-axis changes are in turn reflected in Bragg peak properties, in diffraction, and in image contrast for features such as bend contours and contrast fringes.

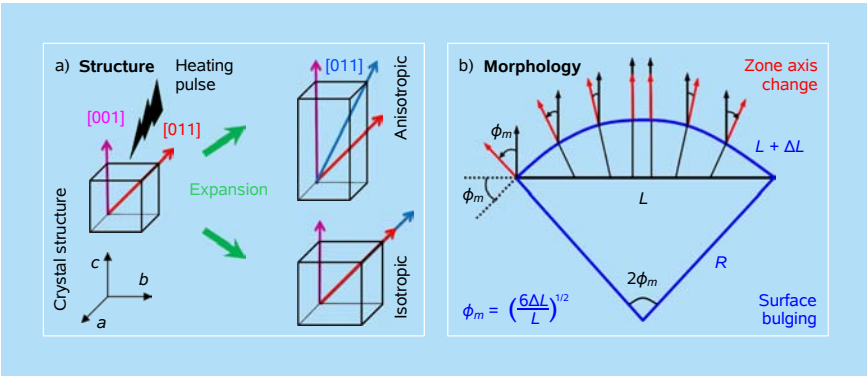


Figure 6.15 Temporal changes in (a) crystal structure and (b) film surface morphology induced by heating.[31]

Bulging of a thin film induced by expansion between fixed end points of an initially flat section of lateral dimension L can be modeled as shown in Fig. 6.15.[31] The surface after expansion is an arc of circumferential length $L+\Delta L$, subtending a central angle of $2\phi_{\mathrm{m}}$, where ϕ_{m} is the maximum tilt angle caused by the bulge. We then have $\sin\phi_{\mathrm{m}} = L/(2R)$ and $\phi_{\mathrm{m}} = (L+\Delta L)/(2R)$. Expanding $\sin\phi_{\mathrm{m}}$ to second order ($\sin\phi_{\mathrm{m}} \cong \phi_{\mathrm{m}} - \phi_{\mathrm{m}}^3/6$) allows us to immediately reach the approximation:

$$\frac{\phi_{\mathrm{m}}^3}{6} \cong \frac{\Delta L}{2R} = \frac{\Delta L}{L+\Delta L}\phi_{\mathrm{m}}, \tag{6.5}$$

which leads to the expression for the maximum angle:

$$\phi_{\mathrm{m}} \cong \sqrt{\frac{6\Delta L}{L+\Delta L}}. \tag{6.6}$$

In the limit of $\Delta L \ll L$, then

$$\phi_{\mathrm{m}} \cong \sqrt{6\Delta L/L} = \sqrt{6\alpha_{\mathrm{eff}}\Delta T_{\mathrm{eff}}}. \tag{6.7}$$

When thermal equilibrium is attained, α_{eff} becomes the coefficient of linear thermal expansion, but at ultrashort time the expansion is far from equilibrium and ΔT_{eff} is determined by electron and lattice temperatures, as discussed below. We note that the bulging is the result of boundary constraints and thus the relevant expansion is the lateral one. When the material is thick or on a substrate as typically used in UEC,[10a,28] the longitudinal expansion is prominent. The morphology effect, which may influence intensities of diffraction, is different from the well-known Debye–Waller description (see below).

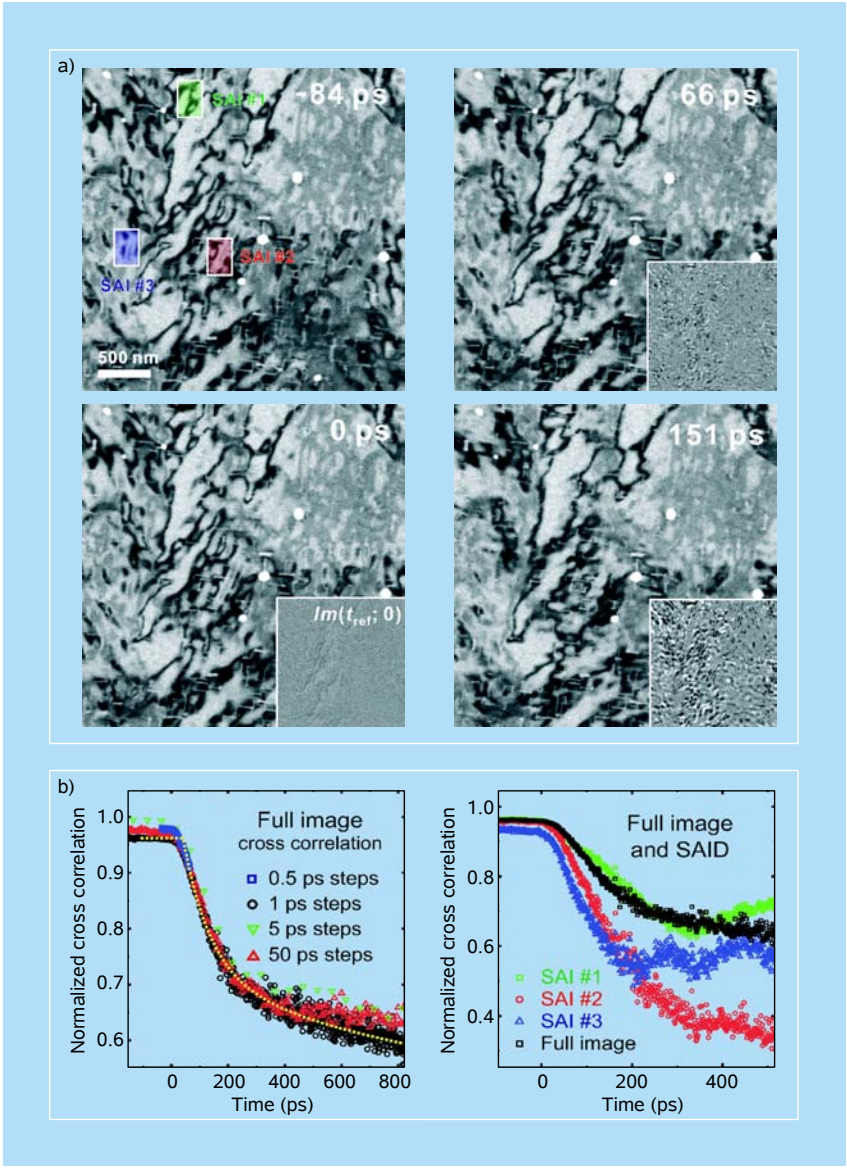


Figure 6.16 Representative real-space UEM image frames at different times and temporal change in UEM. Shown are electron images at four time delays after heating with the femtosecond pulse (fluence of 1.7 mJ/cm^2). In the images, the bend contours and holes in the sample show up as dark bands and bright white circles, respectively (a). The insets are image-difference frames with respect to the image taken at −84 ps. The specimen is a gold single crystal film mounted on a standard 3 mm 400-mesh grid with polymeric support film. The gold thickness was measured by electron energy loss spectroscopy in UEM-2. (b) The time dependence of image cross-correlations of the full image from four independent scans taken with different time steps. A fit to biexponential rise of the 1 ps step scan is drawn, yielding time constants of 90 ps and 1 ns (left). Time dependence of image cross-correlations at 1 ps time steps for the full image and for selected regions of interest SAI #1, #2, and #3, as shown in the −84 ps image, is displayed on the right.[24]

6.3.3 Proof of principle: Gold dynamics

With UEM, the results obtained for gold are shown as images (Figs. 6.16 and 6.17) and diffraction patterns (Figs. 6.18 and 6.19). In Fig. 6.19 two single-pulse selected-area-diffraction (SAD) patterns recorded at different time delays with respect to a high fluence nanosecond excitation pulse (40 mJ/cm^2) are presented; at this fluence, permanent damage is induced in the specimen so only single-pulse experiments are appropriate. The surface normal of the 11 nm thick gold film is parallel to the [100] zone axis, and the diffraction pattern at negative time displays a square array of Bragg reflections that are properly indexed by the fcc structure projected along the [100] zone axis at zero tilt angle. From their positions (interplanar spacings of 1.248 and 0.951 Å), the indicated transient spots measured at 50 ns after arrival of the excitation pulse are reflections from the {113} and {133} planes, forbidden in the [100] zone-axis viewing.

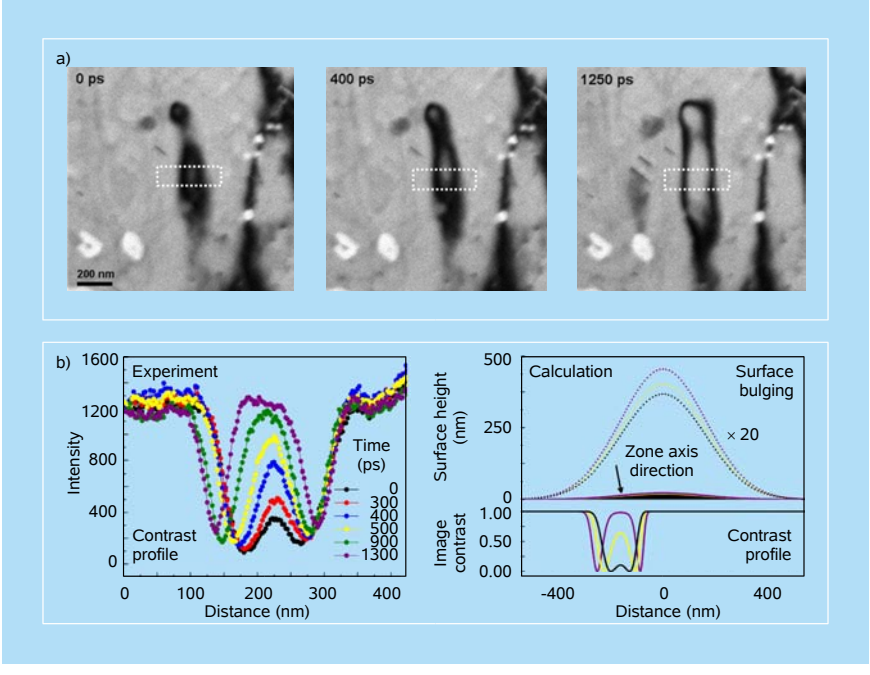


Figure 6.17 High-magnification nanoscale image changing with time. The sample was tilted at a 10° angle to the microscope axis and a femtosecond excitation of 1.7 mJ/cm^2 pulse fluence was used to induce the change. (a) Note the dramatic change in the penguin-like contrast with time, as discussed in the text.[24] (b) Experimental contrast profiles, for a series of different delay times, from the selected area indicated by the dotted white boxes in the images above. Also shown are hypothetical Gaussian surface bulge (360 nm FWHM) and corresponding calculated contrast profiles, at three bulge heights (18.9, 20.7, and 23.4 nm). Positions where the direction of the surface normal aligns with the indicated arrow have minimum contrast, and the contrast reduction has a Gaussian angular dependence of 1° FWHM about this direction.[31]

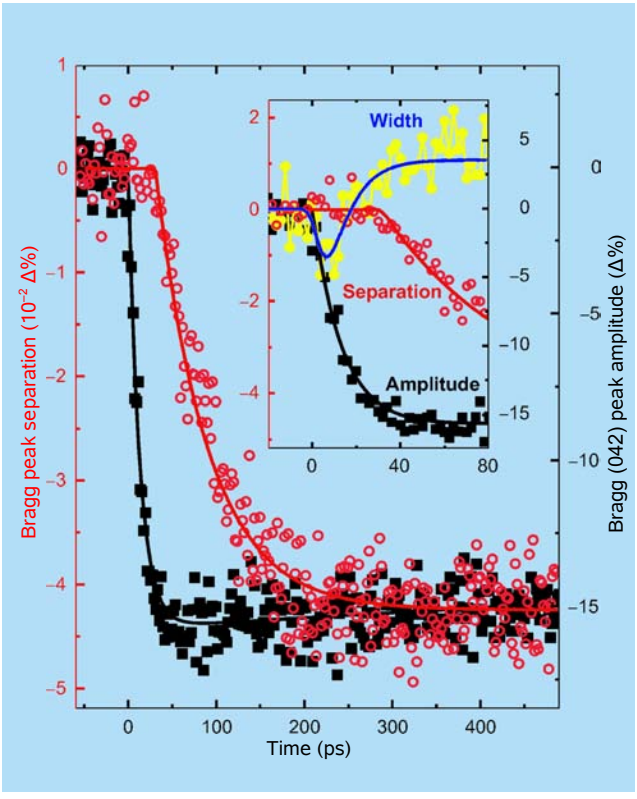


Figure 6.18 UEM Bragg-peak separation and amplitude change with time: SADD for femtosecond excitation at 1.7 mJ/cm^2 peak fluence (519 nm). The separation for all Bragg peaks and the amplitude of the (042) peak are shown in the main panel; the inset gives an expanded time scale, along with the width change for all peaks. The peak amplitude has been normalized to the transmitted beam amplitude, and the full width of diffracted peaks were calculated as excess contributions to the width with respect to the full width of the transmitted beam in each time frame. The time dependence of amplitude and separation is fit as an exponential rise, and a delay with rise, respectively.[24]

In order to reproduce the appearance of the forbidden peaks without laser heating, we need to tilt the sample by an angle of more than 10°. Therefore, we see unambiguously that the gold specimen has undergone severe heat-stress related distortion at the time of the diffraction exposure. The symmetric appearance of the new peaks also shows that tilts with all azimuthal orientations are produced, consistent with bulging of the surface. For a substantial contribution to the total diffraction to come from angles greater than 10° about any given tilt axis, ϕ_m for a bulging surface must have a considerably higher value. Thus, if $\Delta L/L$ is roughly proportional to heating fluence, even fluences as low as 0.4 mJ/cm^2 are expected to produce lattice tilts of greater than 1°. Such tilting may have a substantial effect on diffraction patterns, an effect that will, however, be moderated by angular averaging.

From ultrafast stroboscopic SAD patterns recorded at an excitation fluence of 1.7 mJ/cm^2, the average Bragg peak separations and amplitudes shown in Fig. 6.18 were obtained as a function of time. The average amplitude of all {042} diffraction peaks drops significantly, with the time constant of the exponential rise equal to 12.9 ps. The plotted Bragg peak separation is the average of the relative separation changes of all 20 planes observed in the [100] zone axis having $d \geq 0.85$ Å. It drops by 0.043% with a rise time of 60 ps and an onset delayed by 31 ps relative to that of the {042} amplitude. The average Bragg peak width rises because of the induced structural inhomogeneity, as expected,[10a] and near $t = 0$ an initial decrease is noted, which indicates that the overall temporal response is much faster than that of the amplitude or separation rise. The delay in the onset of separation change with respect to amplitude change is similar to the time scale for the amplitude to reach its plateau value.

In order to determine the recovery time of the structure, we carried out stroboscopic (and also single-pulse) experiments over the time scale of

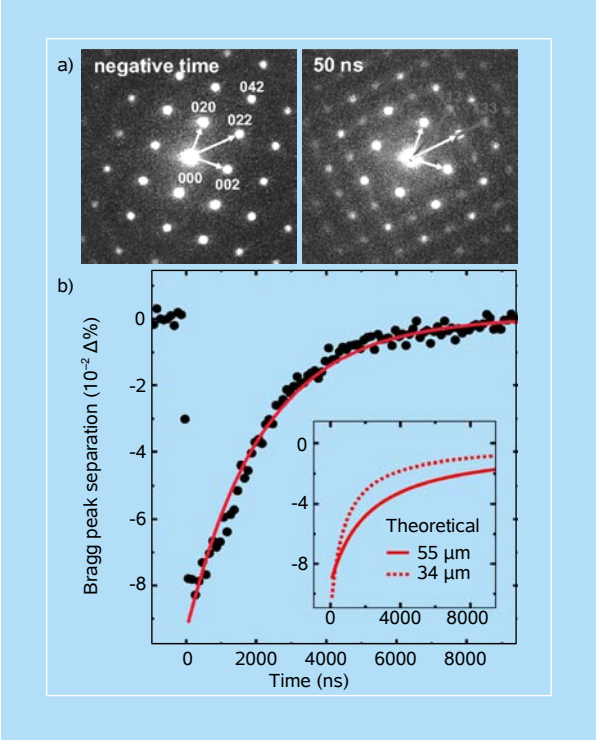


Figure 6.19 Structural recovery. (a) Diffraction patterns of gold as obtained using a single pulse of 6×10^5 electrons at high peak fluence (40 mJ/cm^2) and selected-area aperture of 25 μm diameter; two frames are given to indicate the change. (b) 2D lateral heat transport calculations are presented in the inset for two different laser heating spot sizes as indicated in the figure.[24,31] The single-exponential fit to the data (red) gives the microsecond scale for recovery (see text).

microseconds. In Fig. 6.19, the stroboscopically measured Bragg peak separation transient is plotted for nanosecond pulses at a fluence of 7 mJ/cm^2. An exponential fit gives a time constant for full recovery of the specimen to its prepulse condition of 2.2 μs, a time scale that is reproduced by calculations of 2D lateral heat transport (see inset of Fig. 6.19) in an infinite sheet,[22,24] with thermal conductivity of 3.17 W/(cmK) at 300 K and laser spot diameter of 60 μm. The exponential fit and the theoretical cooling transient are also plotted in the figure. The closer agreement of the experimental result to an exponential functional form (at longer times) may be attributed to the presence of a copper grid acting as a heat sink near the edge of the initially heated specimen area.

In the images, the change in morphology is evident. Shown in Fig. 6.16b are the cross-correlation values between the image at each measured time point and a reference image recorded before the arrival of the clocking pulse. The experiments were repeated for different time-delay steps (500 fs, 1 ps, 5 ps, and 50 ps) and similar results were obtained, showing that morphology changes are completely reversible and reproducible, retracing the same evolution over each 5 μs inter-pulse interval. The adjacent-time cross-correlations display the time scales for intrinsic changes in the images, which disappear for time steps below 5 ps, consistent with full-image rise in time. Over all pixels, the time scale for image change covers the full range of time delay, from picoseconds to nanoseconds, indicating the collective averaging over sites of the specimen; as shown in Fig. 6.16b the fit to the transients gives two prominent time constants of 90 ps and 1 ns.

The power of SAID is illustrated when the nanoscale dynamics of the bend contours are followed in different SAI regions, labeled in the image as SAI #1, 2, and 3. The corresponding image cross-correlations (Fig. 6.16b) have different shapes and amplitudes from one another and from the full-image correlation. The dramatic differences observed here and for other data sets include onsets delayed in time and sign reversals, which are indicative of the variation in local deformation dynamics. In Fig. 6.17, a time-resolved SAI at higher magnification is depicted. A broad and black “penguin-like” contour is observed as the dominant feature of this area. As shown in the frames, a colossal response to the femtosecond heating is noted. The gray region inside the black contour appears and broadens with time. Also, a new black contour above the large central white hole begins to be evident at 1200 ps and gains substantial intensity over the following 50 ps; all frames taken construct a movie of SAID. Here, in the UEM measurements, the changes in local tilt are transient in nature, reflecting the temporal changes of morphology and structure. Indeed, when the experiments were repeated in the TEM mode of operation, i.e., for the same heating laser pulse and same scanning time but with continuous electron probe beam, no image change was observed.

In Fig. 6.17, transient image contrast changes in a bend contour of gold are shown for stroboscopic femtosecond excitation at the fluence of 1.7 mJ/cm^2 with the

specimen tilted by 10°. Experimental contrast profiles across the changing contour are plotted at the bottom left for a series of different delay times. At the bottom right, a calculation of contrast profiles for a growing surface bulge illustrates the mechanism by which surface topology may contribute to image changes. The maximum diffraction loss, and thus minimum image contrast, occurs where the surface is normal to the indicated zone-axis direction, and the loss is given a full width at half height of 1° in alignment error. As the surface expands laterally and the bulge grows, the surface slope changes, the points of best zone axis alignment move, and the minimum contrast follows. The agreement of observed trend with the theoretical calculation supports the model of morphological dynamics.

The atomic-scale motions, which lead to structural and morphological changes, can now be elucidated. Because the specimen is nanoscale in thickness, the initial temperature induced is essentially uniform across the atomic layers and heat can only dissipate laterally. It is known that for metals the lattice temperature is acquired following the large increase in electron temperature.[32] The results in Fig. 6.18 give the temperature rise to be 13 ps. From the known electron and lattice heat-capacity constants [$C_1 = 70$ J/(m^{3}K^2) and $C_2 = 2.5\times10^6$ J/(m^{3}K), respectively] and the electron–phonon coupling [$g = 2\times10^{16}$ W/(m^{3}K)] we obtained[33] the initial heating time ($\tau_1 = C_1T_1/g$, where T_1 is the electron temperature) to be ~10 ps for $T_1 \sim 2500$ K, in good agreement with the observed rise. Optical reflectivity measurements do not provide structural information but they give the temperature rise.[34] We note that for the bulk material, the time scale for heating (~1 ps) is shorter than that of the nanoscale specimen (~10 ps) due to confinement in the latter, which limits the ballistic motion of electrons in the specimen,[34] and this is evident in the UEM studies.

Up to 30 ps the lattice is hot but, because of the macroscopic lattice constraint, the atomic stress cannot lead to changes in lateral separations, which are the only separations visible for the [100] zone-axis probing. However, the morphology warping change is correlated with atomic (lateral) displacements in the structure as it relieves the structural constraint. Indeed the time scale of the initial image changes seen in imaging experiments is similar to that of plane separations in diffraction (60–90 ps). Finally, the calculated Debye–Waller factor (see Eq. 2.50 and, e.g., Ref. 35) for structural changes gives a temperature of 420 K ($\Delta T = 125$ K), of the same order as lattice temperature derived under similar conditions, noting that for the nanoscale material the temperature is higher than in the bulk.[34] It follows that both zone-axis tilting and thermal lattice effects influence morphological and structural dynamics, and in this regard the separation of time scales is important for elucidating each contribution, as shown, for example, in Figs. 6.17–6.19, where we notice that the structural change precedes the warping because of the needed lateral restructuring. The value of ϕ_m, for $\Delta L/L = 0.178\%$, from Eq. (6.7), is equal to 5.9°.

6.3.4 Prototypical case: Graphite in 4D space

In this section, we present a case study that covers all four domains (Chap. 1) of UEM exploration of structures, morphologies, mechanics, and bonding. The system under consideration is the quasi-2D graphite material. The findings over the range of time scales (femtosecond to microsecond), length scales (angstrom to micron), and energy scales (0–50 eV) are lucid and strikingly reminiscent of the role played by the diatomic molecule of sodium iodide in the development of femtochemistry (Chap. 5); as such it represents a paradigm shift for 4D electron imaging.

6.3.4.1 *Atomic motions*

The layered structure of graphite is the key factor which determines its unique physicochemical properties. The structure and diffraction of graphite are shown in Fig. 6.20. When subjected to a shock, the structure distorts, and at high fluences ablates to form graphene (individual sheets of hexagonal sp^2-bonded carbons) as

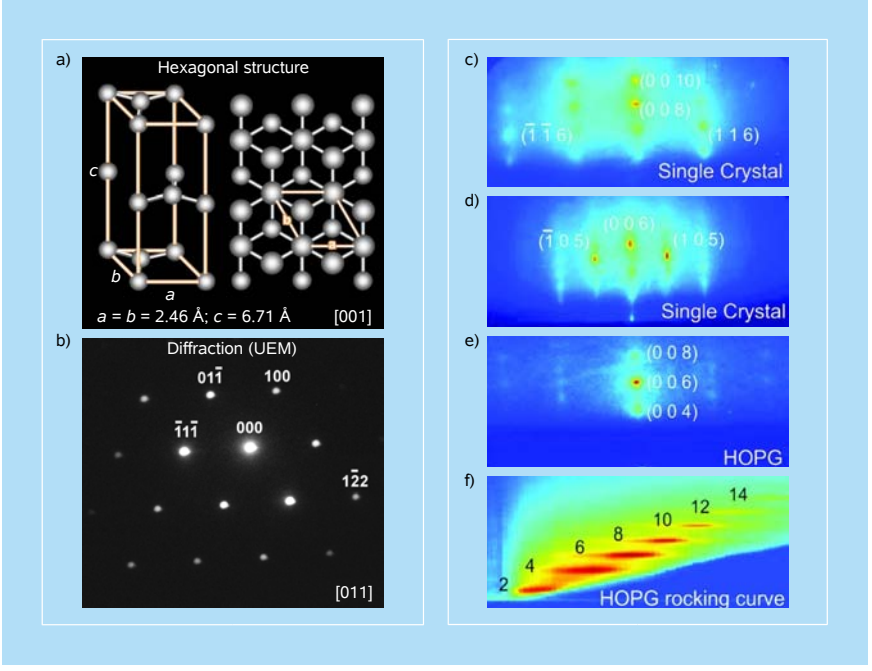


Figure 6.20 Graphite structure and diffraction patterns. (a,b) Hexagonal structure of graphite and its diffraction pattern as obtained in UEM. (c,d) Static electron diffraction (UEC) from a single crystal along two different zone axes. The lattice constants are a = 2.46 Å and c = 6.7 Å and all Bragg spots can be identified. (e) Diffraction pattern of a highly ordered pyrolitic graphite crystal (HOPG) also showing the direct attenuated nondiffracted spot (lower edge). (f) Rocking curve (θ dependence, in the range θ = 0 to 4°) for measuring the c-axis value of a highly ordered graphite crystal. For this material, c is 6.7 Å.[31,36]

a result of this distortion. Early optical studies focusing on the carrier dynamics induced by a laser excitation revealed features of phonon propagation and the coherent modes involved. The atomic-scale structural dynamics and the electronic structure of graphite, following an impulsive near-infrared excitation, was extensively studied with UEM. Moreover, UEC was also invoked[36] to probe the structure of the surface layers. First we discuss these experiments and then detail UEM and EELS studies.

The femtosecond laser pulse optically excites the π-system of graphite. Typical diffraction patterns obtained in UEC are displayed in Fig. 6.20. Following the laser excitation, Bragg spots of the diffraction patterns show unique changes in position, intensity and width, corresponding to changes in the crystal structure. The measured position of, e.g., the (0014) Bragg spot as a function of the time delay between the exciting laser pulse and the probing electron pulse, and at different excitation fluences, is displayed in Fig. 6.21a. The results depict an initial lattice compression on the 3 ps time scale followed by an expansion along the *c*-axis. The corresponding evolution of the spot intensity is displayed in Fig. 6.21c.

Notably, the compression and expansion processes display different dependencies on excitation fluence. The maximum of the compression occurs with progressive delay as the fluence deposited into the crystal is decreased, as indicated in Fig. 6.21a. The time constant of the expansion is not considerably affected by the laser fluence, and a common inflection point is noted at 3–4 ps, which marks a transition for the dynamics of processes involved: one occurring at early times and another at longer time, which results in the expansion. To account for an overall change of rates and the fluence dependence, a simple model describing population flow as a sum of two-rate processes for the compression and expansion was invoked (Fig. 6.21b).[36] It was found from the solution that the expansion has a constant rate of 7 ps, consistent with a direct exponential rise to the transient at long time. The expansion amplitude scales linearly with the fluence, suggesting that the excited carriers directly contribute to the nonthermal expansion observed in 7 ps. For the compression, the change of the rate with fluence can account for the observed trend in Fig. 6.21.

The initial step involves the anisotropic carrier excitation and select phonon dynamics, while the expansion is the result of a nonequilibrium lattice temperature. Indeed, the maximum observed expansion of the lattice corresponds to ~1.25% of the *c*-axis equilibrium value, 6.7 Å, which is beyond any thermal value. Given the thermal linear expansion coefficient for graphite, 7.9×10^{-6} K^{-1}, such change would correspond to a temperature jump of more than 1500 K (cf. the maximum expected temperature rise of ~40 K as obtained from the laser fluence and the heat capacity of graphite). Moreover, what we expected to observe was the expansion and its increased amplitude with fluence, but what we actually saw was, in addition, the contraction which is significant in describing structural instabilities and ablation of

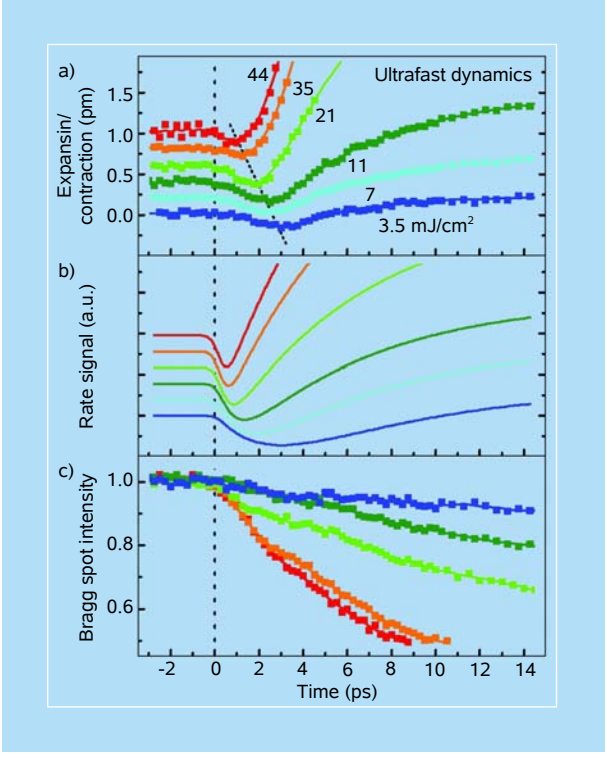


Figure 6.21 Ultrafast diffraction of graphite. (a) Time dependence of lattice compression or expansion for different excitation fluences in UEC. The curves are vertically shifted for better visibility, and the given scale is for the 3.5 mJ/cm^2 case. (b) The results of a dynamic model involving two processes: expansion driven by the fraction of energy delivered to the lattice, and a fluence dependent rate constant describing the early time compression. The characteristic time of the compression is found to range between 0.2 ps and 2.5 ps at the lowest and highest fluences, respectively. The time constant for expansion is 7 ps but the amplitude increases as fluence increases. (c) Time dependence of the (0014) Bragg spot intensity. This spot becomes the most intense when θ is optimized.[36] In Sec. 6.4.4.4, further information is provided.

graphite. Finally, the changes in the Bragg-spot intensity, Fig. 6.21c, are consistent with the behavior of the shift; observed is an increased intensity depletion as fluence increases. The spot width (not shown) increases with time, reflecting the expected dynamical inhomogeneity of the lattice. We shall return to these features of diffraction toward the end of this chapter in Sec. 6.4.4.4.

The mechanism can now be elucidated. With impulsive laser excitation the electronic bands are populated anisotropically and, because energy and momentum have to be conserved, these carriers excite the large-momentum lattice vibrations. Such selective coupling is known to lead to an efficient excitation of a subset of phonons, or so-called *strongly-coupled optical phonons* (SCOP), which ultimately decay to yield other vibrations. Because of electron–electron

scattering, the initial femtosecond carrier excitation acquires a broadening in the energy distribution. This distribution relaxes by electron–phonon coupling, leading to SCOP in the case of graphite. The time scale has been estimated to be on the order of 500 fs, for the electron–phonon-coupling process, and 5–7 ps for the decay of SCOP into other modes (see Ref. 36 and references therein). The resulting coherent oscillations are shown in Fig. 6.22 and display the near-1-ps period of the modes involved. The initial compression of the interplanar distance, which occurs with a characteristic time as fast as 500 fs depending on the laser fluence, suggests that such a phenomenon is mediated by the cooperative motion induced by the phonon subset, prior to their decay in 5–7 ps, and the out-of-equilibrium electronic structure driven by the potential of the excited carriers.[10] Because of the band anisotropy of the electronic structure, the electrons excited from the π to the π^* band preferentially weaken the *c*-axis bonding, which causes the observed

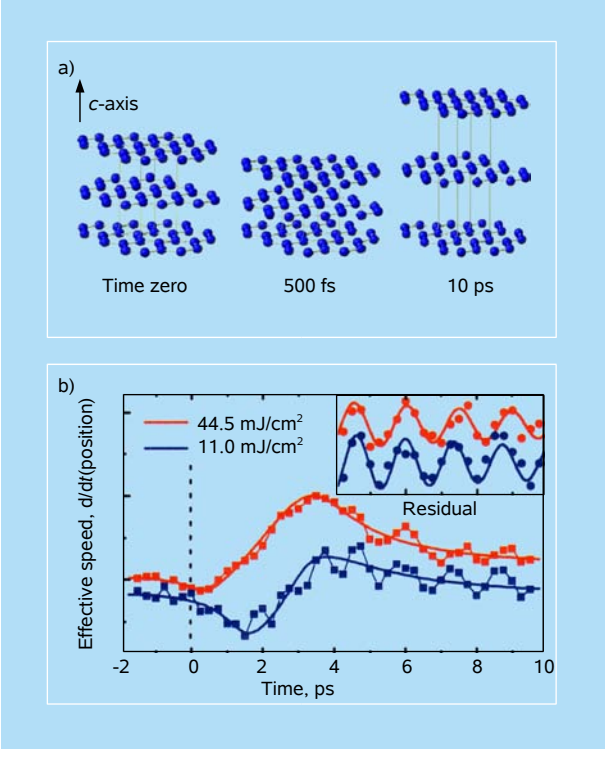


Figure 6.22 The coherent lattice oscillations and atomic motions of graphite at different times. The schematic (a) displays the movements of planes, exaggerated for clarity; the time scales in this graph correspond to the observation at high laser fluences. The time scales at different fluences can be seen in Fig. 6.21. The derivative as a function of time of the (0014) Bragg-spot position is displayed in panel (b) for the 44.5 mJ/cm^2 (red) and 11.0 mJ/cm^2 (blue) excitation fluences. The blue trace is magnified by a factor of six for display. The inset shows the residual of an exponential fit to the data between 4 and 10 ps.[36]

c-axis dynamics (see Fig. 6.22). It is the fs-to-ps-scale collective carrier and phonon dynamics that determines the extent of layer collisions and ultimately leads to graphene sheet ablation. These structural compression/expansion dynamics are also evident in more recent optical and STM studies of graphite[37]; see Sec. 6.4.4.4.

6.3.4.2 *Coherent resonances in diffraction: Longitudinal Young's modulus*

The visualization of resonances caused by coherent atomic motions is illustrated by the observations made in a series of femtosecond UEM diffraction experiments carried out on graphite. An electron diffraction pattern captured at negative time, i.e., before arrival of an excitation pulse, in UEM mode is displayed in Fig. 6.20 together with the graphite unit cell and its hexagonal lattice structure. The crystal structure corresponds to the space group $P6_3/mmc$, having a hexagonal unit cell with lattice dimensions of $a = b = 2.46$ Å and $c = 6.71$ Å. Stacked layers of 2D graphene sheets, where the atoms of each sheet are covalently bonded (in-plane bond length 1.42 Å), are weakly held together by van der Waals forces with a separation between sheets of 3.35 Å. The resulting anisotropy in the physical properties of graphite provides a strong contrast with the isotropic properties of gold. For example, the interplanar or *c*-axis of the graphite lattice is known to have a very large coefficient of thermal expansion, $\alpha_c = 25\times10^{-6}\ \mathrm{K}^{-1}$, whereas the corresponding value for the (*a*, *b*) plane is small, and even negative: $\alpha_{(a,b)} \cong -1.5\times10^{-6}\ \mathrm{K}^{-1}$. The geometric consequence of heating in graphite is therefore represented by the example of anisotropic expansion given in Fig. 6.15.

With UEM, diffraction experiments on graphite were performed for both tilted and untilted geometries at excitation fluences of 1.7 to 9 mJ/cm^2. For the femtosecond diffraction pattern of Fig. 6.20, recorded at negative delay time, the specimen was tilted at a 19.4° angle to the microscope axis, and the diffraction spots were indexed along the [011] zone axis, which has a nominal tilt angle from the crystal *c*-axis ([001] zone axis) of 20.1°, with tilt axis along the (000)−(100) direction. Figure 6.23 gives the temporal change, following a 7 mJ/cm^2 fs excitation pulse, of the separation (from the transmitted beam) and of the integrated peak intensity, normalized by the integral of the transmitted beam, of the $(\bar{1}1\bar{1})$ Bragg reflection (see Fig. 6.20b). Clear, regular oscillations occur in both separation and intensity, which when fitted give a very well-defined oscillation period of 20.6 ps (±100 fs). Oscillations of a similar nature but of longer period were previously reported for a specimen of different thickness,[24] and as shown in Fig. 6.24, the period does change with thickness.

The intensity integral of the $(\bar{1}1\bar{1})$ Bragg spot displays a huge increase of up to 100%. The behavior of the spot position indicates an apparent average decrease in plane separation in a simple Bragg law interpretation. Unlike the Bragg spot changes discussed above for gold, the response seen here does not conform at first

glance to expectations from lattice heating, since the direction of the intensity change contradicts the Debye–Waller effect and the direction of Bragg angle change is opposite to that of thermal expansion. Moreover, the intensities and separations of other Bragg peaks in the same graphite experiments show a wide variety of behaviors, with some measurements of each quantity increasing and others decreasing after excitation.

These observations can be rationalized by considering the zone-axis (and structural) change that anisotropic lattice expansion causes for the tilted geometry

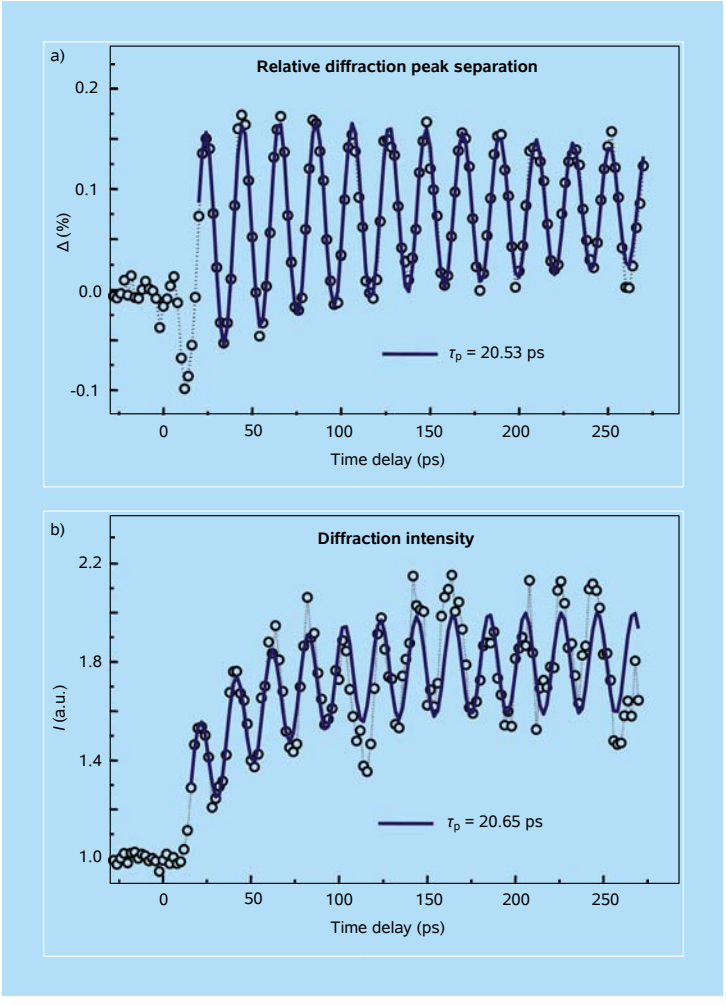


Figure 6.23 Resonance dynamics in graphite diffraction. Temporal change of the intensity integral (b) and separation of the graphite $(\bar{1}1\bar{1})$ Bragg spot (a) indicated in the diffraction pattern in Fig. 6.20. The integral was normalized by that of the transmitted peak. The oscillation periods are as indicated. An EELS log ratio of 0.258 was measured for the specimen at zero degrees tilt, yielding a thickness of 39.2 nm for $\bar{\lambda} = 152$ nm. The excitation fluence was 7 mJ/cm^2.[24,31]

of the experiments, as illustrated in Fig. 6.15, and more details can be found in Ref. 31.

In order to confirm that the period scales with thickness and derive the elastic properties of the nanofilm, we repeated the experiments, but for diffraction peaks of a number of different specimen areas. The thicknesses of these areas has been determined by the EELS log ratio: $\ln(I_t/I_0)$, where I_0 and I_t are the integrated intensities of, respectively, the EELS zero-loss peak and the total spectrum. A plot of oscillation period versus EELS log ratio for all these data sets is displayed in Fig. 6.24, revealing a linear relationship. The line in the figure is a weighted linear fit of the measured data points, and the slope of the line is 76.1 ± 2 ps. Since the thickness d is directly proportional to the EELS log ratio, with the constant of proportionality being the mean-free path of inelastic scattering ($\bar{\lambda}$), the data indicate a linear dependence of the oscillation period on thickness. This dependence is expected only in the case of oscillations due to a resonant modulation in film thickness (d), for which the period is given by the round-trip time for propagation of acoustic waves between the two film surfaces, in a "cavity-type" behavior:

$$\tau_p = \frac{2d}{\upsilon}, \tag{6.8}$$

where υ is the velocity of sound in the material. The plot of Fig. 6.24, therefore, definitively establishes the nature of the dynamics underlying the observed oscillations.

The Young's modulus (Y) of the c-axis (c_{33} elastic strain modulus) can be derived from the oscillation period through the relation:

$$\upsilon = \left(\frac{Y}{\rho}\right)^{1/2}, \tag{6.9}$$

yielding

$$Y = \frac{4\rho d^2}{\tau_p^2}, \tag{6.10}$$

where ρ is the density. A quantitative evaluation of υ and thus of Young's modulus from the oscillations depends on the value of $\bar{\lambda}$. To determine $\bar{\lambda}$ for our EELS measurements of graphite at 200 kV and collection semi-angle of 7.3 mrad, we rely on measured $\bar{\lambda}$ or σ (scattering cross-section) values under different conditions: σ is related to $\bar{\lambda}$ by

$$\bar{\lambda} = \frac{Am}{\rho\sigma}, \tag{6.11}$$

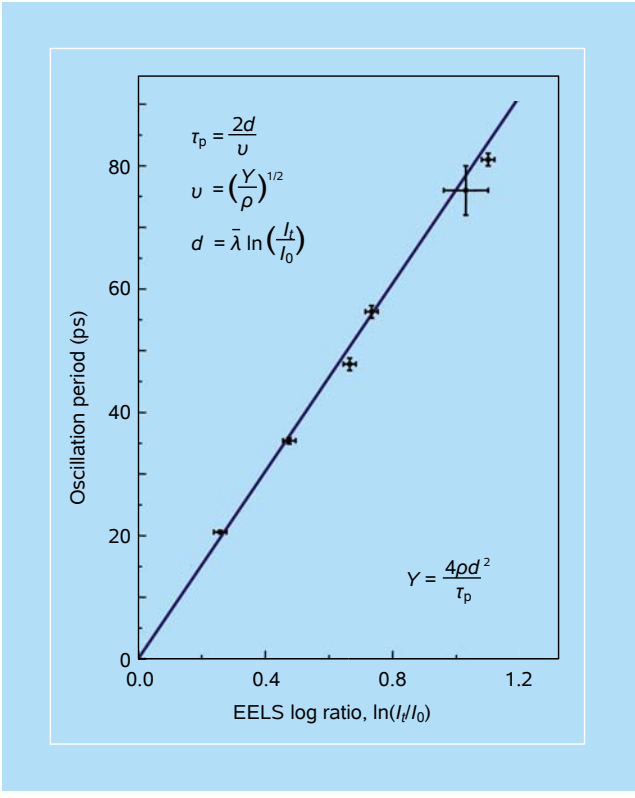


Figure 6.24 Resonance period dependence on specimen thickness. The periods of oscillation of Bragg spot properties from different specimen areas are plotted against the corresponding measured EELS log ratios. A linear fit yields the blue line with a slope of 76.1 ± 2 ps. The thickness of the film is equal to the EELS log ratio multiplied by 152 nm, the mean-free path of inelastic scattering. Thus, the coherent oscillation period scales linearly with the film thickness.[31]

where A is the atomic weight and m is the atomic mass unit. The slope of the linear fit in Fig. 6.24, together with the value of $\bar{\lambda} = 152$ nm (further details are given in Ref. 31), allows us to relate τ_p to d and obtain a sound velocity υ from Eq. (6.8) of 3990 m/s. For a graphite density of $\rho = 2.26$ g/cm^3, Y equals 36.0 GPa for the c-axis strain, quite close to the value of 36.5 GPa reported in the literature for studies of specimens of macroscopic thickness, and well within the reported range of 20–47 GPa for different graphites.

6.3.4.3 *Resonances in images: Longitudinal elasticity*

The mechanism for diffraction change through c-axis oscillation of graphite thin films, as described in the previous section, can also be expected to produce changes in UEM images, since image contrast effects are closely tied to variation

in diffraction efficiencies. One example of the detection of the graphite thickness oscillation in UEM images is shown in Fig. 6.25; others will be seen below in the discussion of moiré fringes and mechanical drumming.

The dark-field (DF) image shown at lower left is obtained by stroboscopic femtosecond UEM of a nanometer-thick specimen, at a tilt angle of 21° ([011] zone axis), selecting the $(1\bar{2}2)$ diffraction spot (Fig. 6.20b) by the objective aperture. A time series of such images was recorded, and three frame-referenced images constructed from this series are shown, representing the image change at the intervals indicated. Presented in Fig. 6.25 is a plot of the time dependence of selected-area image intensity for 20 different 1×100 pixel image areas, all within the red box shown on each image. The reversal in contrast for a time change of 30 ps is apparent in the different images, and the intensity plots give an oscillation period of ~55 ps. The period derived from diffraction measurements on the same area was 56.3 ± 1 ps. The fact that the period of the diffraction is the same as that

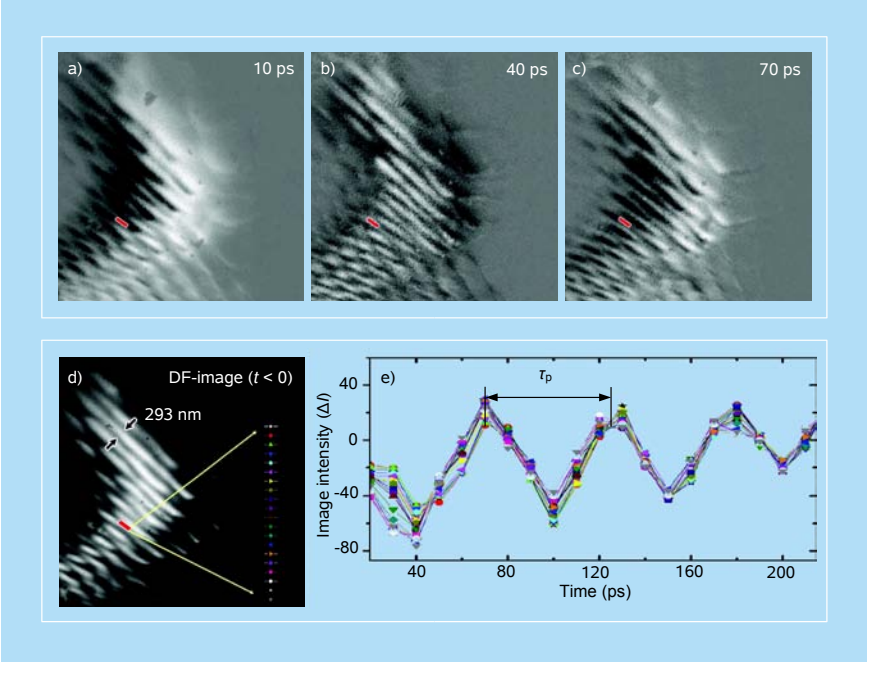


Figure 6.25 Resonance dynamics in graphite images. Dark-field (DF) image from a femtosecond UEM time scan (negative times; d) and three difference images constructed from images acquired in the same scan at the indicated times (a–c). The DF images were obtained by selecting the Bragg $(1\bar{2}2)$ spot. Contrast fringes due to physical buckling are present. The red boxes indicate the selected areas used for intensity plots (d,e). Also shown is time dependence of intensity difference (DF) for 20 adjacent selected areas of 1 × 100 pixels within the red boxes highlighted in the images (e). Each color in the intensity plot corresponds to a distinct selected area that is a 1 × 100-pixel slice parallel to the contrast fringes in the images. The intensities were derived from the referenced images for a constant time delay of 30 ps.[24,31]

of the image confirms the common origin of the two dynamic observations and provides the longitudinal Young's modulus directly from the contrast change of the image.

6.3.4.4 *Emergence of mechanical drumming: Transverse ellasticity*

In addition to the *c*-axis resonance, another form of resonance behavior has been observed in graphite, with a much longer period of oscillation. Because of the time scale involved, nanosecond imaging was used for the detection of the resonance.[23] Although also evident in both diffraction and image dynamics, we discuss here only the image measurements presented in Fig. 6.26. Typical images of a time series are shown in panels (a) and (b) of the figure. All images were constructed stroboscopically at 5 kHz excitation frequency, and therefore with a recovery time between pulses of 200 μs. The time evolution of each of the selected-area images indicated in Fig. 6.26c is traced in three different time regimes, in the form of normalized image cross-correlation [see Eq. (6.4)]. The cross-correlation for the full image is also plotted in black for each time regime.

The time scale for image change covers the full range of time delays, from 10 ns after impulsive heating to hundreds of microseconds. In regime I, for $t < 5$ μs, the image cross-correlation changes rapidly in a complex ("chaotic") fashion, with a great deal of variation among the selected areas. Beyond the first 10 μs, in regime II for example (25–29 μs), the cross-correlation change begins to exhibit periodicity, but with distinct patterns in different areas of the specimen. Finally, after several tens of microseconds, a well-defined global resonance oscillation emerges (regime III). The shape of these image cross-correlation dynamics was robust at different fluences from 2 to < 10 mJ/cm^2, but with varying amplitude.

The decay of the oscillation, $Q/(\pi \cdot f_0) = 44$ μs, as derived below, where f_0 is the resonant frequency and Q is the cavity quality factor, is not limited by cooling but rather is determined by the damping of mechanical motions. The overall decay of the transients is complete on a time scale shorter than the separation between pulses, as evidenced by the lack of modulation at negative time in the recorded transients. Given the nature of the free-standing specimen, cooling of the film must occur by lateral heat transport. With an initial *z*-independent heat profile by absorption of the heating pulse in graphite, we estimated, using a 2D heat diffusion in a homogeneous medium,[22,24] the time scale for an in-plane transfer, with thermal conductivity of 704 W/(m K) [cf. ~5000 W/(m K) for graphene] for natural graphite flakes, ρ of 2.26 g/cm^3, and specific heat c_V of 700 J/(K kg). For the radius at half-height of the initial pulse heat distribution $r_0 = 30$ μm, the time for the axial temperature to drop to a half of its initial value, $t_{1/2}$, is equal to ~720 ns (cf. ~100 ns for graphene), much shorter than the duration of the observed dynamics.

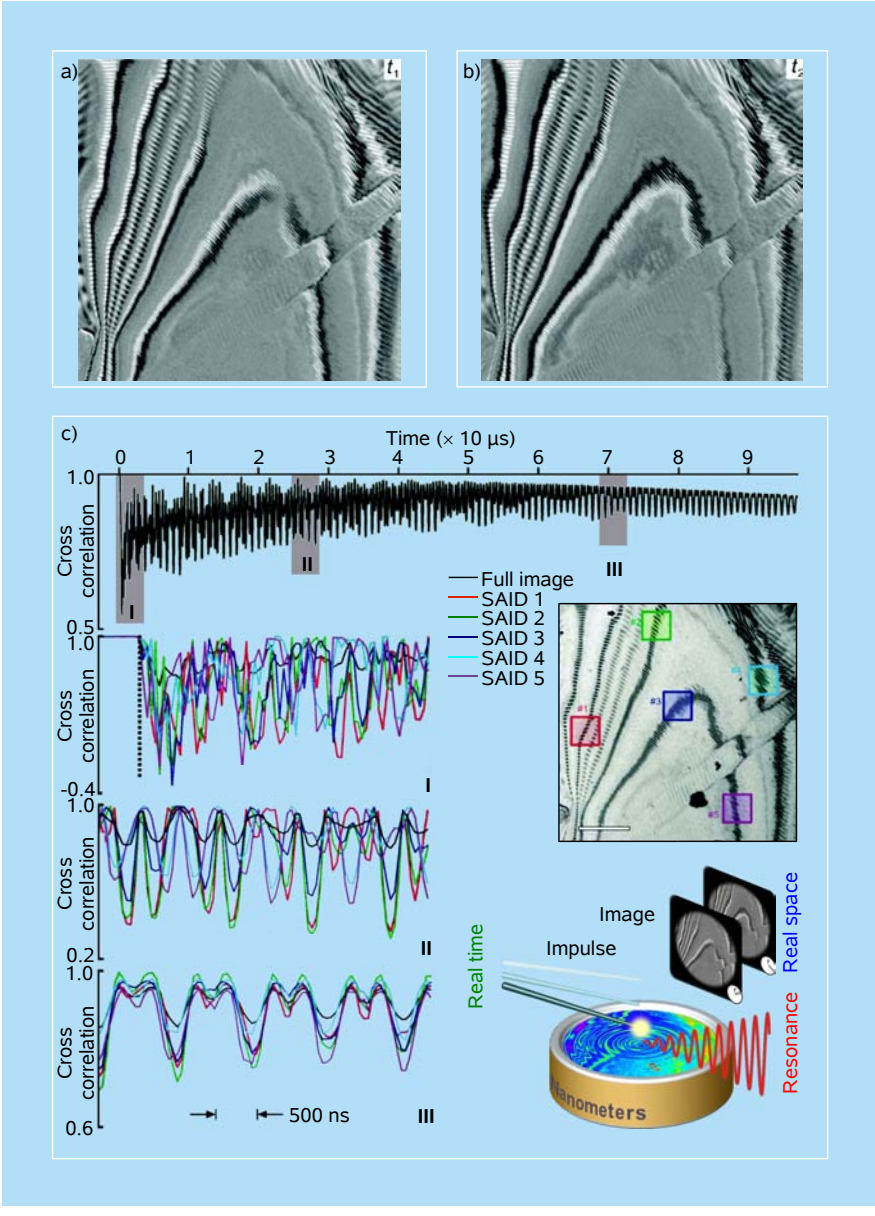


Figure 6.26 Drumming dynamics in graphite images. (a, b) Difference frames with respect to an image taken at negative times, which show clearly the image contrast change in time. The reversal in contrast displays the oscillatory (resonance) behavior. (c) The whole scan for 100 μs is made of 2000 images taken at 50 ns steps. Also depicted are the zoomed-in image cross-correlations of three representative time regimes (I, II, and III). In each zoomed-in panel, the selected-area image dynamics of five different regions are included. Note the evolution from the complex ("chaotic") to the global resonance (drumming) behavior at long times.[23]

When the specimen absorbs an intense pulse of energy, the lattice energy, converted from carrier (electron) energy by electron–phonon coupling in a few picoseconds (as discussed above), builds up in the illuminated spot on the surface within the duration of the laser pulse. As a consequence, the irradiated volume expands rapidly on the time scale of tens of picoseconds. The resulting thermal stress can induce mechanical vibration in the material, but a coherent oscillatory behavior will only emerge in the image if the impulsive stress is on a time scale shorter than the period. To resolve the full dynamics, probing of images should be over the entire time scale of the process, in this case 100 μs.

The frequency of the asymptotic global resonance was determined by Fourier transformation of the temporal evolutions of cross-correlations and selected-area image intensities. The dominant peak yields a resonance frequency f_0 of 1.08 MHz and a cavity quality factor ($Q = f_0/\Delta f$) of 150 ± 30. Normally, such resonances are strong in amplitude at short times. Here, we observe the rise of a resonance with time, and two cases are relevant. If the plate modes of high frequencies decay faster than those of low frequencies, then it is possible that at longer times only the lower-frequency modes would survive. However, it is quite possible that nonlinear effects will lead to such emergence because of the anharmonic nature of modes. For the latter, the process would be analogous to the so-called "mode locking" in lasers, for which nonlinear interactions (e.g., through refractive-index change with intensity) bring the modes in phase; see Chap. 5.

Mechanical modes have well-defined shapes, and their frequencies depend on elasticity of the material. For a theoretical model, we considered a square mechanical resonator clamped at three edges, since the film in question had one free edge. Such a resonating plate has a fundamental resonance mode of f_0, which is given by

$$f_0 = A\frac{d}{L^2}\left[\frac{Y}{(1-\nu^2)\rho}\right]^{1/2} + f(T), \tag{6.12}$$

where $f(T)$ due to tension T is taken to be zero in this case. Here, Y is the Young's modulus for the lateral dimension, ν is the Poisson's ratio, L is the dimension of the square, d is the thickness of the plate, and A is a constant equal to 1.103. We measured d to be 76 nm from EELS in our microscope ($\bar{\lambda} = 152$ nm). With $\nu = 0.16$ for graphite and $L = 40$ μm, the dimension of the vibrating area, we obtained from the observed resonance frequency the Young's modulus to be 0.94 TPa, which is in good agreement with the in-plane average value of 1.02 TPa obtained using bulk stress–strain measurements. This value is more than an order of magnitude greater than the *c*-axis value given above. The observation of the striking emergence of drumming modes suggests other studies of the phenomena, both from the experimental and theoretical perspective.

6.3.4.5 *Moiré fringe dynamics*

With the UEM spatiotemporal resolutions it was possible to visualize in images a diffraction contrast phenomenon, that of moiré fringes, and its dynamics. In Fig. 6.27, two representative moiré fringe images of graphite are shown, which were taken by using the DF-imaging method for better contrast. From these and other measurements made at various points of the specimen, fringes with spacings from 2.6 to 107.5 nm were found. It is known that rotational moiré fringes result from the misorientation between two overlapping crystals or two layers. In the case of rotational moiré fringes, the rotation angle θ between two misoriented layers is derived from the measured spacing S by the following equation:

$$S = \frac{d}{2\sin(\theta/2)} \cong \frac{d}{\theta}. \tag{6.13}$$

For the interplanar spacing d of the $(\bar{1}1\bar{1})$ plane of graphite (2.03 Å), the misorientation angle versus fringe spacing given by Eq. (6.13) is plotted in panel (c) of Fig. 6.27. In the inset of that panel, the formation of optical rotational moiré fringes is illustrated, defining S, d, and θ for this case.

The fringes in the BF image are formed due to the interference between the intense undeviated beam (transmitted beam) and relatively weak doubly diffracted beams that have undergone one scattering event in each of the misoriented layers. The fringes in the DF image arise from the interference of the electron waves from primary diffracted spots from each of the two layers having roughly equal intensity. Therefore, the DF method provides better contrast than the BF one since only the beams (planes) producing the fringes are represented, which is especially significant at higher magnification.

The time-resolved imaging experiments were carried out by stroboscopic femtosecond UEM for a specimen area with moiré fringes, which yielded the diffraction pattern of Fig. 6.20b and the diffraction dynamics of Fig. 6.23. Both BF and DF image time scans were recorded for selection of the (000) and $(\bar{1}1\bar{1})$ diffraction peaks, as indicated in Fig. 6.20b. The BF and DF images of Fig. 6.27 are frames from these scans at negative time, i.e., before the laser excitation. The average moiré fringe spacing in the indicated boxed areas is about 108 nm, corresponding to a rotation angle of 0.11° by Eq. (6.13). This rotation angle is too small to enable us to resolve two spots in the diffraction pattern (Fig. 6.20b). (The different spacings in DF and BF images may be attributable to a change in the specimen between experimental runs; see below.) To measure the spacing of moiré fringes, we constructed the profiles representing the transversely-averaged intensities along the length of the regions outlined by the white rectangles in both the BF and DF images of Fig. 6.27.

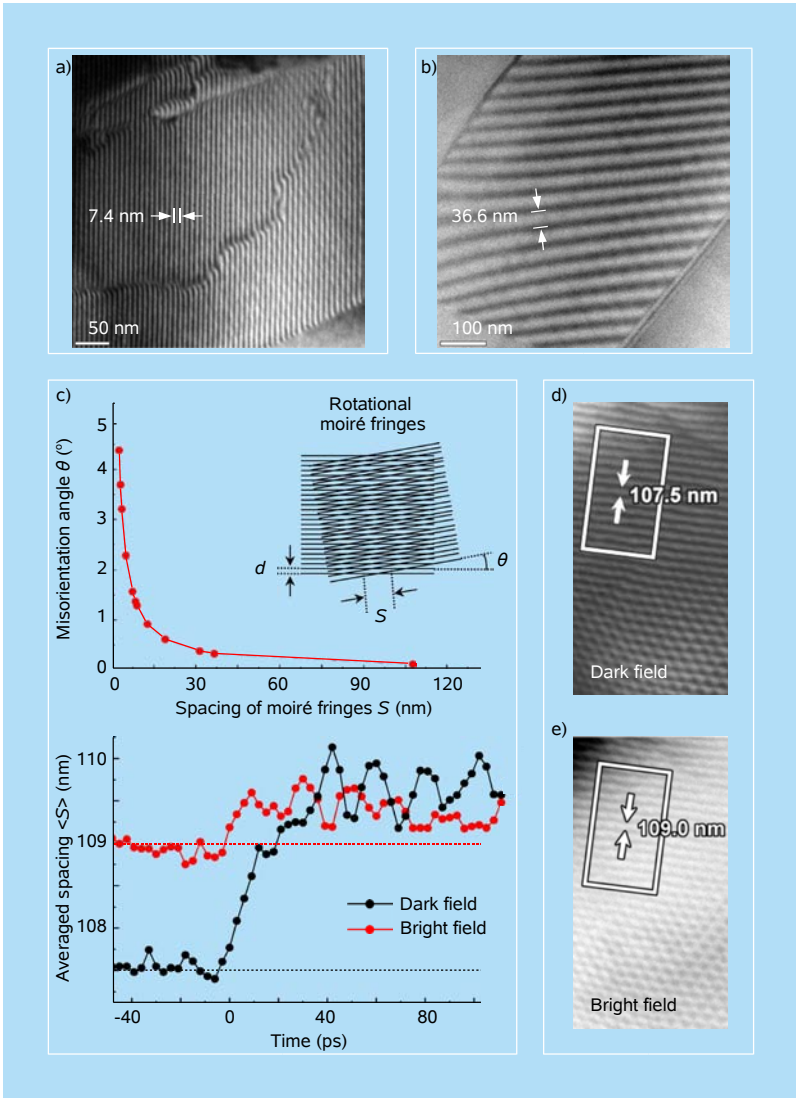


Figure 6.27 Moiré fringe dynamics in graphite. (a,b) Two examples of moiré fringes in graphite. (c) Rotational moiré-fringe spacing versus misorientation angle plotted for the $(1\bar{1}1)$ direction in graphite (interplanar spacing d = 2.03 Å; upper panel) and time dependence of the moiré-fringe spacing as obtained stroboscopically with femtosecond heating pulses (laser fluence: 7 mJ/cm^2). The average fringe spacing as a function of time is plotted for the rectangular area indicated in panels (d) and (e), which represent DF and BF moiré-fringe images, respectively.[31]

The evolution of the moiré fringe spacing following pulsed excitation at a fluence of 7 mJ/cm^2 is plotted in Fig. 6.27c. An increase in fringe spacing upon excitation and strong oscillatory behavior are observed in both BF and DF images. The oscillation period is ~20 ps, matching that measured from the $(\bar{1}1\bar{1})$ diffraction peak (Fig. 6.23). Based on Eq. (6.13), the change of fringe spacing could be

interpreted as the result of either an angle change between graphene layers after a femtosecond impulsive heating, i.e., $\Delta\theta = 0.003°$ for the maximum change and $\Delta\theta = 0.001°$ for the oscillatory change in the DF image, or a change in d, or a combination of the two. As was the case for the spacings at negative time, the temporal changes observed vary from area to area across the image, so their origins must be local in nature. Given the dramatic changes that the excitation causes in the $(\bar{1}1\bar{1})$ diffraction peak, the theoretical connection between moiré fringes and diffraction makes a change in the fringes not unexpected.

In addition to moiré fringe spacing, other quantities used to characterize the DF images were selected-area image cross-correlation and integrated intensity. Both the cross-correlations and the intensities show oscillatory behavior with the same period ($\tau_p \sim 20$ ps) as appears in the moiré-fringe-spacing dynamics, with the phase of the oscillation depending on the selected area and the type of measurement considered. It is significant that the moiré fringes exhibit the same elastic-wave dynamical behavior of the material and that the strain amplitude can now be obtained for different stresses. Such studies can be extended, for example, to examine stacking faults and dislocations. In all of these studies, selected area imaging is critical for the visualization of local, and not averaged, dynamics.

6.3.4.6 *FEELS: Femtosecond EELS and chemical bonding*

Recently, femtosecond-resolved EELS (FEELS) was reported for the first time.[25] FEELS is conceptually different from time-resolved EELS (termed TREELS) as the time resolution in FEELS is not limited by detector response and sweep rate. The resolution is that of UEM, i.e., 10 orders of magnitude higher than that of TREELS. Moreover, both real-space images and energy spectra can be recorded *in situ* in UEM, and with energy filtering the temporal resolution can be made optimum, as discussed in Chaps. 5 and 8.

For graphite, we studied changes in the energy spectrum with femtosecond-delay-time steps and charted the change from zero to 50 eV in a 3D surface, time–energy–amplitude maps. A typical 3D FEELS map for the plasmon portion of the spectrum (up to 35 eV) is shown in Fig. 6.28, together with the EEL spectrum taken at negative time. The map reflects the difference for all energies and as a function of time, made by referencing to the EEL spectrum at negative time. The relatively strong enhancement of the energy loss in the low energy (electron–hole carrier) region is visible and the change is near the energy of the laser excitation. This feature represents the energy-loss enhancement due to the femtosecond laser excitation into the $\pi\pi^*$ band structure, as discussed below and in Sec. 6.6. At higher energy, the 7 eV π plasmon peak remains relatively unperturbed by the excitation. For the 27 eV $\pi + \sigma$ bulk plasmon, an increased spectral weight at positive time is visible as a peak in the 3D plot.

In order to elucidate the type of bonding involved in the temporal evolution, the spectrum was divided into three regions: the low energy region between 2 and 5 eV, the π plasmon region between 6 and 8 eV, and the $\pi + \sigma$ plasmon region between 20 and 30 eV. The integrated intensity is then plotted versus time, as in

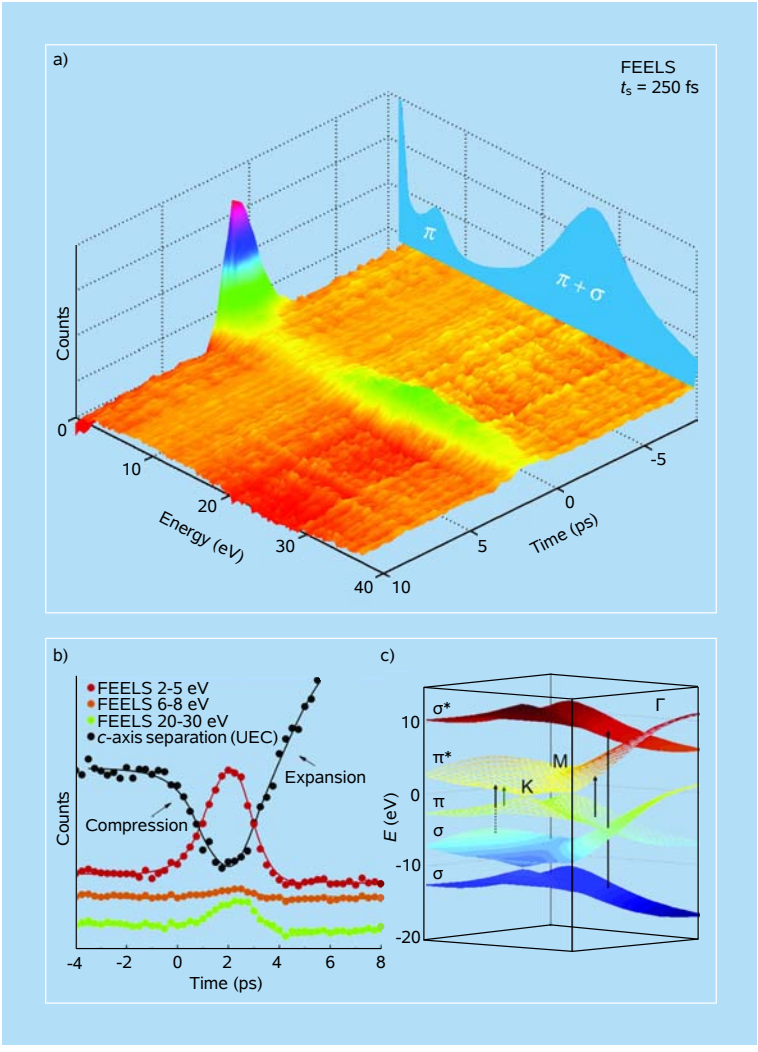


Figure 6.28 FEELS time–energy–amplitude maps of graphite. (a) The spectral changes at different time delays are displayed in a 3D surface plot. Shown also is the EEL spectrum at negative time. (b) FEELS and UEC temporal behavior. Three distinct energy ranges of the spectrum are analyzed: the excitation region (red), the π plasmon region (orange), and the $\pi + \sigma$ bulk plasmon region (green); the zero of time is relatively positioned (see Fig. 6.29).[25] The FEELS data are compared to the ultrafast response of graphite *c*-axis compression and expansion[36] (black), which was recorded at a comparable fluence. (c) Calculated bands of graphite in a portion of the Brillion zone. Shown are the excitation in the $\pi\pi^*$ band and the $\pi\pi^*/\sigma\sigma^*$ contributions to the plasmon resonance.[25] See also Sec. 6.6.

Fig. 6.28b, with the decay being on the fs time scale (see below). When, in the same figure, we display the temporal evolution of the interlayer spacing of graphite (obtained by UEC at a similar fluence, as discussed above), the time scale of the ultrafast compression/expansion tracks that of FEELS.

In time-integrated EELS, the most prominent and studied peaks are those at 7 eV and the much stronger one at 27 eV. From consideration of the in-plane and out-of-plane components of the dielectric tensor, it was established that the 7 eV band is a π plasmon in origin, resulting from interband $\pi\pi^*$ transitions in the energy range of 2–5 eV, whereas the 27 eV band is a $\pi + \sigma$ plasmon dominated by $\sigma\sigma^*$ transitions beyond 10 eV (Fig. 6.28c). We note that in this case the plasmon frequencies are not directly given by the $\pi\pi^*$ and $\sigma\sigma^*$ transition energies as they constitute tensorial quantities. For example,

$$\omega_{\pi+\sigma}^2 = \omega_{\mathrm{p}}^2 + \frac{1}{4}(\Omega_\pi^2 + 3\Omega_\sigma^2), \tag{6.14}$$

where $\omega_{\mathrm{p}} = (n\rho e^2/\varepsilon_0 m)^{1/2}$ is the free-electron-gas-plasma frequency and Ω_π and Ω_σ are the excitation energies for $\pi\pi^*$ and $\sigma\sigma^*$ transitions, respectively. For ω_{p}, the electron density is $n\rho$, where n is the number of valence electrons per atom and ρ is the density of atoms, and ε_0 is the vacuum dielectric constant. It follows that the density of occupied and empty (π, σ, π^* and σ^*) states is critical and that the π plasmon arises from the collective excitation of the π-electrons (one electron in the p-orbital, with screening corrections), whereas the $\pi + \sigma$ plasmon is the result of all four valence electrons collectively excited over the coherent length scale of bulk graphite; there are also surface plasmons but at different energies.

It has been shown both theoretically and experimentally that the π and $\pi + \sigma$ plasmons are sensitive to the interlayer separation, but whereas the former plasmon shows some shift of its peak, the latter plasmon undergoes a dramatic reduction in its intensity; when reaching the graphene limit, only a relatively small peak at ~15 eV survives.[38,39] This is particularly evident when the momentum transfer is perpendicular to the c-axis, the case at hand and for which the EEL spectrum[39] is very similar to that presented in Fig. 6.28. Thus, a connection between the selective femtosecond atomic motions and changes in the dielectric properties of plasmon resonances, the electron density, is possible.

As discussed earlier in this chapter, the initial compression and coherent oscillations suggest that the process is a cooperative motion and is governed by the out-of-equilibrium electronic structure change in the potential of excited carriers; in this case, the $\pi\pi^*$ excitation weakens the c-axis bonding. The compression, or "hard graphite" effect, is expected to cause an increase in the amplitude of the $\pi + \sigma$ plasmon peak.[40] In other words, just as the spectral weight decreases when the layers separate (to ultimately give graphene),[39] it increases when the layer separation

is reduced because of the enhanced collectiveness of all four valence electrons of carbon. The change involves shear motions and it is not surprising that the $\pi + \sigma$ peak (dominated by $\sigma\sigma^*$ excitation) is sensitive to such distortions. The π peak is less influenced presumably because only one electron is involved, but the amplitude change is relatively small. The recovery of EEL peaks in hundreds of fs is, accordingly, a consequence of the expansion which "decouples" the π and σ system.

More recently, it was possible to resolve in graphite these dynamical changes at the millielectronvolt (sub-picometer motion) scale. In this way, the influence of surface and bulk atoms motion was examined (Fig. 6.29). Among the most notable features of the dynamics, besides the correlations found in FEELS and diffraction, are the distinct behaviors of surface and bulk plasmons, both in the phase of the signal change and in the change in energy with time on the meV scale.[25a] Graphite is an ideal test case for investigating the correlation between structural and electronic dynamics. Single-layered graphene has the orbitals on carbon as sp^2-hybrids, and in graphite the π-electron is perpendicular to the molecular plane. Strongly compressed graphite transforms into diamond, whose electron density pattern is a three-dimensional network of covalent bonds with sp^3-hybrid orbitals. Thus, any structural perturbation on the ultrashort time scale will lead to changes in the chemical bonding and is therefore observable in UEM. Comparison with electron density maps, as illustrated in Fig. 6.29, can thus be made.[25a] Moreover, surface atoms have unique binding and they too are distinguishable in their influence from bulk atoms dynamics. Finally, FEELS provides a direct, *in situ* measurement of $t = 0$, when the zero-loss peak is filtered and its temporal response is examined (see Ref. 25a and Sec. 6.6).

In concluding the graphite story, our paradigm case study, we hope that the exploration made of structural dynamics with the femtosecond resolution in real space (images), Fourier space (diffraction), and energy space (spectra) now demonstrate the potential of 4D electron microscopy. This table-top UEM methodology is poised for other applications and extensions, some of which are discussed below and in Chap. 8. We also believe that FEELS should attract investigations which have traditionally been made using synchrotrons, especially in the UV and soft X-ray regions. Lastly, because of the orders-of-magnitude large cross-section of electron interactions with matter, it is possible to examine nanometer-scale molecular interfaces. Here again, graphite becomes a player, in this case as a substrate; see discussion in Sec. 6.4.3.

6.4 Selected Other Applications

In what follows we highlight other applications of 4D electron imaging with a focus on phase transitions, molecular interfaces, and nano(micro)structures during their function.

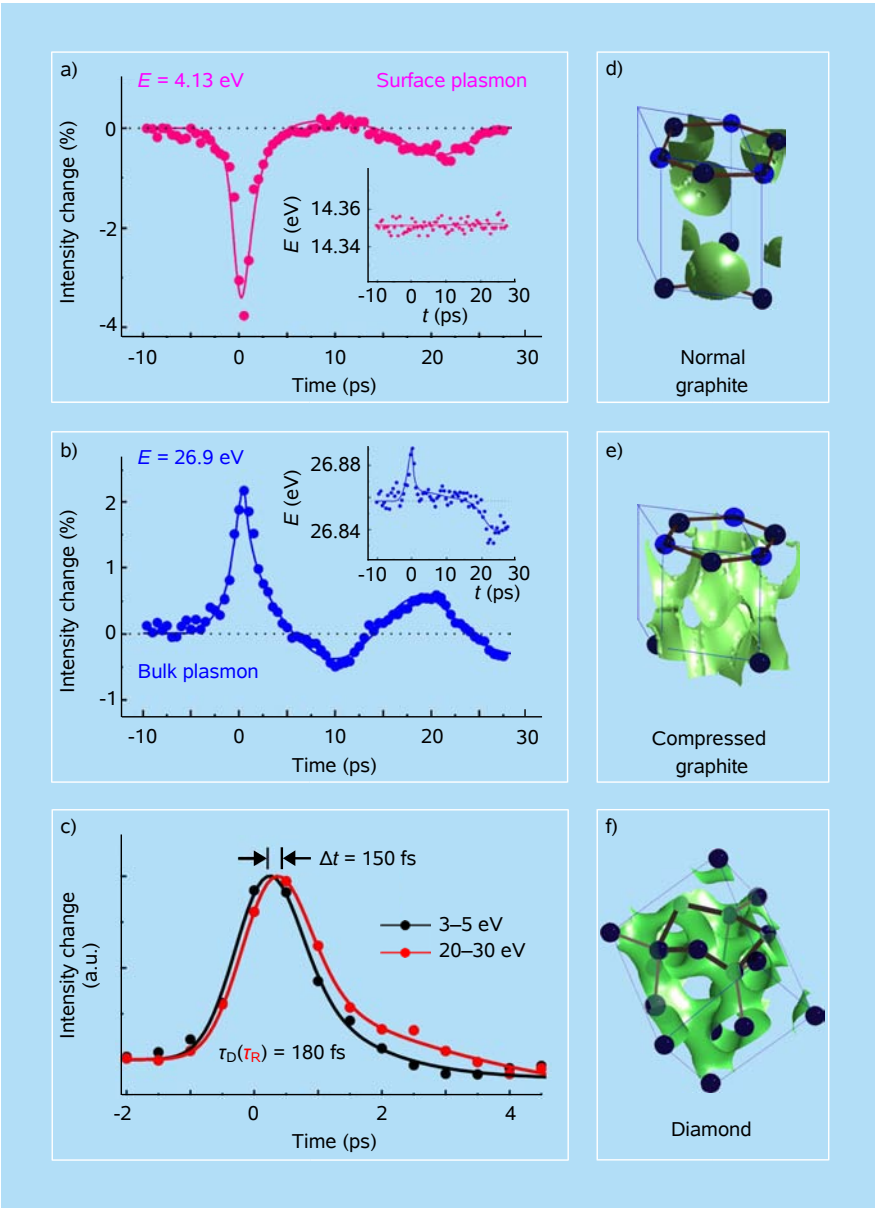


Figure 6.29 FEELS of graphite surface and bulk plasmons. Left: Peak intensity changes of (a) surface and (b) bulk plasmons as a function of time. Solid lines are guides to the eyes. Note that the increase in intensity for the bulk corresponds to a decrease in intensity for the surface plasmon (i.e., they are nearly out of phase). The time dependence shown at the bottom (c) defines the time zero (excitation region) and plasmon rise and decay. Right: Electron density distributions calculated for (d) graphite crystal with $c/a = 2.7$, (e) graphite crystal with $c/a = 1.5$, and (f) diamond. All electron densities are visualized as constant value surfaces corresponding to approximately one third of the maximum value.[25]

6.4.1 Structural phase transitions

Two examples are given here for studies of the metal–insulator phase transition in vanadium dioxide and of the nonequilibrium structural transitions in superconducting cuprates.

6.4.1.1 *Metal–insulator transformation*

Metal–insulator phase transitions are abundant in many systems. Two fundamental pertinent questions are the following: Is the transition driven by electronic redistribution or does it occur as a result of collective restructuring by nuclear motions? For either, what are the time and length scales involved? UEM imaging of films of vanadium dioxide (VO_2) reveals that such transition occurs in crystallites of the nanometer scale (Figs. 6.30 and 6.31). Both the ultrafast insulator-to-metal transition and the slower recovery were studied, examining, with theoretical models, the effects of connectivity and diffusion on the observed behavior.[19,21] Figure 6.30 depicts the images of the film and the diffraction pattern of a single crystal. The results of the hysteresis, the temporal change, and the apparatus

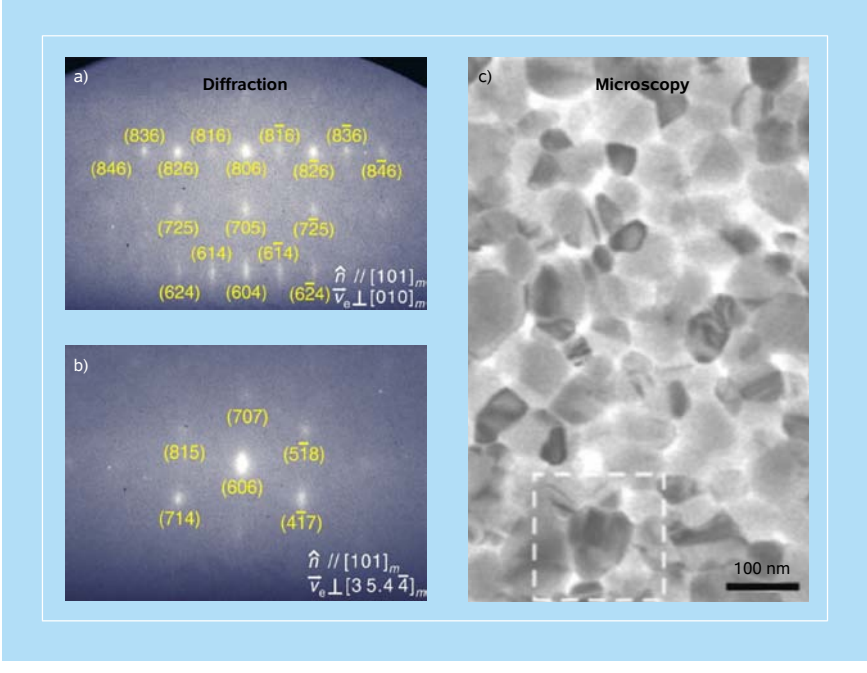


Figure 6.30 UEC and UEM of vanadium dioxide. Bragg diffractions (left) of different directions (***n***) and zone axes ($\boldsymbol{v_e}$). Real-space image as obtained using UEM-1 (right) is shown to indicate the nanometer scale of crystallites.[13a,19]

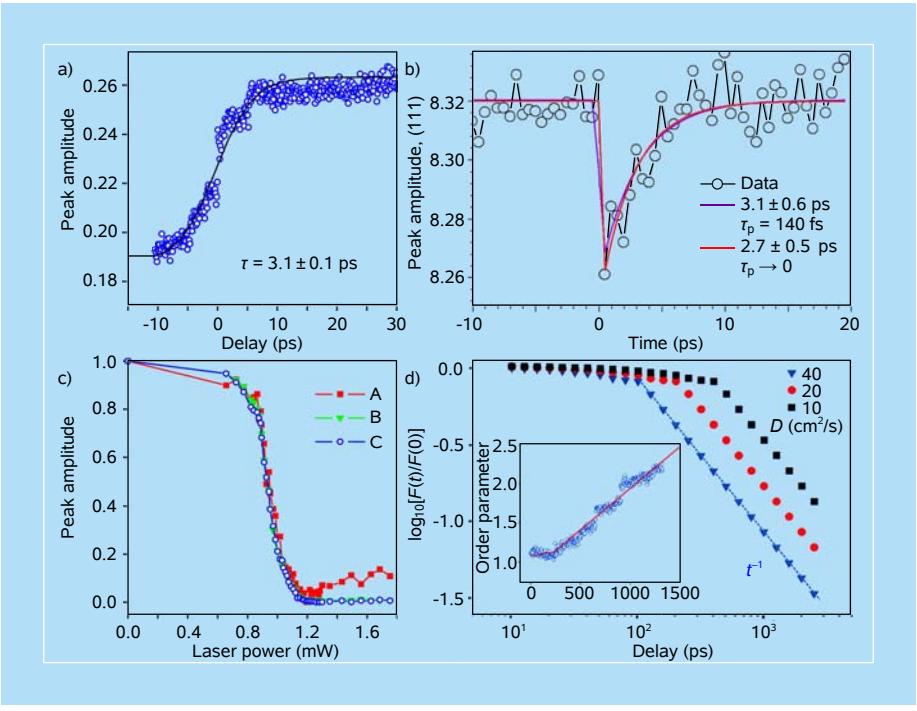


Figure 6.31 UEM of the metal–insulator phase transition in vanadium dioxide. (a) The temporal evolution, (b) the apparatus (UEM-1) response (taken by V. A. Lobastov using gold particles; τ_p = 140 fs from the fit), and (c) the change in the peak amplitude as a function of fluence. The experimental results for structural recovery are compared with the theoretical ones in panel (d). The curves (log–log plot) for the normalized survival probability $F(t)$ are shown for different diffusion constants D and display an initial plateau, reflecting connectivity among particles excited, and a long-time decay of t^{-1}, which is a characteristic of 2D diffusion. The inset gives the temporal behavior for the recovery of the order parameter with a fit for D = 20 cm^2/s. However, this value could be smaller by an order of magnitude if diffusion is within an e-probed zone.[19,21]

response (UEM-1) are given in Fig. 6.31. The transition is indeed ultrafast, but the recovery is significantly slower.

To calculate the survival probability of the metallic nanocrystallites (NCs) within the probed zone of an electron beam, $\rho(X, Y; t)$ was defined as the density distribution of metallic NCs, for any (X, Y) distances. The change with time follows the 2D diffusion equation:

$$\frac{\partial}{\partial t}\rho(X, Y; t) = D\left(\frac{\partial^2}{\partial X^2} + \frac{\partial^2}{\partial Y^2}\right)\rho(X, Y; t), \tag{6.15}$$

where D is the diffusion constant. The experimental observable, which is the survival probability, $F(t)$, of these excited metallic NCs within the electron-beam

zone can then be calculated as

$$F(t) = \int_0^R Rdr \int_0^{2\pi} d\theta \rho(X, Y; t), \tag{6.16}$$

where the integration is over the circular zone, with R being the radius of the probed region, and $X = r\cos\theta$ and $Y = r\sin\theta$ are the polar coordinates. In Fig. 6.31, the theoretical curves shown are the numerical solutions of Eq. (6.16) for different D values. The best fit gives $D = 20$ cm^2/s at the fluence used, when diffusion is to regions outside the electron zone area; if within, D becomes much less in magnitude (0.2–2 cm^2/s). The obtained values are significantly larger than that of the heat diffusion in VO_2 ($D_{heat} \approx 0.02$ cm^2/s) but close to the carrier diffusion value calculated from mobility measurements ($D \sim 0.2$ cm^2/s); for semiconductors, D reaches about 20 cm^2/s, while for metals, say, gold, $D \sim 150$ cm^2/s. (We note that the diffusion in two dimensions has the t^{-1}-dependence that is different from that of one-dimensional ($t^{-1/2}$) or three-dimensional ($t^{-3/2}$) diffusion.) At high fluences and/or for large D values, the plateau in Fig. 6.31d shrinks and the apparent recovery appears faster.

The phase transition exhibits a well-defined hysteresis (Fig. 6.31c) between the two thermodynamically-stable structures (Fig. 6.32). In order to map pathways of motion, we studied UEC of single crystals. All observed Bragg diffractions of different planes and zone axes were examined on the femtosecond to nanosecond time scale for the single crystal.[13a] As with chemical reactions, the concept of concerted (or concurrent) versus consecutive nuclear motions, which deal with reorganization of the lattice and microscopic restructuring within unit cells, becomes important to understanding the elementary steps of the mechanism and the energy landscape. Because the transformation takes place in a strongly correlated system, the dependence on excitation fluence was observed in a threshold behavior, and such dependence at short and long times was studied to elucidate the nonequilibrium transition from local atomic motions to shear at sound wave (and carrier) velocity.

The 3D sampling of diffraction and long-range order studied allow for the separation of different nuclear motions, which are mirrored in the temporal change of the structure factor for various indices (Chap. 2). For two different kinds of investigated Bragg spots (hkl), characterized by $h \neq 0$ and $h = 0$, two different types of dynamics were observed: a femtosecond process ($\tau_1 = 307$ fs) and another on the picosecond timescale ($\tau_2 = 9.2$ fs).[13a] This distinct behavior in dynamics indicates stepwise atomic motions along different directions. Because an atomic movement along a certain direction can only affect such Bragg spots that have nonzero contributions in the corresponding Miller indices, it was evident that the

initial femtosecond motion was along the *a*-axis (Fig. 6.32), which is the direction of the V–V bond in the monoclinic (insulator) structure. The full temporal behavior is illustrated in the lower panel of Fig. 6.32.

The observed structural dynamics now elucidate the mechanism and the energy landscape. From a chemical-bonding perspective, the excitation is to an antibonding state, which instantly results in a repulsive force on the atoms, and they separate along the bond direction resulting in dilation of the V–V bond on the femtosecond timescale. In sequence and on a slower time scale, the unit cell transforms toward the configuration of the rutile phase. On the nanosecond time scale, the system reaches equilibration nearly at sound wave shear motion. We have also examined the temperature dependence (of the initial state) of the dynamics and found a correspondence between thermally- and optically-driven transitions, but, of course, with different rates. These observations of stepwise atomic motions indicate that the phase transition proceeds by a nondirect pathway on the multidimensional potential energy surface and not by a direct structural conversion, thus defining a transition-state intermediate for the insulator-to-metal transition (see Fig. 6.33).

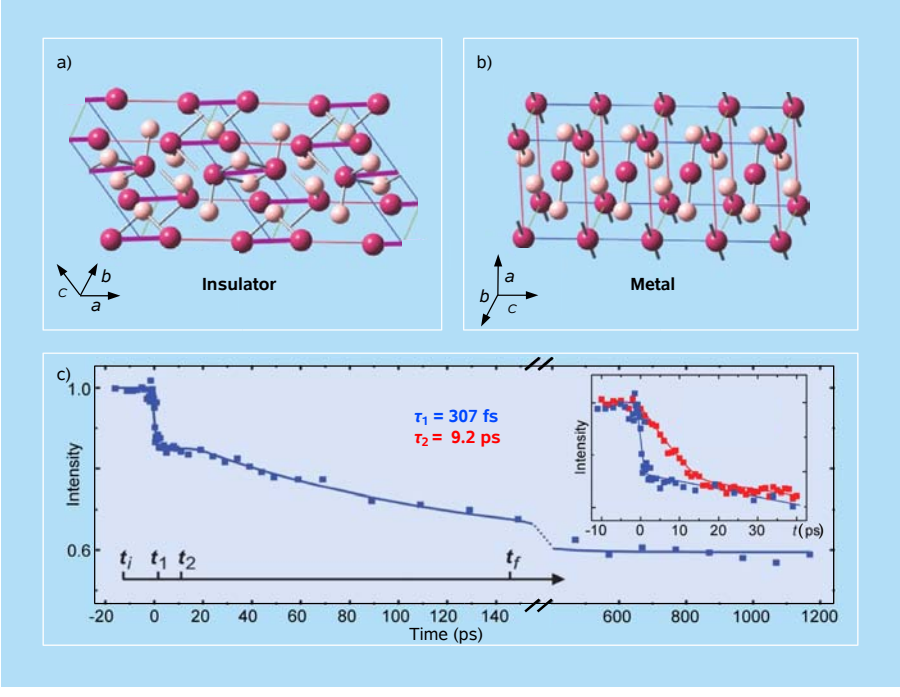


Figure 6.32 Structures and dynamics of the insulating and metallic phases of vanadium dioxide. (a, b) Structures with (insulator) and without (metal) V–V pairing. The full temporal behavior discussed in the text is given in panel (c).[13a]

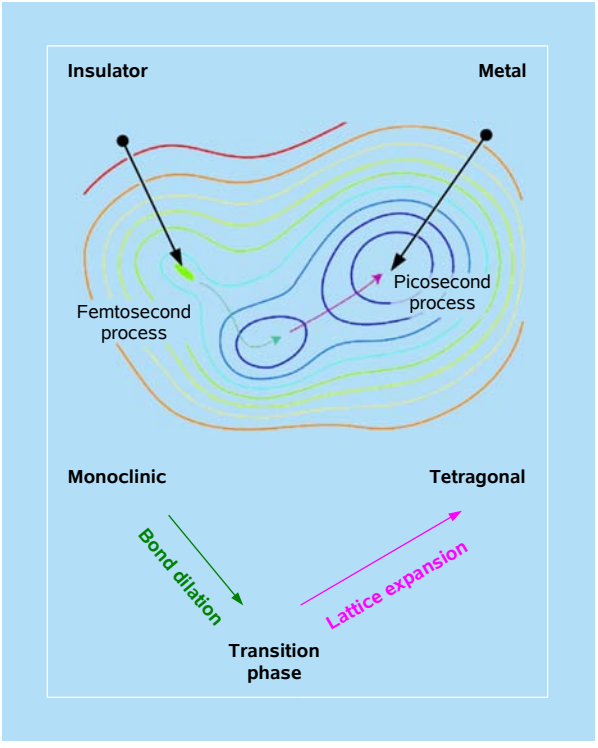


Figure 6.33 Energy landscape and coordinates pertinent to phase transition in vanadium dioxide. The phase transition between the initial (monoclinic) and the final (tetragonal) structure occurs stepwise on the femtosecond and picosecond time scales.[13]

6.4.1.2 *Transient phases of superconducting cuprates*

The second example for nonequilibrium structural phase transitions is that of superconducting cuprates.[12a] The specific material studied is oxygen-doped $La_2CuO_{4+\delta}$ (LCO); the undoped material is an antiferromagnetic Mott insulator, whereas doping confers superconductivity below 32 K and metallic properties at room temperature. From the observed Bragg spots, the unit-cell parameters were obtained using the patterns for different zone axes: $a = b = 3.8(1)$, $c = 13.1(1)$ Å. Structural dynamics were then obtained by recording the diffraction frames at different times, before and after the arrival of the optical excitation pulse. What was expected, as observed in previous UEC studies in this laboratory, is that the peak shifts continuously and the intensity decreases with time. Instead, all curves obtained at different times cross at a single value of s-coordinate which parametrizes the momentum-transfer space. This intensity sharing with a common crossing point, a structural *isosbestic point*, indicates a transition from the initial phase to a new (transient) phase.

The structural interconversion displaying the isosbestic point is illustrated in Fig. 6.34. The frame-referenced diffraction profile, as a function of time, reveals the depletion of the initial structure and the accumulation of the transient-phase structure. The population of the initial (transient) phase decays (builds up) with a time constant of 27 ps, but the formed phase restructures on a much longer time scale (307 ps). Because the linear expansion coefficient is $\alpha_l \leq 1.0\times10^{-5}$ K^{-1}, the observed 2.5% increase in the lattice constant would correspond to an unphysical 2500 K rise in the lattice temperature at equilibrium. Another striking feature of this structural phase transition is its dependence on the fluence of the initiating pulse. A threshold was observed, above which the lattice constant of the transient-phase structure changes linearly with the fluence (Fig. 6.34).

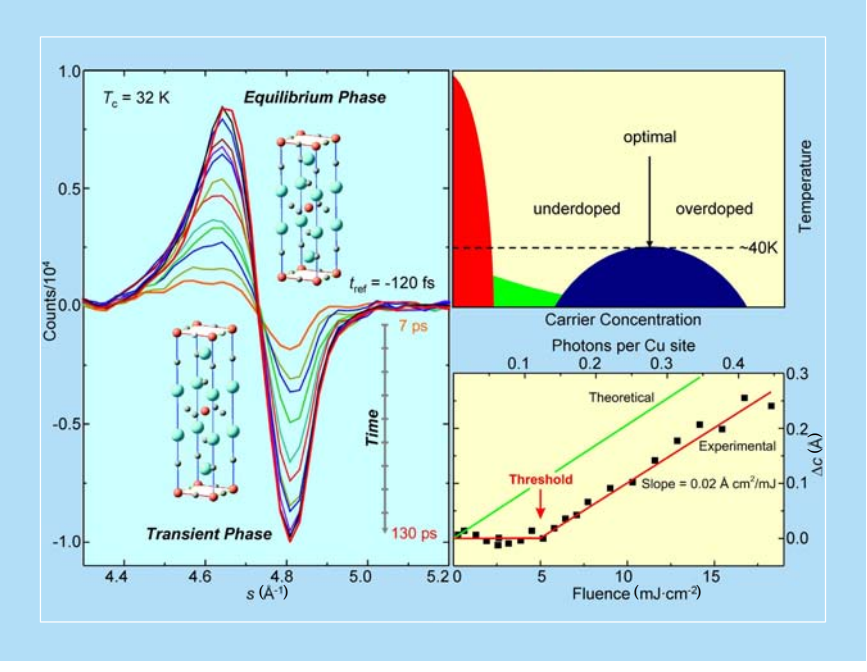


Figure 6.34 Structural phase transition in oxygen-doped cuprate $La_2CuO_{4+\delta}$. Schematic phase diagram (upper right) and the fluence dependence (lower right) are also shown. The temporal evolution of the referenced diffraction profile indicates the depletion of the initial structure and the build up of the transient-phase structure (left). Lattice constant change as a function of fluence of the initiating pulse is shown at lower right. Note the presence of the isosbestic point at ~4.76 $Å^{-1}$ and the threshold at ~5 mJ/cm^2, which corresponds to a critical value for the number of photons (0.12) per copper site.[12a]

The transformation at 1.55 eV is believed to be the result of a charge transfer from oxygen (O^{2-}) to copper (Cu^{2+}) in the *a*–*b* copper–oxygen planes, as reported in the literature. With the lattice relaxation being involved, the excitation is shared

microscopically (exciton type), and finally a transition to a transient phase is made (macroscopic domain). The net charge distribution in the transient phase results in the weakening of interplanar Coulomb attractions, leading to expansion along the *c*-axis. The behavior is nonlinear in that when the number of transformed sites is below a critical value, the macroscopic transition is not sustainable. The crystal domain is greater than $20\,nm^2$ and symmetry breaking is not evident because charge transfer occurs in a plane perpendicular to the *c*-axis expansion. By consideration of Madelung energy and charge distributions, the linear dependence on fluence and the large values of expansion were accounted for.[12a,12d]

The transient-phase structures, which are inaccessible by means of equilibrium methods, are undetectable by optical probes with wavelengths longer than lattice spacings. Moreover, the time scales of optical response and structural changes are very different. For the cuprate studied, the observed phase transition is the result of electronic charge redistribution and lattice collective interactions to form domains. One final note regarding the observed threshold effect can now be made. The similarity of the apparent threshold for "photon doping" at ~0.12 photons per copper site and the "chemical doping" at fractional charge of 0.16 per copper site, required for superconductivity, may have its origin in the nature of the photoinduced inverse Mott transition. The implications are clear, both from the perspective of condensed-matter physics of correlated systems and for possible use in changing the effective temperature.

In a more recent study,[12b] the mechanism of electron pairing (spin-exchange versus electron–phonon interaction) in high-temperature superconductors, which is still the subject of intense debate, was examined in a series of investigations involving different types of cuprates. The transient structures were determined using time-resolved electron diffraction, following carrier excitation, but, in this study, with polarized femtosecond heating pulses. The effect of different dopings and temperatures was investigated. The results provide evidence of the role of structural dynamics, with selective atomic motions (buckling of copper–oxygen planes), in the anisotropic electron–lattice coupling. The deformation amplitude reaches 0.5% of the *c*-axis value of 30 Å when the light polarization is in the direction of the copper–oxygen bond, and the decay of diffraction intensity slows down at 45°. These findings suggest a selective dynamical lattice involvement with the anisotropic electron–phonon coupling being on a time scale of 1–3.5 ps, depending on direction. It is of the same order of magnitude as that of the spin exchange of electron pairing in the superconducting phase. The results, we hope, should stimulate the development of theoretical models that explicitly incorporate the role of structural dynamics in the mechanism of high-temperature superconductivity.

6.4.2 Nucleation and crystallization phenomena

The transformation of amorphous structures, such as liquids or random-coiled proteins, into ordered structures involves complex dynamical processes that ultimately lead to the final native state. The mechanism is determined by the scales of time, length, and energy as they define the nature of the elementary steps involved. Using the single-pulse mode of UEM, we reported[22] the *in situ* 4D visualization of the irreversible process involved in the transformation of amorphous silicon, from the liquid phase to its crystalline phase, a complex transition involving nucleation and growth on the nanosecond-to-microsecond time scale. Earlier, surface melting was observed on the picosecond time scale in one of the first UEC studies from this laboratory (see below).

To initiate crystallization from the amorphous (metallic) phase, a single optical pulse at a fixed energy was used, while a single pulse of electrons imaged selectively in space the transformation as the structure was continuously changing with time. With selected-area diffraction, the amorphous and crystalline phases were identified, and the temporal profile of (structural) crystallization in the embryonic stages was obtained for the entire image. Amorphous nanoscale surface melting is ultrafast within 10 ps,[8] but crystallization takes place on a significantly longer time scale(s). On the nanosecond time scale, there is no significant heat diffusion within the probed volume, and two main processes were identified, that of a *nondiffusive* nanoscale layer propagation at ~15 m/s (explosive growth), and another that occurs with much slower velocity. Both the nanosecond and femtosecond pulses of UEM were used in this study, and below we discuss the significance in the description of the dynamics observed *in situ* as the transformation unfolds.

The central panel of Fig. 6.35 depicts a low-resolution image obtained after a single laser pulse heated the material and induced regions of change (*a*, *c*, *C*, and *i* denote amorphous, small-grain crystalline, large-grain crystalline, and interfacial phases, respectively). The structures in those regions were studied more in detail through real-space images and Fourier-space diffraction patterns. A broad halo ring characteristic of the amorphous phase is evident in *a* and, after laser pulse heating, the diffuse amorphous patterns transform to sharp Debye–Scherrer rings or Bragg spots. The rings after crystallization are the averaged diffraction spots from a large number of diamond-type structures (nanocrystals) with random orientations. Nucleated nanocrystallites with sizes of 10 nm are observed in phase *c*, whereas larger grains with sizes of several hundred nanometers are observed in phase *C*, which is made in the area with the highest laser fluence (as noted in the diffraction frames, Bragg spots become most pronounced in phase *C*). At the interface (*i*), the amorphous phase and the crystalline phase coexist. Frames 1–5 taken with femtosecond irradiation show the onset of nucleation and grain growth,

as evidenced in the increased grain size noted by dotted circles and by the change in contrast.

To quantify the spatiotemporal behavior, diffraction profiles were constructed from the radial average of respective diffraction images, all normalized to total electron count (Fig. 6.36). The images reflect the growth of crystallinity, i.e., the enhancement of diffraction peaks over the diffuse background from $t < 0$ to $t \gg 0$ (asymptote). Notably, a transient diffraction profile captured at, for example, 600 ns after the arrival of the heating pulse is fit to a linear combination of the two images at negative time (58%) and at the asymptotic (42%) time point (reminiscent of Vegard's law and indicative of the interconversion between the two phases). Accordingly, all obtained frames of transient change were handled similarly to provide the actual temporal change of the structure with time (the *kinetic profile*). As seen in Fig. 6.37, the frame-averaged fraction of structural crystallinity increases with time in a nonuniform manner. If totally averaged, in a simple exponential rise fit, a characteristic time of 880 ± 140 ns for the final transformation

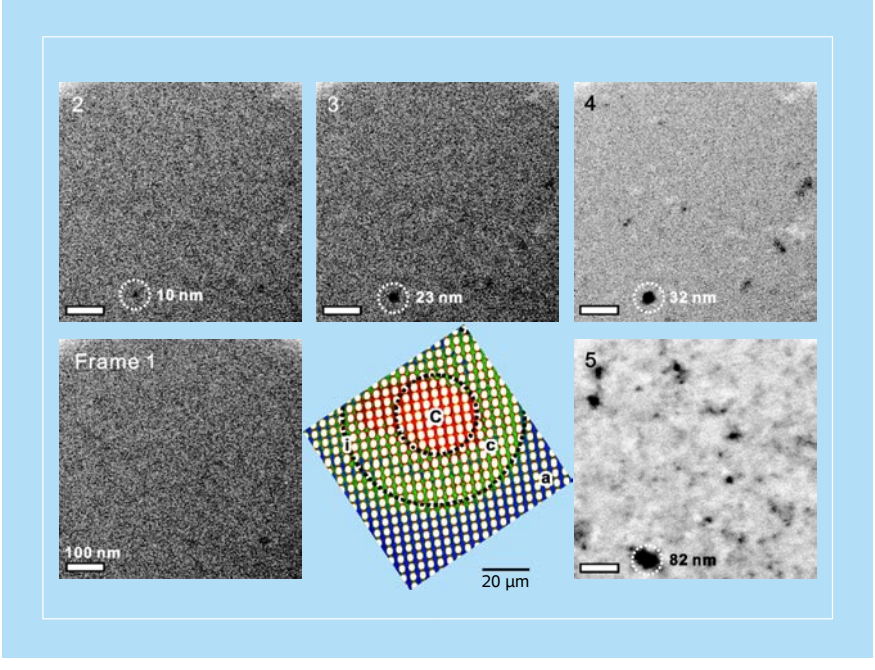


Figure 6.35 Imaging of crystallization. Image frames obtained by using femtosecond and nanosecond pulses. Frames 1–5 show the growth of crystallite size as the number of femtosecond pulses increases, with the transition from (1) amorphous to (5) polycrystalline being evident in the images. The central colored image shows the response after a single laser-pulse heating. The zones are identified by circles to indicate the amorphous (*a*), the interfacial (*i*), and the two crystalline (*c* and *C*) phases. The *C*-zone diameter is 35 μm and the *c*-zone diameter is 77 μm. The specimen in the microscope was thin films of amorphous silicon deposited on a grid by e-beam evaporation of a silicon wafer.[22]

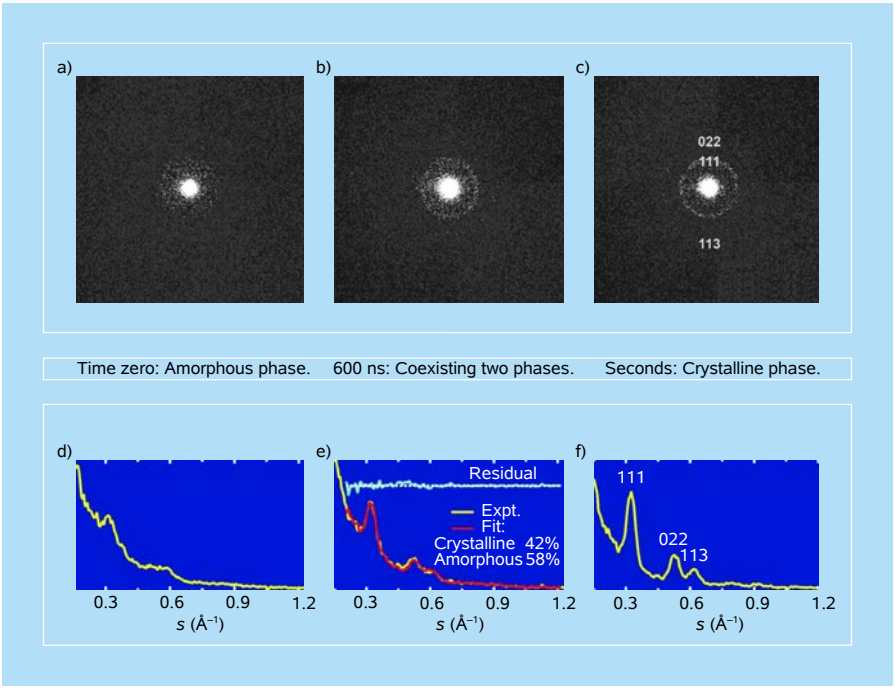


Figure 6.36 Diffraction representing three phases in crystallization. Top: Diffraction (a) before, (b) during, and (c) after the transformation, taken from the same sample area of irradiation. Three prominent diffraction rings are indexed in the image. Bottom: Radially averaged diffraction profiles of respective images shown above. To fit a transient-frame diffraction profile (e), the negative-time (d) and asymptotic profile (f) are combined in proportion to their contributions. For the quality of a fit, the difference between the transient-frame profile and the fitted one is also depicted. The high signal-to-noise ratio proved essential for the accuracy displayed.[22]

is obtained. Finally, the single-pulse diffraction of a single crystal is shown in Fig. 6.37 on the right. Also included in the same figure are two frames for surface melting taken at ±10 ps.

The maximum extent of zone *C* melting must be reached by the end of the heating pulse. At the same time, an outwardly propagating wave of explosive crystallization is launched (see below). The recorded structural crystallization transient of Fig. 6.37 reflects the fractional crystalline content of the entire area probed by electron pulses. Increase in the signal comes both from the super-lateral growth from liquid pockets (zone *C*) and explosive crystallization (zone *c*), with each having a distinct time evolution. At the interface, nucleated crystallites form the explosive front because of balance between latent heat of their crystallization and melting in neighboring amorphous structure. In zone *C*, the crystallization will occur at later times, but the released latent heat most probably will be nonexplosive given the time scale involved in cooling rates.

A simple model of explosive crystallization starting at the edge of zone C at $t = 0$ and traveling at 15 m/s across zone c yields a duration of 1.2 μs for contributions to the total signal from zone c alone, which is the time scale recorded in Fig. 6.37. Taking a propagation time of ~0.4 μs gives a distance in zone c of ~6 μm and a 2D area accounting for ~20% of the final signal, which is consistent with the explosive wave being effective to nearly half of the measured value at that time. The step observed in the kinetic profile appears to be from a delayed process, and we attribute it to the super-lateral growth that is of slower nature. Thus, from the evolution of crystallinity in real time and the changes in morphology, for nanosecond and femtosecond pulse heating, two types of processes are identified, one that occurs at early time and involves a nondiffusive motion and another that takes place on a longer time scale. Similar mechanisms of two distinct time scales may perhaps be important in biomolecular folding, as suggested in Ref. 22.

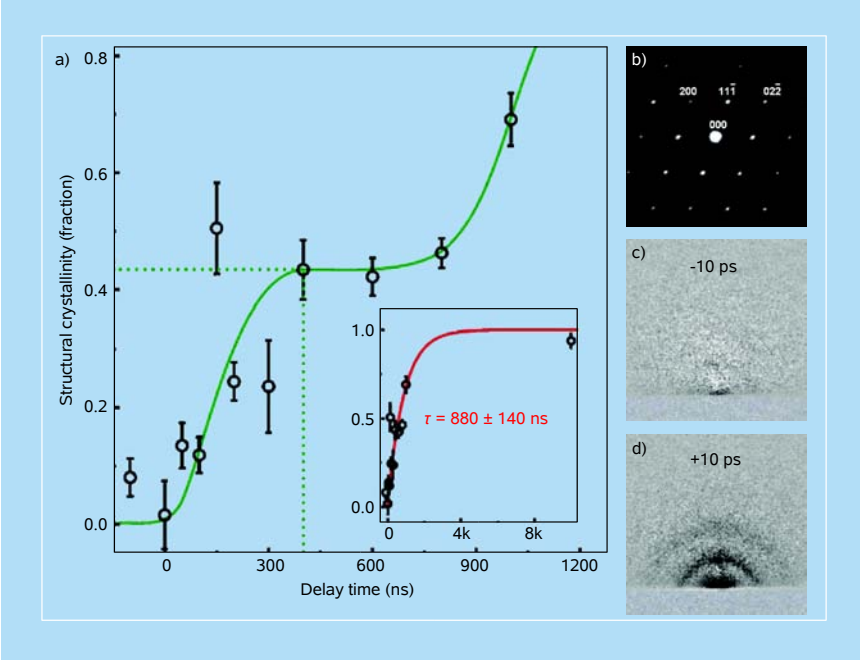


Figure 6.37 Time-dependent growth of structural crystallinity. The fraction (circles) versus time gives the kinetic profile of the transformation (a). Zoomed-out kinetics over a longer time scale: the "averaged" exponential rise fit of τ = 880 ns is depicted by a solid line (inset). Single-pulse diffraction pattern of single-crystal silicon was taken with the incident electron beam parallel to the [011] zone axis (b). Two diffraction difference frames as obtained at the indicated delay times are shown (c,d). Referencing to the pattern at negative time, the frame at $t = -10$ ps displays no pattern (c), whereas the one at $t = +10$ ps shows the surface melting because of the ultrafast phase transition of the amorphous solid to the liquid phase (d).[8] In the kinetic profile of crystallization, we note both the (averaged) exponential rise and the structured one.[22]

6.4.3 Interfaces and biological assemblies

6.4.3.1 *Water on hydrophobic and hydrophilic substrates*

Nanometer-scale assemblies at interfaces are challenging structures to determine, yet their structures and dynamics are of the utmost importance. Perhaps water and ice are the supreme examples. With UEC, the problem of interfacial water was addressed by determining both the structure and dynamics using hydrophobic or hydrophilic surface substrates.[11] The interfacial and ordered (crystalline) structure was evident from the Bragg diffraction, and the layered and disordered (polycrystalline) structure was identified from the Debye–Scherrer rings (Fig. 6.38). The temporal evolution of interfacial water and layered ice after the temperature jump to the substrate was studied with monolayer sensitivity.

On the hydrophilic surface substrate (chlorine-terminated silicon), the structure was found to be cubic (I_c), not hexagonal (I_h); on the hydrophobic surface (hydrogen-terminated silicon), the structure was still cubic, but very different in the degree of order. As detailed elsewhere,[11] structural dynamics is distinctively different for the two phases. The interface is dominated by polycrystalline I_c, but coexisting in this phase are crystallite structures not adjacent to the surface of the substrate. The change in the structure of I_c has different temporal behaviors reflecting differences in the transfer of energy to polycrystalline and crystallite I_c. The issues of interest are the coexistence of these structures, their different dynamics, and the time scales for energy transfer and disruption of the hydrogen bond network.

At the microscopic level, several conclusions were drawn. First, the reaction coordinate for breaking hydrogen bonds involves a significant contribution from the O···O distances. This is evidenced in the depletion with time of the corresponding peak in the radial distribution function. Second, the time scale of energy dissipation in the layered structure must be faster than that of desorption, as no loss of water molecules was observed. Third, the time scale of the dynamics at the interface is similar to that of water at protein surfaces. Finally, the order of water molecules at the interface is of high degree. Using molecular dynamics (MD) simulations, the nature of forces (the atomic-scale description) and the degree of ordering were examined in collaboration with Parrinello's group, elucidating features of crystallization, amorphization, and the time scales involved.[41]

For water on graphite (Fig. 6.39),[11b] one of the unexpected observations made was a structural isosbestic point, similar to the one found in the study of cuprates. In the case of the ice–graphite composite, however, the surface structure plays a crucial role. As discussed in Sec. 6.3.4, the lattice of graphite has been shown to undergo a contraction followed by a large expansion along the *c*-axis with a time constant of ~7 ps.[36] For water on graphite, the observed "old-structure"

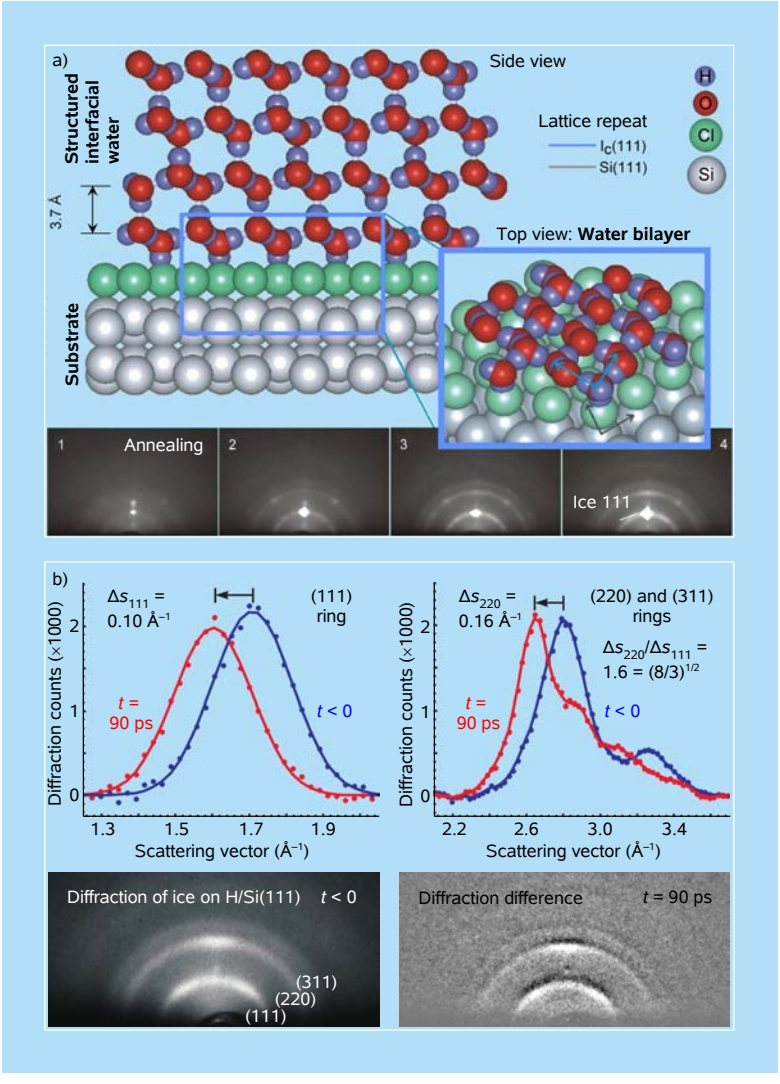


Figure 6.38 Structural dynamics of interfacial water at hydrophilic (hydrophobic) interfaces. The chlorine termination on a Si(111) substrate (a) forms a hydrophilic layer that orients the water bilayer. The ordered stacking persists for three to four bilayers (~1 nm) before disorientation takes place and results in crystallite islands, forming the layered structure. Note the coexistence of two phases as evidenced by Bragg spots and Debye–Scherrer rings.[11a] For the hydrophobic (hydrogen-terminated) substrate (b), two frames are shown together with the change in Bragg diffraction position. Note the extent of the shift Δs and its scaling with the order of diffraction (Miller indices); see also Sec. 6.4.4.4.

disappearance time (~10 ps) nearly matches the 7 ps time constant, indicating that excitation of the original assembly is due to the vibrational coupling of ice with underlying graphite. In other words, the ice film receives the vibrational energy from graphite and undergoes ultrafast softening ("melting"), which causes the

diffraction intensity to decrease by ~20%. To accommodate such large-amplitude structural perturbation and the undulating substrate, the stack of (111) water bilayers expands in a phase transformation, with a rise time of ~20 ps. Later on, when the vibrational motion of graphite comes to an end, the interfacial assembly

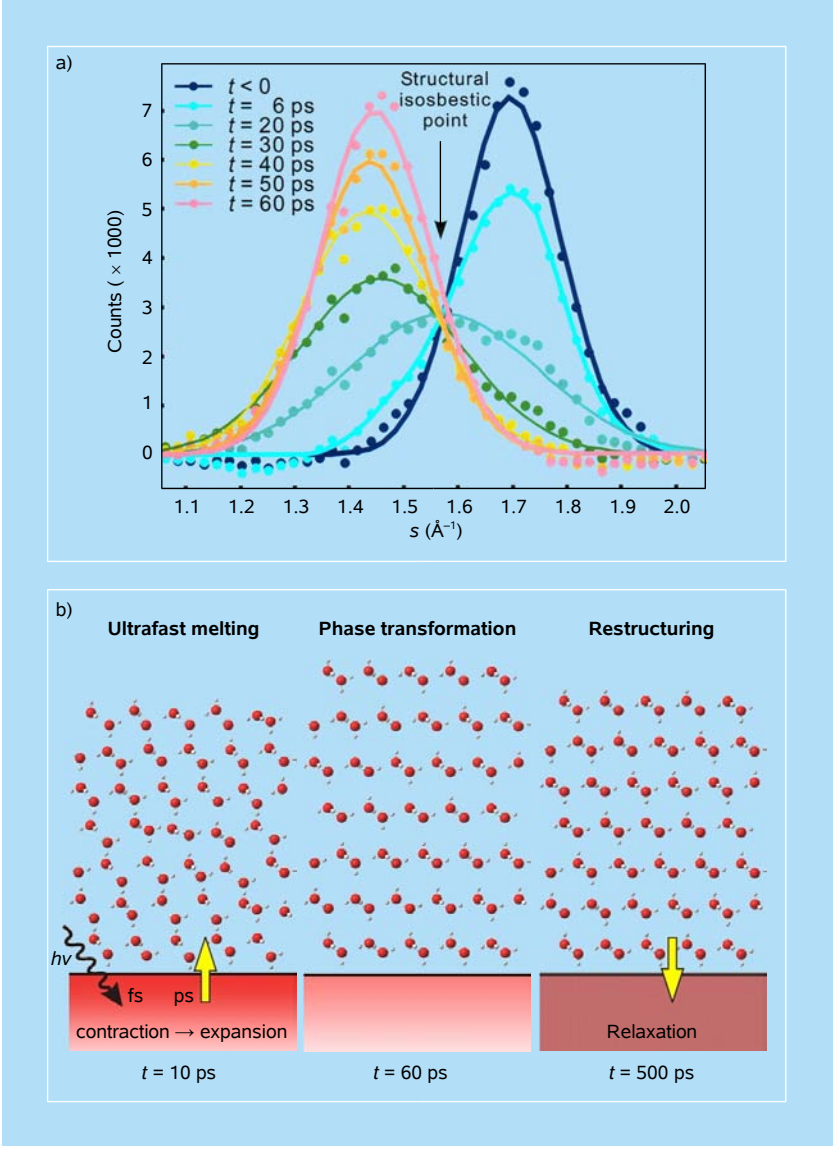


Figure 6.39 Structural isosbestic point as obtained for (ordered) water on graphite. (a) The shift of the (111) diffraction peak of the water layers as a function of time and (b) representation of the dynamics for the ice layers on graphite. Yellow arrows denote the direction of energy transfer. Note the formation of the expanded ice phase involving the structural isosbestic point at a well-defined value of *s* in panel (a).[11b]

is able to form an even better conformation (structural annealing) that enhances the diffraction intensity. When this stage is reached, the transformed ice structure behaves collectively and begins its recovery as a single unit, which occurs with two time constants of 75(5) and 390(35) ps. The overall restructuring can be reproduced with heat diffusion to the substrate. Accordingly, there is no bottleneck in the energy exchange between water and graphite, neither in cooling nor in the initial expansion of the water assembly. The transformations span the picosecond-to-nanosecond time scales.

In comparing the behavior on graphite with that on another hydrophobic surface, the hydrogen-terminated Si (111), it is concluded that the potential energy interactions (simple bonding energetics) of water–water and water–substrate are insufficient to describe the relatively high degree of water ordering observed on graphite. Substrate morphology at the interface could be as important in confining the structure on the nanometer scale. For ice on Pt (111), both the surface steps and screw dislocations are critical in the formation of metastable cubic structure (and not the hexagonal one) and in the growth. Spectroscopic investigation of interfacial (D_2O) ice, using sum-frequency generation with longer time resolution, suggested the presence of melted regions, but neither the structure nor the order was possible to observe with atomic-scale resolutions. Unlike in conventional heating, the lack of sublimation of water on the time scale of observed dynamics is due to ultrafast melting and energy exchange in the water assembly; see Fig. 6.39b.[11b] Questions remain that are pertinent to details of the topology of the energy landscape, and work is continuing on this problem, which is relevant to many fields.

6.4.3.2 *Bilayers, phospholipids, and cells*

Perhaps the simplest of membrane-type structures is a bilayer of fatty acids. These long hydrocarbon chains self-assemble on surfaces (substrates) and can also be made as "2D crystals." To achieve crystallinity, the methodology of Langmuir–Blodgett (LB) films is invoked, providing control over pH, thickness, and pressure. Chain structures formed by self-assembly were also examined. In these studies,[42] the *T*-jump of the substrate, or the expansion force of surface atoms, is exploited to heat up the adsorbed layers deposited on either a hydrophobic or hydrophilic substrate. The femtosecond infrared pulse has no resonance for absorption to the adsorbate. The studies made for monolayers, bilayers, and multilayers of fatty acids and phospholipids provide an opportunity to determine structural dynamics and to examine nanoscale changes due to the transition from 2D to 3D dimensionality.

Four types of measurements were made: the change of position of Bragg spots with time, the temporal evolution of the diffraction intensity, the increase/decrease in diffraction spot width, and the change of diffraction with the angle of incidence (rocking curve) and azimuthal angle for the position of the electron pulse relative

to the zone axis of the substrate. The static, time-averaged crystal structures of the adsorbates were established by determining the orientation of the chains relative to the surface plane, and the $-CH_2-CH_2-CH_2-$ chain distances, which define the subunit-cell dimensions a_0, b_0 and c_0 (Fig. 6.40). For arachidic fatty acid, depending on pH and deposition conditions, a_0 ranges from 4.7–4.9 Å, b_0 ranges from 8.0–8.9 Å, and c_0 ranges from 2.54–2.59 Å. For dimyristoyl phosphatidic acid (DMPA), $c_0 = 2.54$ Å.

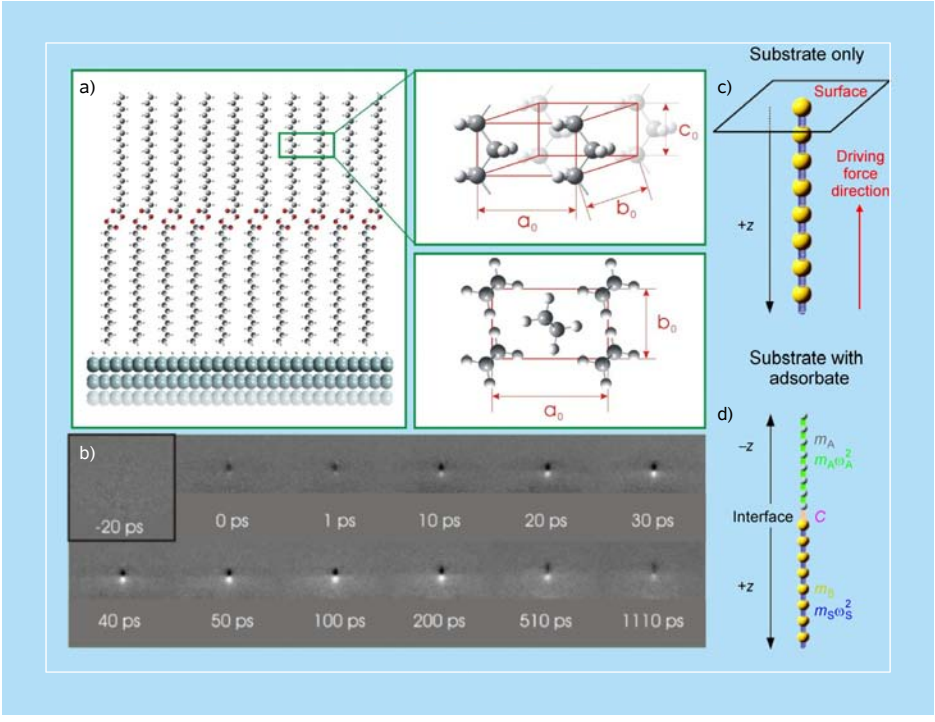


Figure 6.40 UEC of 2D fatty-acid bilayers. (a) The subunit cell structure determined and (b) the dynamics obtained by frame referencing, following the heat pulse transfer to the substrate. The frames at 0 and 1 ps show the onset of structural changes, and the frames at longer times display the restructuring toward the equilibrium state. Also shown are the 1D atomic chains which represent (c) substrate-only and (d) substrate–adsorbate lattices used in the theoretical model. Phospholipids were similarly studied.[33a,42]

The transient anisotropic change in c_0 of fatty-acid and phospholipid layers is vastly different from that observed in steady state. At equilibrium, the observed changes are in a_0 and b_0 (not in c_0), and the diffraction intensity monotonically decreases, reflecting the thermal, incoherent motions (Debye–Waller effect) and phase transitions; it is known that, for phospholipids, different phases ("gel" and

"liquid") exist. On the ultrashort time scale, the expansion is along c_0, unlike in the thermal case, and the amplitude of change is much larger than that predicted by incoherent thermal expansion. The expansion amplitude depends on layer thickness and the nature of bonding to the substrate (hydrophilic versus hydrophobic). The changes in Bragg-spot intensity and width are very different from those observed by equilibrium heating.

Following the ultrafast *T*-jump, the structure first expands (because of atomic displacements) along the *c*-direction (Fig. 6.40). As evidenced in the increased diffraction intensity and narrowing beyond the initial values of the width of the diffraction feature, the motions with the acquired energy in the layers result in transient structural ordering through "annealing" and/or chain-decreased friction. On the nanosecond and longer time scale, the structure reaches the quasi-equilibrium or equilibrium state (incoherent movement of atoms), and the original configuration is recovered by heat diffusion on the millisecond time scale between pulses. This behavior is in contrast with that observed at steady state, as mentioned above.

The net change in displacement is determined by the impulsive force of the substrate (including coupling to adsorbate), and the maximum value of the extension depends on elasticity and heat capacity. If heating occurs in an equilibrated system, the change in the value of c_0 with temperature, Δc_0, should be independent of the number of $-CH_2-CH_2-CH_2-$ subunits in the chain; because of anharmonicity, $\Delta c_0/c_0$ is determined simply by α, the thermal expansion coefficient, which is typically very small, $\Delta c_0/c_0 \sim 10^{-5}$ K^{-1}. For a 10-degree rise, this expansion would be on the order of 10^{-4} Å, while the observed transient change is as large as 0.01 Å.

In the nonequilibrium regime, the large amplitude of expansion is understood, even for harmonic chains. The impulsive force at short times induces a large change in the value of Δc_0 as the disturbance (wave-type) accumulates to give the net effect that is dependent on the number of C atoms. In other words, as the disturbance passes through the bonds, the diffraction amplitude builds up and exhibits a delay, ultimately giving rise to a large total amplitude for the change. This picture also explains the dependence of expansion on the total length of the chains, the increase in the initial maximum amplitude as the temperature of the substrate increases, and the effect of strong (hydrophilic) versus weak (hydrophobic) binding to the substrate. Quantification of the total change must take into account the form of the substrate force and variation in the density of the LB films upon going from single to multiple layers, as well as the (zone-axis) morphology discussed earlier in this chapter.

The fact that the initial change in intensity (and elongation) occurs on the 10 ps time scale and that the distance traveled is approximately 20 Å (for a

monolayer), the speed of propagation should be sub-kilometers per second, which is reaching that of the propagation of sound waves. Because the substrate is heated through optical and acoustic phonons, the rise is convoluted with the process of phonon generation, which is on the time scale of 10 ps. Accordingly, the speed could be of a higher value, reflecting the actual speed of sound in the layers. Future experiments will further investigate this region in order to elucidate the maximum extension possible and the expected features of coherent motion. The model of coherent coupling among bonds in the underdamped regime of harmonic motions yields results vastly different from those of the diffusive behavior in the overdamped regime, but it has some common characteristics to the well-known Fermi–Pasta–Ulam model of anharmonically coupled chain dynamics,[33] and to recent optical study of thermal conductance of molecular chains.[43]

Besides the above analytical theoretical work, MD simulations of a prototype system were performed in collaboration with T. Shoji and colleagues in Japan. The model used is that of a silicon substrate with the adsorbates made of $C_{20}H_{42}$ chains, covering a total combined length of 95 Å. The potentials for the chains, substrate, and the interaction at the interface were obtained from *ab initio* calculations. The time step was 0.5 fs and the total number of steps was 200,000. The heat pulse was modeled based on the kinetic energy of the substrate atoms. The radial distribution function and the actual vibration motions of the atoms were then obtained at different times. These calculations correctly predict the structural cell dimensions observed experimentally and elucidate the coherent motion in the chain bonds and their time scales. Preliminary results show the increase in $-CH_2-CH_2-CH_2-$ distance near the silicon surface by 0.08 Å in about 5 ps. Using the same approach, studies of the self-assembly were made to elucidate the formation of interchain stacking with void channels in between at zero pressure and in a confined "box" of chains.

In biological imaging, as discussed in Chap. 8, the pulsed characteristic may lead to a better resolution of single-particle imaging and, just as importantly, will provide the time scale of structural change. The first effort in this regard was directed toward the imaging of a stained cell with UEM-1.[16] For such a proof of principle of real-space imaging, the issues involved are the extent of lateral coherence and sensitivity to contrast imaging. UEM images for cells derived from the small intestines of a four-day-old rat were recorded. The specimen was prepared using standard thin-section methods. The cells were positively stained with uranyl acetate, causing them to appear dark on a bright background.

Figure 6.41 shows the UEM images of the cells at two different magnifications. The images were obtained using ultrashort pulses of UEM with exposure times of a few seconds; such exposure times compare well with those of standard EM imaging. In the figure, both the microvilli as well as the subcellular vesicles of the epithelial cells can be visualized. More recently, in UEM-2, it was possible to

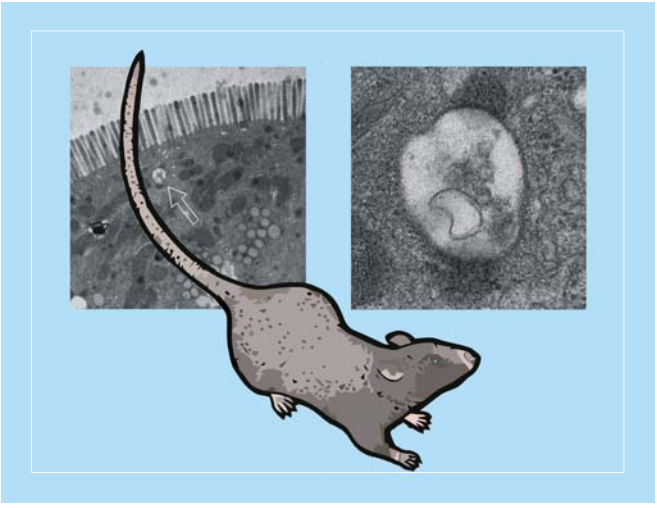


Figure 6.41 UEM images of positively stained rat intestinal cells. The images show the microvilli in the intestinal epithelium of the neonatal rat along with numerous small vesicles throughout the cytoplasm. The arrow on the left indicates the vesicle magnified on the right.[16,29]

obtain *phase-contrast* images of bacterial cells in vitreous water (cryo-microscopy) showing the membrane structure and flagells of the cell, as displayed in Fig. 6.42. In the same figure, we also show the results of phase-contrast imaging of a crystal of catalase protein, which displays the resolution of the crystal-plane separation (9.3 nm). We will address the opportunities and challenges of bio-UEM imaging in Chap. 8.

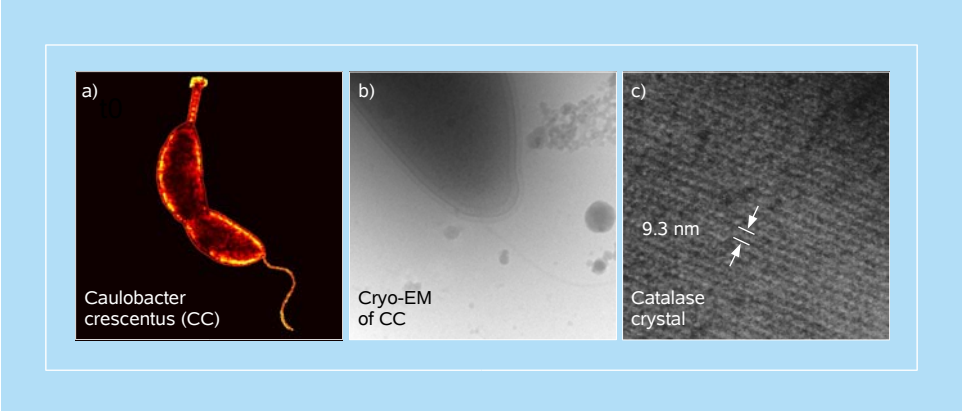


Figure 6.42 Cryo-UEM of proteins and cells. (a) Bacterial cell (*Caulobacter crescentus*, CC, adapted from MicrobeWiki) and a high-resolution image of CC as obtained using cryo-UEM in vitreous ice (b), showing membrane structure and flagellum. (c) High-resolution image of a catalase protein crystal showing the planes with the interplane separation of 9.3 nm.

6.4.4 Nanomechanical and optoelectronic systems

In the lexicon of nanoscience and nanotechnology, the interest has primarily focused on the development of nanoscale "devices" with functional purposes, such as the potential for nanoelectronics and possibly the making of mimics of biomachines, which operate with impressive molecular scale precision. Uncovering the mechanisms of the function of nanoscale phenomena requires direct imaging at sufficiently high spatial and temporal resolutions. Equally importantly, such visualization provides understanding of the fundamental nature of the physical forces which derive the directed function in these complex systems.

6.4.4.1 *Channel gating*

The first of our examples comes from a study of transitions induced in a quasi-one-dimensional semiconductor, copper-7,7,8,8-tetracyanoquinodimethane (Cu–TCNQ), carried out using UEM-1. By serendipity it was discovered, in micron-length crystals, that the material exhibits a mechanical nanoscale molecular phenomenon, a switchable channel or *gate* (Fig. 6.43).[20] The control of the gate was made using the excitation shock induced in the sample by near infrared laser pulses. Remarkably, the switching, after a shock, is not only controllable with the pulses being on or off, but also returns the material in real space to the original structure. The functional behavior is robust in the relatively low-fluence regime. At significantly higher fluences, in the microscope, the internal dilation and the reduction of the copper ions was observed to form islands of neutral copper metal structures.

The crystallinity of the specimen was confirmed by selected-area electron diffraction in UEM. Cu–TCNQ is known to exist in both a kinetically favored state ("phase I," needle-crystal habit) and a thermodynamically favored state ("phase II," platelet-crystal habit). The material synthesized was that of phase I, as determined by the observed crystal habit and confirmed by indexing of the experimental diffraction pattern.

Figure 6.43 shows a Cu–TCNQ single crystal that has been shocked by initially applying a burst of the laser light. Following this shock, the crystal was observed to fracture and separate by tens of nanometers, depending upon the fluence of the initial pulse sequence. When the shutter was open (i.e., the sample was irradiated with laser pulses), the crystal expanded along the long axis and closed the gap completely; that is, the two exposed faces of the crystal came into full contact. When the shutter was closed (i.e., the specimen was not irradiated with laser pulses), the material returned to its original structure in which the channel was again clearly observed.

The experiments were then repeated by switching between the on and off modes. Four typical frames of the images obtained when the laser pulses are on or

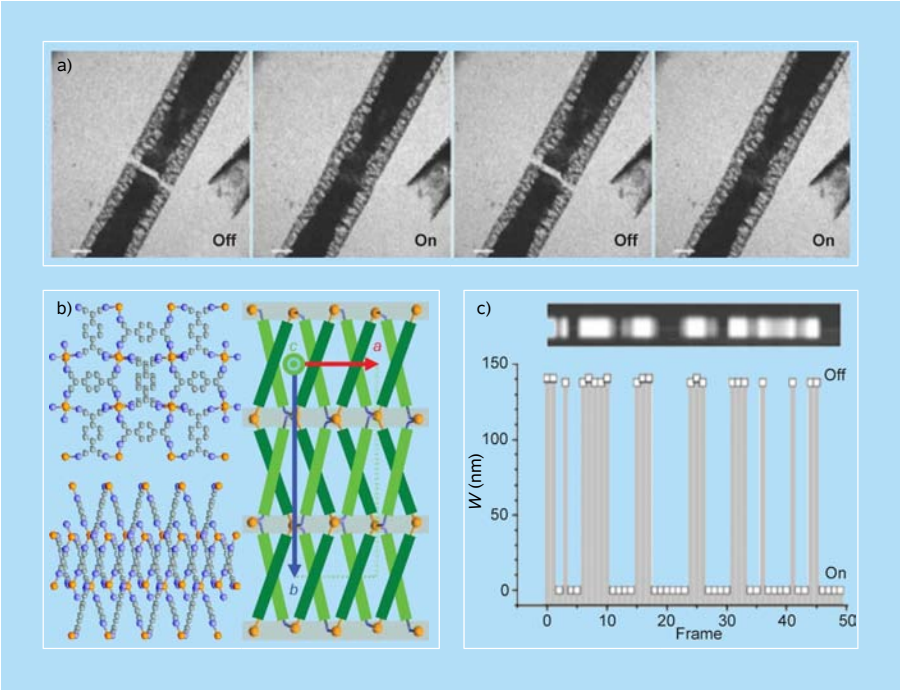


Figure 6.43 UEM images of channel gating. Images taken in the already shocked single crystal of Cu–TCNQ in the absence ("off") and in the presence ("on") of pulsed-laser irradiation (a), and a plot showing the results of a sequence of "on" and "off" cycles. The channel width *W* varies from 0 (pulsed laser irradiation) to 140(5) nm (no pulsed-laser irradiation) over a series of 50 frames. The bar in the inset is to illustrate the modulation in a memory recording (c). Crystal structure of Cu–TCNQ is shown in panel (b).[20,26]

off are given in the upper panel of Fig. 6.43. The width of the nanoscale channel that was formed in this case was equal to ~140 nm. The reversibility of gating is demonstrated in Fig. 6.43 by measuring the actual magnitude (distance) of the channel opening/closing while varying the number of frames recorded. The crystal was observed to expand and contract along the long axis in synchrony with the shutter. However, the contrast of the image in the open zone changes to give a relative value of 0.5 over the course of the recording time of 1 s, but the sharpness of the channel edges is maintained. From these experiments the time scale of the channel switching was estimated to be less than 10 ms, and more recently was accurately measured on nanosecond and longer time scales. The expansion/contraction was also observed in intact single crystals (i.e., not photo-shocked) as changes at the edges of the material.

The mechanism as obtained from the microscopic images and diffraction patterns involves a nonthermal electron transfer process. In single crystals of Cu–TCNQ charge-transfer complex, Cu^{+} and $TCNQ^{-}$, form discrete columnar

stacks in a face-to-face configuration with strong overlap in the π-system. Further, the copper atoms are bound in a four-coordinate, highly distorted tetrahedral geometry to the nitrogen atoms on the cyano groups of the TCNQ moieties. The strong through-space interactions between the π-electrons result in quasi-1D structure of the material in the solid state. As seen in the crystal structure in Fig. 6.43, the *a*-axis is unique for stacking. Hence, the large anisotropy of gating must reflect the unique changes along the stacking axis [100] and not perpendicular to it. The modulation (i.e., closing and opening) of the channel is a modulation of the π-electron interactions of the stacked TCNQ molecules as a result of the charge transfer from $TCNQ^-$ to Cu^+, and elasticity of the material. This same reduction process is responsible for the formation of metallic copper islands at higher fluences.

For expansion along the *a*-axis, the change is nearly linear until a relatively high fluence is reached. For the perpendicular direction, there is no significant change when compared with the longitudinal one. The "microscopic," sub-angstrom, changes at the atomic scale were checked against the "macroscopic," nanometer, channel formation. In the linear regime of expansion, it was found that the channel width decreased linearly as a function of fluence (i.e., the crystals expanded linearly along the *a*-axis) up to 0.28 $\mu J/cm^2$ per pulse, at which point the rate of crystal expansion became ill-defined owing to loss of crystallinity and formation of discrete domains corresponding to the individual atomic and molecular components of the material.

These results illustrate the significant change of the channel width with energy and the clear anisotropy along different directions, whereby the dominant change is along the stacking axis. They also explain why, at higher fluences, the Cu–TCNQ crystals are observed to lose all crystallinity and undergo a reductive formation of copper metal domains (along with strands of neutral TCNQ lying approximately perpendicular to the long axis of the crystal). In UEM, selected-area electron diffraction of these domains shows that they are composed of crystalline copper metal; thus the Cu^+ reduction process can be imaged at high fluences. The controlling excitation at 1.6 eV is not only the key to gating of the channel at lower fluences, but it is also sufficient to facilitate charge transfer and hence weakening of copper bonding and enhancement of its mobility at higher fluences. The surface energy of metal clusters composed of a few atoms is quite large, and there is a strong driving force for the smaller clusters to coalesce (reminiscent of Ostwald ripening). This behavior greatly enhances the formation of larger copper clusters and further reduces the ordered structure of the Cu–TCNQ material until finally, at relatively high fluence, the crystal separates into its constituent atomic and molecular components. The findings, especially at low fluences, may be of value in applications involving molecular nanoswitches and channels, as well as optical pulse memory.

6.4.4.2 *Functional cantilevers*

Another dimension to UEM is in the study of *in situ* structural transformations during a macroscopic function. The change is directly visualized in the variation of the properties of the material itself as it functions! As such, they can be used to discover and optimize nano(micro)electromechanical systems (NEMS and MEMS).[44,45] Beyond mass measurement and analytic detection, changes in the dynamics of these nanoscale structures have been shown to be sensitive to very weak external fields. The response to external stimuli is manifested in deflections of the nanoscale, and a variety of techniques have been used to both actuate and detect the small-amplitude deflections. While current methods of detection provide insight into the movement amplitude, real-space and time visualization of modes of oscillations at frequencies pitched in the ultrasonic range (i.e., kilohertz to gigahertz) has not so far been possible. In principle, the use of optical methods (e.g., reflection/interferometry) can provide the ultrafast temporal resolution but not the 3D spatial structure. On the other hand, conventional electron microscopy can image the structure in real space but has a temporal resolution of milliseconds, being limited by the frame rate.

The visualization of nano- and microscale cantilevers and the (resonance) oscillations of their mechanical motions has recently been reported.[26] The static structures were constructed from a tomographic tilt series of images, whereas the *in situ* temporal evolution was determined using the stroboscopic configuration of UEM. Cu–TCNQ formed the nano- and microscale systems under study. The optomechanical motions are triggered by the charge transfer from the TCNQ radical anion ($TCNQ^-$) to copper (Cu^+), as discussed above. More than a thousand frames were recorded to provide a movie of the 3D movements of cantilevers in time. The expansions were found to be colossal, reaching the micrometer scale, and the spatial modes are resolved on the nanoscale in the images (and angstrom scale in diffraction) with resonances of megahertz frequencies for the fixed–free cantilevers. From these results, Young's modulus, the force, and the energy stored in the cantilevers were obtained.

The tilt-series (tomography-like) images are shown in Fig. 6.44a; they provide the 3D coordinates of the cantilevers. The dimensions and protrusion angles of these free-standing crystals were characterized by taking static frames at different rotational angles of the substrate and made to construct the movie of their 2D projections. The 4D space–time evolution of cantilevers is shown in Figs. 6.44b,c. The referenced (to negative time, $t_{ref} = -10$ ns; i.e., before the arrival of the clocking pulse) images of the microscale (Fig. 6.44b) and nanoscale (Fig. 6.44c) free-standing single crystal clearly display modes of expansion on the megahertz scale. Each image illustrates how the spatial location of the crystal has changed

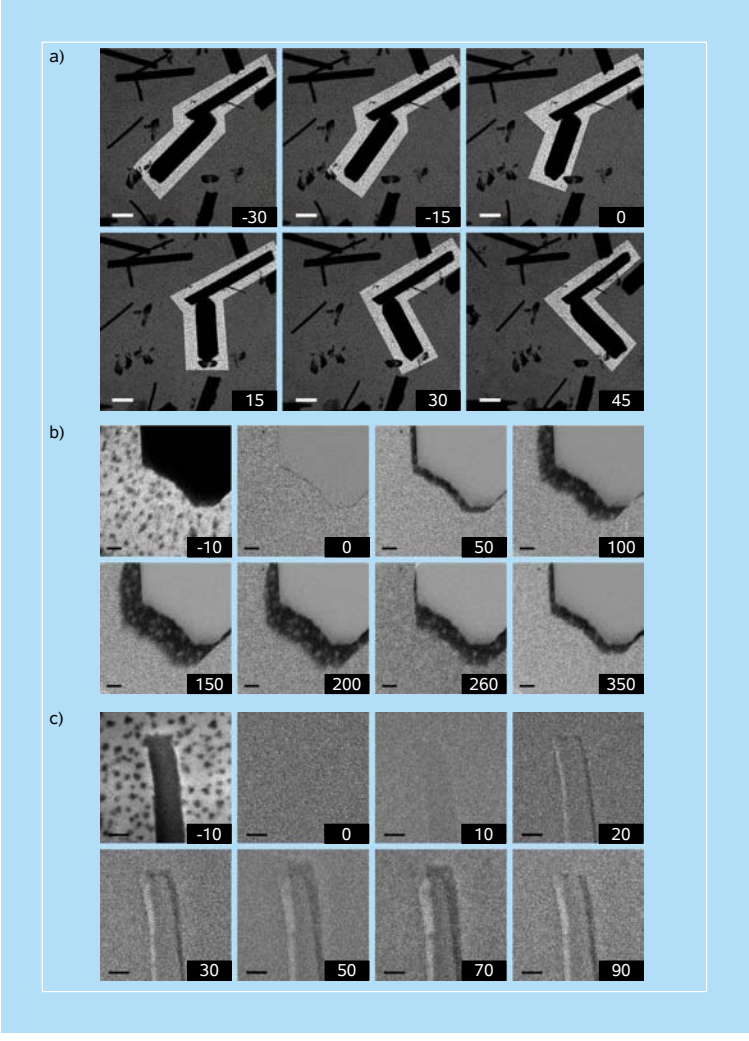


Figure 6.44 Tomography-like UEM micrographs of nanoscale and microscale cantilevers. (a) The tilt series of images with the frames showing the 2D projections of the Cu–TCNQ single crystals acquired at different tilt angles of the specimen substrate. The highlighted region illustrates a large change in the position of the free-standing microscale crystal relative to another, which is lying flat on the substrate, as the tilt angle is changed. The tilt angle at which each image was acquired is shown in the lower-right corner of each frame (in degrees). 4D electron micrographs of the microscale (b) and nanoscale (c) cantilevers are presented. The frames show referenced images with respect to the frame recorded at a negative time. The images illustrate the characteristic dynamics.[26]

relative to the reference image as a function of the time delay, elucidating both the longitudinal and transverse displacements relative to the at-rest position.

Shown in Figs. 6.45a–d is the displacement of the nano(micro)scale cantilevers as a function of time, in both the longitudinal and transverse directions, along with the Fourier transforms (FT) of the observed spatial oscillations for the

time range shown (i.e., 0–3.3 μs). The motions in both directions of measurement are characterized by a large initial displacement from the at-rest position. The scale of expansion is enormous. The maximum longitudinal expansion possible (after accounting for the protrusion angle) for the 10 μm crystal is 720 nm or over 7 percent of the total length.

Because of the boundary conditions of a fixed–free beam, the vibration modes are not evenly spaced and the overtones are not simple integer multiples of the fundamental flexural frequency (f_1) but rather occur at 6.26, 17.5, and 34.4 for f_2, f_3, and f_4, respectively. This is in stark contrast to the integer multiples of the fundamental frequency of a fixed–fixed beam. Taking 3 MHz to be the main fundamental flexural frequency of the microscale crystal, we can deduce Young's elastic modulus of the material. The expression for the frequencies of transverse (flexural) vibrations of a fixed–free beam is given by

$$f_n = \eta \frac{\pi\kappa}{8L^2} c \equiv \eta \frac{\pi\kappa}{8L^2} \sqrt{\frac{Y}{\rho}}, \tag{6.17}$$

where f_n is the frequency of the nth mode in hertz, L is the beam length at rest, Y is Young's modulus, and ρ is the density. The radius of gyration of the beam cross-section is κ and is given as $d/12^{1/2}$, where d is the thickness of the beam with rectangular cross-section. The values of η for the beam are 1.194^2, 2.988^2, 5^2, 7^2,…, $(2n-1)^2$, approaching whole numbers for higher η values. From the above equation, and knowing $\rho = 1.802$ g cm^{-3}, Young's modulus was obtained to be 2 GPa, with the speed of sound, therefore, being 1100 m/s. This value of Young's modulus (N/m^2) is in the range of that measured for the related material of TTF–TCNQ using a millimeter length vibrating reed under an alternating voltage.[46]

The potential energy stored and the force exerted at the moment of full extension along the long axis just after time zero (cf. Fig. 6.45a) can be estimated from the amplitudes and using Hooke's law

$$V = \frac{1}{2}\left(\frac{YA}{L}\right)\Delta L^2, \tag{6.18a}$$

$$F = \left(\frac{YA}{L}\right)\Delta L, \tag{6.18b}$$

where V and F are the potential energy and force, respectively, and A is the cross-sectional area of the crystal. The bracketed term in Eq. (6.18a) is the spring constant (assuming harmonic elasticity, below the plasticity range). By simple substitution, the values of 200 pJ and 600 μN were obtained for the potential

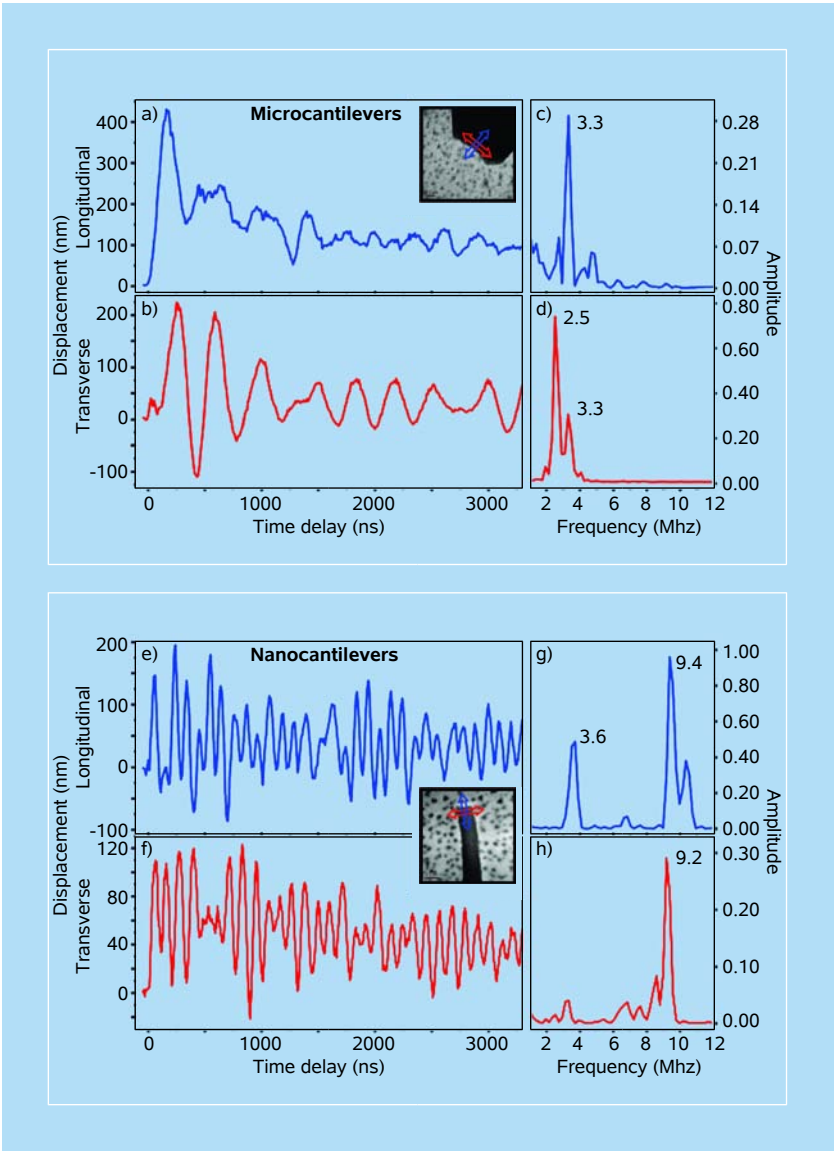


Figure 6.45 Oscillation (resonance) dynamics and frequencies of the cantilevers. Shown are the resonances, and their Fourier transforms, for the micro (a–d) and nano (e–h) cantilevers. The initial displacement and subsequent oscillatory motion of the Cu–TCNQ crystal were quantified in the longitudinal (a,e) and transverse (b,f) directions. The longitudinal and transverse directions are defined as being parallel and perpendicular to the crystallographic *a*-axis, respectively, and are illustrated by the blue (longitudinal) and red (transverse) directional arrows in the figure insets. All measured crystal positions are relative to the at-rest position at a negative time delay. Panels (c,g) and (d,h) display the frequencies in the longitudinal and transverse directions, respectively, which are obtained by FT of the displacement responses shown in panels (a,e) and (b,f). The prominent frequencies are labeled, and the accuracy of the weaker peak amplitude depends on sampling resolution, which in this case is 0.1 MHz.[26]

energy and force, respectively, considering the maximum possible expansion of 720 nm; even when the amplitude is at its half value (see Fig. 6.45a), the force is very large (~300 μN). In other words, owing to molecular stacking, the force is very significant. Also, because of the microscale cross-section, the pressure of expansion translates to 0.1 GPa, only a few orders of magnitude less than the pressures exerted by a diamond anvil.

Knowing the dimensions of the material, and the laser fluence and absorptivity, the maximum pulse energy absorbed by the crystal is 30 nJ, and this means that, of the initial optical energy, a minimum of ~1 percent is converted into mechanical motion. But in fact, it could reach 10 percent or more as determined by the projection of the electric field of light on the crystal. The resonance frequency should scale with L, and indeed when these studies were extended to other sets of beams, namely, those of reduced dimensions, higher frequency nanoscale (cantilevers) resonance modes were observed and studied similarly (see Figs. 6.45e–h). The observed function is robust, at least for 10^7 continuous pulse cycles (~10^{11} oscillations for the recorded frames), with no damage or plasticity. With these imaging methods in real time, and using other variant techniques of UEM, it should now be possible to test the various theoretical models involved in MEMS and NEMS, and to design relevant structures.

6.4.4.3 *Optoelectronic nanorods*

Quantum dots and nanotubes are examples of nanostructures which have unique length scale and architecture and possess the potential for unique applications. Among the materials that show promising features for optoelectronics, such as blue-green laser diodes and photonic devices, the wide-gap semiconductor zinc oxide (ZnO) has been one of the most investigated nanorod materials.[47] At the nanoscale, their design requires the fundamental understanding of electronic and nuclear degrees of freedom because they control the effective carrier mobility and properties along the rods. With optical methods, the transient response can be probed but, because of the wavelengths involved, the atomic-scale structural changes are not determined.

Structural dynamics of vertically aligned ZnO nanorods (Figs. 6.46 and 6.47) fabricated through physical vapor deposition (Fig. 6.47a, high-density assembly) and hydrothermal synthesis (Fig. 6.47c, low-density assembly) on ZnO-covered mounting substrates were recently visualized.[14] Besides characterization by scanning electron microscopy (SEM), Figs. 6.47a,c, the specimens were also studied using TEM and X-ray diffraction. The heating pulse precedes in time the (probing) electron pulse and is used to excite the carriers of the rods. From the intensities, positions, and widths of the Bragg diffractions, the change of atomic positions with time, the structure factor, and the transient inhomogeneity in the rods were

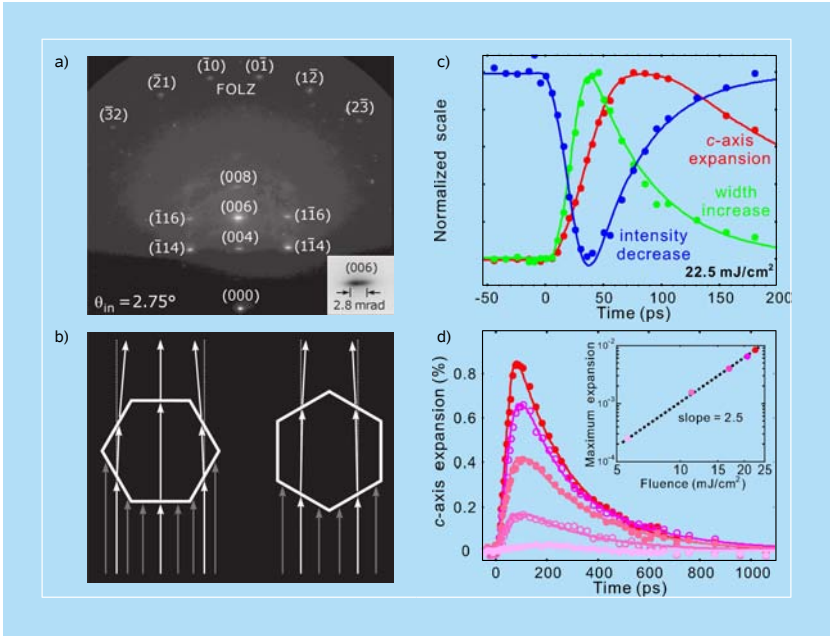


Figure 6.46 Structural dynamics of ZnO nanorods. (a) Pancake-type diffraction and shape effect of the nanorods. The diffraction patterns were obtained at an incidence angle of $\theta_{in} = 2.75°$, with the electron beam propagating along the [110] direction. From the well-indexed Bragg spots and FOLZ diffractions, the wurtzite ZnO structure was determined; the undiffracted electrons form the (000) spot. The inset shows the enlarged horizontal width of the (006) Bragg spot. (b) Refraction of the probing electrons at the vacuum-material interface and the overall beam deflection caused by the different orientations of a nanorod (top view). (c,d) Changes of the *c*-axis expansion, diffraction intensity and width of the (006) Bragg spot with time. (c) The intensity decrease (blue dots) and width increase (green dots) behave similarly, but with different time constants, and precede the build up of structural expansion (red dots). In ~200 ps the former two diffraction features almost recover the original values, whereas the decay of expansion appears on a longer time scale (nanosecond). (d) A significantly larger *c*-axis expansion was obtained at higher excitation fluences. The fitted slope in the log–log plot (inset) indicates that the maximum expansion is proportional to the fluence to the power of 2.5, and similar results were obtained for the intensity change. At all fluences, however, the expanded nanorods nearly return to their original structure in ~1 ns. A more significant intensity decrease and a larger width increase are observed following a stronger excitation. Care was taken to check for surface charging (see Sec. 6.4.4.4).[14]

obtained (Fig. 6.46). By varying the fluence, the influence of carrier density was also examined, and in this case the rods were directly excited. Finally, by using two arrays of nanorods, the effect of the deposition density was elucidated.

We first determined the hexagonal static geometry of nanorods through observation of "pancake-type" diffraction.[14] The pancake horizontal shape of the Bragg spots (Fig. 6.46a) is distinct from the spots seen in bulk crystal diffraction and mirrors the hexagonal nanostructured rod. Because of the rods' nanoscale geometry and the attractive inner potential V_I inside the material, the diffracted electron beam is refracted and gives rise to a horizontal spread. Quantitatively, this

geometry effect can be described by considering the analog of Snell's law in optics, giving

$$\sin\theta_{\text{vac}} = \sqrt{1 + eV_{\text{I}}/E_{\text{k}}}\,\sin\theta_I, \tag{6.19}$$

where θ_{vac} and θ_I are, respectively, the angles in the vacuum and inside the rods, with respect to the surface normal, and E_{k} is the kinetic energy of the incident electron, which in our case is 30 keV (Chap. 2). In general, when the surface layers are perfectly parallel and $eV_{\text{I}} \ll E_{\text{k}}$, as in typical UEC experiments, this effect is small, but for the nanorods the situation is different (Fig. 6.46b), and the refraction can be invoked to obtain the geometry of the rods.

For a hexagonal rod, with the two orientations of 30° rotation difference (Fig. 6.46b), there are two angles of deflection, $\sqrt{3}V_{\text{I}}/E_{\text{k}}$ and $V_{\text{I}}/\sqrt{3}E_{\text{k}}$. Thus, the intensity profile of the pancake is the sum of the contributions from the center and the two sides of the hexagon. For a cylindrical rod, the intensity profile is

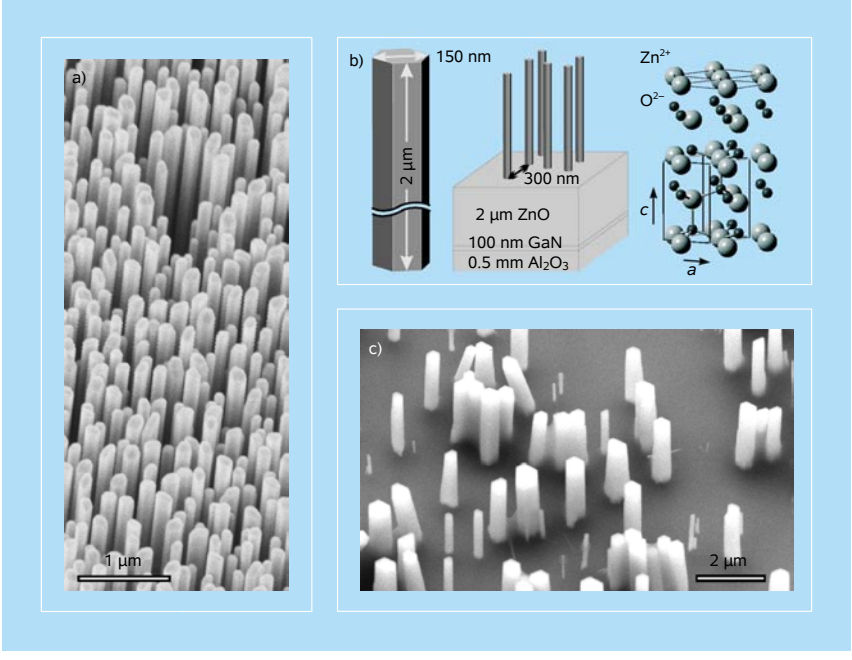


Figure 6.47 Scanning electron micrographs and structure of ZnO nanorods. (a) The SEM image of the array fabricated through physical vapor deposition. The high-density rods are vertically aligned. Gold nanoparticles may exist dispersively as a result of the catalyst-assisted synthesis. (b) The structure together with a schematic showing the average diameter and length of a single rod, the average spacing between rods, and the layer composition below the ZnO array sample. (c) The SEM image of the array fabricated through hydrothermal synthesis. The vertically aligned rods are thicker, on average, and form an array with a much lower density. No gold nanoparticles exist in the nanorods since a metallic catalyst is not required for this wet-chemistry synthesis method.[14]

determined by the deflections, $\tan\theta_{vac}(V_I/E_k)$. The difference between the two geometries is easy to discern: if circular, the rotation of the nanostructure would give no variation in the horizontal spot width, which is inconsistent with the experimental observation. For ZnO, V_I is 15.9 V. With the camera being 16.8 cm away from the nanorods, additional diffraction intensities near ±0.92 mrad and ±0.31 mrad (Figs. 6.46a,b) result in the horizontal spread from the spot center. Accordingly, this shape effect properly accounts for the observations made for the two zone axes studied. Moreover, from the fact that the probing direction in Fig. 6.46a is [110] we determine the facets of nanorods to be of {110}, which is again consistent with the preferential growth. It should be noted that the vertical spot width did not increase, a consequence of the unique elongated geometry.

After electronic excitation of the specimen, the rods were found to exhibit large expansions, two orders of magnitude higher than that expected at thermal equilibrium; the expansion is highly anisotropic, quasi-one-dimensional behavior which is facilitated by the induced antibonding character. When the maximum *c*-axis expansion (or Bragg diffraction intensity) is plotted as a function of fluence in a log–log plot (Fig. 6.46), the dependence follows a linear relationship with a slope value near 3; similar behavior was found in the intensity change. The change of the expansion, intensity or width at different fluences occurs immediately after the zero of time, which is indicative of the fact that the rods are excited *directly* (i.e., not through the substrate). Finally, the *c*-axis expansion is higher in amplitude for the low-density array and rises faster than that found for the high-density array. When carrier density is relatively high, especially at high fluences, care has to be taken to examine the effect of surface charging (potential) on the outgoing electron beam, particularly for low incident angles; but such interferences are examined experimentally as discussed in the following section for nanorods and other systems.

With this in mind, the observations made for ZnO suggest that the rods display collective, nonequilibrium behavior, and this behavior is anisotropically driven by the potential of the carriers generated by the excitation, i.e., by the electronic change in bonding, the antibonding Zn–O character along the rod direction. Besides the temporal behavior observed, several observations support this conclusion. First, the restructuring for the expansion occurs on a much longer time scale than those of the intensity and width, and it is similar to the decay time of photoluminescence (radiative recombination) of ~1–2 ns. Second, the horizontal expansion is absent at lower fluences and at the highest fluence changes on a shorter time scale. Finally, as the excitation fluence increases, the density of carriers increases and the larger the observed induced longitudinal expansion and intensity decrease. Other types of nanostructures exhibiting different optical and electronic properties should be studied similarly to elucidate the relationship of structural dynamics to properties.

6.4.4.4 *Diffraction and materials surface charging*

In this section, we address issues of significance to studies that may involve large structural changes, and also when poor-conducting materials are under investigation, especially at high fluences. Care must be taken to examine the effect of surface charging on the deflection of the outgoing electron beam, particularly for low-incident angles. We shall illustrate these points by considering the studies made on graphite, gallium arsenide, and ZnO nanorods.

Schäfer and Liang of Caltech have made careful and systematic studies of diffraction of graphite (Sec. 6.3.4), exploring the dependence of the shift and intensity change of Bragg spots with time on the pulse fluence. They also examined the influence of crystal quality and inhomogeneity on diffraction patterns, invoking both LEED and SEM for single crystals and HOPG specimens. Of relevance here are the results of Fig. 6.48 in which two representative diffraction orders are shown in panel (a); other orders were similarly studied. In the same figure, the change of the (008) and (0014) diffraction spots is displayed in panel (b) as a function of time (raw data), for the single crystal where inhomogeneity is nearly absent. It is evident that both the intensity change and the peak shift

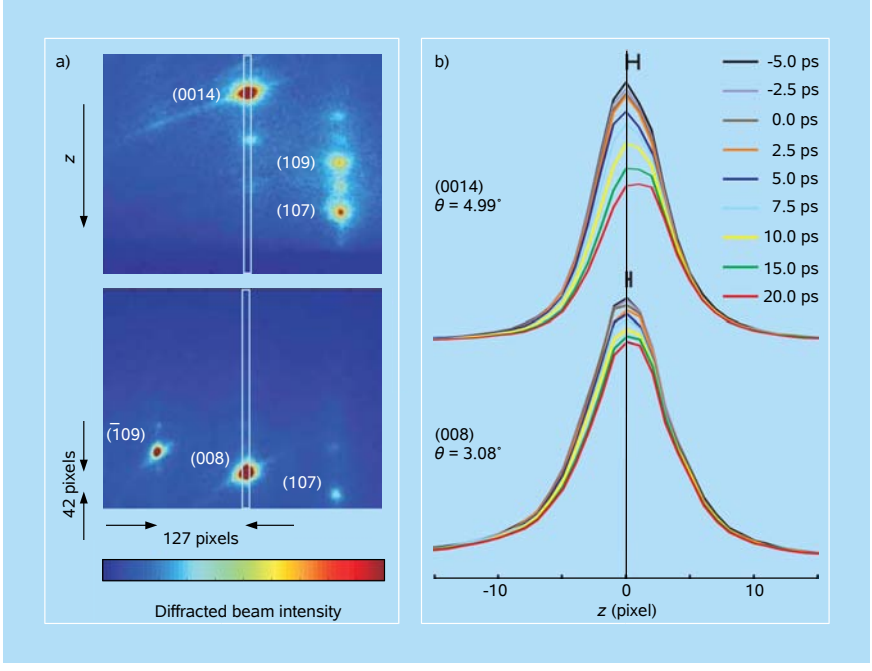


Figure 6.48 Diffraction patterns and temporal evolution of Bragg spots. Shown are the results for single-crystal graphite, and for two orders of diffraction: (008) and (0014). Other spots were examined similarly.

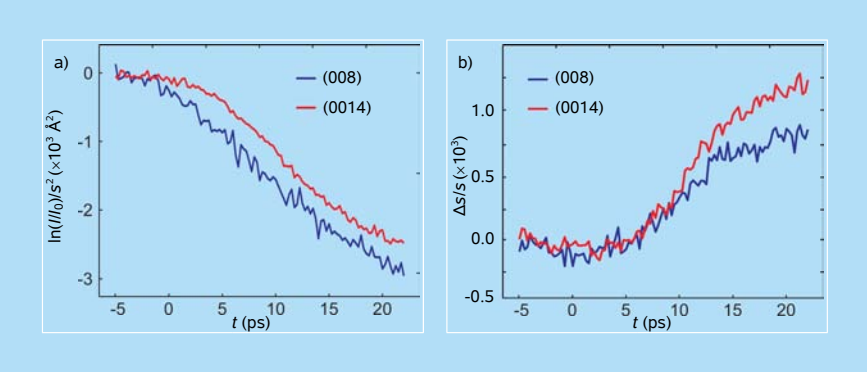


Figure 6.49 The temporal behavior of Bragg diffraction. Shown are the movements of spots and decrease in intensity, both when properly scaled according to the theoretical dependence on the scattering vector for different diffraction orders.

depend on the order of diffraction, being larger for the higher order peaks and consistent with results obtained for other systems; see Figs. 6.14 and 6.38. Further quantification of the data is given in Fig. 6.49 where plots are made of the change, but taking into account the expected order dependence of intensity [Debye-Waller s^2 functional form; Eq. (2.50)] and shift [the $\Delta s/s$ functional form; Eq. (2.48)]. Indeed, within the signal-to-noise ratio (being somewhat lower for the lower orders) the agreement elucidates that the observed dynamics is that of the actual structural change, and not due to other surface effects at the fluence used (~10 mJ/cm^2). The fluence was determined from measurements of pulse energy and the longitudinal and transverse beam profile at the specimen.

For the (0012) Bragg spot, the dependence of the temporal behavior on fluence was similarly examined; see Fig. 6.50. The behavior is consistent with the

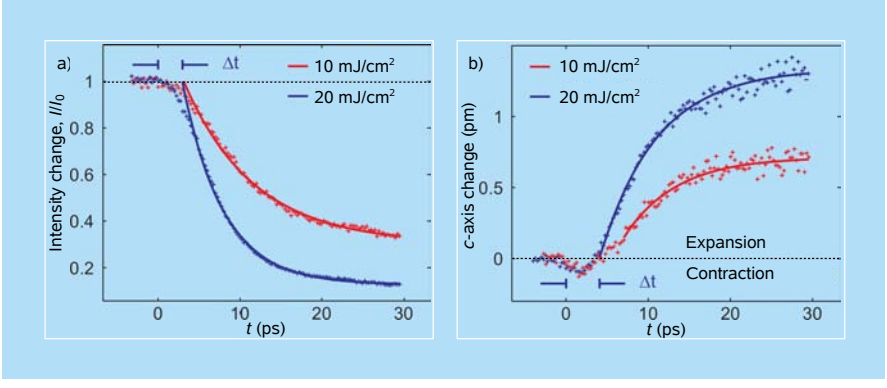


Figure 6.50 The dependence of the temporal evolution on the fluence. Note the change in amplitude, and shift in time (Δt).

results reported in Sec. 6.3.4.1, displaying the contraction/expansion features and the increase of the amplitude at higher fluences. However, the form of transient is robust at both values of 10 and 20 mJ/cm^2 fluences. The time constant of the expansion is reproduced to be ~7 ps. The contraction begins at $t = 0$ and persists over ~3 ps, whereas the intensity lags for that amount of time. As mentioned earlier, the contraction is a manifestation of collective (coherent) interactions prior to any expansion, and intensity changes require incoherent lattice motions to occur only following the electron-phonon coupling.

Results of recent optical studies of graphite are also consistent with these findings. Using optical reflectivity,[37a] Petek and his group have shown that the phonon frequency (involving in-plane motions) increases, not decreases, upon excitation (0.1–1 mJ/cm^2) of graphite and this unexpected observation was supported by theoretical calculations. The Raman shift of modes is behaving similarly[37b] and the decay of the modes is in ~2 ps, similar to that reported in Ref. 37a. One of the most interesting observations that relates to our finding in graphite was made by Nasu and colleagues[37c] in a STM study with 10^4 laser shots of ~60 mJ/cm^2 fluence. Regions were formed irreversibly of diamond-type structures (reminiscent of our FEELS findings in Sec. 6.3.4), only when femtosecond, not nanosecond, pulses were utilized. At about 20 mJ/cm^2 fluence, Ruan and his group[37d] identified transient sp^3-type similar structures from analysis of UEC diffraction. In one publication,[37e] they have shown that the contraction indeed occurs in the first few ps, and in another publication[37d] they reported such behavior but on the 100 ps time scale, at relatively high fluences.

On such long time scales, and at very high fluences, one must consider electron beam deflections from surface charging as a result of possible ionization and recombination of electrons in the material. During such processes introduced by the laser pulse, transient "plasma-type" charge separation occurs and causes a deflection of the electron beam. However, the time scale of such processes is far from the fs–ps regime, and in fact this effect was used earlier in UED studies[48] to determine, from the onset of the change, the *in situ* $t = 0$ of the studied transients. In Fig. 6.51, displayed is the deflection behavior for both GaAs and ZnO, emphasizing that for all studies made, the direct (undiffracted) beam is monitored as a function of time to ensure that its change in amplitude is negligible on the scale of the observed diffraction.[49] As shown in the figure, neither the magnitude nor the time scale (or phase of the signal) is relevant to the structural dynamics studied at the fluences used. Moreover, in UEM, with its unique geometry and very high energy of acceleration (200 keV), such a concern is of no substance, as the deflection angle reaches micro (to nano) radians at 10 mJ/cm^2. Thus, with the discussed experimental care and understanding of such effects, the reported results by Park and Zuo[50] become of interest from a theoretical point of view.

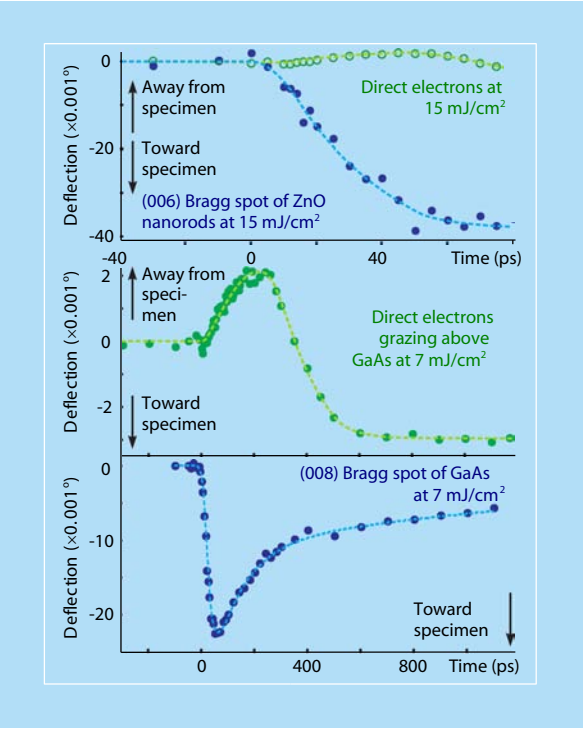


Figure 6.51 Behavior of direct (unscattered) and diffracted electrons for two cases, ZnO and GaAs. Note the relatively small magnitude and difference in time scale (and phase) depending on fluence.[49]

6.5 4D Convergent Beam UEM: Nanodiffraction

As discussed throughout this chapter, atomic-scale determination of structures can be achieved by using Fourier-space diffraction data, but this approach suffers from the averaging over the selected specimen area which is typically on the micrometer scale. Significant progress in techniques has enabled localization of diffraction to nanometer and even angstrom-sized areas by focusing a condensed electron beam onto the specimen. Parallel illumination with a single electron wavevector is reshaped to a convergent beam with a span of incident wavevectors. This method of convergent beam electron diffraction (CBED), or electron microdiffraction,[51–53] has made possible determination of structures in 3D with highly precise localization to areas reaching below one unit cell. The applications have been wide-ranging, from revealing bonding charge distribution[51,52] and local defects and strains in solids[54] to detecting local atomic vibrations and correlations.[55] Today, aberration-corrected, atomic-sized convergent electron beams enable analytical probing using EELS[56a] and STEM.[56b] The above-mentioned studies all probed static, or time averaged, structures.

With time resolution, the spatial resolution in the diffraction investigations discussed here and reported from other laboratories[37d,57] corresponds to atomic (and sub-atomic) length-scale, but the probed specimen's area is square microns large, as mentioned above, albeit a significantly reduced volume when compared with X-ray diffraction (see overviews in Refs. 9 and 58). Thus, these techniques, as with those based on parallel beam illumination, lack the capability of resolving nanometric heterogeneity, for instance in nm-sized particles or specific sites in an extended material. Real-space UEM, as shown above, can resolve such objects by selected-area imaging, but the diffraction from a selected area is typically averaged over microns or more. Accordingly, the analytical and quantitative information available from this parallel-beam diffraction within such areas is somewhat limited. It follows that nanoscale (and finer) time-resolved, convergent-beam diffraction would be of great use in materials science and biology for structural dynamics studies in heterogeneous media.

The development of CB-UEM was recently reported[59] with applications in the study of nanoscale, site-selected structural dynamics initiated by ultrafast laser heating (10^{14} K/s). With the time resolution being ten orders of magnitude improved from that of conventional CBED, and because of beam convergence, high-angle Bragg scatterings can be spatiotemporally resolved with their intensities being very sensitive to both the 3D structural changes and amplitudes of atomic vibrations. The CB-UEM configuration is shown in Fig. 6.52; the chosen specimen was a crystalline silicon slab, a prototype material for such investigations. From these experiments, it was found that the structural change within the locally probed site occurs with a time constant of 7.3 ± 3.5 ps, which is on the time scale of the rise of lattice temperature known for bulk silicon from the spectroscopic studies of Sjodin *et al.*[60] For these local sites, the temperatures measured at different laser fluences range from 299°K to 366°K, corresponding to vibrational amplitude changes from 0.077 Å to 0.084 Å, respectively. The reported results would be impossible to obtain with conventional, parallel beam diffraction.

In these and subsequent studies, the electron packets were accelerated to 200 keV, de-magnified, and finally focused to various areas, from 10 to 300 nm in diameter, on the specimen, as shown in Fig. 6.52. A wide range of thicknesses, starting from ~2 nm, was accessible simply by moving the electron beam laterally. In the microscope, Kikuchi lines were observed and used as a guide to orient the specimen with the [011] zone axis either parallel or tilted relative to the incident electron beam direction. In Fig. 6.52, typical high-magnification (high-value camera length) CB-UEM patterns are displayed. Unlike parallel-beam diffraction which yields spots, convergent-beam diffraction produces discs in reciprocal space (back focal plane of the objective lens) with their diameter given by the convergence angle (2α) of the electron pulses. These discs form the Zero Order Laue Zone (ZOLZ) of the pattern; they show white contrast with thin specimens

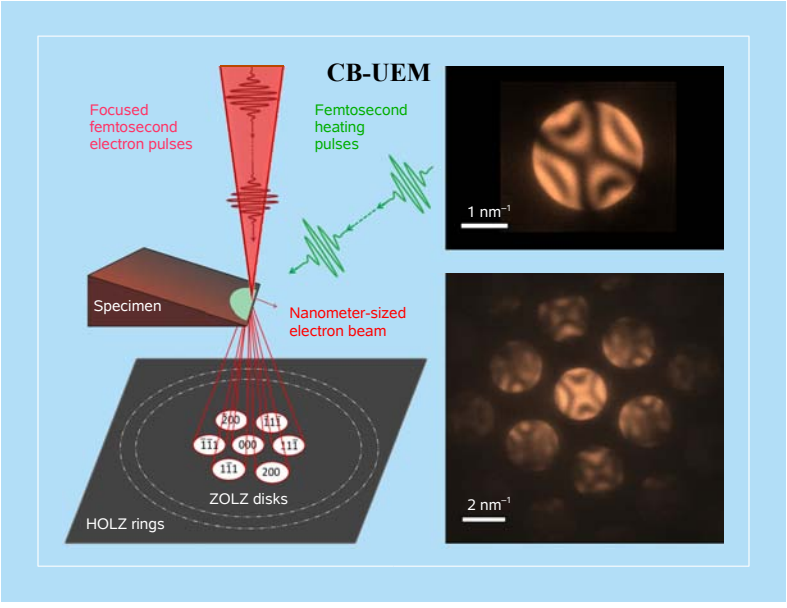


Figure 6.52 Schematic of the CB-UEM set-up, and observed low-angle diffraction discs. Femtosecond electron pulses are focused on the specimen to form a nanometer-sized electron beam. Structural dynamics are determined by initiating a change with a laser pulse and then observing the consequences using electron packets delayed in time. Insets (right) show the CB-UEM patterns taken along the Si [011] zone axis at different magnifications. At the high camera length used, only the ZOLZ discs indexed in the figure are visible; the kinematically-forbidden 200 disc appears as a result of dynamic scattering. Figures 6.52–6.55 are adapted from Ref. 59.

and exhibit the interference patterns displayed in Fig. 6.52 when the thickness is increased, as expected.

In the reciprocal space, the effective thickness of the Ewald sphere is 2α (Fig. 6.53), giving rise to multiple spheres that can intersect with Higher Order Laue Zones (HOLZ) reflections (Fig. 6.54), the key to 3D structural information; the first and second zones, FOLZ and SOLZ, are examples of such zones or rings. The interference patterns in the disks are the result of dynamical scattering in silicon[51–53] and are reproduced in our CB-UEM patterns (Fig. 6.52). Figure 6.54a presents the HOLZ ring taken with the CB-UEM. The scattering vector of the Bragg points of the ring, from the direct beam position, was measured to be 2.2 $Å^{-1}$, close to the value of 2.22 $Å^{-1}$, giving the known lattice separation of 5.4 Å for silicon. In addition to the SOLZ ring, Kikuchi lines and some oscillatory bands are also visible in the CB-UEM, as seen in Fig. 6.54a. Kikuchi lines arise from elastic scatterings of the inelastically scattered electrons, whereas the oscillatory bands in the thermal diffuse scattering (TDS) background result from correlations between the atoms.[55b] We also observed deficit HOLZ lines and interference fringes in ZOLZ discs for a two-beam condition.[53]

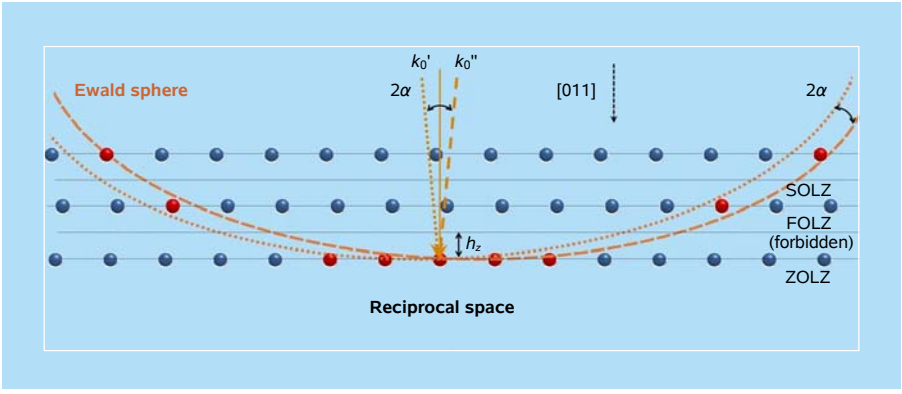


Figure 6.53 Reciprocal space representation of CB diffraction. The Ewald sphere has an effective thickness of 2α, the convergence angle of the electron beam. The diamond-like structure of Si forbids any reflections from odd numbered Laue planes when the zone axis is [011].

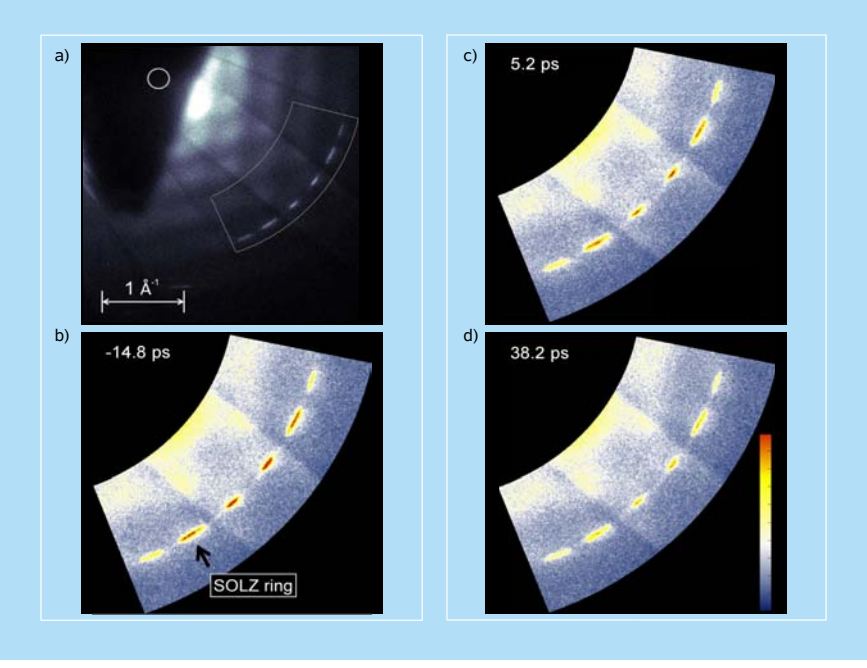


Figure 6.54 CB-UED temporal frames of nanoscale speciemens. (a) High-angle SOLZ rings obtained for a tilt angle of 5.15° from the [011] zone axis. Besides SOLZ, Kikuchi lines and periodic bands (due to atomic correlations) are visible. The ZOLZ discs are blocked (top left) to enhance the dynamic range in the area of interest; the disc of the direct beam (the center one in Fig. 6.52 discs) is indicated by a circle. (b–d) Time frames of the SOLZ ring are shown by color mapping for visualization of dynamics. The intensity of the ring changes within picoseconds, but the surrounding background remains at the same level.

The temporal behavior is displayed in Fig. 6.54, with three CB-UEM frames taken at time delays of $t = -14.8$ ps, +5.2 ps, and +38.2 ps, together with a static image; the zero of time is defined by the coincidence of the pump and probe pulses in space and time. The frame at negative time has higher ring intensity than that observed at +38.2 ps, whereas the +5.2 ps frame shows an intermediate intensity value. It is clear from the results that the intensity change is visible within the first 5 ps of the structural dynamics. For quantification, the intensities in each frame were normalized to the area of azimuthally integrated background. The normalization of the HOLZ ring intensities to the TDS background makes the atomic vibration estimations insensitive to the thickness changes of the probed area which may result from slight beam jittering.

Figure 6.55a depicts the transient behavior of the SOLZ ring intensity for two different laser powers, 10 mW and 107 mW, corresponding to pulse fluence of 1.7 and 19 mJ/cm^2, respectively; the heating laser beam diameter on the specimen is 60 μm.[24] The intensities were normalized to the average value obtained at negative times. Whereas the intensity change is essentially absent in the 10 mW data, the results for the 107 mW set shows a transient behavior with a characteristic time of 7.3 ± 3.5 ps, obtained from the mono-exponential fit shown in red in the figure. As discussed in this chapter, the temporal response of UEM-2 is on the fs time scale, as obtained by EELS, and it is much shorter than the 7 ps reported here. The effective lattice temperature is ultimately established depending on density of carriers or fluence. However, in CB-UEM measurements the lattice-temperature rise could be slower than in bulk, depending on the dimension of the specimen relative to the mean free path of electrons in the solid.

The dynamical change can be quantified by considering a time-dependent Debye-Waller factor (see Chap. 2) with an effective temperature describing the decrease in the Bragg spot intensity with time. If the root-mean-square displacement of the atoms, $\langle u_x^2 \rangle^{1/2}$, along one of the three principle axes is denoted by u_x for simplicity, then the HOLZ ring intensity can be expressed as

$$I_{\text{Ring}}^{F}(t) = I_0(t_-) \exp[-4\pi^2 s^2 u_x^2(t)], \tag{6.20}$$

where $I_{\text{Ring}}^{F}(t)$ is the measured intensity for a given fluence, F, and the vibrational amplitude is now time dependent. Note that u_x^2 is 1/3 of the total, u_{total}^2.[b]

[b]In the standard formula for the Debye-Waller effect, three components of vibrational amplitudes are considered and the exponent becomes $\exp[-(4/3)\pi^2 s^2 u_{\text{total}}^2]$. All three components in the Einstein model are equal and, hence, $u_x^2 = 1/3 u_{\text{total}}^2$. From an experimental point of view, the intensities could be expressed as $I_{\text{Ring}}^{F}(t) \sim \exp[-4\pi^2 s^2 u_x^2(t)]/(1 - \exp[-4\pi^2 s^2 u_x^2(t)])$, where the denominator represents the TDS background.[61] However, under our experimental conditions, this background is insignificant in the narrow angular range studied. Moreover, we note that TDS and other background contributions, such as inelastic plasmon scattering, which can be filtered out,[61] are not of concern here as we are monitoring the temporal change in reference to frames recorded at, e.g., negative delay times.

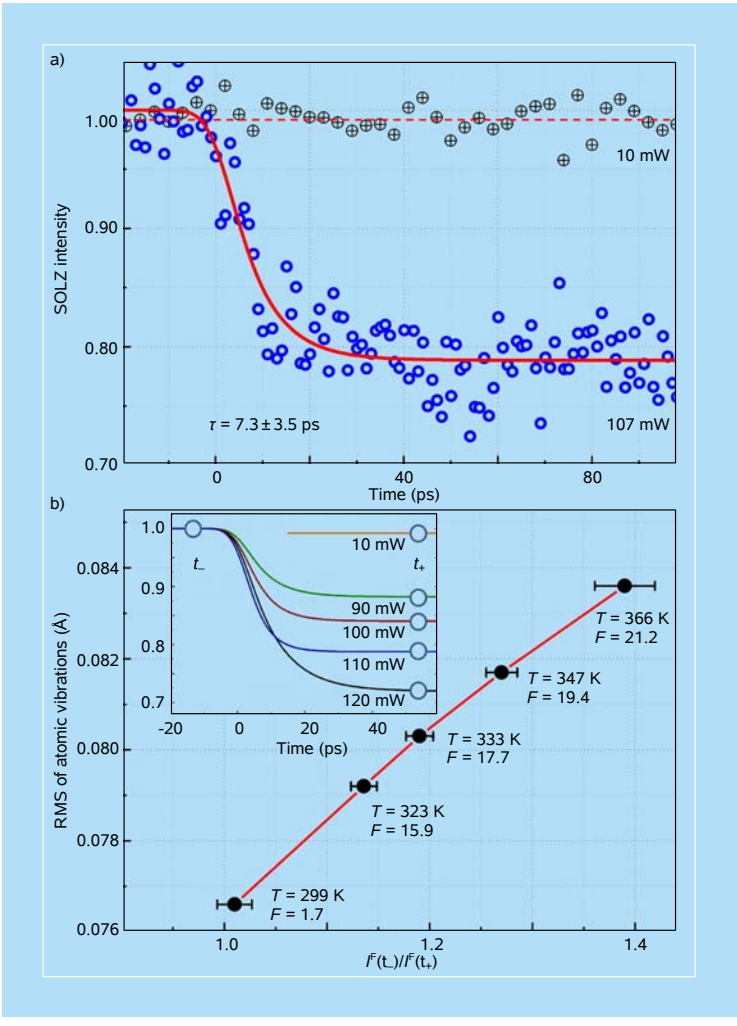


Figure 6.55 Nanodiffraction intensities, and amplitudes of atomic vibrations. (a) Shown are normalized, azimuthally-integrated intensity changes of the SOLZ ring. Whereas the 10 mW response does not show noticeable dynamics, the 107 mW transient has a clear intensity change with a characteristic time of 7.3 ± 3.5 ps. The red curve is a mono-exponential fit based on the Debye–Waller effect, as discussed in the text. The red dashed line through the 10 mW data is an average of the points after +20 ps. The dependence on fluence is given in panel (b), where root-mean-square amplitudes of atomic vibrations are plotted against the observed intensity change. The inset shows the mono-exponential temporal behavior, with the asymptotes highlighted (circles) for their values at different fluences. For each fluence, the temperature represents the effective value for the lattice structural change (see text).

In the Einstein model of atomic vibrations, which has been used successfully for silicon,[55a,61] the atoms are treated as independent harmonic oscillators, with the three orthogonal components of the vibrations decoupled. As a result, a single frequency (ω) is sufficient to specify the energy eigenstates of the oscillators.

The relationship of the vibrational amplitude to temperature can be established by simply considering the Boltzmann average over the populated eigenstates. Consequently, the probability distribution of atomic displacements is derived to be of Gaussian form, with a standard deviation corresponding to the root-mean-square (u_x) of the vibration involved[55a]

$$u_x = [(\hbar/2\omega m)\coth(\hbar\omega/2k_{\mathrm{B}}T_{\mathrm{eff}})]^{1/2} \tag{6.21}$$

where $\hbar$ is Planck's constant, k_{B} the Boltzmann constant, T_{eff} in this case the effective temperature, and m the mass of the oscillator. In the high temperature limit, i.e., when $\hbar\omega/2k_{\mathrm{B}}T \ll 1$, Eq. (6.21) simplifies to $m\omega^2 u_x^2 = k_{\mathrm{B}}T$, which is the classical limit for a harmonic oscillator; the zero-point energy, which contributes almost half of the mean vibration amplitude at room temperature, is included in Eq. (6.21). The value of $\hbar\omega$ is 25.3 meV. Despite its simplicity, the Einstein model in Eq. (6.21) was remarkably successful in predicting the HOLZ rings and TDS intensities by multi-slice simulations,[55a] although it was not as successful in predicting the oscillatory bands in the background which are due to correlations.[55b]

In Fig. 6.55b, the change in the asymptotic intensity with fluence (inset), and the derived vibrational amplitudes for the different temperatures are presented. The amplitudes are directly obtained from Eq. (6.20), as s is experimentally measured. The relative temperature change (from t_- to t_+) is then derived from Eq. (6.21), taking the value of u_x at room temperature (297°K) to be 0.076 Å (from CBED and X-ray measurements) for the unexcited specimen. The amplitude of atomic vibrations, and hence the temperature, increases as the fluence of the initiating pulse increases. Although the trend is expected for an increased u_x with temperature, the absolute values, from 0.077 to 0.084 Å, correspond to a large 3.2% to 3.6% change in nearest neighbor separation; these values are still well below the 15% criterion for a melting phase transition.

The linear thermal expansion coefficient has been accurately determined for silicon and the vibrational amplitudes reported here are much higher than the equilibrium thermal values at the same temperature. This is because the effective temperature applies to a lattice arrested in a picosecond time window; at longer times, the vibrations equilibrate to a lower temperature as discussed in the examples given in this chapter. As such, measuring nanoscale local temperatures on the ultrashort time scale enhances the sensitivity of the probe thermometer by orders of magnitude. Moreover, the excitation per site is significantly enhanced, as emphasized earlier. For a single-photon absorption at the fluence used, the number of absorbed photons per Si atom (for the fs pulse employed) is estimated to be ~0.01, as opposed to 10^{-9} photons per atom if the experiments were conducted in the time-averaged mode.

The achievement of nanoscale diffraction with convergent-beam ultrafast electron microscopy opens the door to exploration of different structural, morphological, and electronic phenomena. The spatially focused and timed electron packets enable studies of single particles and structures of heterogeneous media. Extending the methodology of CB-UEM to other variants, such as EELS, STEM and nanotomography,[62] promises possibilities for mapping individual unit cells and atoms on the ultrashort time scale of structural dynamics. These nano-, and below, diffraction probes in 4D electron microscopy are not achievable by other techniques (see Chap. 7).

6.6 4D Near-Field UEM: Nanostructures and Plasmonics

In UEM, photons are typically used to initiate a change (clocking) for the study of structural or morphological dynamics, which occur on the femtosecond (fs) and longer time scale, but, prior to these structural changes, electronic distributions are altered, with their dynamical changes being on the fs and shorter time scale. It is possible to exploit strong interactions between photons and electrons to open up a new avenue in UEM investigations.

In free space, an electron cannot directly absorb a quantum of electromagnetic energy because of the lack of energy-momentum conservation. However, as Kapitza and Dirac suggested in 1933,[63] absorption followed by stimulated emission can occur when two (counter-propagating) photons are utilized. The experimental verification was made in 2001,[64] and earlier in 1988 it was demonstrated[65] that an intense standing optical wave can result in momentum transfer to free electrons with scattering rates approaching that of the optical frequency. These photon-electron interactions are basic to attosecond pulse generation[66] and to studies of multiphoton harmonics and the laser-assisted surface photoelectric effect.[67] It was theoretically suggested[68] that photon-electron interactions can be exploited in EELS to provide high resolution spectra of plasmons and other features, invoking continuous-wave (CW, no time resolution) lasers.[69]

Recently, photon-induced, near-field (PIN) imaging in UEM was reported,[70] showing that only when the field is induced and probed on the ultrashort time scale would it be visualized and controlled for applications in imaging and spectroscopy. The experiments were performed on an individual and a collection of multiwalled carbon nanotubes with diameters of ~140 nm and lengths of ~7 μm, and on silver nanowires with diameters of ~100 nm and lengths ranging from 2 to 20 μm. Both fs-resolved electron energy spectra and energy-filtered photon-induced images were recorded in UEM-2. Shown in Fig. 6.56 are temporally-resolved electron energy spectra obtained from the carbon nanotubes irradiated with the intense fs laser pulse. For comparison, two separate electron spectra are shown: one obtained when the electron packet arrived before the fs pulse, $t = -2$ ps spectrum, and a

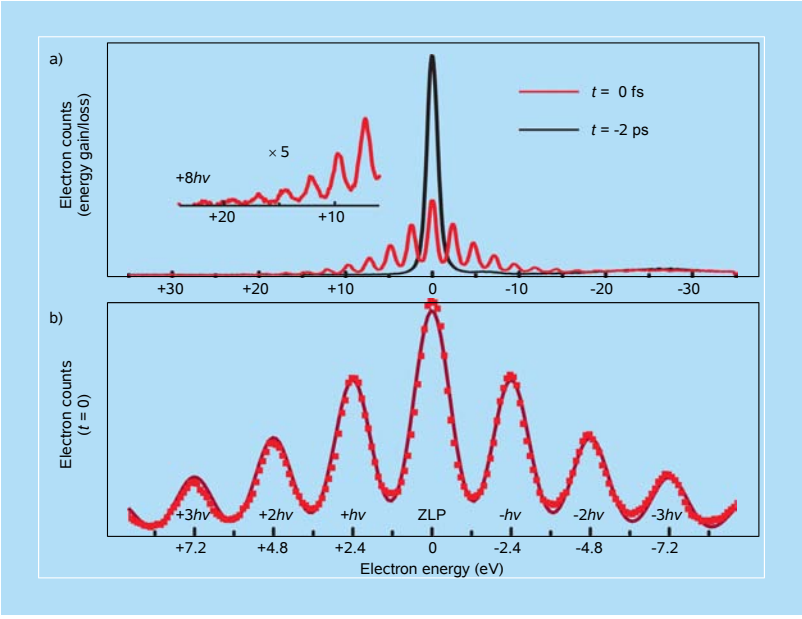


Figure 6.56 Electron energy spectra of carbon nanotubes irradiated with an intense fs laser pulse at two different delay times. Shown in panel (a) is the zero-loss peak (ZLP) of the 200 keV electrons (black) taken when the electron packet arrives before the fs pulse; in this spectrum only the plasmon peaks are present (see text). The energy spectrum at coincidence of the two pulses ($t = 0$ fs) displays the multiple quanta of photon absorption/emission. The inset depicts the positive energy gain region multiplied by 5 for the $t = 0$ spectra indicating that absorption of at least eight quanta of photon energy can be observed at maximum spatiotemporal overlap. Panel (b) gives a magnified view of the electron energy spectrum obtained at $t = 0$. The energy is given in reference to the loss/gain of photon quanta by the electrons with respect to the ZL energy. Figures 6.56–6.59 are adapted from Ref. 70.

second one recorded when the electron and photon pulses were configured for a maximum overlap, the $t = 0$ fs spectrum. Upon inspection of the change, one immediately notices that at the maximum overlap ($t = 0$) the spectrum consists of discrete peaks of decreasing intensity on both the lower- and higher-energy side of the ZLP. For the $t = -2$ ps spectrum, however, the discrete peaks are absent and only the π and $\pi + \sigma$ plasmon peaks at 6 and 25 eV, respectively, are observed.

A magnified view of the electron spectrum obtained at $t = 0$ reveals that the discrete peaks on both sides of the ZLP occur at integer multiples of the photon energy of the exciting fs pulse (2.4 eV; cf. Fig. 6.56b). From these spectra, it is apparent that the discrete peaks occur as a consequence of the interaction of the 200 keV ultrafast electron packet with the 2.4 eV fs photon pulse. On this time scale, the large influence of the PIN effect is illustrated by the substantial decrease in the ZLP intensity at maximum overlap. From the recorded electron-energy spectra, we found that electrons of the ultrafast packets can absorb more than eight photons during the brief interaction with the nanostructure (cf. inset of Fig. 6.56a).

It is important to note that the spectra shown in Fig. 6.56 (loss/gain) are observed only in the presence of the nanostructure.

On the length scale (d) of the nanostructure, relative to that of the photon (wavelength λ), the interaction between photons and electrons (i.e., free-free transitions) is greatly enhanced by the evanescent field that is created through the excitations of the carbon nanotube, and similarly for the silver nanowires, without which energy-momentum conservation is impossible.[69,71] The probability of electron-photon coupling in the presence of a third body (e.g., atom, molecule, or a surface) increases as the electron energy *increases* for a fixed laser intensity and wavelength,[67b] and such a characteristic is ideal for UEM at 200 keV. Whereas plasmonic fields induced by the fs pulse can follow the laser electric field,[72] the relaxation time for metallic nanoparticles is on the order of tens to hundreds of fs, depending on the damping (recombination) processes involved.[73] Multiwalled carbon nanotubes are metallic in nature,[74] as are silver nanowires, and for such metallic nanostructures absorption is enhanced at lower energies.[75]

In Fig. 6.57, the effect of energy filtering on imaging is displayed at different times. By using an energy filter to select only those electrons that have absorbed energy to form an image, the evanescent field generated by the fs excitation pulse became evident in real-space images of the isolated (individual) nanotube. Further, by varying the arrival time of the electron packet at the carbon nanotube relative to the clocking fs pulse, the ultrafast evolution of the evanescent field was followed in real time. As can be seen in Fig. 6.57, image contrast appears only within the local vicinity of the surface of the carbon nanotube; no energy gain occurs far from the nanostructure or within the tube itself. The energy-filtered image generated by selecting only energy-gained electrons and obtained at $t = -600$ fs shows almost no contrast. As the temporal overlap increases, however, the contrast count due to the evanescent optical field increases and reaches a maximum at $t = 0$ (i.e., maximum overlap) before decreasing again to almost zero at $t = +600$ fs. In addition to revealing the rise time of the evanescent field to be much less than 1 ps, the sequence of images in Fig. 6.57 shows that the field extends at the interface to ~ 50 nm ($1/e$) into vacuum on either side of the nanotube. The image length scale is consistent with theoretical considerations of optically-excited plasmons.[76] The apparent decay constant of image contrast in Fig. 6.57 is 130 fs, reflecting the rate of change. From the frames recorded, a movie of the time-dependent behavior of the carbon nanotubes was made.

Besides the temporal and energy domain observations, the spatial distribution of the near-field was examined from images taken for different polarizations of the fs pulse relative to the orientation of an individual nanotube (or wire). Shown in Fig. 6.58b,c are UEM images obtained at $t = 0$ with the E-field of the fs laser pulse polarized either parallel (left frame) or perpendicular (right frame) to the long-axis of the carbon nanotube. These frames display striking changes in the images:

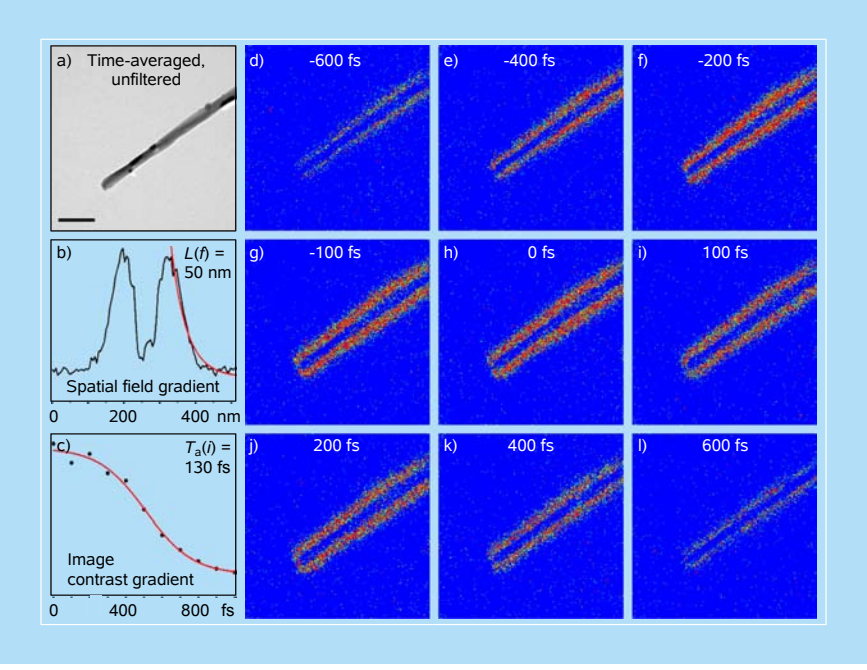


Figure 6.57 Photon-induced near-field electron microscopy of an individual nanotube. Panel (a) is a bright-field image shown for reference; the scale bar represents 500 nm. Panels (b) and (c) depict the spatial field gradient, $L(f)$, of image contrast and the rate of change, $T_a(i)$, of image contrast, respectively. To the right of the left-most panels are the energy-filtered UEM images acquired by using only the electrons that have gained energy relative to electrons with zero-loss energy. Blue areas indicate regions of the CCD where no counts were recorded because only the $+n\hbar\omega$ region was selected. The time of arrival of the electron packet at the nanotube relative to the clocking laser pulse is shown in the upper part of each panel. The electric field of the clocking laser pulse was linearly polarized perpendicular to the long-axis of the nanotube (cf. Fig. 6.58). The counts indicated in orange represent the gradient of the fields created by the fs pulse around the surface of the nanostructure and their decay with time.

when the laser polarization is positioned appropriately relative to the nanotube orientation, a spatial enhancement of the evanescent field is observed in the UEM images. This is because for this case the confinement is in the regime of $d < \lambda$. On the other hand, for the other polarization, when $d > \lambda$, the tip enhancement is seen when the polarization changes by 90°. For this case, certain spatial modes may be formed with unique distributions (cf. Fig. 6.58). The precise distribution of the field is dependent upon the nanoscale geometry of the specimen,[73,76] and the fact that the apex of the tip has a decrease in counts at perpendicular polarization is consistent with nanometric-scale field calculations.[77] When the above experiments were repeated using silver nanowires, the filtered energy gain images, the electron spectra, and the polarization dependence were obtained similarly, but with an order of magnitude lower fluence, consistent with the stronger near-field formed in the metallic nanowire.

From the above UEM results, the PIN picture can be illustrated by considering the spatiotemporal coordinates of the three bodies involved (Fig. 6.59). At negative times ($t < 0$), we can visualize the fs laser pulse as not impinging on the nanostructure, and no discrete electron energy gain or loss would be observed. When the laser pulse encounters the nanostructure ($t = 0$), it creates the near-field excitations, and this interaction causes the surface field to oscillate with the

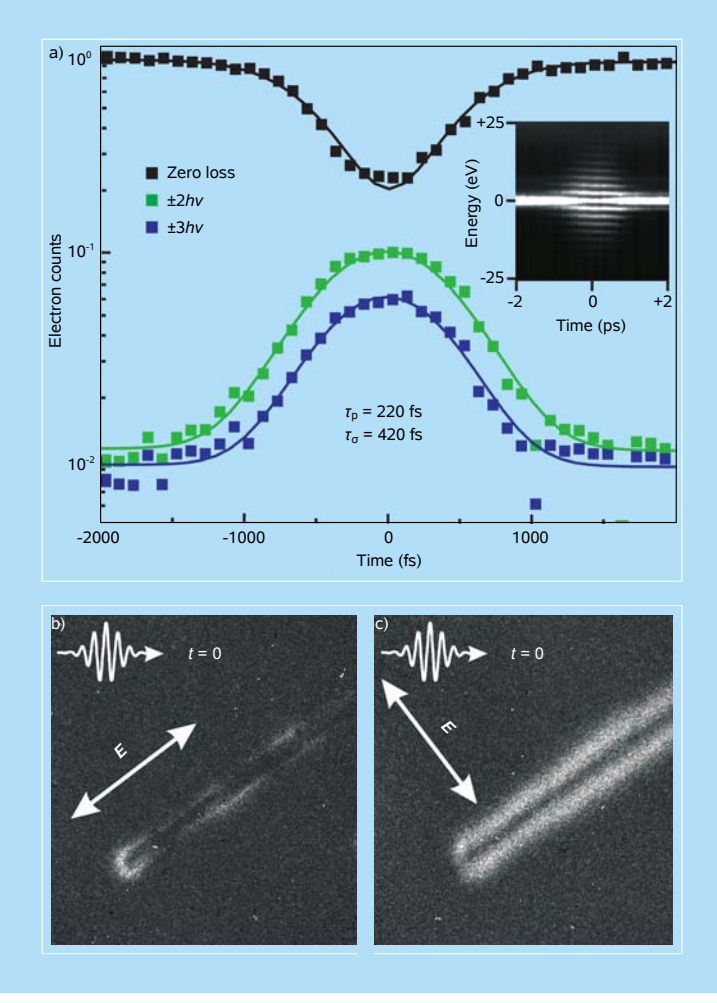


Figure 6.58 Temporal response and polarization dependence of the imaged interfacial fields. Panel (a) shows the temporal dependence of the ZLP and representative peaks of $\pm n\hbar\omega$ plotted on a log scale. Given are the values for τ_p of the fs optical pulse and τ_σ of the observed transient. The $t = 0$ position is determined by the temporal response of the ZLP. The inset is a contour plot of the data fit by the Gaussians. Lower panels display the two images taken when the *E*-field polarization of the fs laser pulse is parallel (b) and perpendicular (c) to the long axis of the nanotube. Both polarization frames were taken at $t = 0$, when the interaction between electron, photons, and the evanescent field is at a maximum.

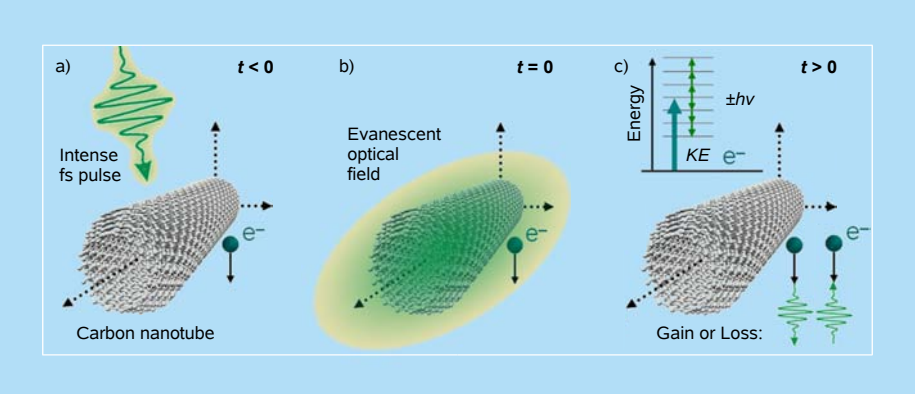


Figure 6.59 Physical depiction of the interaction between the electron, photon, and the nanostructure. Panel (a) represents a frame when the electron packet arrives at the nanotube before the fs laser pulse ($t < 0$); no spatiotemporal overlap has yet occurred. Panel (b) depicts the precise moment at $t = 0$ when the electron packet, fs laser pulse, and evanescent field are at maximum overlap at the carbon nanotube. Panel (c) illustrates the process during and immediately after the interaction ($t > 0$) when the electron gains/loses energy equal to integer multiples of fs laser photons. The inset in the right panel displays the possible final energies in the continuum due to the free-free transitions between the imaging electron and the photons in the fs laser pulse.

electric field of the laser. Because the nanotube diameter ($\sim$140 nm) is much less than the wavelength of the light (519 nm), the field is confined ($d < \lambda$) by the dimensions of the tube (wire), and this confinement sets up an oscillating dipole in the structure. The intensity of the evanescent field extends beyond the structure of the nanotube and falls off exponentially with distance from the surface. Thus, evanescent fields effectively mediate the interaction between the 200 keV electron and the 2.4 eV photons in the fs excitation pulse, but the absorption/emission processes only occur when both the electron and photon are overlapped in space at the nanostructure and in time at $t = 0$.

The orders of magnitude enhancement achieved in UEM may be appreciated when comparing with the time-averaged (CW) mode of excitation. For a tightly-focused CW laser (10^6 W/cm^2), the number of excitations on the timescale of the field is nearly five orders of magnitude less than that achieved in UEM, an irradiance of $\sim$100 GW/cm^2 being used. Further, for CW powers of about 10 W, it would be necessary for the nanostructures to dissipate the energy without significant structural damage. In UEM, typically the average power is on the order of 100 mW. Perhaps most importantly, the precise overlap of pulses in UEM allows for signal acquisition times of only a few seconds, as every electron contributes to the gain/loss signal on the timescale of field's existence. In contrast, for CW electron spectroscopy, the signal will be overwhelmed by a background whose magnitude is conditioned by the repetition rate and other factors. Finally, we note

that the process of $\pm n\hbar\omega$ absorption/emission takes place for each single-electron, timed packet.

Photon-induced near-field imaging with electrons is made possible by the precise overlap of ultrafast electron packets, intense ultrafast laser pulses, and nanostructures. These structure-mediated (electron-photon interaction) phenomena, as well as the spatiotemporal properties of the evanescent electric fields, can now be imaged in real space on the fs timescale. By knowing the distribution of the field and the control over its polarization and temporal behavior, it is possible to explore the nature of interfacial fields and their role in a variety of applications at the nanoscale of materials. In optics, near-field imaging is now known to break the diffraction limit,[76] but with limited spatial and temporal resolutions. PIN-UEM promises to take the resolutions into the domain of electron microscopy, the atomic scale. Moreover, with PIN-UEM, which exploits inelastic interactions, the real-space images (and diffraction), which are the result of elastic interactions, can easily be obtained by removing the energy filter; the scanning requirement of optical near-field methods is not of concern here. Because the method involves surface electrons, it also can bring about attosecond imaging (see Chaps. 5 and 8).

It is our hope that the above scope of applications will stimulate the reader to share our enthusiasm for the potential of UEM in various fields of study. In Chap. 8, we highlight some emerging developments and explore some aspects of the commentaries made following the birth of UEM.[27,78,79]

References

1. A. H. Zewail, *Faraday Discuss. Chem. Soc.*, **91**, 207 (1991).
2. J. C. Williamson and A. H. Zewail, *Proc. Natl. Acad. Sci. USA*, **88**, 5021 (1991).
3. J. M. Thomas, *Nature*, **351**, 694 (1991).
4. J. C. Williamson, J. Cao, H. Ihee, H. Frey and A. H. Zewail, *Nature*, **386**, 159 (1997).
5. H. Ihee, V. A. Lobastov, U. M. Gomez, B. M. Goodson, R. Srinivasan, C.-Y. Ruan and A. H. Zewail, *Science*, **291**, 458 (2001).
6. R. Srinivasan, V. A. Lobastov, C.-Y. Ruan and A. H. Zewail, *Helv. Chim. Acta*, **86**, 1763 (2003), and references therein.
7. (a) R. Srinivasan, J. S. Feenstra, S. T. Park, S. Xu and A. H. Zewail, *Science*, **307**, 558 (2005).
   (b) J. S. Baskin and A. H. Zewail, *Chem. Phys. Chem*., **7**, 1562 (2006).
   (c) J. S. Baskin and A. H. Zewail, *Chem. Phys. Chem*., **6**, 2261 (2005).
8. C.-Y. Ruan, F. Vigliotti, V. A. Lobastov, S. Chen and A. H. Zewail, *Proc. Natl. Acad. Sci. USA*, **101**, 1123 (2004).
9. J. M. Thomas, *Angew. Chem*., *Intl. Ed*., **43**, 2606 (2004).
10. (a) D.-S. Yang, N. Gedik and A. H. Zewail, *J. Phys. Chem. C*, **111**, 4889 (2007).

(b) F. Vigliotti, S. Chen, C.-Y. Ruan, V. A. Lobastov and A. H. Zewail, *Angew. Chem*., *Intl. Ed*., **43**, 2705 (2004).
11. (a) C.-Y. Ruan, V. A. Lobastov, F. Vigliotti, S. Chen and A. H. Zewail, *Science*, **304**, 80 (2004).
(b) D.-S. Yang and A. H. Zewail, *Proc. Natl. Acad. Sci. USA*, **106**, 4122 (2009).
12. (a) N. Gedik, D.-S. Yang, G. Logvenov, I. Božović and A. H. Zewail, *Science*, **316**, 425 (2007).
(b) F. Carbone, D.-S. Yang, E. Giannini and A. H. Zewail, *Proc. Natl. Acad. Sci. USA*, **105**, 20161 (2008).
(c) Y. L. Chen, W. S. Lee and Z. X. Shen, *Proc. Natl. Acad. Sci. USA*, **106**, 963 (2009).
(d) Z. Radović, N. Božović and I. Božović, *Phys. Rev. B*, **77**, 092508 (2008).
13. (a) P. Baum, D.-S. Yang and A. H. Zewail, *Science*, **318**, 788 (2007).
(b) A. Cavalleri, *Science*, **318**, 755 (2007).
14. D.-S. Yang, C. Lao and A. H. Zewail, *Science*, **321**, 1660 (2008).
15. A. H. Zewail, *Phil. Trans. R. Soc. A*, **364**, 315 (2005).
16. V. A. Lobastov, R. Srinivasan and A. H. Zewail, *Proc. Natl. Acad. Sci. USA*, **102**, 7069 (2005).
17. A. H. Zewail and V. A. Lobastov, *US Pat*. 7 154 091 (2006).
18. H. S. Park, J. S. Baskin, O.-H. Kwon and A. H. Zewail, *Nano Lett*., **7**, 2545 (2007).
19. V. A. Lobastov, J. Weissenrieder, J. Tang and A. H. Zewail, *Nano Lett*., **7**, 2552 (2007).
20. D. J. Flannigan, V. A. Lobastov and A. H. Zewail, *Angew. Chem*., *Intl. Ed*., **46**, 9206 (2007).
21. M. S. Grinolds, V. A. Lobastov, J. Weissenrieder and A. H. Zewail, *Proc. Natl. Acad. Sci. USA*, **103**, 18427 (2006).
22. O.-H. Kwon, B. Barwick, H. S. Park, J. S. Baskin and A. H. Zewail, *Proc. Natl. Acad. Sci. USA*, **105**, 8519 (2008).
23. O.-H. Kwon, B. Barwick, H. S. Park, J. S. Baskin and A. H. Zewail, *Nano Lett.*, **8**, 3557 (2008).
24. B. Barwick, H. S. Park, O.-H. Kwon, J. S. Baskin and A. H. Zewail, *Science*, **322**, 1227 (2008).
25. (a) F. Carbone, O.-H. Kwon and A. H. Zewail, *Science*, **325**, 181 (2009).
(b) F. Carbone, B. Barwick, O.-H. Kwon, H. S. Park, J. S. Baskin and A. H. Zewail, *Chem. Phys. Lett*., **468**, 107 (2009).
26. D. J. Flannigan, P. C. Samartzis, A. Yurtsever and A. H. Zewail, *Nano Lett*., **9**, 875 (2009).
27. J. M. Thomas, *Angew. Chem*., *Intl. Ed*., **44**, 5563 (2005).
28. A. H. Zewail, *Annu. Rev. Phys. Chem*., **57**, 65 (2006), and references therein.
29. D. Shorokhov and A. H. Zewail, *Phys. Chem. Chem. Phys*., **10**, 2879 (2008), and references therein.
30. A. Gahlmann, S. T. Park and A. H. Zewail, *Phys. Chem. Chem. Phys*., **10**, 2894 (2008).
31. H. S. Park, J. S. Baskin, B. Barwick, O.-H. Kwon and A. H. Zewail, *Ultramicroscopy*, **110**, 7 (2009).
32. G. Ertl, *Adv. Catal*., **45**, 1 (2000), and references therein.

33. (a) J. Tang, D.-S. Yang and A. H. Zewail, *J. Phys. Chem. C*, **111**, 8957 (2007).
    (b) See also: J. Tang, *J. Chem. Phys*., **128**, 164702 (2008).
    (c) Erratum: J. Tang, *J. Chem. Phys*., **129**, 179901 (2008).
34. J. Hohlfeld, S.-S. Wellershoff, J. Güdde, U. Conrad, V. Jähnke and E. Matthias, *Chem. Phys*., **251**, 237 (2000).
35. L. Reimer, "*Transmission Electron Microscopy: Physics of Image Formation and Microanalysis*," 2nd ed., Springer, Berlin–Heidelberg (1989).
36. F. Carbone, P. Baum, P. Rudolf and A. H. Zewail, *Phys. Rev. Lett.*, **100**, 035501 (2008).
37. (a) K. Ishioka, M. Hase, M. Kitajima, L. Wirtz, A. Rubio and H. Petek, *Phys. Rev. B* **77**, 121402 (2008).
    (b) H. Yan, D. Song, K. F. Mak, I. Chatzakis, J. Maultzsch and T. F. Heinz, *Phys. Rev. B*, **80**, 121403 (2009).
    (c) J. Kanasaki, E. Inami, K. Tanimura, H. Ohnishi and K. Nasu, *Phys. Rev. Lett.*, **102**, 087402 (2009).
    (d) R. K. Raman, Y. Murooka, C.-Y. Ruan, T. Yang, S. Berber and D. Tománek, *Phys. Rev. Lett.*, **101**, 077401 (2008).
    (e) C.-Y. Ruan, Y. Murooka, R. K. Raman and R. A. Murdick, *Nano Lett.*, **7**, 1290 (2007).
38. A. G. Marinopoulos, L. Reining, V. Olevano, A. Rubio, T. Pichler, X. Liu, M. Knupfer and J. Fink, *Phys. Rev. Lett*., **89**, 076402 (2002).
39. T. Eberlein, U. Bangert, R. R. Nair, R. Jones, M. Gass, A. L. Bleloch, K. S. Novoselov, A. Geim and P. R. Briddon, *Phys. Rev. B*, **77**, 233406 (2008).
40. W. L. Mao, H.-K. Mao, P. J. Eng, T. P. Trainor, M. Newville, C.-C. Kao, D. L. Heinz, J. Shu, Y. Meng and R. J. Hemley, *Science*, **302**, 425 (2003).
41. O. Andreussi, D. Donadio, M. Parrinello and A. H. Zewail, *Chem. Phys. Lett.*, **426**, 115 (2006).
42. (a) S. Chen, M. T. Seidel and A. H. Zewail, *Proc. Natl. Acad. Sci. USA*, **102**, 8854 (2005).
    (b) M. T. Seidel, S. Chen and A. H. Zewail, *J. Phys. Chem. C*, **111**, 4920 (2007).
    (c) S. Chen, M. T. Seidel and A. H. Zewail, *Angew. Chem., Intl. Ed*., **45**, 5154 (2006).
43. Z. Wang, J. A. Carter, A. Lagutchev, Y. K. Koh, N.-H. Seong, D. G. Cahill and D. D. Dlott, *Science*, **317**, 787 (2007).
44. H. G. Craighead, *Science*, **290**, 1532 (2000).
45. K. L. Ekinci and M. L. Roukes, *Rev. Sci. Instrum*., **76**, 061101 (2005).
46. M. Barmatz, L. R. Testardi, A. F. Garito and A. J. Heeger, *Solid State Commun*., **15**, 1299 (1974).
47. (a) Z. L. Wang, *J. Phys. Condens. Mater*., **16**, R829 (2004).
    (b) Z. L. Wang, *Mater. Today*, **10**, 20 (2007).
48. (a) M. Dantus, S. B. Kim, J. C. Williamson and A. H. Zewail, *J. Phys. Chem*., **98**, 2782 (1994).
    (b) See Ref. 4 and references therein.
    (c) See Ref. 6 and references therein.
49. D.-S. Yang, *Ph.D. Thesis*, California Institute of Technology, Pasadena, California (2009).
50. H. Park and J. M. Zuo, *Appl. Phys. Lett.*, **94**, 251103 (2009).

51. (a) J. C. H. Spence and J. M. Zuo, "*Electron Microdiffraction*," Springer, Berlin–Heidelberg (1992).
(b) J. M. Zuo, M. Kim, M. O'Keefe and J. C. H. Spence, *Nature*, **401**, 49 (1999).
52. (a) P. A. Midgley, M. Saunders, R. Vincent and J. W. Steeds, *Ultramicroscopy*, **59**, 1 (1995).
(b) P. A. Midgley and M. Saunders, *Contemp. Phys*., **37**, 441 (1996).
53. D. B. Williams and C. B. Carter, "*Transmission Electron Microscopy: A Textbook for Materials Science*," Vol. II, Plenum Press, New York (1996).
54. P. Zhang, A. A. Istratov, E. R. Weber, C. Kisielowski, H. He, C. Nelson and J. C. H. Spence, *Appl. Phys. Lett.*, **89**, 161907 (2006).
55. (a) R. F. Loane, P. R. Xu and J. Silcox, *Acta Cryst. A*, **47**, 267 (1991).
(b) D. A. Muller, B. Edwards, E. J. Kirkland and J. Silcox, *Ultramicroscopy*, **86**, 371 (2001).
56. (a) D. A. Muller, L. Fitting Kourkoutis, M. Murfitt, J. H. Song, H. Y. Hwang, J. Silcox, N. Dellby and O. L. Krivanek, *Science*, **319**, 1073 (2008).
(b) K. A. Mkhoyan, P. E. Batson, J. Cha, W. J. Schaff and J. Silcox, *Science*, **312**, 1354 (2006).
57. (a) H. E. Elsayed-Ali, T. B. Norris, M. A. Pessot and G. A. Mourou, *Phys. Rev. Lett*., **58**, 1212 (1987).
(b) M. Harb, R. Ernstorfer, C. T. Hebeisen, G. Sciaini, W. Peng, T. Dartigalongue, M. A. Eriksson, M. G. Lagally, S. G. Kruglik and R. J. D. Miller, *Phys. Rev. Lett.* **100**, 155504 (2008).
(c) S. Nie, X. Wang, H. Park, R. Clinite and J. Cao, *Phys. Rev. Lett*., **96**, 025901 (2006).
58. M. Chergui and A. H. Zewail, *Chem. Phys. Chem*., **10**, 28 (2009).
59. A. Yurtsever and A. H. Zewail, *Science*, **326**, 708 (2009).
60. T. Sjodin, H. Petek and H. L. Dai, *Phys. Rev. Lett.* **81**, 5664 (1998).
61. P. Xu, R. F. Loane and J. Silcox, *Ultramicroscopy*, **38**, 127 (1991), and references therein.
62. P. A. Midgley, E. P. W. Ward, A. B. Hungría and J. M. Thomas, *Chem. Soc. Rev.*, **36**, 1477 (2007).
63. P. L. Kapitza and P. A. M. Dirac, *Proc. Cambridge Phil. Soc*., **29**, 297 (1933).
64. D. L. Freimund, K. Aflatooni and H. Batelaan, *Nature*, **413**, 142 (2001).
65. P. H. Bucksbaum, D. W. Schumacher and M. Bashkansky, *Phys. Rev. Lett.*, **61**, 1182 (1988).
66. F. Krausz and M. Ivanov, *Rev. Mod. Phys*., **81**, 163 (2009).
67. (a) L. Miaja-Avila, C. Lei, M. Aeschlimann, J. L. Gland, M. M. Murnane, H. C. Kapteyn and G. Saathoff, *Phys. Rev. Lett.*, **97**, 113604 (2006).
(b) G. Saathoff, L. Miaja-Avila, M. Aeschlimann, M. M. Murnane and H. C. Kapteyn, *Phys. Rev. A* **77**, 022903 (2008).
68. A. Howie, in "*Electron Microscopy and Analysis 1999*," Ed. C. J. Kiely, IOP Publishing, Bristol, p. 311 (1999).
69. F. J. García de Abajo and M. Kociak, *New J. Phys*., **10**, 073035 (2008).
70. B. Barwick, D. Flannigan and A. H. Zewail, *Nature*, in press (2009).
71. R. Ishikawa, J. Bae and K. Mizuno, *J. Appl. Phys*., **89**, 4065 (2001).
72. S. Kim, J. Jin, Y.-J. Kim, I.-Y. Park, Y. Kim and S.-W. Kim, *Nature*, **453**, 757 (2008).

73. C. Burda, X. Chen, R. Narayanan and M. A. El-Sayed, *Chem. Rev.*, **105**, 1025 (2005).
74. A. Bachtold, C. Strunk, J.-P. Salvetat, J.-M. Bonard, L. Forró, T. Nussbaumer and C. Schönenberger, *Nature*, **397**, 673 (1999).
75. S. A. Maier and H. A. Atwater, *J. Appl. Phys.*, **98**, 011101 (2005).
76. S. Kawata, Y. Inouye and P. Verma, *Nat. Photon.*, **3**, 388 (2009).
77. L. Novotny, R. X. Bian and X. S. Xie, *Phys. Rev. Lett.*, **79**, 645 (1997).
78. K. D. M. Harris and J. M. Thomas, *Cryst. Growth Des.*, **5**, 2124 (2005).
79. (a) M. Quirk, *Anal. Chem.*, **77**, 336A, (2005).
(b) H. Hogan, *Biophoton. Int.*, **12**, 61 (2005).
(c) D. L. Shenkenberg, *Photon. Spectra*, **II**, 102 (2008).
(d) M. Jacoby, *Chem. Eng. News*, **87**, 9 (2009).
(e) J. M. Thomas, *Angew. Chem., Intl. Ed.*, **48**, 8824 (2009).

This page intentionally left blank

Chapter 7

The Electron Microscope and the Synchrotron

A Comparison

7.1 Introduction

Since the synchrotron is the only other sophisticated, single scientific instrument that rivals the electron microscope for structural investigation, it is prudent to outline its role in modern imaging and other related endeavors as we ponder further the power and future potential of electron microscopy. Many sharp differences and some similarities exist between the two: the synchrotron, which is a large-scale facility, and the electron microscope, which is a table-top instrument. In the words of an expert practitioner of both instruments, J. C. H. Spence,[1a] "*Electrons give much more signal, which is very difficult to quantify; X-rays quantify beautifully, but the images are rather poor.*" An additional comment, the validity of which will emerge later in this chapter, also expressed by Spence, is the striking similarity between soft X-rays (i.e., those of energy less than 2 keV) and electrons. They can each be: very intense, highly coherent, exhibit strong interaction with matter as well as giving rise to multiple scattering. In the case of synchrotron-generated X-rays, image resolution is limited by wavelength, there being a value of 1 nm at 1.2 keV. But if one selects harder X-rays (so as to decrease their wavelength), the intensity falls off dramatically, and there is the added problem that the zone plates (see later) required to focus X-rays of high energy are either not available or not good enough for imaging purposes in X-ray microscopy. No X-ray spectroscopy or X-ray microscopy can compare for spatial resolution with STEM-EELS of the kind discussed in Chaps. 3, 4 and 6.

We now proceed to illustrate the effectiveness of synchrotron radiation as a means of retrieving structure-imaging information in a range of different circumstances by first recalling the intrinsic properties of both X-rays and electrons as probing radiation (see Table 7.1). This table was compiled in 2004: since then many improvements in performance have been made, some of which are detailed in notes relating to Table 7.1.

Table 7.1 Comparison of Some Key Characteristics Exhibited by Synchrotron X-ray Radiation and High-Energy Electrons (compiled in 2004; courtesy of A. I. Kirkland and J. C. H. Spence)

Radiation	Source Brightness (particles/cm^2/eV/Sr)	Elastic Mean-Free Path (nm)	Absorption Length (nm)	Minimum Probe Size (nm)
X-rays[1,2]	10^{26}	10^3	10^5	10^2
Electrons[3]	10^{29}	10^1	10^2	10^{-1}

[1]The Linac Coherent Light Source (LCLS) at Stanford promises to deliver (by September 2009) laser-like pulses of X-rays of about 100 fs duration and 10^{12} photons per pulse. At the FLASH facility in Hamburg it is envisaged that pulses of 50 fs (or smaller) duration will become available.
[2]Focusing devices for X-rays, whether by zone plates or of the Kirkpatrick–Baez-mirror type, are rapidly improving. Soon to be commissioned are 10 nm X-ray probes at about 5 keV.
[3]Higher-order aberration-corrected STEM microscopes will have a brightness that surpasses the value achieved in 2004.

Reverting to the definitions of 4D electron microscopy in our preface, where we talked of real space, reciprocal space, spectroscopy (electron-energy-loss) and the time domain, we shall examine here some of the strengths and achievements of synchrotron radiation in retrieving the kind of structural and dynamic information that we have already demonstrated electron microscopy can achieve. The sequence of topics that we shall discuss involve:

- imaging in real space,
- imaging via reciprocal space,
- X-ray absorption spectroscopy, and
- time-resolved measurements.

In dealing with all these topics using the intense beams of X-rays that synchrotron sources provide, we also recall the advantages of utilizing such sources (especially for *in situ* investigations). These are[1b]:

- high fluxes (of both nonfocused and nanofocused) monochromatic, as well as high fluxes of polychromatic, X-rays;
- high penetrating power of (high-energy) synchrotron radiation so that thick specimens and those in contact with liquids may be investigated;[2]
- high degree of energy resolution (*ca.* 0.1 eV);
- ease of effecting X-ray microscopic tomography of both small and large specimens, coupled with the advantage of being able to determine the chemical composition (of, e.g., catalysts)[3,4] from X-ray absorption peaks through the tunability of the wavelength;
- the in-principle possibility of combining the recently developed technique of coherent (X-ray) diffraction imaging, CDI (or diffraction imaging microscopy, DIM — these two acronyms signify the same

principle) using nanofocused illumination[3,5,6] (the term DIM is fast disappearing, and CDI is nowadays often expanded as CXDI);

- high brightness of the infrared region of synchrotron radiation also allows considerable scope for *in situ* microspectroscopic studies of reactions taking place in solids — see, for example, the work of Stavitski *et al.*[7] on the oligomerization of styrene derivatives at single-site Brønsted acid active centers [necessarily, the spatial resolution in such work is not high (2 to 10 μm) and incapable, therefore, of affording mechanistic insights];
- the feasibility of combining wide-angle, small-angle and ultrasmall-angle X-ray scattering (WAXS, SAXS and USAXS, respectively) for the imaging of structures — see, for example, the study by de Moor *et al.*[8]

The principal disadvantage of X-rays as primary probes, however, is their large beam area, which is several orders of magnitude more than that of very sharply focused (*ca.* < 1.5 Å dia.) high-energy electrons. It follows that spatial characterization with X-rays cannot compete with that attainable using STEM or convergent-beam electron diffraction (CBED) (see Chap. 6). But, like positron emission computerized tomography (PET)[9] or magnetic resonance imaging (MRI),[10] it is well-suited for the *in situ* investigation of systems operating in reactions typical of those used industrially.

In the preceding chapters, the advantages of the electron microscopy approach have been frequently highlighted. It is, however, instructive to recite the comparisons between (soft) X-ray sources and field-emission electron sources made by Spence and Howells,[11] who, *inter alia*, draw attention to the differing interactions of electrons and photons with matter. In particular, these workers assess the Advanced Light Source synchrotron facility at Berkeley, USA. They also deal with upper limits on source brightness, as well as relationships among brightness, coherence parameters and degeneracy, and the vexed question of radiation damage inflicted by both X-rays and electrons.

7.2 Transmission X-ray Microscopy and X-ray Microscopic Tomography

Although much effort has been involved over several decades in the use of powerful laboratory sources of X-rays (such as rotating anodes) and in the construction of Fresnel zone-plate lenses, where a circular diffraction grating has alternating transparent and opaque concentric rings,[12] it is the availability of synchrotron X-ray sources, where high fluxes of controllable monochromatic rays may be readily generated (through the use of dual crystal monochromators), that has greatly transformed the prospects of utilizing transmission X-ray microscopy (TXM) for

a range of 3D imaging. As in conventional medical computed tomography (CT) practice, X-ray tomography on the nanoscale entails computing (using proven mathematical procedures) from a series of projected 2D images, taken over as a wide range of angular orientations as possible, the 3D image of the object under investigation.

In the soft X-ray region (100 eV to 2 keV) a zone-plate-based TXM[13,14] has achieved a spatial resolution of 15 nm.[15] In the hard X-ray region, because of the difficulty in fabricating the requisite zone-plate lenses, such spatial resolution is not yet attainable, and values of around 60 nm are more typical.[15,16] Several other parallel imaging measurements may be carried out using TXM. A useful recent summary exists[16b] of near-edge X-ray absorption microscopy, as well as micro-X-ray absorption fine structures — both of the extended (EXAFS) and near-edge (XANES) varieties. (Note that NEXAFS, i.e., near-edge X-ray absorption fine structure, is synonymous with XANES, but the two terms arose under different circumstances.)

The chemical sensitivity of TXM tomography (TXMT), which comes about because it is readily possible to tune the X-ray energy to be just above or just below a prominent absorption edge of a particular element, is a great advantage especially in the examination of biological materials and hydrated samples. Thus, with soft X-ray TXMT, one may operate in such a manner that oxygen (and hence water), where the K-edge energy of the atom is 542 eV, is transparent whereas carbon (K-edge, 284 eV) is opaque if the X-ray energy in the tomography is, say *ca.* 500 eV.[16] Recently, detailed 3D maps of *ca.* 60 nm spatial resolution of several cryogenically fixed, whole fully-hydrated and unstained biological samples, including individual yeast and bacterium cells have been produced (Fig. 7.1; see Ref. 2 and references therein).

In the hard X-ray region (typically 3–11 keV) nanotomographic examination encounters rather more experimental difficulties since the achievable magnification varies with X-ray energy. Benefiting from this range of X-ray energies, however, one may readily access the absorption energies of a wide range of elements that are utilized in the semiconductor industry, *viz.* Cu, Zn, Ga, Ge, As, Ta, W, Au, Hg, Pb, etc.[15] Using the X-ray energy range given above, Yin *et al.*[15] achieved 60 nm spatial resolution in two dimensions; and they obtained 3D images (plus chemical sensitivity) of tungsten plugs such as those used as electrical interconnects between layers in semiconductor integrated circuits (Fig. 7.2).

In scanning transmission X-ray microscopy (STXM), a fine X-ray probe is passed across the specimen. The resulting "image" can represent any of the wide range of signals — in addition to density information, for example, it is possible to measure local diffraction, fluorescence, or absorption. Figure 7.3 illustrates the chemical sensitivity possible in a tomographic reconstruction based on STXM. Schroer *et al.*[4] combined XANES with scanning microtomography in the hard

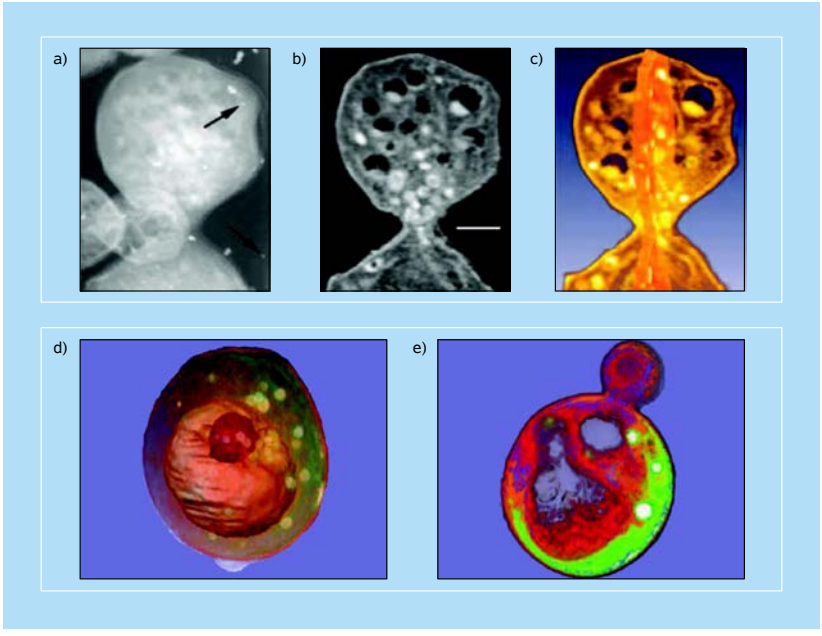


Figure 7.1 Tomography of whole yeast cells using TXM. (a) Single projection image of a rapidly frozen budding yeast. The data for the complete 3D reconstruction were composed of 45 such images, collected through a total of 180° of rotation. (b) Computer generated section through a tomographic reconstruction of the raw data shown in (a). The scale bar is 0.5 μm. (c) Cropped volume rendered view of the reconstructed data from (b). (d) An edge enhancement gradient algorithm was used to volume segment the 3D data used to produce this volume rendered image, showing the nucleus (purple), vacuole (pink) and lipid droplets (white). (e) Color-coded reconstruction of a budding yeast. Lipid droplets, which are the densest structures, were color coded white, the least dense vacuoles were color coded gray, and numerous other subcellular structures or intermediate densities were colored shades of green, orange and red. (See Ref. 2 and references therein.)

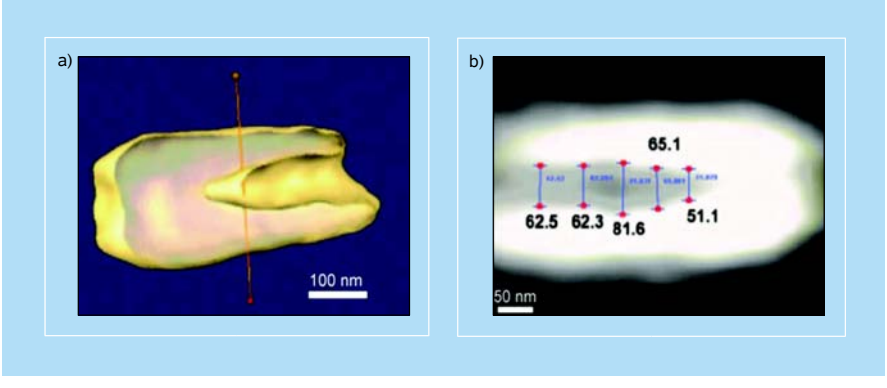


Figure 7.2 3D imaging of a tungsten plug with a keyhole. (a) The side view of a reconstructed tungsten plug displayed using an isosurface, which is a 3D-surface contour drawn at equal intensity (brightness) values. (b) The measurement of the size of the keyhole. The diameters of the keyhole range from 50 to 80 nm.[15]

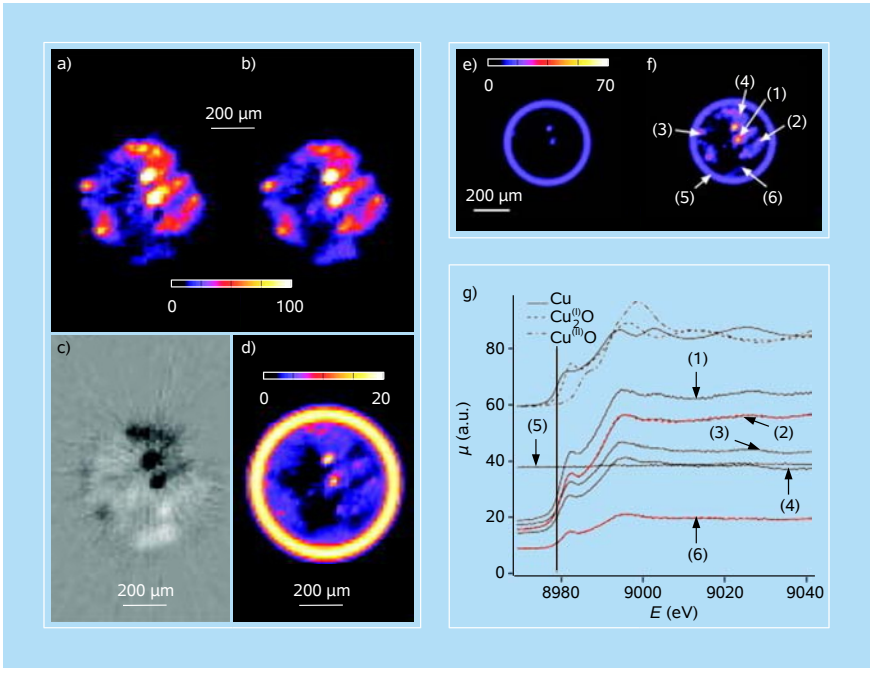


Figure 7.3 Catalyst-filled capillary imaging. (a–d): Relative concentrations of (a) metallic and (b) monovalent Cu inside a glass capillary, on the same scale. (c) Difference image between (b) and (a). The shade of gray outside the specimen corresponds to zero difference. Darker areas depict a higher concentration of metallic Cu, while brighter regions signify a higher concentration of $Cu(I)_2O$. (d) Attenuation generated by elements other than Cu. (e–g): Reconstructed tomogram of a capillary filled with a CuO/ZnO catalyst (e) below (E = 8970 eV) and (f) above the Cu K-edge (E = 8995 eV). (g) Reconstructed XANES spectra at different locations marked in (f) and (offset) reference spectra of metallic Cu, $Cu(I)_2O$, and Cu(II)O. The dashed red curves in (g) represent two fits using the reference spectra. Color bars represent the attenuation coefficient μ in arbitrary units.[4]

X-ray range. The sample was scanned in translation and rotation through an intensive pencil beam generated at the European Synchrotron Radiation Facility (ESRF) in Grenoble, France. After each full translational scan, during which a full absorption spectrum was recorded at each position, the sample was rotated through 3.6° until 101 projection data sets had been recorded. After filtered back projection,[17] a full XANES spectrum is available at each location on any virtual slice through the sample, allowing chemical states to be mapped throughout the sample volume.

In a recent *in situ* study of a simulated Fischer–Tropsch catalyst (consisting of minute particles of Fe_2O_3 dispersed over SiO_2) heated up to *ca.* 350°C, a Fresnel-type zone plate was used to focus monochromatic X-rays down to a spot size of 35–40 nm diameter. De Smit *et al.*[18] employed the STXM technique to track the formation of the intermediate solid-state phases involved in the catalytic conversion of CO and H_2. By scanning through different X-ray energies (at Berkeley, USA) and traversing the X-ray absorption edges of known elements in the catalysts, they

mapped absorption across the solid, thereby arriving at an "image" of the spatial distribution of each element. They found that the minute particles of Fe_2O_3 were converted at *ca.* 350°C (and a gas pressure of 1 bar) into metallic Fe, some iron carbide and also Fe_3O_4, as well as a trace of Fe_2SiO_4. The spatial distribution (150 Å) is not capable of yielding an atomic-scale view of the structure of the working catalyst. Considerable advances are required in this technique for that to be accomplished.

Further mechanisms of image contrast remain to be fully exploited in TXM. For example, phase-contrast imaging in addition to (or in place of) absorption may be used.[16] Notwithstanding the difficulties still encountered in zone-plate fabrication, much may be expected from the sophisticated high-resolution X-ray tomographic equipment now being developed at the Advanced Light Source at Berkeley in the USA[16] and similar facilities (capable of cryo-fixed biotomography) in Japan, Germany and also in Taiwan (see Ref. 15, Ref. 2, and references therein).

7.2.1 X-ray tomography of biological cells

X-ray tomography has features that combine those associated with both optical and electron microscopy. It is fast and relatively easy to accomplish (like optical microscopy) and it produces relatively high-resolution, absorption-based images (like electron microscopy). One of its supreme advantages is that it can examine whole, hydrated cells (at better than 35 nm resolution, currently) and reveal details of sub-cellular structures. In the energy range of the photons used, typically between the K-shell absorption edges of carbon (284 eV) and oxygen (542 eV), organic material absorbs approximately ten times as much as does water, thus producing a quantifiable natural contrast and dispensing with the need for contrast enhancement procedures (such as staining).

Computed tomography, which involves reconstructing projections of an object viewed from different directions[19] (see Fig. 3.13), derives from the mathematical principles first described by Radon.[20] A projection of an object at a given angle in real space is a central section through the Fourier transform of that object, the so-called "central slice" theorem.[21] The method of back projection, used in X-ray tomography (as in electron tomography), is based on inverting the set of recorded images, projecting each image back into an object space at the angle at which the original image was recorded. Using a sufficient number of back projections, from different angles, the superposition of all the back projected "rays" will construct the original object.[21]

The recent work of Larabell and Le Gros[2,22,23] using tomography (with a setup as depicted in Fig. 7.4, where a Fresnel zone-plate lens is used to focus the monochromatic X-rays from a synchrotron), has yielded detailed images of whole yeast cells (*Saccharomyces cerevisiae*) such as those shown in Fig. 7.5. (For whole-

cell imaging, the destructive effects of radiation damage must be attenuated by cryo-cooling, as first demonstrated by Glaeser.[24]) These workers have also shown how X-ray tomography has uncovered complex structural details of whole bacterial cells (*Escherichia coli*).[2] Other workers using soft X-rays have recently achieved a spatial resolution of 15 nm.[15] In the hard X-ray region, because of the difficulties mentioned earlier (Sec. 7.2), spatial resolutions are inferior (around 60 nm at best).[15,16] It is thought that, in the near future, work at places such as the US National Center for X-ray Tomography (NCXT), will contribute greatly to cell biology because the achievable spatial resolution should be extended to between 10 and 15 nm; and the development and deployment of multiple wavelength (multicolor) X-ray probes will enable the simultaneous localization of multiple proteins to be undertaken.[2]

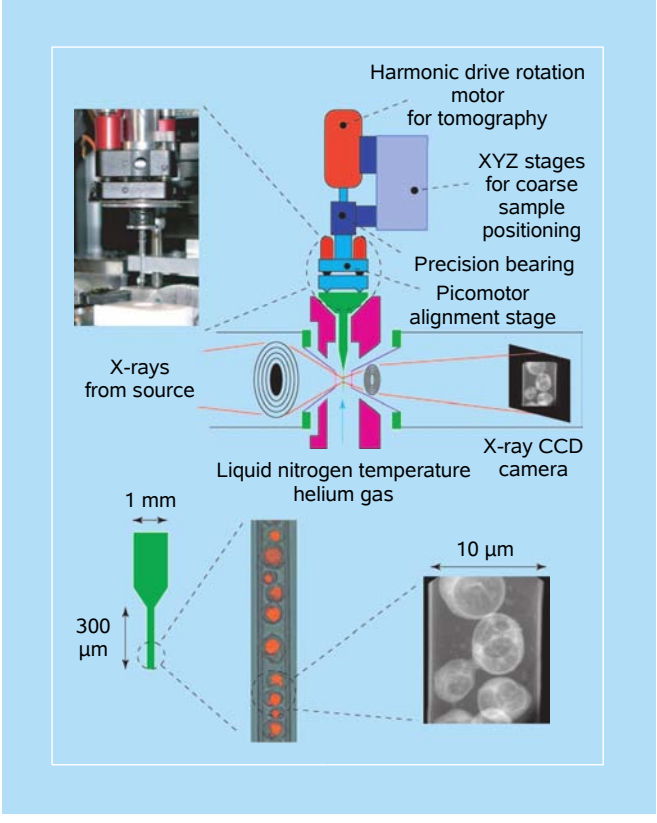


Figure 7.4 Principles used at the National Center for X-ray Tomography (NCXT), Berkeley, USA, for the 3D imaging of cells.[2,22,23]

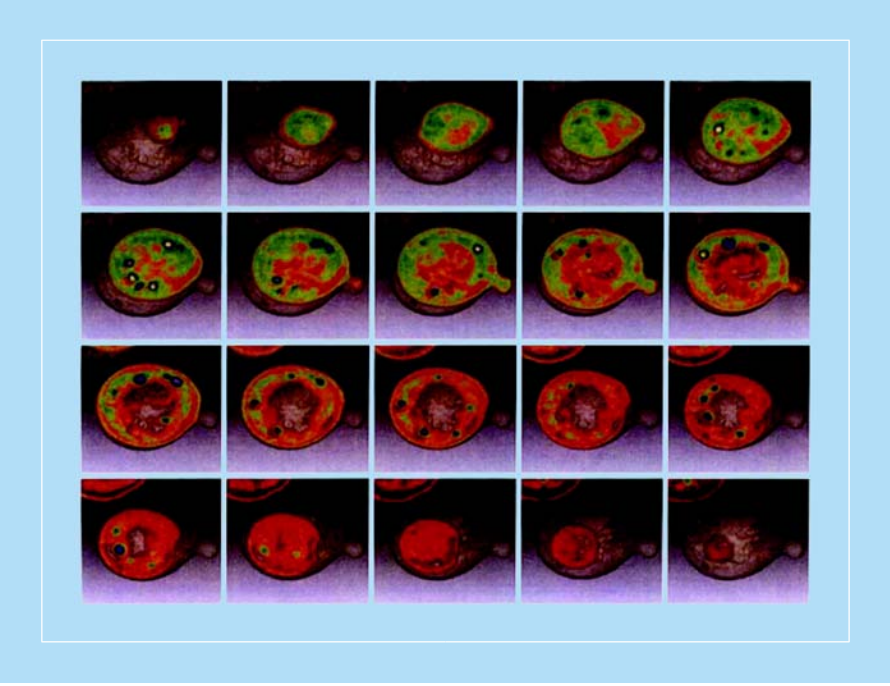


Figure 7.5 Low-resolution X-ray tomograms of early budding yeast cells (see Ref. 2 and references therein): computer-generated sections through the tomographic reconstruction. Structures have been assigned different colors, which indicate degree of X-ray adsorption. Dense lipid droplets appear white and other cell structures are colored shades blue, green, and orange with decreasing density. The yeast cell is 5 μm in diameter.[22]

7.3 Coherent X-ray Diffraction Imaging

Although CXDI (also known as DIM; see Sec. 7.1) does not use lenses of any kind, it is nevertheless described next as it is now designated as microscopy, made possible by the availability of highly coherent radiation.

With the significant improvements in attaining coherence of synchrotron radiation that has been achieved recently,[25] the feasibility of embarking upon CXDI, already demonstrated in principle,[5,26,27] has become particularly attractive. Coherent X-ray imaging, which uses in-phase X-rays, offers a "lensless" method of imaging in place of ordinary X-ray microscopy (that requires Fresnel zone plates). This approach was first suggested by Sayre[26a] and first carried out by Miao *et al.*[5,26b] at the Brookhaven Synchrotron in 1999.

The CXDI method is, in principle, an optimal method for X-ray imaging, as there is no intensity loss of high spatial frequency information arising from a lens contrast transfer function, and the theoretical resolution limit is close to the wavelength used to create the diffraction pattern. However, limitations concerning

sample configuration — isolated particles are required — and long data acquisition times (many hours) render the technique one of rather low throughput at present as compared with what may be achieved by X-ray microscopy based on zone-plate imaging. As with the latter, cryogenic sample stages must be used to mitigate the deleterious effects of radiation damage inflicted during data collection.

CXDI uses a computer instead of a lens to form an image. The phase problem is solved numerically by iterative methods for the continuous distribution of scattering which results from a noncrystalline sample. Although the phase problem can be solved by several methods for crystalline samples, its solution for an isolated, 3D nonperiodic sample has become possible only recently. The solution developed from work by Saxton and Gerchberg[27] in electron microscopy, and later from an important paper by Fienup,[28b] which forms the basis of most modern algorithms. This is a rapidly expanding field.[28b,28c,29]

The procedure using CXDI entails three steps[28]:

- a tilt series of diffraction patterns is generated using coherent X-rays, providing the amplitude of the diffracted wave field;
- the phases of the wave field are obtained using variants of phase retrieval algorithms[30] developed in other branches of optics; and
- the series of tilt images are recorded by means of Fourier transformation of each individual (fixed angle) data set.

In a significant report by Chapman, Spence, Marchesini *et al.*,[29] the authors showed how to accomplish *ab initio* 3D X-ray diffraction microscopy using coherent X-rays. They achieved a relatively high resolution in the three dimensions: 10 nm in x and y directions, and 50 nm in z. Recently, it was indicated (see Ref. 6 and references therein) that coherent X-ray diffraction imaging is an aspect of X-ray microscopy with the potential of reaching spatial resolutions well beyond the diffraction limits of X-ray microscopes based on optics. This report describes the attainment of a resolution of about 50 Å that is achieved in an exposure time of 600 s (with a small gold particle <1000 Å and illuminated with a hard X-ray nanobeam of primary energy 15.25 keV, $\lambda = 0.813$ Å, and beam dimensions of *ca.* 1000×1000 Å^2).

As outlined above, from the (coherent) diffraction pattern, the wave field behind the object is reconstructed by iteratively solving the phase problem.[28–30] 3D imaging is possible by recording a (tomographic) series of diffraction patterns (see Refs. 30, 31 and also 26b and references therein). Coherent illumination of the object is crucial to this technique, and the coherent dose on the sample determines the spatial resolution. As the coherent flux at modern synchrotron radiation sources is rather limited, CXDI experiments require long exposure time, and the spatial resolutions obtained so far have been similar to those of direct X-ray microscopic imaging techniques, lying in the range of a few 100 Å.

What Schroer *et al.* have done in their recent CXDI experiment with nanofocused illumination, from which the small Au particle under investigation was reconstructed with 50 Å resolution (see Ref. 6 and references therein), is to demonstrate proof of principle. As a result of the nanofocusing, the coherent flux on the sample was sufficiently increased, thereby reducing considerably the exposure time at high spatial resolution. This work paves the way to combining spatial resolution well beyond that of each X-ray technique taken by itself[32,33] and is of vital importance to single-particle diffraction experiments with future free-electron laser sources.[34]

Very recently,[35] Song *et al.*, for the first time, recorded and reconstructed X-ray diffraction patterns from single unstained viruses that have a molecular mass about three orders of magnitude smaller than those specimens investigated by the DIM approach previously. Using the setup shown in Fig. 7.6, these workers separated the diffraction pattern of the virus particles from that of their surroundings (Fig. 7.7), thereby producing quantitative and high-contrast imaging of a single virion with a resolution of 22 nm. In this way the structure of the viral capsid inside the virion was identified. The actual samples studied were murine herpesvirus-68 (MHV-68) virions.

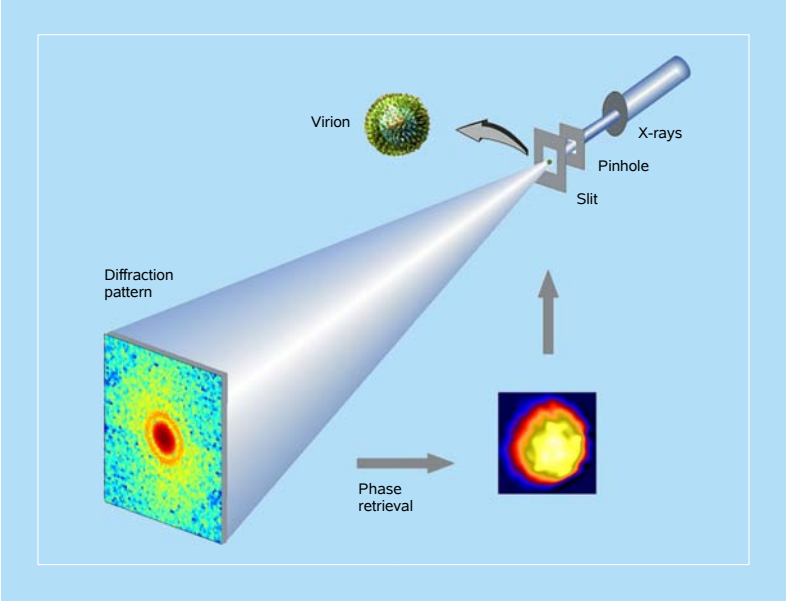


Figure 7.6 Schematic layout of the X-ray diffraction microscope. A 20 μm pinhole was used to define the incident X-ray beam. The virion specimen was positioned at a distance of 1 m from the pinhole. A silicon guard slit with beveled edges was used to eliminate the parasitic scattering from the pinhole. The oversampled diffraction pattern, recorded on a liquid nitrogen-cooled CCD camera, was directly inverted to a high-contrast image using an iterative algorithm.[35]

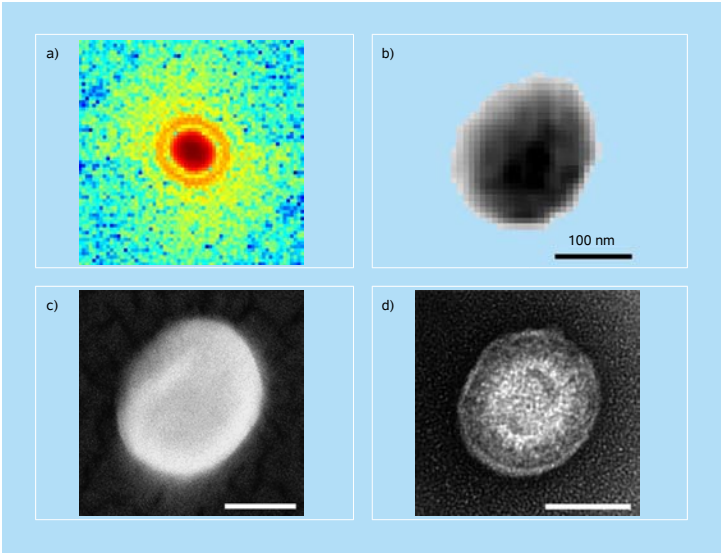


Figure 7.7 X-ray diffractive imaging of single herpesvirus virions. (a) X-ray diffraction pattern obtained from a single unstained virion. (b) High-contrast image reconstructed from (a), where the background and the surroundings of the virion were completely removed. (c) SEM image of the same virion. (d) Negative stain TEM image of a similar herpesvirus virion.[35]

The experiment was carried out on an undulator beam line at SPring-8 in Japan. Unfocused, monochromatic X-rays with an energy of 5 keV were filtered by a 20 μm diameter pinhole (Fig. 7.6), placed *ca.* 1 m upstream from the sample to block the parasitic scattering from the upstream optical components. To obtain the diffraction pattern only from single virions, Song *et al.* measured[35] two sets of diffraction intensities with the specimens in and out of the X-ray illumination, and then subtracted the two diffraction patterns. Figure 7.7a shows the diffraction pattern of a single, unstained virion which was added up from three independent diffraction patterns, each having a radiation dose of *ca.* 3.5×10^7 J kg^{-1}. [The phase retrieval of the diffraction pattern was carried out by the so-called guided hybrid-input-output algorithm (GHIO).[36]]

With more brilliant synchrotron radiation sources, and especially after the advent of X-ray free electron lasers,[11] much higher resolutions may be achievable. With these prospects in mind, it is not idle to speculate that coherent X-rays and coherent electron waves will ultimately be used to image living cells, organelles, some viruses and many important macromolecules that are difficult or impossible to crystallize, and which thereby render them inaccessible to structural crystallography.

Whilst the achievement of Song *et al.*[35] is commendable with respect to characterizing virus structures at relatively low resolution using synchrotron

radiation, it must not be forgotten that the very first elucidation of virus structure by electron microscopy, in the hands of Crowther (see Chap. 3) 38 years ago, led to high-resolution imaging of single virus particles with the icosahedral structure clearly determined. And we should also not forget the present state of the art of cryo-electron microscopy (see Chaps. 3 and 4, Figs. 3.16 and 3.17, and Figs. 4.22 and 4.23 especially) where whole-cell imaging in vitreous ice is reported at 2 nm resolution, far beyond the current capability of CXDI imaging using a synchrotron.

7.4 Extraction of Structures from Powdered Specimens

It is nowadays almost routinely possible, provided the material is phase pure (in the crystallographic sense), to deduce the 3D structure of a solid in atomic detail from X-ray powder diffractograms.[37] This may be done via reciprocal space[37a] or via real space.[37b] Because of the nature of synchrotron radiation, higher resolution is achieved in the diffractograms, so that it becomes easier to unravel overlapping peaks (that are not usually resolved in diffractograms generated using ordinary laboratory X-ray sources). Incommensurate structures, and those that yield weak diffraction peaks with conventional equipment, are also readily identified by use of synchrotron radiation. An instructive example of how modern X-ray powder diffractometry (employing the direct-space genetic algorithm technique for structure solution followed by Rietveld refinement)[37] is seen in the work of Guo *et al.*,[38] who charted the spontaneous induction of chirality in the preparation of Werner's complex *cis*-$[CoBrNH_3(en)_2]Br_2$. A similar approach was recently invoked to obtain structures from gas-phase electron diffraction.[39]

Ever since the first *ab initio* structure determination was carried out[40a] from powder synchrotron X-ray data (on α-$CrPO_4$), many unique new insights of a structural nature have been achieved by taking advantage of the tunability of the wavelength. It is, for example, possible to identify ions of differing valence states in mixtures of orthorhombic Eu_3O_4 and cubic Eu_2O_3. This is achieved by using wavelengths below and above an X-ray absorption edge, so that anomalous scattering is observed. This is the basis of Attfield's work known as valence-specific diffraction.[40b]

Recently, a new algorithm has been developed, based on the so-called "charge-flipping" method (introduced by Oszlányi and Sütő)[40c] that has been modified for application to powder diffraction.[40d] By working with the complex scattered amplitudes within the iterative loop (rather than with intensities), these workers have solved new oxide structures and found significant advantages over existing methods. As with the "charge-flipping" method for single crystals, no prior knowledge of composition or space group is needed, and this *ab initio* phasing method for powders is well suited to poor quality data, prior to subsequent refinement.

7.4.1 Extraction of structures from ultramicrocrystalline specimens

When crystals are so small that they cannot be satisfactorily probed with conventional four-circle diffractometers (for the standard crystallographic study of single crystals), they are amenable, using synchrotron sources, to X-ray powder diffractometry, as demonstrated by Sankar *et al.*[41] with the templated nanoporous catalyst known as DAF-5 (Davy Faraday number 5). Their diffraction technique employed the synchrotron radiation source at Daresbury, UK. A wavelength of 1.39994 Å was used, and the Rietveld powder refinement method[37a] yielded the precise location of the incarcerated template molecule (4-piperidinopiperidine) inside the CoALPO (DAF-5) catalyst. The structure was solved by direct methods.[42]

7.4.2 Energy-dispersive X-ray diffraction

The rudiments of energy-dispersive X-ray diffraction (EDXRD) (see Fig. 7.8) are given in the following expressions: $\lambda = hc/E$, $n\lambda = 2d\sin\Theta$, and $E/n = hc/(2d\sin\Theta)$. Several characteristics thus define the approach:

- The detector angle 2Θ is kept fixed at typically 1.5°; calibrated with a radioactive source and a silicon standard; the *wavelength is varied* as opposed to conventional XRD where wavelength is fixed and 2Θ varied.
- *White radiation* permits simultaneous measurement over a wide range of energies (*d*-spacing); the measurement range is dependent on the detector angle.
- The high intensity of X-rays allows the beam to pass through the stainless steel *in situ* hydrothermal cell, PTFE liner, and sample.
- The *time-resolved technique* typically used 1–2 min per scan.
- Poor resolution and intensity variation contribute towards the EDXRD data being unsuitable for Rietveld refinement since structure factors are currently unobtainable.

EDXRD may be readily applied to study the appearance and growth of new crystalline phases from a nutrient solution or gel.[43] One convenient arrangement, which uses three separate detectors set at different values of 2Θ so as to cover a wide range of *d*-spacings, is shown in Fig. 7.8a. The time-resolved (minute-scale) diffractograms in Fig. 7.8b show the relatively sudden transition (at *ca.* 100 min) from an amorphous precursor gel to a crystalline zeolite (type A). Such data allow the detailed kinetics of crystallization to be quantitatively evaluated on this time scale. A typical setup for such measurements, emphasizing the ease of coping with high-flux, high-energy synchrotron radiation, is shown in Fig. 7.9.[44]

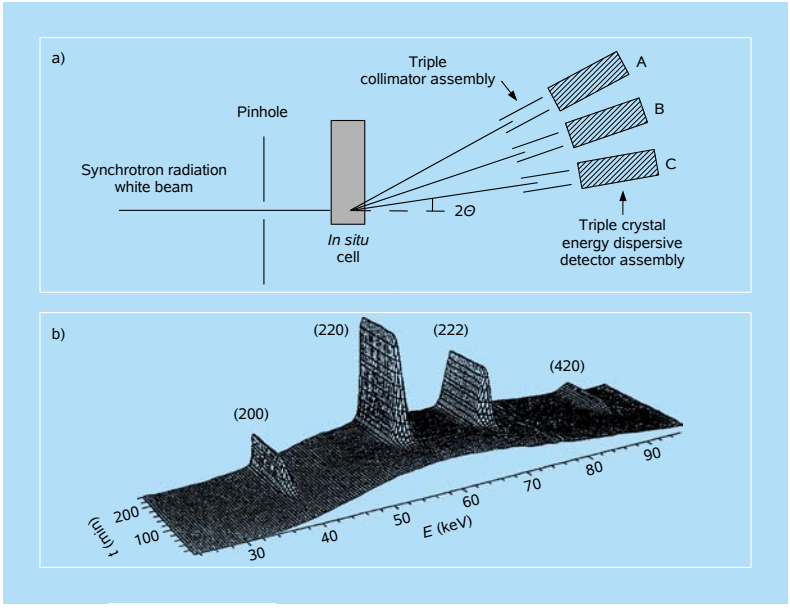


Figure 7.8 Energy-dispersive X-ray diffraction. (a) Setup used for energy-dispersive X-ray diffraction (EDXRD) in *in situ* measurements of the kinetics of crystallization of microporous materials. A bank of three detectors, each at fixed values of 2Θ, ensures that a wide range of *d*-spacings is explored. (b) Typical stacked, time-resolved EDXRD plots (time in minutes versus energy in keV) for the synthesis of zeolite A at 367 K.[43]

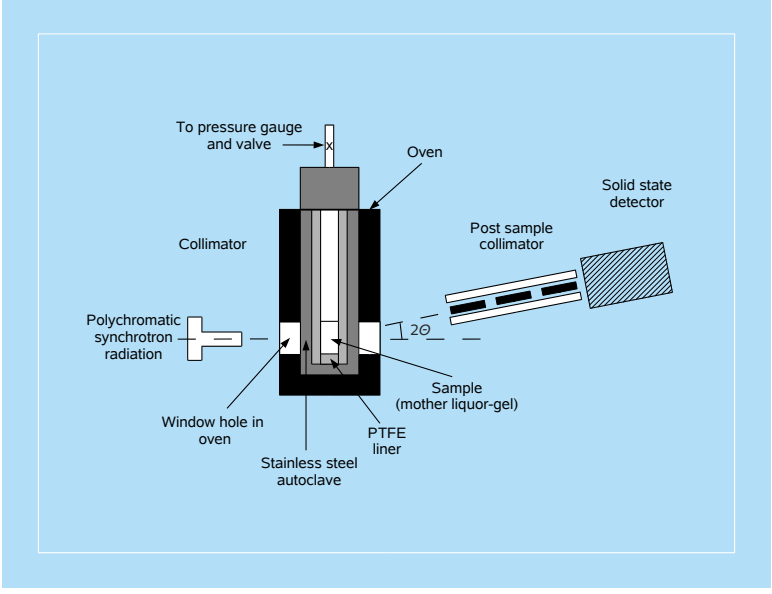


Figure 7.9 Schematic diagram of the experimental setup for EDXRD used to follow nucleation and formation of a catalyst inside a stainless steel autoclave. The single-element (Canberra) detector is inclined at a fixed 2Θ angle of 1.4566 ± 0.0001° and calibrated using a silicon standard.[44]

7.4.3 X-ray absorption fine structure spectroscopy

For the retrieval of structural changes that occur during the course of various transformations [e.g., a solid behaving as a catalyst in contact with fluid reactants or a solution (or gel) about to nucleate a new solid phase], both XANES and EXAFS are invaluable.[43,45,46] Whereas XANES and EXAFS yield interatomic distances, as well as coordination numbers and Debye–Waller factors, of nearest-neighbor atoms, XANES in addition yields information on the valence states of the absorbing atom. To illustrate how significant (and relatively easy to measure) the electronic changes that accompany alterations in valence are, we show the situation pertaining to the element vanadium in Figs. 7.10a,b.[46,47] Similar shifts in EXAFS spectral features accompany most transition metal ions, notably those of cobalt and titanium, two elements that figure eminently in powdered heterogeneous catalysts.

Figure 7.11a shows the way that the Ti K-edge spectra reveal how the coordination and environment of the central Ti^{iv} ion becomes tetrahedral (from the shape of the corresponding XANES fingerprint) on being bonded to the siliceous nanoporous solid known as MCM-41.[48] It is not only the location and nature of

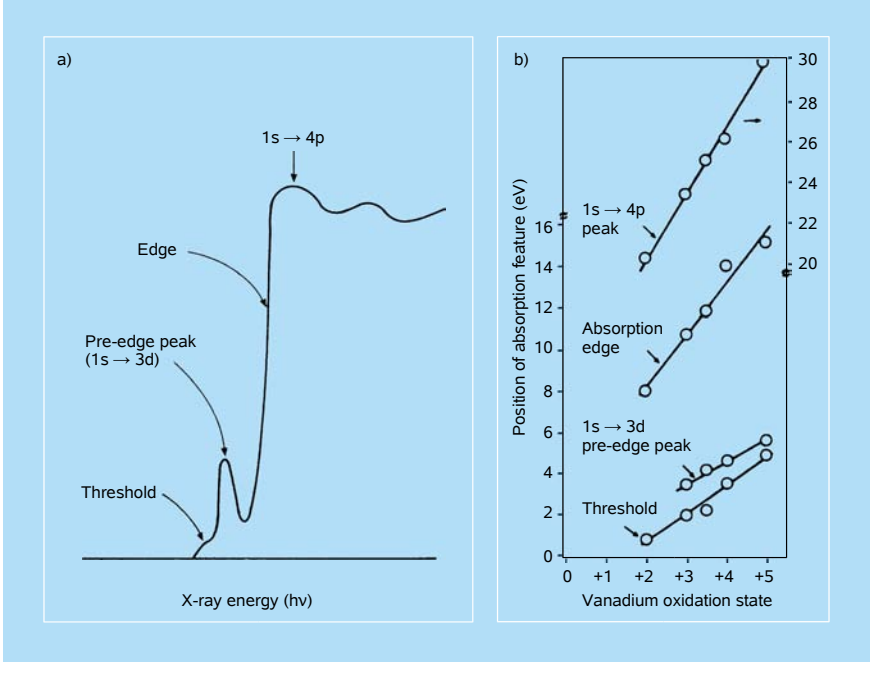


Figure 7.10 X-ray edge absorption spectra. (a) Representation of the X-ray K-edge absorption spectra (XANES) obtained from solids containing first-row transition metals such as vanadium, chromium or titanium.[46] (b) A strong correlation exists between the oxidation state of vanadium and the position of its X-ray absorption edges.[47]

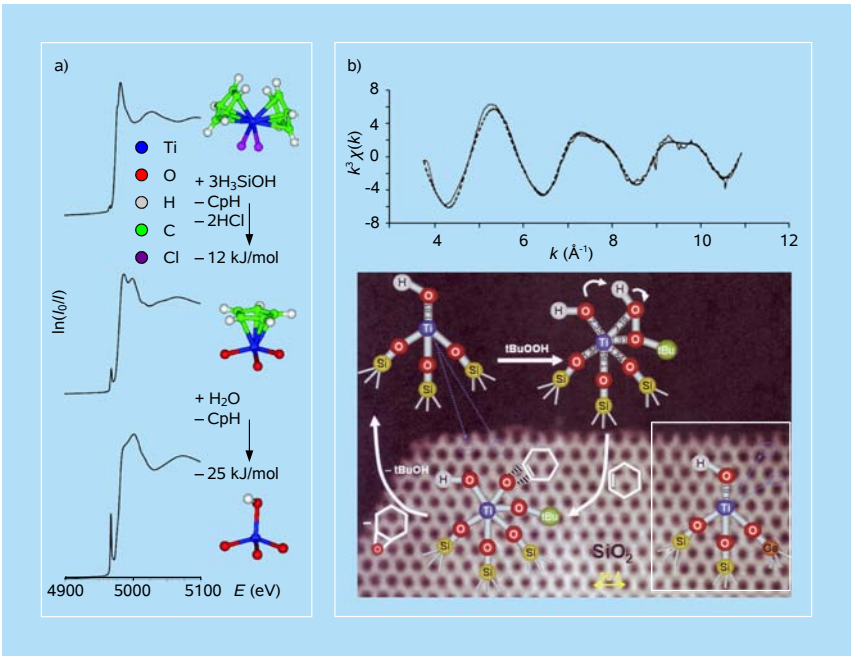


Figure 7.11 XANES and EXAFS of Ti-based catalysts. (a) The titanium K-edge XANES spectra show how the multiply-bonded Ti^{iv} ion (in titanocene dichloride, $Ti(cp)_2Cl_2$) changes on being anchored to silica via Ti–O–Si bonds, into a four-coordinated, tetrahedrally bonded state (bottom right).[48] (b) *In situ* EXAFS studies reveal the structure of the empty active site (a titanol group tripodally attached to mesoporous silica, center-left) for the epoxidation of cyclohexene. The steady-state structure is six-fold coordinated (center) from which the mechanism may be deduced. It is of the Eley–Rideal kind; only one of the two reactants (the *t*BuOOH) is adsorbed. The "free" cyclohexene plucks an oxygen from the hydro-peroxide chemisorbed at the active site. When $HOTi–(OSi)_3$ active centers are replaced by $HOTi–(OSi)_2(OGe)$ (bottom right) a superior catalyst results. Inset (top): Correspondence between experimentally determined EXAFS (full line) and calculated (dashed line) for the steady state structural situation, with six-fold coordination (center).[1b]

a catalytically active center in a solid catalyst that is readily obtained from EXAFS studies, the mechanism of a heterogeneous catalytic reaction may also be deduced from *in situ* measurements of the X-ray absorption spectra under steady-state conditions. When, for example, the tetrahedrally bonded Ti^{iv} active center $[HOTi–(OSi)_3]$ shown at the center left of Fig. 7.11b is involved in the catalytic conversion of cyclohexene by *t*BuOOH (tert-butyl hydroperoxide), the steady-state structure, as deduced from the EXAFS spectrum, is as shown in the six-coordinated state at the center of Fig. 7.11b. It has the hydroperoxide bound to the Ti^{vi} active center. The cyclohexene, in a quasi-free state, upon reaching the surface, plucks away one of the bound oxygens (according to the classic Eley–Rideal mechanism)[1,46] thereby yielding the cyclohexyl epoxide and restoring the active center to its "free" tetrahedral state. (All this occurs when a few of the Si^{iv} tetrahedral centers in the mesoporous silica are replaced by Ti^{iv}.)

The precise location, environment and bond distances associated with a framework-substituted Co^{II} ion in the parent nanoporous catalyst known as CoALPO-18 and in its active state involving the Co^{III} ion (for the oxyfunctionalization of alkanes in air) are shown in Figs. 7.12, along with relevant FTIR and EXAFS spectra.[49]

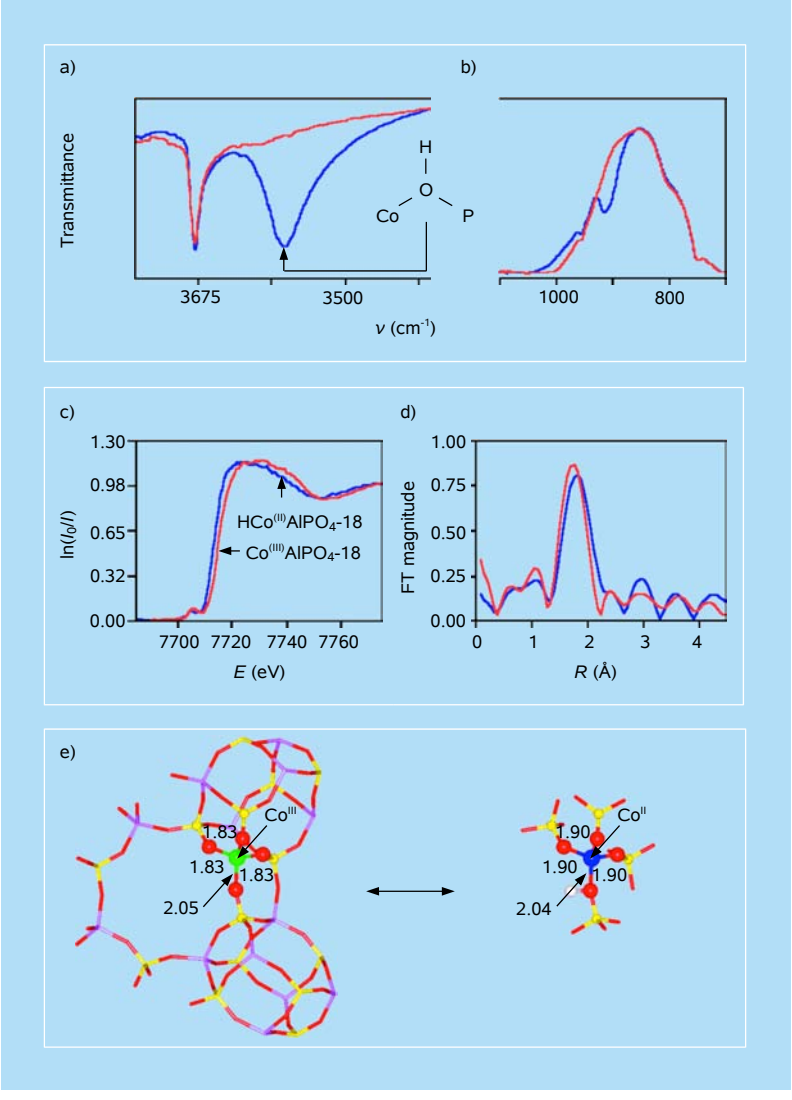


Figure 7.12 Proof (from FTIR and EXAFS) that when the microporous solid Co^{II}AlPO-18 (blue curves) is converted, by calcination in oxygen, to Co^{III}AlPO-18 (red curves), the proton is lost from the framework (a,b), and there is a shift to increasing energy in the Co X-ray K-edge (c), and a decrease in the Co–O bond length (d) signifies a change of the oxidation state from Co^{II} to Co^{III}. The structural details are summarized in (e).[49]

Whereas XANES and EXAFS information using synchrotron radiation is necessarily retrieved from rather large areas of a sample (owing to the large area of the X-ray beam), the complementary techniques of ELNES (electron-loss near-edge structrure) of EXELFS (extended electron-loss fine structure) that are readily recorded on the sub-nanometer scale electron microscopically yield the same kind of information — bond distances, coordination numbers and Debye–Waller factors — but in much faster time (see Fig. 3.18). Spence[1a] has shown that ELNES and EXELFS can greatly elucidate multi-component glasses on an ultrasmall sample.

7.4.4 Combined X-ray absorption and X-ray diffraction for *in situ* studies of powdered catalysts

Combined techniques always tend to yield more than the individual ones separately or than their simple addition, and this is also the case for the combination of X-ray absorption and diffraction of powdered crystals.[50–52] For uniform powdered catalysts the experimental setup is as shown in Fig. 7.13.[43] By bringing these two techniques together one may chart, in parallel, both the short-range order (of a minority constituent of the solid, for example, the active sites in a solid catalyst) and the long-range order of the host solid in which it is embedded. This has been a

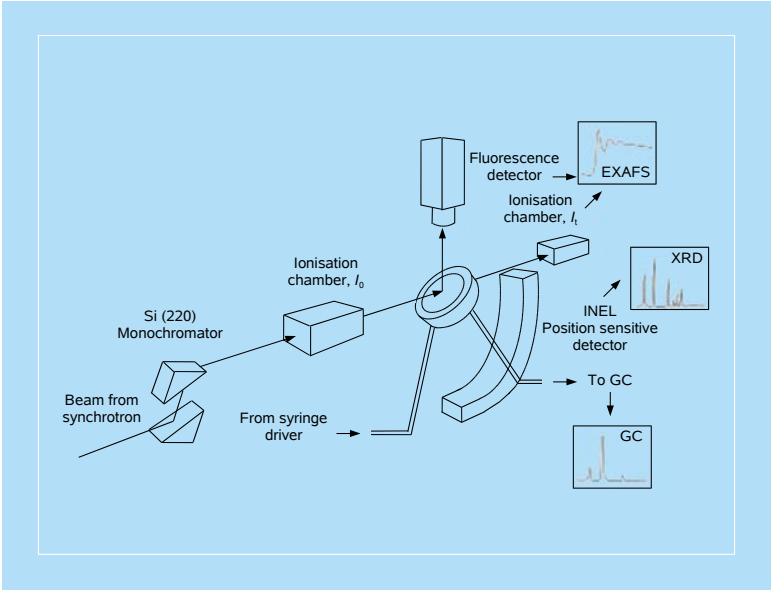


Figure 7.13 Schematic representation of the setup used for the *in situ* parallel recording of X-ray absorption (XRA) spectra and X-ray diffraction (XRD) of solid catalysts in contact with either liquid of gaseous reactants. The fluorescence detector makes it possible to probe absorption edges of elements present in ppm levels in the solid catalyst.[43,46]

considerable step forward in the *in situ* characterization of heterogeneous catalysts in the presence of high pressures of gases and of liquid reactants.[53]

7.5 Studies of Species in Solution

Although XANES and EXAFS measurements provide less precise information on the architecture of atomic structure than X-ray diffraction does, when applied to solutions they avoid many problems associated with background scattering from the solvent.

Here, we begin with one example that may illustrate the structural features that can be obtained with electrons and X-rays[54] for systems randomly oriented in the gas and solution phases. The specific reaction studied (Fig. 6.11), the halogen elimination in dihaloethanes, involves the breakage of two bonds from the reactant ($C_2F_4I_2$; 1,2-diiodotetrafluoroethane) to give the product (C_2F_4; tetrafluoroethylene). The structures of all intermediates were unknown and ultrafast electron diffraction (UED) and spectroscopy were invoked to determine these structures.[55] The temporal evolution of the two steps of the reaction revealed for the final step a rise of 25(7) ps, and for the intermediate a decay of 26(7) ps; the first step occurs in 200 fs. As noted in Fig. 6.11, when the diffraction frame is referenced before and after the first C–I bond breakage, the peak of the I···I bond is either present or absent, respectively. For the molecular structure of the C_2F_4I intermediate, both the bridged and classical C_2F_4I structures were considered in the analysis of the diffraction data. The theoretical curves for the classical structures reproduce the experimental data very well, whereas the fit provided by the theoretical bridged structure is vastly inferior. Thus, the structure of the C_2F_4I radical intermediate is, in fact, classical in nature — the iodine atom does not bridge the two carbons (Fig. 6.11). This case is the first example of resolving such complex structures in isolated reactions during the transition to final products, and for this phase only electrons can be used to map out the transformation.

Recently, the same class of reactions was studied in liquid methanol by Ihee and co-workers, using X-ray scattering (XRS).[56] The transient structure of CF_2CF_2I was also found to be classical, but the secondary dissociation slows down by a factor of six in methanol because of the solvent influence. In this case, the solvent could lead to faster energy redistribution, change in reaction barrier height, and/or caging of fragments. With a spatial resolution of *ca.* 0.01 Å and *ca.* 100 ps time resolution, Ihee and co-workers proposed a similar mechanism to that of the isolated $C_2F_4I_2$ molecule, and suggested the following mechanism for $C_2H_4I_2$ in solution: the loss of one I atom from $C_2H_4I_2$ leads to the C–I–C triangular geometry of CH_2ICH_2 (see Fig. 7.14). This transient C_2H_4I then binds to an I atom to form a new species, the C_2H_4I–I isomer, which eventually decays into C_2H_4 and I_2.

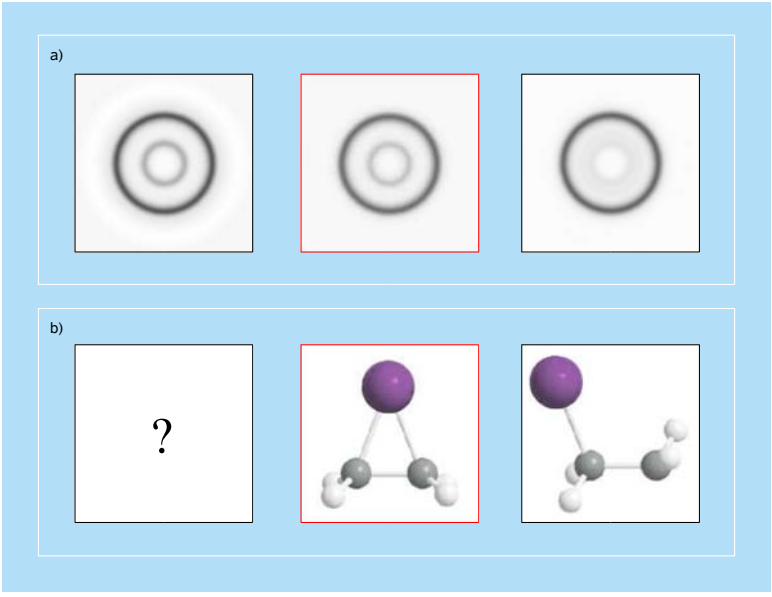


Figure 7.14 Time-resolved X-ray scattering can reveal changes in the pair-distribution function (histogram of interatomic distances) during the course of a chemical reaction in solution. Snapshot of $C_2H_4I_2$ in methanol taken 100 ps after photolysis (upper-left). The calculated pair distribution function for the bridged $C_2H_4I_2$ radical (lower-middle) is a better match than that for the anti radical (lower-right). After Ref. 60b.

Earlier, such experiments were carried out on small molecules in solution, and since then a number of studies have been made.[57–62] The complexity of the analysis of X-ray scattering in solution stems from the fact that the scattering signal contains contributions from all the species present in the probed volume, and in particular from a distribution of excited state configurations, from hot and cold ground-state molecules, and from the solvent heated by the photochemical reaction. On the other hand, X-ray absorption spectroscopy (XAS) offers the advantage of chemical selectivity. One of the early examples that illustrates the approach can be found in Ref. 63 and in Fig. 7.15.[64]

The state of the art in XAS is illustrated by the cogent examples from the work of Chergui and his group.[65–67] They determined the structural change upon light-induced spin crossover (singlet-to-quintet conversion) in iron(II)-based molecular complexes (Fig. 7.16). Iron-tris-bipyridine ($[Fe^{II}(bpy)_3]^{2+}$) is a typical representative of this family. It has low-lying metal-centred (MC) states that provide a channel for nonradiative relaxation, such that upon excitation of the singlet metal-to-ligand-charge-transfer (^{1}MLCT) state in the visible-UV, population reaches the lowest excited high-spin metal-centred (MC) ^{5}T state in <1 ps (Fig. 7.16a; see also Refs. 68 and 69). The latter then relaxes to the ground state nonradiatively in ~ 650 ps. In the quintet state, it is known that the Fe–N

bonds elongate, but the structure was not determined for $[Fe^{II}(bpy)_3]^{2+}$ due to its short lifetime. Using 70 ps long X-ray pulses, Gawelda *et al.* captured the structure of the quintet state after laser excitation.[66] Figure 7.16b shows the Fe K-edge XANES of the molecule in the ground state and in the quintet state at 50 ps time delay, as retrieved from the measurements of the difference spectra shown in Fig. 7.16b (lower panel). From the structural analysis, the Fe–N bond elongation ΔR_{Fe-N} is 0.20(5) Å (Fig. 7.16c).[66] Similar picosecond XAS studies were also carried out on $[Fe^{II}(tren(py)_3)]^{2+}$, which has a longer-lived (~60 ns) quintet state.[69] More recently,[65] the Chergui group, in a tour de force investigation with a small number of photons per pulse, extended their XAS studies into the femtosecond time domain, probing the system evolution from the initially excited (^{1}MLCT) state to the quintet electronic state (Fig. 7.16d), showing that the quintet state is populated in ~150 fs and resolving the long-standing issue of ultrafast spin change in such systems. Other illustrative examples on longer time scales come from the work of Chen *et al.*, and details can be found in Refs. 70–72.

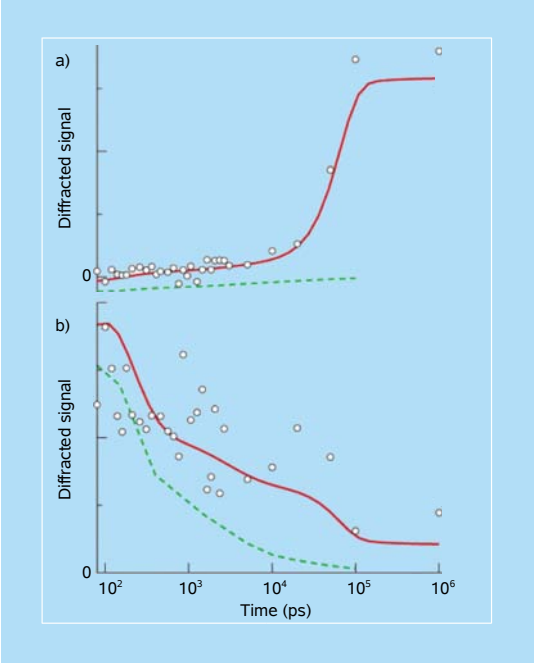


Figure 7.15 Dissociation of I_2 molecules monitored using X-ray pulses applied at varying time intervals after the laser excitation. The diffracted signal is shown for two different spatial scales: (a) 0.6 nm and (b) 0.15 nm. The rise of the data in (a) is attributed to the solvent taking up the energy released by the iodine molecules as they relax to their ground state, and expanding. The behavior is quite different in (b): on this scale, both the relaxation of the iodine molecules and the recombination of dissociated atoms contribute to the signal. In both cases, the red curve indicates the predicted signal, taking all the relevant effects into account; the green curve represents the contribution from iodine atoms that do not escape their solvent cage.[64]

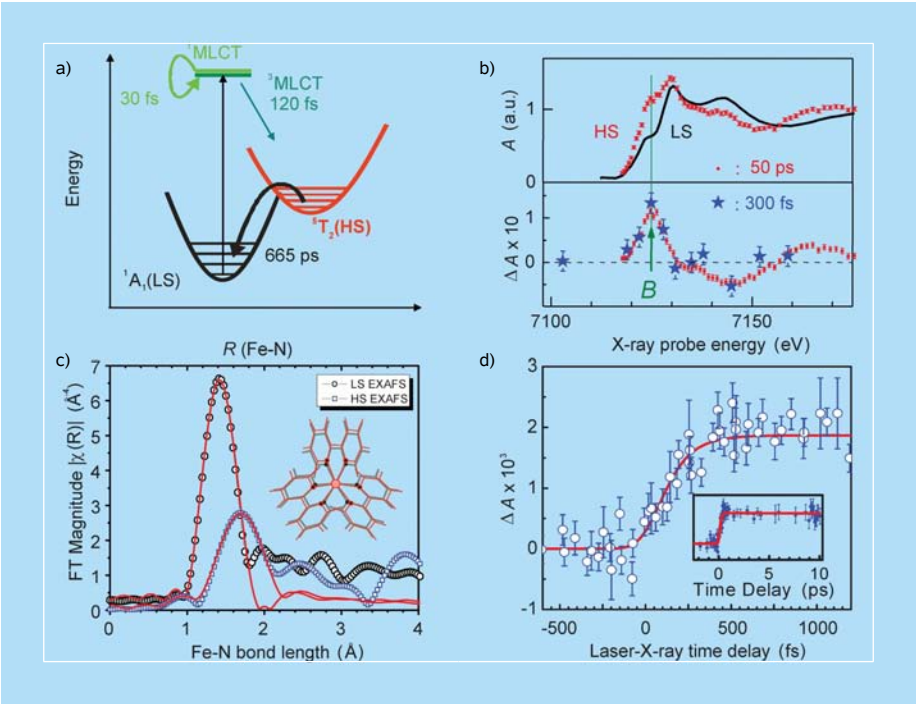


Figure 7.16 The iron complex structure ($[Fe^{II}(bpy)_3]^{2+}$) and its photocycle.[54] (a) Schematic potential energy curves of $[Fe^{II}(bpy)_3]^{2+}$ and its photocycle as determined by laser optical experiments.[66] (b) XANES spectra of the ground, low-spin (LS) state and excited, high-spin (HS) state of aqueous $[Fe^{II}(bpy)_3]^{2+}$. The HS spectrum is generated from the ground and the transient spectrum recorded at a time delay of 50 ps after laser excitation (lower panel). The stars (lower panel) represent an energy scan obtained at 300 fs time delay.[65] (c) Radial distribution functions of the LS and HS states obtained from a structural analysis of the EXAFS region. The inset shows the change of structure of the molecule. (d) Femtosecond optical pump-X-ray probe at the XANES feature labeled B in panel (b). The inset shows a time scan up to a time delay of 10 ps.[65]

7.6 Laue Crystallography: Static and Dynamic

Crystal structures were almost invariably solved by the static crystal-multiwavelength X-ray source devised by von Laue in the first decade or so after the discovery of X-ray diffraction; then the method gave way to the more convenient Bragg method using monochromatic X-rays. The synchrotron has, however, led to a renaissance of the Laue method, partly because of the development of new, more intense and brilliant synchrotron sources and partly because of vast improvements in rapid readout CCD detectors. As a consequence, synchrotron radiation protein crystallography has expanded enormously as a field of activity in the last two decades.[73–78] The Laue method permits much shorter exposure times than monochromatic techniques by three to four orders of magnitude. It is

therefore ideally suited for time-resolved studies of structural changes in biological macromolecules whilst they undergo their chemical conversions. But it also yields good quality structural images of proteins, enzymes and a host of other biologically significant materials to a resolution of better than 1.7 Å.

Examples of prototypical studies can be found in Refs. 73–79, which shed considerable light on the transient structures of the macromolecules in question (e.g., concanavalin-A glucoside and mannoside complexes, hydroxymethylbilane synthase, and photoactive yellow protein, PYP, are but a few). To illustrate the power of being able to visualize a "protein as it functions," there are few better macromolecules to describe than myoglobin.[76–78] The time scale of such experiments is sub-ns or longer.

Schotte *et al.*[76] reported a time-resolved X-ray diffraction (from 0.1 ns to 3 μs) of the myoglobin (Mb) mutant in which the Leu29 amino acid is replaced by Phe (otherwise known as L29F-mutant). The frame-by-frame structural evaluation, resolved to 1.8 Å, together with time-resolved mid-IR spectroscopy, allows one to, in effect, "watch" the protein as it executes the transformation. The time-resolved mid-IR spectroscopy of the flash-photolyzed adduct L29F MbCO revealed a short-lived carbon monoxide intermediate, the 140 ps lifetime of which is shorter than that found in wild-type Mb by a factor of 1000. The electron density maps of the protein unveil transient conformational changes far more dramatic than the structural differences between the carboxy and deoxy states and depict the correlated side-chain motion responsible for rapidly sweeping CO away from its primary docking site. One of the 155 time-resolved Laue diffraction patterns recorded from a crystal of the mutant L29F MbCO is shown in Fig. 7.17. With ~150 ps X-ray and 2.2 ns optical pulses, the relaxation of the protein from 3.2 ns to 3 ms after photodissociation was studied by Bourgeois *et al.*,[77] using Laue crystallography.

Another impressive example of Laue crystallography is seen in the work of Ihee *et al.*,[79a] who succeeded in extracting short-lived intermediate structures of PYP, from one nanosecond to one second. In so doing they elucidated signal transduction in the blue light photoreceptor from *Halorhodospira halophila* (see also Refs. 79b, 79c). By analyzing a comprehensive set of Laue data during the PYP photocycle (47 time points), they tracked PYP during its photocycle, an overview of which is shown in Fig. 7.18, and observed how absorption of a blue light photon (by its p-coumaric acid chromophore) triggers a reversible photocycle. The structural changes at the chromophore in the early, red-shifted intermediates are transduced to the exterior of the protein in the late, blue-shifted intermediates through an initial so-called "volume-conserving" isomerization of the chromophore and the progressive disruption of the hydrogen bonds between the chromophore and its surrounding binding pocket.

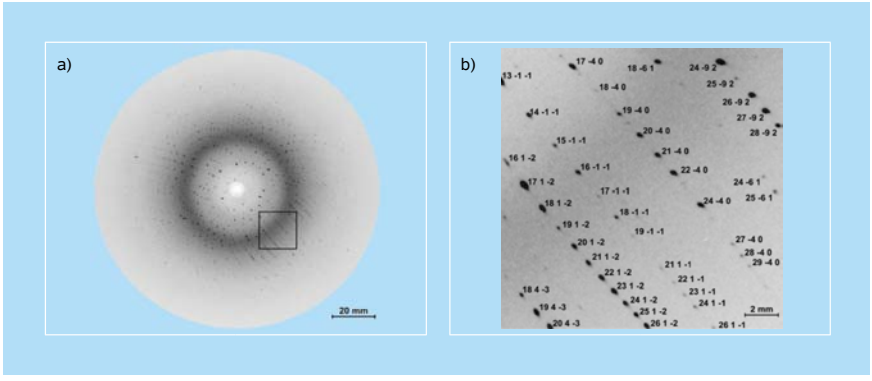


Figure 7.17 One of 155 time-resolved Laue diffraction patterns recorded from a single crystal of L29F MbCo. The image shown in panel (a) contains about 3000 diffraction spots each of which corresponds to a Bragg reflection from a set of crystallographic planes. Many spots are simultaneously excited. Only a small fraction of these spots would appear when probed with monochromatic X-ray radiation. The enlarged view (b) identifies Miller indices associated with the crystallographic planes that generated the corresponding Bragg reflections. This 10 s exposure, acquired midway through the series, represents the cumulative diffracted intensity from 32 X-ray pulses with the pump probe delay set to 100 ps.[76]

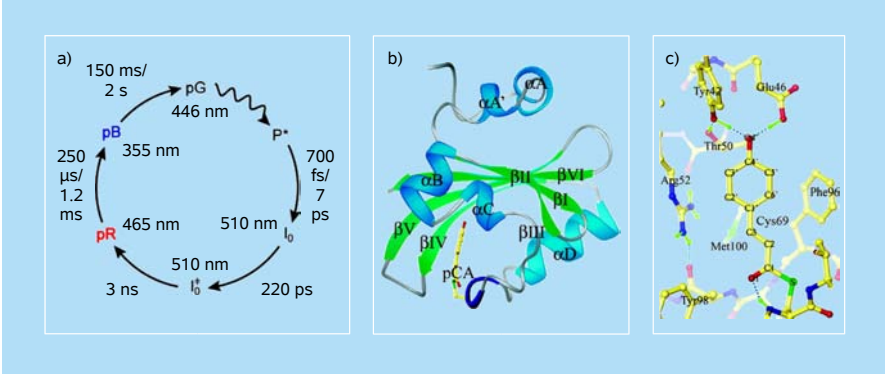


Figure 7.18 Overview of photoactive yellow protein (PYP). (a) The room-temperature PYP photocycle. (b) Structure of PYP. The pCA chromophore is shown in yellow; secondary structure is labeled using the notation of Rubinstenn *et al.*[79b] (c) Structure of the chromophore binding pocket of PYP. Hydrogen bonds are shown as green dotted lines and chromophore atoms are labeled according to Borgstahl *et al.*[79c] Adapted from Ref. 79a.

7.7 The Perennial Problem of Radiation Damage

The key factor governing the structural (and hence imaging) information retrievable both by electron microscopy and synchrotron radiation, especially of biological macromolecules, is radiation damage. Whereas the degree of structure disruption is almost infinitesimal in the case of metals, semiconductors, most oxides, and certain

other so-called "radiation-hard" inorganic materials, biological macromolecules are extremely vulnerable to radiation damage.[3,80] The (desirable) elastic scattering of electrons by the atoms making up a biological structure is accompanied by a multiplicity of inelastic interactions — of which we take advantage in EELS — which deposit energy in the covalent bonds that hold the macromolecules together. Each C–C single bond in a protein, for example, upon rupture, suffers a change of bond distance that increases from 1.5 to 3.4 Å (the latter being the sum of the van der Waals radii of the atoms) with concomitant generation of free radicals which, in turn, set off a cascade of other reactions that produce a variety of compounds very different from the original specimen under investigation. These radiolytic products, which are very similar whether the incident radiation is an electron or an X-ray, include small, mobile fragments as well as a few with additional cross-links. The smaller fragments can be restrained from moving from their initial location by cooling the specimen to low temperatures, as we have described in earlier chapters.

Ultimately, the structural resolution achievable via electrons or X-rays is governed by the crucial question of damage. With the impending inauguration of the Linac Coherent Light Source (LCLS) at Stanford, and powerful X-ray free-electron lasers (XFELs) in several other geographical locations, the question of radiation damage as a critical determinant looms large.

Howells *et al.*[81] have, for example, assessed the resolution limitation due to radiation damage in X-ray diffraction microscopy (notably of the CXDI kind described in Sec. 7.3). These workers distinguish, *inter alia*, between the radiation dose required for imaging, on the one hand, from the question of how much dose can a specimen tolerate before unacceptable degradation occurs to images at a given resolution on the other. They used a statistical analysis based on the so-called "dose-fractionation theorem" of Hegerl and Hoppe[82] to calculate the dose needed to make an image of a single "life-science" sample by CXDI with a given resolution. It transpires that the needed dose scales with the inverse fourth power of the resolution, i.e., the required dose for imaging as a function of resolution (d) is approximately $1/d^4$; compare with Eq. (5.48) for the case of 2D. This significant result shows that there is a very severe damage penalty incurred through small increases in resolution. Howells *et al.*[81] conclude that (synchrotron) radiation damage will limit X-ray imaging of protein in water to about 100 Å resolution.

Solem[83] made the suggestion, early on, that one should use pulses shorter than the time scale of destruction. Recent calculations along these lines, using a more complete model of the X-ray interaction that occurs in a single-shot, 2D image, indicate that spatial resolution below 10 Å could be achieved on whole unsectioned cells irradiated by soft X-ray pulses (see the articles and commentary in Ref. 84). Image resolution can be extended even further, as recently pointed out by Chapman,[85] perhaps to the interatomic length scale, so as to be able to resolve

the secondary structure of proteins, by combining many diffraction patterns from single particles in a stream of identical structures, such as macromolecules, protein complexes, or viruses. These particles may be in a liquid droplet stream. The merits of such an approach have been outlined by Spence and colleagues at Arizona State University.[86]

7.8 Summarizing Assessment

So far as the examination of microscopic samples ranging in size from 10^3 Å downwards is concerned, the electron microscope yields far more information pertaining to structural imaging and related properties than does the best available focused illumination currently achievable with synchrotron radiation. In large measure this is because the probe size for electrons can be made smaller by three orders of magnitude than that achievable with synchrotron beams (see Table 7.1). It follows that the number of unit cells that are interrogated (for structure determination) is far smaller (by orders of magnitude) than the corresponding number associated with the probes using synchrotron sources. Very recently, several synchrotron centers have developed means of producing smaller diameter (*ca.* 100 Å) probes. In addition to imaging by electron microscopy, a cornucopia[3] of other characteristic properties of minute objects are also readily recorded, such as chemical composition (by EELS, EFTEM or EDX), crystallographic symmetry, unit cell dimensions and crystallographic phase (see Chaps. 3 and 4). The applications in biological imaging are supreme. The power of electron microscopy can be seen by its use in numerous disciplines, from chemistry, physics and biology to materials science and mineralogy, and by the time resolution now achievable.

Where imaging studies using synchrotrons reign supreme is in the examination of thick specimens and those that are placed in contact with liquids or gases at high pressure and temperature. Environmental electron microscopy has already advanced greatly (see Chaps. 3 and 4) and biological macromolecules are now routinely studied in vitreous ice, as previously discussed. The synchrotron has much else to commend it. With the near-atomic resolution it can achieve via Laue and Bragg crystallography (i.e., through reciprocal space), the structures of inorganic, organic, and biological molecules can be straightforwardly determined. With the anticipated laser-like pulses of X-rays of about 100 fs duration and 10^{12} photons per pulse (from X-ray free-electron facilities), time-resolved Laue crystallography (as discussed in Sec. 7.6) could lead to a significant advance in X-ray imaging, especially in studies of proteins; however, radiation damage is a serious consideration, as discussed above. The hope is that the extreme intensity of the pulses will also enable much smaller sizes of crystals to be investigated. There

has been speculation that the required "crystal size may shrink all the way down to single molecules, giving 3D movies of conformational dynamics and chemical reactions, and allowing the imaging of macromolecules that cannot be easily crystallized."[85]

Electron crystallography, when dynamical scattering is taken into account, has been shown to provide structures from the combined real-space and Fourier-space approach.[87] Atoms can also be directly visualized in real space, as discussed in previous chapters. In proteins, deep insights pertaining to the visualization of intermediate structures involved in intricate reaction pathways (from nanoseconds to seconds) have been attained via time-resolved Laue crystallography, and efforts are now directed to studying them with femtosecond resolution in UEM (Chaps. 6 and 8).

On the industrial side of structural investigation, the synchrotron holds prime place because it permits a wide variety of *in situ* studies of solid catalysts (via EXAFS and XRD, as well as the feasibility of combining wide-angle, small-angle, and ultrasmall angle X-ray scattering — WAXS, SAXS and USAXS, respectively — for the direct imaging of open-structure and hybrid catalysts as demonstrated by de Moor *et al.*)[8] that can be examined under conditions identical to those that are used in commercial applications. The recent availability of ultrasensitive X-ray CCD cameras enables the changing chemical composition along a catalytic reactor tube to be probed with commendable spatiotemporal resolution and at least as good as PET or MRI.

As noted by Spence and Howells,[11] the unique capability of field-emission STEM instruments is their ability to combine atomic resolution imaging with energy-loss spectroscopy on a similar scale of spatial resolution, using magnetic lenses with sub-angstrom resolution to form a focused probe. They also emphasize that a fundamental limitation of the brightness of both field-emission electron sources and undulator soft X-ray sources is the Coulomb interaction between electrons. With timed, single-electron packets of high degeneracy, as discussed in this text for UEM, this problem is circumvented for different aspects of electron microscopy. The recent advance (see Ref. 88 and Chap. 6) of energy and time (down to the femtosecond) resolution in EELS-UEM and of CB-UEM (see Ref. 89 and Chap. 6) promises numerous applications, and when combined with imaging and diffraction capabilities of TEM and STEM the four dimensions discussed here are all within exploratory reach for inorganic materials and biological structures.

References

1. (a) J. C. H. Spence, private communication, May 2009.
   (b) J. M. Thomas and J.-C. Hernandez-Garrido, *Angew. Chem., Intl. Ed.*, **48**, 3904 (2009).

2. M. A. Le Gros, G. McDermott and C. A. Larabell, *Curr. Opin. Struct. Biol.*, **15**, 593 (2005).
3. J. M. Thomas, in "*Physical Biology: From Atoms to Medicine*," Ed. A. H. Zewail, Imperial College Press, London, p. 51 (2008).
4. C. G. Schroer, M. Kuhlmann, T. F. Günzler, B. Lengeler, M. Richwin, B. Griesebock, D. Lützenkirchen-Hecht, R. Frahm, E. Ziegler, A. Mashayekhi, D. R. Haeffner, J.-D. Grunwaldt and A. Baiker, *Appl. Phys. Lett.*, **82**, 3360 (2003).
5. J. Miao, P. Charalambous, J. Kirz and D. Sayre, *Nature*, **400**, 342 (1999).
6. C. G. Schroer, P. Boye, J. M. Feldkamp, J. Patommel, A. Schropp, A. Schwab, S. Stephan, M. Burghammer, S. Schöder and C. Riekel, *Phys. Rev. Lett.*, **101**, 090801 (2008).
7. E. Stavitski, M. H. F. Kox, I. Swart, F. M. F. de Groot and B. M. Weckhuysen, *Angew. Chem.*, *Intl. Ed.*, **47**, 3543 (2008).
8. (a) P.-P. E. A. de Moor, T. P. M. Beelen, B. U. Komanschek, L. W. Beck, P. Wagner, M. E. Davis and R. A. van Santen, *Chem. Eur. J.*, **5**, 2083 (1999).
   (b) B. M. Weckhuysen, *Angew. Chem., Intl. Ed.*, **48**, 4910 (2009).
9. G. Jonkers, K. A. Vonkeman, S. W. A. van der Wal and R. A. van Santen, *Nature*, **355**, 63 (1992).
10. (a) A. J. Sederman, M. D. Mantle, C. P. Dunckley, Z. Huang and L. F. Gladden, *Catal. Lett.*, **103**, 1 (2005).
    (b) L.-S. Bouchard, S. R. Burt, M. S. Anwar, K. V. Kovtunov, I. V. Koptyug and A. Pines, *Science*, **319**, 442 (2008).
11. J. C. H. Spence and M. R. Howells, *Ultramicroscopy*, **93**, 213 (2002).
12. D. T. Attwood, "*Soft X-rays and Extreme Ultraviolet Radiation: Principles and Applications*," Cambridge University Press, Cambridge (2000).
13. E. Di Fabrizio, F. Romanato, M. Gentili, S. Cabrini, B. Kaulich, J. Susini and R. Barrett, *Nature*, **401**, 895 (1999).
14. W. Yun, B. Lai, A. A. Krasnoperova, E. Di Fabrizio, Z. Cai, F. Cerrina, Z. Chen, M. Gentili and E. Gluskin, *Rev. Sci. Instrum.*, **70**, 3537 (1999).
15. G.-C. Yin, M.-T. Tang, Y.-F. Song, F.-R. Chen, K. S. Liang, F. W. Duewer, W. Yun, C.-H. Ko and H.-P. D. Shieh, *Appl. Phys. Lett.*, **88**, 241115 (2006).
16. (a) D. Attwood, *Nature*, **442**, 642 (2006).
    (b) H. Ade and H. Stoll, *Nature Mater.*, **8**, 281 (2009).
17. A. C. Kak and M. Slaney, "*Principles of Computerized Tomographic Imaging*," IEEE Press, New York (1988).
18. (a) E. de Smit, I. Swart, J. F. Creemer, G. H. Hoveling, M. K. Gilles, T. Tyliszczak, P. J. Kooyman, H. W. Zandbergen, C. Morin, B. M. Weckhuysen and F. M. F. de Groot, *Nature*, **456**, 222 (2008).
    (b) See also: E. de Smit, I. Swart, J. F. Creemer, C. Karunakaran, D. Bertwistle, H. W. Zandbergen, F. M. F. de Groot and B. M. Weckhuysen, *Angew. Chem.*, *Intl. Ed.*, **48**, 3632 (2009).
19. P. A. Midgley, E. P. W. Ward, A. B. Hungría and J. M. Thomas, *Chem. Soc. Rev.*, **36**, 1477 (2007).
20. J. Radon, *Ber. Saechs. Akad. Wiss. Leipzig*, *Math.-Phys. Kl.*, **69**, 262 (1917).
21. S. R. Deans, "*The Radon Transform and Some of Its Applications*," Wiley, New York (1983).
22. C. A. Larabell and M. A. Le Gros, *Mol. Biol. Cell*, **15**, 957 (2004).

23. W. Gu, L. D. Etkin, M. A. Le Gros and C. A. Larabell, *Differentiation*, **75**, 529 (2007).
24. R. M. Glaeser, *J. Microsc*., **12**, 133 (1971).
25. J. Evans, *Phys. Chem. Chem. Phys*., **8**, 3045 (2006).
26. (a) D. Sayre, in "*Imaging Processes and Coherence in Physics*," Eds. M. Schlenker, M. Fink, J. P. Goedgebuer, C. Malgrange, J. C. Viénot and R. H. Wade, Springer, Berlin, p. 229 (1980).
    (b) J. Miao, T. Ishikawa, Q. Shen and T. Earnest, *Annu. Rev. Phys. Chem*., **59**, 387 (2008).
27. R. W. Gerchberg and W. O. Saxton, *Optik*, **34**, 277 (1971).
28. (a) S. Marchesini, H. He, H. N. Chapman, S. P. Hau-Riege, A. Noy, M. R. Howells, U. Weierstall and J. C. H. Spence, *Phys. Rev. B*., **68**, 140101 (2003).
    (b) J. R. Fienup, *Appl. Opt*., **21**, 2758 (1982).
    (c) J. C. H. Spence in "*Science of Microscopy*," Vol. 2, Eds. P. W. Hawkes and J. C. H. Spence, Springer, New York, p. 1196 (2007).
    (d) S. Marchesini, *Rev. Sci. Instrum*. **78**, 011301 (2007).
29. H. N. Chapman, A. Barty, S. Marchesini, A. Noy, S. P. Hau-Riege, C. Cui, M. R. Howells, R. Rosen, H. He, J. C. H. Spence, U. Weierstall, T. Beetz, C. Jacobsen and D. Saphiro, *J. Opt. Soc. Am. A*, **23**, 1179 (2006).
30. G. J. Williams, M. A. Pfeifer, I. A. Vartanyants and I. K. Robinson, *Phys. Rev. Lett*., **90**, 175501 (2003).
31. J. Miao, C.-C. Chen, C. Song, Y. Nishino, Y. Kohmura, T. Ishikawa, D. Ramunno-Johnson, T.-K. Lee and S. H. Risbud, *Phys. Rev. Lett*., **97**, 215503 (2006).
32. H. M. L. Faulkner and J. M. Rodenburg, *Phys. Rev. Lett*., **93**, 023903 (2004).
33. P. Thibault, M. Dierolf, A. Menzel, O. Bunk, C. David and F. Pfeiffer, *Science*, **321**, 379 (2008).
34. R. Neutze, R. Wouts, D. van der Spoel, E. Weckert and J. Hajdu, *Nature*, **406**, 752 (2000).
35. C. Song, H. Jiang, A. Mancuso, B. Amirbekian, L. Peng, R. Sun, S. S. Shah, Z. H. Zhou, T. Ishikawa and J. Miao, *Phys. Rev. Lett*., **101**, 158101 (2008).
36. C.-C. Chen, J. Miao, C. W. Wang and T. K. Lee, *Phys. Rev. B*, **76**, 064113 (2007).
37. (a) H. M. Rietveld, *J. Appl. Crystallogr.*, **2**, 65 (1969).
    (b) K. D. M. Harris and E. Y. Cheung, *Chem. Soc. Rev.*, **33**, 526 (2004).
38. F. Guo, M. Casadesus, E. Y. Cheung, M. P. Coogan and K. D. M. Harris, *Chem. Commun*., 1854 (2006).
39. S. Habershon and A. H. Zewail, *Chem. Phys. Chem*., **7**, 353 (2006).
40. (a) J. P. Attfield, A. W. Sleight and A. K. Cheetham, *Nature*, **322**, 620 (1986).
    (b) J. P. Attfield, *Nature*, **343**, 46 (1990).
    (c) G. Oszlányi and A. Sütő, *Acta Cryst. A*, **61**, 147 (2005).
    (d) J. Wu, K. Leinenweber, J. C. H. Spence and M. O'Keeffe, *Nature Mater*., **5**, 647 (2006).
41. G. Sankar, J. K. Wyles, R. H. Jones, J. M. Thomas, C. R. A. Catlow, D. W. Lewis, W. Clegg, S. J. Coles and S. J. Teat, *Chem. Commun*., 117, (1998).
42. G. M. Sheldrick, SHELXS-86, Program for Crystal Structure Determination, University of Göttingen, Germany, 1986.
43. (a) J. M. Thomas, *Angew. Chem*., *Intl. Ed*., **38**, 3588 (1999).

(b) A. T. Davies, G. Sankar, C. R. A. Catlow and S. M. Clark, *J. Phys. Chem. B*., **101**, 10115 (1997).
44. F. Rey, G. Sankar, J. M. Thomas, P. A. Barrett, D. W. Lewis, C. R. A. Catlow, S. M. Clark and G. N. Greaves, *Chem. Mater*., **7**, 1435 (1995).
45. C. Bressler and M. Chergui, *Chem. Rev*., **104**, 1781 (2004).
46. J. M. Thomas and W. J. Thomas, "*Principles and Practice of Heterogeneous Catalysis*," VCH, Weinheim, pp. 195, 196 (1996).
47. J. Wong, F. W. Lytle, R. P. Messmer and D. H. Maylotte, *Phys. Rev. B*, **30**, 5596 (1984).
48. T. Maschmeyer, F. Rey, G. Sankar and J. M. Thomas, *Nature*, **378**, 159 (1995).
49. J. M. Thomas, R. Raja, G. Sankar and R. G. Bell, *Acc. Chem. Res*., **34**, 191 (2001).
50. J. W. Couves, J. M. Thomas, D. Waller, R. H. Jones, A. J. Dent, G. E. Derbyshire and G. N. Greaves, *Nature*, **354**, 465 (1991).
51. J. M. Thomas and G. N. Greaves, *Science*, **265**, 1675 (1994).
52. B. S. Clausen, H. Topsøe and R. Frahm, *Adv. Catal*., **42**, 315 (1998).
53. J. M. Thomas, *Chem. Euro. J*., **3,** 1557 (1997).
54. M. Chergui and A. H. Zewail, *Chem. Phys. Chem*., **10**, 28 (2009), and references therein.
55. (a) H. Ihee, V. A. Lobastov, U. M. Gomez, B. M. Goodson, R. Srinivasan, C.-Y. Ruan and A. H. Zewail, *Science*, **291**, 458 (2001).
(b) H. Ihee, J. Kua, W. A. Goddard, III and A. H. Zewail, *J. Phys. Chem. A*, **105**, 3623 (2001).
56. (a) J. H. Lee, T. K. Kim, J. Kim, Q. Kong, M. Cammarata, M. Lorenc, M. Wulff and H. Ihee, *J. Am. Chem. Soc*., **130**, 5834 (2008).
(b) T. K. Kim, J. H. Lee, M. Wulff, Q. Kong and H. Ihee, *Chem. Phys. Chem*., **10**, 1958 (2009).
57. R. Neutze, R. Wouts, S. Techert, J. Davidsson, M. Kocsis, A. Kirrander, F. Schotte and M. Wulff, *Phys. Rev. Lett*., **87**, 195508 (2001).
58. M. Wulff, S. Bratos, A. Plech, R. Vuilleumier, F. Mirloup, M. Lorenc, Q. Kong and H. Ihee, *J. Chem. Phys*., **124**, 034501 (2006).
59. Q. Kong, J. Kim, M. Lorenc, T. K. Kim, H. Ihee and M. Wulff, *J. Phys. Chem. A*, **109**, 10451 (2005).
60. (a) H. Ihee, M. Lorenc, T. K. Kim, Q. Y. Kong, M. Cammarata, J. H. Lee, S. Bratos and M. Wulff, *Science*, **309**, 1223 (2005).
(b) P. Anfinrud and F. Schotte, *Science*, **309**, 1192 (2005).
61. S. Bratos, F. Mirloup, R. Vuilleumier and M. Wulff, *J. Chem. Phys*., **116**, 10615 (2002).
62. M. Cammarata, M. Lorenc, T. K. Kim, J. H. Lee, Q. Y. Kong, E. Pontecorvo, M. Lo Russo, G. Schiró, A. Cupane, M. Wulff and H. Ihee, *J. Chem. Phys*., **124**, 124504 (2006).
63. A. Plech, M. Wulff, S. Bratos, F. Mirloup, R. Vuilleumier, F. Schotte and P. A. Anfinrud, *Phys. Rev. Lett*., **92**, 125505 (2004).
64. S. A. Rice, *Nature*, **429**, 255 (2004).
65. C. Bressler, C. Milne, V.-T. Pham, A. ElNahhas, R. M. van der Veen, W. Gawelda, S. Johnson, P. Beaud, D. Grolimund, M. Kaiser, C. N. Borca, G. Ingold, R. Abela and M. Chergui, *Science*, **323**, 489 (2009).

66. (a) W. Gawelda, V.-T. Pham, M. Benfatto, Y. Zaushytsin, M. Kaiser, D. Grolimund, S. L. Johnson, R. Abela, A. Hauser, C. Bressler and M. Chergui, *Phys. Rev. Lett*., **98**, 057401 (2007).
(b) W. Gawelda, A. Cannizzo, V.-T. Pham, F. van Mourik, C. Bressler and M. Chergui, *J. Am. Chem. Soc*., **129**, 8199 (2007).
67. C. Bressler, R. Abela and M. Chergui, *Z. Kristallogr*., **223**, 307 (2008); see also Ref. 45.
68. J. E. Monat and J. K. McCusker, *J. Am. Chem. Soc*., **122**, 4092 (2000).
69. M. Khalil, M. A. Marcus, A. L. Smeigh, J. K. McCusker, H. H. W. Chong and R. W. Schoenlein, *J. Phys. Chem. A*, **110**, 38 (2006).
70. L. X. Chen, X. Zhang, E. C. Wasinger, K. Attenkofer, G. Jennings, A. Z. Muresan and J. S. Lindsey, *J. Am. Chem. Soc*., **129**, 9616 (2007).
71. L. X. Chen, G. Jennings, T. Liu, D. J. Gosztola, J. P. Hessler, D. V. Scaltrito and G. J. Meyer, *J. Am. Chem. Soc*., **124**, 10861 (2002).
72. L. X. Chen, G. B. Shaw, I. Novozhilova, T. Liu, G. Jennings, K. Attenkofer, G. J. Meyer and P. Coppens, *J. Am. Chem. Soc*., **125**, 7022 (2003).
73. Z. Ren, D. Bourgeois, J. R. Helliwell, K. Moffat, V. Šrajer and B. L. Stoddard, *J. Synchrotron Rad*., **6**, 891 (1999).
74. "*Time Resolved Diffraction*," Eds. J. R. Helliwell and P. M. Rentzepis, Oxford University Press, Oxford (1997).
75. K. Moffat, *Chem. Rev*., **101**, 1569 (2001).
76. F. Schotte, M. Lim, T. A. Jackson, A. V. Smirnov, J. Soman, J. S. Olson, G. N. Phillips, Jr., M. Wulff and P. A. Aufinrud, *Science*, **300**, 1944 (2003).
77. D. Bourgeois, B. Vallone, F. Scholte, A. Arcovito, A. E. Miele, G. Sciara, M. Wulff, P. Anfinrud and M. Brunori, *Proc. Natl. Acad. Sci. USA*, **100**, 8704 (2003).
78. H. Frauenfelder, B. H. McMahon and P. W. Fenimore, *Proc. Natl. Acad. Sci. USA*, **100**, 8615 (2003).
79. (a) H. Ihee, S. Rajagopal, V. Šrajer, R. Pahl, S. Anderson, M. Schmidt, F. Schotte, P. A. Anfinrud, M. Wulff and K. Moffat, *Proc. Natl. Acad. Sci. USA*, **102**, 7145 (2005).
(b) G. Rubinstenn, G. W. Vuister, F. A. A. Mulder, P. E. Düx, R. Boelens, K. J. Hellingwerf and R. Kaptein, *Nature Struct. Biol*., **5**, 568 (1998).
(c) G. E. O. Borgstahl, D. R. Williams and E. D. Getzoff, *Biochemistry*, **34**, 6278 (1995).
80. R. Henderson, *Quart. Rev. Biophys*., **37**, 3 (2004).
81. M. R. Howells, T. Beetz, H. N. Chapman, C. Cui, J. M. Holton, C. J. Jacobsen, J. Kirz, E. Lima, S. Marchesini, H. Miao, D. Sayre, D. A. Shapiro, J. C. H. Spence and D. Starodub, *J. Electron Spectrosc. Rel. Phenom*., **170**, 4 (2009).
82. R. Hegerl and W. Hoppe, *Z. Naturforsch. A*, **31**, 1717 (1976).
83. J. C. Solem, *J. Opt. Soc. Am. B*, **3**, 1551 (1986).
84. (a) M. Bergh, G. Huldt, N. Tîmneanu, F. R. N. C. Maia and J. Hajdu, *Quart. Rev. Biophys*., **41**, 181 (2008).
(b) See Ref. 34 and references therein.
(c) R. Henderson, *Nature*, **415**, 833 (2002).
(d) J. C. H. Spence, *Nat. Photon*., **2**, 390 (2008).
85. H. N. Chapman, *Nature Mater*., **8**, 299 (2009).

86. D. P. DePonte, U. Weierstall, K. Schmidt, J. Warner, D. Starodub, J. C. H. Spence and R. B. Doak, *J. Phys. D*, **41**, 195505 (2008).
87. X. Zou and S. Hovmöller, *Acta Cryst. A*, **64**, 149 (2008), and references therein.
88. (a) F. Carbone, B. Barwick, O.-H. Kwon, H. S. Park, J. S. Baskin and A. H. Zewail, *Chem. Phys. Lett*., **468**, 107 (2009).
(b) F. Carbone, O.-H. Kwon and A. H. Zewail, *Science*, **325**, 181 (2009).
(c) J. M. Thomas, *Angew. Chem.*, *Intl. Ed*., **48**, 8824 (2009).
89. A. Yurtsever and A. H. Zewail, *Science*, **326**, 708 (2009).

This page intentionally left blank

Chapter 8

4D Visualization
Past, Present, and Future

From the very beginning of mankind's efforts to pursue scientific inquiry into the structures and laws of nature, inexorable progress has been made first to visualize processes and then to conceptualize the structures and dynamics that constitute the essence of the physical and biological worlds we inhabit. In the microscopic world of inanimate matter and biological systems, the fundamental scales of length and time required to resolve the nature of structure and dynamics are, at the atomic scale, the picometer and femtosecond. These are the dimensions that form the basis for defining and understanding the intrinsic function at larger and longer scales. Once it becomes feasible to visualize complexity in the four dimensions of space and time, we are better placed to fathom fully the emergent behavior of collective and coherent function.

In this concluding chapter, we extrapolate from the evolutionary and revolutionary accomplishments made so far into projections concerning future opportunities and challenges in the myriad applications that are now possible. First, we begin by addressing the actual meaning of "complexity" and the importance of visualization.[1]

8.1 Visualization and Complexity

Realization of the importance of visualization and observation is evident in the exploration of natural phenomena from the very small to the very large. A century ago, the atom appeared complex, a "raisin or plum pie of no structure," until it was visualized on the appropriate length and time scales. Similarly, with telescopic observations a central dogma of the cosmos was changed and complexity yielded to the simplicity of the heliocentric structure and motion in the entire solar system. From the atom to the universe, the length and time scales span extremes of powers of 10 (Fig. 8.1). The electron in the *s*-orbital of hydrogen has a "period" of sub-femtoseconds, and the size of atoms is on the nanometer scale or less. The lifetime of our universe is ~13 billion years and, considering the light year ($\sim 10^{16}$ m), its length scale is on the order of 10^{26} m. In between these scales lies the world of

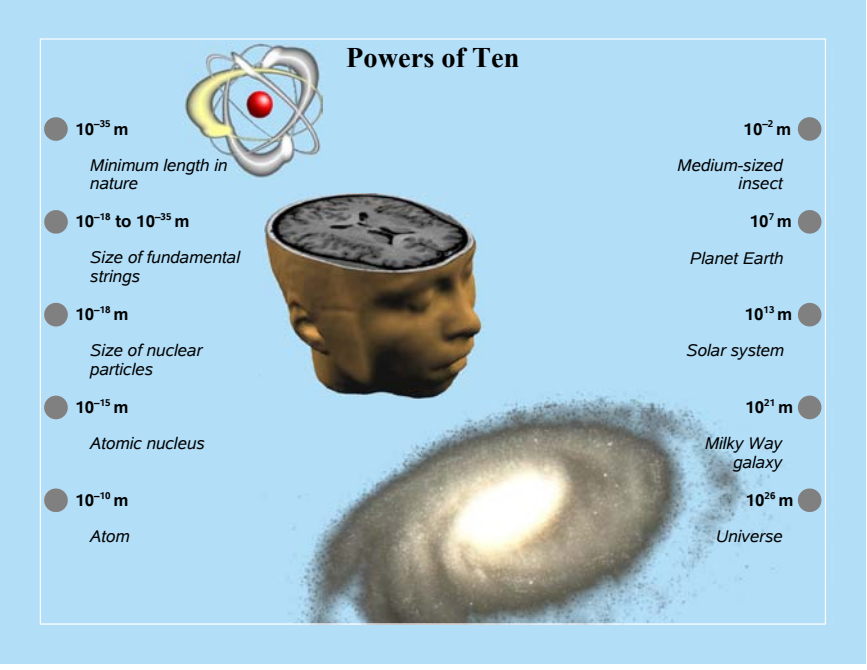


Figure 8.1 Powers of ten, from the atom to the universe.[1a]

life processes with scales varying from nanometers to centimeters and from femtoseconds to seconds (Fig. 8.2).

For the atom, major advances have been made: not only has the language (quantum mechanics) been developed to describe its behavior and interactions, but also the atom itself has been tamed. Since the first conceptual idea that "*there are only atoms and the void*" postulated by Democritus more than two millennia ago, we now know the components of atoms and how to detect them, count each one, cool them to sub-kelvin temperatures or trap and suspend them with light. Of major impact in molecular sciences is the ability to observe atoms at rest at angstrom resolution and atoms in motion at femtosecond resolution. X-ray and electron techniques have led to structural determination, beginning with the structures of ionic solids consisting of just two atoms (sodium chloride or zinc sulfide), and now culminating in the determination of structures of molecular machines with the number of atoms exceeding 10^5 (Fig. 8.3). Similarly, for dynamics, the time scales at the beginning of the 20th century were practically seconds (hydrolysis of sugars); now, it has reached the atomic scale of femtoseconds (Fig. 8.4) and attoseconds (see Fig. 8.5).

In the early days of DNA structural determination (1950s), a cardinal concept, in vogue at that time, was encapsulated in Francis Crick's statement: "*If you want to know the function, determine the structure.*" This view pervaded the thinking

at the time, and it was what drove Max Perutz and John Kendrew earlier in their studies of proteins. But as we learn more about complexity, it is clear that the so-called "structure–function" correlation is insufficient to establish the mechanisms that hold sway in complex systems. For example, the structures of many proteins have been determined, but we still do not understand how they fold, how they selectively recognize other molecules, and how the matrix water assists folding and the role it plays in directionality, selectivity, and recognition. The proteins hemoglobin and myoglobin (a subunit of hemoglobin) have unique functions: the former is responsible for transporting oxygen in the blood of vertebrates, while the latter carries and stores oxygen in muscle cells. The 3D structure of both proteins has been determined (by Perutz and Kendrew), but we still do not understand the differences in behavior in the oxygen uptake by these two related proteins, the role of hydration, and the exact nature of the forces that control the dynamics of oxygen binding and liberation from the heme group. Visualization of the changing structures during the course of its functional operation is what is needed.

Superficially, it may seem that the simplest degree of structural change in complexity would be that of chemical reactions involving a mere 10 or so atoms. However, such transformations result from many-body interactions, and, until recently, the determination of their isolated intermediate structures was impossible

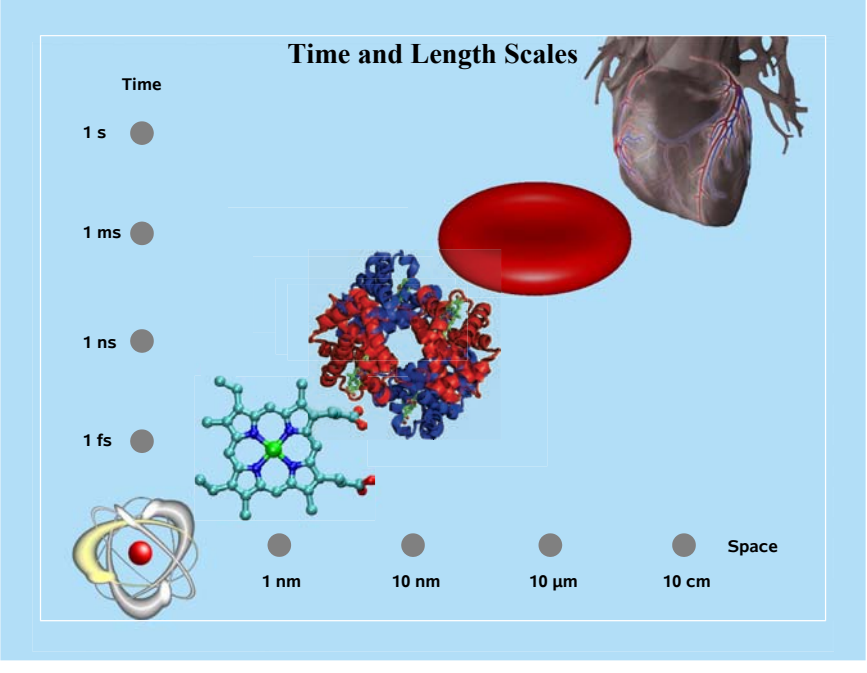


Figure 8.2 Time and length scales, from atoms to organs. Shown are schematically represented atom, heme molecule, hemoglobin, red blood cell, and a heart.[1a]

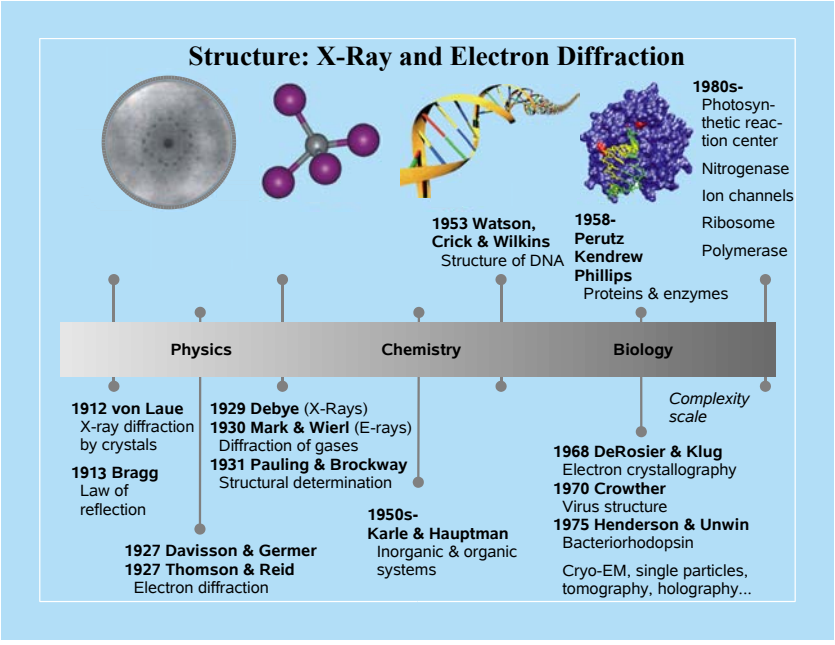


Figure 8.3 Developments in structural determination, from sodium chloride crystals to DNA and proteins, including the Nobel Prizes awarded (date given is that of development).[1a]

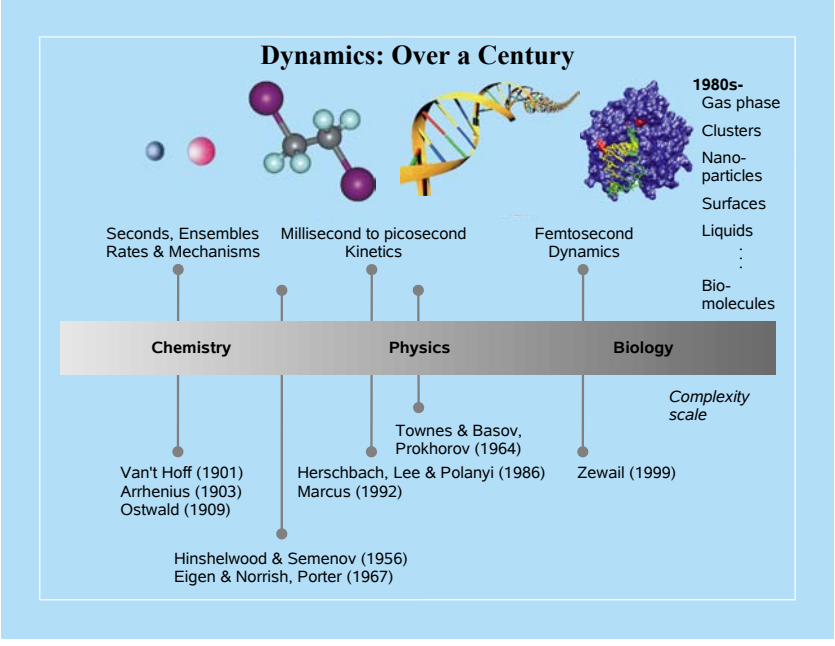


Figure 8.4 Developments in studies of kinetics and dynamics, from seconds to the femtosecond atomic scale as cited (date indicated) by Nobel Prizes awarded over the past century.[1a]

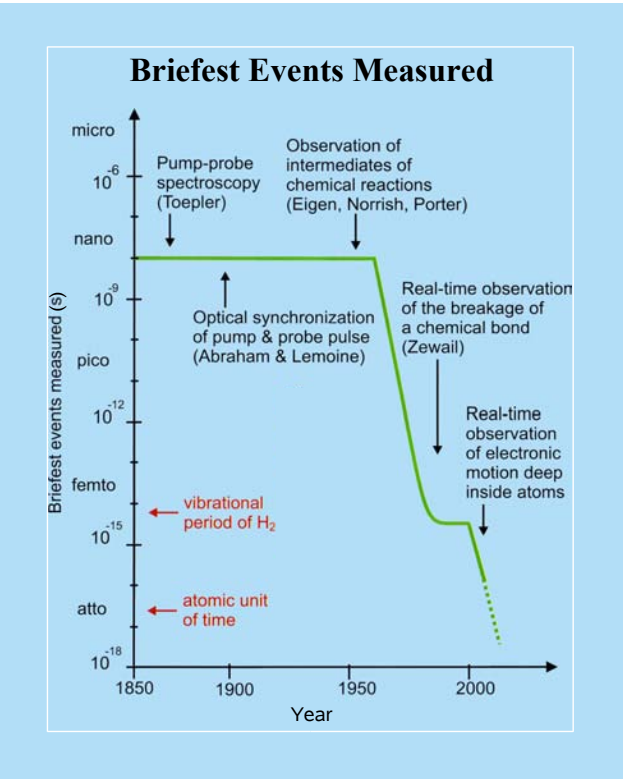


Figure 8.5 Krausz–Ivanov summary of the briefest events measured, from 1850 to the present;[15c] see also Ref. 14 and Chap. 5.

because of their fleeting nature on the time scale of a picosecond or less. These transient structures are optically "dark" in that they undergo radiationless transitions into reactive or nonreactive channels, and in most cases they do not emit light. The bifurcation obscures the mechanism and only through studies of both their structures and dynamics (see, for example, Refs. 2 and 3) can one resolve the complexity of pathways and elucidate mechanisms — and hopefully control the behavior. It is important to realize that even for a system of three atoms, trajectories of motion can result in different final structures of several channels: an ABC system is, in principle, destined to produce AB + C, A + BC, AC + B, and A + B + C.

A supreme example of large-scale complexity is evident in correlated physical systems exhibiting phase transitions or self-assembly, and in biological systems with emergent behavior.[1] For the latter, the energy landscape involves very complicated pathways, including those that lead to a multitude of conformations with some that are "active" and others that are "inactive" in the biological function. Moreover, the landscapes define "good" and "bad" regions, the latter being

descriptive of the origin of molecular diseases, as noted below. It is remarkable that the robustness and function of these "molecules of life" are the result of a balance of weak forces — hydrogen bonding, electrostatic forces, dispersion, and hydrophobic interactions — all of energy on the order of a few kcal/mole, or ~0.1 eV or less. Determining time-averaged molecular structures is important and has led to an impressive list of achievements, for which more than 10 Nobel Prizes have been awarded, but the structures relevant to function are those that exist in the nonequilibrium state. Understanding their behavior requires an integration of the trilogy: structure, dynamics, and function.

8.2 Complexity Paradox: Coherence and Creative Chaos

How does a complex system, supposedly chaotic in origin, selectively yield a coherent function? And what drives selectivity in dynamics to yield a particular native structure? Among the many profound questions, these two are the most challenging in the understanding of complexity. The questions can be further illuminated if we consider how structure, dynamics, and function are correlated in a system of thousands of atoms. The mechanism for the vision of our eyes — designed with incredible molecular precision — may illustrate the pertinence of such questions.

Vision is the result of the conversion of light energy to an electrochemical impulse. The impulse is transmitted through neurons to the brain where signals from all the visual receptors are interpreted. The initial receptors contain the pigment rhodopsin, which is located in the rods of the retina. This pigment consists of an organic molecule, retinal, in association with the protein opsin. A change in the shape of retinal, which involves twisting of a chemical double bond, apparently gives the signal to opsin to undergo a sequence of dark (thermal) reactions involved in triggering neural excitations.

The speed of twisting is exceptionally rapid. The primary event of twisting in rhodopsin[4] is basically that of a double bond after photoexcitation.[5] It occurs in about 200 fs, and such rapidity indicates selectivity with a nonstatistical distribution of energy among all nuclear motions: if energy is to be statistically distributed, the rate of twisting would be orders of magnitude slower. Following light absorption, the large retinal molecule twists on the time scale of molecular vibrations; the entire process is accomplished in a coherent manner. This coherence is credited for making possible the high (70 percent or more) efficiency of the initial step, despite the large size of the rhodopsin molecule and the many possible channels for dissipation of energy. Large molecular systems are usually described by statistical theories and only when such systems are directly and coherently activated can we hope to observe nonstatistical behavior,[6a] and such coherence is pertinent to the function.[6b] The indicated selectivity is a paradox, as the protein is designed to

assist the retinal molecule to twist efficiently; retinal in solution loses this unique feature. How this occurs is a question still under scrutiny.

Another paradox of complexity relates, in general, to the dynamics of protein folding, i.e., how does the protein acquire the native state, and in such a relatively short time? One might think that a protein would try random visits of all possible conformations until it eventually settles into the correct structure. But how long would it take to do so? For proteins with N residues, each with the ability to twist or contort, there exist about 10^N possible conformations (ignoring other possible motions involving side chains). With 100 residues (amino acids), the protein would, in principle, need time that far exceeds the age of the universe (~13 billion years) — the so-called Levinthal's paradox. Obviously, if this were true, we would not be here!

The picture is further complicated by the matrix in which the folding takes place. Proteins function in water, and the time scale for the relevant motion of water molecules (picoseconds) is on the scale of elementary motions of the protein, especially in the stages of secondary structure formation. Accordingly, the role of water, through weak hydrophilic/hydrophobic and hydrogen-bonding interactions, is an integral part of this complexity.[7] Despite the numerous many-body interactions, a collective guiding bias is needed to direct the system into the native conformational structure in a reasonable time, which is much shorter than a second (Fig. 8.6).

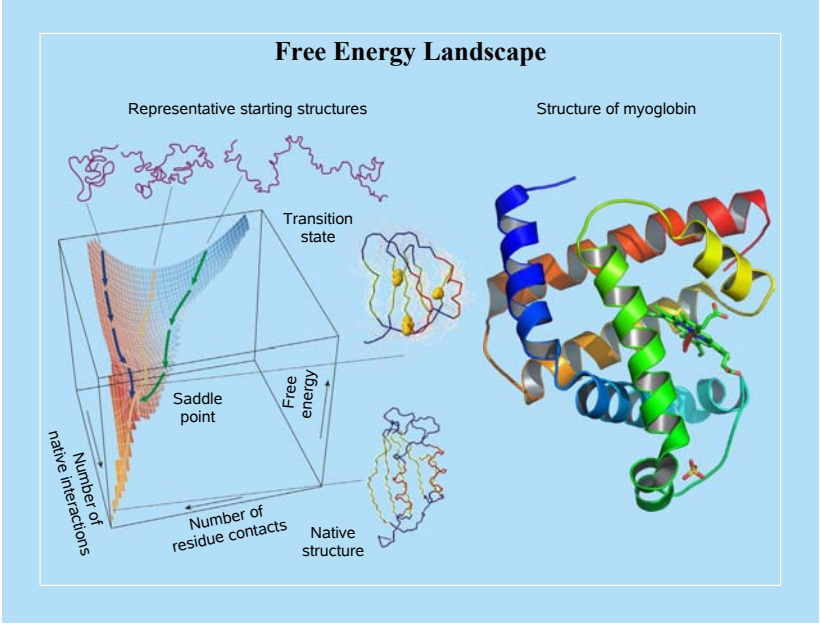


Figure 8.6 Free energy landscape[8] and the structure of myoglobin (adapted from *wikipedia.org*).

If, in the process of folding, a protein does not follow the "good" path, i.e., misfolds, then the outcome is detrimental. A biopolymer protein becomes abiological — a chemical polymer — and this misfolding leads to the emergence of diseases such as Alzheimer's, a neurodegenerative brain disease. The deposition of proteins in the form of amyloid fibrils or plaques as a result of misfolding and aggregation is now considered to be associated with other diseases such as Parkinson's, type 2 diabetes, and amyloidosis.[8] The deposits, depending on the disease, can be in the brain, in skeletal tissue, or in other organs. Clearly, a wide variety of questions is pertinent to this important and fascinating world of complex molecular systems, and without direct visualization many of these questions will remain unanswered.

8.3 From 2(3)D to 4D Microscopy

As discussed at the beginning of this book, the spatial resolution achieved in electron microscopy is ~0.8 Å, and by the year 2010 it is expected to be at about 0.5 Å. This state of the art and steps of progress in improving resolution (Fig. 8.7)

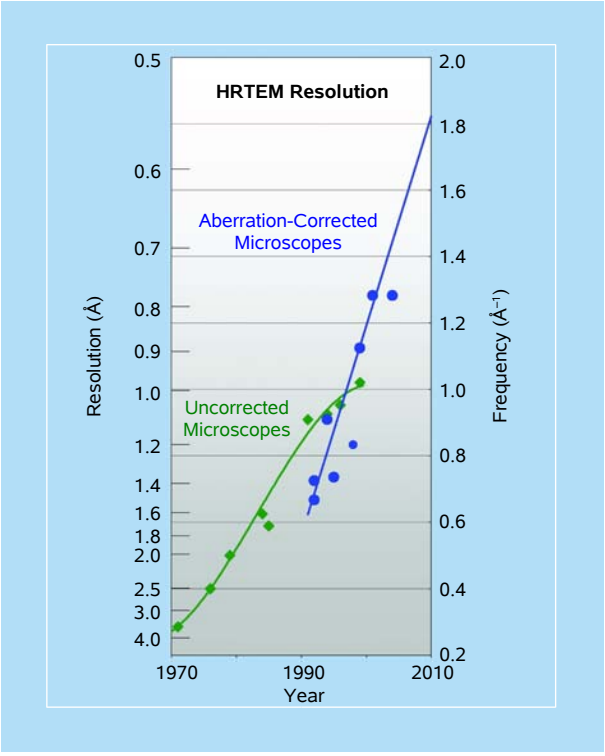


Figure 8.7 Progress in high-resolution transmission electron microscopy (HRTEM), from 1970 to the present.[9b]

are indeed impressive. The resolution of the Knoll–Ruska microscope was sub-micrometer, and with Scherzer focus (Chaps. 1 and 2) the nanometer scale was reached. The lens defects limit the resolution, but with the development of aberration-corrected microscopes, it became possible to reach the sub-angstrom scale in high-resolution imaging[9] for both TEM and STEM instruments. A major turning point in the evolution of EM was Crewe's imaging of atoms[9c] with his field-emission STEM techniques, as discussed below.

In principle, even better resolution can be reached if we consider the Rayleigh criterion (or the one modified for coherent sources, as discussed in Chap. 2) and the additional fact that the wavelength of the electron is in the picometer range. At present, the picometer resolution is achieved via selected-area and convergent-beam diffraction. However, we must recall that, unlike real-space images (that form at the image plane) and which result from interferences between scattered and unscattered rays, diffraction patterns (formed at the back focal plane) result from interferences between scattered rays only. Hence, in diffraction from complex objects the phase retrieval is not straightforward (the so-called *phase problem*; see Chap. 2).

One of the most powerful methods of imaging involves the use of "phase contrast," which makes possible the study of thin specimens with small changes in refractive index (potential) across the object (Chap. 5). This is the method of choice for biological specimens and for high-resolution imaging. For 2D crystals, electron crystallography is a powerful method in which both imaging and diffraction are jointly utilized to overcome the phase problem and to extract 3D structures. The Fourier transform of an image grid provides the phases for diffraction in the reciprocal (Fourier) space, as discussed in Chap. 2. Another equally important approach is single-particle cryo-EM, which is crystallography without crystals, done in vitreous ice to simulate the *in vivo* environment (see Chap. 4). With these and other variant methods, as illustrated in Fig. 1.7, it is now possible to obtain high-resolution images of complex materials and extended structures including cells and viruses.[10]

The time resolution of milliseconds or longer in conventional microscopes is limited by their detector response and electron counts. The approach of high-speed microscopes, using electronic triggers and nanosecond lasers, improved the temporal resolution, but with limits on the spatial one.[11] With 10^8 electrons in a pulse, the temporal resolution is nanoseconds and the spatial resolution was initially sub-micrometer,[11] and is now reaching tens of nanometers.[12a] Clearly, with nanosecond time resolution, it should be impossible to observe the elementary dynamics of processes (Fig. 8.8). Recent work using these nanosecond, single-pulse imaging techniques has examined kinetics of moving reactive fronts of phase separation[12b] and the embryonic stages of nucleation and crystallization.[13] However, for such single-shot image formation, the electron sources do not provide sufficient

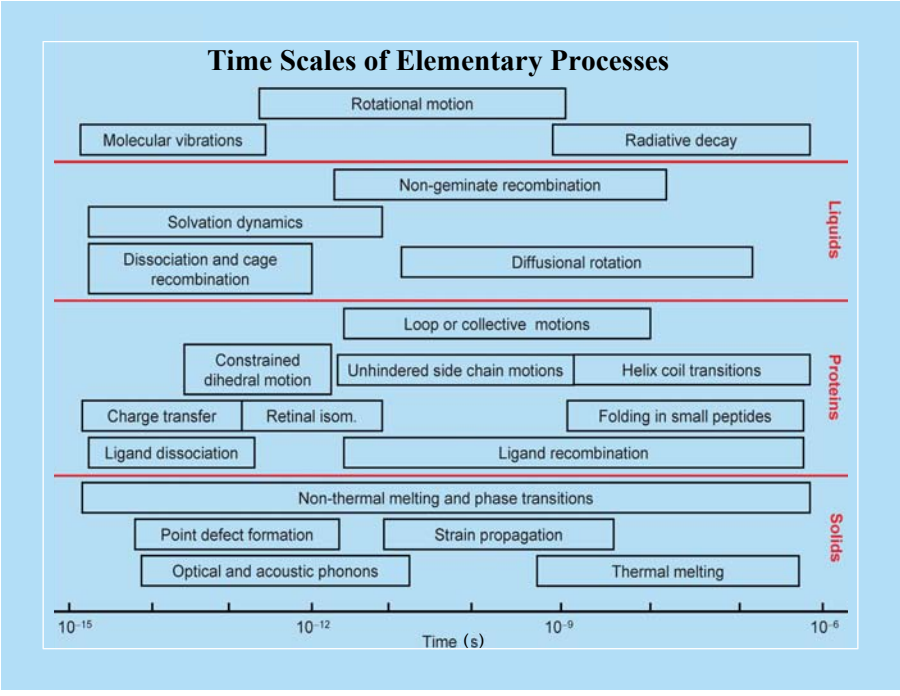


Figure 8.8 Time scales of typical processes in liquids, solids, proteins, and other biological systems.[16]

spatial coherence for high-resolution imaging with adequate numbers of electrons in durations of picoseconds or femtoseconds, although, as discussed below, there exist several schemes for reaching such a mode of imaging. Our interest at Caltech (see Chap. 6) has been in reaching the joint high spatial resolution and ultrashort time domain in microscopy and diffraction. Such domains require the machinery of the ultrafast techniques outlined in Chap. 5, which provides the precision of femtosecond resolution using "pump-probe" schemes and are not limited by detector response or electronic trigger circuitry.

With such techniques, the ultimate for spatiotemporal resolution is achievable in timed, single-electron microscopy. For single electrons, Dirac's famous dictum "*each electron interferes only with itself*" takes center stage in coherent stroboscopic imaging in ultrafast electron microscopy (UEM). The spatial resolution, as discussed in Chaps. 2 and 5, remains determined by the source characteristics without (transverse) Coulombic enlargement and (longitudinal) Boersch space–charge effect. UEM high-resolution imaging was confirmed by resolving 3.4 Å separations between the atomic planes in a graphitized carbon specimen (Fig. 6.3), at 0.9 Scherzer's focus. This resolution was obtained without using a field-emission gun (FEG) because the electrons were launched through the

photoelectric effect. However, in order to reach the 200 keV acceleration regime, the UEM was designed with an FEG-type arrangement. As outlined in Chaps. 5 and 6, in the single-electron mode of operation the degeneracy factor can be increased by several orders of magnitude, which opens up possibilities for studies in electron quantum optics.

The temporal characteristics in ultrafast imaging are not only determined by the duration of the pulse but also by its coherence time and length (see Chap. 5). Pulses which are transform-limited are coherent, and this feature is a unique trait of ultrafast electron packets as opposed to fast (nanosecond or longer) electron pulses, which do not possess such coherence properties — hence the boundaries that could be drawn in Fig. 1.6, which define the range of time resolution and coherence characteristics. The importance of initiating-pulse coherence is in creating spatially-localized (atomic-scale) wave packets which permit high-resolution imaging of dynamics, as detailed in Chap. 5. Such ultrafast optical and electron pulses can now be generated, but it is essential to distinguish between "pulse generation," as there exists a scientific community racing to achieve the shortest pulse duration, and the utilization of such pulses in exploring new scientific frontiers and gaining new knowledge.

Optical pulse generation, which is an integral part of UEM, has witnessed revolutionary advances in both duration and stability of the pulse. Together with the progress made in nonlinear optics, computation, and electronics, these advances have led to the attainment of femtosecond and attosecond pulses, which became the source for utilization in different studies, as discussed in Chap. 5. A review published by one of us[14] has highlighted important developments up to the femtosecond regime. The current state of the art is presented in more recent reviews and books that are highly recommended for further reading.[15] Figure 8.5, which is adapted from the review by Krausz and Ivanov,[15c] covers the most recent developments and includes attosecond pulse generation and spectroscopy (see also Ref. 14 and Chap. 5). Figure 8.8 shows the time scales for pertinent physical, chemical, and biological processes in different phases (liquids, solids, and macromolecular assemblies).[16] The studies of electron motions, for example, those characteristic of core electron excitation/recombination, can now be added to the lexicon of such processes.

In 4D UEM imaging, the central concept emphasized throughout this book is that of timed electron packets consisting of single (or few) noninteracting electrons; they provide the ultrashort temporal resolution because of the absence of the space–charge broadening (see Chaps. 2, 5 and 6). These packets are temporally coherent, as mentioned before, and, in contrast to conventional microscopes, their trains are not randomly spaced, or bunched/antibunched, in time (Fig. 5.18). However, the dispersion energy carried by the electrons, relative to the energy of acceleration, determines pulse coherence and duration. Because the energy carried

by photoelectrons (*ca*. 1 eV) is much smaller than the acceleration energy of the microscope (typically, hundreds of keV), and the initiating optical pulse is now typically 100–200 fs long, the broadening of the pulse is still on this hundreds-of-femtoseconds time scale, as demonstrated by temporal image change and electron-energy spectroscopy, which has resolved a temporal shift of 150 fs in UEM (see Chap. 6). To achieve the tens-of-femtoseconds and possibly attosecond electron pulse duration, different schemes for UEM have been reported[17] (see below).

As mentioned above, in contrast to single-electron imaging, single-pulse experiments require a large number of electrons to be delivered to the specimen during a short period of time. For diffraction studies, contributions in pulse generation and characterization have been published from Caltech, Brown University, University of Toronto, Max-Planck-Institut für Quantenoptik in Garching, UCLA, and other laboratories.[17,18] It has already been demonstrated that a ~1 ps pulse could be packed with $\sim 10^4$ electrons,[2,17c] but providing $\sim 10^6$–10^7 electrons in femtosecond pulse duration requires modifications involving compression, reversal of chirp, grating fields, megavolt acceleration, and possibly others. From Eq. (6.3), it follows that the longitudinal broadening Δt_{KE} can be reduced by decreasing d or increasing V.

In developing UEM-1 and UEM-2, and the preceding generations of UEC and UED (generations 1, 2, 3, and now 4), several technical hurdles had to be overcome in order to establish the methodology as a viable approach to 4D imaging. These include the *in situ* determination of the zero of time for clocking of the change, the high-precision frame referencing for the needed sensitivity, and the adjustable timing between the clocking and probing electron pulses in the microscope, with robust electron trajectories throughout the recording. Moreover, in UEM, the entire imaging process cannot exceed seconds (Chap. 6). Electron (and, earlier, photon) interferences had been observed in a double-slit-type experiment, but real-space imaging with timed, single-electron packets was accomplished only recently in UEM (Chap. 5). Other developments in conventional EM, including the important environmental TEM and holography, can now be augmented in UEM. Recent reviews[1c,d] and a patent to Caltech[17e] summarize some details of the approach and applications.

UEM together with its variant techniques (scanning, focusing, filtering, and tomographic capabilities, as well as EELS) provides a universal table-top approach for many investigations. In Chap. 7, we discussed the power of large-scale synchrotron facilities in determining structures of (thick) crystals and species in solution. The X-ray methods of absorption, diffraction, and low-resolution "imaging" are developing well and, depending on the method of generation of X-ray pulses, the time resolution versus flux requirement determines the type of application. In Chap. 7 as well as in a recent article,[16] the methods of electron and X-ray structural determination have been reviewed, and in Fig. 8.9, the current

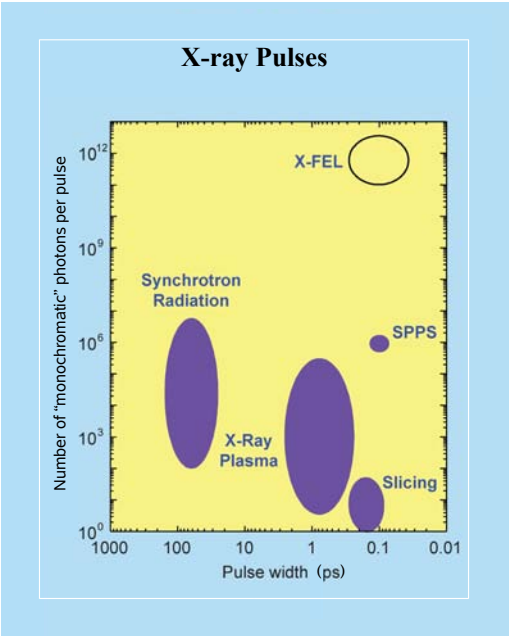


Figure 8.9 Comparison of different sources of hard (2–20 keV) X-ray pulses. Because the pulse is very broad in energy, the number of photons is calculated at the central energy of the pulse for an energy width of 0.1% of the central energy (so-called 0.1% bandwidth, b.w.) for the purpose of comparing different machines. The anticipated free-electron X-FEL characteristics are indicated by an empty ellipsoid (see Ref. 16 and Chap. 7).

status of X-ray pulse generation is displayed as a time versus flux plot.[16] The utilization of X-rays from synchrotrons for the study of protein crystals using the Laue method, and absorption spectroscopy in solution, is still at the forefront of structural determination (Chap. 7).

8.4 Emerging Developments

Building on the concepts, techniques, and applications given in Chaps. 2, 5 and 6, and recalling the numerous static structures examined in Chaps. 3 and 4, in what follows we explore potential areas of studies, which are defined by three major directions of research at Caltech. We also outline the theoretical machinery needed for the determination of structural dynamics and for the visualization of complex systems.

8.4.1 Materials science

From the applications highlighted in Chap. 6 — visualization, from atomic motions (femtosecond, angstrom) to mechanical motion (microsecond, micron) — the kind

of experiment now rendered feasible by 4D UEM should enable unprecedented detail to be retrieved in the study of inorganic and organic materials. For example, the femtosecond-scale heating (or photoexciting) pulse provides for the possibility of studying of a wide range of phase transitions including such phenomena as martensitic transformations,[19] the elementary dynamics of which have hitherto proved inordinately difficult to record directly by other experimental means.

Very recently (in the proof stage), the martensitic phase transition of iron was studied using 4D diffraction and imaging.[20a] One of the most revealing findings is that the observed direct transformation is accomplished through collective atomic motions, but by involving incoherent barrier-crossing processes with dependence on the initial temperature. This transformation in iron on the ns time scale should be contrasted with that of the ultrafast metal-insulator transition in vanadium dioxide (Sec. 6.4.1.1) and for which the initial coherent bond dilation and unit-cell expansion are critical to the subsequent macroscopic phase growth. For solid-state and materials chemists, the phenomenon of diffusionless reactions (such as photostimulated dimerizations and polymerizations)[20b] in organic molecular crystals, with its related strategy of crystal engineering,[20c] should similarly be a rich pasture to explore. Also, with open-structure, single-site heterogeneous catalysts[20d] the breathing of nanopores and nanocages and the amplitudes of atomic motion of the active center grafted onto them, which are crucial determinants in separations based on adsorption and in catalytic performance, should be directly recordable. Because of the high sensitivity of 4D UEM, it should be possible to uncover new facts about the enigmatic catalytic activity of nanoparticle gold[21] (supported on certain oxides).

Bulk gold is catalytically inactive, but nanoparticles of the metal (~5 nm in diameter) supported typically on TiO_2 catalytically oxidize carbon monoxide in air at below room temperature. What are the structural dynamics of the terrace, step, and kink sites of this remarkable material?[21] And what are the ultrastructural characteristics of macromolecular assemblies, nanocomposites, grain boundaries, and quantum dots? Because the cross-section for electron scattering is about six orders of magnitude larger than that of X-ray scattering, the electron imaging of catalysts, with site selectivity (Chap. 6), should be of major interest in further applications of UEM. Other examples are given in the publications by Thomas and by Harris and Thomas,[22] which include photochemical processes in diacyl peroxides and other materials, chemical reactions, and phase transitions (urea inclusion compounds), and crystal nucleation.

The recent success in the study of melting and the embryonic stages of nucleation and crystallization, together with structural and morphological dynamics (see Chap. 6), certainly holds promise as to the elucidation of these and other applications in crystallography and catalysis in particular, and in materials science in general.

8.4.2 Biological UEM

As discussed in earlier chapters, biological imaging basically involves two approaches[23]: electron crystallography, which relies on "redundancy" of the unit-cell repetition in crystals, and imaging of an ensemble of single (isolated) particles. Here, we recall that, by precisely aligning and superimposing the individual images, or by obtaining a tilt series of images of a single particle (electron tomography), "crystallography without crystals" can be achieved. Cryo-electron microscopy (cryo-EM) is an essential part of this endeavor (Chaps. 3 and 4). In the "single-particle-analysis" technique, a purified protein or phage sample is spread into thin films across EM grids, plunged into liquid ethane to immobilize the proteins within vitreous ice, and imaged in an electron microscope. The key advantages of the single-particle cryo-EM-based methods are several. First, because the particles in the samples are imaged one-by-one, in isolation rather than in a crystal, no time-consuming searches for crystallization conditions are necessary; second, the proteins can be imaged in physiologically relevant buffers and concentrations rather than those special conditions that produce crystals; and third, a mere $\sim 10^8$ particles are estimated to be needed to determine the structure.

Despite steady progress, the highest resolution achieved to date by single particle cryo-EM is 4–8 Å, but typical structures are still determined at a resolution of 1–2 nm. The limit on resolution is not fundamental since the current resolution is determined by radiation damage among other factors, such as particle heterogeneity and image contrast. High energy electrons break covalent bonds, deposit thermal energy, and occasionally even dislodge atomic nuclei. Thus, radiation damage limits the useful dose to ~20 $e^-/Å^2$, a flux issue. Because near-atomic resolution reconstructions require pixel sizes of approximately 1 $Å^2$, the quantum noise ("shot noise") present in the images makes it difficult to align the images precisely; misalignment directly degrades reconstruction resolution.

As we saw earlier (Chaps. 3 and 4), a major consequence of radiation damage is the inter- and intramolecular movement of the specimen during the exposure, which causes blurring of images. At about 80 K, proteins in vitreous ice are essentially motionless. Nevertheless, when struck by an electron beam, and covalent bonds are broken, electrons are liberated and heat is deposited in the macromolecule and surrounding ice. Pressure builds up within the ice as radiolytic fragments move (at least a van der Waals radius) apart. Residual positive charge accumulates as secondary electrons (electrons originally present in the sample) are emitted, causing internal repulsions. Consequently, the deteriorating macromolecule may move and rotate appreciably.

If the exposure time is diminished by many orders of magnitude (from $\sim 10^0$ to somewhere between 10^{-6} and 10^{-12} s), beam-induced specimen movement becomes negligible on the time scale of the UEM experiment. Also, with

single-pulse techniques we do expect to be able to effect successful imaging before fragmentations or reactions take place. Whereas reducing the dosage per frame did not prevent the beam-induced movement in paraffin,[24] our recent theoretical work (at Caltech) makes it clear that there exist optimum conditions for pulse duration and dosage. This research on resolution improvement is currently being pursued in a collaborative effort with Grant Jensen and his co-workers.

So far, with cryo-UEM, stroboscopic images of the protein catalase crystals were obtained, resolving the 9.3 nm lattice planes, and bacterial cells (*Caulobacter crescentus*) displaying the membrane structure and flagellum were imaged (Chap. 6). For the purpose of a single-pulse recording, the challenge is in ensuring that the lateral coherence is sufficient not only to see the object as a whole but also the structure within. Because of the cryo-imaging condition, pulses of picosecond duration may be sufficient. For the optimum lateral coherence, the contrast transfer function must persist to permit the survival of high spatial frequencies and at the same time have a significant value for the low frequency ones (Chap. 2). Such coherent properties require studies of focusing and cathode design, and work is in progress to optimize, under different conditions, the phase contrast of images.

UEM also opens up the possibility of recording femtosecond (or longer) time scale dynamics. In this regard, for the biological systems studied using X-ray diffraction[25] and electron crystallography[26] methods, it should be possible to image their structural dynamics in real time. It is also conceivable to obtain time frames of single particles after an excitation pulse warms the sample, excites a conformational change, or releases a photocaged reactant. There are no fundamental physical reasons why multiple electron pulses could not be configured simultaneously to record whole tilt series, producing dynamically-resolved 3D tomographs of particles or cells. The regularity of pulsed dosing may result in the control of energy redistribution and heat dissipation, especially with substrates. Finally, the investigation of redox metalloenzymes by UEM should shed new light on the mode of action because it will be possible to record electron-energy-loss spectra (Chap. 6) and to reveal oxidation states of the metal ions at active centers, even though the measurements may be complicated by the reducing environment created by the secondary electrons within the sample. These and other promising features of UEM were realized since the first UEM publication[27] (Sec. 6.1) and currently represent one of the major areas of study at Caltech.

8.4.3 Structural dynamics: Theory and experiment

Perhaps one of the most important messages of this book is the strong interplay between theoretical concepts and experimental findings. There should not be a "gulf" between the two if theory is to impact upon experimental progress and

if experiments are so designed as to search for new phenomena and discoveries. For example, in the subsequent section, the vectorial properties associated with molecular structure and rotational dynamics, which add another dimension to 4D electron imaging, are shown to arise through the imposition of spatial order on otherwise randomly oriented ensembles, or through the initiation of a particular material change along a unique crystallographic direction. A theoretical examination of the time dependency reveals that the isotropic (scalar) contribution of scattering depends only on population dynamics and internuclear separations, whereas higher-moment, vectorial components of the distribution evolve in time due to rotational motion and decoherence. In these studies, for both spectroscopy and diffraction, theoretical and experimental advances are interconnected. Similarly, for more complex systems where structures could exist in a nonequilibrium state, experimental observations together with numerical diagonalizations of molecular Hamiltonians and Monte Carlo methods have elucidated possible conformational dynamics.[28]

For crystals and interfaces, it is important to identify the nature of the microscopic forces that trigger dynamical change, and theoretical models are essential in this regard. For example, a recent generic theoretical study of nonequilibrium structural dynamics for a quasi-1D, anharmonically-coupled substrate–adsorbate lattice elucidated the importance of coherent wave propagation in the nonequilibrium regime and provided the atomic-scale description of both surface and bulk dynamics (Chap. 6). Another example is that of a model describing nanometer-scale crystallites that undergo a change, such as a phase transition. One has to consider the time scale of connectivity in the material and the density of carriers involved (Chap. 6). Lastly, as the time scale becomes shorter, and into the attosecond regime (see below), the electron density distribution in space $\rho(\boldsymbol{r})$, $\boldsymbol{r} = (x, y, z)$, which determines both the spatial structure and the structural change of a molecule, crystal, or biological assembly, should be examined theoretically and experimentally, as temporal snapshots of $\rho(\boldsymbol{r})$ provide a quantitative topological insight into the dynamics.

With the increased complexity of systems studied, a significant effort was devoted at Caltech to developing the computational machinery required for the frame-by-frame deciphering of the ultrafast structural changes recorded in UED, UEC, and UEM experiments. Due to the complex nature of the ephemeral phenomena under study, experiment and theory often have to go hand in hand in unraveling details of the microscopic forces involved. For structural determination, the initial (trial) structure is usually obtained from theoretical calculations. For example, in the UED of isotropic gases, a 3D molecular structure is projected onto a 1D momentum transfer space as a superposition of damped oscillations, each of which represents a particular internuclear separation r_{ij} (Chap. 2). A molecular model based on some estimated structure parameters and vibrational correction

terms is then constructed and further refined using a least-squares fitting procedure until a reasonable agreement between the experimental and calculated scattering intensities is achieved.[28a] Different algorithms can be used to refine the structure for each particular point in time (temporal frame).[1c]

In imaging, thousands of frames are involved in making movies of the actual overall structural change. Similarly, for macromolecules there exist a myriad conformations and populations. As a guide for experimental design and development of sound structural models, as well as for handling of large volumes of data, supercomputer facilities are essential. With the Linux-based supercomputer cluster at Caltech, which currently features 32 dual quad-core Intel E5345 compute nodes, 12 GB RAM/node, ~20 TB of network-attached disk storage, and a 1-GigE interconnection mesh, ultrafast nonequilibrium dynamics of complex energy landscapes were modeled for processes involving protein and DNA (un)folding. With the advent of such numerical experiments (see below), the explicit dynamical trajectories of the "specimen" can be pictured with an atomic resolution, and the ensemble convergence can be achieved at increasingly longer time scales.

Even in the case of a relatively small macromolecule, such as a DNA hairpin, the complexity of the energy landscape demands these new computational tools. Stimulated by recent observations of denatured intermediate states for a 5′-CCCCTT-X_{13}-AAGGGG-3′ DNA hairpin at temperatures higher than that of the melting point, as evidenced by absorption/fluorescence-quenching experiments,[29] ensemble-convergent MD simulations were performed on a similar, benchmark 5′-ATCCTA-X_4-TAGGAT-3′ DNA hairpin. But, concurrently, we developed a simple *kinetic intermediate structure* (KIS) *model* of DNA unzipping (Fig. 8.10) based on tabulated pairing-stacking thermodynamic parameters and loop entropy.[30a] Specifically, the model elucidates the key sequence and loop-dependent behavior of unzipping as well as the identity of intermediate states of unzipping kinetics. After verifying the assumptions and predictions of the model via MD simulations on the benchmark hairpin, the model was applied as a function of stem sequence and loop length permutations of the 5′-CCCCTT-X_{13}-AAGGGG-3′ hairpin so as to study the entropic and enthalpic factors that determine hairpin stability in aqueous solution. For a wide range of such permutations of the hairpin, it was important to determine the temperature range for which the two-state-unfolding hypothesis breaks down as well as the base-pairing configuration of the intermediate state.[30a] Similar MD studies were carried out for elucidating the temporal behavior during T-jump-induced helix-coil transitions of polypeptides in aqueous solution.[30b]

Of special interest to us are the manifestations of biological complexity in diffraction, first *in vacuo* and then after incorporation of the effect of hydration. Here again, we embarked on both the analytical theoretical approach and numerical experiments, especially when considering the myriad of structural configurations and the unique features of diffraction. This complexity could, naively, suggest the

masking of any significant change in diffraction. However, an accurate theoretical mapping of order–disorder transitions in large molecular ensembles indicated that the problem of tracking down the disruption of the structural ordering in space and time may be experimentally tractable due to the unique coherence in scattering (*spatial resonance*) associated with the quasi-periodic secondary structure in the macromolecule.[31] Although biological macromolecules often tend to undergo local distortions, the structures they commonly assume are relatively robust, persisting in a wide variety of environments. Thus, according to electrospray mass-spectrometric observations, DNA retains its major structural features even in the absence of the hydrating water layer (see, for example, Ref. 32).

Very recently, the structural dynamics of DNA unfolding *in vacuo* — as obtained from ensemble-convergent MD simulations carried out for a number of protonation states and temperature jumps — was reported along with a comparison of the unfolding dynamics of the same macromolecule in aqueous solution.[31a] In

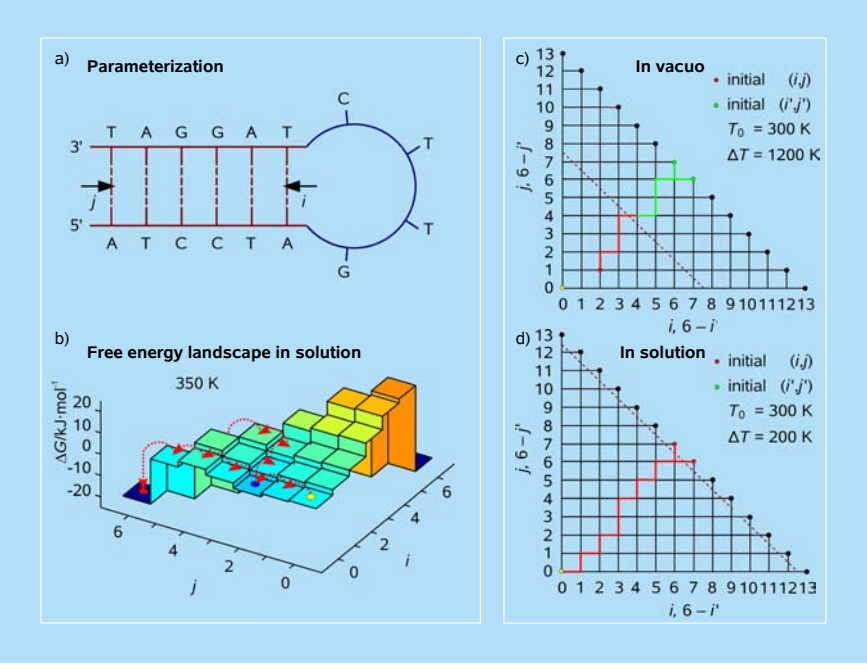


Figure 8.10 Kinetic intermediate structure (KIS) model. (a,b) Model parametrization, free energy landscape, and unfolding trajectory for a small DNA hairpin.[30a] (c,d) Unfolding trajectories of 5′-d(CGCGGTGTCCGCG)-3′ DNA duplexes *in vacuo* and in aqueous solution. In the gas phase the helices are, on average, partially unwound at $t = 0$. Because of the formation of internal bubbles at (i', j'), with the origin defined to be the center of the helix, the unfolding trajectories bifurcate into (i, j)- and (i', j')-related pathways which are associated with unzipping from the ends (red solid line) and from the middle of the helix (green solid line), respectively. The barrier to the unfolding transition is represented by a red dashed line. In the presence of water, continuous unzipping from the ends of the duplex is favored against internal bulge formation.[31a]

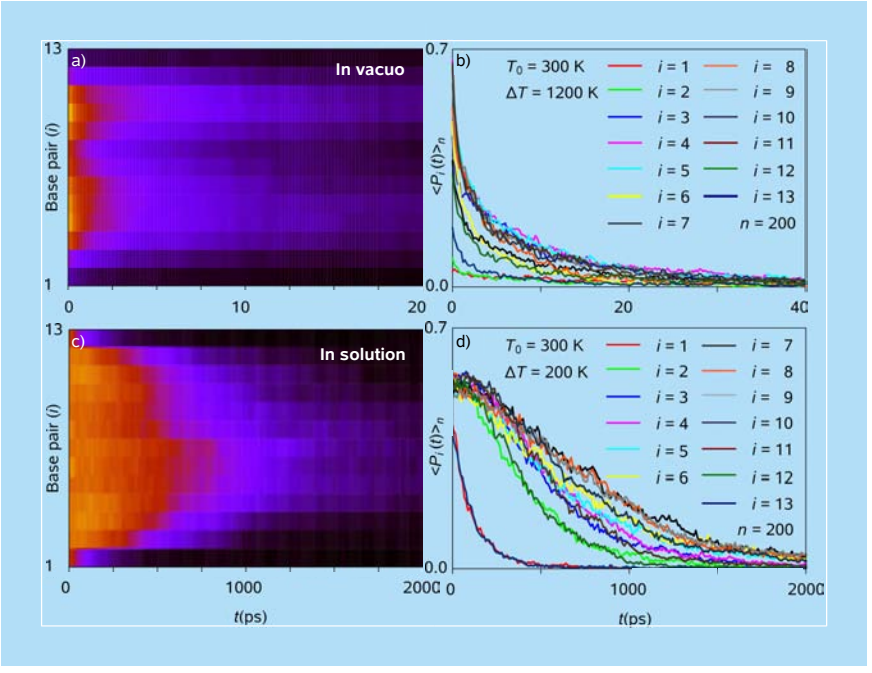


Figure 8.11 Base-pair separation time scales *in vacuo* and in aqueous solution for DNA duplexes of Figs. 8.10c,d. Shown is the temporal evolution of the native base pairing in macromolecular ensembles upon (a,b) 1200 K and (c,d) 200 K temperature jumps. The initial temperature T_0 of the ensembles was equal to 300 K. For each Watson–Crick base-pairing contact i, $1 \le i \le 13$, the probability density of finding the contact intact at time t upon the temperature jump ($\langle P_i(t) \rangle_n$; averaged over $n = 200$ independent MD trajectories) is plotted as (a,c) a color map (yellow: 0.7, black: 0.0); and (b,d) a one-dimensional profile. Note the formation of internal bubbles characteristic of unfolding *in vacuo* as opposed to aqueous solution.[31a]

addition to the use of native-contact KIS metrics to parametrize the (local) base-pairing disruption (Fig. 8.10), we also elucidated the (global) conformational dynamics of the duplex via the radial distribution functions of UED simulations — which were invoked as an intuitively appealing coarse-graining approach. The simulations revealed that, though ensemble-averaged unfolding trajectories of free DNA were largely insensitive to both protonation-state and temperature changes, the denaturation dynamics *in vacuo* was vastly different from that found in aqueous solution (Fig. 8.11). Strikingly, the "stickiness" (or self-bonding bias) characteristic of free DNA macromolecules makes order–disorder transitions in the gas phase about two orders of magnitude slower than in solution, which is caused by repetitive bonding recombination of broken Watson–Crick base pairs in the absence of water. It is also noteworthy that, in the absence of surrounding water, the DNA duplexes tend to unwind from both the interior and the ends (Figs. 8.10 and 8.11). An overview of DNA unfolding dynamics *in vacuo* as obtained from the ensemble-convergent MD simulations is presented in Fig. 8.12,

and further studies are currently underway to unravel the intermediate structures and dynamics pertinent to protein denaturation and cognate phenomena.

The approach is applicable to other systems and to experiments of UED, UEC, and UEM. The ensemble-convergent MD simulations, when aided with adequate coarse-graining approaches which permit one to explore the key features of the dynamical landscape without sacrificing important structural details, provide unprecedented insights into both structure and dynamics of macromolecular ensembles. As with experiments[33] designed to probe the spatial and temporal coherences of the specimen, the numerical experiments highlighted in this section make use of both spatial resonance *and* ensemble averaging to extract accurate and reliable structural data at each particular point in time.

A reminder regarding large-scale computations in relation to imaging can now be made: because electrons have a large cross-section of interaction with matter, it is clear that, for electron-scattering processes, the kinematic approximation for intensity calculations breaks down in some cases (see Chaps. 3 and 4). Under these conditions, a full dynamical theory is needed in order to reproduce the observed

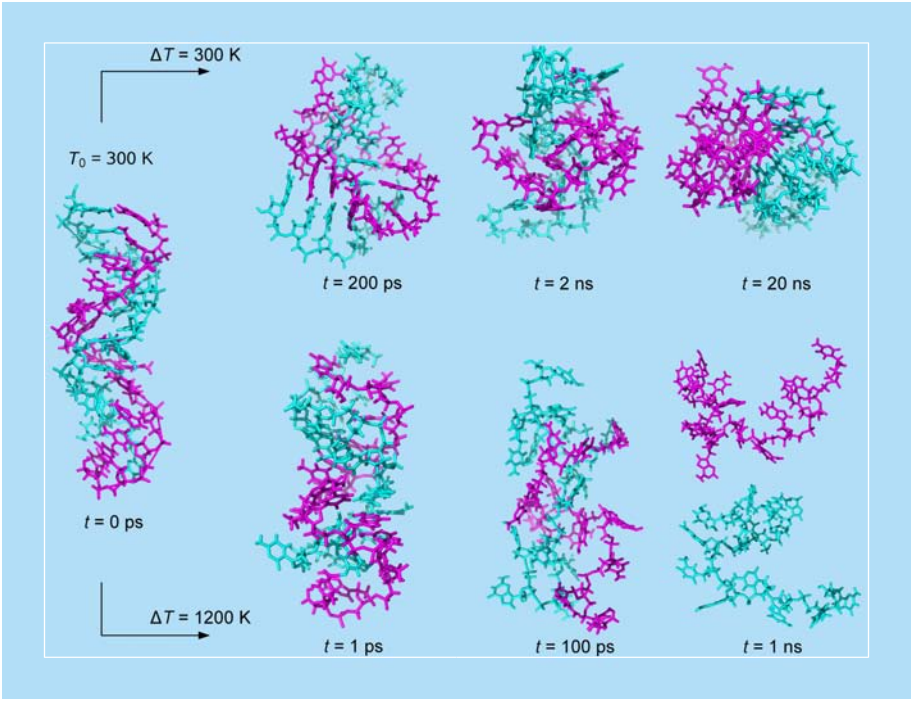


Figure 8.12 DNA unfolding dynamics: characteristic structures *in vacuo*. Shown are macromolecular structures representative of different stages of DNA order–disorder transitions as obtained from MD simulations for 300 K and 1200 K temperature jumps. Note the extended helical structure at $T_0 = 300$ K and formation of a "blob" for $\Delta T = 300$ K.[31a]

diffraction intensities. It would be of interest to examine the theory as the time of observation in imaging reaches the (sub)femtosecond domain.

8.4.4 Aligned- and single-molecule imaging

Many chemical, physical and biological properties of molecules are inherently dependent on spatial orientation, and therefore the degree of orientation, as well as its temporal evolution, significantly affects studies of molecular systems in all physical phases, from dilute gases to solids. The type of orientational motion encountered varies from generally negligible in solids, to diffusive in liquids, to hindered in dense gases, to purely inertial and coherent in isolated molecules. For decades we have been concerned with some basic concepts of molecular orientational dynamics, with an emphasis on their role in femtosecond and picosecond spectroscopic studies and their relationship to molecular structures in diffraction.[34] The femtosecond time scale is unique in that the orientation of molecular samples, even when undergoing free rotation, can be followed as it evolves, providing direct information on the nature of the motion and the forces that govern it.

With these techniques, known as *rotational coherence spectroscopy* (RCS), a variety of molecular structures were determined from the knowledge of moments of inertia (rotational constants) directly available from the measured periods of recurrences.[35] Besides these structural features, the temporal decay (dephasing) of rotational anisotropy allows for studies of phenomena associated with free rotational motion of isolated molecules; orientation in the presence of collisions and the effect of solvent density in the transition from inertial to diffusive motion, spanning gases and liquids; orientation in reactions; and some effects of restricted geometry in biological systems.[34,35] The classical and quantum origins of orientational dynamics and the differences to be expected between experiments using femtosecond pulses and those using nanosecond or longer pulses (continuous wave excitation) were part of these studies.

When an isotropic sample is excited by a linearly polarized pump pulse via an electric-dipole-allowed vibronic transition, a spatial alignment of the excited-state population is created at the instant of excitation ($t = 0$; Fig. 8.13). The probability for absorption depends upon the relative orientation of the polarization vector and the molecular transition dipole. The excited ensemble is created with its transition dipoles exhibiting a cosine-squared distribution around the polarization vector (Fig. 8.13). Subsequent rotational motion causes an evolution of that distribution, as represented in the figure. Recurrences in the orientation are understandable classically only for molecules at a fixed angular momentum, but the recurrences shown in the figure at the indicated times occur in thermal samples as a quantum effect.

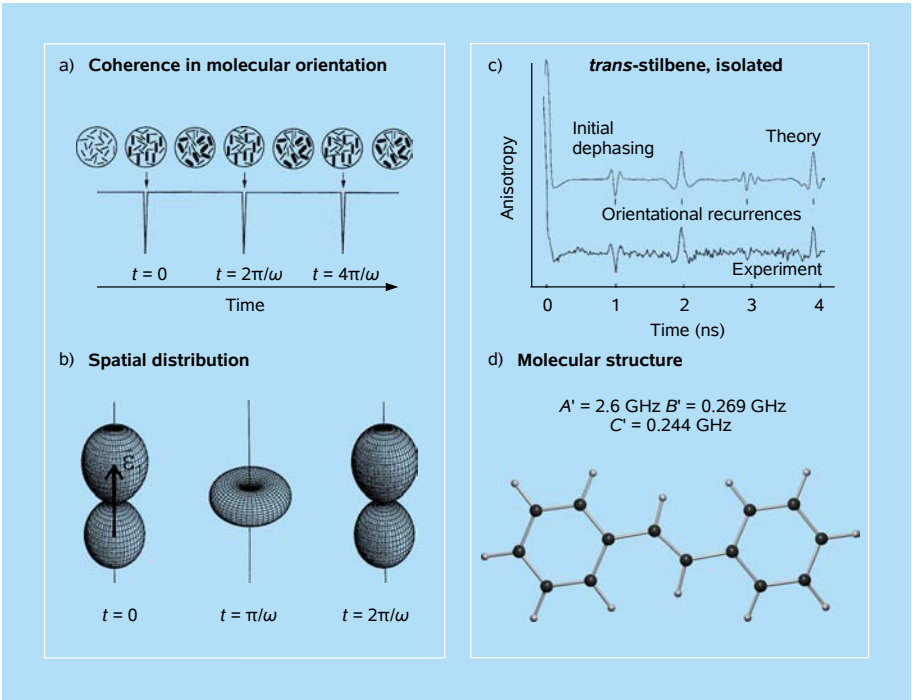


Figure 8.13 The concept of free molecular alignment with ultrafast lasers, and experimental observations. (a) An isotropic distribution of dipoles is excited at $t = 0$ by a vertically polarized light pulse. Thicker lines represent transition dipoles of excited molecules, which follow individual rotational trajectories. (b) Spatial distributions of transition dipole orientations for the excited-state population at selected times for a simple rotor with a parallel dipole. The pump polarization vector is shown. (c) Comparison of experimental and theoretical anisotropies of jet-cooled samples of *trans*-stilbene in a rare gas. (d) Molecular structure of *trans*-stilbene as obtained from rotational-coherence measurements.[34]

Two significant observations are obtainable from RCS: the recurrences which provide the information for structure determination and the loss of phase coherence (dephasing) — which gives the effective temperature and the extent of free rotation (free molecules versus molecules in collision with the solvent medium). In Fig. 8.13, we give an example of such studies for the first molecule (*trans*-stilbene) that we used in determining the structure and dephasing with RCS. The same methodology was applied to dense fluids, chemical reactions (uni- and bimolecular), and to biological systems.[34]

In 1994, Williamson *et al.*[36] considered the diffraction of aligned molecules, and showed that for a random (isotropic) distribution, which gives rise to Debye diffraction rings, the alignment in the field of laser radiation provides an additional dimension of imaging — in a sense, Debye's ring pattern begins to approach the diffraction from a crystalline specimen. This is optimally manifested when ultrafast pulses are used to induce the coherence of the aligned molecules which is well

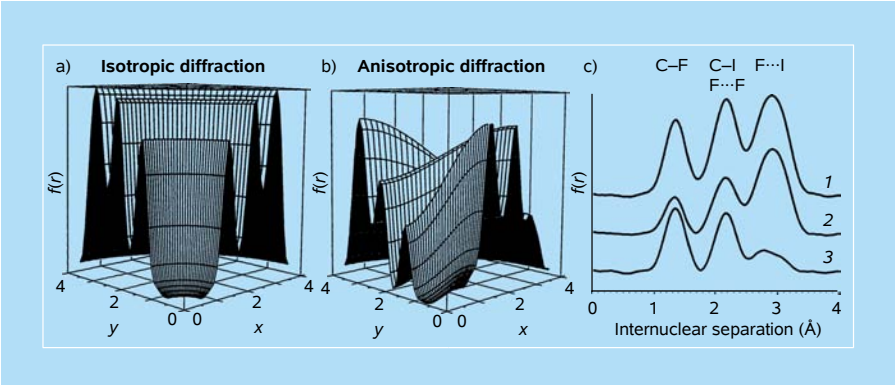


Figure 8.14 Alignment in diffraction: two-dimensional radial distribution curves for CF_3I. (a) Isotropic sample. The inner ring corresponds to the C–F separation, the middle ring is a sum of C–I and F···F separations, and the outer ring corresponds to the F···I separation. Only one quadrant is shown. (b) Perpendicular/unexcited subset of the molecular sample after laser excitation along the C–I bond axis. Only one quadrant is shown. The inner ring is most intense along the *x*-axis because ***r***(C–F) is almost perpendicular to ***r***(C–I). The outer ring is most intense along the *y*-axis because ***r***(F···I) is almost parallel to ***r***(C–I). The two components of the middle ring have been separated along the axes because ***r***(F···F) ⊥ ***r***(C–I). Scattering from ***r***(C–I) is found along the *y*-axis. This peak is less intense than the ***r***(F···F) peak along the *x*-axis because scattering from ***r***(C–I) is damped. (c) Radial distribution profiles *1*, and *2*, *3* represent isotropic CF_3I and *y*-, *x*-perpendicular/unexcited CF_3I subsets, respectively.[36]

defined by the structural parameters discussed above. The induced order provides a detailed (3D) structural picture since bond angles can then be obtained, instead of being modeled. Figure 8.14 depicts some of the results reported, along with the manifestations of 3D structure in patterns of aligned molecular ensembles, which should be involved in future UED studies.

Baskin *et al.*,[37] in generalizing the theoretical treatment to include arbitrary polarization directions and molecular distributions, showed that the scattering can be rigorously separated into isotropic (population-specific) and angular (orientation-specific) contributions. This separation allows for the extraction of the structural evolution and bond population dynamics without consideration of the rotational dynamics of the sample, in a manner reminiscent of the use of "magic angle" detection in spectroscopy. Such capability, which was demonstrated for a five-atom chemical reaction, is of significance in the studies of complex reactions and conformational changes, given the dependence of the rotational motion of products on a range of structural and energetic factors which are inherently difficult to fully characterize. UED of the elimination reaction highlighted in Fig. 6.11 is an example of such studies of structure and population dynamics,[2] and of alignment.[38] For crystals, the polarization effect in diffraction is of special value in revealing the directionality of specific couplings involved, as shown in the study of, for example, correlated systems in the condensed phase.[39]

For solvated molecules, Spence and Doak[40] suggested a single-molecule diffraction methodology for specimens that are difficult to crystallize, such as a small hydrated protein in a cold jet of doped liquid-helium droplets. Here, a continuous high-energy electron beam is used to obtain the diffraction and the alignment is induced using an elliptically polarized laser beam. The (nonresonant) laser field "fixes" the orientation of passing molecules (~60 percent of the molecules can have their axes located within a cone of 30°, compared with ~3 percent for random orientations) and a large number of 2D diffraction patterns can, in principle, be added together from different molecules of similar orientation. The resolution becomes simply the product of molecular length and the angle.

From the above studies, it is clear that a fruitful direction for research would be to include alignment as a method in the study of diffraction and imaging.

8.4.5 Imaging with attosecond electrons

We have already discussed techniques for generating electron and optical pulses, and the possible extensions of these techniques to the attosecond regime. Here, we highlight one of these approaches, namely, the "temporal-lens" concept for sub-femtosecond electron imaging,[17a] which is now being implemented in collaboration with Herman Batelaan and his co-workers at the University of Nebraska.[17f]

A central requisite for reaching this degree of temporal resolution with electron imaging and diffraction is the generation of attosecond electron pulses in "free space," so that diffraction from freely chosen specimens of interest can take place independently of the mechanisms of pulse generation. As discussed previously, electrons with energies of 30–300 keV are ideal for imaging and diffraction because of their high scattering cross-sections, convenient diffraction angles, and the appropriate de Broglie wavelength (0.02–0.07 Å) to resolve atomic-scale changes. Moreover, they have a high degree of monochromaticity. For example, electrons accelerated to $E_0 = 30$–300 keV with pulse duration of 20 as (bandwidth of $\Delta E \approx 30$ eV) have $\Delta E/E_0 \approx 10^{-3}$–$10^{-4}$, making diffraction and imaging possible without a spread in angle and resolution. Optical attosecond pulses have typically $\Delta E/E_0 \approx 0.5$ and, because of this close proximity of ΔE and E_0, their duration is Fourier-limited to ~100 as. Free electron pulses of kiloelectron-volt central energy can, in principle, have much shorter duration, down to sub-attoseconds, while still consisting of many wave cycles.[17a]

As emphasized earlier, pulses with a large number of electrons suffer from the effect of space–charge, which deteriorates both the spatial and the temporal resolution. This can be avoided by using packets of single, or only a few, electrons in a high repetition rate, as discussed above. (Figure 5.15 depicts the relation of single electron packets to the effective envelope due to statistics.) A second obstacle is the high dispersion for electrons of nonrelativistic energy. The small but

unavoidable bandwidth of an attosecond electron pulse causes the pulse to disperse during propagation in free space, even when no space charge forces are present. For example, a 20 as pulse with $\Delta E/E_0 \approx 10^{-3}$ would stretch to picoseconds after just a few centimeters of propagation.

The approach for the suppression of dispersion and the generation of free attosecond electron pulses is based on the initial preparation of negatively chirped electron packets.[17a] Femtosecond electron pulses are generated by photoemission and accelerated to kiloelectron-volt energies in a static electric field. Preceding the experimental interaction region, optical fields are used to generate electron packets with a velocity distribution, such that the higher-energy parts are located behind the lower-energy ones. With a proper adjustment of this chirp, the pulse then self-compresses to an extremely short duration while propagating towards the point of diffraction. To achieve attosecond pulses, the chirp must be imprinted to the electron pulse on a nanometer length scale. Optical waves provide such fields.

Because nonrelativistic electrons move significantly slower than the speed of light (e.g. $\sim 0.3c$ for 30 keV), the direct interaction with an optical field will, therefore, cancel out over time and cannot be used to accelerate and decelerate electrons for compression. In order to overcome this limitation, use was made of the ponderomotive force, which is proportional to the gradient of the optical intensity to accelerate electrons out of regions with high intensity. By optical wave synthesis, intensity profiles can be made to propagate with less than the speed of light and, therefore, allow for co-propagation with the electrons.

Figure 8.15 depicts a schematic of the concept (see also Chap. 5). A synthesized optical field of two counter-propagating waves of different wavelengths

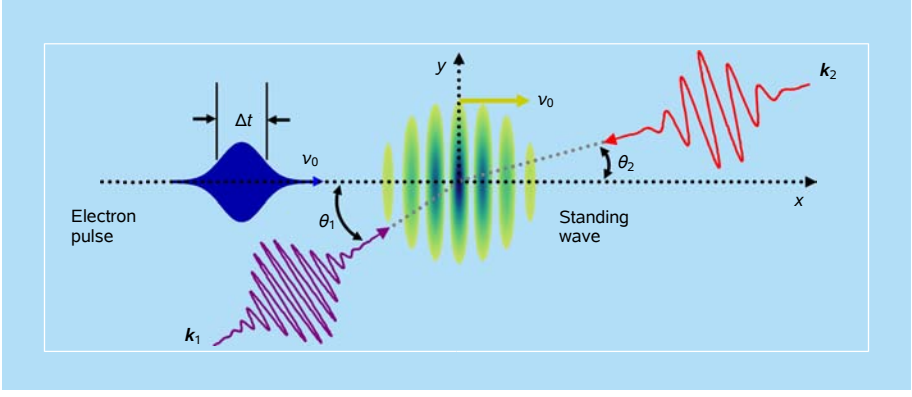


Figure 8.15 Two-dimensional schematic of the tilted laser pulse concept for a standing wave used for attosecond electron compression (thick temporal lens). When $\theta_1 = \theta_2 = 0$, then $v_0 = c/3$ for ω and 2ω beams.[17a] With frequencies of ω and 2ω the velocity of the standing wave can be tuned by tilting the two laser pulse angles θ_1 and θ_2. When the two angles are chosen according to $\theta_2 = \arcsin(2\sin\theta_1)$, the wells in the standing wave are perpendicular to the electron propagation.[17f,g]

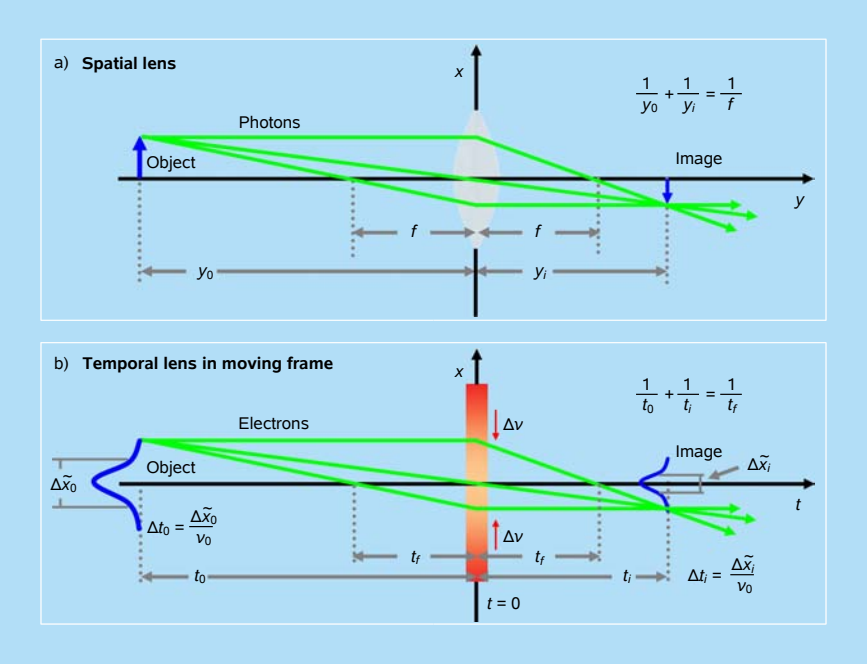


Figure 8.16 Ray diagrams for (a) spatial and (b) temporal lenses. The top panel depicts three primary rays for an optical thin spatial lens. The object is located at y_0, and the spatial lens has a focal length, *f*. A real image of the object is created at the image plane, position y_i. The bottom panel is a ray diagram for a temporal thin lens. The diagram is drawn in a frame moving with the average speed v_0 of the electron packet. The slopes of the different rays in the temporal diagram correspond to different initial velocities that are present in the electron packet. As shown in the diagram, a temporal image of the original electron packet is created at the image time t_i. The initial packet (object) is created at a time t_0 with $\Delta t_0 = \Delta\tilde{x}_0/v_0$, where the spatial extent of the pulse is directly related to the temporal duration of the object. The lens is pulsed on at $t = 0$ and the temporal focal length of the lens is t_f. The lens represents the ponderomotive potential and in this case is on for the very short time τ.[17f]

results in an effective intensity grating, similar to a standing wave, which moves with a speed slower than the speed of light. Electrons can, therefore, co-propagate with a matched speed and are accelerated or decelerated by the ponderomotive force according to their position within the wave. After the optical fields have faded away, this velocity distribution results in self-compression; the attosecond pulses are formed in free space. Depending on the optical pulse intensity, the electron pulse duration can be made as short as 15 as,[17a] and, in principle, shorter durations are achievable. If the longitudinal spatial width of the initial electron pulse is longer than the wavelength of the intensity grating, multiple attosecond pulses emerge that are located with well-defined spacing at the optical minima. This concept of compression, which was tested by simulation,[17a] can be rigorously described analytically as a "temporal-lens effect"[17f] (see Fig. 8.16, with relation to

other schemes of electron pulse shortening and to the Kapitza–Dirac effect).[41] The concept was generalized to address noncollinear geometries, space–charge effects, and space–time phase matching with tilted attosecond electron pulses.[17g]

With this methodology, two types of applications were considered,[17g] one that is concerned with changes in chemical bonding, which is reminiscent of the VO_2 phase transition study highlighted in Chap. 6, and another which maps an electron distribution oscillation in an electric field. The results suggest that 4D imaging methods with diffraction have the potential of revealing electron dynamics on the attosecond time scale. However, given the proposed pump-probe experiments with free attosecond electrons, the issue of the pump characteristics must be considered. In femtoscience, especially femtochemistry and femtobiology, the pump-pulse energy width is a small fraction of the dissociation energy of molecules and the dynamics is that of a nuclear motion in a given force field (electronic state).[14] For the attosecond regime, because of the uncertainty in energy (> 10 eV), such conditions are not met.

Three different avenues of pump-probe studies were suggested.[17g] The first is to use soft X-ray pump pulses that coherently cover multiple electronic states to form an electronic wave packet, the analogue of a nuclear packet. Charge density of "some superposition" will rapidly oscillate in a frozen nuclear framework. Because diffraction is insensitive to the resolution of states, it can follow the dynamics of possible electron–electron interactions and transient repulsion effects. The second approach would be to excite a well-defined state with a few-femtosecond pulse duration and then map the electronic distribution changes on this time scale while the nuclear motion is evolving. The third prospect, as outlined above, is the study of "off-resonance" effects due to electron displacements and oscillations, which are responsible for physical properties such as the refractive index and nonlinear optical effects. In all such scenarios, diffraction with free electron packets allows for 4D imaging of charge density, independent of the mechanism of pulse generation.

In the regime of stroboscopic recording with "thin" and "thick" temporal lenses,[17f] the number of electrons cannot be very high. Recently, however, it was proposed that tens-of-femtosecond electron pulses[17d] could be sustained through the use of chirp in combination with energy filtering. Space–charge forces in multi-electron pulses accelerate leading electrons and retard trailing ones, thus inducing correlations of momentum and time. By resolving the diffraction images with an energy analyzer, well-defined temporal slices of the long electron packet can be selected. Numerical simulations showed that conventional electron sources are sufficient to reach the 30 fs domain of resolution without electron pulse compression. They also reveal the influence of pulse shape, electron density, and energy bandwidth on the achievable time resolution (see Fig. 5.28).

8.5 Epilogue

In the course of this monograph, we have discussed the major developments consequential upon the invention of the electron microscope. We have identified the milestones reached from the early days of the birth of 2D and 3D electron microscopy, and traced the progress that has made possible the visualization of atoms, high-resolution charge density maps, complex materials, hydrated proteins, and single-particle-ribosome structures, viruses, and cells (see previous chapters and Fig. 1.7).[42–45] With the development of 4D UEM, which introduces the fourth dimension of time into static electron microscopy, and by demonstrating the power of variant techniques (convergent-beam, near-field, FEELS, cryo-UEM, etc.) and elucidating different applications of the methodology (see Chaps. 6 and 8, and Fig. 1.8), this book will hopefully stimulate research in many disciplines, including extensions into the domains of holography, tomography, environmental and photoemission electron microscopy. The speed of imaging is now nearly 10 orders of magnitude as fast as that of conventional microscopes, thus enabling the charting of the elementary processes involved in the function of complex materials structures as they unfold in real time. Moreover, there exist other emerging developments, including those of biological UEM. Over the past 70 years of developments, since its invention, the table-top approach of electron microscopy has been unmatched in impacting various fields of science, and with the fourth (temporal) dimension added, the future should be as rich as the past!

References

1. (a) A. H. Zewail, in "*Physical Biology: From Atoms to Medicine*," Ed. A. H. Zewail, Imperial College Press, London, p. 23 (2008), and references therein.
   (b) A. H. Zewail, in "*Visions of Discovery: New Light on Physics, Cosmology, and Consciousness*," Eds. R. Y. Chiao, M. L. Cohen, A. J. Leggett, W. D. Phillips and C. L. Harper, Jr., Cambridge University Press, Cambridge, in press, and references therein.
   (c) D. Shorokhov and A. H. Zewail, *Phys. Chem. Chem. Phys.*, **10**, 2879 (2008), and references therein.
   (d) A. H. Zewail, *Annu. Rev. Phys. Chem.*, **57**, 65 (2006), and references therein.
   (e) A. H. Zewail, *Chem. Phys. Chem.*, **10**, 23 (2009).
2. H. Ihee, V. A. Lobastov, U. M. Gomez, B. M. Goodson, R. Srinivasan, C.-Y. Ruan and A. H. Zewail, *Science*, **291**, 458 (2001).
3. R. Srinivasan, J. S. Feenstra, S. T. Park, S. Xu and A. H. Zewail, *Science*, **307**, 558 (2005).
4. Q. Wang, R. W. Schoenlein, L. A. Peteanu, R. A. Mathies and C. V. Shank, *Science*, **266**, 422 (1994).
5. S. Pedersen, L. Bañares and A. H. Zewail, *J. Chem. Phys.*, **97**, 8801 (1992).
6. (a) E. W.-G. Diau, J. L. Herek, Z. H. Kim and A. H. Zewail, *Science*, **279**, 847 (1998).

(b) Y.-C. Cheng and G. R. Fleming, *Annu. Rev. Phys. Chem*., **60**, 241 (2009).
7. (a) G. M. Whitesides, P. W. Snyder, D. T. Moustakas and K. A. Mirica, in "*Physical Biology: From Atoms to Medicine*," Ed. A. H. Zewail, Imperial College Press, London, p. 189 (2008).
(b) P. Ball, *Chem. Rev.*, **108**, 74 (2008).
8. (a) C. M. Dobson, in "*Physical Biology: From Atoms to Medicine*," Ed. A. H. Zewail, Imperial College Press, London, p. 289 (2008).
(b) C. M. Dobson, *Nature*, **426**, 884 (2003).
9. (a) "*Science of Microscopy*," Eds. P. W. Hawkes and J. C. H. Spence, Springer, New York (2007), and references therein.
(b) M. A. O'Keefe, *Ultramicroscopy*, **108**, 196 (2008), and references therein.
(c) A. V. Crewe, J. Wall and J. Langmore, *Science*, **168**, 1338 (1970).
(d) J. M. Thomas, *Nature*, **281**, 523 (1979).
(e) T. T. Tsong, *Phys. Today*, **59**, 31 (2006).
(f) P. E. Batson, N. Dellby and O. L. Krivanek, *Nature*, **418**, 617 (2002).
10. J. M. Thomas, in "*Physical Biology: From Atoms to Medicine*," Ed. A. H. Zewail, Imperial College Press, London, p. 51 (2008), and references therein.
11. O. Bostanjoglo, *Adv. Imaging Elecron Phys*., **121**, 1 (2002).
12. (a) T. LaGrange, G. H. Campbell, B. W. Reed, M. Taheri, J. B. Pesavento, J. S. Kim and N. D. Browning, *Ultramicroscopy*, **108**, 1441 (2008), and references therein.
(b) J. S. Kim, T. LaGrange, B. W. Reed, M. L. Taheri, M. R. Armstrong, W. E. King, N. D. Browning and G. H. Campbell, *Science*, **321**, 1472 (2008).
13. O.-H. Kwon, B. Barwick, H. S. Park, J. S. Baskin and A. H. Zewail, *Proc. Natl. Acad. Sci. USA*, **105**, 8519 (2008).
14. A. H. Zewail, in "*Les Prix Nobel: The Nobel Prizes 1999*," Ed. T. Frängsmyr, Almqvist & Wiksell, Stockholm, p. 110 (2000), and references therein.
15. (a) N. Bloembergen, *Rev. Mod. Phys*., **71**, S283 (1999).
(b) N. Bloembergen, in "*Progress in Optics*," Vol. 50, Ed. E. Wolf, Elsevier, Amsterdam, p. 1 (2007).
(c) F. Krausz and M. Ivanov, *Rev. Mod. Phys*., **81**, 163 (2009), and references therein.
(d) C. Altucci and D. Paparo, in "*Developments in Surface Contamination and Cleaning: Fundamentals and Applied Aspects*," Eds. R. Kohli and K. L. Mittal, William Andrew, Norwich, p. 25 (2008).
16. M. Chergui and A. H. Zewail, *Chem. Phys. Chem*., **10**, 28 (2009), and references therein.
17. (a) P. Baum and A. H. Zewail, *Proc. Natl. Acad. Sci. USA*, **104**, 18409 (2007).
(b) P. Baum and A. H. Zewail, *Proc. Natl. Acad. Sci. USA*, **103**, 16105 (2006).
(c) A. Gahlmann, S. T. Park and A. H. Zewail, *Phys. Chem. Chem. Phys*., **10**, 2894 (2008).
(d) P. Baum and A. H. Zewail, *Chem. Phys. Lett*., **462**, 14 (2008).
(e) A. H. Zewail and V. A. Lobastov, *US Pat*., 7 154 091 (2006).
(f) S. A. Hilbert, C. Uiterwaal, B. Barwick, H. Batelaan and A. H. Zewail, *Proc. Natl. Acad. Sci. USA*, **106**, 10558 (2009).
(g) P. Baum and A. H. Zewail, *Chem. Phys*., doi:10.1016/j.chemphys.2009.07.013 (2009), in press.

18. (a) J. B. Hastings, F. M. Rudakov, D. H. Dowell, J. F. Schmerge, J. D. Cardoza, J. M. Castro, S. M. Gierman, H. Loos and P. M. Weber, *Appl. Phys. Lett*., **89**, 184109 (2006).
(b) C. T. Hebeisen, G. Sciaini, M. Harb, R. Ernstorfer, T. Dartigalongue, S. G. Kruglik and R. J. D. Miller, *Opt. Express*, **16**, 3334 (2008).
(c) M. Harb, R. Ernstorfer, T. Dartigalongue, C. T. Hebeisen, R. E. Jordan and R. J. D. Miller, *J. Phys. Chem. B*, **110**, 25308 (2006).
(d) X. J. Wang, D. Xiang, T. K. Kim and H. Ihee, *J. Korean Phys. Soc*., **48**, 390 (2006).
(e) P. Musumeci and J. Moody, *Proc. 22-nd PAC Conf*., June 25–29, 2007, Albuquerque, NM, USA, p. 2751.
(f) L. Veisz, G. Kurkin, K. Chernov, V. Tarnetsky, A. Apolonski, F. Krausz and E. Fill, *New J. Phys*., **9**, 451 (2007); see also: E. Fill, L. Veisz, A. Apolonski and F. Krausz, *ibid*. **8**, 272 (2006).
19. (a) J. W. Christian, "*The Theory of Transformations in Metals and Alloys*," 3rd ed., Part 2, Pergamon Press, Oxford (2002).
(b) C. M. Wayman and H. K. D. H. Bhadeshia in "*Physical Metallurgy*," 4th ed., Eds. R. W. Cahn and P. Haasen, North Holland, Amsterdam, p. 1507 (1996).
(c) W. Jones, J. M. Thomas and J. O. Williams, *Phil. Mag*., **32**, 1 (1975).
20. (a) H. S. Park, O.-H. Kwon, J. S. Baskin, B. Barwick and A. H. Zewail, *Nano Lett.*, **9**, 3959 (2009).
(b) J. M. Thomas, *Phil. Trams. R. Soc. Lond. A*, **277**, 251 (1974).
(c) J. M. Thomas, *Nature*, **289**, 633 (1981).
(d) J. M. Thomas, *J. Chem. Phys*., **128**, 182502 (2008).
21. (a) M. Haruta, T. Kobayashi, H. Sano and N. Yamada, *Chem. Lett*., **16**, 405 (1987).
(b) R. Meyer, C. Lemire, S. K. Shaikhutdinov and H.-J. Freund, *Gold Bull*., **37**, 72 (2004).
(c) H. Falsig, B. Hvolbæk, I. S. Kristensen, T. Jiang, T. Bligaard, C. H. Christensen and J. K. Nørskov, *Angew. Chem*., *Intl. Ed*., **47**, 4835 (2008).
(d) P. P. Edwards and J. M. Thomas, *Angew. Chem*., *Intl. Ed*., **46**, 5480 (2007).
(e) M. Chen, Y. Cai, Z. Yan and D. W. Goodman, *J. Am. Chem. Soc*., **128**, 6341 (2006).
22. (a) K. D. M. Harris and J. M. Thomas, *Cryst. Growth Des*., **5**, 2124 (2005).
(b) J. M. Thomas, *Angew. Chem*., *Intl. Ed*., **44**, 5563 (2005).
23. (a) R. M. Glaeser, K. Downing, D. DeRosier, W. Chiu and J. Frank, "*Electron Crystallography of Biological Macromolecules*," Oxford University Press, New York (2007).
(b) J. Frank, "*Three-Dimensional Electron Microscopy of Macromolecular Assemblies: Visualization of Biological Molecules in Their Native State*," Oxford University Press, New York (2006).
24. D. Typke, C. J. Gilpin, K. H. Downing and R. M. Glaeser, *Ultramicroscopy*, **107**, 106 (2007).
25. (a) F. Schotte, M. Lim, T. A. Jackson, A. V. Smirnov, J. Soman, J. S. Olson, G. N. Phillips, Jr., M. Wulff and P. A. Anfinrud, *Science*, **300**, 1944 (2003).
(b) H. Frauenfelder, B. H. McMahon and P. W. Fenimore, *Proc. Natl. Acad. Sci. USA*, **100**, 8615 (2003).

(c) H. Ihee, S. Rajagopal, V. Šrajer, R. Pahl, S. Anderson, M. Schmidt, F. Schotte, P. A. Anfinrud, M. Wulff and K. Moffat, *Proc. Natl. Acad. Sci. USA*, **102**, 7145 (2005).
26. (a) S. Subramaniam and R. Henderson, *Nature*, **406**, 653 (2000).
(b) W. Kühlbrandt, *Nature*, **406**, 569 (2000).
27. V. A. Lobastov, R. Srinivasan and A. H. Zewail, *Proc. Natl. Acad. Sci. USA*, **102**, 7069 (2005).
28. (a) D. Shorokhov, S. T. Park and A. H. Zewail, *Chem. Phys. Chem*., **6**, 2228 (2005).
(b) C.-Y. Ruan, V. A. Lobastov, R. Srinivasan, B. M. Goodson, H. Ihee and A. H. Zewail, *Proc. Natl. Acad. Sci. USA*, **98**, 7117 (2001).
29. H. Ma, C. Wan, A. Wu and A. H. Zewail, *Proc. Natl. Acad. Sci. USA*, **104**, 712 (2007).
30. (a) M. M. Lin, L. Meinhold, D. Shorokhov and A. H. Zewail, *Phys. Chem. Chem. Phys.*, **10**, 4227 (2008).
(b) O. F. Mohammed, G. S. Jas, M. M. Lin and A. H. Zewail, *Angew. Chem., Intl. Ed.*, **48**, 5628 (2009).
31. (a) M. M. Lin, D. Shorokhov and A. H. Zewail, *Phys. Chem. Chem. Phys*., doi:10.1039/b910794k (2009), in press.
(b) M. M. Lin, D. Shorokhov and A. H. Zewail, *J. Phys. Chem. A*, **113**, 4075 (2009).
(c) M. M. Lin, D. Shorokhov and A. H. Zewail, *Chem. Phys. Lett.*, **420**, 1 (2006).
32. M. Rueda, S. G. Kalko, F. J. Luque and M. Orozco, *J. Am. Chem. Soc*. **125**, 8007 (2003), and references therein.
33. (a) D.-S. Yang, N. Gedik and A. H. Zewail, *J. Phys. Chem. C*, **111**, 4889 (2007).
(b) F. Vigliotti, S. Chen, C.-Y. Ruan, V. A. Lobastov and A. H. Zewail, *Angew. Chem*., *Intl. Ed*., **43**, 2705 (2004).
34. See the Feature Article entitled "*Molecular Structure and Orientation: Concepts from Femtosecond Dynamics*" [J. S. Baskin and A. H. Zewail, *J. Phys. Chem. A*, **105**, 3680 (2001)], which reviews the work begun in early 1980s.
35. (a) P. M. Felker and A. H. Zewail, in "*Femtosecond Chemistry*," Eds. J. Manz and L. Wöste, VCH, Weinheim, p. 193 (1995).
(b) P. M. Felker and A. H. Zewail, in "*Jet Spectroscopy and Molecular Dynamics*," Eds. J. M. Hollas and D. Phillips, Blackie Academic and Professional, London, p. 222 (1995); see also the publisher's erratum enclosed with the book.
(c) P. M. Felker and A. H. Zewail, in "*Jet Spectroscopy and Molecular Dynamics*," Eds. J. M. Hollas and D. Phillips, Blackie Academic and Professional, London, p. 181 (1995); see also the publisher's erratum enclosed with the book.
36. J. C. Williamson and A. H. Zewail, *J. Phys. Chem*., **98**, 2766 (1994).
37. (a) J. S. Baskin and A. H. Zewail*, Chem. Phys. Chem*., **7**, 1562 (2006).
(b) J. S. Baskin and A. H. Zewail, *Chem. Phys. Chem*., **6**, 2261 (2005).
38. P. Reckenthaeler, M. Centurion, W. Fuβ, S. A. Trushin, F. Krausz and E. E. Fill, *Phys. Rev. Lett*., **102**, 213001 (2009).
39. (a) F. Carbone, D.-S. Yang, E. Giannini and A. H. Zewail, *Proc. Natl. Acad. Sci. USA*, **105**, 20161 (2008).
(b) Y. L. Chen, W. S. Lee and Z. X. Shen, *Proc. Natl. Acad. Sci. USA*, **106**, 963 (2009).

40. J. C. H. Spence and R. B. Doak, *Phys. Rev. Lett*., **92**, 198102 (2004).
41. (a) M. A. Monastyrskiy, S. V. Andreev, D. E. Greenfield, G. I. Bryukhnevich, V. A. Tarasov and M. Y. Schelev, in "*26th International Congress on High-Speed Photography and Photonics*", *Proc. SPIE*, Vol. 5580, Eds. D. L. Paisley, S. Kleinfelder, D. R. Snyder and B. J. Thompson, p. 324 (2005).
(b) P. L. Kapitza and P. A. M. Dirac, *Proc. Cambridge Phil. Soc*., **29**, 297 (1933).
(c) P. H. Bucksbaum, D. W. Schumacher and M. Bashkansky, *Phys. Rev. Lett*., **61** 1182 (1988).
(d) D. L. Freimund, K. Aflatooni and H. Batelaan, *Nature*, **413**, 142 (2001).
42. A. Sali, R. Glaeser, T. Earnest and W. Baumeister, *Nature*, **422**, 216 (2003).
43. (a) J. M. Zuo, M. Kim, M. O'Keeffe and J. C. H. Spence, *Nature*, **401**, 49 (1999).
(b) E. R. Scerri, *J. Chem. Educ*., **77**, 1492 (2000).
44. J. Frank, *Bioessays*, **23**, 725 (2001).
45. (a) K. A. Taylor, *J. Microsc*., **112**, 115 (1978).
(b) K. A. Taylor and R. M. Glaeser, *Science*, **186**, 1036 (1974).

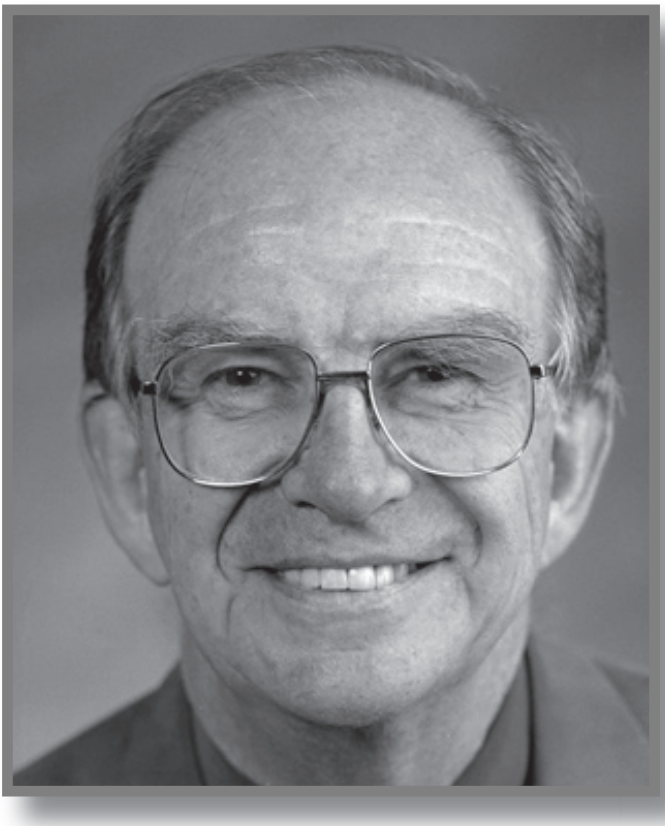
This page intentionally left blank

Biographical Profiles

This page intentionally left blank

Ahmed H. Zewail

Ahmed Hassan Zewail is the Linus Pauling Chair Professor of Chemistry and Professor of Physics, and Director of the Physical Biology Center at Caltech. In 1999, he was awarded the Nobel Prize for the development of the field of *Femtochemistry*. Since then, the main focus of his research group at Caltech has been on the development of another field — *4D Electron Microscopy* — for the direct imaging in the four dimensions of *space* and *time*. He holds Honorary Degrees from forty universities and is an elected member of academies and societies around the world. Postage stamps were issued in recognition of his contributions to science and humanity. Recently, he was appointed to President Obama's Council of Advisors on Science and Technology. More information can be found at http://zewail.caltech.edu.



John M. Thomas

John Meurig Thomas is one of the world's foremost solid-state chemists who pioneered the chemical applications of 2D and 3D electron microscopy since the 1960s. He is also a leading authority on the design of solid catalysts. A former Head of Physical Chemistry, and former Master (Head) of Peterhouse (the oldest college) in the University of Cambridge, he is also a former Director of the Royal Institution of Great Britain, where he occupied the Chair created for Michael Faraday. A Sesquicentenary Honorary Fellow of the Royal Microscopic Society, he is also a foreign Honorary Fellow of numerous national academies and the recipient of several international awards. In 1991, he was knighted by Queen Elizabeth for services to chemistry and the popularization of science.