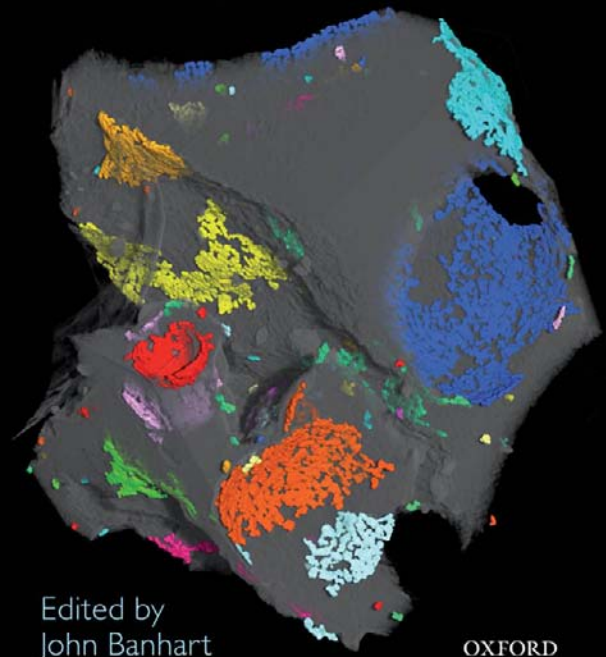
Advanced Tomographic Methods in Materials Research and Engineering



Edited by
John Banhart

OXFORD

MONOGRAPHS ON THE
PHYSICS AND CHEMISTRY OF
MATERIALS

MONOGRAPHS ON THE PHYSICS AND CHEMISTRY OF MATERIALS

Theory of dielectrics H. Frohlich
Strong solids (Third edition) A. Kelly and N. H. Macmillan
Optical spectroscopy of inorganic solids B. Henderson and G. F. Imbusch
Quantum theory of collective phenomena G. L. Sewell
Principles of dielectrics B. K. P. Scaife
Surface analytical techniques J. C. Rivière
Basic theory of surface states Sydney G. Davison and Maria Steslicka
Acoustic microscopy G. A. D. Briggs
Light scattering: principles and development W. Brown
Quasicrystals: a primer (Second edition) C. Janot
Interfaces in crystalline materials A. P. Sutton and R. W. Balluffi
Atom probe field ion microscopy M. K. Miller, A. Cerezo, M. G. Hetherington, and G. D. W. Smith
Rare-earth iron permanent magnets J. M. D. Coey
Statistical physics of fracture and breakdown in disordered systems B. K. Chakrabarti and L. G. Benguigui
Electronic processes in organic crystals and polymers (Second edition) M. Pope and C. E. Swenberg
NMR imaging of materials B. Blümich
Statistical mechanics of solids L. A. Girifalco
Experimental techniques in low-temperature physics (Fourth edition) G. K. White and P. J. Meeson
High-resolution electron microscopy (Third edition) J. C. H. Spence
High-energy electron diffraction and microscopy L.-M. Peng, S. L. Dudarev, and M. J. Whelan
The physics of lyotropic liquid crystals: phase transitions and structural properties A. M. Figueiredo Neto and S. Salinas
Instabilities and self-organization in materials, Volume 1: Fundamentals of nanoscience, Volume 2: Applications in materials design and nanotechnology N. Ghoniem and D. Walgraef
Introduction to scanning tunneling microscopy (Second edition) C. J. Chen
Smart structures V. K. Wadhawan
Advanced tomographic methods in materials research and engineering J. Banhart

ADVANCED TOMOGRAPHIC METHODS IN MATERIALS RESEARCH AND ENGINEERING

Edited By

JOHN BANHART

Hahn-Meitner Institute, Berlin, Germany

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS
Great Clarendon Street, Oxford OX2 6DP
Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide in
Oxford New York
Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto
With offices in
Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam
Oxford is a registered trade mark of Oxford University Press
in the UK and in certain other countries
Published in the United States
by Oxford University Press Inc., New York

First published 2008

British Library Cataloguing in Publication Data
Data available
Library of Congress Cataloging in Publication Data
Data available
Printed in Great Britain
on acid-free paper by
Biddles Ltd., www.biddles.co.uk
ISBN 978–0–19–921324–5
1 3 5 7 9 10 8 6 4 2

Dedicated to the memory of Attila Kuba (1953–2006)

This page intentionally left blank

PREFACE

This book was inspired by the success of a special topical symposium at the annual spring meeting of the German Physical Society in Berlin in 2005 for which the editor was one of the organizers. Such symposia usually attract 10 to 30 papers but this particular one on 'Tomographic Methods in Materials Research' set a new record with 45 papers. This showed that there is a lot of interest in three-dimensional imaging methods both in materials science and in physics. Moreover, it became apparent that a concise treatment of 'advanced' tomographic methods is still lacking.

In contrast to the symposium, this book omits all three-dimensional imaging methods based on destroying the specimen to be investigated, e.g. by cutting, polishing, or milling. There is a reason for this, besides the obvious lack of space: although the information provided by destructive imaging can be similar to that of images obtained non-destructively, the underlying physics and mathematical algorithms differ, and the book would not have had a cohesive theme with all these methods included. With the word 'advanced' in the book title, some of the established methods based on conventional X-ray tubes, nuclear magnetic resonance scanners or other laboratory-based equipment are also excluded. This does not imply that these methods are unsophisticated: rather, the word 'advanced' is used here to reflect the fact that novel third-generation synchrotrons, modern neutron sources, and electron guns in the most sophisticated transmission electron microscopes are also being used to capture three-dimensional images of materials and components. This represents new opportunities for a large community of engineers, materials scientists, or applied physicists, who are the target group of this book in terms of profession. The book should be useful for senior undergraduates and postgraduates, as well as for postdoctoral workers.

Before beginning a description of the individual tomographic methods, some fairly general introductory chapters on the principles and mathematics of tomography, and the basic interactions between matter and the three radiation types considered, are given. Ample reference lists in each chapter allow the reader to seek out information that is not contained in the book. However, the book is conceived such that the concepts of advanced tomography can be understood without necessarily having to read further, more specialized literature.

The field of three-dimensional tomographic imaging is developing quickly, and the more the community of materials researchers gets to know the available methods, the more exciting applications will emerge from their work. One aim of this book is to stimulate such work.

John Banhart, Berlin, June 2007

ACKNOWLEDGEMENTS

The editor would like to thank Jörg Bajorat, Jakob Slaby, Simon Zabler, Markus Wollgarten and Andre Hilger for helping with formatting, for valuable discussions, or for proof reading. Gerd Schneider and Nelia Wanderka provided images, Claudio Ferrero, Michael Godehardt, Alessandro Mirone, and Christof Reinhart gave their support to the compilation of examples on the CD-ROM.

CONTENTS

AUTHORS

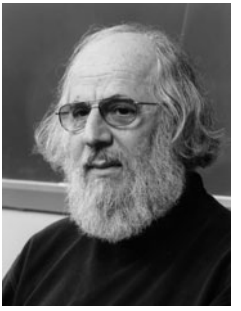
Prof Dr John Banhart
Hahn-Meitner-Institut (HMI)
Glienicker Strasse 100
14109 Berlin
Germany

Dr Pierre Bleuet
European Synchrotron Radiation Facility (ESRF)
6 rue Jules Horowitz
38043 Grenoble
France

Dr Bruno Golosio
Struttura Dipartimentale di Matematica e Fisica
Universita' di Sassari
via Vienna, 2
07100 Sassari
Italy

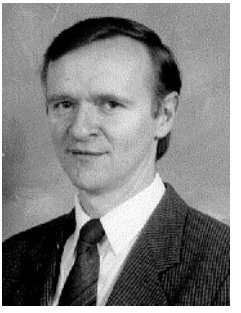
Dr Heinz Graafsma
Deutsches Elektronen Synchrotron (Hasylab)
Notkestr. 85
22607 Hamburg
Germany

Dr Astrid Haibel
Deutsches Elektronen Synchrotron (Hasylab)
Notkestr. 85
22607 Hamburg
Germany



Prof Gabor T Herman, PhD
Distinguished Professor of Computer Science
The Graduate Center
City University of New York
USA

Dr Nikolay Kardjilov
Hahn-Meitner-Institut (HMI)
Glienicker Strasse 100
14109 Berlin
Germany



Prof Dr Attila Kuba
Image Processing and Computer Graphics
University of Szeged
Hungary

Dr Eberhard H Lehmann
Spallation Neutron Source Division
Paul Scherrer Institut (PSI)
5232 Villigen PSI
Switzerland

Dr Thierry Martin
Instrument Support Group
Detector Systems Unit
European Synchrotron Radiation Facility (ESRF)
38043 Grenoble
France

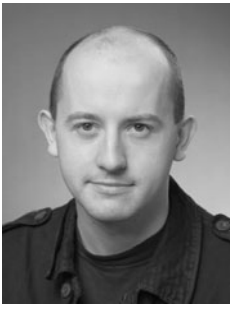
Dr Paul A Midgley
Department of Materials Science and Metallurgy
University of Cambridge
Pembroke Street
Cambridge CB2 3QZ
UK

Prof Dr Keith A Nugent
School of Physics
The University of Melbourne
Melbourne, Victoria, 3010
Australia

Prof Dr Joachim Ohser
University of Applied Sciences
Schfferstrasse 3
64295 Darmstadt
Germany

Dr Andrew G Peele
Department of Physics
La Trobe University
Bundoora, Victoria, 3086
Australia

Prof Dr Henning Friis Poulsen
Risø National Laboratory
4000 Roskilde
Denmark



Dr Alexander Rack
Forschungszentrum Karlsruhe
Institute for Synchrotron Radiation – ANKA
76344 Eggenstein-Leopoldshafen
Germany

Dr Katja Schladitz
Fraunhofer Institut für Techno- und Wirtschaftsmathematik (ITWM)
Fraunhofer-Platz 1
67663 Kaiserslautern
Germany

Prof Dr Christian Schroer
Institut für Strukturphysik
Technische Universiät Dresden
01062 Dresden
Germany

Prof Dr Alexandre Simionovici
Observatoire des Sciences de l'Univers de Grenoble
Lab. Géophysique Interne & Tectonophysique
UMR5559, Maison des Geosciences, BP 53
38041 Grenoble Cedex 9
France

Dr Heiko Stegmann
Carl Zeiss NTS Innovationslabor Dresden
Manfred von Ardenne-Ring 20
01099 Dresden
Germany

Dr Yoshio Suzuki
Japan Synchrotron Radiation Research Instutute
SPring-8
Sayo, Hyogo, 679-5198
Japan

Prof Dr Hiroyuki Toda
Department of Production Systems Engineering
Toyohashi University of Technology
Hibarigaoka, Tempaku, Toyohashi, Aichi, 441-8580
Japan

Prof Dr Wolfgang Treimer
University of Applied Sciences (TFH)
Department II: Physics, Mathematics, Chemistry
Luxemburger Strae 10
13353 Berlin
Germany

Dr Matthew Weyland
Department of Materials Engineering
Monash University
Victoria, 3800
Australia

LIST OF ACRONYMS

2D,3D	two(three)-**d**imensional
3DAP	**3D A**tom **P**robe
3DXRD	**3D X**-**r**ay **D**iffraction
AFM	**A**tomic **F**orce **M**icroscopy
APS	**A**dvanced **P**hoton **S**ource
ART	**A**lgebraic **R**econstruction **T**echnique
BESSY	**B**erliner **E**lektronen**s**peicherring-Gesellschaft für **Sy**nchrotronstrahlung
BF	**B**right **F**ield
BGO	**B**ismuth **G**ermanate
CAT	**C**omputer-**A**ssisted **T**omography
CCD	**C**harge-**C**oupled **D**evice
(C)MOS	(**C**omplementary) **M**etal **O**xide **S**emiconductor
CONRAD	**Co**ld **N**eutron **Rad**iography
CRL	**C**omposite **R**efractive **L**ense
CT	**C**omputerized **T**omography
CTF	**C**ontrast **T**ransfer **F**unction
DCT	**D**iffraction **C**ontrast **T**omography
DEI	**D**iffraction-**E**nhanced **I**magingg
DQE	**D**etective **Q**uantum **E**fficiency
DESY	**D**eutsches **E**lektronen**sy**nchrotron
DF	**D**ark **F**ield
DFT	**D**iscrete **F**ourier **T**ransform
DPR	**D**ifferential **P**hase **R**esidual
DT	**D**iscrete **T**omography
EBSD	**E**lectron **B**ackscatter **D**iffraction
EDT	**E**uclidean **D**istance **T**ransformation
EDX	**E**nergy-**D**ispersive **X**-ray Spectroscopy
EFTEM	**E**nergy-**F**iltered **T**ransmission **E**lectron **M**icroscopy
EIT	**E**lectrical **I**mpedance **T**omography
ERT	**E**lectrical **R**esistance **T**omography
ESRF	**E**uropean **S**ynchrotron **R**adiation **F**acility

FBP	**F**iltered **B**ack**p**rojection
FEM	**F**inite **E**lement **M**ethod
FET	**F**ield **E**ffect **T**ransistor
FFT	**F**ast **F**ourier **T**ransform
FIB	**F**ocused **I**on **B**eam
FO	**F**ibre **O**ptic
FOV	**F**ield **o**f **V**iew
FReLoN	**F**ast **Re**adout **Lo**w **N**oise
FSC	**F**ourier **S**hell **C**orrelation
FWHM	**F**ull **W**idth at **H**alf-**M**aximum
FZP	**F**resnel **Z**one **P**late
Gadox	**Gad**olinium **Ox**ysulphide
GGG	**G**adolinium **G**allium **G**arnet
HAADF	**H**igh-**A**ngle **A**nnular **D**ark-**F**ield
ICON	**I**maging with **Co**ld **N**eutrons
ITT	**I**ntegrated **T**omographic **T**echniques
KB	**K**irkpatrick–**B**aez
LAG	**L**utetium **A**luminium **G**arnet
LEAP	**L**ocal **E**lectrode **A**tom **P**robe
LSF	**L**ine-**S**pread **F**unction
LSO	**L**utetium Oxyortho**s**ilicate
MAVI	**M**odular **A**lgorithms for **Vi**sualization
MAP	**M**aximum **a** **P**osteriori
μ-CT	**M**icro-**C**omputed **T**omography
ML(EM)	**M**aximum **L**ikelihood **M**ethod (with **E**xpectation **M**aximization)
MMSE	**M**inimum **M**ean **S**quare **E**rror
MOS	**M**etal **O**xide **S**emiconductor
MRI	**M**agnetic **R**esonance **I**maging
MS	**M**ultiple **S**cattering
MTF	**M**odulation **T**ransfer **F**unction
NA	**N**umerical **A**perture
NDT	**N**on-**D**estructive **T**esting
NEUTRA	**Neu**tronen **Tra**nsmissions Radiography
NMR	**N**uclear **M**agnetic **R**esonance
NSLS	**N**ational **S**ynchrotron **L**ight **S**ource

PET	**P**ositron **E**mission **T**omography
PIN	**p**ositive **i**ntrinsic **n**egative (diode)
PyHST	**Py**thon **H**igh-**S**peed **T**omographic Reconstruction
pixel	**Pi**cture **El**ement
PSF	**P**oint-**S**pread **F**unction
PScF	**P**oint-**Sc**attered **F**unction
PSI	**P**aul **S**cherrer **I**nstitute
RF	**R**adio **F**requency
RGB	**R**ed **G**reen **B**lue
SANS	**S**mall-**A**ngle **N**eutron **S**cattering
SAXS	**S**mall-**A**ngle **X**-ray **S**cattering
SCF	**S**ingle-**C**rystal **F**ilm
SEM	**S**canning **E**lectron **M**icroscopy
SART	**S**imultaneous **A**lgebraic **R**econstruction **T**echnique
SA	**S**imulated **A**nnealing
SINQ	**S**chweizer**i**sche **N**eutronen**q**uelle
SIRT	**S**imultaneous **I**terative **R**econstruction **T**echnique
SSD	**S**olid-**S**tate **D**etector
(S)SNR	(**S**pectral) **S**ignal-to-**N**oise **R**atio
SPring-8	**S**ynchrotron Ring 8 GeV
STEM	**S**canning **T**ransmission **E**lectron **M**icroscopy
TDS	**T**hermal **D**iffuse **S**cattering
TEM	**T**ransmission **E**lectron **M**icroscopy
TFT	**T**hin-**F**ilm **T**ransistor
ULK	**U**ltra-**l**ow κ
voxel	**vo**lume picture **el**ement
WBP	**W**eighted **B**ack**p**rojection
XANES	**X**-ray **A**bsorption **N**ear-**E**dge **S**tructure
XAS	**X**-ray **A**bsorption **S**pectroscopy
XCFT	**X**-ray **C**omputed **F**luorescence **T**omography
XRD	**X**-**r**ay **D**iffraction
XRF	**X**-**r**ay **F**luorescence
XTM	**X**-ray **T**omographic **M**icroscopy
YAG	**Y**ttrium **A**luminium **G**arnet

This page intentionally left blank

PART I

BASIC CONCEPTS

In the first part of this book general facts that are important for tomography shall be described independent of the type of radiation used and the specific experimental method chosen.

First, some general definitions and concepts will be described, after which the fundamental mathematical part of tomography will be reviewed: this includes the mathematics of reconstruction of tomographic data, i.e. the formalism needed to calculate 3D images from tomographic projections, and the mathematics needed to display, manipulate and to analyse 3D images by the computer.

In a later chapter the ways to generate the different types of radiation described in this book and the various interactions of this radiation with matter are reviewed.

This page intentionally left blank

1

INTRODUCTION

John Banhart

1.1 History and motivation

Among the many ways to display information about real objects, the use of images is the most direct, easy to understand and widespread method. Images in the original sense represent real-space information about optical properties of objects, e.g. brightness, colour, shininess. In a more general sense images describe spatial distributions of physical properties such as density, water content or element distribution. Most images are two-dimensional (2D) and appear on a piece of paper or on a computer screen. A sculpture is an example of a three-dimensional image of a three-dimensional object, mostly made by an artist and often with deliberate deviations from the true 3D object. Because making sculptures is not a practical way, one attempts to visualize three-dimensional information in two dimensions by perspective drawing or by showing many two-dimensional slices through an object shown individually or stacked on top of each other.

Often, images are obtained by using radiation that passes through an object during which the radiation properties are modified. Radiation imaging can make use of electromagnetic radiation all through the spectrum from microwaves to γ-rays, particles such as protons, electrons, neutrons, or other waves such as sound waves. The interactions between the rays and the object can be manifold: radiation can be simply attenuated, shifted in phase, deflected, scattered, or changed in energy. After passing through the objects, suitable detector systems detect and record the changes. Alternatively, the incoming radiation can excite molecules or atoms in the object that send out their own radiation that is then monitored.

One of the most simple and successful ways of imaging is X-ray radiography. An X-ray beam penetrates an object and is attenuated due to interactions with the electrons of the material. Effects other than absorption can be neglected in most cases so that an X-ray-sensitive emulsion or detector behind the object can pick up a two-dimensional image of the object representing the projected electron density in the object. One of the first ever images of this kind is shown in Fig. 1.1. Owing to the geometry of the arrangement, features within the object behind each other with respect to the beam direction are superimposed and cannot be separated. Doctors try to partially overcome this problem by turning, e.g. the hand of the patient, and to acquire more than one radiogram from different

FIG. 1.1. X-ray radiogram made by Wilhelm Conrad Röntgen (1845–1923) of the hand of a colleague in 1896. Röntgen was awarded the first Nobel prize for physics in 1901. Element 111 has been named 'Roentgenium' in 2004 to honour his discovery of X-rays in 1895.

angles. This allows them to separate features that are otherwise superimposed, see Fig. 1.2.

From this procedure a further important step was taken with the invention of medical tomography by Godfrey N. Hounsfield (1919-2004) (Hounsfield, 1972, Hounsfield, 1973): Instead of merely taking radiographs from a small number of directions, the object is rotated around an axis through 360°, and radiographs are acquired every given angular increment. A mathematical reconstruction algorithm is then applied that approximately calculates the X-ray attenuation coefficient of any spatial volume inside the object, see Chapter 2. The mathematical foundations were first laid by Johann Radon (1917), but the contribution of Allan Cormack (1963) who many years later formulated some of the mathematical principles, is considered equally important. Hounsfield and Cormack received the Nobel prize in Medicine in 1979. The result of mathematical reconstruction is a 3D image of the electron density inside an object that can be

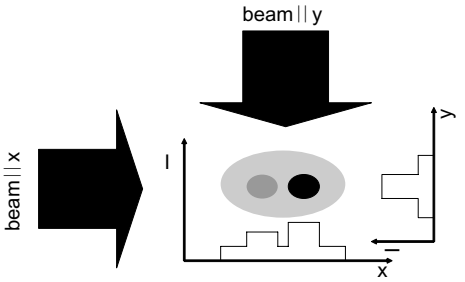


FIG. 1.2. Radiography from two different directions. Arrows denote the incoming radiation, $I(x)$ is the intensity profile measured behind the sample. The object contains three components with different densities.

displayed and analysed in many different ways, as will be described in Chapter 3.

Historically, medical X-ray tomography was developed first. Terms like *computed tomography* (CT), *computer-assisted tomography* (CAT), *computed axial tomography* (also CAT), or *computerized tomography* were used to specify the methods. Medical X-ray tomography is described in the literature, see e.g. Kalender (2000) and Hsieh (2003). The fact that computers are used to calculate the 3D image is almost trivial nowadays so that the prefix 'computer' is often omitted.

CT was also used for the investigation of non-living structures just a few years later. Some early experiments were, e.g., conducted at the Bundesanstalt für Materialforschung (BAM) in Berlin in 1978 using a commercial medical CT scanner and X-ray energies up to 120 keV (Reimers *et al.*, 1984). From these experiments *industrial X-ray tomography* emerged. Although CT seems an ideal tool for problems occurring in industry, e.g. in quality management or even production, since it provides a precise 3D image of the internal structures of a part and is non-destructive, it has to compete with ordinary radiography in many cases. Inspection of castings, e.g., can be carried out by CT in much more detail than by radiography but time and cost are considerably higher for CT so that routine X-ray inspection in production is still done by radiography. However, CT has proven valuable in the areas of rapid prototyping, reverse engineering and metrology.

Rapid prototyping means building parts in a variety of materials from computer models using various methods such as fused deposition modelling, selective laser sintering or multiphase jet solidification (Kai *et al.*, 2004). Data obtained by CT can be used to create the 3D data set needed as an input for the rapid prototyping machine.

Reverse engineering comprises analysing existing parts for which no blue prints are available. CT can be used to scan the internal structure of a part and to create a computer-aided design (CAD) model that can be further engineered. CT is an alternative to co-ordinate measuring machines that are otherwise used

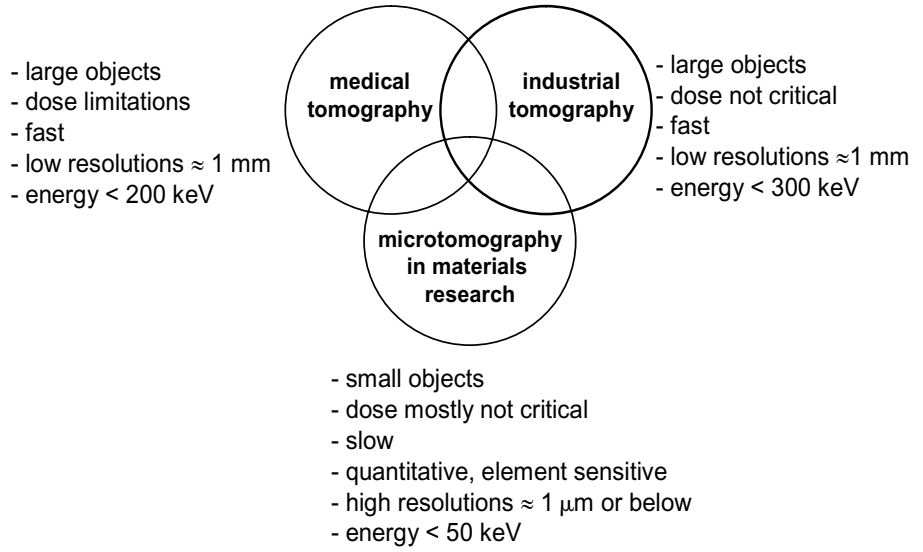


FIG. 1.3. Features of different types of X-ray tomography.

for reverse engineering, except with the additional option of not only obtaining surface information but also accurate information of the internal structure.

Metrology is a task that comprises the precise determination of distances within a part. CT can be used for it since resolution is high and it is straightforward to calculate distances from 3D data sets in a computer.

Figure 1.3 lists the main characteristics of tomography in different fields. Obviously it matters whether CT is to be used in medicine, materials research or for testing components in industry.

Despite some success, X-ray tomography is still not very common in materials science and engineering compared to the number of applications in medicine. One reason is the rather low resolution available until ten years ago. In materials analysis the relevant features to be detected are often much smaller than the ones in the body of humans or animals. Now, the μm or even sub-μm resolution range is accessible by using novel X-ray tubes or synchrotrons, methods that are frequently called *microtomography* (μCT) and applications in materials research are emerging. As resolutions are comparable to light microscopy, tomography can act as '3D-metallography' in metals research. This is nicely demonstrated by Fig. 1.4, which shows a 2D slice through a holotomographic image (see Chapter 6) of an alloy and compares it to a traditional metallographic section. Obviously, the features visible are very similar, but the tomographic data set contains information from full space and allows for a true 3D assessment of the structure. As synchrotron radiation is many orders of magnitude more intense than that of X-ray tubes, is parallel, can be monochromatized and shows partial coherence, many novel opportunities, such as phase contrast, fluorescence or

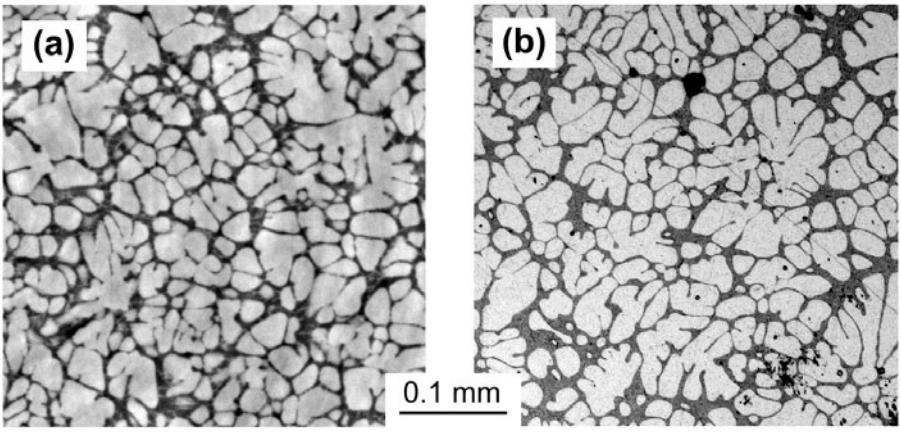


FIG. 1.4. A step towards 3D metallography: AlSi7Mg (A357) alloy with rosette-shaped α-Al grains imaged by (a) holotomography – a 2D slice through the 3D data set is displayed – and, (b) optical microscopy of a polished section. Courtesy: Simon Zabler, Hahn-Meitner-Institute Berlin.

diffraction tomography have been implemented, and yield valuable information. The high flux of synchrotron sources allows for fast measurements of radiograms in the ms range, of tomograms in the <10 s range, as compared to 30 min or more for conventional tomography. *In-situ* investigation of processes in great detail is therefore possible. Examples include the foaming of metals (Banhart *et al.*, 2001) the discharge of batteries (Manke *et al.*, 2007) – both studied in two dimensions – and sintering of metals, solidification of alloys or imbibition of liquid into granular material studied in three dimensions (Di Michiel *et al.*, 2005). Acquisition of an individual X-ray tomogram in less than 10 s has now become possible.

1.2 What is tomography?

To start with semantics: The word 'tomography' comes from the Greek words 'to cut or section' (tomos) and 'to write' (graphein). In this book, a device used in tomography is called a *tomograph*, while the image produced is a *tomogram*, the person making the tomograph would then be the *tomographer*. The use of the words *radiography*, *radiograph* and *radiogram* should be analogous, but there are many slightly different uses for these terms. Both the Greek suffixes '-graph' and '-gram' stand for 'something written or drawn' and there is therefore some ambiguity in these definitions. The word 'tomography' is nowadays used for many 3D imaging methods, not just for those based on radiographic projections, but also for a wider range of techniques that yield 3D images.

In the following, the various tomographic methods are subdivided into three groups:

1.2.1 *Non-destructive techniques using simple projections*

These methods are tomography in the original sense: an object is imaged non-destructively[1] in many directions by passing rays through it and by measuring the modification of these rays caused by their interaction with the matter of the sample. The physical properties that can be characterized vary, but the *projection requirement* assumed states that the signal used for tomographic reconstruction is a monotonic function of a projected physical property. The result of the *tomographic reconstruction* is a 3D map of the respective physical property. A counterexample are the Bragg reflections occurring in electron microscopy, whenever a crystallite within a sample is rotated into an orientation such that the Bragg peak illuminates the detector, see Section 11.2.1. Another example is electrical conductance. The path of the electrical current depends on the neighbourhood of a straight line through an object and not just on the properties along the path.

1.2.2 *Non-destructive techniques using information beyond simple projection data*

Tomographic methods that create a 3D image by using a more complex signal than a simple projection extend the original meaning of tomography. An example are the Bragg reflections mentioned above. On applying a 3D X-ray diffractometer (3D XRD) a 3D map of crystal orientations can be generated, see Chapter 9. The mathematical treatment of such problems where the projection requirement is not fulfilled, is very challenging.

1.2.3 *Destructive techniques*

Parts of the object to be imaged are removed step-by-step. Either the newly created surfaces are imaged and/or the removed parts of the object are analysed. From this information a 3D image is created. These techniques will not be described in this book, just some examples will be given in Section 1.4.

The term 'non-destructive' should be further explained. Many interactions between the probing rays and matter modify the sample. Neutrons used in neutron tomography lead to nuclear reactions, high-energy electrons hitting a sample may cause radiation damages that can be so severe that a special dose management has to be applied (see e.g. Section 11.2.2).

Most tomographic methods require a certain degree of sample preparation, e.g. the preparation of thin foils for transmission electron microscopy that actually destroys the original sample. Despite these limitations, we call a method *non-destructive* if the damage caused during measurement is an unwanted side effect and can be kept to a level that still allows the relevant 3D information sought to be obtained. *Destructive* techniques are those where the object is deliberately destroyed to get access to its interior or to analyse the material removed. Sample preparation is not considered destructive.

[1] In medicine the term *non-invasive* is used instead of non-destructive.

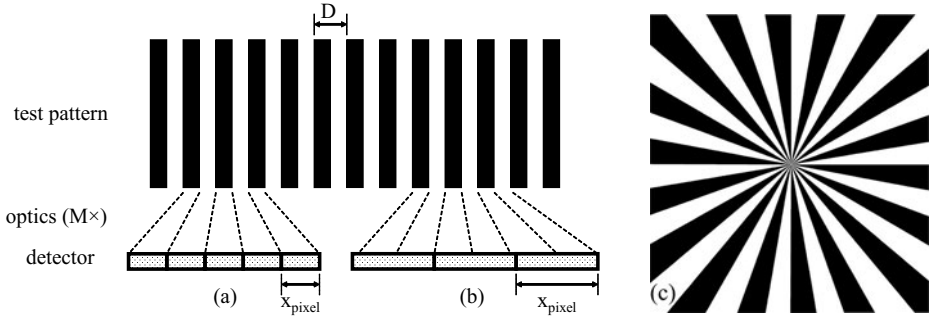


FIG. 1.5. Test patterns to verify the resolution of an imaging system. (a,b): principle of imaging of such a pattern using a detector with small (a) or large pixels (b). (c) layout of a 'Siemens star', a pattern use for real tests.

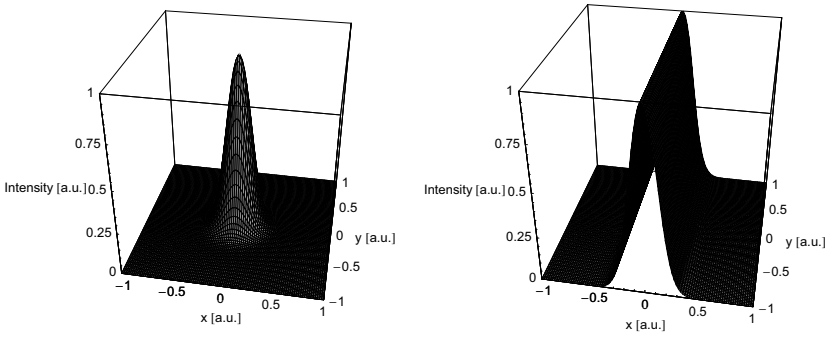
In this book we concentrate on methods that are based on a mathematical reconstruction of 3D images from 2D projections non-destructively. This we call *tomography*.

1.3 What is resolution?

The term 'resolution' is often used for the description of radiographic and tomographic images, but frequently in a rather sloppy way. Resolution is the capability of an imaging system to separate two distinct objects that are located close to each other and to make them visible as separate objects. The definition is facilitated by considering simple objects such as pairs of lines of equal width that are either 'black', i.e. absorbing, or 'white', i.e. transparent, to the probing radiation, see Fig. 1.5. Optical imaging theory provides criteria that determine whether such line pairs (or point sources) can be separated by the observer and do not merely occur as a uniform grey area in the microscope (see Section 7.1 for further definitions of resolution).

A practical way to verify the resolution of an imaging system is to use test patterns such as the ones shown in Figs. 1.5(a,b), Fig. 7.13, or the 'Siemens star' shown in Fig. 1.5(c). Test imaging provides an estimate for spatial resolution.

The resolution of an imaging system is not only limited by the optical components, i.e. lenses, mirrors, *etc.*, but also by the detector used, which regardless of whether it is the human eye or a CCD camera consists of an array of discrete picture elements (pixels). This can limit resolution, as demonstrated in Fig. 1.5. In Fig. 1.5(a) pixel size is just sufficient to reproduce the pattern projected onto the detector by an optical system, in Fig. 1.5(b) the pixel size is too large (or magnification too low) and the detector will not resolve the pattern even if the optics reproduce the image perfectly. Obviously the relationship $D \geq 2x_{\text{pixel}}/M$ must hold in this case, where M is the magnification of the optical system. A further source of blurring is beam divergence. Only a perfectly parallel beam or a



(a) (b)

FIG. 1.6. Schematic representation of the, (a) point-spread function (PSF) and, (b) the line-spread function (LSF).

beam from an ideal point source can produce a sharp image, a beam with divergence cannot, regardless of how good the imaging system is (see Section 13.2.3 for examples from neutron imaging).

Mathematically, the quality of an imaging system is described by various functions: The point-spread function (PSF) defines the apparent shape of an object point as it appears in the image created, see Fig. 1.6. The image of an arbitrary object function $O(\boldsymbol{r})$ then appears as the convolution of the true object function and the PSF. The PSF would be a δ-function for a perfect imaging system, but imaging artefacts lead to a deviation from this ideal shape. Noise further deteriorates the image signal $I(\boldsymbol{r})$ and one measures:

$$I(\boldsymbol{r}) = O(\boldsymbol{r}) * \mathrm{PSF}(\boldsymbol{r}) + \mathrm{noise}(\boldsymbol{r}),$$

where the asterisk stands for the convolution integral. The Fourier integral of the PSF is the *modulation transfer function* (MTF), which also contains a full description of the imaging process and can be used alternatively to the PSF. The MTF specifies as a function of frequency $2\pi\omega$ which Fourier components of an image are transmitted by the imaging system. The better the resolution, the higher frequencies pass. By determining the cutoff frequency ω_c of the MTF(ω), one can calculate the resolution $\Delta = 1/(2\omega_c)$. The MTF can be decomposed into a product of various MTFs, each one describing one component of an optical system, e.g. the scintillator, the lenses and other components. The 'weakest link' then dominates the overall performance.

Practical measurements of the PSF (or MTF) do not start from a point source because such a source is hard to realise. Instead, a line-shaped object, a wire or a sharp edge, is used as an object. Blurring of the image of such a line-shaped object represents the *line-spread function* (LSF) that is the convolution of the PSF and the edge function describing the object. Thus, from the LSF the PSF and MTF can be calculated.

In summary, the resolution in tomographic imaging is limited by a number of factors that will be discussed later, including:

- deviation of the beam from parallelity or ideal conical shape, see e.g. Section 13.2.3,
- imperfect mechanical rotation or tilting stage: tumbling errors, drift effects etc, see e.g. Section 5.4.4 and Section 11.3,
- blurring in the conversion system (scintillator), see Chapter 10,
- performance limitation of optical system, see e.g. Chapter 10,
- too large pixel size of the detector (undersampling),
- insufficient number of angular projections, see Section 2.3.1.

Development of improved tomographic techniques is ongoing. One key objective clearly is to improve resolution.

1.4 Tomographic methods not further treated in this book

As pointed out previously there are many more 3D imaging techniques that can be used in materials research than the ones treated in this book. Some of these are tomographic methods in the sense that they reconstruct images from projections but they either make use of conventional radiation sources or are still exotic and in an experimental stage of development, such as

- *Proton tomography*, e.g. using 250-MeV protons (Johnson *et al.*, 2003),
- *Terahertz tomography*, inspired by X-ray tomography, but using THz (submillimetre) waves to image materials that are penetratable, e.g. polystyrene (Wang *et al.*, 2004),
- *Ultrasound tomography*, used in industry (Gilmore, 1996) and medicine (West, 2003).

Other 3D imaging methods are not based on 2D projections. Of these, a few shall be briefly mentioned in the following to create an impression of related methods that, however, are different from the viewpoint of the mathematical approach and the measurement procedure.

1.4.1 *Positron emission tomography*

Positron emission tomography (PET) is based on radiation emitted by a living body (Bailey *et al.*, 2005). To create such radiation a tracer liquid is injected. The tracer is distributed in the body by blood circulation and metabolism. Depending on the biochemical properties of the particular molecule contained in the tracer, enrichment in certain regions of the body takes place, e.g. in tumours. The tracer liquid comprises molecules that contain radioactive atoms such as ^{18}F having a short half-life and emitting positrons when they decay. The positrons travel just for a few millimetres and then annihilate with electrons, forming pairs of photons, each having energies of about 511 keV. Two such photons radiate in opposite directions and are registered by a position-sensitive detector spanning a 180° ring around the object. Whenever two events are detected within a time span

of a few nanoseconds, the event is accepted (coincidence), otherwise discarded. By determining the line between the locations of the two detector pixels that registered the coincident photons and calculating the intersection with the object, one can conclude on which line through the object the annihilation took place. By accumulating over many positron-annihilation events many such lines can be constructed and a strongly emitting region within the object will be at the intersection of many such lines. Thus, density enrichments can be located in three dimensions. The spatial resolution of the method is of the order of a few millimetres and the signal obtained represents the local density of the tracer molecule. By combining X-ray tomography and PET one can increase the level of information: The X-rays yield the density distribution of the body with a good resolution, PET the local concentration of the radioactive tracer molecule, and both can be superimposed graphically to facilitate clinical analysis.

1.4.2 *Electrical impedance or resistance tomography*

Electrical impedance or resistance tomography (EIT/ERT) is an imaging technique in which an image of the conductivity or permittivity of a conducting object is derived from surface electrical measurements. In medicine, conducting electrodes are attached to the skin of the body under study, and small alternating currents are applied to the electrodes. The resulting electrical potentials are measured, and the process is repeated for numerous different configurations of applied current. Checks of the lung function and detection of skin cancer are applications.

In geophysics, ERT is based on using electrodes attached to the surface of the earth or in bore holes to locate resistance anomalies. Industrial processes can be monitored by, e.g., putting arrays of electrodes onto vessels or pipes containing conductive fluids (Plaskowski *et al.*, 1995). Further applications include stirred tank reactors, centrifugal separators, conveying systems, voidage mapping in porous beds, tracer migration in porous beds and soils (West, 2003, University of Manchester, 2007).

Mathematically, the problem of recovering the conductivity from surface measurements of current and potential is a non-linear inverse problem and is severely ill-posed. There is extensive mathematical research on the problem of uniqueness of solution and numerical algorithms to solve this problem.

1.4.3 *Electrical capacitance tomography*

ECT is a method for the determination of the dielectric permittivity distribution in the interior of an object from external capacitance measurements. It is a close relative of electrical impedance tomography and is proposed as a method for industrial process monitoring (University of Manchester, 2007). Potential applications include measuring and displaying the concentration distribution of a mixture of two insulating (dielectric) fluids, such as oil, gas, plastic, glass and some minerals, located inside a vessel. The technique differs from conventional tomographic methods, in which high-resolution images are formed of slices of

a material. The metallic plates used to pick up the signal must be sufficiently large to give a measureable change in capactiance, implying that the number of electrodes is usually very small, ≈ 10. This means that the technique is limited to producing very low resolution images but it is fast and inexpensive.

1.4.4 *Magnetic resonance imaging*

Magnetic resonance imaging (MRI) is based on an effect called *nuclear magnetic resonance* (NMR) (Slichter, 1994). In the easiest case a sample is placed in a uniform magnetic field. As most nuclei have a magnetic moment, the moment aligns with the applied field. To flip an aligned nucleus one must apply an energy $m\hbar\omega_L$, where $\omega_L = \gamma B_0$ is the Larmor frequency that depends both on the gyromagnetic ratio γ of the specific nucleus and the applied field B_0, and m is an integer. The frequency lies in the range of radio frequencies (RF) for magnetic field strengths of a few teslas as, e.g. for protons (^{1}H) $\gamma/2\pi = 42.576\,\mathrm{MHz/T}$. By applying RF fields with the Larmor frequency, one can create resonance effects and manipulate the nuclear spin ensembles of a particular nucleus. The nuclear spins can be polarized and their decay caused by spin–lattice and spin–spin interactions can be measured. The associated characteristic relaxation times T_1 and T_2 depend not only on the type of nucleus but also on the chemical environment of the atoms containing the nucleus under investigation. Thus, NMR signals provide chemical sensitivity.

In medical MRI, T_1 and T_2 are measured for hydrogen, which is abundant in living bodies and that provides an excellent signal (McRobbie *et al.*, 2002). By applying magnetic fields that are not uniform but have a gradient in all three directions of space one can locate the resonance signals in space and construct 3D maps corresponding to the T_1 and T_2 times. These times are converted to and displayed as greyscales or colour-coded signals. Empirical knowledge exists on how different features in the body influence T_1 and T_2. Tumours, e.g., tend to have longer T_2 times than healthy tissue and a doctor can therefore locate them using MRI maps representing T_2. Spatial resolutions of commercial systems are of the order of a millimetre.

MRI has found other applications outside of the medical and biological fields such as the measurement of rock permeability to hydrocarbons or the characteriztion of timber. MRI was developed by Paul Lauterbur and Peter Mansfield in the early 1970s. They were awarded the Nobel prize in Medicine in 2003 for their achievements.

1.4.5 *Seismic tomography*

The probing signal for seismic tomography are seismic waves created by earthquakes that travel through the planet and 'illuminate' its interior (Iyer and Hirahara, 1993). The time that the waves arrive at seismic stations scattered over the world can then be used to calculate the waves' velocity through the Earth. By combining analyses from many earthquakes in different places around the Earth, a three-dimensional map of wave velocity through the Earth can be

constructed. Seismic tomography is considered as a non-linear inversion problem in which the seismic data are given and the goal is to determine the attributes of the medium such as wave speed and medium density.

1.4.6 *Tomographic imaging by sectional slicing*

A traditional way to analyse a material in 3D is to slice it sectionally and to image the sample's surfaces. In metallography this well-established, although labour-consuming method, is carried out by polishing in roughly equal steps, after which a freshly prepared surface is imaged, e.g. by light microscopy. Development in both slicing and imaging techniques has enabled researchers to carry out such destructive investigations with much higher precisions (resolutions) and with a much larger variety of probing instruments than 10 years ago.

Focussed ion beam milling allows one to mill samples in equal steps using a programmable interface that repeats the slicewise milling procedure automatically. Modern FIB instruments are equipped with a range of electron imaging and analysis systems, so that the sample can remain on the milling stage during analysis.

To give two examples, FIB milling combined with electron backscatter diffraction (EBSD) has allowed for a 3D investigation of the texture below a nanoindent with spatial resolutions around 1 µm (Zaafarani *et al.*, 2006). By sequential sectioning of a polymer sample by plasma etching and subsequent characterization of the surfaces by scanning force microscopy, topographic 3D data sets with a step width of 7.5 nm were realized (Magerle, 2000). Further combinations of sectioning methods and imaging methods are feasible.

1.4.7 *3D atom probe tomography*

3D atom probe tomography (3DAP) is a method yielding three-dimensional images of individual atoms in solid materials with atomic resolution (Cerezo *et al.*, 1988, Blavette *et al.*, 1993). The method is explained in Fig. 1.7(a): the sample has to be prepared as a sharp tip with a curvature of a few tens of nanometres. An applied DC voltage of a few kV creates very large field gradients at the curved surface. The field gradients are such that atoms are just not extracted, but a short additional pulse eventually removes atoms that start to fly away in a radial direction. The atoms are detected at some distance by means of a combination of multichannel plate (MCP) system and a position-sensitive detector. The geometrical magnification defined by the tip curvature and the distance to the detector is very large, up to 10^8. The time of flight is recorded for each individual atom, allowing one to calculate their masses, knowing the initial kinetic energy given by the applied voltages. The experiment is carried out in ultra-high vacuum at temperatures typically around 50 K. Mathematical treatment of the measured data leads to graphical representations such as the one shown in Fig. 1.7(b) where precipitated atom clusters of a complex composition occur in a 8-component maraging steel (Leitner *et al.*, 2004). The method is destructive since all atoms measured are removed. Its resolution is the highest of all

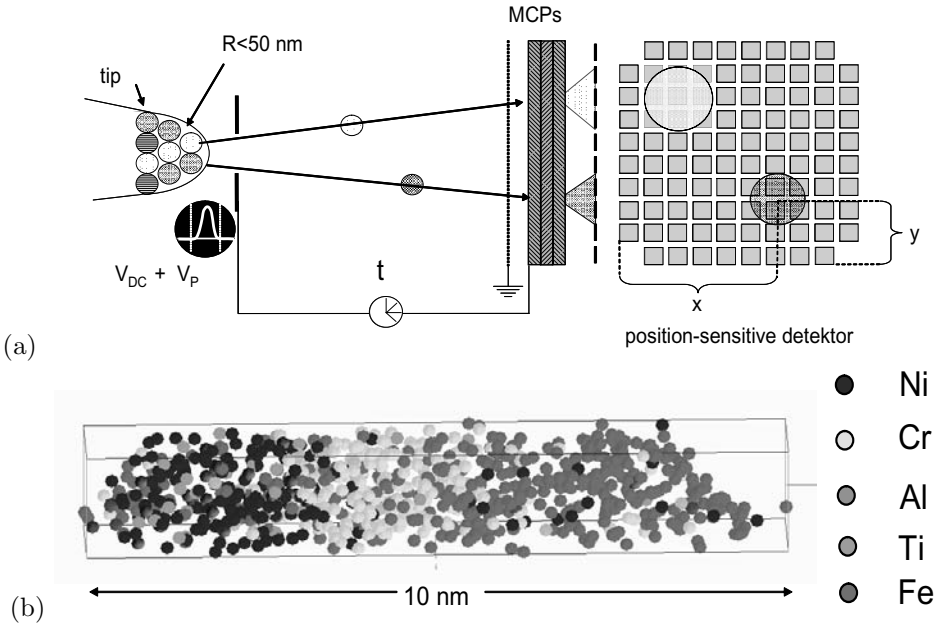


FIG. 1.7. 3D atom probe tomography (3DAP). (a) principle of method; (b) tomogram showing a region in a sample made of the stainless maraging steel Fe-13.1Cr-8.4Ni-1Si-1Ti-0.7Al-0.6Mo-0.1Mn (all concentrations given in at.%). Courtesy: N. Wanderka (HMI Berlin) Image (b) is reproduced in **colour** in the central part of the book.

tomographic methods.

1.5 Summary

Tomography adds a further dimension to traditional imaging methods operating in two dimensions only. The variety of 3D imaging methods is large. The known methods differ by the probing rays they use, by the physical quantity they measure and by the mathematical reconstruction techniques they are based on.

Figure 1.8 shows that tomographic methods range from nanoprobes for submicroscopic samples to engineering probes for entire components. Five distinct methods are listed and visualized with an example from materials science. Corresponding resolutions (left end of each bar) and maximum sample sizes (right end of bar) for each method are specified. The typical length scale spans nine orders of magnitude of resolution and sample size. A great wealth of information is contained in these images, depending on the specific method. Synchrotron X-ray, electron and neutron tomography will be explained in detail in Parts II, III and IV of this book.

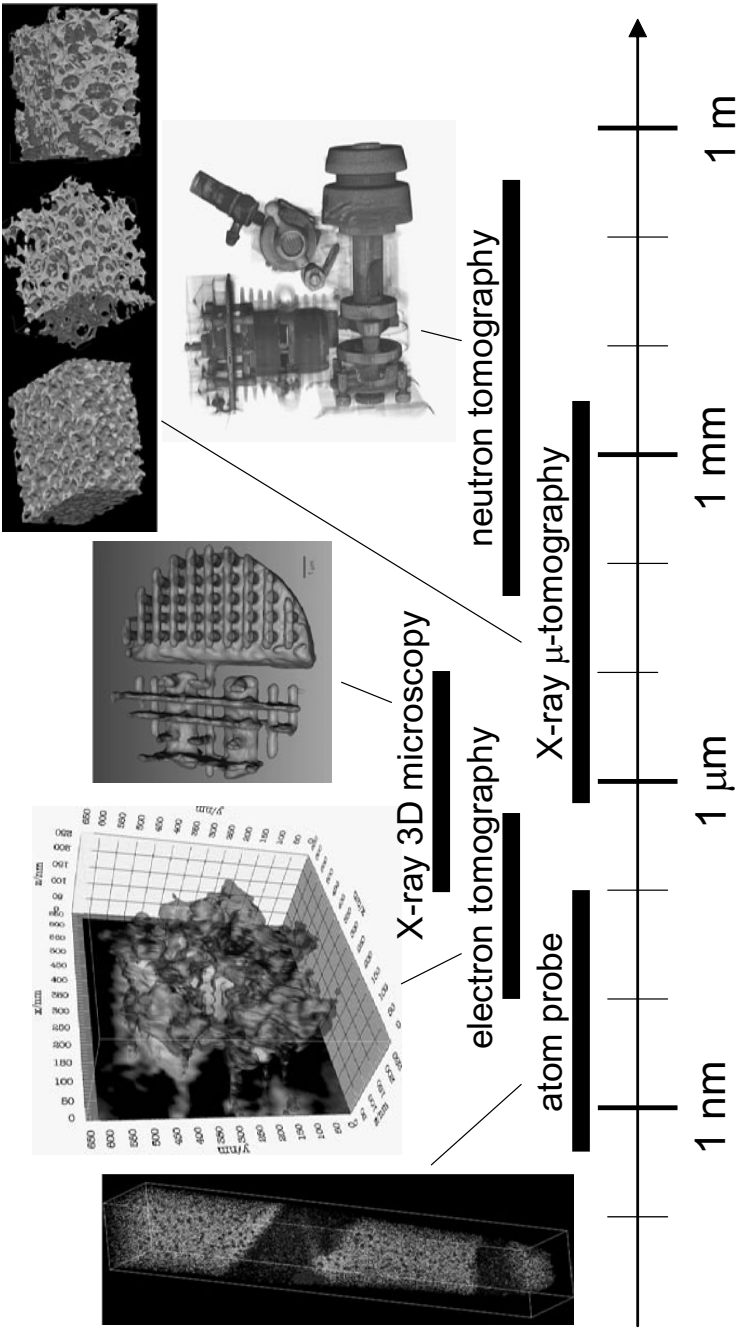


FIG. 1.8. Tomography over 9 orders of magnitude.

1.6 References

Bailey, D.L., Townsend, D.W., Valk, P.E., and Maisey, M.N. (2005). *Positron emission tomography*. Springer.

Banhart, J., Stanzick, H., Helfen, L., and Baumbach, T. (2001). Metal foam evolution studied by synchrotron radioscopy. *Applied Physics Letters*, **78**, 1152.

Blavette, D., Deconihout, B., Bostel, A., Sarrau, J.M., Bouet, M., and Menand, A. (1993). The tomographic atom probe: A quantitative three-dimensional nanoanalytical instrument on an atomic scale. *Review of Scientific Instruments*, **64**, 2911.

Cerezo, A., Godfrey, T.J., and Smith, G.D.W. (1988). Application of a position-sensitive detector to atom probe microanalysis. *Review of Scientific Instruments*, **59**, 862.

Cormack, A.M. (1963). Representation of functions by its line integrals with some radiological applications. *Journal of Applied Physics*, **34**, 2722.

Di Michiel, M., Merino, J.M., Fernandez-Carreiras, D., Buslaps, T., Honkimäki, V., Falus, P., Martin, T., and Svensson, O. (2005). Fast microtomography using high energy synchrotron radiation. *Review of Scientific Instruments*, **76**, 043702.

Gilmore, R.S. (1996). Industrial ultrasonic imaging and microscopy. *Journal of Physics D: Applied Physics*, **29**, 1389.

Hounsfield, G.N. (1972). A method and apparatus for examination of a body by radiation such as X or Gamma. British Patent 1283915.

Hounsfield, G.N. (1973). Computerized transverse axial scanning (tomography). *The British Journal of Radiology*, **46**, 1016.

Hsieh, J. (2003). *Computed tomography*. SPIE Press, Washington.

Iyer, H.M. and Hirahara, K. (1993). *Seismic tomography: theory and practice*. Kluwer.

Johnson, L., Keeney, B., Ross, G., Sadrozinski, H.F.-W., Seiden, A., Williams, D.C., Zhang, L., Bashkirov, V., Schulte, R.W., and Shahnazi, K. (2003). Initial studies on proton computed tomography using a silicon strip detector telescope. *Nuclear Instruments and Methods in Physics Research A*, **514**, 215.

Kai, Chua Chee, Fai, Leong Kah, and Chu-Sing, Lim (2004). *Rapid prototyping: principles and applications* (2nd edn). World Scientific, Singapore.

Kalender, W.A. (2000). *Computed tomography*. Publicis MCD Verlag, München.

Leitner, H., Clemens, H., Höring, S., Wanderka, N., Banhart, J., Staron, P., and Jamnig, B. (2004). Characterisation of precipitates in a stainless maraging steel by three-dimensional atom probe and small-angle neutron scattering. *Zeitschrift für Metallkunde*, **95**, 644.

Magerle, R. (2000). Nanotomography. *Physical Review Letters*, **85**, 2749.

Manke, I., Hartnig, C., Grünerbel, M., Lehnert, W., Kardjilov, N., Haibel, A., Hilger, A., and Banhart, J. (2007). Investigation of water evolution and

transport in fuel cells with high resolution synchrotron X-ray radiography. *Applied Physics Letters*, **90**, 174105.

McRobbie, D.W., Moore, E.A., and Graves, M.J. (2002). *MRI from Picture to Proton*. Cambridge University Press, Cambridge.

Plaskowski, A., Beck, M.S., Thorn, R., and Dyakowski, T. (1995). *Imaging industrial flows*. Institute of Physics, Bristol.

Radon, J. (1917). Über die Bestimmung von Funktionen durch ihre Integralwerte längs gewisser Mannigfaltigkeiten. *Berichte über die Verhandlungen der Königlichen Sächsischen Akademie der Wissenschaften*, **69**, 262.

Reimers, P., Goebbels, J., Weise, H.-P., and Wilding, K. (1984). Some aspects of industrial non-destructive evaluation by X and γ-ray computed tomography. *Nuclear Instruments and Methods in Physics Research*, **221**, 201.

Slichter, C.P. (1994). *Principles of magnetic resonance*. Springer, Berlin.

University of Manchester (2007). Industrial process tomography. `http://www.tomography.manchester.ac.uk/`. [Online; accessed 19 June 2007].

Wang, S., Zhang, X.-C., and Ferguson, B. (2004). Pulsed terahertz tomography. *Journal of Physics D: Applied Physics*, **37**, R1.

West, R. (2003, June). In industry, seeing is believing. *Physics World*.

Zaafarani, N., Raabe, D., Singh, R.N., Roters, F., and Zaefferer, S. (2006). Three-dimensional investigation of the texture and microstructure below a nanoindent in a Cu single crystal using 3D EBSD and crystal plasticity final element simulations. *Acta Materialia*, **54**, 1363.

2

SOME MATHEMATICAL CONCEPTS FOR TOMOGRAPHIC RECONSTRUCTION

Attila Kuba and Gabor T. Herman

Tomography is a procedure for imaging sections of an object. There are analogue methods to achieve this (Bocage, 1922), but in this chapter we deal only with computerized tomography in which a computational procedure is used to create cross-sectional images from their projections. The projections are usually taken with the help of some rays (e.g. X-rays, neutron rays) and can be thought of mathematically as the line integrals of the function representing the cross-section. The mathematical problem is to reconstruct the cross-section / function from its (noisy and incomplete) projections. In this chapter we present some of the mathematical concepts associated with computerized tomographic reconstruction from projections; more complete coverage can be found in standard books such as by Herman (1980) and Natterer (1986).

2.1 Foundations of reconstruction from projections

2.1.1 *Notation, definitions, the reconstruction problem*

We represent the cross-section to be reconstructed as a function $f(x, y)$. Then the projection of f along the vertical line $x = x_0$ is defined as

$$\int_{-\infty}^{\infty} f(x_0, y)\, dy.$$

More generally, let us define the projection of f taken from an arbitrary direction $\vartheta \in [0, \pi)$ as follows. Let (s, u) denote the co-ordinates of the point (x, y) in the co-ordinate system rotated by ϑ (see Fig. 2.1), i.e.

$$\begin{aligned} s &= x \cos\vartheta + y \sin\vartheta, \\ u &= -x \sin\vartheta + y \cos\vartheta. \end{aligned}$$

Then, the *projection of f taken from the direction ϑ* (or *the ϑ-projection of f*) is defined as

$$[\mathcal{R}f](s, \vartheta) = \int_{-\infty}^{\infty} f(x, y)\, du. \tag{2.1}$$

It can be said that the transform $\mathcal{R}$ defined by eqn 2.1 gives the ϑ-projections of f for any $\vartheta \in [0, \pi)$. The transform $\mathcal{R}$ is called the (2-dimensional) *Radon*

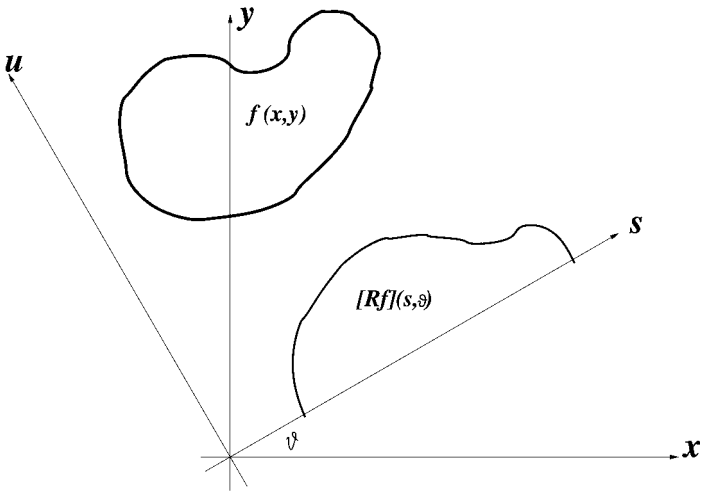


FIG. 2.1. Function f and its θ-angle projection.

transform of f, named after J. Radon who first studied this kind of transform (Radon, 1917).

Let us suppose for now that we have the projections of f for all directions, i.e. that $[\mathcal{R}f](s,\vartheta)$ is available for all $-\infty < s < \infty$ and $\vartheta \in [0,\pi)$. Then our task is to determine f from $\mathcal{R}f$, that is, to *invert* the Radon transform. In other words, the tomography problem can be expressed mathematically as the following *reconstruction problem*: Given a function $g(s,\vartheta)$, find an $f(x,y)$ such that

$$[\mathcal{R}f](s,\vartheta) = g(s,\vartheta) \quad \text{for all} \quad -\infty < s < \infty,\ \vartheta \in [0,\pi).$$

2.1.2 *Theoretical background to the solution of the reconstruction problem*

In order to give the inverse of the Radon transform, we have to introduce further transforms.

Let $g(s,\vartheta)$ be a real function defined on $(-\infty,\infty)\times[0,\pi)$. The *backprojection transform* $\mathcal{B}$ is applied to such functions and is defined as

$$[\mathcal{B}g](x,y) = \int_0^\pi g(x\cos\vartheta + y\sin\vartheta,\vartheta)\ d\vartheta, \tag{2.2}$$

for all points (x,y) in the plane. It is easy to see that in general $\mathcal{B}\mathcal{R}f \neq f$, that is, the backprojection transform is not the inverse of the Radon transform.

We need also the 1- and 2-dimensional Fourier transforms. The *1-dimensional* Fourier *transform* $\mathcal{F}_S$ is applied only to the first variable of the function $g(s,\vartheta)$ and it is defined as

$$[\mathcal{F}_S g](S,\vartheta) = \int_{-\infty}^{\infty} g(s,\vartheta)\, e^{-2\pi i s S} ds. \tag{2.3}$$

The *inverse* ${\mathcal{F}_S}^{-1}$ of $\mathcal{F}_S$ is defined by

$$[{\mathcal{F}_S}^{-1}g](s,\vartheta) = \int_{-\infty}^{\infty} g(S,\vartheta)\, e^{2\pi i s S} dS.$$

The *2-dimensional Fourier transform* $\mathcal{F}_2$ and its *inverse* ${\mathcal{F}_2}^{-1}$ are

$$[\mathcal{F}_2 f](X,Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\, e^{-2\pi i(xX+yY)} dx dy, \tag{2.4}$$

$$[{\mathcal{F}_2}^{-1} f](x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(X,Y)\, e^{2\pi i(xX+yY)} dX dY. \tag{2.5}$$

We need also the 2-dimensional inverse Fourier transform expressed in polar co-ordinates:

$$[{\mathcal{F}_2}^{-1} f](x,y) = \int_0^{\pi}\int_{-\infty}^{\infty} f(S\cos\vartheta, S\sin\vartheta)\, e^{2\pi i(x\cos\vartheta + y\sin\vartheta)S} |S| dS d\vartheta. \tag{2.6}$$

The (1-dimensional) *convolution* of two functions $g_1(s,\vartheta)$ and $g_1(s,\vartheta)$ is defined as

$$[g_1 * g_2](s,\vartheta) = \int_{-\infty}^{\infty} g_1(\sigma,\vartheta) g_2(s-\sigma,\vartheta) d\sigma.$$

In the following we make use of the well-known *convolution theorem*:

$$\mathcal{F}_S\,(g_1 * g_2) = \mathcal{F}_S g_1 \cdot \mathcal{F}_S g_2, \tag{2.7}$$

$$\mathcal{F}_S^{-1}\,(g_1 \cdot g_2) = \mathcal{F}_S^{-1} g_1 * \mathcal{F}_S^{-1} g_2. \tag{2.8}$$

The *projection theorem* (also known as the *Fourier slice theorem*) gives the connection between the Radon and Fourier transforms (see Fig. 2.2):

$$[\mathcal{F}_S \mathcal{R} f](S,\vartheta) = [\mathcal{F}_2 f](S\cos\vartheta, S\sin\vartheta). \tag{2.9}$$

This is easily proved:

$$\begin{aligned}
[\mathcal{F}_S \mathcal{R} f](S,\vartheta) &= \int_{-\infty}^{\infty} [\mathcal{R} f](s,\vartheta)\, e^{-2\pi i s S} ds \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(s\cos\vartheta - u\sin\vartheta, s\sin\vartheta + u\cos\vartheta) du\, e^{-2\pi i s S} ds \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y) e^{-2\pi i(x\cos\vartheta + y\sin\vartheta)S} dx dy \\
&= [\mathcal{F}_2 f](S\cos\vartheta, S\sin\vartheta).
\end{aligned} \tag{2.10}$$

The inversion of the (2-dimensional) Radon transform can be written as

$$\mathcal{R}^{-1} g = \mathcal{B}\mathcal{F}_S^{-1}[\mathrm{abs}\cdot \mathcal{F}_S g],$$

where 'abs' is a function of two variables and is defined by $\mathrm{abs}(S,\vartheta) = |S|$. Thus, as $F_S g$ is a function of two variables (see eqn 2.3), $\mathrm{abs}\cdot F_S g(S,\vartheta) = |S|[F_S g](S,\vartheta)$.

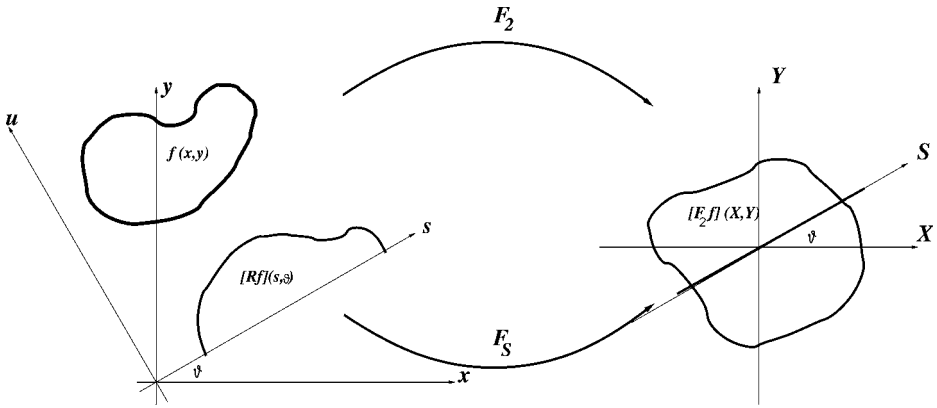


FIG. 2.2. The illustration of the projection theorem.

The formal proof of eqn 2.1.2 is the following. Let us start the derivation with the identity

$$f(x,y) = [{\mathcal{F}_2}^{-1}\mathcal{F}_2 f](x,y).$$

Using the polar co-ordinate form of the inverse 2-dimensional Fourier transform eqn 2.6, we can write that

$$f(x,y) = \int_0^{\pi}\int_{-\infty}^{\infty} [\mathcal{F}_2 f](S\cos\vartheta, S\sin\vartheta)\, e^{2\pi i(x\cos\vartheta + y\sin\vartheta)S}|S| dS d\vartheta.$$

Applying the projection theorem eqn 2.9, we get

$$\begin{aligned} f(x,y) &= \int_0^{\pi}\int_{-\infty}^{\infty} [\mathcal{F}_S\mathcal{R}f](S,\vartheta)\, e^{2\pi i(x\cos\vartheta + y\sin\vartheta)S}|S| dS d\vartheta \\ &= \int_0^{\pi}\left(\int_{-\infty}^{\infty} [\mathcal{F}_S\mathcal{R}f](S,\vartheta)\,|S|\, e^{2\pi i(x\cos\vartheta + y\sin\vartheta)S} dS\right) d\vartheta \qquad (2.11) \\ &= \int_0^{\pi}\left(\int_{-\infty}^{\infty} (\mathrm{abs}(S,\vartheta)\,[\mathcal{F}_S\mathcal{R}f](S,\vartheta))\; e^{2\pi i(x\cos\vartheta + y\sin\vartheta)S} dS\right) d\vartheta. \end{aligned}$$

The inner integral is just the inverse 1-dimensional Fourier transform of the product $\mathrm{abs}\cdot\mathcal{F}_S\mathcal{R}f$, that is

$$\begin{aligned} f(x,y) &= \int_0^{\pi} [{\mathcal{F}_S}^{-1}(\mathrm{abs}\cdot\mathcal{F}_S\mathcal{R}f])\,(x\cos\vartheta + y\sin\vartheta,\vartheta) d\vartheta \\ &= \left[\mathcal{B}{\mathcal{F}_S}^{-1}(\mathrm{abs}\cdot\mathcal{F}_S\mathcal{R}f)\right](x,y). \qquad (2.12) \end{aligned}$$

In the last step, we used the backprojection transform defined by eqn 2.2. The expression eqn 2.12 is a *reconstruction formula*, since it gives a mathematical expression for recovering a function f from its Radon transform $\mathcal{R}f$.

2.2 Examples of reconstruction methods

In this section we introduce two widely used reconstruction approaches: filtered backprojection and the algebraic reconstruction techniques (and also some variants of these approaches). This is followed by a discussion of a relatively new development: discrete tomography. We warn the reader that there is a very serious terminological problem in the field of image reconstruction from projections: many different actual algorithms are identified in the literature by the same name (and, conversely, some algorithms go under a variety of names). One has to be careful in referring to the results of a paper that identifies an algorithm by its name but without giving the exact specification that is understood by that name in that paper, since the reader's assumption as to what exactly that algorithm does may be incorrect!

2.2.1 *Filtered backprojection*

One of the most frequently used reconstruction methods in tomography is *filtered backprojection* (FBP) (Ramachandran and Lakshminarayanan, 1971). It can be derived from the reconstruction formula eqn 2.12. Since

$$f = \mathcal{B}\left({\mathcal{F}_S}^{-1}\left(\mathrm{abs}\cdot\mathcal{F}_S\mathcal{R}f\right)\right), \tag{2.13}$$

we can apply the convolution theorem eqn 2.8 to the inverse 1-dimensional Fourier transform of the product $\mathrm{abs}\cdot\mathcal{F}_S\mathcal{R}f$ and obtain

$$f = \mathcal{B}\left({\mathcal{F}_S}^{-1}\mathrm{abs} * {\mathcal{F}_S}^{-1}\mathcal{F}_S\mathcal{R}f\right) = \mathcal{B}\left({\mathcal{F}_S}^{-1}\mathrm{abs} * \mathcal{R}f\right). \tag{2.14}$$

The problem with eqn 2.14 is that ${\mathcal{F}_S}^{-1}\mathrm{abs}$ does not exist (there is no real-valued function of a real variable whose Fourier transform is 'abs', and so in practice it has to be replaced by some function whose Fourier transform is 'abs' in only some approximate sense. Ignoring this matter for the moment, we see that eqn 2.14 can be implemented in two steps: first the convolution

$$g'(s,\vartheta) = [{\mathcal{F}_S}^{-1}\mathrm{abs} * \mathcal{R}f](s,\vartheta) \tag{2.15}$$

is computed for all $\vartheta \in [0,\pi)$ and then the function g' is backprojected to produce

$$f(x,y) = [\mathcal{B}g'](x,y). \tag{2.16}$$

In the rest of this subsection we consider the problems connected to the fact that instead of all (infinitely many) line integrals of the function to be reconstructed, we can measure only finitely many. Let us suppose that the projection data are available from N projections along M lines in each projection (for simplicity of the following formulas we assume that M is odd), in discrete positions $s = m\Delta s$ and $\vartheta = n\Delta\vartheta$. More precisely, the projection $g = \mathcal{R}f$ of f is sampled as

$$g_n(m) = g(s_m,\vartheta_n) = g(m\Delta s, n\Delta\vartheta), \qquad \begin{array}{l} m = -(M-1)/2,\ldots,(M-1)/2, \\ n = 0,1,\ldots,N-1, \end{array} \tag{2.17}$$

where $\Delta s > 0$ denotes the sampling distance of s and $\Delta\vartheta = \pi/N$. We now need to approximate the functions g' and f using the discrete versions of the operations given in eqn 2.15 and eqn 2.16.

The convolution in eqn 2.15 can be replaced by the *discrete convolution* of two vectors $h(m)$ and $g_n(m)$, where h is a discrete approximation of ${\mathcal{F}_S}^{-1}\text{abs}$, providing

$$g'_n(m\Delta s) = \sum_{i=-(M-1)/2}^{(M-1)/2} h(m-i)g_n(i), \tag{2.18}$$

for $m = -(M-1)/2, -(M-1)/2+1, \ldots, (M-1)/2$. Several discrete filters h can be applied here. For example, Shepp and Logan (1974) suggest

$$h(m) = \begin{cases} \dfrac{1}{4\Delta s}\,, & \text{if } m = 0, \\ 0\,, & \text{if } m \text{ is even and non-zero,} \\ \dfrac{-1}{\pi^2 m^2 \Delta s}\,, & \text{otherwise.} \end{cases}$$

The backprojection operator given in eqn 2.16 can be replaced by its discrete version

$$[\mathcal{B}g'](x,y) \approx \frac{\pi}{N} \sum_{n=0}^{N-1} g'_n \left(x\cos(n\frac{\pi}{N}) + y\sin(n\frac{\pi}{N}) \right). \tag{2.19}$$

Note that eqn 2.18 gives the values of g'_n only at the discrete points of $\{m\Delta s\}$, but for the numerical implementation of eqn 2.19 we need its values at the points $\{x\cos(n\pi/N) + y\sin(n\pi/N)\}$, and hence some interpolation is necessary. In most cases, linear interpolation provides satisfactory results.

In Herman (1980), FBP is considered to be one of the family of *transform methods,* whose distinguishing feature is that they are numerical implementations of mathematical formulas based on the material in Section 2.1. A particularly simple example of this is the *Fourier method*, see Section 9.2 of Herman (1980), also known as *direct Fourier inversion*, which is a straightforward application of the projection theorem, as illustrated in Fig. 2.2. In its numerical implementation, the 1-dimensional fast Fourier transform (FFT) is used to estimate from the given projection data $\mathcal{F}_2 f$ a finite number of equally placed points on a finite number of lines through the origin. Then, by interpolating, the values of $\mathcal{F}_2 f$ are estimated on a regular square grid, after which we can apply the 2-dimensional inverse FFT, which is an efficient implementation of eqn 2.5, to recover f. A problem with this method is that if a simple interpolation (such as bilinear) is used, then the reconstruction quality is much inferior to that of FBP. More sophisticated methods using the same basic principle include the technique of *linograms* (Edholm and Herman, 1987, Edholm *et al.*, 1988) that also uses FFTs but does not require any interpolations for special modes of data collection and *weighted backprojection* (WBP) (Radermacher, 1988) that allows,

in principle, reconstruction of a function of three variables from arbitrarily oriented 2-dimensional projections by compensating for the shape of blurring that one would obtain by simple backprojection of the projections.

2.2.2 *Algebraic reconstruction techniques*

The *algebraic reconstruction techniques* (ART) take a very different approach (Gordon *et al.*, 1970). The 2-dimensional cross-section to be reconstructed is represented as a *digital image*, i.e. as a linear combination of a finitely many, say J, fixed *basis functions*. A popular way of defining the basis functions is by using a finite regular array of unit squares (called *picture elements or pixels*) in the plane, each giving rise to a single basis function that has value 1 in the pixel and value 0 elsewhere (see Fig. 2.3). Clearly, any digital image that uses this kind of basis functions will be constant valued within the pixels. Given the fixed basis functions, any digital image is completely described by a column vector $x = (x_1, \ldots, x_J)^T$, in which x_j is the coefficient of the jth basis function in the linear combination. Similarly, a collection of I measurements can be described by a column vector $b = (b_1, \ldots, b_I)^T$. For a digital image f, it is easily seen to be the case that if b_i is the line integral in eqn 2.1 for the ith among the I lines for which measurements are taken, then

$$\sum_{j=1}^{J} a_{ij} x_j = b_i, \quad \text{for } i = 1, \ldots, I, \tag{2.20}$$

which can be abbreviated as

$$Ax = b, \tag{2.21}$$

where a_{ij} is the integral along the ith line of the jth basis function (see Fig. 2.3). The linear equation system eqn 2.21 is typically of a large size; e.g. if the image is based on 1000×1000 pixels, i.e. $J = 10^6$, and we have 1000 projections with 1000 measurements in each, i.e. $I = 10^6$, then the size of A is $10^6 \times 10^6$. On the other hand, the matrix A is typically very sparse; e.g. in the example just discussed we will have that, for $i = 1, \ldots, I$, the number of j for which $a_{ij} \neq 0$ is always less than 2000. In ART one makes essential use of this fact.

In the previous paragraph we presented ART for the situation in which a 2-dimensional object is to be reconstructed (this can be referred to as 2D-ART), but the same basic mathematics is applicable to reconstructing 3-dimensional objects (3D-ART). Indeed, mathematically speaking, ART can be considered as a general approach to solving systems of equations, applicable to many different physical interpretations of the meaning of those equations.

The ART approach to solving the system of equations 2.20, or eqn 2.21, produces a sequence $x^{(0)}, x^{(1)}, x^{(2)}, \ldots$ of J-dimensional column vectors. A typical ART method is one in which, for $k \geq 0$,

$$x^{(k+1)} = x^{(k)} + \frac{b_{i_k} - \langle a_{i_k}, x^{(k)} \rangle}{\|a_{i_k}\|^2} a_{i_k}, \tag{2.22}$$

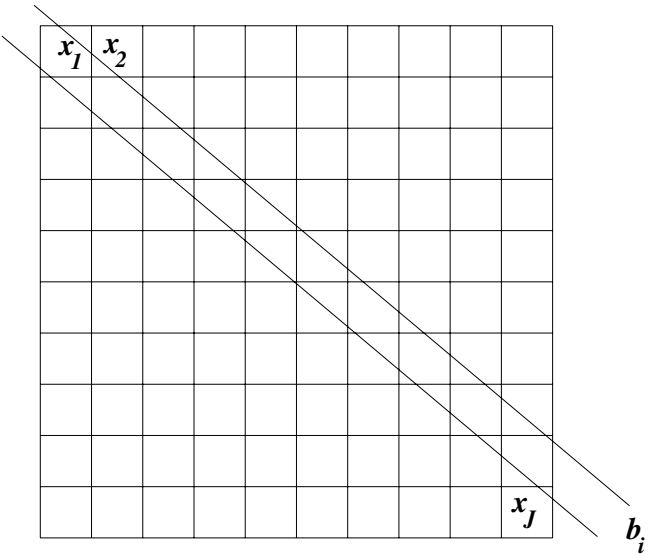


FIG. 2.3. The object to be reconstructed, represented as a digital image using pixels, and two lines for which we have measurements.

where i_k is $k \pmod I + 1$. (As usual, the *inner product* $\langle u, v\rangle$ of two vectors u and v is defined as the sum of the products of their matching components; e.g. the left-hand side of eqn 2.20 is the inner product of a vector $a_i = (a_{i1}, \ldots, a_{iJ})^T$ with the vector x. Also the *norm* of a a vector u is defined as $\|u\| = \sqrt{\langle u, u\rangle}$.) It is shown in Herman (1980), p. 188, that if eqn 2.20 has a solution at all and $x^{(0)}$ is selected to be the vector of all zeroes, then the sequence $x^{(0)}, x^{(1)}, x^{(2)}, \ldots$ converges to the solution of eqn 2.20 that among all the solutions has the smallest norm. From the point of view of computing, what is good about such a method is that the cost of performing the iterative step eqn 2.22 is proportional not to J, which is large, but to the number of non-zero components in a_{i_k}, which is typically much smaller (see above). In fact, in some software implementations of ART – e.g. CUNY Institute for Software Design and Development (2007) – the matrix A of eqn 2.21 is never stored at all, but rather the locations and sizes of the non-zero components of a_{i_k} are calculated based on the geometry of the situation (see Fig. 2.3) just before executing eqn 2.22.

In practice, it is likely that the system eqn 2.21 will not have a solution, both because a real cross-section can be only approximated by a digital image (and hence the sum in eqn 2.20 can only approximate the integral in eqn 2.1) and because of measurement errors (the b_i provided by an instrument will only be an approximation of the line integral of f). Due to this, it is better to aim at finding an x for which eqn 2.21 is satisfied only approximately. There are a number of ways that the notion of 'approximately' can be made mathematically precise,

and for many of them there is an ART method that converges to a solution with the specified property, as illustrated by the following example.

Let us say that we would like to find an x for which the error in the ith equation of eqn 2.20 is not greater than some non-negative ϵ_i, more formally that

$$|\sum_{j=1}^{J} a_{ij}x_j - b_i| \leq \epsilon_i, \tag{2.23}$$

for $i = 1, 2, \ldots, I$. The system of inequalities eqn 2.23 can be rewritten as

$$\sum_{j=1}^{J} a_{ij}x_j \leq b_i + \epsilon_i, \tag{2.24}$$

$$-\sum_{j=1}^{J} a_{ij}x_j \leq -b_i + \epsilon_i, \tag{2.25}$$

for $i = 1, 2, \ldots, I$; that is in the form

$$\sum_{j=1}^{J} c_{ij}x_j \leq d_i, \tag{2.26}$$

for $i = 1, 2, \ldots, 2I$. An ART-type reconstruction method for the inequality system eqn 2.26 is as follows. Let, for $i = 1, 2, \ldots, 2I$, $c_i = (c_{i1}, \ldots, c_{iJ})^T$. Using $i_k = k \pmod{2I} + 1$, we define

$$x^{(k+1)} = \begin{cases} x^{(k)}\,, & \text{if } \left\langle c_{i_k}, x^{(k)} \right\rangle \leq d_{i_k}, \\ x^{(k)} + \lambda^{(k)} \frac{d_{i_k} - \left\langle c_{i_k}, x^{(k)} \right\rangle}{\|c_{i_k}\|^2} c_{i_k}\,, & \text{otherwise,} \end{cases} \tag{2.27}$$

where $\lambda^{(k)}$ is a real number, called a *relaxation parameter*. It is proved in Section 16.8 of Herman (1980) that, irrespective of the choice of $x^{(0)}$, the sequence of $x^{(0)}, x^{(1)}, x^{(2)} \ldots$ generated by eqn 2.27 with

$$0 < \delta_1 \leq \lambda^{(k)} \leq \delta_2 < 2,$$

for some δ_1 and δ_2, converges to a solution of eqn 2.26, provided only that there is a solution at all.

Such mathematical results are true independently of the choice of the basis functions. In practice, it turns out that the basis functions illustrated above (those based on pixels in the plane and their analogues, called voxels, in three-dimensional space) are far from optimal. A much superior choice is to use the generalized Kaiser–Bessel window functions, usually referred to as *blobs*, proposed for this purpose by Lewitt (1990). Such superiority has been repeatedly demonstrated in the literature, an example from positron emission tomography is Matej *et al.* (1994).

In Herman (1980), ART is considered to be one of the family of *series expansion methods*, whose distinguishing feature is that they aim to find the coefficients of the expansion of the function to be reconstructed into a linear combination of fixed basis functions. Many series-expansion methods can be considered as special cases of *block-ART*, first introduced by Eggermont *et al.* (1981). Here, we describe a simplified variant of it, based on the way it is presented in Carazo *et al.* (2006).

Assume that the I equations of the system eqn 2.20 can be divided into N blocks of M equations each, and so $I = MN$. The iterative ART step in eqn 2.22 is replaced by

$$x^{(k+1)} = x^{(k)} + \lambda^{(k)} \sum_{i=n_k M+1}^{(n_k+1)M} \frac{b_i - \left\langle a_i, x^{(k)} \right\rangle}{\|a_i\|^2} a_i, \tag{2.28}$$

where n_k $(0 \leq n_k < N)$ is the index of the block to be used in the kth iterative step and $\lambda^{(k)}$ is a relaxation parameter. Note that if $M = 1$ and $\lambda^{(k)} = 1$, for all k, then this is just eqn 2.22. If $M = I$ (and hence $N = 1$), then this is what is referred to as a *fully simultaneous* iterative method (see Censor and Zenios (1997), p. 100), and is closely related to various variants of the *simultaneous iterative reconstruction technique* (SIRT), introduced by Gilbert (1972). A frequently used in-between case is when the blocks are formed by all the equations associated with a single projection (in which case M and N are exactly as they are defined in Section 2.2.1), the *simultaneous algebraic reconstruction technique* (SART) of Andersen and Kak (1984) is an approach of this type. A straightforward implementation of eqn 2.28 loops through the various values of i, from the lower limit to the upper one. It has been reported in Herman and Meyer (1993) that with such an implementation using block size $N = 1$ (i.e. ordinary ART with a relaxation parameter) is superior to using larger block sizes, as measured by the accuracy of the early reconstructions in the iterative process. (Warning: This result is valid having carefully selected the indexing, by i, of the equations and the relaxation parameters. The details are explained in Herman and Meyer (1993).) However, there are more sophisticated ways of implementing eqn 2.28, especially if blobs are used, either in software – e.g. by using footprints as in Matej and Lewitt (1996) – or hardware – e.g. by using texture mapping as in Mueller and Yagel (1996). Using such implementations, it is most efficient to use blocks that are associated with single projections. Going all the way to a fully simultaneous method such as SIRT is not advisable; careful studies, e.g. by Sorzano *et al.* (2001), demonstrated that the quality that can be achieved is not improved by going to SIRT, but the computational cost needed to achieve that quality becomes much greater.

Variations of such series-expansion methods can be designed with the aim of optimizing various functions defined on the vector x, including things such as entropy, likelihood, *etc.*; see Censor and Zenios (1997) and Herman (1980) and also the Section 2.2.3 that follows. A development that became particularly pop-

ular in the emission tomography community was initiated by Shepp and Vardi (1982): it involves the use of the general technique called *expectation maximization to maximize likelihood*, and it is referred to as *MLEM*. In our hands, the performance of ART was found to be superior, or at worst similar, to that of MLEM, even in some emission-tomography-oriented tests (Herman and Meyer, 1993, Matej *et al.*, 1994).

A recent careful comparative study in the context of electron microscopy of block-ART (with each projection providing a single block) with blobs and WBP is reported in Carazo *et al.* (2006). The general conclusion there is that the reconstruction quality of block-ART with blobs is slightly, but statistically significantly, better than that of WBP. This conclusion is similar to those of earlier comparative studies.

2.2.3 *Discrete tomography*

There are situations in which, in addition to the projection data, we have some further information about the object to be reconstructed. An example is when we know that the object consists of homogeneous regions of known materials, which is often the case in non-destructive testing. Such information can be included into the reconstruction process, and this makes it possible to use fewer projections than would be needed otherwise to achieve the same quality reconstruction.

If the object consists of homogeneous regions of known materials, then it can be represented by a function that takes its values from a known finite set D. *Discrete tomograph*y (DT) is the field of research dealing with problems related to the reconstruction of functions from their projections when the functions has such a known range D (Herman and Kuba, 1999, Herman and Kuba, 2007).

The simplest instance is when the object to be reconstructed is binary, that is, $D = \{0, 1\}$. In other words: there are only two materials in the object, for example, steel and air. Binary objects can be represented in different ways depending on their complexity. When the object is the constellation of regular geometric forms such as circles, as few as 2 to 5 projections have been shown to be sufficient for a good-quality reconstruction (Kuba *et al.*, 2005). More general shapes can be described using polygons and deformable models (Mohammad-Djafari and Sauer, 1998). The most general class of objects needs to be represented by binary matrices. The reconstruction of binary matrices from their projections is a well-understood area with powerful theoretical foundations, reconstruction algorithms, and applications (Herman and Kuba, 1999, Herman and Kuba, 2007).

In many cases, a DT reconstruction problem is reformulated as an optimization problem. Indeed, the same can be said for the general reconstruction problem, especially in conjunction with the use of digital images. Note that one way of defining the DT problem is: find the digital image that satisfies the given projections, in at least some approximate sense, in which all the coefficients in the linear combination of basis functions come from D. In the binary case this results in the linear equation system eqn 2.21 but with the binary constraint:

$$Ax = b, \qquad x \in \{0, 1\}^J. \tag{2.29}$$

We note that such a system of equations may not have a solution and that, even if it does have one, an exact solution may not be a desirable one since it would fit the noise in the measurements b. So, instead of demanding an exact solution, we request that some cost function C be minimized. We now mention three specific approaches that have been used.

2.2.3.1 *Regularization* One general form of the cost function is

$$C(x) = \|Ax - b\|^2 + \Phi(x),$$

where $\Phi(x)$ is a real-valued function whose value indicates how undesirable a solution x is from the point of view of our application. One approach is to select $\Phi(x)$ to be $\gamma\|x\|^2$, which results in

$$C(x) = \|Ax - b\|^2 + \gamma\|x\|^2, \tag{2.30}$$

where γ is a positive real number called the *regularization parameter*. The minimizer of eqn 2.30 over the set of real-valued vectors is known to be

$$\left(A^T A + \gamma I\right)^{-1} A^T b, \tag{2.31}$$

where A^T denotes the transpose of A. A practical difficulty here is that a very large matrix may have to be inverted or, as one is more likely to do instead, a large system of linear equations will have to be solved numerically. More importantly, the x provided by eqn 2.31 will not in general have the property that $x_j \in D$, for all j, and so it is not a solution of the DT reconstruction problem.

An alternative approach is provided by *simulated annealing* (SA) (Kirkpatrick *et al.*, 1983), which is a *stochastic* (also called *Monte Carlo*) *minimization* algorithm derived from thermodynamics. It consists of sampling the state variables (in our case these can be the coefficients assigned to the pixels) in order to reach the thermodynamic equilibrium at a given temperature. By lowering the temperature, the system can be frozen at a solution of minimal energy. The essential point for us here is that we start with an x satisfying $x_j \in D$, for all j, and all through the process we allow only such changes that preserve these constraints. There are important tricks for efficacious implementation of SA, involving the choices of the initialization, the annealing schedule, and the stopping criterion. Since this is a potentially very time consuming method, much work has been done to speed it up (Vardi *et al.*, 2001). Such Monte–Carlo approaches have been recently applied to reconstruction of polycrystals (Alpers *et al.*, 2006), see also Chapter 9.

2.2.3.2 *Probabilistic models* Sometimes, probabilistic models can be used. Suppose that we know the probability $L(b\,|\,x)$ of measuring projection b if the image is x. A *maximum likelihood* (ML) solution is an x for which $L(b\,|\,x)$ attains a maximal value. Suppose further that we know $\Pi(x)$, the (prior) distribution of possible objects. Then, the posterior distribution of x, having observed b, is

$$P(x \mid b) = \frac{\Pi(x)L(b \mid x)}{P(b)},$$

where $P(b)$ is the probability of observing the projection data b. A *maximum a-posteriori* (MAP) solution of the DT reconstruction problem is an x that maximizes

$$\Pi(x)L(b \mid x), \tag{2.32}$$

(the $P(b)$ has been dropped, since it is independent of x). The maximization of eqn 2.32 can be difficult. Running an optimization method such as SA in such cases may be very time consuming for objects with sizes used in real applications.

Another estimator that has been applied is the *minimum mean square error* (MMSE) estimator that minimizes the posterior mean (the expected object over the posterior distribution), defined as

$$E[x \mid b] = \sum P(x \mid b)x.$$

The calculation of the MMSE estimate is also difficult because of the complexity of the function $P(x \mid b)$. For additional information about these estimators and approaches, see Chan *et al.* (1999).

2.2.3.3 *Linear integer programing* This kind of method can be applied if the cost function C is a linear function of x; for a collection of examples, see Herman and Kuba (2006). In most cases we are interested in a binary solution. One method is to apply a linear programming algorithm to find a solution x in $[0, 1]^J$ and then round this solution so that it (approximately) satisfies eqn 2.29.

2.3 Some problems associated with the application of reconstruction algorithms

In this section, we mention some of the practical difficulties that arise when applying a reconstruction algorithm to real data. One important relevant topic that we do not treat here is that of the noise in the data: the nature of such noise depends on the kind of rays that are used for collecting the projection data. The chapter on X-ray tomography describes how the problem of noise (and other problems) can be treated in practice, see Section 5.4.

2.3.1 *Resolution and sampling*

Sometimes, the resolution of a digital image is defined as the number of pixels it contains. Accepting this definition, from the model of linear equation system eqn 2.20, we get a very simple relation between the resolution (J) and the number of (independent) projection data (I):

$$J \leq I,$$

that is, the number of pixels should not be greater than the number of measurements. For example, for the reconstruction of a 100×100 image from 180

projections, we need at least $100 \times 100/180 \approx 56$ measurements per projection. However, this is a somewhat simplistic point of view: we have already mentioned in Section 2.2.3 that having prior information about the kind of image that we need to reconstruct sometimes allows us to achieve useful reconstructions from a relatively small number of projection data items.

Another way of defining resolution is connected with the size of small details that can be identified in the reconstruction (e.g. are two nearby points distinguishable?). This has a strong relation with the frequencies available in the Fourier transform: high-frequency components are responsible for small details in the image. For this reason, resolution can be defined as a function of the maximal available frequency. A reasonably detailed discussion between this approach to resolution and the associated sampling requirements is given in Natterer (1986), in the next paragraph we give the flavour of one aspect of this.

For any $W > 0$, the *W-bandlimited version* of f is obtained by applying an approximation of the 2-dimensional inverse Fourier transform expressed in polar co-ordinates, eqn 2.6, to $\mathcal{F}_2 f$. The approximation is the replacement of the infinite inner integral in eqn 2.6 by an integral from $-W$ to W (equivalent to removing from f all frequencies above W). The projection theorem eqn 2.9 implies that, in order to recover the W-bandlimited version of f, we need the information in the projection data up to the bandlimit W. By the well-known Shannon sampling theorem, this implies that the Δs in eqn 2.17 must satisfy

$$\Delta s \leq \frac{1}{2W}.$$

2.3.2 *Incomplete data*

Let $r > 0$ be such that the function f to be reconstructed is zero-valued outside the circle of radius r around the origin. We say that the projection data $g(s, \vartheta)$ (possibly in discretized form) is incomplete if it is not given on all of the domain $[0, \pi) \times (-r, r)$. This situation is quite realistic in industrial applications. For example, the object can be too long in a certain direction and from that direction the collection of projection data is practically impossible. Another case of incomplete data is when the object to be reconstructed is too large, larger than the field of view of the imaging system. This is the case if the segment of the detectors, say $[-\rho, \rho]$, is not long enough to contain the support of the projections, i.e. if $\rho < r$, and we are interested in the reconstruction of only that part of the object that is in the circle around the origin with radius ρ (see Fig. 2.4). This is sometimes referred to as the *interior problem.*

Another case of incomplete data is the 'missing wedge' problem that occurs when a sample cannot be rotated through a full 180° angle, e.g. because it is a sheet or film. This happens frequently in electron tomography, see Section 11.2.2 and Figs. 11.1 and 11.2, where TEM foils can hardly be rotated by more than 140° to 150°, but also in industrial tomography of large flat components.

The ART methods and the FBP algorithm perform differently in the case of incomplete data. Based on an earlier stated result, we see that ART can be

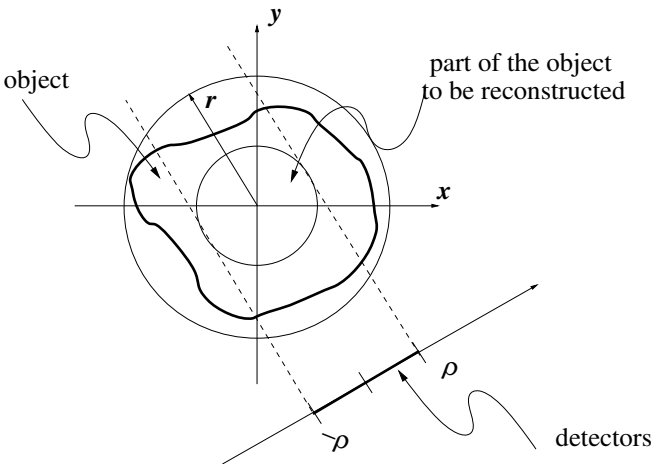


FIG. 2.4. The interior problem: projection data are available only for an inner part of the object to be reconstructed.

used to provide us with (an approximation to) the minimal norm solution of the interior problem. Images reconstructed by FBP have artifacts (Natterer, 1986) that show up mainly in the neighbourhood of lines that are tangents to the circle of radius ρ outside which the values of $\mathcal{R}f$ are missing. A possible approximate solution of the interior problem (Natterer, 1986) is to find a complete extension g^c of the projection data g such that there is a solution f^c of the reconstruction problem $\mathcal{R}f^c = g^c$. Then, by reconstructing such an f^c, we know that

$$|(f(x,y) - f^c(x,y) - K| \leq C\|g - g^c\|,$$

where K and C are an unknown and a known constant, respectively. The inequality shows that if we are interested in the changes of the values of f (rather than in f itself), then we can get them from f^c. A severe case of the interior problem and a demonstration that useful reconstructions can be obtained from such data are illustrated in Figs. 14.9 and 14.11 of Herman (1980).

2.3.3 *Beam hardening*

In the discussion to this point we have ignored the physical meaning of the values of the function that we are reconstructing from its projections. For many applications, it is supposed to be the *absorption coefficient* that indicates how much the radiation is absorbed at that point during the physical measurement process. However, in many cases, this absorption depends not only on the material at the point, but also on the energy spectrum of the ray. To make this more precise, it is not sufficient to consider a single function of the spatial variables x and y, but we need to be aware that there is a different function f_E for every energy level E. Under these circumstances, the physically correct model for the total detected intensity is

$$O(s,\vartheta) = \int I(E) e^{-[\mathcal{R}f_E](s,\vartheta)} dE, \tag{2.33}$$

where $I(E)$ is the intensity of the radiation of energy E at the source. We may declare that all that we wish to reconstruct is $f_{\bar{E}}$ for some typical energy level $\bar{E}$, but to do so from data provided by the model eqn 2.33 is much more difficult than simply inverting the Radon transform. As the rays traverse an object, the lower-energy photons tend to be attenuated more than the higher-energy ones, resulting in the phenomenon referred to as *beam hardening*. If it is ignored, a 'cupping' in the reconstructed image will be seen; see Fig. 5.9 of Herman (1980). Beam hardening can result in serious inaccuracies in the presence of high-absorption structures in the object.

Many methods have been suggested (Herman, 1979, Jennings, 1988, Joseph and Spittal, 1978) to overcome this problem. One of them is the linearizing of eqn 2.33, in which a polynomial is estimated based on physical experiments to allow us to map the actual measurement $O(s,\vartheta)$ into an estimate of $\mathcal{R}f_{\bar{E}}$. This linearizing procedure with polynomials has the advantage that the polynomial coefficients are easy to calculate and, once they are calculated, it is easy to correct for the beam-hardening effect.

Figure 5.10, shows an example of beam hardening. In synchrotron X-ray or neutron tomography beam hardening can be avoided by using monochromatic radiation.

2.4 Summary

We have indicated the mathematical foundations of tomographic reconstruction. Further details can be found in the references that follow.

2.5 References

Alpers, A., Poulsen, H.F., Knudsen, E., and Herman, G.T. (2006). A discrete tomography algorithm for improving the quality of three-dimensional X-ray diffraction grain maps. *Journal of Applied Crystallography*, **39**, 582.

Andersen, A.H. and Kak, A.C. (1984). Simultaneous algebraic reconstruction technique (SART): A superior implementation of the ART algorithm. *Ultrasonic Imaging*, **6**, 81.

Bocage, A.E.M. (1922). Procédé et dispositif de radiographie sur plaque en mouvement. *Brevet francais*, **534**, 464.

Carazo, J-M., Herman, G.T., Sorzano, C.O.S., and Marabini, R. (2006). Algorithms for three-dimensional reconstruction from the imperfect projection data provided by electron microscopy. In *Electron tomography: methods for three-dimensional visualization of structures in the cell* (2nd edn) (ed. J. Frank), pp. 217. Springer, New York.

Censor, Y. and Zenios, S.A. (1997). *Parallel optimization: theory and algorithms*. Oxford University Press, New York.

Chan, M.T., Herman, G.T., and Levitan, E. (1999). Probabilistic modeling of discrete images. In *Discrete tomography. Foundations, algorithms, and applications* (ed. G. Herman and A. Kuba), pp. 213. Birkhäuser, Boston.

CUNY Institute for Software Design and Development (2007). Snark05 [programming system]. `http://www.snark05.com`. [Online; accessed 26 April 2007].

Edholm, P.R. and Herman, G.T. (1987). Linograms in image reconstruction from projections. *IEEE Transactions on Medical Imaging*, **6**, 301.

Edholm, P.R., Herman, G.T., and Roberts, D.A. (1988). Image reconstruction from linograms - Implementation and evaluation. *IEEE Transactions on Medical Imaging*, **7**, 239.

Eggermont, P.P.B., Herman, G.T., and Lent, A. (1981). Iterative algorithms for large partitioned linear systems with applications to image reconstruction. *Linear Algebra and its Applications*, **37**, 40.

Gilbert, P. (1972). Iterative methods for three-dimensional reconstruction of an object from projections. *Journal of Theoretical Biology*, **36**, 105.

Gordon, R., Bender, R., and Herman, G.T. (1970). Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and X-ray photography. *Journal of Theoretical Biology*, **29**, 471.

Herman, G.T. (1979). Correction for beam hardening in computed tomography. *Physics in Medicine and Biology*, **24**, 81.

Herman, G.T. (1980). *Image reconstruction from projections. The fundamentals of computerized tomography.* Academic Press, New York.

Herman, G.T. and Kuba, A. (1999). *Discrete tomography. Foundations, algorithms, and applications.* Birkhäuser, Boston.

Herman, G.T. and Kuba, A. (2006). Optimization in the medical applications of discrete tomography. *Society of Industrial and Applied Mathematics Activity Group — SIAG/Optimization Views-and-News*, **17**, 2.

Herman, G.T. and Kuba, A. (2007). *Advances in discrete tomography and its applications.* Birkhäuser, Boston.

Herman, G.T. and Meyer, L.M. (1993). Algebraic reconstruction techniques can be made computationally efficient. *IEEE Transactions on Medical Imaging*, **12**, 600.

Jennings, R.J. (1988). A method for comparing beam hardening filter materials for diagnostic radiology. *Medical Physics*, **15**, 588.

Joseph, P.M. and Spittal, A. (1978). A method for correcting bone induced artifacts in computed tomography scanners. *Medical Physics*, **2**, 100.

Kirkpatrick, S., Gelatt, C.D., and Vecchi, M.P. (1983). Optimization by simulated annealing. *Science*, **220**, 671.

Kuba, A., Ruskó, L., L.Rodek, and Kiss, Z. (2005). Preliminary studies of discrete tomography in neutron imaging. *IEEE Transactions on Nuclear Science*, **52**, 380.

Lewitt, R.M. (1990). Multidimensional digital image representation using generalized Kaiser-Bessel window functions. *Journal of the Optical Society of*

America A, **7**, 1834.

Matej, S., Herman, G.T., Narayan, T.K., Furuie, S.S, Lewitt, R.M., and Kinahan, P.E. (1994). Evaluation of task-oriented performance of several fully 3D PET reconstruction algorithms. *Physics in Medicine and Biology*, **39**, 355.

Matej, S. and Lewitt, R.M. (1996). Practical considerations for 3-D image reconstruction using spherically symmetric volume elements. *IEEE Transactions on Medical Imaging*, **15**, 68.

Mohammad-Djafari, A. and Sauer, K. (1998). Shape reconstruction in X-ray tomography from a small number of projections using deformable models. In *Maximum entropy and Bayesian methods* (ed. J. Rychert and G. Erickson), pp. 1. Kluwer Academic, Norwell.

Mueller, K. and Yagel, R. (1996). Rapid 3D cone-beam reconstruction with SART using texture mapping hardware. *IEEE Transactions on Medical Imaging*, **19**, 1227.

Natterer, F. (1986). *The mathematics of computerized tomography.* Teubner, Stuttgart.

Radermacher, M. (1988). Three-dimensional reconstruction of single particles from random and nonrandom tilt series. *Journal of Electron Microscopy Technique*, **9**, 359.

Radon, J. (1917). Über die Bestimmung von Funktionen durch ihre Integralwerte längs gewisser Mannigfaltigkeiten. *Berichte über die Verhandlungen der Königlichen Sächsischen Akademie der Wissenschaften*, **69**, 262.

Ramachandran, G.N. and Lakshminarayanan, A.V. (1971). Three-dimensional reconstruction from radiographs and electron micrographs: Application of convolutions instead of Fourier transforms. *Proceedings of the National Academy of Sciences of the USA*, **68**, 2236.

Shepp, L.A. and Logan, B.F. (1974). The Fourier reconstruction of a head section. *IEEE Transactions on Nuclear Science*, **21**, 21.

Shepp, L.A. and Vardi, Y. (1982). Maximum likelihood reconstruction in emission tomography. *IEEE Transactions on Medical Imaging*, **1**, 113.

Sorzano, C.O.S., Marabini, R., Boisset, N., Rietzel, E., Schröder, R., Herman, G.T., and Carazo, J.M. (2001). The effect of overabundant projection directions on 3D reconstruction algorithms. *Journal of Structural Biology*, **133**, 108.

Vardi, E., Herman, G.T., and Kong, T.Y. (2001). Speeding up stochastic reconstructions of binary images from limited projection directions. *Linear Algebra and its Applications*, **339**, 75.

3

VISUALIZATION, PROCESSING AND ANALYSIS OF TOMOGRAPHIC DATA

Joachim Ohser and Katja Schladitz

We give an overview of methods providing a geometric characterization of structures based on tomographic data. A quantitative analysis usually requires some processing of the image data and – except for spectral analysis – *segmentation*. That is, an essential pre-requisite of the analysis is to find the image segment or objects of interest in the image.

Methods for processing and analysis of 3D image data are not just straightforward generalizations of classical 2D methods. The much larger amount of data requires more efficient algorithms and rules out many that are still acceptable in 2D. The much higher complexity, e.g. due to the exponential increase in directions given by pairs of vertices of the lattice unit cell, causes various ambiguities and problems. For example, there is still no complete description of all discrete neighbourhoods or connectivities available that can be used consistently for the description of both the foreground and the background of a 3D image. Furthermore, visualization of 3D data is a research field on its own that is essential since visual inspection or evaluation of image-processing or image-segmentation results is hard or even impossible.

This chapter summarizes methods for image segmentation and image processing particularly useful for 3D images of materials and structures. The second focus of this chapter is on methods of analysis developed for extracting the spatial structural information from tomographic data. The intrinsic volumes and their densities are introduced as basic and versatile geometric characteristics of image objects and image components or segments, respectively. An efficient algorithm for their measurement is described. As a by-product, this algorithm yields orientation information, too. Covariances and correlations are investigated using spectral methods.

The chapter is completed by a survey of visualization techniques as well as an introduction to the simulation of materials behaviour based on CT data and computing of macroscopic properties.

3.1 Lattices, adjacency of lattice points, and images

Image data are usually given on homogeneous point lattices, e.g. the *cubic primitive* lattice $\mathbb{L}^3 = a\mathbb{Z}^3$, $a > 0$, where $\mathbb{Z}$ denotes the set of integers and a is the lattice distance.

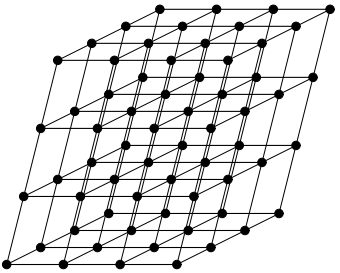


FIG. 3.1. A non-orthogonal homogeneous lattice $\mathbb{L}^3$. There is no orthogonal lattice basis.

Attributes of images are the underlying point lattice defining the pixel distances, the *window* through which the data are observed (given by the image size and offset, in image processing: the region of interest) and the type of pixel values. Furthermore, the connectivity of pixels is the input of many image-processing and analysis algorithms. Hence, connectivity is usually considered a further image attribute. For images of dimensions greater than two, the connectivity has particular importance. Therefore, as the theoretical foundation of the present chapter, the concepts 'connectivity' and 'connected-components' are introduced in detail.

3.1.1 *Homogeneous lattices*

In the following, we also consider lower-dimensional section lattices that are not necessarily orthogonal. See Fig. 3.1 for a non-orthogonal lattice. A n-dimensional *homogeneous lattice* $\mathbb{L}^n$ is a subset of the n-dimensional Euclidean space $\mathbb{R}^n$, $n = 1, 2, 3$, with

$$\mathbb{L}^n = \{x \in \mathbb{R}^n : x = \sum_{i=1}^{n} \lambda_i u_i,\ \lambda_i \in \mathbb{Z}\} = U\mathbb{Z}^n, \tag{3.1}$$

where $u_1, \ldots, u_n \in \mathbb{R}^n$ form a basis of $\mathbb{R}^n$ and $U = (u_1, \ldots, u_n)$ is the matrix of the u_i as column vectors. See Fig. 3.2 for an example. Given the basis $u_1, \ldots, u_n$ the (topologically closed) *unit cell* of $\mathbb{L}^n$ is

$$C = \{x \in \mathbb{R}^n : x = \sum_{i=1}^{n} \lambda_i u_i,\ \lambda_i \in [0,1]\} = U \cdot [0,1]^n,$$

where $[0,1]^n$ is the unit cube in $\mathbb{R}^n$. The volume of C is $\operatorname{vol} C = |\det U| > 0$ independent of the choice of the basis. The set of translates of the lattice cell $\{C + x : x \in \mathbb{L}^n\}$ covers $\mathbb{R}^n$, i.e. $\mathbb{R}^n = \bigcup_{x \in \mathbb{L}^n} (C + x)$. For a cubic lattice all edges of the unit cell have the same length, the lattice spacing, called *pixel size* in the following. Note that most lattices known from crystallography are homogeneous in the above sense.

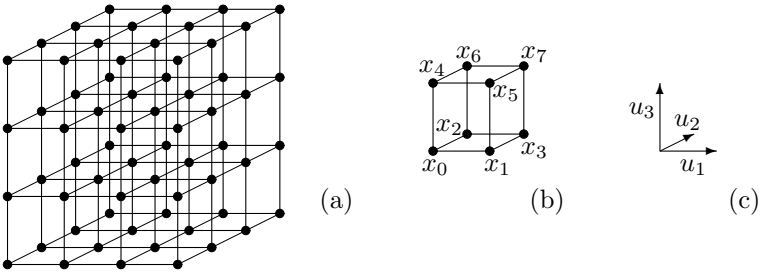


FIG. 3.2. (a) A part of a homogeneous lattice $\mathbb{L}^3$, (b) a unit cell with the vertices $x_0, \ldots, x_7$, (c) the corresponding basis u_1, u_2, u_3.

In this chapter, we consider homogeneous lattices, only. Thus, we call them just 'lattices' for short in the following.

3.1.2 *Image data*

A three-dimensional *image* is a mapping from a part of the lattice points to a domain of pixel values V. Let W be a cuboidal window with edges parallel to the co-ordinate axes. The mapping

$$\mathbb{L}^3 \cap W \mapsto V,$$

with $V = \{0, 1\}$, is a three-dimensional *binary image*, also said to be a *Boolean image*. The pixels of value 1 belong to the *foreground* of the image and those with the value 0 form the background. Analogously, for $V = \{0, \ldots, 255\}$ the image is an 8-bit *grey-value image* (an image of pixel type `unsigned char`).

In principle, the pixel values can be arbitrary. Important in practice are the cases where V are the sets of 16-bit signed or unsigned and 32-bit unsigned integers. *Label images* are grey-value images where the pixel value is the number of the object or region this pixel belongs to. Label images result from segmentation algorithms such as labelling or the watershed transform, see Section 3.4.2 and 3.4.3.

Further types of pixel values are `float`, `double`, `long double`. The Fourier transform of a grey-value image has complex-valued pixels and colour images are multiple-channel images with vector-valued pixels usually with $V = \{0, \ldots, 255\}^3$. If the image data are given on a lattice $\mathbb{L}^3$ with the matrix U of the basis vectors, the basis vectors of the Fourier-transformed image are the column vectors of $(U')^{-1}$ (the inverse of the transpose of U).

In the above setting, a *pixel*(picture element) is a pair (x, p) with $x \in \mathbb{L}^n \cap W$ and $p \in V$. Often, a pixel of a three-dimensional image is also called a *voxel* (volume element).

3.1.3 *Adjacency and Euler number*

The connectivity or adjacency of the pixels is needed e.g. for the meshing of surfaces, for labelling, watershed transform, skeletonization, and measurement of features. In the literature, the connectivity is usually characterized by a neighbourhood graph of the pixels. Here, we use the alternative concept of an adjacency system *adjacency system* motivated by the computation of the Euler number of a discretization and thoroughly introduced by Nagel *et al.* (2000), Ohser *et al.* (2002, 2003) and Schladitz *et al.* (2006*a*).

3.1.3.1 *Discretization with respect to an adjacency system* The data $X \cap \mathbb{L}^n$ can be interpreted as the foreground of a binary image. Many of the processing and analysis methods described in the following need the adjacency of the lattice points. Therefore, we introduce the concept *discretization* of the set X with respect to a given adjacency system $\mathbb{F}$.

In order to provide a clear definition of adjacency, the vertices $x_j = \sum_{i=1}^n \lambda_i u_i$ of the unit cell C are indexed with $j = \sum_{i=1}^n 2^{i-1}\lambda_i$, $\lambda_i \in \{0,1\}$. Clearly, the unit cell C has 2^n vertices, $x_i \in \mathcal{F}^0(C)$, $i = 0, \ldots, 2^n - 1$, where $\mathcal{F}^0(C)$ denotes the set of vertices of C. In a similar way we introduce the index of a subset $\xi \subseteq \mathcal{F}^0(C)$. Let $\mathbb{I}$ denote the indicator function of a set, i.e. $\mathbb{I}(x \in \xi) = 1$ if $x \in \xi$ and $\mathbb{I}(x \in \xi) = 0$ otherwise. We assign the index ℓ to ξ and write ξ_ℓ if

$$\ell = \sum_{j=0}^{2^n-1} 2^j \mathbb{I}(x_j \in \xi), \tag{3.2}$$

i.e. $\ell \in \{0, \ldots, \nu\}$ with $\nu = 2^{2^n} - 1$. Note that $\xi_0 = \emptyset$, $\xi_\nu = \mathcal{F}^0(C)$, and $\xi_{\nu-\ell} = \xi_\nu \setminus \xi_\ell$. The set ξ_ℓ can be seen as a local pixel configuration of the foreground of a binary image. In the following, we use pictograms to illustrate configurations. For example, in the 2D case, the configuration $\xi_{11} = \{x_0, x_1, x_3\}$ is represented by  where the full discs mark the foreground pixels and the empty discs denote background pixels. In the 3D case, the configuration ξ_{113} is given by  . Finally, we introduce the convex hulls $F_\ell = \operatorname{conv} \xi_\ell$ forming convex polytopes with $F_\ell \subseteq C$ and $\mathcal{F}^0(F_\ell) \subseteq \mathcal{F}^0(C)$, $\ell = 1, \ldots, \nu$. Let $\mathcal{F}^j(F)$ denote the set of all j-dimensional faces of a convex polytope F: $\mathcal{F}^0(F)$ is the set of vertices of F, $\mathcal{F}^1(F)$ is the set of edges, $\mathcal{F}^2(F)$ is the set of faces, and $\mathcal{F}^3(F)$ is the set of polytopes. For a set $\mathbb{F}$ of convex polytopes write $\mathcal{F}^j(\mathbb{F}) = \bigcup\{\mathcal{F}^j(F) : F \in \mathbb{F}\}$.

Now we are able to equip the lattice $\mathbb{L}^n$ with a (homogeneous) adjacency system defining the neighbourhood of lattice points.

Definition Let $\mathbb{F}_0 \subseteq \{F_0, \ldots, F_\nu\}$ be a set of convex polytopes $F_\ell = \operatorname{conv} \xi_\ell$, and $\mathbb{F}$ the union over all lattice translations of $\mathbb{F}$, that is $\mathbb{F} = \bigcup_{x \in \mathbb{L}^n} \mathbb{F}_0 + x$. If

(i) $\emptyset \in \mathbb{F}_0$, $C \in \mathbb{F}_0$,

(ii) if $F \in \mathbb{F}_0$ then $\mathcal{F}^i(F) \subset \mathbb{F}_0$ for $i = 0, \ldots, \dim F$,

(iii) if $F_i, F_j \in \mathbb{F}$ and $\operatorname{conv}(F_i \cup F_j) \notin \mathbb{F}$ then $F_i \cap F_j$, $\overline{F_i \setminus F_j}$, $\overline{F_j \setminus F_i} \in \mathbb{F}$.

(iv) if $F_{i_1}, \ldots, F_{i_m} \in \mathbb{F}_0$ and $F = \bigcup_{j=1}^{m} F_{i_j}$ is convex then $F \in \mathbb{F}_0$, $m = 2, \ldots, \nu$,

then the system $\mathbb{F}_0$ is called a *local adjacency system* and $\mathbb{F}$ is said to be an *adjacency system* of the lattice $\mathbb{L}^n$.

From condition (i) in this definition it follows immediately that $\mathcal{F}^0(\mathbb{F}) = \mathbb{L}^n$. The pair $\Gamma = (\mathcal{F}^0(\mathbb{F}), \mathcal{F}^1(\mathbb{F}))$ is the *neighbourhood graph* of $\mathbb{F}$; it consists of the set $\mathcal{F}^0(\mathbb{F})$ of nodes and the set $\mathcal{F}^1(\mathbb{F})$ of edges. All nodes are of the same order since Γ is homogeneous, $\Gamma + x = \Gamma$, $x \in \mathbb{L}^n$. The order of the nodes is called the *connectivity* of $\mathbb{L}^n$.

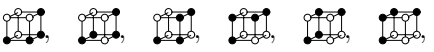
In the simplest case where the adjacency system is generated by the unit cell C, the order of the nodes is $2n$, and $\mathbb{F}_{2n} = \bigcup_{x\in\mathbb{L}^n} \bigcup_{j=0}^{n} \mathcal{F}^j(C+x)$. The maximal adjacency system consisting of the convex hulls of all point configurations is a (3^n-1)-adjacency, $\mathbb{F}_{3^n-1} = \bigcup_{x\in\mathbb{L}^n}\{F_0 + x, \ldots, F_\nu + x\}$. Notice that for all adjacency systems $\mathbb{F}$ of $\mathbb{L}^n$ it follows that

$$\mathbb{F}_{2n} \subseteq \mathbb{F} \subseteq \mathbb{F}_{3^n-1}.$$

Now we recall the adjacency systems for $\mathbb{L}^3$ considered in detail by Ohser *et al.* (2002, 2003):

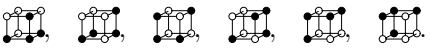
6-adjacency. The 6-adjacency is used as a standard in image processing, $\mathbb{F}_6 = \mathbb{F}_{2\cdot 3} = \bigcup_{x\in\mathbb{L}^3} \bigcup_{j=0}^{3} \mathcal{F}^j(C+x)$.

14.1-adjacency. This adjacency system is generated from the tessellation of C into the 6 tetrahedra F_{139}, F_{141}, F_{163}, F_{177}, F_{197}, F_{209}, which are the convex hulls of the configurations



i.e. $\mathbb{F}_0$ consists of all j-faces of the tetrahedra, $j = 0, \ldots, 3$, and their convex unions. The edges of the corresponding neighbourhood graph Γ are the edges of C, the face diagonals of C containing the origin 0, the space diagonal of C containing 0, and all their lattice translations. The order of the nodes of Γ is 14.

14.2-adjacency. The 14.2-adjacency system is generated from the tetrahedra F_{43}, F_{141}, F_{147}, F_{169}, F_{177}, and F_{212}, which are the convex hulls of



The corresponding neighbourhood graph Γ differs from the one for 14.1 in containing a face diagonal of C that does not include 0.

26-adjacency. This is the maximal adjacency for $n = 3$, $\mathbb{F}_{26} = \bigcup_{x\in\mathbb{L}^3}\{F_0 + x, \ldots, F_{255} + x\}$.

It should be noted that for $n > 2$ there can be two or more adjacency systems having the same neighbourhood graph. In other words, an adjacency system $\mathbb{F}$ is not uniquely determined by Γ, see the example on the following page.

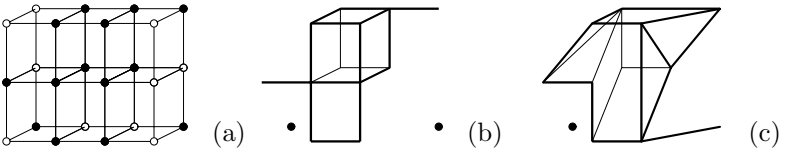


FIG. 3.3. Two discretizations of image data $X \cap \mathbb{L}^3$: (a) the foreground pixels $X \cap \mathbb{L}^3$, (b) the discretization $X \sqcap \mathbb{F}_6$, (c) the discretization $X \sqcap \mathbb{F}_{14.1}$.

The set $X \subset \mathbb{R}^3$ (e.g. a constituent of a material structure or a particle) is observed[2] on a lattice $\mathbb{L}^3$ or a section lattice $\mathbb{L}^n$, $n = 1, 2$. Here, we consider only the simplest case where the observation is modelled as the intersection $X \cap \mathbb{L}^n$ of the set X with the lattice $\mathbb{L}^n$ (Gauss digitization).

Definition The *discretization* $X \sqcap \mathbb{F}$ of a compact subset $X \subset \mathbb{R}^3$ *with respect to a given adjacency system* $\mathbb{F}$ is defined as the union of all j-faces of the elements of $\mathbb{F}$ for which all the vertices hit X, i.e.

$$X \sqcap \mathbb{F} = \bigcup \{F \in \mathbb{F} \,:\, \mathcal{F}^0(F) \subseteq X\}. \tag{3.3}$$

This means that a 'brick' $F \in \mathbb{F}$ is a subset of the discretization of X if and only if all vertices of F belong to X, see Fig. 3.3. The set $X \sqcap \mathbb{F}$ is an approximation of X by a polyhedron obtained from discrete data $X \cap \mathbb{L}^n$. In particular, we have $X \sqcap \mathbb{F} = (X \cap \mathbb{L}^n) \sqcap \mathbb{F}$, i.e. the discretization eqn 3.3.

Other digitization modes might be more realistic, see e.g. Klette and Rosenfeld (2004), but our discretization model is mathematically easily tractable. Nevertheless, most considerations made in the following are in principle independent of the chosen digitization mode.

3.1.3.2 *Euler number* By definition, the *Euler number* $\chi(X)$ of a (compact) convex set X is 1, and $\chi(\emptyset) = 0$ for the empty set. For the union of two convex sets X_1 and X_2 one obtains $\chi(X_1 \cup X_2) = \chi(X_1) + \chi(X_2) - \chi(X_1 \cap X_2)$. This means that the Euler number of the union $X_1 \cup X_2$ is 2 if the intersection $X_1 \cap X_2$ is empty, and $\chi(X_1 \cup X_2) = 1$ otherwise. More generally, let X be the union of m convex sets $X_1, \ldots, X_m$. Then the Euler number of X can be computed via the so-called *inclusion-exclusion principle*,

$$\chi(X) = \sum_{i=1}^{m} \chi(X_i) - \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} \chi(X_i \cap X_j) + \ldots - (-1)^{m+1} \chi\Big(\bigcap_{i=1}^{m} X_i\Big).$$

Furthermore, if P is a polyhedron, the Euler number of P can be computed via the *Euler–Poincaré formula* as an alternating sum of the numbers of edges, vertices, faces, and cells,

[2]The term 'observed' is used here to express that just a discrete subset of a continuous set is considered.

$$\chi(P) = \#\,\text{vertices} - \#\,\text{edges} + \#\,\text{faces} - \#\,\text{cells}.$$

Now, since the set $X \sqcap \mathbb{F}$ forms a (not necessarily convex) polyhedron, the number of elements of $\mathcal{F}^j(X \sqcap \mathbb{F})$ is finite and therefore the Euler number $\chi(X \sqcap \mathbb{F})$ can be computed via the *Euler–Poincaré formula*,

$$\chi(X \sqcap \mathbb{F}) = \sum_{j=0}^{3} (-1)^j \, \#\mathcal{F}^j(X \sqcap \mathbb{F}). \tag{3.4}$$

For the simple example in Fig. 3.3, we obtain

$$\begin{aligned} \chi(X \sqcap \mathbb{F}_6) &= 14 - 17 + 7 - 1 = 3 \qquad \text{for (b)}, \\ \chi(X \sqcap \mathbb{F}_{14.1}) &= 12 - 18 + 9 - 1 = 2 \qquad \text{for (c)}. \end{aligned}$$

In order to apply a 'local method' for measuring the Euler number we derive now a local version of the Euler–Poincaré formula eqn 3.4. For the local configuration $\xi_\ell = X \cap C \cap \mathbb{L}^n$ we have $\xi_\ell \sqcap \mathbb{F} = (X \sqcap \mathbb{F}) \cap C = X \sqcap \mathbb{F}_0$. We define

$$\mu F_\ell = \min\{j : \text{there is an } F \in \mathcal{F}^j(C) \text{ with } F_\ell \subseteq F\}.$$

Then

$$2^{n-\mu F_\ell} = \sum_{x \in \mathbb{L}^n} \mathbb{I}(F_\ell \subseteq C + x)$$

is the number of lattice cells covering F_ℓ, and its reciprocal is the weight for the edge correction. Now, the edge-corrected *local Euler number* χ_0 can be defined as

$$\chi_0(\xi_\ell \sqcap \mathbb{F}) = \sum_{j=0}^{n} (-1)^j \sum_{F \in \mathcal{F}^j(\xi_\ell \sqcap \mathbb{F})} 2^{\mu F - n}, \qquad \ell = 0, \ldots, \nu \tag{3.5}$$

$$\text{and} \quad \chi_0((\xi_\ell + x) \sqcap \mathbb{F}) = \chi_0(\xi_\ell \sqcap \mathbb{F}) \quad \text{for all} \quad x \in \mathbb{L}^n.$$

3.1.3.3 *Two adjacency systems with the same neighbourhood graph* Consider for $n = 3$ the 12.1-adjacency generated by the convex polytopes

$$F_{113} = \text{conv}\,\square, \quad F_{142} = \text{conv}\,\square, \quad F_{231} = \text{conv}\,\square$$

and the 12.2-adjacency generated by

$$F_{113} = \text{conv}\,\square, \quad F_{142} = \text{conv}\,\square, \quad F_{103} = \text{conv}\,\square, \quad F_{230} = \text{conv}\,\square.$$

Both these adjacencies have the same neighbourhood graph Γ with order of the nodes 12. However, the discretization of $F_{102} = \text{conv}\,\square$ in reference to 12.1 consists of 4 edges and 4 vertices only, while the quadrangle is completely contained in the discretization in reference to 12.2. Thus, for dimensions larger than two,

an adjacency system $\mathbb{F}$ is not uniquely determined by the neighbourhood graph Γ. In fact, the different discretizations yield different Euler numbers, too. For the Euler numbers as defined in eqn 3.4 and the local Euler numbers as defined in eqn 3.5 we obtain

$$\begin{aligned}
\chi_0(F_{102} \sqcap \mathbb{F}_{12.1}) &= 4\frac{1}{8} - (2\frac{1}{4} + 2\frac{1}{2}) = -1 \\
\chi_0(F_{102} \sqcap \mathbb{F}_{12.2}) &= 4\frac{1}{8} - (2\frac{1}{4} + 2\frac{1}{2}) + 1 = 0 \\
\chi(F_{102} \sqcap \mathbb{F}_{12.1}) &= 4 - 4 = 0 \\
\chi(F_{102} \sqcap \mathbb{F}_{12.2}) &= 4 - 4 + 1 = 1.
\end{aligned}$$

Additivity and translation invariance of the Euler number and the fact that $X \sqcap \mathbb{F} = (X \cap \mathbb{L}^n) \sqcap \mathbb{F}$ yield

$$\begin{aligned}
\chi(X \sqcap \mathbb{F}) &= \sum_{x \in \mathbb{L}^n} \chi_0(C \cap ((X \sqcap \mathbb{F}) - x)) \\
&= \sum_{x \in \mathbb{L}^n} \sum_{\ell=0}^{\nu} \chi_0(\xi_\ell \sqcap \mathbb{F}) \mathbb{I}(\xi_\ell + x \subseteq X) \mathbb{I}(\xi_{\nu-\ell} + x \subseteq X^c) \\
&= \sum_{\ell=0}^{\nu} \chi_0(\xi_\ell \sqcap \mathbb{F}) \underbrace{\sum_{x \in \mathbb{L}^n} \mathbb{I}(\xi_\ell + x \subseteq X) \mathbb{I}(\xi_{\nu-\ell} + x \subseteq X^c)}_{h_\ell}. \qquad (3.6)
\end{aligned}$$

Thus, the Euler number can be written as a scalar product, $\chi(X \sqcap \mathbb{F}) = wh$, where the components $w_\ell = \chi_0(\xi_\ell \sqcap \mathbb{F})$ of the vector $w = (w_\ell)$ depend on the adjacency system $\mathbb{F}$, but not on the structure X. On the other hand, the vector $h = (h_\ell)$ depends on X but not on $\mathbb{F}$; its components h_ℓ can be computed very efficiently from 'local information' about $X \cap \mathbb{L}^n$.

3.1.3.4 *Complementarity* It is well known from image processing that if one chooses an adjacency system $\mathbb{F}$ for the discretization of X then an $\mathbb{F}_c$ for the discretization of the complementary set X^c is implicitly chosen, too. In other words, if the 'foreground' $X \cap \mathbb{L}^n$ is $\mathbb{F}$-connected then the 'background' $\overline{X^c} \cap \mathbb{L}^n$ must be $\mathbb{F}_c$-connected. For $n > 2$ it is not sufficient to consider connectivity, thus further criteria have to be regarded. Usually, the adjacency of lattice points (or pixels) is characterized by a neighbourhood graph Γ, and the complementarity of adjacencies is defined via the Jordan surface theorem (Jordan–Brouwer theorem), which says that for every $X \subset \mathbb{R}^3$ there exists a closed surface separating $X \sqcap \mathbb{F}$ from $X^c \sqcap \mathbb{F}_c$ 'in a unique way'. Here, we present an alternative approach: We use an adjacency system $\mathbb{F}$ and introduce 'complementarity' by means of the Euler number of the discretization $X \sqcap \mathbb{F}$.

The pair $(\mathbb{F}, \mathbb{F}_c)$ is called a *pair of complementary adjacency systems* if $(X \sqcap \mathbb{F}) \cap (X^c \sqcap \mathbb{F}_c) = \emptyset$ and

$$\chi(X \sqcap \mathbb{F}) = (-1)^{n+1} \chi(X^c \sqcap \mathbb{F}_c) \tag{3.7}$$

for all compact $X \subset \mathbb{R}^n$. An adjacency system $\mathbb{F}$ is called *self-complementary* if $\chi(X \sqcap \mathbb{F}) = (-1)^{n+1} \chi(X^c \sqcap \mathbb{F})$ for all compact X.

So far, there is no method known for constructing the complementary adjacency system $\mathbb{F}_c$ for a given adjacency system $\mathbb{F}$. Even worse, for a given adjacency system $\mathbb{F}$ there is not necessarily an adjacency system $\mathbb{F}_c$ such that eqn 3.7 holds. Nevertheless, four pairs of complementary adjacency systems are known:

$(\mathbb{F}_6, \mathbb{F}_{26})$, $(\mathbb{F}_{14.1}, \mathbb{F}_{14.1})$ *and* $(\mathbb{F}_{14.2}, \mathbb{F}_{14.2})$. The 6-adjacency is complementary to the 26-adjacency and vice versa. There are two self-complementary adjacency systems, the 14.1-adjacency and the 14.2-adjacency, see Ohser *et al.* (2002, 2003).

18-adjacency. A remarkable observation in this context is that the 18-neighbourhood cannot be used for a consistent estimation of the Euler number. This counteracts the fact that (18,6) is considered to be a 'good pair' in digital topology, see e.g. Klette and Rosenfeld (2004, Chap. 7).

The Euler number of the discretization $X \sqcap \mathbb{F}$ cannot be expected to be equal to the Euler number of X. However, if the surface of X is sufficiently smooth, then $\chi(X \sqcap \mathbb{F})$ converges to $\chi(X)$ for lattice distances going to zero. Thus $\chi(X \sqcap \mathbb{F})$ is multigrid convergent in this special case. See Ohser and Nagel (1996) for details.

3.1.4 *Connected components*

In order to describe image-processing techniques such as skeletonization or labelling, it is necessary to introduce the topological concepts 'connectedness' and 'connected component'. A two-dimensional digital topology was introduced by Rosenfeld (1970) who defined connectedness on lattices and stated a discrete Jordan–Veblen curve theorem. Rosenfeld's definition of connectedness can simply be extended to lattices of arbitrary dimensionality, see e.g. Lachaud and Montanvert (2000). Obviously, connectedness of components is closely related to the connectivity of lattice points (or pixels) and, hence, we introduce connectedness with respect to a given adjacency system.

First, we consider the continuous case and introduce path-connectedness for the Euclidean space $\mathbb{R}^3$. The connected components of a bounded set $X \subset \mathbb{R}^3$ can be considered as the equivalence classes of $X \subseteq \mathbb{R}^3$ concerning an appropriately chosen equivalence relation $\sim$ defined for point pairs in $\mathbb{R}^3$. Here, a path in $\mathbb{R}^3$ from the point x to the point y is a continuous mapping $f : [0,1] \mapsto \mathbb{R}^3$ with $f(0) = x$ and $f(1) = y$. Now, a non-empty set X is called *path-connected* if for every $x, y \in X$ there exists a path f from x to y such that f belongs completely to X, $f(\cdot) \subseteq X$.

We write $x \sim y$ for path-connected points $x, y \in \mathbb{R}^3$. It can be shown that the binary relation $\sim$ is an equivalence relation, i.e. $\sim$ is reflexive, symmetric,

and transitive. The equivalence classes $X_1, \ldots, X_m$ of X under $\sim$ are called the *path-connected components* of X. For more details see e.g. Rotman (1993). The equivalence classes form the 'objects' of the foreground, e.g. the particles of a constituent. A path-connected component without holes or tunnels is called a *simply connected component*.

Contrary to an image object, an *image segment* is not necessarily a path-connected component. The concept 'image segment' is usually used for the result of a binarization, see Section 3.4.1, that is, a whole constituent of the material considered. An image segment can consist of a set of isolated objects such as pores in a closed foam or be multiply connected like the strut system of an open foam.

Now we consider a discrete set Y on a homogeneous lattice $\mathbb{L}^3$ equipped with a pair of complementary adjacency systems $(\mathbb{F}, \mathbb{F}_c)$. Let $x, y \in \mathbb{L}^3$ be lattice points. A discrete path from x to y with respect to the adjacency system $\mathbb{F}$ is a sequence of lattice points $(x_i)_{i=0}^m \subset \mathbb{L}^3$, $m \in \mathbb{N}$, with $x_0 = x$, $x_m = y$, and $[x_{i-1}, x_i] \in \mathbb{F}$, $i = 1, \ldots, m$. Here, $\mathbb{N}$ is the set of natural numbers.

A non-empty discrete set $Y \subseteq \mathbb{L}^3$ is called $\mathbb{F}$-path-connected if Y consists of just one equivalence class concerning path-connectedness or if for all pairs $(x, y) \in Y^2$ with $x \neq y$ there exists a discrete $\mathbb{F}$-path from x to y. $\mathbb{F}$-*connectedness* in Y is an equivalence relation defining the *connected components* of Y.

3.2 Visualization

Visualization of volume image data is an important tool for exploratory analysis. That is, visualization is indispensable to provide first ideas about appropriate further processing and analysis steps. Moreover, visualization helps to verify and interpret results. Finally, high-quality 3D visualizations are still 'eye catchers' for presentations. The user should, however, always be aware that any visualization of 3D image data is based on some reduction of the spatial information. The human eye is easily misled. For instance, it is nearly impossible to judge connectivity in 3D just visually.

The easiest way to visualize volume images is to cut slices parallel to the coordinate directions. This is usually the default view mode. Spatial impressions are obtained by volume-rendering techniques.

3.2.1 *Volume rendering*

Direct volume-rendering methods yield a visualization of the 3D data without previously determining foreground and background or identifying objects. This allows exploratory visual inspection without imposing additional knowledge.

3.2.1.1 *Physical background* Direct volume rendering is based on solving the equation for radiative transfer (Hege *et al.*, 1993) for the entire 3D data set: Denote by $I(x, \theta, \nu)$ the intensity radiated from point x in direction θ with frequency ν. Absorption consists of true absorption κ, transforming radiant energy to thermal energy, and scattering σ, causing a change in direction θ of radiation. Emission is described by a source term q and the scattering part j.

The equation of radiative transfer captures the fact that the difference between the energy at two points on a line going through a lattice cell must be equal to the difference between the energy emitted and the energy absorbed through that cell. Consider a ray of light travelling along a straight line with direction θ, parameterized by s, entering the volume at position s_0 and leaving at position s_m. For a point on this line we then have

$$\frac{\partial}{\partial s} I = -(\kappa + \sigma) I + q + j.$$

In order to make this equation solvable, the volume is imagined to be filled with small light-emitting particles leading to the *emission-absorption* (Hege *et al.*, 1993) or *density-emitter* model (Sabella, 1988), which allows us to ignore the contributions of scattering and frequency dependence. The equation for radiative transfer thus simplifies to

$$\frac{\partial}{\partial s} I = -\kappa I + q. \tag{3.8}$$

This equation can be solved analytically by

$$I(s) = I(s_0) \mathrm{e}^{-\tau(s_0, s)} + \int_{s_0}^{s} q(\hat{s}) \mathrm{e}^{-\tau(\hat{s}, s)} \, d\hat{s}, \tag{3.9}$$

where $\tau(s_1, s_2) = \int_{s_1}^{s_2} \kappa(s) \, ds$ denotes optical depth. Discretization yields an iterative numeric solution with updating equation

$$I(s_k) = I(s_{k-1}) v_k + b_k, \;\; k = 1, \ldots, m, \tag{3.10}$$

where $v_k = \mathrm{e}^{-\tau(s_{k-1}, s_k)}$ denotes the *transparency* of the material and $b_k = \int_{s_{k-1}}^{s_k} q(s) \mathrm{e}^{-\tau(s, s_k)} \, ds$ is the *emission* term. This is the fundamental equation for almost all methods of direct volume rendering. Often, *opacity* $1 - v_k$ is used instead of transparency.

The sampling points are equidistant on the ray and hence usually located between pixels, as in general the volume is not aligned with the ray of sight. Therefore, it is necessary to interpolate the values of the sampling points from the surrounding pixels.

3.2.1.2 *Transfer function* Transfer functions map the scalar (grey-)values given by a sample of the raw data to colour (emission) and opacity (absorption) of light at this point. The transfer function is usually defined and modified interactively by the user. Often, this is a tedious and time-consuming procedure. Automatic and semi-automatic data-driven methods are still topics of active research.

In practice, the transfer function is realized as a lookup table of fixed size. The emitted radiance is usually represented as an RGB value (red, green, blue). The absorption coefficient α is a scalar value between 0 and 1, where a pixel with α-value 0 is fully transparent (invisible), whereas a pixel with α-value 1 is fully opaque.

The transfer function can be applied before or after interpolation. In *pre-classification*, the transfer function is applied to the discrete data. That is, the emission and absorption coefficients for a sample point that does not lie on the lattice are calculated by interpolating between the coefficients of the local neighbours. In *post-classification*, however, the value of the transfer function at the interpolated grey-value of the sample point is used. Clearly, the image quality created with post-classification methods is highly superior to the quality of pre-classification images.

Complex transfer functions incorporate illumination and shadows. Illumination and shadows not only highly influence visualization quality but also increase insight into the visualized structure. However, solving the equation of radiative transfer in its full generality, incorporating illumination and shadowing of objects by other objects and the specific reflection and refraction properties of the material is computationally infeasible. Moreover, in CT data, the necessary information about the material is not available. Alternatively, local illumination models can be used. Ambient and diffuse reflection as well as specular highlights are computed for each pixel. Shadows are generated separately. See Akenine-Moeller and Haines (2002) and Rezk-Salama (2001) for further reference.

In the following, two popular direct volume-rendering techniques are described briefly – ray casting and 3D texture mapping. Both are image-order techniques considering each pixel of the resulting 2D image separately and computing the contribution of the entire volume to this pixel's final colour.

3.2.1.3 *Ray casting* Rays of sight are cast from the viewer into the volume – one ray for each pixel in the resulting 2D image. Ray casting is a special case of ray tracing, ignoring reflection and refraction of the ray. For each ray, the pixel intensity values, as given by eqn 3.10, are calculated and summed up (Akenine-Moeller and Haines, 2002). This can be done either front-to-back or back-to-front, yielding the same results. However, front-to-back allows early ray termination: As soon as transparency reaches zero, computation for this ray can be stopped as none of the following sampling points contributes to the final pixel value.

A computationally very simple variation is *maximum intensity projection*, just keeping the maximal intensity value along the ray and thus emphasizing bright structures but occluding depth information.

Ray casting is capable of visualizing isosurfaces by testing every ray for intersection with the surface defined by the corresponding grey-value interval. Scattering can be included, too. Levoy (1990) combines a local illumination model with Phong shading to a computationally feasible algorithm yielding a satisfactory illusion of smooth surfaces.

Images generated by ray casting represent the reference results in terms of image quality. However, ray casting is expensive in terms of memory and computation time. Nevertheless, parallelization and the use of data structures for very efficient access (so-called 'acceleration structures') bring interactive ray casting

into reach.

3.2.1.4 *3D texture mapping* Modern consumer graphics cards support 3D textures and are able to perform a *trilinear interpolation* within the volume. This allows us to render a stack of polygon slices parallel to the image plane for the current viewing direction (Rezk-Salama, 2001). This viewport-aligned stack has to be recomputed every time the viewpoint changes. Every polygon is assigned a slice of the volume as a texture. Finally, textured polygons are blended onto the image plane in back-to-front order. Recent graphics cards allow us to use short programs (shaders) for fast application of the transfer function as well as calculation of local illumination and shadows for each pixel. Isosurfaces can be visualized, too.

3.2.2 *Surface rendering*

Indirect methods extract an image segment or single objects from a binary or label image and display their surfaces. This means, contrary to direct methods, indirect methods require a segmentation step (see Section 3.4) prior to visualization. Surface rendering comprises all approaches that transform the corresponding data into surface representations. The resulting surface is finally displayed, usually supported by hardware. Thus, within the visualization process, the major and most complex part of indirect approaches is to map a set of lattice points to a (closed) surface. The surface ∂X of a set X is reconstructed from the discrete data $X \cap \mathbb{L}^3$ leading to an approximation of ∂X. The reconstructed surface depends on the chosen adjacency system $\mathbb{F}$, and is therefore denoted by $\partial(X, \mathbb{F})$ in the following.

3.2.2.1 *Properties of the reconstructed surface*

(i) *Regular and polygonal.* The reconstructed surface should be polygonal. More precisely, if $X \cap \mathbb{L}^3$ is non-empty, there exists a regular polyhedron $P \subset \mathbb{R}^3$ with $\partial(X, \mathbb{F}) = \partial P$. The faces of P are called the *meshes* of $\partial(X, \mathbb{F})$.

(ii) *Topology preserving.* The surface $\partial(X, \mathbb{F})$ should fit the topology of $X \sqcap \mathbb{F}$. More precisely, if $X \cap \mathbb{L}^3 \neq \emptyset$, there should exist a compact and topologically regular set $Y \subset \mathbb{R}^3$ with $\partial(X, \mathbb{F}) = \partial Y$ and

$$X \sqcap \mathbb{F} \subseteq Y, \qquad X^c \sqcap \mathbb{F}_c \subseteq \overline{Y^c}.$$

(iii) *Complementarity.* Meshing the point set $X \cap \mathbb{L}^n$ should lead to the same result as meshing the complement $X^c \cap \mathbb{L}^n$. It is supposed that

$$\partial(X, \mathbb{F}) = \partial(X^c, \mathbb{F}_c).$$

(iv) *Locality.* The surface $\partial(X, \mathbb{F})$ should be a collection of local contributions belonging to the lattice cells,

$$\partial(X - x, \mathbb{F}) \cap C = \partial((X - x) \cap C, \mathbb{F}) \cap C, \qquad x \in \mathbb{L}^3.$$

A surface fulfilling conditions (ii) and (iii) can be described in the following way: Let $\operatorname{dist}(x, Y)$ denote the shortest distance between a point $x \in \mathbb{R}^3$ and a set $Y \subset \mathbb{R}^3$, $\operatorname{dist}(x, Y) = \inf\{\|x - y\|_p : y \in Y\}$ with the p-norm $\|\cdot\|_p$, $p \in [1, \infty]$. Since the intersection $(X \sqcap \mathbb{F}) \cap (X^c \sqcap \mathbb{F}_c)$ is a set of measure 0, the set

$$\{x \in \mathbb{R}^3 : \operatorname{dist}(x, X \sqcap \mathbb{F}) = \operatorname{dist}(x, X^c \sqcap \mathbb{F}_c)\}$$

forms a surface $\partial(X, \mathbb{F})$ of X. For $p = 1$ this surface is additionally polygonal. However, so far there is no algorithm known that yields $\partial(X, \mathbb{F})$. Instead, algorithms for surface construction are based on tessellating the unit cell. The prototype is the marching-cube algorithm described below.

In most applications the surface $\partial(X, \mathbb{F})$ is oriented in order to distinguish between the interior and the exterior side. Let $\partial^+ X$ denote the positively oriented surface with normals equivalent to the outer surface normals of X. The negatively oriented surface of X is denoted by $\partial^- X$. Then, the complementarity condition (iii) takes the form $\partial^+(X, \mathbb{F}) = \partial^-(X^c, \mathbb{F}_c)$. The edges of the triangles and polygons belonging to a polygonal surface $\partial^+(X, \mathbb{F})$ are listed in counterclockwise order (according to a right-handed system).

3.2.2.2 *Marching-cube-type algorithms* Applying the locality, a marching-cube-type algorithm can be formulated, typically consisting of the following steps:

(i) *Pre-processing.* The local contributions $\partial(\xi_\ell, \mathbb{F}) \cap C$ are computed for the local point configurations ξ_ℓ described in Section 3.1.3.1, $\ell = 0, \ldots, 255$. Each local contribution consists of a set of triangles that are united – depending on the applied visualization technique – to a set of convex polygons (meshes), a set of triangle fans, *etc.*, in order to reduce the data and to speed up the visualization. To enable fast access, the local contributions are usually listed in a look-up table.

(ii) *Surface meshing.* The index ℓ of the local configuration $\xi_\ell = (X - x) \cap \mathcal{F}^0(C)$ is determined and the corresponding local contribution $\partial(\xi_\ell, \mathbb{F}) \cap C$ shifted by x is collected in a set (container) of polygons, where x runs through all lattice points. (In the case of a cubic lattice, the unit cell forms a cube that 'marches' through the image.)

(iii) *Post-processing.* Even if pre-processing includes some data reduction, the meshing step produces a large number of triangles or convex polygons that can lead to an overtessellation and, finally, to problems in displaying and in further processing the mesh data. This overtessellation can be reduced further by a so-called mesh simplification where coplanar or nearly coplanar polygons are joined into a simpler structure.

The most popular indirect method based on a marching-cube-type algorithm originates from Lorensen and Cline (1987), who derived local contributions empirically. Unfortunately, a meshing using these local contributions does not give a closed surface in all cases, see also Lohmann (1998, p. 184). Obviously, the most critical point of a marching-cube-type algorithm is pre-processing. Here,

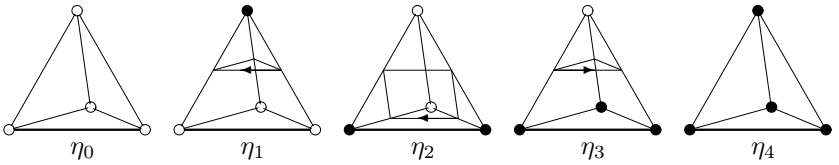


FIG. 3.4. The five configurations $\eta_0, \ldots, \eta_4$ of the vertices of a regular tetrahedron and the corresponding surface patches. The arrow marks the orientation of the surface.

we follow the approach of Gueziec and Hummel (1994) based on a tessellation of the unit cell into tetrahedra. This approach – called the *wrapper algorithm* – leads to closed polygonal surfaces.

3.2.2.3 *The wrapper algorithm* Let $\mathcal{G}_0 = \{P_1, \ldots, P_m\}$ be a tessellation of the unit cell C into m tetrahedra $P_1, \ldots, P_m$ with non-empty interior. It is assumed that the tessellation is space filling, $\cup_{i=1}^m P_i = C$, the polyhedra do not overlap, $\operatorname{int} P_i \cap \operatorname{int} P_j = \emptyset$ for $i \neq j$, and the vertices of the tetrahedra are vertices of the unit cell $\mathcal{F}^0(P_i) \subset \mathcal{F}^0(C)$, $i = 1, \ldots, m$. Furthermore, we assume that $\mathcal{G}_0$ is a face-to-face tessellation, i.e. $P_i \cap P_j$ is either empty or a k-face of P_i and P_j, $k \leq 3$.

The idea behind the wrapper algorithm is to use 'building blocks' smaller than the unit cell of the lattice and to reduce the complexity of local configurations and to simplify the construction of surface patches in this way. Consider a uniform tetrahedron P. There exist (up to rotations) only 5 configurations $\eta_0, \ldots, \eta_4 \subseteq \mathcal{F}^0(P)$ with number of edges $\#\eta_\ell = \ell$. The configurations η_0 and η_4 do not contribute to the surface, the contributions of η_1 and η_3 are triangles, and the contribution of η_2 is a square. The vertices of the triangles and the square are centres of edges of P, $\frac{1}{2}(x + y)$, where x is a vertex of P belonging to the foreground and the vertex y is a vertex belonging to the background, see Fig. 3.4.

Since each tetrahedron P_i is a linear mapping of P, the surface patches of the P_i are linear mappings of the surface patches of P corresponding to the configuration η_ℓ. Finally, the surface patches of the unit cell with respect to a configuration ξ_ℓ are formed by the contributions of the tetrahedra.

Gueziec and Hummel (1994) describe the wrapper algorithm for the tessellation of the unit cell into the five tetrahedra

$$P_1 = \operatorname{conv}\,\square, \quad P_2 = \operatorname{conv}\,\square, \quad P_3 = \operatorname{conv}\,\square,$$
$$P_3 = \operatorname{conv}\,\square, \quad P_5 = \operatorname{conv}\,\square.$$

The wrapper algorithm can be applied to the tessellations of the unit cell into 6 tetrahedra that are used to construct the two self-complementary adjacency systems $\mathbb{F}_{14.1}$ and $\mathbb{F}_{14.2}$ introduced in Section 3.1.3.1, too. This leads to a polygonal surface with the properties (i)–(iv) considered on page 49.

3.2.2.4 *Merging and simplification of surface meshes* The time necessary for a ray tracing (in particular for animation of the data) is linear in the number of meshes. Therefore, merging surface meshes is almost as important as the mesh generation itself. A merge leading to considerable data reduction is very complicated and tedious. It can be eased considerably by knowledge about neighbouring meshes where the indices of the neighbours of each mesh are temporarily included in the structure of the rendering data.

Two coplanar neighbours are merged to a more complex polygon by removing the touching edges. Then, both lists of indices of further neighbours are united. The merging is continued until the neighbour lists of all meshes are empty. A considerable reduction of the rendering data can be achieved if smoothing techniques are involved in the merging process. For instance, a plane can be fitted to an assembly of nearly coplanar meshes, see Schroeder *et al.* (1992), or some of the vertices of the polygonal surface are slightly moved, see Hoppe *et al.* (1993).

Notice that the merged meshes are not necessarily convex. Depending on the requirements of the ray tracing, complex meshes are simplified after merging e.g. by a Delaunay triangulation, (Watson, 1981). For further details on merging and simplification of meshes see e.g. Gueziec and Hummel (1994) and Peiró (1999).

3.3 Processing of image data

This section provides a wide variety of tools for image enhancement or image transformation, e.g. for denoising, emphasizing structures of interest, reducing the image information to that needed in further analysis steps. While some – for instance filters – can be seen as pure pre-processing steps, others – such as morphological or distance transforms – are methods of analysis in their own right, too.

3.3.1 *Fourier transform*

The Fourier transform is a classical tool in image processing. On the one hand, as in signal analysis and physics, the Fourier transform is used to detect periodicities. On the other hand, the Fourier transform is the basis for image enhancement by removing undesired frequencies directly in the Fourier-transformed image (low-, high-, or bandpass filtering) or by filtering using the convolution with a large filter mask (Jähne, 2005).

3.3.1.1 *Continuous Fourier transform* Denote by $\mathbb{C}$ the space of complex numbers. In the continuous case, the *Fourier transform* $\hat{f} = \mathcal{F}f$ of a measurable function $f : \mathbb{R}^3 \mapsto \mathbb{C}$ is defined by

$$\mathcal{F}f(\xi) = \frac{1}{(2\pi)^{3/2}} \int_{\mathbb{R}^3} f(x) \mathrm{e}^{-i\xi x} dx, \qquad \xi \in \mathbb{R}^3.$$

Notice that the function $\hat{f}$ is usually complex-valued, $\hat{f} : \mathbb{R}^3 \mapsto \mathbb{C}$, even in the case when f is only real-valued. Analogously, the inverse *Fourier transform* (or Fourier cotransform) $\bar{\mathcal{F}}f$ of f is defined by

$$\bar{\mathcal{F}}f(x) = \frac{1}{(2\pi)^{3/2}} \int\limits_{\mathbb{R}^3} f(\xi) \mathrm{e}^{ix\xi} d\xi, \qquad x \in \mathbb{R}^3.$$

The Fourier transform and its inverse are linear and for any integrable function f, the inversion formulae $\bar{\mathcal{F}}\mathcal{F}f = f$ and $\mathcal{F}\bar{\mathcal{F}}f = f$ are valid. Furthermore, let $\overline{f(x)}$ denote the complex conjugate of $f(x)$. Then, for two sufficiently smooth functions f and g of rapid decay the *Parseval identity*

$$\int\limits_{\mathbb{R}^3} f(x)\overline{g(x)} dx = \int\limits_{\mathbb{R}^3} \hat{f}(\xi)\overline{\hat{g}(\xi)} d\xi \tag{3.11}$$

and the *Plancherel identity*

$$\int\limits_{\mathbb{R}^3} |f(x)|^2 \, dx = \int\limits_{\mathbb{R}^3} \left|\hat{f}(\xi)\right|^2 d\xi$$

hold. It should be noted that in the literature different definitions of the Fourier transform and inverse are used. Sometimes, the factor $\frac{1}{(2\pi)^{3/2}}$ is omitted for the Fourier transform and the inverse is normalized by $\frac{1}{(2\pi)^3}$. Also, in the exponents, i and $-i$ are sometimes switched or replaced by $\pm 2\pi i$, in the latter case the factor $\frac{1}{(2\pi)^3}$ disappears, as is illustrated by the different definition of the Fourier transform in Section 2.1.2, eqn 2.3.

3.3.1.2 *Discrete Fourier transform of an image* Now, let $\mathbb{L}^3$ denote a cuboidal lattice with lattice spacings $\Delta = \{\Delta_1, \Delta_2, \Delta_3\}$. That is, the unit cell C is a cuboid of edge length Δ. Let the data be given in the cuboidal window $W = [0, \Delta_1 m_1] \times [0, \Delta_2 m_2] \times [0, \Delta_3 m_3]$, where $m_i \in \mathbb{N}, i = 1, 2, 3$ are the numbers of pixels in x-, y-, and z-directions, respectively. Let

$$f : \mathbb{L}^3 \cap W \to \mathbb{R}$$

represent a real-valued image. Denote by $\hat{\mathbb{L}}^3$ the inverse lattice with unit cell

$$\hat{C} = [0, \frac{2\pi}{\Delta_1 m_1}] \times [0, \frac{2\pi}{\Delta_2 m_2}] \times [0, \frac{2\pi}{\Delta_3 m_3}] \quad \text{and} \quad \hat{W} = [0, \frac{2\pi}{\Delta_1}] \times [0, \frac{2\pi}{\Delta_2}] \times [0, \frac{2\pi}{\Delta_3}].$$

The discrete Fourier transform (DFT) of f is then defined by

$$(\mathcal{F}_{\mathbb{L}^3 \cap W} f)(w) = \frac{1}{\sqrt{\mathrm{vol}\, W}} \sum_{x \in \mathbb{L}^3 \cap W} f(x) \mathrm{e}^{-2\pi i w x} \quad \text{for} \qquad w \in \hat{W}. \tag{3.12}$$

The DFT of f thus yields an image with $m_0 m_1 m_2$ complex entries, whereas the window in frequency space is $\hat{W}$. The DFT defined in this way is $\hat{W}$-periodic in the sense that $(\mathcal{F}_{\mathbb{L}^3 \cap W} f)(w + k\frac{2\pi}{\Delta}) = (\mathcal{F}_{\mathbb{L}^3 \cap W} f)(w)$ for $k \in \mathbb{Z}^3$ and $w \in \hat{W}$. Here, $\frac{2\pi}{\Delta} = \left(\frac{2\pi}{\Delta_1}, \frac{2\pi}{\Delta_2}, \frac{2\pi}{\Delta_3}\right)$. Note that for increasing resolution but constant

window W (i.e. $\Delta_i \to 0,\ m_i \to \infty,\ \Delta_i m_i = \text{const.},\ i = 1,2,3$), the discrete Fourier transform $\mathcal{F}_{\mathbb{L}^3 \cap W}$ approximates the Fourier transform on W wrapped into a torus instead of the continuous Fourier transform $\mathcal{F}$ on $\mathbb{R}^3$. This can blur the result significantly if there is a period in the original that does not fit the window. Padding with zeroes overcomes this problem but increases memory use considerably. For a more detailed discussion see Koch *et al.* (2003).

The DFT as defined in eqn 3.12 maps the origin in frequency space onto the pixel $(0,0,0)$. Due to periodicity and symmetry ($\mathcal{F}_{\mathbb{L}^3 \cap W} f(w) = \overline{\mathcal{F}_{\mathbb{L}^3 \cap W} f(\frac{2\pi}{\Delta} - w)}$) of the DFT, this choice has no significance. However, this way of displaying is not consistent with the Fourier transform of macroscopically homogeneous microstructures (see Section 3.5.3) where periodicity is not expected and a surrounding of the zero frequency is of interest. Therefore, we use

$$(\mathcal{F}_{\mathbb{L}^3 \cap W} f)(w) = \frac{1}{\sqrt{\operatorname{vol} W}} \sum_{x \in \mathbb{L}^3 \cap W} f(x) \mathrm{e}^{-2\pi i (w - \frac{\pi}{\Delta})(x - \frac{m\Delta}{2})}, \quad \text{for} \quad w \in \hat{W}. \tag{3.13}$$

The translation moves the origin to the image centre. In particular, the zero frequency is in the image centre. Here, division by 2 means integer division and $m\Delta = (m_1\Delta_1, m_2\Delta_2, m_3\Delta_3)$.

The popularity of the discrete Fourier transform is mainly due to the fact that the *fast Fourier transform* provides a way to compute it very efficiently even for large images (Marcotte, 1996). The complexity of the discrete Fourier transform is $\mathcal{O}(M \log M)$ where M is the pixel number, $M = m_1 m_2 m_3$. There are various algorithms for its computation. It depends highly on the size of the image which algorithm performs best. The fastest ones work for pixel numbers that are powers of 2 or other prime numbers. Therefore, Frigo and Johnson (1998*a*, 1998*b*) developed a sophisticated meta-algorithm, which very roughly speaking chooses the best algorithm on its own – the FFTW ('fastest Fourier transform of the West'). The FFTW is a collection of fast C routines for computing the DFT. It adapts itself to both the problem to be solved (type and size) and to the hardware used.

3.3.2 *Morphological transforms*

We now introduce morphological transforms as a special class of filters that have a close correspondence to image analysis, i.e. the generation of features of segments or objects. In this sense, morphological transforms are a bridge between image processing and analysis. The theoretical background – mathematical morphology – has a strong flavour of analysis in its use in other fields of mathematics such as set theory and topology, see e.g. Serra (1982, 1988) and Soille (1999). We follow the line of classical morphology and introduce the morphological transforms most important for the filtering of segments or objects in binary images.

3.3.2.1 *Minkowski addition and dilation* The definitions of the basic morphological transforms go back to the definition of the Minkowski sum of sets. The expression

$$X \oplus Y = \{x + y \,:\, x \in X,\ y \in Y\}$$

is called the *Minkowski sum* of X and Y, the *Minkowski addition*, or *direct sum*.

By definition, the Minkowski addition $\oplus$ is commutative, $X \oplus Y = Y \oplus X$. Nevertheless, in image processing, the sets X and Y do not play the same role. By convention, the set X is associated with an image segment or an object and the set Y is called the *structuring element*. It has a meaning similar to the support of a filter mask. In order to embed morphology consistently into the theory of filtering we introduce the dilation of one set by another.

For a set $Y \subseteq \mathbb{R}^n$, denote by $\check{Y} = \{-y :\ y \in Y\}$ the reflection of Y at the origin. For $X, Y \subseteq \mathbb{R}^n$ the set $X \oplus \check{Y}$ is called the dilation of X by Y. Notice that in image processing the structuring element is in most cases symmetric, $\check{Y} = Y$. Then, the dilation is equivalent to the Minkowski addition, and, hence, one tends to ignore the differences between both transforms.

Both Minkowski addition and dilation can be applied to fill holes (caves) in sets (objects) and to close gaps between sets.

3.3.2.2 *Minkowski subtraction and erosion* The Minkowski subtraction of two sets is defined using the Minkowski addition. For $X, Y \subseteq \mathbb{R}^3$ the set

$$X \ominus Y = (X^c \oplus Y)^c \tag{3.14}$$

is called the *Minkowski subtraction* of the sets X and Y. It should be noted that Minkowski subtraction is not the inverse of the Minkowski addition.

Analogously to dilation, *erosion* is introduced as the Minkowski subtraction of the reflected set, $X \ominus \check{Y}$. Clearly, if Y is symmetric, then erosion is the same as Minkowski subtraction.

Morphological erosion (or the Minkowski subtraction) can be used to remove small (noisy) objects from an image and open up space between just-touching objects.

3.3.2.3 *Mean co-ordination number of sintered copper* Separating just-touching objects by morphological erosions is demonstrated in an application. Consider a copper sinter material with spherical sinter particles in an early stage of the sintering process, see Fig. 3.5(a). The sinter material may consist of m convex particles $X_1, \ldots, X_m$ that overlap partially. The set $X = \bigcup_{i=1}^{m} X_i$ forms the solid constituent, while X^c is the pore space. If $X_i \cap X_j \neq \emptyset$, a sinter neck is formed between the ith and the jth particle. Otherwise, if $X_i \cap X_j = \emptyset$, both particles do not touch, $i, j = 1, \ldots, m$, $i \neq j$. In the early stage of a sinter process, one can assume that $X_i \cap X_j \cap X_k = \emptyset$ for pairwise different indices. In this example the problem consists in the determination of the co-ordination number κ of the sinter necks, i.e. the mean number of sinter necks per particle.

Now we choose a structuring element Y (e.g. a ball of radius r centred at the origin, $Y = B_r$) such that in $X \ominus Y$ the space between all objects is opened but none of the objects is completely removed from the image, see Figs. 3.5(b)–(d).

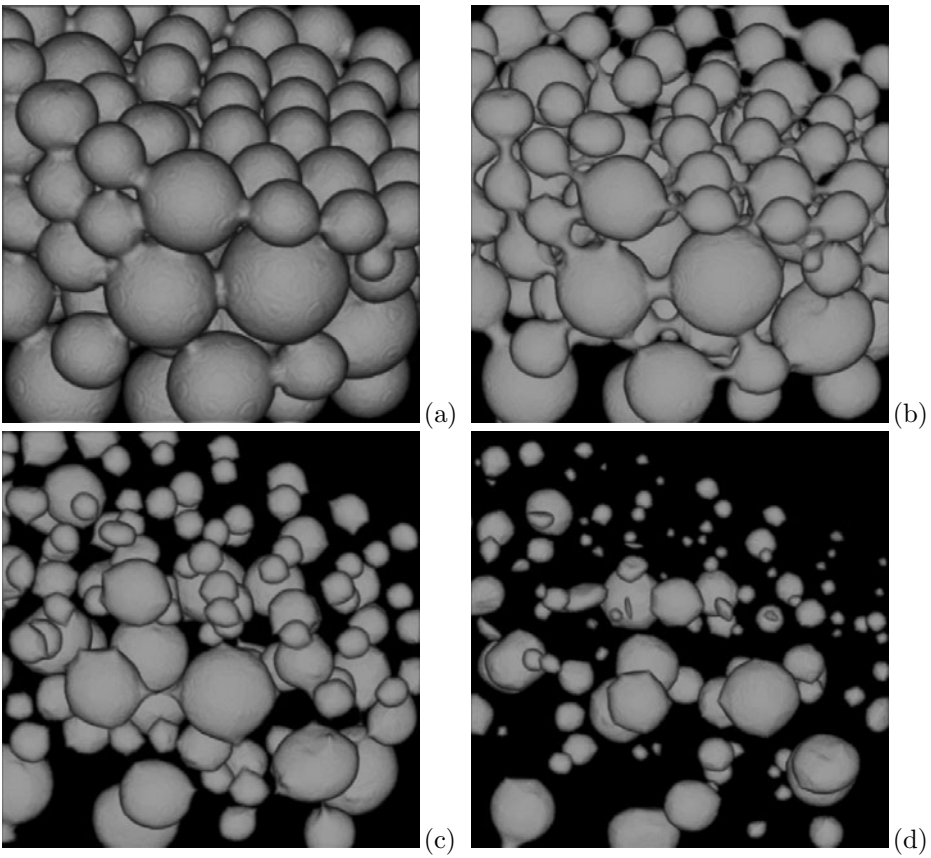


FIG. 3.5. Stepwise erosion of a copper sinter specimen: (a) original image, (b), (c), (d) eroded with a ball of increasing radius. Original image: X-ray CT, Jürgen Goebbels, BAM, pixel size 13.9 µm. Sample: Karsten Pischang, TU Dresden, ball diameter approx. 1mm.

Then $\chi(X \ominus Y)$ is the number of particles and $\chi(X \ominus Y) - \chi(X)$ is the number of sinter necks. Since each sinter neck belongs to two particles, the co-ordination number κ is

$$\kappa = 2\frac{\chi(X \ominus Y) - \chi(X)}{\chi(X \ominus Y)},$$

where the quantities $\chi(X)$ and $\chi(X \ominus Y)$ are measured directly from the corresponding binary images.

3.3.2.4 *Morphological opening and closure* For two sets $X, Y \subset \mathbb{R}^3$ the *opening* $X \circ Y$ of the set X with the set Y is an erosion followed by a Minkowski addition and the *closure* $X \bullet Y$ is a dilation followed by a Minkowski subtraction,

$$X \circ Y = (X \ominus \check{Y}) \oplus Y,$$
$$X \bullet Y = (X \oplus \check{Y}) \ominus Y.$$

A morphologically opened set has an appearance similar to that of the original set, but it is based only on the points of the set that survived the initial erosion. Morphological opening is one of the most commonly used image-processing techniques for removing pixel noise and thus cleaning up binary images. For the fibre-felt data in Fig. 3.6 a morphological opening was used to extract the thick fibres.

Opening tends to separate small objects. Sequential use of morphological openings with increasing size of the structuring element simulates a sieving procedure with increasing mesh width. For such a 'sieving procedure' a ball of radius r is used as the structuring element, where r is the parameter of the mesh width. The result is a (spherical) granulometry yielding local pore sizes or local thickness values.

Morphological closure is the corresponding processing of the complementary set. By means of a closure, small holes in objects or gaps between neighbouring objects are closed.

3.3.2.5 *Top-hat transforms* The difference

$$X \setminus (X \circ Y)$$

is called the *top-hat* transform of the set X with respect to Y where $\setminus$ denotes the set difference. The set $X \setminus (X \circ Y)$ is always part of the foreground and emphasizes small foreground structures. For the fibre felt from Fig. 3.6, a top-hat transform yields the system of the thin fibres while removing the thick ones. The transform

$$(X \bullet Y) \setminus X$$

is said to be a *top-hat*, too. The result $(X \bullet Y) \setminus X$ belongs to the background X^c and features small holes in the structure X as well as small background structures.

3.3.3 *Geodesic morphological transforms*

In this section, two complex morphological transforms on grey-value images are introduced. In contrast to the morphological transforms described so far, geodesic morphological transforms operate on two input images. Typically, one input image is the original one, while the other is a morphologically transformed version of it.

There is a wide range of geodesic morphological transforms serving various purposes such as filling holes or detecting regional extrema, for an overview see Soille (1999, Chapter 6). Here, the h-minima transform and reconstruction by dilation are explained. The h-minima transform is the best filtering method for the reconstruction of foam cells, see Section 3.4.3. Reconstruction by dilation is an efficient way of hysteresis thresholding, see Section 3.4.1.3.

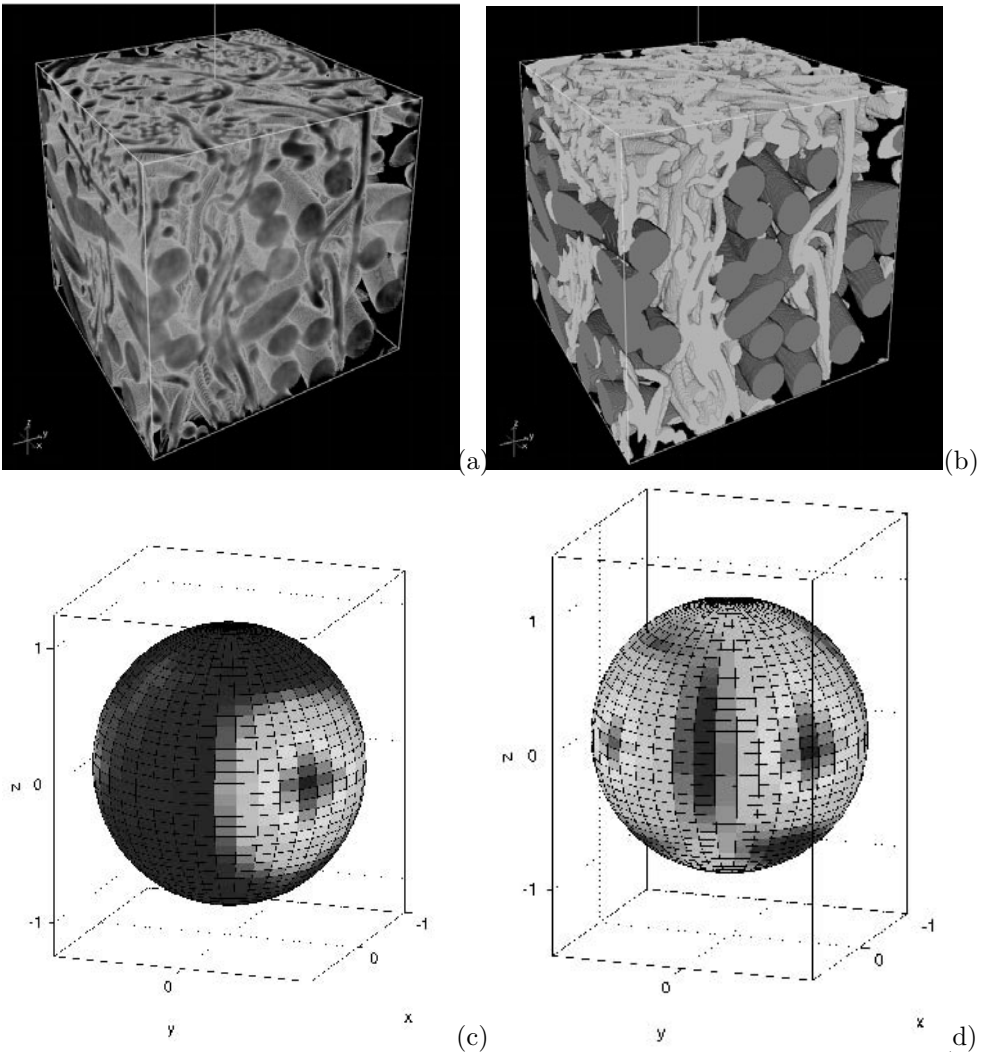


FIG. 3.6. Fibre felt. (a) original, (b) result of morphological separation of the thin (blue) and the thick fibres (red), (c,d) visualizations of the directional distribution of the fibres in a typical point for the thick and the thin fibres. The thick fibres are concentrated in the x-direction, while the thin fibres have two preferred directions – the x- and the y-direction. Original image: phase–contrast method, Lukas Helfen, ESRF, pixel size 6.65 μm. Sample: Heimbach Düren GmbH & Co. KG, Düren. Images (b–d) are reproduced in **colour** in the central part of the book.

3.3.3.1 *Reconstruction by dilation* Reconstruction by dilation performs successive morphological dilations on a so-called *marker image* restricted by a second image – the *mask image*. The grey-values of the mask image have to be pixelwise higher than those of the marker image. The marker image is dilated stepwise with the chosen neighbourhood as structuring element. More precisely, a maximum filter (see Section 3.3.5.1) is applied, where the filter mask is given by all pixels being connected to the current pixel by an edge of the neighbourhood graph. After each dilation step, the pixelwise minimum of mask image and dilated marker image is taken. The reconstruction procedure stops when the marker image cannot be modified any further.

Reconstruction by dilation serves as the basis for many other geodesic morphological transforms.

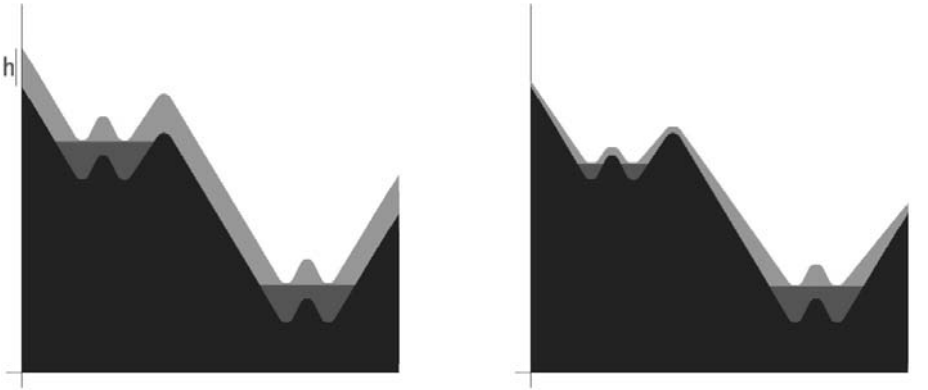
3.3.3.2 *h-minima transform* The h-minima transform removes local grey-value minima depending on their *dynamic*. In order to explain the idea of this transform, interpret the grey-values as heights in a topological relief with bright pixels corresponding to mountain tops and dark pixels corresponding to valleys.

The dynamic of a local minimum is the minimal height (number of grey-levels) one has to overcome in order to reach a lower local grey-value minimum. The local minima with dynamic lower than the parameter h are removed in the following way: 'valleys' in the grey-value relief are filled until either the local minimum is increased by h or water would spill into a valley of dynamic higher than h. Thus, spurious local minima are removed, while the remaining image information remains intact. The h-minima transform is a special case of reconstruction by dilation with the original image $\{f(x) : x \in \mathbb{L}^3 \cap W\}$ serving as the marker image and $\{f(x)+h : x \in \mathbb{L}^3 \cap W\}$ as the mask image, see Fig. 3.7 for a sketch of the principle.

For structures with strongly varying pore or cell sizes such as the ceramic grain in Fig. 3.15, the parameter h can not be chosen such that non-relevant local minima in large cells disappear without removing local minima corresponding to centres of small cells. This problem can be overcome by adapting the parameter h to the total grey-value, see Godehardt and Schladitz (2006). After inversion of the distance image, centres of small pores correspond to local minima at a high absolute grey-value. Centres of large pores are given by local minima at low absolute grey-values. Thus, the filtering should be strong at low grey-values and very cautious at high grey-values. Choosing $h = h(f(x))$ as a linearly decreasing function of the grey value fulfils these requirements. See Fig. 3.7 for a sketch, Fig. 3.15 for the application, and Section 3.4.3 for the description of the pore-reconstruction method.

3.3.4 *Linear filters*

Linear filters are applied e.g. to reduce noise in an image, to detect edges, and to obtain a reference signal containing the shading of the image.



(a) (b)

FIG. 3.7. h-minima (a) and height adaptive h-minima transforms (b) on a 1D example. Dark grey: original image. Light grey: grey-values increased by h or $h(f(x))$, respectively. Medium grey: difference between original image and result. The fixed value for h removes the local minima both at high and low grey-values, (a). The adaptive transform filters the low grey-value minima while conserving the high grey-value minima, (b).

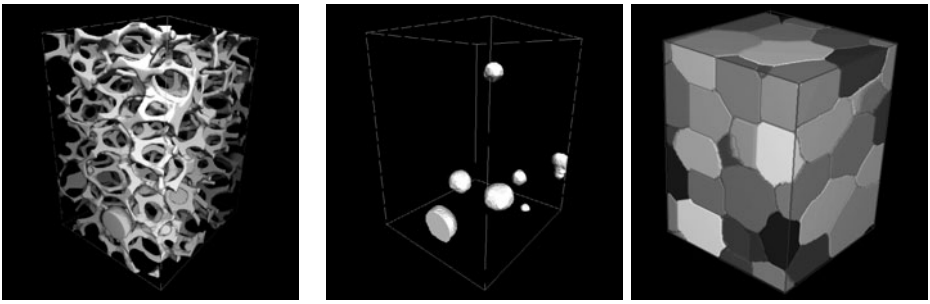
3.3.4.1 *Continuous case* First, we consider the continuous case and we introduce *linear filtering* as the convolution $f * g$ of real-valued functions f and g where the first function represents our signal (a grey-value image) and the second function is interpreted as the *filter mask*,

$$(f * g)(x) = \int_{\mathbb{R}^3} f(y)g(x-y)\,dy, \qquad x \in \mathbb{R}^3.$$

Roughly speaking, the signal f is filtered with g and the convolution of f and g is called the *filter response*. In some cases it is useful to perform the convolution in inverse space. In particular for large filter masks, i.e. for a filter function g with large support, the use of Fourier techniques can reduce computation time considerably. We write $g'(x) = g(-x)$, $x \in \mathbb{R}^3$ for the reflection of g at the origin. If both functions are integrable, their Fourier transforms $\hat{f} = \mathcal{F}f$ and $\hat{g} = \mathcal{F}g$, exist and, furthermore, if f and g are 'smooth' functions of rapid decay, then the inverse Fourier transform $\bar{\mathcal{F}}$ of the convolution $f * g$ can be written in the form

$$f * g = \bar{\mathcal{F}}[\hat{f} \cdot \widehat{g^*}]. \tag{3.15}$$

From the last equation it follows that $(f * g) * h = f * (g * h)$, i.e. the convolution is associative. This leads to the term 'separability' of linear filters. The function g is called *separable* if there exists an integer $m > 1$ and functions $g_1, \ldots, g_m$



(a) (b) (c)

FIG. 3.8. (a) Open-cell aluminium foam containing spherical inclusions in some of the cells, (b) the inclusions extracted from the original structure by an opening. Original image: X-ray CT, Sven Gondrom, Fraunhofer EZRT, pixel size 74 µm, visualized are 250 × 220 × 300 pixels. Sample: Dieter Girlich, m-pore GmbH, Dresden, (c) reconstructed cells without the strut system in a subvolume.

such that $g = g_1 * \ldots * g_m$. The separability of filters can be utilized to reduce computation time. In particular for three-dimensional images and for large filter masks, the saving is considerable.

Linear filters, in particular linear smoothing filters, should have the following properties, see Jähne (2005, 11.2):

(i) The application of a smoothing filter is free of translation if the filter mask is *symmetric*, $g = g'$.

(ii) The filter mask g is called *isotropic* if $g(x) = g(y)$ for all $x, y \in \mathbb{R}^3$ with $\|x\| = \|y\|$.

(iii) The filter mask g is called *normalized* if $\|g\| = 1$.

(iv) The Fourier transform $\hat{g}$ is monotonously decreasing, i.e. $\hat{g}(x) \leq \hat{g}(y)$ for $\|x\| \geq \|y\|$.

The last requirement is motivated by the fact that smoothing filters should affect small scales stronger than coarse scales, see Jähne (2005, 11.2.3).

3.3.4.2 *Discrete case* The signal f is now observed on a homogeneous lattice $\mathbb{L}^3$ with base $\{u_1, u_2, u_3\}$ and unit cell C. The digitization $f_{\mathbb{L}^3} : \mathbb{L}^3 \mapsto \mathbb{C}$ of f may be given by $f_{\mathbb{L}^3}(x) = f(x)$ for $x \in \mathbb{L}^3$ (Gauss digitization). Consider a discrete version $g_{\mathbb{L}^3}$ of g, which can be seen as the filter function g scattered on $\mathbb{L}^3$. Then the discrete convolution $\tilde{*}$ of $f_{\mathbb{L}^3}$ and $g_{\mathbb{L}^3}$ can simply be obtained from eqn 3.15 by applying the rectangular quadrature rule corresponding to $\mathbb{L}^3$,

$$(f_{\mathbb{L}^3} \tilde{*} g_{\mathbb{L}^3})(x) = |C| \sum_{y \in \mathbb{L}^3} f(y) g(x - y), \qquad x \in \mathbb{L}^3. \tag{3.16}$$

Obviously, $f_{\mathbb{L}^3} \tilde{*} g_{\mathbb{L}^3}$ can be considered as an approximation of $(f * g)$ in the following sense. If the functions f, g are bounded and continuous and $(f * g)(x)$ is finite for all $x \in \mathbb{R}^3$, then

$$f_{a\mathbb{L}^3} \tilde{*} g_{a\mathbb{L}^3} = (f * g)_{a\mathbb{L}^3}$$

as the lattice distance tends to zero, $a \to 0$. For bounded support Y, the filter mask $g_{\mathbb{L}^3}$ can be represented as a matrix A consisting of the values of $g_{\mathbb{L}^3}$ for at least those x belonging to $Y \cap \mathbb{L}^3$.

In order to give examples of the three-dimensional case, we introduce the convolution $A * B$ of matrices $A = (a_{ijk})$ and $B = (b_{ijk})$, where A and B may be of dimensions (m_1, m_2, m_2) and (m_1', m_2', m_3'), respectively. Then $A * B$ is a matrix of dimension $m_1 + m_1' - 1, m_2 + m_2' - 1, m_3 + m_3' - 1$ with coefficients

$$\sum_i \sum_j \sum_k a_{ijk} b_{i'-i, j'-j, k'-k}.$$

The summation over i is from $\max\{1, i' - m_1'\}$ to $\min\{m_1, i' - 1\}$, and analogously over j and k.

3.3.4.3 *Examples of filters* Consider the case of the cubic lattice with lattice distance 1, i.e. $\mathbb{L}^3 = \mathbb{Z}^3$.

(i) *The* $2 \times 2 \times 2$*-mean value filter* is a smoothing filter represented by the matrix

$$A = \frac{1}{8} \left(\begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \right).$$

The mean-value filter is not separable, e.g. the convolution $A * A$ does not yield the mask of the $3 \times 3 \times 3$-mean value filter.

(ii) *The* $3 \times 3 \times 3$*-mean value filter* is represented by the matrix

$$\frac{1}{27} \left(\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \right).$$

This filter mask is the convolution of the three mean-value filters given by the matrices

$$\frac{1}{4} \left(\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 1 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \right),$$

$$\frac{1}{4} \left(\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \right),$$

$$\frac{1}{4} \left(\begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \right),$$

which represent linear filtering in the directions of the basis vectors u_1, u_2, and u_3, respectively. In this sense, the mean-value filter is separable.

(iii) *Binomial filters.* The matrix A given in (i) can also be interpreted as the $2 \times 2 \times 2$-binomial filter, and the matrix $B = A * A$ corresponds to the $3 \times 3 \times 3$-binomial filter,

$$B = \frac{1}{64} \left(\begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{pmatrix}, \begin{pmatrix} 2 & 4 & 2 \\ 4 & 8 & 4 \\ 2 & 4 & 2 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{pmatrix} \right).$$

Binomial filters are separable and the coefficients on the edges of the matrices are binomial coefficients. Notice that binomial filters are (in some sense) discrete versions of Gauss filters.

(iv) *Gradient filters* are used for edge detection. Let $x = (x_1, x_2, x_3)$ be a point in the Cartesian co-ordinate system. The gradient filter with the mask $G_1 = (1, -1, 0)$ can be considered as a discrete version of the first derivative $\frac{\partial}{\partial x_1}$. It works in the lattice direction u_1 and detects edges in $f_{\mathbb{Z}^3}$ that are parallel to the plane orthogonal to u_1. Smoothing related to the orthogonal direction, e.g. with the $1 \times 3 \times 3$-binomial filter leads to the mask

$$\frac{1}{16} \left(\begin{pmatrix} 1 & -1 & 0 \\ 2 & -2 & 0 \\ 1 & -1 & 0 \end{pmatrix}, \begin{pmatrix} 2 & -2 & 0 \\ 4 & -4 & 0 \\ 2 & -2 & 0 \end{pmatrix}, \begin{pmatrix} 1 & -1 & 0 \\ 2 & -2 & 0 \\ 1 & -1 & 0 \end{pmatrix} \right).$$

(v) *The $3 \times 3 \times 3$-Laplace filter.* First, we remark that the convolution $G_1 * G_1 = (1, -2, 1)$ is a discrete version of the second derivative $\frac{\partial^2}{\partial x_1^2}$. Let G_2 and G_3 be the gradient filters in the lattice directions of u_2 and u_3, respectively. Since the Laplace operator is the sum

$$\Delta = \frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} + \frac{\partial^2}{\partial x_3^2},$$

the mask of a Laplace filter is

$$\begin{aligned} &G_1 * G_1 + G_2 * G_2 + G_3 * G_3 \\ &= \left(\begin{pmatrix} 0 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & -1 & 0 \\ -1 & 6 & -1 \\ 0 & -1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \right). \end{aligned}$$

A Laplace filtering including a smoothing can be obtained from $\Delta(f * g)$ where g is the filter mask of a Gauss filter. Since $\Delta(f * g) = f * \Delta G$, a smoothed Laplace filter is a discrete version of the mask Δg. In the literature, various discretizations can be found. As an example, we present the mask

$$\frac{1}{64} \left(\begin{pmatrix} -1 & -2 & -1 \\ -2 & -4 & -2 \\ -1 & -2 & -1 \end{pmatrix}, \begin{pmatrix} -2 & -4 & -2 \\ -4 & 48 & -4 \\ -2 & -4 & -2 \end{pmatrix}, \begin{pmatrix} -1 & -2 & -1 \\ -2 & -4 & -2 \\ -1 & -2 & -1 \end{pmatrix} \right).$$

Note that none of the filter masks fulfil the properties (i) to (iv) on page 61. In particular, the filter masks are not isotropic and in most cases their discrete Fourier transforms include considerable contributions of higher frequencies. Hence, in the literature, various versions of smoothing filters and edge-detection filters with 'improved filter responses' are given, see e.g. Jähne (2005).

Using the discrete convolution eqn 3.16, linear filtering with a cubic mask of edge length ℓ is of complexity $\mathcal{O}(\ell^3)$ per pixel. If separability can be exploited, the complexity reduces to $\mathcal{O}(3\ell)$ per pixel. For large masks, linear filtering should be performed via the inverse space and using the fast Fourier transform. Consider an image with m^3 pixels. Its Fourier transform is of complexity $\mathcal{O}(3m^3 \log m)$, i.e. $\mathcal{O}(3 \log m)$ per pixel. This means that a filtering of the image with a cubic mask of edge length ℓ via the inverse space is advantageous if $3 \log m \ll \ell^3$ for non-separable masks and if $\log m \ll \ell$ for separable masks.

3.3.5 *Non-linear filters*

3.3.5.1 *Morphological filters* First, we remark that there is a close correspondence between the Minkowski addition of sets and the supremum of a function. For sets $X, Y \subset \mathbb{R}^3$

$$\mathbb{I}_{X \oplus Y}(x) = \sup\{\mathbb{I}_X(x+y) \,:\, y \in Y\}.$$

Analogously, the Minkowski subtraction is related to the infimum by

$$\mathbb{I}_{X \ominus Y}(x) = \inf\{\mathbb{I}_X(x+y) \,:\, y \in Y\}.$$

This consideration leads to a generalization of the morphological transforms where the characteristic function $\mathbb{I}_X$ is replaced with an arbitrary bounded function $f : \mathbb{R}^3 \mapsto \mathbb{R}$. We write

$$f^Y(x) = \sup\{f(x+y) \,:\, y \in Y\},$$
$$f_Y(x) = \inf\{f(x+y) \,:\, y \in Y\}.$$

The supremum with respect to a set Y is called the *maximum filter* and the infimum with respect to Y is said to be the *minimum filter*. Both the minimum filter and the maximum filter belong to the class of morphological filters containing also the dilation filter (the supremum related to the reflection $\check{Y}$ of Y) and the erosion filter (the infimum related to $\check{Y}$), see also Soille (1999). As the corresponding morphological transforms the supremum and the infimum are separable in the sense that

$$f^{X \oplus Y} = (f^X)^Y, \qquad f_{X \oplus Y} = (f_X)_Y$$

for all sets $X, Y \in \mathbb{R}^3$ and all bounded functions $f : \mathbb{R}^3 \mapsto \mathbb{R}$. The relations between the supremum and the infimum on the one hand and the corresponding morphological transforms on the other hand can also be shown by the following

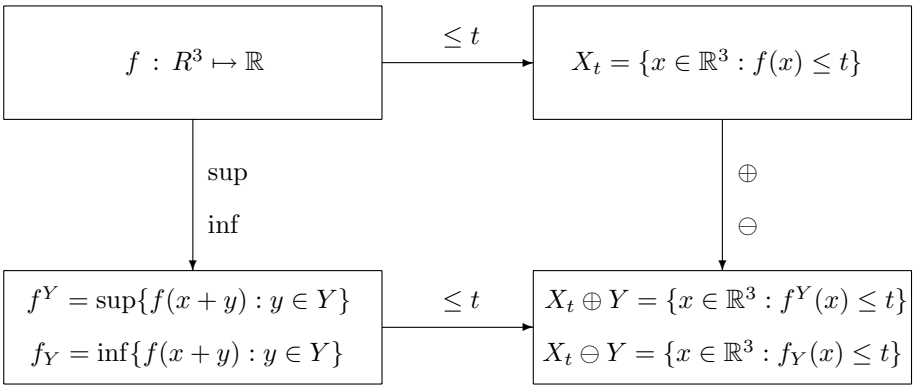


FIG. 3.9. Scheme showing relations between morphological transforms of sets and morphological filters.

consideration. Let a bounded function $f : \mathbb{R}^3 \mapsto \mathbb{R}$ be given and define the set X_t as the thresholding of f at the level t,

$$X_t = \{x \in \mathbb{R}^3 \,:\, f(x) \leq t\}, \qquad t \in \mathbb{R}.$$

Then it follows that

$$X_t \oplus Y = \{x \in \mathbb{R}^3 \,:\, f^Y(x) \leq t\},$$
$$X_t \ominus Y = \{x \in \mathbb{R}^3 \,:\, f_Y(x) \leq t\}.$$

This means that thresholding a signal followed by a Minkowski addition (Minkowski subtraction) leads to the same result as taking the supremum (infimum) followed by the thresholding, see also Fig. 3.9.

Finally, we note that for a digitization $f_{\mathbb{L}^3}$ of f and a subset Y' of lattice points the filters defined by

$$(f_{\mathbb{L}^3})^{Y'}(x) = \max\{f_{\mathbb{L}^3}(x+y) \,:\, y \in Y'\},$$
$$(f_{\mathbb{L}^3})_{Y'}(x) = \min\{f_{\mathbb{L}^3}(x+y) \,:\, y \in Y'\}$$

are discrete versions of the morphological filters, the *maximum filter* and the *minimum filter*, respectively.

As the corresponding morphological transforms and some of the linear filters, minimum and maximum filters are separable filters. In order to give an impression of the impact of separability we consider a cubic mask with edge length ℓ. If separability is not exploited, the complexity of a morphological filter is of order $\mathcal{O}(\ell^3)$ per pixel. If separability is used, the complexity reduces to $\mathcal{O}(\ell)$ per pixel.

3.3.5.2 *Diffusion filters* Smoothing filtering removes noise from the image but it also destroys valuable information about edges of segments (constituents) or

objects (particles). Thus, we are interested in constructing filters that reduce the noise but keep edge information. In this context we consider so-called *locally adaptive smoothing filters* where the sizes of the filter masks are adapted to local characteristics of the image. If the centre of the filter mask is close to an edge, the smoothing effect should be low, and hence we choose a small mask. Otherwise, strong smoothing is desirable in the interior of the segments or objects and thus we adjust a large mask for positions far from edges. This means the first step in edge-preserving smoothing is an appropriate edge detection.

There are various techniques for edge-preserving smoothing. Most of them have been developed for special applications and thus they can be applied successfully only under restrictive conditions. An important class of edge-preserving smoothing filters that are of common interest are *diffusion filters*. Diffusion filtering is based on *Fick's second law of diffusion*. Consider a function $f : \mathbb{R}^3 \times \mathbb{R}_+ \mapsto \mathbb{R}$ that models the (continuous) processing of a real function, i.e. $f(x,t)$ is the (real) grey-value at position $x \in \mathbb{R}^3$ and processing step $t \in \mathbb{R}_+$. In terms of diffusion, f is a concentration depending on location x and time t. Fick's second law states

$$\frac{\partial f}{\partial t} = \nabla(D\nabla f), \tag{3.17}$$

where ∇ denotes the Nabla operator and $D > 0$ is the diffusion coefficient that usually depends on the concentration, $D = D(f)$. This inhomogeneous partial differential equation simplifies if the diffusion coefficient D is independent of f. In this case Fick's second law can be rewritten as

$$\frac{\partial f}{\partial t} = D\Delta f, \tag{3.18}$$

where Δ is the Laplace operator. For given initial values $f(x,0) = f_0(x), x \in \mathbb{R}^3$ the solution of eqn 3.18 can be given in a closed form; it is a convolution of f_0 with a function $g_{Dt} = g(x/\sqrt{2Dt})$ depending on the parameter Dt, $f(x,t) = \big(f_0 * g_{Dt}\big)(x)$, $x \in \mathbb{R}^3$, where g is the probability density function of the 3-dimensional standard normal distribution.

It should be noted that diffusion filtering presented in the literature is in any case more or less empirical. It is only motivated by Fick's second law but in no sense is it based on a solution of eqn 3.17 or eqn 3.18. Nevertheless, there are various attempts to exploit ideas from modelling of diffusion processes in order to close the gap between empirical methods and theory, see e.g. Perona and Malik (1990) and Weickert (1998). In principle, diffusion filtering works as follows: The Gauss filter

$$g_\Sigma(x) = \frac{1}{\sqrt{(2\pi)^3 \det \Sigma}} \exp\left(-\frac{1}{2} x' \Sigma^{-1} x\right) \qquad x \in \mathbb{R}^3$$

is applied to the function f_0 where the covariance matrix Σ itself is considered as a function of f_0. In order to stop (or to reduce) diffusion at the edges in f, Σ is made dependent on the gradient ∇f_0, i.e. $\Sigma = \Sigma(\nabla f_0)$. This corresponds

to the assumption that the diffusion coefficient in eqn 3.17 is a function of ∇f, $D = D(\nabla f)$. Notice that there is no Gauss function with $\Sigma(\nabla f_0)$ that solves eqn 3.17 even if $D = D(\nabla f)$.

In the simplest case of inhomogeneous but isotropic diffusion filtering one chooses an isotropic Gauss filter with covariance matrix $\Sigma = \sigma^2 I$, where σ is a function of $\|\nabla f_0\|$. Following Perona and Malik (1990) we use the setting

$$\sigma(\nabla f_0) = \frac{\sigma_0 \alpha^2}{\|\nabla f_0\|^2 - \alpha^2}, \tag{3.19}$$

where $\sigma_0 > 0$ and $0 < \alpha < \|\nabla f_0\|$ are parameters of the filter to be adjusted for applications. Obviously, $\sigma \approx \sigma_0$ for $\|\nabla f_0\| \ll \alpha$ and $\sigma \approx 0$ for $\|\nabla f_0\| \gg \alpha$. Clearly, in discrete cases where f_0 is known only at lattice points, we apply a discrete gradient filter and the Gauss filter is approximated by binomial filters with discrete mask sizes. Furthermore, often a smoothed gradient is applied instead of ∇f_0, which usually stabilizes the response of the diffusion filter. For further ideas of regularization, see Weickert (1998).

Diffusion filtering can be summarized in the following two steps:

1. Computing a reference image with the pixel entries $\sigma(\nabla f)$. This usually involves a smoothing of $\|\nabla f\|$ with a binomial filter of constant mask size.
2. Applying a locally adaptive binomial filter to the original image where the mask size is computed from the local entry of the reference image.

Since diffusion is stopped totally at the edges, isotropic diffusion filtering cannot remove noise along the edges. To overcome this problem diffusion is stopped on the edge only along its normal direction, which is the direction of the gradient. Then, diffusion perpendicular to the normal direction leads to a smoothing along the edges. Before introducing anisotropic diffusion filtering we first have a look at the partial differential equations eqn 3.17 and eqn 3.18. In both equations, the diffusion coefficient is replaced with a positive-definite diffusion tensor D characterizing the anisotropic properties of the diffusion process. Again, for an initial setting $f(x,t) = f_0(x)$ the solution of eqn 3.18 can be given in a closed form. Then $f(x,t) = \big(f_0 * g_{Dt}\big)(x)$ solves eqn 3.18, where g_{Dt} is now the probability density function of the 3-dimensional Gauss distribution with covariance function $\Sigma = Dt$.

This is the motivation to assume that the diffusion tensor in eqn 3.17 depends on ∇f. As a consequence we choose Σ with the eigenvalues $\lambda_1 = \sigma(\nabla f)$ according to eqn 3.19, and $\lambda_i = \sigma_0$, $i = 2, 3$. The first eigenvector x_1 of Σ belonging to λ_1 is chosen as the normalized gradient, $x_1 = \nabla f / \|\nabla f\|$, while the others are arbitrary (but orthogonal to each other as well as orthogonal to x_1).

For further details about anisotropic diffusion filters and their implementation, see Weickert (1998) and Jähne (2005). Here, we only note that there is a significant difference between the implementations of isotropic and anisotropic diffusion filters. In the case of an anisotropic diffusion the reference image stores

the gradient itself, i.e. the reference image is now a vector-valued image. In particular, for higher-dimensional images this can need huge amounts of memory. Since, in the discrete case, there is only a finite number of binomial filters used to approximate the anisotropic Gauss filters, the reference image should contain only the indices of the corresponding binomial filters.

3.3.5.3 *Rank-value filters* The maximum and minimum filters introduced in Section 3.3.5.1 belong to the class of the so-called *rank-value filters* that are introduced now in their general setting.

Let $f : \mathbb{R}^3 \mapsto \mathbb{R}$ be a locally integrable function and let $Y \subset \mathbb{R}^3$ be a compact set with non-empty interior. In order to introduce rank-value filters we first consider the function

$$F_Y(t) = \frac{1}{\operatorname{vol} Y} \int_Y \mathbb{I}(f(x) \leq t)\, dx,$$

which is monotonously increasing and takes values between 0 and 1. For given t, $F_Y(t)$ is the volume fraction of the subset of points $x \in Y$ with $f(x) \leq t$.

Furthermore, we introduce the quantity t_α such that $F_Y(t_{Y,\alpha} - 0) \leq \alpha \leq F_Y(t_{Y,\alpha})$, where α is a real number, $0 \leq \alpha \leq 1$. The quantity $t_{Y,\alpha}$ can be interpreted as a threshold.

Since F_Y has the same properties as a probability distribution function, the quantity $t_{Y,\alpha}$ is called the *quantile of* F_Y *of order* α. In particular the threshold $t_{Y,1/2}$ is called the *median*.

This motivates us to introduce quantile filters with mask Y and of order α. If Y is shifted by x, the quantile also depends on the position x. Now, the filter response of the *quantile filter* of a locally integrable function f is defined by

$$F_{Y+x}(t_{Y,\alpha}(x) - 0) \leq \alpha \leq F_{Y+x}(t_{Y,\alpha}(x)), \qquad x \in \mathbb{R}^3.$$

For $\alpha = 0$ and $\alpha = 1$ we obtain the minimum and maximum filter from Section 3.3.5.1, respectively. If f is a continuous function, then $t_{Y,0} = f_Y$ and $t_{Y,1} = f^Y$. The responses of the minimum and maximum filters are discrete versions of $t_{Y,0}$ and $t_{Y,1}$. Furthermore, since the order is an adjustable parameter, the combination $t_{Y,\alpha} \circ t_{Y,1-\alpha}$ extends the morphological opening and closure of functions defined in Section 3.3.2 considerably. For an α between 0 and 1 this combination can be applied successfully if the use of morphological opening or closure is hampered by noise.

In the special case $\alpha = \frac{1}{2}$ we obtain the *median filter*, which is the simplest and most widely used edge-preserving smoothing filter. The function $t_{Y,1/2}(x)$ is the corresponding filter response.

In order to sketch the implementation of quantile filters we consider the discretization $f_{\mathbb{L}^3}$ on a homogeneous lattice $\mathbb{L}^3$. Implementations are based on sorting the set $\{f_1, \ldots, f_m\}$ of values $f_i = f(x_i)$ on the set of lattice points $(Y + x) \cap \mathbb{L}^3 = \{x_1, \ldots, x_m\}$ in ascending order. The sorting can be speed

up considerably by Shell's method – a quick sorting algorithm of complexity $\mathcal{O}(m \log m)$. Finally, from the set $\{f'_1, \ldots, f'_m\}$ of sorted pixel values one easily obtains the responses corresponding to the filters described above. In particular, the minimum f'_1 is an estimate of $t_{Y,0}(x)$, the maximum f'_m is an estimate of $t_{Y,0}(x)$, and if m is odd then $f'_{(m+1)/2}$ is an estimate of the median $t_{Y,1/2}(x)$. Sorting the values can be seen as a kind of ranking and, hence, the filters introduced above are said to be *rank-value filters*.

3.3.6 *Distance transforms*

Distance transforms operate on binary images and usually assign to each background pixel the distance to the foreground. The distance information can be used to measure pore sizes, to speed up successive dilations or erosions, as input for the watershed transform (Section 3.4.3), skeletonizations (Section 3.3.7) and granulometries (Section 3.3.2.4). Applications beyond the scope of this chapter are in image registration, object matching, and collision detection.

When measuring Euclidean distances there is a trade-off between accuracy of the result and complexity of the algorithm. Therefore, the distances are measured in relation to the Euclidean metric or discrete (Chamfer) metrics approximating the Euclidean. Cuisenaire (1999) gives a comprehensive overview of distance transforms for 3D images. For a more recent summary, see Maurer and Raghavan (2003).

Discrete or Chamfer distance transforms measure the length of digital paths with the set of possible directions as well as the weights for line segments in these directions given by a mask. The distance transform then consists of two sweeps of the mask through the image. The distance value of the current background pixel is the minimum of the distances of its lexicographically smaller (forward sweep) or greater (reverse sweep) 26-neighbours plus the respective value in the mask.

Due to the locality, Chamfer distance transforms are fast for small masks ($(3 \times 3 \times 3)$ or $(5 \times 5 \times 5)$). However, applications such as the reconstruction of foam cells, see Section 3.4.3, or skeletonization (Section 3.3.7) are considerably affected by the approximation error. Even larger masks do not yield a significant gain in accuracy, but slow down computation to an extent that makes exact algorithms competitive. Moreover, determination of the optimal mask values for non-cubic 3D lattices is an open problem.

Saito and Toriwaki (1994) developed the prototype of efficient algorithms for exact Euclidean distance transforms (EDT) in 3D using the fact that the problem is separable. First, by one forward and one backward scan, determine the nearest foreground pixel in each x-column. Then, propagate the distance information in each y-column. Finally, the z-columns are processed:

1. Initialize with $d(i,j,k) = 0$ for foreground pixels and $d(i,j,k) = \infty$ for background pixels.
2. Scan x-columns: $d(i,j,k) = \min\{(i-\ell)^2 : 0 \leq \ell < m_1\}$ for each j,k and for $i = 0, \ldots, m_1 - 1$.

3. Scan y-columns: $d(i,j,k) = \min\{d(i,\ell,k)^2 + (j-\ell)^2 : 0 \le \ell < m_2\}$.
4. Scan z-columns: $d(i,j,k) = \min\{d(i,j,\ell)^2 + (k-\ell)^2 : 0 \le \ell < m_3\}$.

Here, (m_1, m_2, m_3) denotes the number of pixels in x-, y-, and z-directions, respectively. This algorithm results in exact squared Euclidean distances and has complexity $\mathcal{O}(m^4)$ if $m_1 = m_2 = m_3 = m$. Note that for many applications, squared Euclidean distances, which save computation time and – more important – memory, are sufficient.

After generalizing several 2D algorithms to 3D and comparing them empirically, Cuisenaire (1999) suggests a hybrid (Cuisenaire, 1999, Section 6.4), which indeed performs very well in particular on very large images, where it has a nearly linear complexity: First, on the x-y slices, approximate distances are obtained by propagating the vector to the nearest foreground pixel within 4-pixel neighbourhoods (a signed version of the four-point sequential distance mapping by Danielsson (1980)). Subsequently, the approximate distance values are corrected by detecting the corners of the Voronoi tessellation generated by the foreground pixels, see Cuisenaire (1999, Chapter 5). Finally, the algorithm by Saito and Toriwaki (1994) is applied for each column in z-direction.

An algorithm for the Euclidean distance transform in arbitrary dimensions in $\mathcal{O}(M)$ is derived by Maurer and Raghavan (2003): The distance is determined by constructing the intersection of the Voronoi tessellation generated by the foreground pixels with the columns in all co-ordinate directions separately. Moreover, a slight alteration yields not just the distances but the vector to the nearest foreground pixel – a feature needed in various applications.

Note that the analytical complexity is not necessarily the decisive criterion for choosing an algorithm. EDT algorithms of linear complexity but with a very high constant (number of accesses per pixel) perform considerably worse than e.g. Cuisenaire's hybrid.

3.3.7 *Skeletonization*

Skeletonization is the process of reducing foreground regions in a binary image to a lower-dimensional subset preserving the original foreground connectivity. For some structures, in particular fibre or tube systems, the skeleton is an essential step towards the analysis. Moreover, skeletons are an intermediate step in reconstruction of materials by pore-throat systems, see e.g. Lindquist *et al.* (2000) and Benouali *et al.* (2002).

An intuitive definition of the skeleton is the cross-fire or wavefront propagation analogy (see e.g. Soille (1999)): The boundary of an object is set on fire and the fire is spreading uniformly within the body. The skeleton is the set of points where the fires meet. More formally, this is the set of the centres of all maximal inscribed balls. (An inscribed ball is maximal if it is not covered by any other inscribed ball.) The extension of this concept to discrete sets is not straightforward, as discrete skeletons should have the following properties:

(i) *topology preservation:* the Euler number should be the same before and after skeletonization

(ii) *one-pixel thickness:* the skeleton should consist of curves or surfaces and be as thin as possible

(iii) *medial position:* the skeleton pixels should have the same distance to the closest boundary pixels of the object

(iv) *rotation invariance:* skeletonization and object rotation should commute

(v) *noise immunity:* the skeleton should be fairly insensitive to noise (boundary pixels added or removed)

(vi) *reconstructibility:* it should be possible to reconstruct the original image from the skeleton.

In discrete space, most of these characteristics are mutually exclusive. Hence practical skeletonization methods, which are abundant, are a compromise between them. In many applications, e.g. computer graphics and medical, (ii)–(v) are more important than (i). Therefore, many skeletonization algorithms do not preserve topology in the strict sense of (i), which is nevertheless necessary for consistent quantitative analysis.

Skeletonization algorithms can be classified into three major categories – thinning techniques, skeletonization based on distance transforms, and analytical methods.

Thinning algorithms peel off objects layer by layer removing the pixels not necessary for preserving connectivity. Sequential thinning algorithms check for each current boundary pixel whether it can be removed without destroying the topology and to some extent also the shape of the object. Criteria for deletion are derived either from the simple point condition by Bertrand (1996) supplemented by a criterion for finding end-points that cannot be deleted (Tsao and Fu, 1981) or using a morphological hit-or-miss transform (Soille, 1999). Thinning preserves topology and can be adapted to all adjacencies discussed in Section 3.1.3. However, sequential thinning is very time consuming and noise sensitive. Parallel thinning as proposed e.g. by Tsao and Fu (1981) cannot guarantee topology preservation in 3D.

Alternatives are skeletonization algorithms based on distance transform. Discrete or Euclidean distance transforms can be used either for a thinning in the order given by the distance map or to detect the centres of maximal balls as local maxima in the distance image. However, the latter are not necessarily connected. Meijster *et al.* (2005) propose a modified Euclidean distance transform yielding the skeleton. However, this skeleton is not topology preserving in terms of condition (i).

A promising yet algorithmically demanding approach is proposed by Couprie and Zrour (2005). Following Remy and Thiel (2005) and Borgefors and Nyström (1997) they find the centres of maximal discs in the squared Euclidean distance image using look-up tables. The resulting medial axis is used as a constraint set for thinning the original object to the Euclidean skeleton. More precisely,

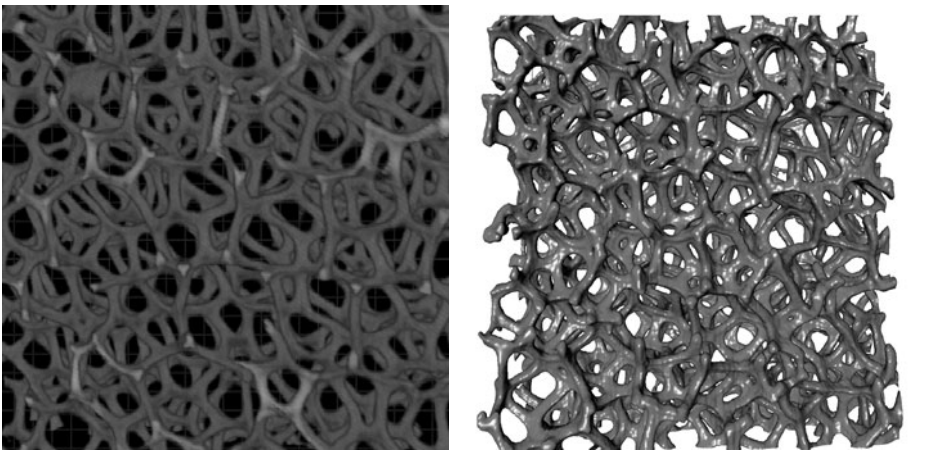


FIG. 3.10. Volume renderings using 3D texture mapping of the reconstructed tomographic image of an open-cell nickel foam. Original image: X-ray CT, Sven Gondrom, Fraunhofer EZRT, Saarbrücken, pixel size 10 µm.

the foreground pixels are checked for being skeleton pixels consecutively moving outwards from the medial axis. In order to deal with discretization effects and noise, the bisector function is computed subsequently. Then two parameters allow a controlled pruning of the skeleton: Pixels with squared radius of the maximal ball $< r$ are removed from the medial axis. Subsequently, pixels with bisector angle $< \alpha$ are removed from the skeleton. Both parameters control the degree of branching of the skeleton – r the large scale and α the small scale. Figure 3.11 shows an example of the skeleton for a subvolume of the nickel foam from Fig. 3.10.

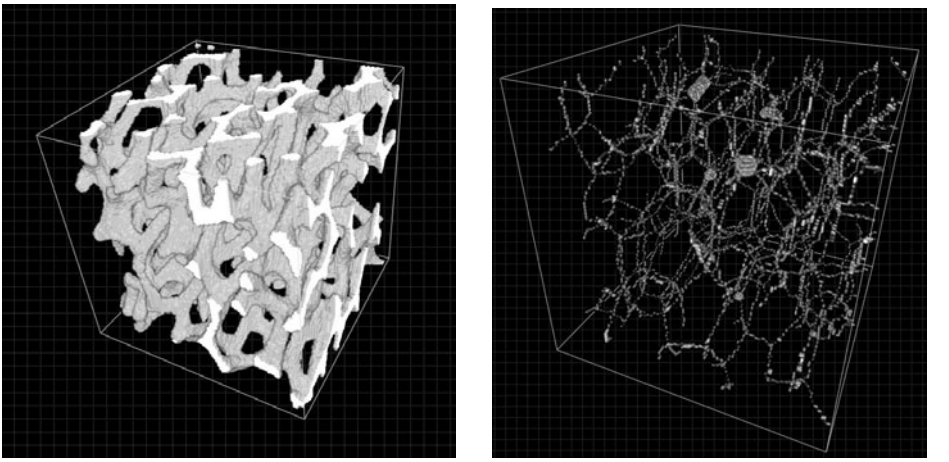
3.4 Segmentation

Usually, the concept *segmentation* is used for two tasks – finding the phase or image segment of interest in a grey-value image and identifying connected objects or regions. The first type of segmentation is called *binarization* in the following, as it results in a binary image having the segment of interest as foreground. Segmentation is crucial in image processing, as most methods of analysis work on segmented images, only. However, segmentation is an ill-posed problem, and often segmentation results are validated just visually. This causes particular problems in dimensions higher than two, where visual inspection can be misleading.

Segmentation methods can be roughly classified as grey-value, region, and shape-based methods. The easiest grey-value-based segmentation methods are thresholding methods using either global or local image information.

3.4.1 *Binarization*

3.4.1.1 *Global thresholding* In thresholding, one assumes that the image segment of interest can be characterized by its brightness. In particular, this as-



(a) (b)

FIG. 3.11. Skeleton for a 128^3 pixel subvolume of the nickel foam from Fig. 3.10. (a) original subvolume, (b) skeleton, skeletonization parameters ($r = \infty, \alpha = \pi$). The skeleton features some small spherical parts. These are the surface of pores within the original strut system. The Euler number for both original and skeleton is -231.

sumption is valid for CT data of porous materials: The material structure with density higher than air will result in bright pixels, indicating large attenuation.

The simplest method to binarize an image with integer grey-values f is to select a grey-value t, $\min_x f(x) \leq t \leq \max_x f(x)$, and create the binary image with values g according to

$$g(x) = \begin{cases} 1 & \text{if } f(x) \geq t \\ 0 & \text{otherwise.} \end{cases} \tag{3.20}$$

The threshold t has to be determined by the user. This can be done interactively, with the visual support of the grey-value histogram of the image and a fast preview showing the result just on one slice.

Otsu (1979) suggests selecting t such that it minimizes the weighted sum of within-class variances, which is equivalent to maximizing the between-class scatter. More precisely, let (p_ℓ) denote the relative frequencies of the grey-values from the grey-value histogram of the image. Then, the optimal threshold according to Otsu is the threshold value maximizing the 'distance' between the (two) classes foreground and background in the following sense

$$t_{\text{Otsu}} = \text{argmax}_t \left\{ \sum_{\ell<t} p_\ell \, (m_b - m)^2 + \sum_{\ell \geq t} p_\ell \, (m_f - m)^2 \right\}, \tag{3.21}$$

where

$$m = \text{mean}\{f(x)\},\ m_f = \text{mean}\{f(x) \mid f(x) \geq t\},\ m_b = \text{mean}\{f(x) \mid f(x) < t\}.$$

In the simple case of 256 grey-values, t_{Otsu} can be determined by just evaluating the right-hand side of eqn 3.21 for each grey-value and then choosing the threshold yielding the global maximum.

However, even if t was chosen optimally using a model fitting to the present image data, global thresholding will give poor results whenever the influence of noise is large compared to the image content or when grey-values fluctuate across the image as in 2D microscopic images with a shading.

3.4.1.2 *Local thresholding* Local thresholding can compensate for this last disadvantage of the global thresholding approach. A common approach for eliminating large-scale grey-value fluctuations is to calculate the mean grey-value within a window about each pixel and to subtract this sliding mean value from its grey-value. (This is equivalent to subtracting a 'reference' image as known from shading correction.) In the same spirit, but instead of modifying the image content prior to binarization, one uses a spatially varying threshold $t(x)$ such that

$$g(x) = \begin{cases} 1 & \text{if } f(x) \geq t(x) \\ 0 & \text{otherwise.} \end{cases} \tag{3.22}$$

Again, eqn 3.22 does not provide any clues on how the threshold values $t(x)$ should be computed. Trier and Jain (1995) obtained the best results among eleven tested algorithms using a method of Niblack (1986). Niblack's algorithm calculates local mean and standard deviation to obtain a threshold:

$$g(x) = \begin{cases} 1 & \text{if } f(x) \geq m_{M(x)}(f) + c\sqrt{\text{var}_{M(x)}(f)} \\ 0 & \text{otherwise.} \end{cases} \tag{3.23}$$

$m_{M(x)}(f)$ and $\sqrt{\text{var}_{M(x)}(f)}$ denote the local empirical grey-value mean and standard deviation, respectively, centred at pixel x and using the local window M. The parameters M and c depend on image content (object size) and noise, respectively, and have to be determined by trial-and-error in most cases. In 3D, this is rather time consuming and should be tried on a small subvolume first.

Mardia and Hainsworth (1988) use a Bayesian approach to find local thresholds. A similar approach is 'indicator kriging' by Oh and Lindquist (1999). The method relies mainly on local covariance estimates for thresholding and is similar to the method of Mardia and Hainsworth (1988) in that it estimates the local threshold value at a pixel using a linear combination of its neighbours, see Wirjadi (2007) for mathematical details.

3.4.1.3 *Hysteresis* Hysteresis, also known as *double thresholding*, was proposed by Canny (1986) as a method to segment connected edges from an edge strength map. Hysteresis thresholding uses two thresholds, $t_1 > t_2$, and starts from a pixel x with $f(x) \geq t_1$. Then, all pixels y neighbouring an already identified foreground pixel and fulfilling $f(y) \geq t_2$ are assigned to the foreground.

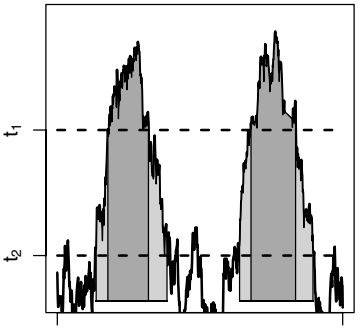


FIG. 3.12. Hysteresis thresholding demonstrated on a 1D example. The two regions are segmented completely, while the noise that would be segmented by just using t_2 as the global threshold is excluded. Dark grey: foreground after global thresholding with t_1. Light grey: regions added to the foreground due to exceeding the lower threshold t_2 and being connected to the dark grey regions. White: background.

This method ensures segmentation of connected image segments, since only definite foreground elements are selected according to the high threshold t_1, while their neighbours may have a lower value. At the same time, noise pixels are suppressed by the higher threshold t_1. See Fig. 3.12 for a sketch of the algorithm. Note that this algorithm can be seen as a special region-growing algorithm, see Section 3.4.1.4 below. Yet, hysteresis thresholding is implemented easiest using *reconstruction by dilation*, see Section 3.3.3, with a mask image given by the low-threshold pixels and the dilation series starting from the high-threshold seeds.

All thresholding techniques, whether global or local, suffer from the fact that they rely on grey-level information, exclusively. No prior information on object shape, structure or localization can be incorporated. Nevertheless, these methods can be very useful for initialization of more sophisticated algorithms such as region-based segmentation algorithms that usually require initialization with so-called seeds.

3.4.1.4 *Region growing* The philosophy behind all region-growing algorithms is that all pixels belonging to one object or image segment are connected and similar concerning some predicate. Starting from some seed pixels, regions are grown by checking whether neighbouring pixels fulfil some homogeneity criterion.

The easiest region-growing criterion is based on the observation that an object's grey-values are usually within some range about a mean value. Thus, while growing a region, its current mean and standard deviation are computed and a new pixel is added if its value does not depart too strongly from the region's mean:

1. Define a set of definite foreground pixels as seeds.

2. Initialize region R as the set consisting of one of the seeds.
3. Check all pixels x neighbouring the region R. Add x to R if and only if
$$|f(x) - m_R(f)| \leq c\sqrt{\mathrm{var}_R(f)}.$$
4. Repeat step 3 until R no longer grows.

Except for the region seed, the only parameter of this method is $c > 0$, the coefficient defining the allowed deviation from the region's mean. Since mean and variance can be updated as one pixel after another is added, this algorithm is easily implemented. It can give reasonable segmentation results where objects are connected and can be characterized by their grey-values.

Clearly, hysteresis as described in Section 3.4.1.3 can be interpreted as a region growing, too. The seeds are the pixels with $f(x) \geq t_1$ and a neighbour y is added to a region R if $f(y) \geq t_2$.

Apart from the choice of an appropriate neighbourhood system and the seed selection, the only difference between the numerous region-growing methods lies in the specification of the criterion for accepting a neighbouring pixel as being part of the region. It can be deduced using

- an edge map derived from an edge-detection algorithm,
- a decision function adapted to the region's size (the bigger the region, the tougher the criterion),
- boundary pixels as seeds.

A region-growing algorithm that can segment non-connected regions was proposed by Revol and Jourlin (1997) and extended to a parameter-free method (Revol-Muller *et al.*, 2002). In seeded region growing, seed selection is crucial but can be seen as an external task, often done by hand in medical image processing. Unseeded region growing has also been proposed but is beyond the scope of this section.

3.4.2 *Connected-component labelling*

Labelling of connected components ('objects') is one of the most important tools of image processing. It is the basis for the generation of object features as well as a kind of filtering, i.e. removing of noisy objects or holes in objects, where the criteria for an object or hole to be removed can be chosen extremely flexibly based on the object features. The task of labelling (object filling, region detection) is to assign labels (in most cases unsigned integers) to the pixels in such a way that all pixels belonging to a connected component of the image are assigned the same label, and pixels belonging to different components have different labels.

Due to its importance in image processing, plenty of literature has been written about labelling and techniques to control (and improve) the processing and memory demands, which can be tremendous for large images that arise frequently in practice.

3.4.2.1 *The basic algorithm*

The prototype of labelling algorithms is the simple and well-known Rosenfeld–Pfaltz method (Rosenfeld and Pfaltz, 1966, Klette and

Rosenfeld, 2004). Here, the image is scanned until a pixel x_k is found that has not yet been labelled. If the pixel x_k does not have any neighbours labelled in relation to the chosen adjacency system, a new label is chosen for x_k. Otherwise, if there is a neighbouring pixel x_j with the label ℓ_j, the label ℓ_j is assigned also to x_k. In the case of more than one neighbouring pixel having different labels, these labels are merged, which is noted in a table of pairs of equivalent labels. Finally, pixels belonging to the same equivalence class of labels form a connected component and are assigned the same label in a final scan of the image. The table of pairs can become very large, which may lead to problems in finding the equivalence classes; the complexity of a corresponding algorithm is $\mathcal{O}(m \log m)$ where m is the number of table entries. There are various versions of the Rosenfeld–Pfaltz method using techniques to keep m as small as possible.

3.4.2.2 *Advanced methods* Recent sophisticated labelling algorithms usually comprise a pre-processing step as an essential part. The pre-processing yields either a decomposition, as in Aguilera *et al.* (2002), or a more compact representation of the input image, and by that allows efficient data access or efficient memory use. The algorithm by Sandfort and Ohser (2007) features a run-length encoding of the input image as an essential pre-processing step. The run-length encoding compresses the data and accelerates the access to it considerably.

Most techniques, including the Rosenfeld–Pfaltz method and (Sandfort and Ohser, 2007), are 2-pass techniques, which means they run through the image twice, where in the first pass preliminary labels are assigned and label correspondences (pairs of equivalent labels) are collected, and in the second pass these correspondences are resolved into equivalence classes and the final labels are set. The resolving step is a critical issue and several methods have been proposed, some of which are described by Di Stefano and Bulgarelli (1999). A very efficient algorithm is explained by Thurfjell *et al.* (1992), see also Sandfort and Ohser (2007).

Instead of building the equivalence classes from all correspondences in the second run, one can do this during the first run by capturing every (preliminary) label in a new class and merging classes each time a correspondence occurs. Here, each label is internally mapped onto its current equivalence class identifier, so redundant equivalences can be avoided. Applying this idea by Di Stefano and Bulgarelli (1999) in classical algorithms, where every pixel is tested for correspondences, can improve performance considerably.

Park *et al.* (2000) discuss an alternative way to reduce the number of label equivalences relying on a partitioning of the image and a divide-and-conquer technique. Borgefors *et al.* (1997) use a recursion step to identify a complete set of connected pixels (an object) and to avoid the explicit construction of equivalence classes.

Single-pass techniques using contour tracing follow a completely different labelling concept. Although this method type eliminates the additional access to pixels for relabelling, the tracing becomes very complex (and inappropriate) for

images of dimensions higher than two. We finally remark that the introduction by Thurfjell *et al.* (1992) gives a short overview of different labelling algorithms, and Chapter 6 by Ritter and Wilson (2001) formalizes labelling operations on the basis of image algebra.

3.4.3 *Watershed transform*

A strong and indispensable tool for segmentation is the watershed transform assigning a connected region to each local minimum in a grey-value image. The transform can be interpreted as the flooding of a topographic surface, where the height is given by the grey-value of the corresponding pixel: All local minima are water sources. The water rises uniformly with growing grey-value. Pixels where waters from different sources meet are watershed pixels. The corresponding immersion algorithm of Vincent and Soille (1991) can be easily extended to 3D and adapted to arbitrary adjacency systems.

Application of the watershed transform is hampered by the oversegmentation caused by the fact that *each* local minimum is assigned an image region. Strategies for overcoming this problem are pre-processing, modification, and post-processing. However, the parameters for post-processing such as merging of basins are hard to deduce from the data. Therefore, only pre-processing and modification are discussed here.

Smoothing of the original grey-value image removes local minima but also image information. A more appropriate alternative are morphological transforms on grey-value images, in particular the h-minima and the height adaptive h-minima transforms, see Section 3.3.3. The latter enables correct segmentation of objects of different scales, see the example in Fig. 3.15.

A modification with straightforward interpretation is the volume-constrained pre-flooded watershed, where the grey-values of basins with a volume below the volume threshold are increased by one during immersion. Generalizations to other sets of constraints (e.g. relative or absolute height of minimum) are possible. Modification of the watershed algorithm is cheaper than pre-processing using grey-value morphology in terms of both memory and time. However, results are difficult to compare as the h-minima transforms work globally on the image, while the constraints during immersion only have local effects. Figure 3.13 compares the pore-reconstruction results for the open nickel foam (Fig. 3.10) without any smoothing, after smoothing with a binomial filter or the h-minima transform as well as with use of the pre-flooded watershed.

Combined with the Euclidean distance transform (Section 3.3.6), the watershed transform can be used for separation of connected particles. Examples are the alumina inclusions in refractory concrete or the pores in an open foam. To reconstruct the cells in an open foam, the binarization yielding the strut system is followed by the Euclidean distance transform on the pores. After filtering, the inverted distance image to remove unwanted local minima, the watershed transform is applied. (Inverted means here that each pixel gets as a new grey-value the difference of the maximal grey-value for the given image type and its old

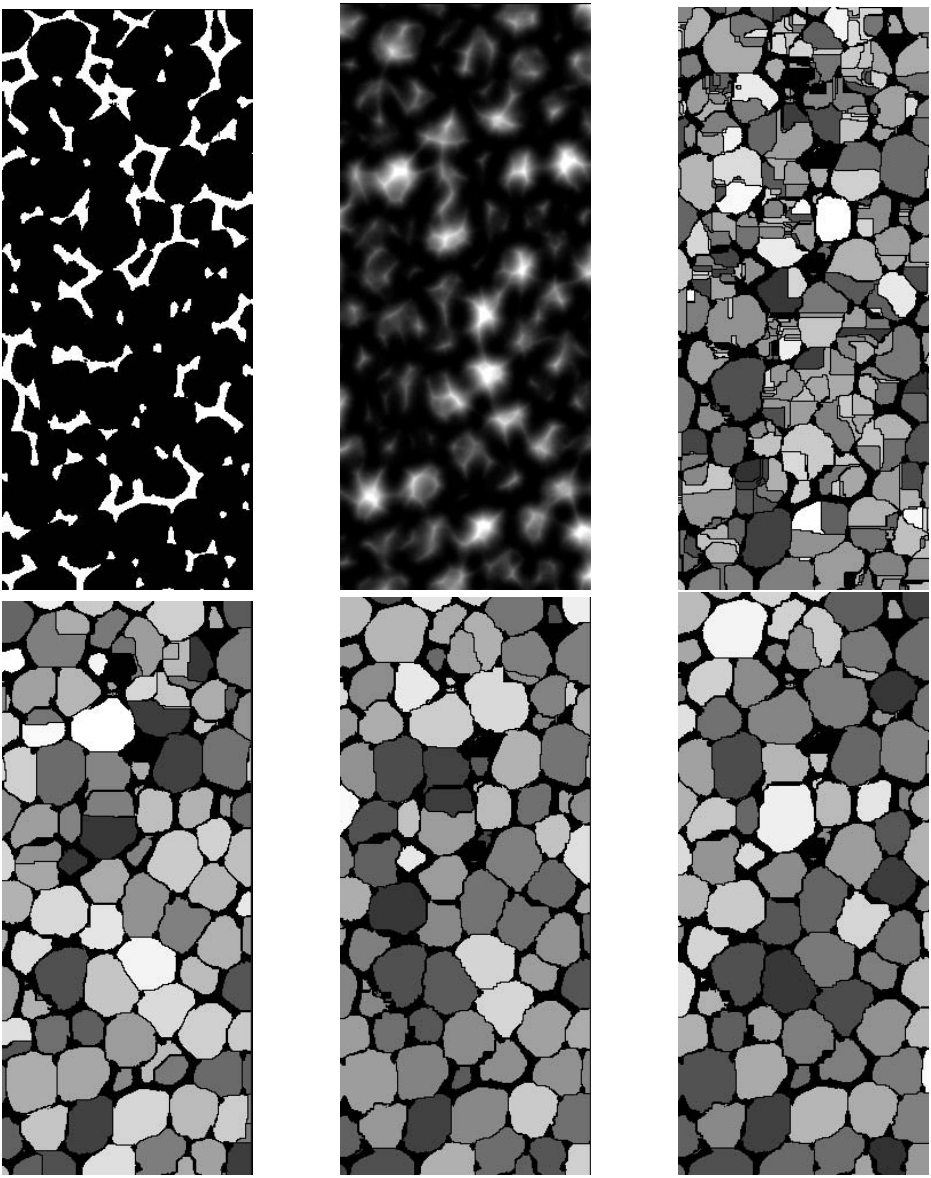


FIG. 3.13. Influence of different pre-processing on pore reconstruction. Top line: binary image of the strut system of the open-cell nickel foam of Fig. 3.10, EDT on pore space, reconstructed cell system without pre-processing. Bottom line: reconstructed cell system smoothed with a mean filter, h-minima, and pre-flooding with size constraint. Four of these image are reproduced in **colour** in the central part of the book.

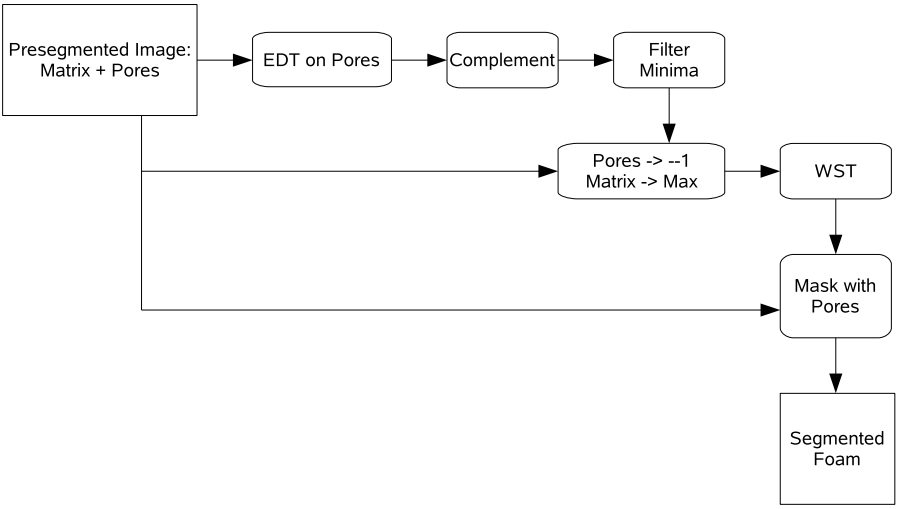


FIG. 3.14. Algorithm for the reconstruction of the cells of an open-cell foam. EDT – Euclidian distance transform, WST – watershed transform.

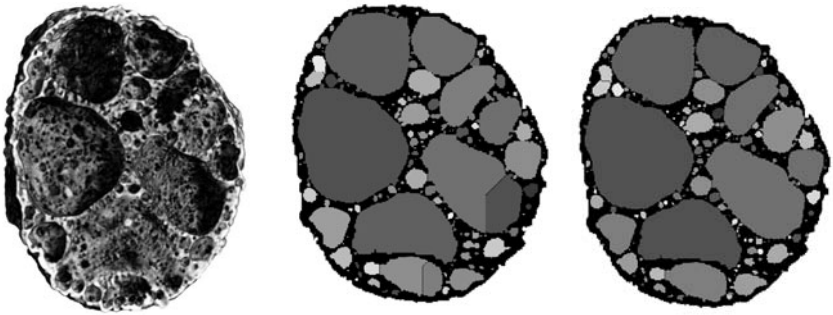
grey-value.) Finally, masking with the pore system yields the reconstructed cells. The complete cell-reconstruction procedure is sketched in Fig. 3.14. A volume rendering of a reconstructed cell system is shown in Fig. 3.8.

3.4.4 *Further segmentation methods*

There is a wide variety of segmentation methods, which cannot be covered here. A few are, nevertheless, mentioned briefly:

A widely used approach in image segmentation is to put up a 'cost' functional that is minimized by the optimal segmentation. The best-known variant of this approach is the one by Mumford and Shah (1989), with a cost function encouraging homogeneous grey-values within image segments and allowing discontinuities at the boundaries, only. Active contours (or snakes) are parametric curves or surfaces that are fit to the image by energy minimization. Both approaches can be formulated as level-set methods. For level-set segmentation, the image is embedded into a higher dimension by use of an auxiliary functional, with the original image as the zero level set. The functional is evolved according to a partial differential equation, the level-set equation. Finally, the segmentation is found as the zero level set of the steady-state solution (Sethian, 1999).

Bayesian image segmentation assigns a label by choosing the maximum *a-posteriori* probability estimate of a label given the pixel's value. Images are modelled by Markov random fields, where the full conditional distribution of the random field at a certain pixel depends on some neighbourhood of that pixel, only. The Hammersley–Clifford theorem ensures that Markov chain Monte Carlo methods such as Gibbs samplers or Metropolis–Hastings can be applied, see the



(a) (b) (c)

FIG. 3.15. Influence of different pre-processing procedures on pore reconstruction. (a) volume rendering of a ceramic grain; (b,c) slices through reconstructed pore system. (b) distance image smoothened using h-minima transform, (c) using height adaptive h-minima transform. Original image: X-ray CT, Bernhard Heneka, RJL Micro&Analytic, pixel size 7.1 µm, approximate grain diameter 2 mm. Sample: Frank Ripplinger, KeraGlas AG, Freiberg. Images (b) and (c) are reproduced in **colour** in the central part of the book.

tutorials by Hurn *et al.* (2003) or Winkler (1995) for mathematical background.

The 'itk' project (Yoo, 2004) provides implementations of many of the methods listed above. Wirjadi (2007) gives a more detailed overview of all segmentation approaches mentioned here.

3.5 Analysis of image data

In our understanding, image analysis is the computation of geometric characteristics of connected components or complete image segments. Connected components or objects, as defined in Section 3.1.3.4 are usually obtained by labelling of the foreground of a binary image (Section 3.4.2) or by the watershed transform (Section 3.4.3). As before, a connected component is modelled as a set $X \subset \mathbb{R}^3$. The characteristics of X are called *object features*. Complete segments are obtained by binarization, see Section 3.4.1. An image segment often represents a material constituent observed through a window W, also called the image *frame* or the *field*. Thus, the characteristics of the segment – the *field features* – can be interpreted as estimates of characteristics of the constituent, in particular if the material is macroscopically homogeneous. In this case the constituent is modelled as a macroscopically homogeneous random set – in the following denoted by Ξ.

The measurement of object features and field features is conceptionally the same. The differences arise only in the interpretation of the measured data and in the handling of edge effects. In the following section, the principle of measurement is explained in detail for the object features, while the measurement of field features is touched only briefly in Section 3.5.2.

Field features such as the surface density (i.e. the mean surface area per unit volume) can be considered as *first moments* of random variables that describe only the 'average or mean behaviour' of a material constituent. Fluctuations of the constituent are captured by *second moments* considered in Section 3.5.3.

3.5.1 *Features of connected components*

A very attractive set of geometric characteristics of connected components are the *intrinsic volumes* or quermass integrals or Minkowski functionals). In 3D, they are, up to constants, *volume*, *surface area*, *integral of mean curvature*, (*Germain's curvature*), and the *integral of total curvature* (*Gaussian curvature*). For example, in the case of fibrous structures, the integral of Germain's curvature yields the total fibre length without need to segment individual fibres. Furthermore, the integral of Gaussian curvature is, up to a constant, equivalent to the Euler number, and in the case of an assembly of simply connected objects it estimates the object number.

Surface area and integral of mean curvature can be measured using Crofton formulae that boil down their computation to computing Euler numbers in lower-dimensional intersections. Discretization of these formulae (see Section 3.5.1.2) combined with an efficient calculation of the Euler numbers in the intersections yield a fast algorithm for simultaneously determining all intrinsic volumes based on observations in a digital image. The backbone of the Euler-number calculation is thorough investigations of digital connectivity and consistency from Nagel *et al.* (2000) and Ohser *et al.* (2002, 2003), as summarized in Section 3.1.3.

The presented method's algorithmic core consists in a convolution of the binary image with a $2 \times 2 \times 2$ mask, resulting in an 8-bit grey-value image (Lang *et al.*, 2001). All further steps are based solely on the grey-value histogram whose size does not depend on image size or content. Thus, the advantage over other methods for computing the intrinsic volumes (Blasquez and Poiraudeau, 2003, Schmidt and Spodarev, 2005) are simplicity and speed of the algorithm. The surface area is measured directly from the binary volume image without need to approximate the surface. Similar methods are from Windreich *et al.* (2003) and Lindblad (2005), see Schladitz *et al.* (2006*a*) for a comparison.

3.5.1.1 *Section lattices and translative complements* We restrict ourselves to the three-dimensional cubic primitive lattice $\mathbb{L}^3 = a\mathbb{Z}^3$ with lattice distance $a > 0$. The Crofton formulae for computing the intrinsic volumes of a set $X \subset \mathbb{R}^3$ use the intersections of X with affine subspaces of $\mathbb{R}^3$. In order to obtain a digitized version, we introduce section lattices of $\mathbb{L}^3$ and their translative complements in analogy to linear subspaces and their orthogonal complements:

Definition A pair $\left(\mathbb{L}^k, {}^T\mathbb{L}^{3-k}\right)$, $k = 1, 2$, is called a *k-dimensional section lattice* $\mathbb{L}^k$ equipped with the *translative complement* ${}^T\mathbb{L}^{3-k}$, if there exists a basis v_1, v_2, v_3 of $\mathbb{L}^3$ with

(i) $\mathbb{L}^k = (v_1, v_k)\mathbb{Z}^k$,
(ii) ${}^T\mathbb{L}^{3-k} = (v_{k+1}, v_3)\mathbb{Z}^{3-k}$,

TABLE 3.1. The bases of the 13 section lattices $\mathbb{L}_i^k$ of $\mathbb{L}^3 = \mathbb{Z}^3$ and a possible translative complement ${}^T\mathbb{L}_i^{3-k}$ for $k = 1$ (left), and $k = 2$ (right).

i	basis of $\mathbb{L}_i^1$	basis of ${}^T\mathbb{L}_i^2$
1	$\{u_1\}$	$\{u_2, u_3\}$
2	$\{u_2\}$	$\{u_1, u_3\}$
3	$\{u_3\}$	$\{u_1, u_2\}$
4	$\{u_1 + u_2\}$	$\{u_1, u_3\}$
5	$\{-u_1 + u_2\}$	$\{u_1, u_3\}$
6	$\{u_1 + u_3\}$	$\{u_1, u_2\}$
7	$\{-u_1 + u_3\}$	$\{u_1, u_2\}$
8	$\{u_2 + u_3\}$	$\{u_1, u_3\}$
9	$\{-u_2 + u_3\}$	$\{u_1, u_3\}$
10	$\{u_1 + u_2 + u_3\}$	$\{u_1, u_2\}$
11	$\{-u_1 + u_2 + u_3\}$	$\{u_1, u_2\}$
12	$\{u_1 - u_2 + u_3\}$	$\{u_1, u_2\}$
13	$\{u_1 + u_2 - u_3\}$	$\{u_1, u_2\}$

i	basis of $\mathbb{L}_i^2$	basis of ${}^T\mathbb{L}_i^1$
1	$\{u_1, u_2\}$	$\{u_3\}$
2	$\{u_1, u_3\}$	$\{u_2\}$
3	$\{u_2, u_3\}$	$\{u_1\}$
4	$\{u_1, u_2 + u_3\}$	$\{u_3\}$
5	$\{u_1, -u_2 + u_3\}$	$\{u_3\}$
6	$\{u_2, u_1 + u_3\}$	$\{u_3\}$
7	$\{u_2, -u_1 + u_3\}$	$\{u_3\}$
8	$\{u_3, u_1 + u_2\}$	$\{u_1\}$
9	$\{u_3, -u_1 + u_2\}$	$\{u_1\}$
10	$\{u_1 + u_3, u_2 + u_3\}$	$\{u_3\}$
11	$\{-u_1 + u_3, u_2 + u_3\}$	$\{u_3\}$
12	$\{-u_1 + u_3, -u_2 + u_3\}$	$\{u_3\}$
13	$\{u_1 + u_3, -u_2 + u_3\}$	$\{u_3\}$

(iii) there is an $x \in \mathcal{F}^0(\check{C})$ with $\{v_1, v_k\} \subset \mathcal{F}^0(C + x)$.

Condition (iii) ensures that integration over 'local knowledge' on the image data is possible. The translative complement ${}^T\mathbb{L}^{3-k}$ has properties similar to those of the orthogonal complement of a linear subspace. In particular, the union of shifts $\mathbb{L}^k + x$ of the section lattice over all x from the corresponding translation lattice ${}^T\mathbb{L}^{3-k}$ yields the cubic lattice $\mathbb{L}^3$,

$$\mathbb{L}^3 = \bigcup_{x \in {}^T\mathbb{L}^{3-k}} \mathbb{L}^k + x.$$

However, the translative complement is not necessarily uniquely determined. Nevertheless, choosing one of the translative complements arbitrarily turns out to work for all considerations presented in the following. There are 13 section lattices $\mathbb{L}_i^k$ for both $k = 1$ and $k = 2$. This restriction is due to condition (iii).

3.5.1.2 *Intrinsic volumes* Consider a compact set $X \subset \mathbb{R}^3$ (i.e. an object or a particle). To fulfil certain regularity conditions we assume that X is a finite union of compact and convex sets. The intrinsic volumes V_j, $j = 0, \ldots, 3$ of X are – up to constant factors – the volume $V(X)$, the surface area $S(X) = 2V_2(X)$, the integral of mean curvature $M(X) = \pi V_1(X)$ and the Euler number $\chi(X) = V_0(X)$. The intrinsic volumes are well ordered according to their physical dimensions m^3, m^2, m^1, and m^0, respectively. The quantities V, S, and M are scaled, while χ is a topological characteristic.

By means of the *Crofton formulae*, the intrinsic volumes can be written as

$$V_{3-k}(X) = 2 \int_{\mathcal{L}^k} \underbrace{\int_{\perp L} \chi(X \cap (L + y))\, d\lambda_{\perp L}(y)}_{p_{3-k}(X)}\, d\mu(L), \tag{3.24}$$

$k = 1, 2$, where $\mathcal{L}^k$ is the set of all k-dimensional linear subspaces of $\mathbb{R}^3$, ${}^{\perp}L$ denotes the orthogonal complement of $L \in \mathcal{L}^k$, $\lambda_{\perp L}$ is the $3-k$-dimensional Lebesgue measure on ${}^{\perp}L$ and μ denotes the invariant measure on $\mathcal{L}^k$ with $\mu(\mathcal{L}^k) = 1$. The interpretation of eqn 3.24 is as follows:

$k = 1$. In this case we consider intersections $X \cap (L + y)$ of the object X with a straight line L shifted by y. The shift y belongs to the plane ${}^{\perp}L$ orthogonal to the straight line L. The intersection $X \cap (L + y)$ consists of an assembly of chords through X and the Euler number $\chi(X \cap (L + y))$ is simply the number of chords. Then the inner integral is over all translations y of the straight line L. The outer integral is over the set $\mathcal{L}^1$ consisting of all straight lines hitting the origin, i.e. the outer integral is over all directions.

$k = 2$. Now L is a plane hitting the origin and $X \cap (L + y)$ is a planar section. The translation y belongs to the straight line ${}^{\perp}L$ orthogonal to L and $\chi(X \cap (L+y))$ is the Euler number of the planar section. The inner integral is over all translations of the plane, while the outer integral is again over all directions that, however, are represented by the set $\mathcal{L}^2$ of all planes hitting the origin. Here, a space direction corresponds to the normal vector of the section plane L.

The observation of X on the lattice $\mathbb{L}^3$ implies that the integrand in the Crofton formulae eqn 3.24 is known for only a finite number of elements of $\mathcal{L}^k$, and the translation $L + y$ is possible for discrete values of y, only. That is, both integrals in eqn 3.24 are approximated by sums. Furthermore, the intersection $(X - y) \cap L$ must be replaced by its discretization $(X - y) \sqcap \mathbb{F}^k$ with respect to an adjacency system $\mathbb{F}^k$ in $\mathbb{L}^k$, where $L = \operatorname{span} \mathbb{L}^k$, and the translations y are from ${}^{T}\mathbb{L}^{3-k}$ instead of ${}^{\perp}L$, where ${}^{T}\mathbb{L}^{3-k}$ is a translative complement according to the definition on page 82.

3.5.1.3 *Discretization of the Crofton formulae* In order to present a discretization of eqn 3.24 we introduce congruence classes of local pixel configurations ξ_ℓ with respect to rigid motions and count $\mathbb{I}(\xi_\ell + x \subseteq X)$ instead of $\mathbb{I}(\xi_\ell + x \subseteq X)\mathbb{I}(\xi_\ell^c + x \subseteq X^c)$, comparable to the procedure given by Windreich *et al.* (2003) and Lindblad (2005).

To this end, choose a pair of complementary adjacency systems $(\mathbb{F}, \mathbb{F}_c)$ for the lattice $\mathbb{L}^3$. Furthermore, the section lattices $\mathbb{L}_i^k$ of $\mathbb{L}^3$ are equipped with pairs $(\mathbb{F}_i^k, \mathbb{F}_{v,i}^k)$ of complementary adjacency systems, $i = 1, \ldots, 13$, $k = 1, 2$. It is assumed that the section lattices $\mathbb{L}_i^k$ and the adjacency systems $\mathbb{F}_i^k$, $\mathbb{F}_{c,i}^k$ are chosen such that for each element F belonging to $\mathbb{F}_i^k$ or $\mathbb{F}_{c,i}^k$ there exists a translation $x \in \mathbb{L}^3$ such that $F + x \subset C$. (Note that not necessarily $C^k \subset C$.) Then it is sufficient to consider local configurations $\xi_\ell \subseteq \mathcal{F}^0(C)$, $\ell = 0, \ldots, 255$.

Now, we average the approximations of the intrinsic volumes related to rotations that leave the lattice $\mathbb{L}^3$ invariant, i.e. related to the symmetry group of the octahedron $\{\theta_1, \ldots, \theta_{48}\}$. Let $D_0, \ldots, D_{21}$ be the congruence classes of $\{\xi_0, \ldots, \xi_{255}\}$ with respect to this group and let $\{\eta_0, \ldots, \eta_{21}\}$ be a system of

TABLE 3.2. The coefficients g_{3j} of the 22 congruence classes of the local pixel configurations in 3D images. (The vertices not marked with a disc can be either foreground or background pixels.) The columns of g_{0j} correspond to $(\mathbb{F}_{26}, \mathbb{F}_6)$, $(\mathbb{F}_{14.1}, \mathbb{F}_{14.1})$, $(\mathbb{F}_{14.2}, \mathbb{F}_{14.2})$, and $(\mathbb{F}_6, \mathbb{F}_{26})$, in this order. $g_{1j} = g_{2j} = 0$ for the congruence classes $j = 11, \ldots, 21$.

j	η_j	g_{0j}	g_{1j}	g_{2j}	j	η_j	g_{0j}
0	ξ_0	0 0 0 0	0	0	11	ξ_{195}	0 0 0 −6
1	ξ_1	1 1 1 1	0.751	0.751	12	ξ_{105}	0 0 0 −2
2	ξ_3	−3 −3 −3 −3	−0.861	−0.275	13	ξ_{99}	0 0 −2 −24
3	ξ_9	0 −3 −3 −6	−1.076	−0.314	14	ξ_{31}	0 0 0 24
4	ξ_{129}	0 −1 −1 −4	−0.314	−0.163	15	ξ_{151}	0 0 0 8
5	ξ_{11}	0 6 6 12	0.549	0	16	ξ_{167}	0 0 0 24
6	ξ_{131}	0 6 4 24	0.628	0	17	ξ_{63}	0 0 0 −12
7	ξ_{41}	0 0 2 8	0.325	0	18	ξ_{159}	0 0 0 −12
8	ξ_{15}	3 0 0 −3	0	0	19	ξ_{231}	0 0 0 −4
9	ξ_{43}	0 0 −2 −8	0	0	20	ξ_{127}	0 0 0 8
10	ξ_{139}	0 −6 −2 −24	0	0	21	ξ_{255}	−1 0 0 −1

representatives, $\eta_\ell \in D_\ell$. Using the coefficients in Tab. 3.2, the approximations of the intrinsic volumes can be rewritten as

$$\tilde{V}_{3-k}(X) = a^{3-k} \sum_{j=0}^{21} \frac{1}{48} \sum_{i=1}^{48} \sum_{x \in \mathbb{L}^3} \mathbb{I}(\theta_i^{-1} \eta_j \subset X - x)\, g_{3-k,j}, \tag{3.25}$$

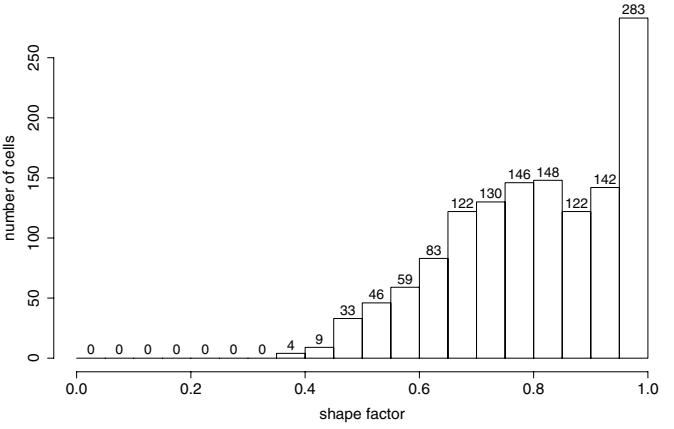
for $k = 1, 2, 3$. The coefficients g_{1j} and g_{2j} are computed for the section lattices listed in Tab. 3.1. The 2-adjacency is applied to one-dimensional section lattices, and the 6-adjacency is applied for two-dimensional section lattices. For further details see Lang *et al.* (2001) and Schladitz *et al.* (2006*a*).

3.5.1.4 *Shape factors* Analogously to the two-dimensional case, shape factors of three-dimensional objects X can be defined based on the isoperimetric inequalities. Since in three dimensions there are three scaled intrinsic volumes, three *shape factors* can be derived,

$$f_1 = 6\sqrt{\pi} \frac{V(X)}{\sqrt{S^3(X)}}, \qquad f_2 = 48\pi^2 \frac{V(X)}{M^3(X)}, \qquad f_3 = 4\pi \frac{S(X)}{M^2(X)},$$

which are normalized such that $f_1 = f_2 = f_3 = 1$, if X is a ball. Deviation from this value describes various aspects of deviation from ball shape. We have $0 \le f_1 \le 1$ and for convex objects X also $0 \le f_2, f_3 \le 1$. Figure 3.16 shows the histograms of shape factor f_1 for the large and small pores of the ceramic grain from Fig. 3.15.

(a)



(b)

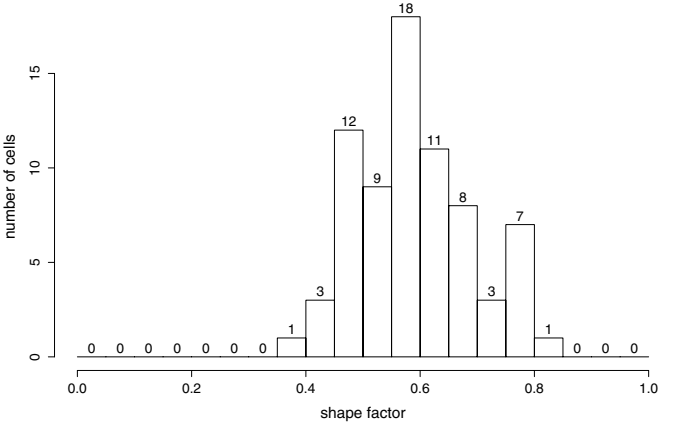


FIG. 3.16. Shape analysis for the pores of the ceramic grain from Fig. 3.15. Histograms of the shape factor f_1 for, (a) the small pores (diameter$\leq$ 0.2 mm) and, (b) the large pores (diameter$>$ 0.2 mm). Clearly, the large pores have a more irregular shape than the small ones.

3.5.2 *Field features*

Field features are usually obtained by integrating over the data of the whole image. Often, they are derived directly from the object features, e.g. the mean volume, the variance of the surface area, a histogram of a shape factor, *etc.* These statistical characteristics are usually computed for all objects contained in the image and, hence, they are related to the corresponding part of the materi-

als sample. In the case of a macroscopically homogeneous microstructure, these characteristics can describe several geometric aspects of a constituent such as porosity, contiguity, or mean grain size.

A problem of 'correct' estimation of statistical characteristics is the handling of edge effects resulting from objects intersected by the boundary of the window W (the edge of the image). The handling of this problem is well known from two-dimensional image analysis: A frame W' is defined inside of W such that all objects having their centre in W' do not hit the edge of W. Then the statistical characteristics obtained from the features of these objects are unbiased, i.e. exact up to fluctuations.

There is a further class of field features including the volume density V_V of a constituent, that is the total volume of this constituent in the window related to the window volume. Its estimation is obvious: The sum of the volumes over all objects of the constituent (including those intersecting the edge of the image) is divided by the volume of the window.

Obviously, the above estimator of V_V is equivalent to the number of pixels belonging to the constituent over the total pixel number. Moreover, counting pixels is much simpler since it does not require labelling.

3.5.2.1 *Densities of intrinsic volumes* The geometry of a materials constituent can be considered as a random subset Ξ of $\mathbb{R}^3$, where macroscopic homogeneity of the constituent is defined as invariance of the distribution of Ξ with respect to translations. In order to ensure the existence of the densities of the intrinsic volumes, certain regularity conditions are supposed: The realizations of Ξ are almost surely locally finite unions of compact and convex sets. For further details about random sets and their applications, see Stoyan *et al.* (1995) and Schneider and Weil (2000).

The *densities of the intrinsic volumes* are now defined as the limit

$$V_{V,k}(\Xi) = \lim_{r\to\infty} \frac{\mathbb{E}V_k(\Xi \cap rW)}{\operatorname{vol} rW}, \qquad k = 0, \ldots, 3,$$

where $\mathbb{E}$ denotes expectation. The limit is due to the fact that for finite r the value $V_k(\Xi \cup rW)$ is edge effected. If r is large enough these edge effects can be ignored. Analogously to the intrinsic volumes, their densities have the following meaning: $V_V = V_{V,3}(\Xi)$ is the *volume density*, $S_V = 2V_{V,2}(\Xi)$ is the *surface density* (or specific surface), $M_V = \pi V_{V,1}(\Xi)$ is the *density of the integral of the mean curvature*, and $\chi_V = V_{V,0}(\Xi)$ is the *density of the Euler number*.

In the above setting, an observation of the constituent Ξ in a cuboidal window W is a realization X of the random set $\Xi \cap W$, and $\Xi \cap W \cap \mathbb{L}^3$ are the foreground pixels of the corresponding binary image. From their definition it immediately follows that using eqn 3.25, the densities of the intrinsic volumes can be estimated from an observation $\Xi \cap W$ of Ξ by

$$\tilde{V}_{V,k}(\Xi) = \frac{\tilde{V}_k(\Xi \cap W)}{\operatorname{vol} W}, \qquad k = 0, 1, 2.$$

These estimators are usually biased (edge effected). However, if the summation in eqn 3.25 is restricted to those $x \in \mathbb{L}^3$ for which $\theta_i^{-1}\eta_j + x \subseteq W$, the estimations are free of edge effects.

(i) It is convenient to identify a constituent Ξ with the random function $\mathbb{I}_\Xi$. Obviously, for macroscopically homogeneous constituents the volume density V_V is independent of the position x. This means that V_V is the expectation of $\mathbb{I}_\Xi(x)$, and, equivalently, V_V is the probability that x belongs to Ξ,

$$V_V = \mathbb{E}\mathbb{I}_\Xi(x) = \mathbb{P}(x \in \Xi), \qquad x \in \mathbb{R}^3.$$

In the language of stochastics, the volume density is also called a *first-order characteristic* (first moment of a distribution). In this sense, the other densities S_V, M_V, and χ_V can be interpreted as first-order characteristics, too.

(ii) Consider the translative integrals p_1 and p_2 in eqn 3.24. The limits

$$\lim_{r\to\infty} \frac{\mathbb{E}p_k(\Xi \cap rW)}{\operatorname{vol} rW}, \qquad k = 1, 2 \tag{3.26}$$

carry directional information about Ξ. For $k = 2$, the inverse cosine transform yields the density of the directional distribution of the surface normals of Ξ. Furthermore, if Ξ is a random system of fibres, then for $k = 1$ the inverse cosine transform of the expression in eqn 3.26 yields the directional distribution of the fibre centres (Schladitz *et al.*, 2006*b*).

(iii) From the quermass densities further characteristics can be derived, e.g. the *mean chord length* $\bar{\ell}$,

$$\bar{\ell} = \frac{V_V}{4S_V},$$

also called the *mean intercept length*. A further characteristic first suggested by Hildebrand and Ruegsegger (1997) for evaluating bone structure is the *structure model index* (SMI) defined as

$$\mathrm{SMI} = 6\frac{S_V' V_V}{S_V^2},$$

where S_V' denotes the derivative of S_V given an infinitesimal dilation with a ball. According to Serra (1982) $S_V' = \frac{\pi}{2} M_V$ and the SMI thus equals

$$\mathrm{SMI} = 3\pi\frac{M_V V_V}{S_V^2}.$$

3.5.3 *Fluctuations*

Popular and powerful tools for describing fluctuation of microstructures are second-order characteristics, in particular the covariance function. These characteristics can be obtained in two ways: On the one hand, scattering (for instance

of X-rays) by the microstructure yields characteristics in frequency space (e.g. the scattering intensity) (Berke *et al.*, 1987). These characteristics are related to second-order characteristics in real space via the Fourier transform (Frank, 1980). On the other hand, second-order characteristics as well as their counterparts in frequency space can be measured from images of the microstructures obtained from μCT (Frank, 1980, Manolakis and Proakis, 1996). The fast Fourier transform allows us to determine these quantities fast and efficiently. The idea to measure the covariance and other second-order characteristics in real space via the Fourier transform has been around for some time (Debye *et al.*, 1957, Torquato, 2002), see also Ohser and Mücklich (2000, Chapter 5). A sound mathematical basis is provided by Koch *et al.* (2003).

It should be noted that the scattering intensity and similar quantities known from small-angle scattering are sometimes called the *power spectrum* or the *spectral density*. These terms are used in Fourier optics, image processing, and stochastic geometry, too. In these contexts, diffraction analysis is synonymous with spectral analysis.

It is well known that the spectral approach and thus the fast Fourier transform can be used to measure second-order characteristics faster than via the convolution of the characteristic function of a random set. More precisely, let cov be the covariance of a macroscopically homogeneous random set observed in a bounded observation window W. The two ways of measuring cov are outlined in the following diagram:

$$\begin{array}{ccc} (\mathbb{I}_\Xi - V_V)\mathbb{I}_W & \xrightarrow{\text{Fourier transform}} & \mathcal{F}((\mathbb{I}_\Xi - V_V)\mathbb{I}_W) \\ \Big\downarrow {\scriptstyle *^*,\,\mathbb{E}} & & \Big\downarrow {\scriptstyle |\ |^2,\,\mathbb{E}} \\ \operatorname{cov} c_W & \xleftarrow{\text{inverse Fourier transform}} & \mathbb{E}|\mathcal{F}((\mathbb{I}_\Xi - V_V)\mathbb{I}_W)|^2. \end{array}$$

Here, $*^*$ indicates the autocorrelation (convolution with reflection) and and c_W is the *window function* of W (the convolution of the characteristic function $\mathbb{I}_W$ with its reflection).

Several problems arise in the diffraction by image processing and are addressed in this section. First, it should be noted that the Fourier transform of a macroscopically homogeneous constituent does not exist, i.e. the function $\mathbb{I}_\Xi - V_V$ is not Fourie transformable in the sense of Section 3.3.1. Hence, the *power spectrum* of a macroscopically homogeneous constituent cannot be defined as the square of the absolute value of the Fourier transform of $\mathbb{I}_\Xi - V_V$. Furthermore, the quantity in the lower right corner of the above diagram is the power spectrum of the windowed function $(\mathbb{I}_\Xi - V_V)\mathbb{I}_W$. It depends on the size and the shape of W. In other words, the usual technique of image processing leads to edge effects in the measurement values of power spectra. Finally, the fast Fourier transform supposes periodicity (related to the window W). However, the microstructure is not periodic (in particular not W-periodic) but macroscopically homogeneous.

3.5.3.1 *Covariance function and Bartlett spectrum* The function $f(x) = \mathbb{I}_\Xi(x) - V_V$ is not transformable, but it corresponds to the density of a locally finite measure μ. Thus, the following extension of the Fourier transform to measures can be applied. A measure $\hat{\mu}$ shall be denoted as the Fourier transform of the measure μ, if for all smooth (infinitely differentiable) functions ψ of rapid decay the Parseval identity

$$\int_{\mathbb{R}^3} \psi(x)\, \hat{\mu}(dx) = \int_{\mathbb{R}^3} \hat{\psi}(\omega)\, \mu(d\omega) \tag{3.27}$$

holds where $\hat{\psi}$ is the Fourier transform of ψ, cf. the Parseval identity eqn 3.11 for functions. In this context Bochner's theorem is very useful. For any continuous and positive-definite function $f : \mathbb{R}^3 \to \mathbb{C}$ there exists a unique positive and finite measure μ with

$$f(x) = \frac{1}{(2\pi)^{3/2}} \int_{\mathbb{R}^3} \mathrm{e}^{i\xi x}\, \mu(d\xi), \quad x \in \mathbb{R}^3. \tag{3.28}$$

The *covariance function* cov of a constituent Ξ with volume density V_V can be defined as the expectation of a product of $\mathbb{I}_\Xi - V_V$,

$$\mathrm{cov}(x - y) = \mathbb{E}\,(\mathbb{I}_\Xi(x) - V_V)\,(\mathbb{I}_\Xi(y) - V_V)$$

for any two positions $x, y \in \mathbb{R}^3$. Because of the macroscopic homogeneity of Ξ the function cov depends on the distance $x - y$ of the positions x and y only.

In general, the covariance function cov of Ξ is not transformable as a function. Hence, its Fourier transform is not defined. This leads to certain problems in the definition of a corresponding quantity in frequency space. To overcome these problems we consider the *covariance measure* Cov of the constituent Ξ defined by

$$\mathrm{Cov}(W) = \int_W \mathrm{cov}(x)\, dx$$

for any set $W \subset \mathbb{R}^3$. Now, since the covariance function is positive-definite and continuous, from Bochner's theorem it follows that there exists a finite measure Γ with

$$\mathrm{cov}(x) = \frac{1}{(2\pi)^{3/2}} \int_{\mathbb{R}^3} \mathrm{e}^{i\xi x}\, \Gamma(d\xi) \qquad x \in \mathbb{R}^3. \tag{3.29}$$

The measure Γ is called the *Bartlett spectrum* of the constituent Ξ.

Note that the covariance function cov as well as the Bartlett spectrum Γ are independent of the sample size and shape (or the size and the shape of the region of interest W). Both cov and Γ characterize exclusively aspects of the fluctuation of the constituent Ξ, and they carry equivalent information.

3.5.3.2 *Power spectrum* By the theorem of Wiener–Khintchine (Manolakis and Proakis, 1996, p. 299) the power spectrum of a signal coincides with the Fourier transform of its autocorrelation function. This is essentially a special case of the well-known convolution theorem from Fourier analysis. A similar result can be established for constituents of macroscopically homogeneous microstructures and their covariance functions.

To ensure that the power spectrum is well defined, the Fourier transform of the signal has to be an integrable function. However, macroscopic homogeneity implies unboundedness of Ξ, and due to the unboundedness of Ξ the signals $\mathbb{I}_\Xi$ as well as $\mathbb{I}_\Xi - V_V$ are almost surely not transformable in this sense. Of course, μCT images represent a (very small) part of the microstructure, i.e. the observed constituent Ξ is restricted to a bounded window W. The random function

$$g(x) = \mathbb{I}_W(x)(\mathbb{I}_\Xi(x) - V_V), \qquad x \in \mathbb{R}^3,$$

is almost surely integrable, thus its Fourier transform $\hat{g}$ is well defined and the *power spectrum* of Ξ related to the window W is given by the expectation

$$\mathrm{pow}_W(\xi) = \mathbb{E}\,|\hat{g}(\xi)|^2, \qquad \xi \in \mathbb{R}^3.$$

This means that the square of the absolute value of $\hat{g}$ is a measurement value of pow_W. Clearly, the function g and, hence, the power spectrum pow depend on the size and shape of the window W. We introduce a *window function* defined by the convolution of $\mathbb{I}_W$ with its reflected counterpart,

$$c_W(x) = \int_{\mathbb{R}^3} \mathbb{I}_W(y)\mathbb{I}_W(y-x)\,dy, \qquad x \in \mathbb{R}^3.$$

As shown by Koch *et al.* (2003), the following analogue to the Wiener–Khintchine theorem holds for all $\xi \in \mathbb{R}^3$

$$(2\pi)^{3/2}\mathrm{pow}_W(\xi) = \mathcal{F}(c_W \mathrm{cov})(\xi). \tag{3.30}$$

Since the window function has bounded support, eqn 3.29 yields

$$(2\pi)^3\,\mathrm{pow}_W(\xi) = \big(\widehat{c}_W * \Gamma\big)(\xi), \quad \xi \in \mathbb{R}^3. \tag{3.31}$$

This relationship links the quantity that can be measured – the power spectrum – to the quantity we are interested in – the Bartlett spectrum.

3.5.3.3 *Measurement* A straightforward approach to the measurement of the covariance function cov of the constituent Ξ observed in a bounded window W is obtained by the convolution equation

$$\mathrm{cov}(x) = \frac{1}{c_W(x)}\,\mathbb{E}\int_{\mathbb{R}^3} g(y)g(x-y)\,dy \tag{3.32}$$

for all $x \in \mathbb{R}^3$ with $c_W(x) > 0$, see Ohser and Mücklich (2000, p. 150).

Using eqn 3.32 and image-analytic methods, the covariance function can be computed from the image data with a complexity in $\mathcal{O}(m^2)$, where m is the number of pixels, i.e. the number of lattice points in $\mathbb{L}^3 \cap W$.

A faster algorithm can be derived using eqn 3.30. Since the function $\mathbb{E}|\hat{g}|^2$ is integrable, the inverse Fourier transform can be applied. This yields

$$(2\pi)^{3/2}\bar{\mathcal{F}}\left(\mathbb{E}|\hat{g}|^2\right) = c_W \mathrm{cov}.$$

Assume now that the origin belongs to the interior of W. Then c_W is positive for all x in the interior and it follows that

$$\mathrm{cov}(x) = \mathbb{E}\ \frac{(2\pi)^{3/2}\left(\bar{\mathcal{F}}|\hat{g}|^2\right)(x)}{c_W(x)}. \tag{3.33}$$

A discretization of this equation gives a measurement of the covariance function. For an image of Ξ with M pixels the covariance can be computed using the fast Fourier transform with a complexity in $\mathcal{O}(M \log M)$. This holds because the FFT has a complexity in $\mathcal{O}(M \log M)$ and the window function can be computed as $c_W = \bar{\mathcal{F}}\,|\mathcal{F}\mathbb{I}_W|^2$.

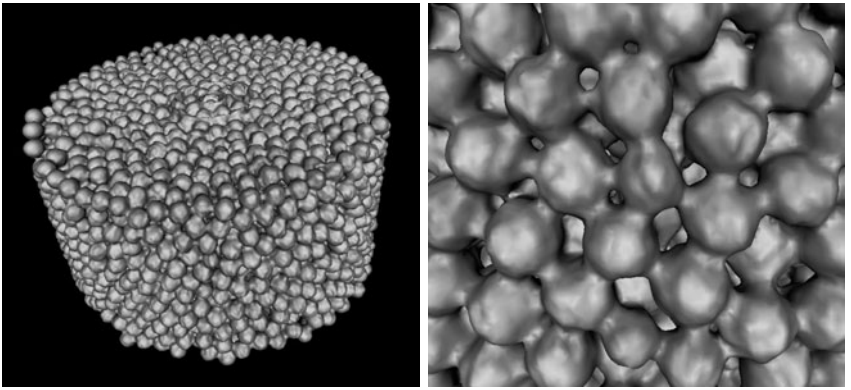
Unfortunately, the assumption of periodicity within the FFT causes an overlapping effect, see Koch *et al.* (2003). This effect can be eliminated by padding with zeros, which increases the number of sample points to $8M$. The complexity still belongs to $\mathcal{O}(M \log M)$, which is a considerable gain compared to the usual estimation of the covariance with complexity $\mathcal{O}(M^2)$.

Further simplifications arise if the microstructure is additionally isotropic, as the Bartlett spectrum then depends only on the radial co-ordinate, see Koch *et al.* (2003).

3.5.3.4 *Mean interparticle distance during sintering* Usually, models for sintering processes are based on observations of sinter particles arranged in a plane (two-particle models (Exner, 1978, Schatt, 1992)). Newer results from sinter experiments with three-dimensional particle systems are obtained from the analysis of CT scans made during the sintering processes (Ohser *et al.*, 2005). In these experiments, spherical copper particles are filled into alumina crucibles and fixed in their positions with a diluted alcoholic solution of polyvinylpyrrolidone. The samples are scanned in the initial stages (not sintered) and at discrete time steps of the sintering process. At these time steps the sintering process was interrupted and the samples were cooled down in H_2 to room temperature. After scanning, the samples were heated to the sintering temperature (heating rate 5 K/s) again.

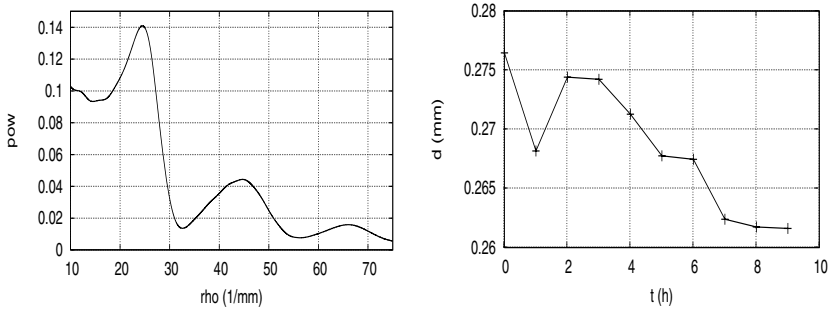
In the following we consider a sample containing about 3750 copper particles of diameters between 200 and 315 µm and sintered at temperatures between 873 and 1223 K. The sintering time between the scans was 1 h. The sintering process was recorded by a sequence of 10 images (including a scan of the initial stage).

Figure 3.17 shows visualizations of a CT scan after a sintering time of 9 h. The corresponding density of the Bartlett spectrum is shown in Fig. 3.18(a). The position of the first maximum corresponds to the *typical* distance d of the particle



(a) (b)

FIG. 3.17. Three-dimensional visualizations of a CT scan of sintered copper spheres after the last sintering period. (a) visualization of the whole sample. (b) a magnified view showing the interparticle necks. Original image (X-ray CT, pixel size 22 µm) and sample preparation: Michael Nöthe, TU Dresden. Reprinted with permission by Carl Hanser Verlag from Ohser *et al.* (2005).



(a) (b)

FIG. 3.18. (a) Density of the Bartlett spectrum measured from the CT scan shown in Fig. 3.17(b), (b) the typical distance d of the centres of the sinter particles depending on the time t.

centres in the sinter material. More precisely, consider an (ordered) quadruple of four neighbouring copper balls randomly chosen from the sample, then the distance is defined as the orthogonal distance between the plane through the centres of the first three balls and the centre of the fourth ball. The 'typical distance' d is the median of the corresponding random variable.

The change of d over the sintering time t is shown in Fig. 3.18(b). During the initial sintering step d decreases because first sinter necks are formed. Then

the distance increases, which is caused by particle rotations. Finally, after rearrangement of the particles, the sinter necks start to grow and the typical distance decreases again.

3.6 Modelling materials properties

The macroscopic properties of solid microscopically heterogeneous materials such as metals, ceramics, composite materials, porous or granular media, *etc.* depend on local properties as well as on the geometry of their constituents. Thus, the macroscopic behaviour of a material can be predicted using local knowledge. Vice versa, local behaviour induced by outer constraints can be investigated. Examples are the computation of flow through porous media and a quantitative description of diffusion processes or thermally induced residual stresses in composite materials.

The geometric data of materials constituents necessary for simulating material's properties can be obtained by various experimental techniques. Imaging by μCT is one of these, providing full information about the three-dimensional geometry in a mesocale (i.e. down to about 1 µm), which is most important in many applications. Now, simulation of material's properties is based whether directly on CT data or on geometric models, where in the latter case the model parameters are fitted to the features measured by CT.

There are various techniques for modelling materials properties. Hornung (1997), Cherkaev and Kohn (1997) and Jeulin (2001) use *homogenization techniques* to compute macroscopic properties from local properties under certain assumptions on the microstructure geometry. Homogenization usually includes 'averaging' over a 'representative volume element' of the structure. In some cases, a geometric model is used for the structure. For example, Schladitz *et al.* (2006*b*) compute the acoustic properties of stacked non-woven fibres and describe the material's geometry as a macroscopically homogeneous random system of straight cylinders (tubes) and numerically simulating the flow determining the acoustic behaviour.

The 'natural' choice for modelling materials properties in reconstructed CT images are *lattice Boltzmann techniques*, since the data are given on homogeneous lattices. Lattice Boltzmann is designed to solve fluid-dynamics problems, see e.g. Succi (2001) for mathematical foundations and applications. Kehrwald (2004) applies lattice Boltzmann schemes for simulating flow through the pore space of porous or granular media, without numerical differentiation for computing the gradient of the velocity field. This is a clear advantage over other methods and carries over to pseudo-elastic fluids.

The probably most popular and widely applied simulation method is the *finite element method* (FEM) that is based on a subdivision of the continuum into cells (meshes, elements). At first glance, this subdivision seems to hamper the application to data given on lattices. However, the surface meshing of the foreground pixels of binary images described in Section 3.2.2 can be considered

a first step in volume meshing. In fact, surface-rendering data can be used as input to most FEM software.

There is rich literature on FEM, its mathematical background, and applications, see the excellent monographs by Bathe (1996) and Zienkiewicz *et al.* (2005) and the references therein. In order to give a short introduction to FEM, we consider the solid mechanics problem of linear elasticity that historically was one of the first that stimulated the impressive development of FEM. Moreover, solid mechanics is still of greatest importance in engineering. For simplicity, we restrict ourselves to materials behaving isotropically. In order to give a glimpse into the numerics behind FEM, the variational form of the principle of virtual work is derived from the equilibrium equations.

3.6.1 *Linear elasticity*

Denote by $X \subset \mathbb{R}^3$ the domain later to be tessellated into cells. The set X models a body, a part, or a constituent. To ensure that the outer normals exist at almost all surface elements, we assume that the surface ∂X of X is sufficiently smooth.

Let ε be a *strain field* given on $X \subset \mathbb{R}^3$. The strain ε is expressed in terms of a *displacement field* $u \in \mathbb{R}^3$ by the linear equation

$$\varepsilon = Su, \tag{3.34}$$

with the corresponding differential operator S. Here, ε, S and u are given by

$$\varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \gamma_{12} \\ \gamma_{23} \\ \gamma_{31} \end{pmatrix}, \qquad S = \begin{pmatrix} \frac{\partial}{\partial x_1} & 0 & 0 \\ 0 & \frac{\partial}{\partial x_2} & 0 \\ 0 & 0 & \frac{\partial}{\partial x_3} \\ \frac{\partial}{\partial x_2} & \frac{\partial}{\partial x_1} & 0 \\ \frac{\partial}{\partial x_3} & 0 & \frac{\partial}{\partial x_1} \\ 0 & \frac{\partial}{\partial x_3} & \frac{\partial}{\partial x_2} \end{pmatrix}, \qquad u = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix},$$

where ε_i and γ_{ij} are the *normal strains* and the *engineering shear strains*, respectively. Analogously, using the stress field σ and the body forces w, the equilibrium equation is written in the form

$$S'\sigma = -w, \tag{3.35}$$

with

$$\sigma = \begin{pmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \\ \tau_{12} \\ \tau_{23} \\ \tau_{31} \end{pmatrix}, \qquad w = \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix},$$

where S' denotes the transpose of S, σ_i are the *normal stresses* and τ_{ij} are *shear stresses*. Stress and strain are related to each other by the *constitutive relation*

$$\sigma - \sigma_0 = D(\varepsilon - \varepsilon_0), \tag{3.36}$$

where σ_0 and ε_0 are the *residual stresses* and *strains*, respectively. In the case of macroscopically isotropic behaviour of the material, the local elastic properties can be described by two parameters, e.g. *Young's modulus* E and *Poisson's ratio* ν. In this setting, the elasticity matrix D is given by

$$D^{-1} = \frac{1}{E}\begin{pmatrix} 1 & -\nu & -\nu & 0 & 0 & 0 \\ -\nu & 1 & -\nu & 0 & 0 & 0 \\ -\nu & -\nu & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2(1+\nu) & 0 & 0 \\ 0 & 0 & 0 & 0 & 2(1+\nu) & 0 \\ 0 & 0 & 0 & 0 & 0 & 2(1+\nu) \end{pmatrix}.$$

Finally, the description of the linear elastic problem is completed by formulating the *boundary* or *transmission conditions*

$$u = u_0, \quad t = t_0 \tag{3.37}$$

specified on the surface ∂X of X, where u_0 and t_0 are known displacement and traction, respectively. In our setting, the displacement u_0 is zero for all surface elements with $t_0 \neq 0$ and vice versa, if $u_0 \neq 0$ then $t_0 = 0$.

The traction $t = (t_1, t_2, t_3)$ may be related to the stress σ by the linear equation $t = A\sigma$, where the matrix A depends on the outer normal unit vector (n_1, n_2, n_3) of X at the current surface element

$$A = \begin{pmatrix} n_1 & 0 & 0 & n_2 & n_3 & 0 \\ 0 & n_2 & 0 & n_1 & 0 & n_3 \\ 0 & 0 & n_3 & 0 & n_1 & n_2 \end{pmatrix}.$$

Instead of the constitutive eqn 3.36 and the boundary conditions eqn 3.37, it is convenient to consider the corresponding *variational form*. In terms of the *virtual strain* $\tilde{\varepsilon}$ (*strain rate*) and the *virtual displacement* $\tilde{u}$ (*virtual velocity*), the *work equation* can be written as

$$\int_X \tilde{\varepsilon}\sigma \, dx - \int_X \tilde{u}w \, dx - \int_{\partial X} \tilde{u}t_0 \, ds = 0 \tag{3.38}$$

(*principle of virtual work*), and the constraint

$$\tilde{u} = u_0 \tag{3.39}$$

on the boundary ∂X. Here, the integration $\int ds$ is related to the surface measure. The virtual strain is arbitrary as long as it is subject to $\tilde{\varepsilon} = S\tilde{u}$, cf. eqn 3.38.

3.6.2 *Finite element method*

FEM is based on a discretization of the domain X implying a discretization of the variational eqn 3.38.

3.6.2.1 *Discretization* The domain X is now approximated by a polyhedral set P that is tessellated into convex polyhedral subsets $P_1, \ldots, P_m$ with non-empty interior. More precisely, the set $\{P_1, \ldots, P_m\}$ have to form a face-to-face tessellation of P, see Schneider and Weil (2000, p. 235). In FEM, the P_i are called the *element domains* (three-dimensional unstructured meshes). Of course, the accuracy of FEM depends on the total number of nodes, but the design of FEM and convergence do not depend on the number of vertices of the individual P_i. Recent techniques are usually based on tetrahedral meshes, see e.g. Weatherill and Hassan (1994).

A three-dimensional mesh can be obtained in the following way:

(i) First, obtain a surface meshing with respect to a given adjacency system $\mathbb{F}$ on the lattice $\mathbb{L}^3$ from the data $X \cap \mathbb{L}^3$ as described in Section 3.2.2. In other words, the mesh data from a surface rendering are 'natural' input to FEM. The meshes can be seen as the faces of the approximation P of X.

(ii) Now, a convex polygonal window W_P is defined containing all vertices of P. Usually, W_P is a cuboid with edges parallel to the co-ordinate axes. Then, the Delaunay tetragonalization formed by the vertices of P and W_P is derived (Watson, 1981). The underlying tetragonalization process is sequential. It starts with the data structure containing a tetragonalization of the cuboid W_P. Then the vertices of P are introduced step-by-step into this structure, in each step forming the local Delaunay tetragonalization, i.e. for the currently included vertex and all neighbouring points just belonging to the structure. This procedure is continued until all vertices of P are included.

(iii) To obtain a further refinement of the tetragonalization, additional points (belonging to the interior of P) are inserted. The point positions are chosen in order to improve the so-called element-size distribution function where the criterion for an improvement depends on the surface meshes, the elastic properties of the material, and the boundary conditions. A clear formulation of the criterion allows an automatic generation of additional points, see Weatherill and Hassan (1994).

(iv) Finally, all tetrahedra outside of P are removed from the structure, while the surface meshes deleted during the previous refinement process are recovered. This means, the tessellation $\{P_1, \ldots, P_m\}$ consists of all non-empty intersections of the tetrahedra obtained from steps (ii) and (iii) with P. It should be noted that an intersection is not necessarily a tetrahedron.

An example of a meshing is shown in Fig. 3.19. Assume that P has the faces F_j (surface meshes). Then, the work eqn 3.38 takes the form

$$\sum_{i=1}^{m} \int_{P_i} \tilde{\varepsilon}\sigma \, dx - \sum_{i=1}^{m} \int_{P_i} \tilde{u} w \, dx - \sum_{j} \int_{F_j} \tilde{u} t_0 \, ds = 0, \tag{3.40}$$

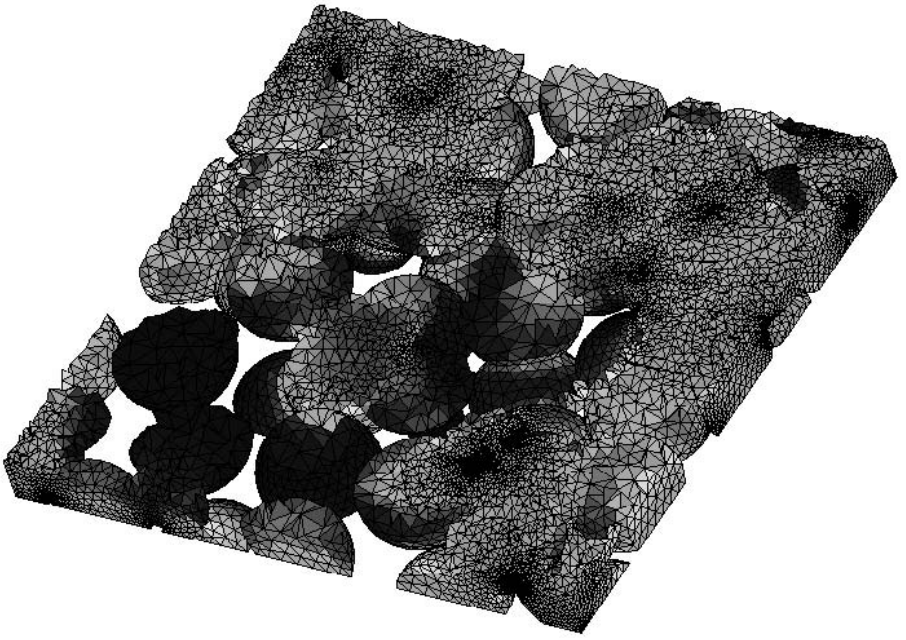


FIG. 3.19. The meshing of a simulated alumina sample (Young's modulus for the solid material is 300 GPa) and the *von Mises equivalent stress* under a shear load. The stresses range from 0 MPa (blue) to 400 MPa (red). The specimen is a sample from a random packing of balls after a uniform growth process used to model intermediate steps during sintering. The parameters of the random packing have been adapted to the alumina sample in such a way that the model represents those aspects of the geometry of the microstructure that are important for the current simulation. Visualized is a thick slice from a cube of side length 1 mm. This image is reproduced in **colour** in the central part of the book.

and the constraint eqn 3.39 is defined on the faces of F_j. The integration $\int ds$ is now related to the Lebesgue measure on F_j and $\sum_j \int_{F_j} ds$ can be rewritten as $\sum_{i=1}^{m} \int_{\partial P_i} ds$ with $t_0 = 0$ for points of ∂P_i in the interior of P.

3.6.2.2 *Numerical solution of the linear elastic problem* Following the approach of Zienkiewicz *et al.* (2005, Chapter 8), we first compute the corresponding strain ε and stress σ from eqn 3.34 and eqn 3.36, respectively. Then, the integration in eqn 3.40 is over each element domain P_i using a suitable quadrature rule (Gauss–Chebyshev quadrature). This leads to a linear equation for a solution of the displacement u, where the coefficient matrix is called the *global stiffness*

array and the right-hand side is known as the *global load array.*

The solution process sketched above includes various numerical aspects. In particular, all of the information is available only on the cells' vertices x_k (nodes) of the tessellation $\{P_1, \ldots, P_m\}$. For example, the continuous displacement u is approximated by corresponding discrete data $\hat{u}_k$ given on the nodes, $u \approx \sum_k f_k \hat{u}_k$, where the coefficients $f_k : \mathbb{R}^3 \mapsto \mathbb{R}$ depend on the shapes of the finite domains and, hence, are called *shape functions.* This means that the shape functions interpolate in some way between the vertices of the cell domains. Analogously, using eqn 3.34 the strains are approximated by $\varepsilon \approx \sum_k S f_k \hat{u}_k$. Now, inserting into eqn 3.40 we get

$$\sum_{i=1}^{m} \hat{\hat{u}}_k \left(\int_{P_i} (Sf_k)' \big(\sigma_0 + D(Sf_\ell \hat{u}_\ell - \varepsilon_0)\big)\, dx - \int_{P_i} f_k w\, dx - \int_{\partial P_i} f_k t_0\, ds \right) = 0.$$

Assuming arbitrary virtual displacements $\hat{\hat{u}}_k$ of the nodes, the last identity leads to a set of linear equations

$$A_{k\ell} \hat{u}_\ell = b_k. \tag{3.41}$$

The global stiffness array and the global load array are given by

$$\begin{aligned} A_{k\ell} &= \sum_{i=1}^{m} \int_{P_i} (Sf_k)' D S f_\ell\, dx, \\ b_k &= \sum_{i=1}^{m} \left(\int_{P_i} \big((Sf_k)'(D\varepsilon_0 - \sigma_0) + f_k w\big)\, dx + \int_{\partial P_i} f_k t_0\, ds \right), \end{aligned}$$

respectively, and eqn 3.41 is solved under the condition eqn 3.38.

As an example, we consider an alumina specimen represented by the data shown in Fig. 3.19. Using eqn 3.38, the displacement field and finally the corresponding von Mises equivalent stress was computed under a given shear load resulting from a displacement of 0.1 % of the edge length of the window (in total 1 µm) where in the case of alumina the isotropic local eleasticty may be characterized by a uniform Young's moduls $E = 300\,\mathrm{GPa}$ and a uniform Poisson ratio of $\nu = 0.22$. Alternative approaches as well as convergence and stability are addressed e.g. by Szabo and Babuška (1991) and Hughes (2000).

Note that the variational equations for linear elastic problems are similar to those for non-linear elastic problems. Hence, methods for their solution are conceptionally the same. Moreover, variational principles serve as a general framework for FEM applied to various other problems such as field problems (heat conduction, diffusion, fluid flow) or fluid dynamics.

Acknowledgements

This survey has only been made possible with the help of many colleagues from Fraunhofer ITWM. We thank in particular Heiko Andrä for providing Fig. 3.19, Tetyana Sych for creating most of the visualizations, Oliver Wirjadi and Michael

Godehardt for essential input to Sections 3.4 and 3.3.3, respectively. Lutz Zybell is acknowledged for input to Section 3.6. K.S. was supported by the Rheinland-Pfalz cluster of excellence 'Dependable Adaptive Systems and Mathematical Modeling' (`http://www.dasmod.de`).

3.7 References

Aguilera, A., Rodríguez, J., and Ayala, D. (2002). Fast connected component labeling algorithm: A non voxel-based approach. Technical Report LSI-02-26-R, Universitat Politècnica, Catalunya.

Akenine-Moeller, T. and Haines, E. (2002). *Real-time rendering* (2nd edn). AK Peters, Wellesley, MA.

Bathe, K.-J. (1996). *Finite element procedures in engineering analysis*. Prentice Hall, London.

Benouali, A.-H., Froyen, L., Delerue, J.-F., and Wevers, M. (2002). Mechanical analysis and microstructural characterization of metal foams. *Materials Science and Technology*, **18**, 489.

Berke, A., Neite, G., Riehmann, W., and Nembach, E. (1987). Characterization of periodic composites by laser-beam diffraction. *Journal of Applied Physics*, **61**, 1263.

Bertrand, G. (1996). A Boolean characterization of three-dimensional simple points. *Pattern Recognition Letters*, **17**, 115.

Blasquez, I. and Poiraudeau, J.-F. (2003). Efficient processing of Minkowski functionals on a 3D binary image using binary decision diagrams. In *11th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision '2003* (ed. V. Skala), Volume 11, Plzen, Czech Republic. Union Agency-Science Press.

Borgefors, G. and Nyström, I. (1997). Efficient shape representation by minimizing the set of maximal discs/spheres. *Pattern Recognition Letters*, **18**, 465.

Borgefors, G., Nyström, I., and Sanniti di Baja, G. (1997). Connected components in 3D neighbourhoods. In *Proceedings of the 10th Scandinavian Conference on Image Analysis* (ed. M. Frydrych, J. Parkkinen, and A. Visa), Lappeenranta, pp. 567. Pattern Recognition Society of Finland.

Canny, J. (1986). A computational approach to edge detection. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Volume 6, Piscataway, pp. 679. Institute of Electrical and Electronics Engineers. PAMI-8.

Cherkaev, A. and Kohn, R. (1997). *Topics in the mathematical modelling of composite materials*. Birkhäuser, Berlin.

Couprie, M. and Zrour, R. (2005). Discrete bisector function and Euclidean skeleton. In *Discrete geometry for computer imagery* (ed. E. Andres, G. Damiand, and P. Lienhardt), Volume 3429 of *Lecture Notes in Computer Science*, Berlin, pp. 216. Springer. Proceedings of the 12th International Conference on Discrete Geometry for Computer Imagery, Poitiers, France.

Cuisenaire, O. (1999). *Distance transformations: fast algorithms and applications to medical image processing*. Ph. D. thesis, Université Catholique de Louvain, Louvain.

Danielsson, P.E. (1980). Euclidean distance mapping. *Computer Vision, Graphics, and Image Processing*, **14**, 227.

Debye, P., Anderson, H.R. Jr., and Brumberger, H. (1957). Scattering by inhomogeneous solid. II. The correlation function and its application. *Journal of Applied Physics*, **28**, 679.

Di Stefano, L. and Bulgarelli, A. (1999). A simple and efficient connected components labeling algorithm. In *10th International Conference on Image Analysis and Processing (ICIAP'99)*, pp. 322. IEEE Computer Society Press.

Exner, H.E. (1978). *Grundlagen von Sintervorgängen*. Gebrüder Bornträger, Stuttgart.

Frank, J. (1980). The role of correlation techniques in computer image processing. In *Computer processing of electron microscope images* (ed. P. Hawkes), Berlin, pp. 187. Springer.

Frigo, M. and Johnson, S.G. (1998*a*). FFTW 2.1.3 (the fastest Fourier transform of the West). `http://www.fftw.org/`. [Online; accessed 12 June 2007].

Frigo, M. and Johnson, S.G. (1998*b*). FFTW: An adaptive software architecture for the fft. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, Volume 3, pp. 1381. IEEE Press.

Godehardt, M. and Schladitz, K. (2006). Geometric characterisation of light weight composites using computer tomographic images. In *9th European Conference on NDT*, Berlin, pp. We.1.6.3. Deutsche Gesellschaft für zerstörungsfreie Prüfung. Proceedings of the 9th ECNDT, Berlin.

Gueziec, A. and Hummel, R. (1994). The wrapper algorithm: surface extraction and simplification. In *IEEE Workshop on Biomedical Image Analysis, Seattle*, Piscataway, pp. 2043. Institute of Electrical and Electronics Engineers.

Hege, H.C., Hoellerer, T., and Stalling, D. (1993). Volume rendering mathematical models and algorithmic aspects. Technical Report TR 93-7, Konrad-Zuse-Zentrum für Informationstechnik Berlin (ZIB), Berlin.

Hildebrand, T. and Ruegsegger, P. (1997). Quantification of bone microarchitecture with the structure model index. *Computer Methods in Biomechanics and Biomedical Engineering*, **1**, 15.

Hoppe, H., DeRose, T., Duchamp, T., McDonald, J., and Stuetzle, W. (1993). Mesh optimization. In *Computer graphics* (ed. T. Kajiya), Volume 27, Providence, pp. 19. ACM Press. SIGGRAPH '93 Proceedings.

Hornung, U. (1997). *Homogenization and porous media*. Springer, New-York.

Hughes, T. J. R. (2000). *Finite element method: Linear static and dynamic analysis*. Dover Publications, New York.

Hurn, M.A., Husby, O.K., and Rue, H. (2003). A tutorial on image analysis. In *Spatial statistics and computational methods* (ed. J. Møller), Volume 173

of *Lecture Notes in Statistics*, Chapter 3, pp. 87. Springer, New York.

Jähne, B. (2005). *Digital image processing* (5th revised and extended edn). Springer, Berlin.

Jeulin, D. (2001). *Mechanics of random and multiscale microstructures*. Springer, Wien, New York.

Kehrwald, D. (2004). Parallel lattice Boltzmann simulation of complex flows. In *Simulation of Complex Flows – Applications and Trends*. Neutrale Interessenvertretung der Anwender der Finite-Elemente-Methode (FEM) und verwandter numerischer Verfahren (NAFEMS). NAFEMS Seminar, Niedernhausen, Germany, 3–4 May 2004.

Klette, R. and Rosenfeld, A. (2004). *Digital geometry*. Morgan & Kaufman, Amsterdam.

Koch, K., Ohser, J., and Schladitz, K. (2003). Spectral theory for random closed sets and estimating the covariance via frequency space. *Advances in Applied Probability*, **35**, 603.

Lachaud, J.-O. and Montanvert, A. (2000). Continuous analogs of digital boundaries: A topological approach to iso-surfaces. *Graphical Models*, **62**, 129.

Lang, C., Ohser, J., and Hilfer, R. (2001). On the analysis of spatial binary images. *Journal of Microscopy*, **203**, 303.

Levoy, M. (1990). Efficient ray tracing of volume data. *Association for Computing Machinery Transactions on Graphics*, **9**, 254.

Lindblad, J (2005). Surface area estimation of digitized 3D objects using weighted local computations. *Image and Vision Computing*, **23**, 111.

Lindquist, W.B., Venkatarangan, A., Dunsmuir, J., and Wong, T.-F. (2000). Pore and throat size distributions measured from sychrotron X-ray tomographic images of Fontainebleau sandstones. *Journal of Geophysical Research*, **105B**, 21508.

Lohmann, G. (1998). *Volumetric image analysis*. Wiley-Teubner, Chichester.

Lorensen, W. E. and Cline, H. E. (1987). Marching cubes: a high resolution 3D surface construction algorithm. *Computer Graphics*, **21**, 163.

Manolakis, D.G. and Proakis, J.G. (1996). *Digital signal processing* (3rd edn). Prentice-Hall, Englewood Cliffs.

Marcotte, D. (1996). Fast variogram computation with FFT. *Computers and Geosciences*, **22**, 1175.

Mardia, K.V. and Hainsworth, T.J. (1988). A spatial thresholding method for image segmentation. In *IEEE Trans. Pattern Analysis and Machine Intelligence*, Volume 6, Piscataway, pp. 919. Institute of Electrical and Electronics Engineers.

Maurer, C. R. and Raghavan, V. (2003). A linear time algorithm for computing exact Euclidean distance transforms of binary images in arbitrary dimensions. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Volume 25, Piscataway, pp. 265. Institute of Electrical and Electronics Engineers.

Meijster, A., Roerdink, J.B.T.M., and Hesselink, W.H. (2005). Euclidean skeletons of 3D data sets in linear time by the integer medial axis transform. *Computational Imaging and Vision*, **30**, 259.

Mumford, D. and Shah, J. (1989). Optimal approximations by piecewise smooth functions and associated variational problems. *Communications in Pure and Applied Mathematics*, **42**, 577.

Nagel, W., Ohser, J., and Pischang, K. (2000). An integral-geometric approach for the Euler-Poincaré characteristic of spatial images. *Journal of Microscopy*, **198**, 54.

Niblack, W. (1986). *An introduction to digital image processing.* Prentice-Hall, Englewood Cliffs, NJ.

Oh, W. and Lindquist, W.B. (1999). Image thresholding by indicator kriging. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Volume 21, Piscataway, pp. 590. Institute of Electrical and Electronics Engineers.

Ohser, J. and Mücklich, F. (2000). *Statistical analysis of microstructures in materials science.* J. Wiley & Sons, Chichester, New York.

Ohser, J. and Nagel, W. (1996). The estimation of the Euler-Poincaré characteristic from observations on parallel sections. *Journal of Microscopy*, **184**, 117.

Ohser, J., Nagel, W., and Schladitz, K. (2002). The Euler number of discretized sets – on the choice of adjacency in homogeneous lattices. In *Morphology of condensed matter* (ed. K. Mecke and D. Stoyan), Lecture Notes in Physics, Berlin, pp. 275. Springer.

Ohser, J., Nagel, W., and Schladitz, K. (2003). The Euler number of discretised sets – surprising results in three dimensions. *Image Analysis and Stereology*, **22**, 11.

Ohser, J., Schladitz, K., Koch, K., and Nöthe, M. (2005). Diffraction by image processing and its application in materials science. *Zeitschrift für Metallkunde*, **96**, 731.

Otsu, N. (1979). A threshold selection method from gray level histograms. In *IEEE Transactions on Systems, Man and Cybernetics*, Volume 9, Piscataway, pp. 62. Institute of Electrical and Electronics Engineers.

Park, J.-M., Looney, C.G., and Chen, H.-C. (2000). Fast connected component labeling algorithm using a divide and conquer technique. Technical report, The University of Alabama.

Peiró, J. (1999). *Surface grid generation*, pp. 19.1. CRC Press Inc., Boca Raton, Florida.

Perona, P. and Malik, J. (1990). Scale space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **12**, 629.

Remy, E. and Thiel, E. (2005). Exact medial axis with Euclidean distance. *Image and Vision Computing*, **23**, 167.

Revol, C. and Jourlin, M. (1997). A new minimum variance region growing

algorithm for image segmentation. *Pattern Recognition Letters*, **18**, 249.

Revol-Muller, C., Peyrin, F., Carrillon, Y., and Odet, C. (2002). Automated 3D region growing algorithm based on an assessment function. *Pattern Recognition Letters*, **23**, 137.

Rezk-Salama, C. (2001). *Volume Rendering Techniques for General Purpose Graphics Hardware*. Ph. D. thesis, Universität Erlangen-Nürnberg.

Ritter, G. X. and Wilson, J. N. (2001). *Handbook of Computer Vision Algorithms in Image Algebra*, Chapter 6, Connected Component Algorithms, pp. 173. CRC Press, Boca Raton.

Rosenfeld, A. (1970). Digital topology. *American Mathematical Monthly*, **86**, 621.

Rosenfeld, A. and Pfaltz, J.L. (1966). Sequential operations in digital picture processing. *Journal of the Association for Computing Machinery*, **13**, 471.

Rotman, J. J. (1993). *An introduction to algebraic topology*. Springer, Berlin.

Sabella, P. (1988). A rendering algorithm for visualizing 3D scalar fields. In *SIGGRAPH '88: Proceedings of the 15th annual conference on Computer graphics and interactive techniques*, New York, pp. 51. ACM Press.

Saito, T. and Toriwaki, J. (1994). New algorithms for Euclidean distance transformations of an n-dimensional digitised picture with applications. *Pattern Recognition*, **27**, 1551.

Sandfort, K. and Ohser, J. (2007). Labeling of n-dimensional images with choosable adjacency of the pixels. to appear in Image and Vision Computing.

Schatt, W. (1992). *Sintervorgänge*. VDI-Verlag, Düsseldorf.

Schladitz, K., Ohser, J., and Nagel, W. (2006*a*). Measurement of intrinsic volumes of sets observed on lattices. In *Discrete geometry for computer imagery* (ed. A. Kuba, L. G. Nyul, and K. Palagyi), Lecture Notes in Computer Science 4245, Berlin, pp. 247. Springer. Proceedings of the 13th International Conference on Discrete Geometry for Computer Imagery.

Schladitz, K., Peters, S., Reinel-Bitzer, D., Wiegmann, A., and Ohser, J. (2006*b*). Design of aoustic trim based on geometric modeling and flow simulation for non-woven. *Computational Materials Science*, **38**, 56.

Schmidt, V. and Spodarev, E. (2005). Joint estimators for the specific intrinsic volumes of stationary random sets. *Stochastic Processes and their Applications*, **115**, 959.

Schneider, R. and Weil, W. (2000). *Stochastische Geometrie*. Teubner-Verlag, Stuttgart, Leipzig.

Schroeder, W.J., Zarge, J.A., and Lorensen, W.E. (1992). Decimation of triangle meshes. *Computer Graphics*, **26**, 650.

Serra, J. (1982). *Image analysis and mathematical morphology, Vol. 1*. Academic Press, London.

Serra, J. (1988). *Image analysis and mathematical morphology, Vol. 2: Theoretical Advances*. Academic Press, London.

Sethian, J.A. (1999). *Level set methods and fast marching methods*. Cambridge University Press.

Soille, P. (1999). *Morphological image analysis*. Springer, Berlin.

Stoyan, D., Kendall, W.S., and Mecke, J. (1995). *Stochastic geometry and its applications* (2nd edn). J. Wiley & Sons, Chichester.

Succi, S. (2001). *The lattice Boltzmann equation for fluid dynamics and beyond*. Clarendon Press, Oxford.

Szabo, B. and Babuška, I. (1991). *Finite element analysis*. J. Wiley & Sons, New York.

Thurfjell, L., Bengtsson, E., and Nordin, B. (1992). A new three-dimensional connected components labeling algorithm with simultaneous object feature extraction capability. *Graphical Models and Image Processing*, **54**, 357.

Torquato, S. (2002). *Random heterogeneous materials: Microstructure and macroscopic properties*. Springer, New York.

Trier, O.D. and Jain, A.K. (1995). Goal-directed evaluation of binarization methods. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Volume 17, Piscataway, pp. 1191. Institute of Electrical and Electronics Engineers.

Tsao, Y.F. and Fu, K.S. (1981). A parallel thinning algorithm for 3-D pictures. *Computer Vision, Graphics and Image Processing*, **17**, 315.

Vincent, L. and Soille, P. (1991). Watersheds in digital spaces: an efficient algorithm based on immersion simulation. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Volume 13, Piscataway, pp. 583. Institute of Electrical and Electronics Engineers.

Watson, D.F. (1981). Computing the n-dimensional Delaunay tessellation with application to Voronoï polytopes. *Computer Journal*, **24**, 167.

Weatherill, N.P. and Hassan, O. (1994). Efficient 3-dimensional Delaunay triangulation with automatic point generation and imposed boundary constraints. *International Journal for Numerical Methods in Engineering*, **37**, 2005.

Weickert, J. (1998). *Anisotropic diffusion in image processing*. Teubner, Stuttgart.

Windreich, G., Kiryati, N., and Lohmann, G. (2003). Surface area estimation in practice. In *11th International Conference on Discrete Geometry for Computer Imagery* (ed. I. Nyström, G. S. di Baja, and S. Svensson), Volume 2886 of *Lecture Notes in Computational Science*, Berlin, pp. 358. Springer.

Winkler, G. (1995). *Image analysis, random fields and dynamic Monte Carlo methods: a mathematical introduction.*, Volume 27 of *Applications of Mathematics*. Springer, Berlin.

Wirjadi, O. (2007). Survey on 3D image segmentation. Technical report, Fraunhofer ITWM, Kaiserslautern.

Yoo, T.S. (2004). *Insight into images: Principles and practice for segmentation, registration, and image analysis*. AK Peters, Wellesley, MA.

Zienkiewicz, O. C., Taylor, R. L., and Zhu, J. Z. (2005). *Finite element method: its basis and fundamentals* (6th edn). Elsevier, Amsterdam.

This page intentionally left blank

4

RADIATION SOURCES AND INTERACTION OF RADIATION WITH MATTER

John Banhart

Advanced tomography requires well-defined beams of the required type of radiation, i.e. X-rays, electrons or neutrons. In this chapter, a brief description of the generation of such beams is given, and the fundamental interactions of these rays with solid matter are outlined.

4.1 General definitions

Each of the types of radiation treated in this book represents two fundamental aspects: that of particles and that of waves. The particle–wave dualism of quantum mechanics implies that it is impossible to explain all phenomena by adopting just one of the two viewpoints, but that both aspects have to be considered.

Moving particles have a rest mass, a mechanical momentum $\boldsymbol{p}$ and a kinetic energy E. The de Broglie relation assigns the wavelength λ to a particle:

$$\lambda = \frac{2\pi\hbar}{|\boldsymbol{p}|}. \tag{4.1}$$

Conversely, an electromagnetic wave of wavelength λ corresponds to a particle – a photon always travelling at the speed of light c – having both momentum and energy:

$$E \approx \frac{2\pi\hbar c}{\lambda}; \quad |\boldsymbol{p}| = \frac{E}{c}.$$

Planck's constant, $\hbar \approx 1.55 \times 10^{-34}$ Js, is the bridge between the particle and the wave description.

Travelling waves and particles can change their direction when they interact with matter or fields. In the case of waves one speaks of *diffraction* or *refraction*, whereas for particles the word *scattering* is more common. These phenomena are further characterized by terms such as *elastic/inelastic* and *coherent/incoherent*, and other attributes.

'Elastic' or 'inelastic' refer to the energy of the particles involved in a scattering event and are best described in the particle picture. The term 'elastic' means that the *total* kinetic energy of two particles remains constant during scattering and no internal degrees of freedom such as atomic excitations or vibrations are excited. Redistribution of kinetic energy between the two particles takes place due to the conservation of momentum end energy as the target particle initially

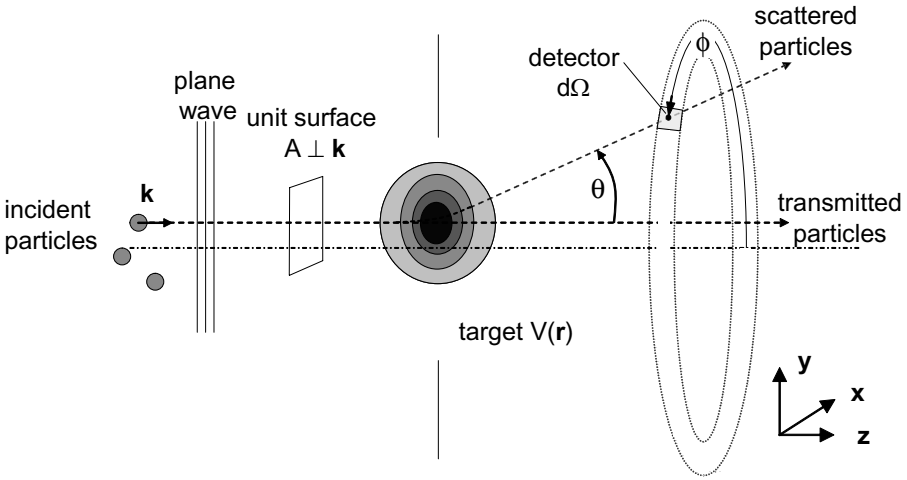


FIG. 4.1. Scattering of particles by a single potential $V(\boldsymbol{r})$.

at rest recoils. Therefore, the frequently heard sloppy statement that 'elastic' means that an incoming particle does not change its kinetic energy is incorrect.

The terms 'coherent'/'incoherent' refer to the wave aspect of particles. A wave is called (spatially) coherent if there is a relationship between its phase at different locations. The *coherence length* specifies the largest distance between such locations. A scattering event is called 'coherent' if there is a phase relationship between the wave before and after scattering or, equivalently, if a coherent wave is still coherent after scattering.

It makes sense to distinguish between scattering by single targets (atoms) and by ensembles of targets (matter) to make clear whether an effect has its origin in the individual scattering event or is an effect of collective scattering by many scattering centres.

In order to describe the interactions between different types of radiation and solid matter it is useful to repeat very briefly some of the important definitions and formulae related to scattering theory. The reader is referred to any textbook on quantum mechanics or scattering theory, such as Cohen-Tannoudji *et al.* (1977) on an elementary level, or Wu and Ohmura (1962), Foderaro (1971) or (McDaniel, 1986) for a more advanced treatment.

4.1.1 *Scattering from a single target*

Figure 4.1 defines the geometry of scattering of non-interacting incident particles by a single target – i.e. an atom, a nucleus, *etc.* – characterized by an interaction potential $V(\boldsymbol{r})$. In the wave picture, these incoming particles are represented by a plane wave and a wavevector $\boldsymbol{k} = \boldsymbol{p}/\hbar$. *Scattering* means that the configuration of the two particles remains the same, as opposed to a *reaction* where the particles are rearranged. Reactions are common when neutrons are involved but do not

play a role for imaging. For a moment we neglect both spin-related interactions and any inelastic mechanism that changes the total kinetic energy of the two particles and correspondingly the internal energy of one or both of the particles, i.e. we consider elastic scattering events only. Moreover, we use a centre-of-mass reference frame in which the scattering problem can be expressed by the relative co-ordinates of the two particles. The *flux* of the incident beam is given by the number of particles N traversing through a unit surface A per unit time t:

$$j = \frac{N}{At}, \quad \text{units: } \mathrm{m}^{-2}\,\mathrm{s}^{-1}.$$

The number of particles per unit time, n, hitting the detector placed outside of the incident beam and having an active area such that it covers a solid angle $d\Omega$ about the direction defined by the angles (θ, ϕ) will then be:

$$n = j\sigma_{\mathrm{diff}}(\theta, \phi)d\Omega,$$

where the quantity $\sigma_{\mathrm{diff}}(\theta, \phi)$ is the *differential elastic scattering cross-section*, units: m^2. The *elastic scattering cross-section* is given by:

$$\sigma_{\mathrm{s}} = \int \sigma_{\mathrm{diff}}(\theta, \phi)d\Omega,$$

or

$$\sigma_{\mathrm{diff}} = \frac{d\sigma_{\mathrm{s}}}{d\Omega}. \tag{4.2}$$

The scattering potential can be described by its action on an incoming plane wave $\psi_0 \propto \exp(ikz)$ that, apart from the wave transmitted without any interaction, is converted into an outgoing scattered spherical wave

$$\psi_s \propto f_k(\theta, \phi)\frac{\exp(i\boldsymbol{kr})}{r}, \quad k = \frac{\sqrt{2\mu E}}{\hbar},$$

for large distances r from the scattering centre. f is the complex *scattering amplitude*, μ the reduced mass of the scattered particle, E its energy in the centre-of-mass system. It is straightforward to show that

$$\sigma_{\mathrm{diff}}(\theta, \phi) = |f(\theta, \phi)|^2. \tag{4.3}$$

In many important cases the potential is a central potential, i.e. $V(\boldsymbol{r}) = V(r)$. In this case scattering is axially symmetric and both $f(\theta)$ and $\sigma_{\mathrm{s}}(\theta)$ are independent of the azimuthal angle ϕ. Moreover, the (quantized) angular momentum ℓ of the incoming particle is a constant of motion and we can decompose in- and outgoing waves into *partial waves* φ_ℓ and completely describe the action of the potential $V(r)$ by a set of real-valued, k-dependent *phase shifts* $\delta_\ell, \ell \in \mathbb{N}_0$. The effect of the potential is to shift the phase of each outgoing partial wave by $2\delta_\ell$ with respect

to the incoming wave, i.e. φ_ℓ is multiplied by $\exp(-2i\delta_\ell)$. For the total elastic scattering cross-section that explicitly depends on k, i.e. on energy, we can write:

$$\sigma_\mathrm{s} = \frac{4\pi}{k^2} \sum_{\ell=0}^{\infty} (2\ell + 1) \sin^2 \delta_\ell. \tag{4.4}$$

σ_s can show rapid variations as a function of energy in the vicinity of the so-called *scattering resonances*.

If the potential has a finite range, i.e. $V(r) = 0$ for $r > r_0$, this summation can be restricted to a finite set of values $\ell = 0 \cdots \approx kr_0$. For very short-ranged potentials or low energies $kr_0 \ll 1$ and δ_0 is sufficient to describe the entire scattering process, a case that is called *pure s-scattering*. In this case, f_k is also independent of θ and scattering is isotropic. Neutron scientists frequently use the *scattering length* $b = -\lim_{k\to 0} f_k$, and $\sigma_\mathrm{s} = 4\pi b^2$.

An important problem not yet treated is that of *absorption*. Herewith, we mean that during scattering particles 'disappear' with a certain probability, which means that incoming particles are no longer registered by the detector as elastically scattered particles since they have been, e.g. converted to other particles in a reaction or removed from the scattering zone in another way. As we are not interested in the details of absorption, a global, phenomenological description is useful. This is done by giving the phase shifts an imaginary part that is equivalent to multiplying the scattered wave by a factor with a modulus smaller 1, i.e. $|\exp 2i\delta_\ell| < 1$, to damp the wave. The scattering cross-section is now split into elastic scattering and absorptive contributions that sum up to the total cross-section:

$$\begin{aligned}
\sigma_\mathrm{s} &= \frac{\pi}{k^2} \sum_{\ell=0}^{\infty} (2\ell + 1) |1 - e^{2i\delta_\ell}|^2 \qquad \textit{elastic scattering},\\
\sigma_\mathrm{abs} &= \frac{\pi}{k^2} \sum_{\ell=0}^{\infty} (2\ell + 1) \left(1 - |e^{2i\delta_\ell}|^2\right) \qquad \textit{absorption},\\
\sigma_\mathrm{tot} &= \sigma_\mathrm{s} + \sigma_\mathrm{abs} \qquad \textit{total cross-section}.
\end{aligned} \tag{4.5}$$

The *optical theorem* of scattering theory then defines an important relation:

$$\sigma_\mathrm{tot} = \frac{4\pi}{k} \mathrm{Im} f_k(0).$$

Thus, the imaginary part of the elastic scattering amplitude f in the forward direction determines the total cross-section including all events.

4.1.2 *Scattering from an ensemble of targets*

Solid matter consists of many atoms and scattering can occur at various scattering potentials simultaneously, whenever the wave packet describing the incoming particle extends over many such potentials. We speak of *coherent scattering*

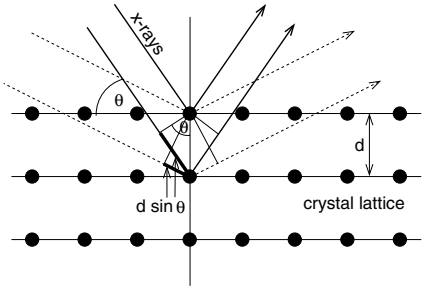


FIG. 4.2. Bragg reflection from an ordered lattice.

whenever scattering leads to interference effects between the individual scattering events. Naturally, a pre-requisite for coherence is that the individual scattering events preserve the phase of the incoming wave that is not the case, e.g. for Compton scattering, as we will see later. The scattered intensity will then be higher in certain directions than in others, depending on the wavevector $\boldsymbol{k}$ or, equivalently, on energy or wavelength. Many materials are crystalline, i.e. identical atoms reside on crystal lattices. For an infinite periodic lattice scattering takes place only in certain discrete directions, as described by Bragg's law (see Fig. 4.2):

$$2d\sin(\theta) = n\lambda. \tag{4.6}$$

If the individual scattered wave packets do not interfere scattering is called *incoherent*. In this case, total scattering is just a sum of all individual scattering events. This can happen when the atoms are too far apart. Other reasons for incoherence can be manifold: positional disorder of the scattering centres due to static displacements from the ideal lattice sites, thermal fluctuations or a variation of the scattering properties of the individual scattering centres due to differences in chemical (different elements), nuclear (different isotopes) or spin properties (different nuclear spin orientations). In practice, scattering will often be a mixture of coherent and incoherent scattering.

Whenever a particle passes through a sample it may be scattered more than once. This phenomenon is called *multiple scattering*. The individual scattering events are independent of each other. Multiple scattering makes data analysis difficult since the detector placed behind the sample measures the total resulting scattering angle and cannot distinguish between particles that have been scattered just once or many times. Use of thin samples helps to minimize this effect and to simplify analysis.

4.1.3 *Absorption and attenuation coefficients*

The absorption cross-section σ_{abs} given in eqn 4.5 provides the information we need to quantify the absorption of a beam by a sample consisting of many atoms. If N_V is the number of atoms per unit volume, the *linear absorption coefficient* μ_{abs} is defined as

$$\mu_{\text{abs}} = N_V \sigma_{\text{abs}}, \quad \text{units: m}^{-1}. \tag{4.7}$$

Usually, neutron scientists do not use SI units but prefer centimetres and barns. In this case, one can use:

$$\mu\ [\text{cm}^{-1}] = 0.6022 \times \frac{\rho\ [\text{g/cm}^3]}{A[\text{g/mol}]} \times \sigma\ [\text{barn}], \tag{4.8}$$

where [...] means that the respective quantity has to be used in the units given. A thin homogeneous sample will modify a parallel incident beam as shown in Fig. 4.3(a). For the intensity change caused by absorption in a slice of thickness dx we write:

$$\frac{dI}{I} = -\mu_{\text{abs}}\, dx; \quad dI = I_t - I_0. \tag{4.9}$$

In practice, a detector will not just measure a decrease in intensity dI caused by absorption in the sense described, but will also note further losses since other particles are scattered away and miss the detector. On the other hand, intensity scattered onto the detector from regions in the sample away from the direct beam trajectory may increase intensity. The extent to which this happens clearly depends on the actual conditions such as the illuminated area and the detector area, see Fig. 4.3(b).

For a point detector and a well-collimated, thin pencil beam (Fig. 4.3(a)) the total attenuation of the beam will be given by the sum of absorption and scattering and the measured *attenuation coefficient* will be

$$\mu = N_V(\sigma_{\text{abs}} + \sigma_{\text{s}}) = N_V \sigma_{\text{tot}}, \tag{4.10}$$

and this is used in eqn 4.9 that by integration from $x = 0$ to $x = d$ yields the well-known exponential *Beer–Lambert law* of attenuation by a sample of thickness d:

$$\frac{I}{I_0} = \exp(-\mu d). \tag{4.11}$$

In this integration, however, we have tacitly assumed that, (i) μ is the same for all particles or wavelengths in the beam and, (ii) that μ is a constant over sample thickness. (i) is usually not true for 'polychromatic' beams, while (ii) is not justified whenever particles are not just simply absorbed or transmitted but lose kinetic energy in multiple and/or inelastic scattering events, after which they continue to travel through the sample. As μ often increases with falling particle energy, μ turns into an implicit function $\mu(x)$ of x in eqn 4.9 and eqn 4.11 no longer holds. This is the case for electrons that undergo cascades of scattering processes during which they lose energy, whereas neutrons and X-rays obey eqn 4.11 quite well.

By defining $\mu_m = \mu/\rho$, where ρ is the sample density, we obtain the *mass attenuation coefficient* (units m^2/kg), the use of which can be more convenient

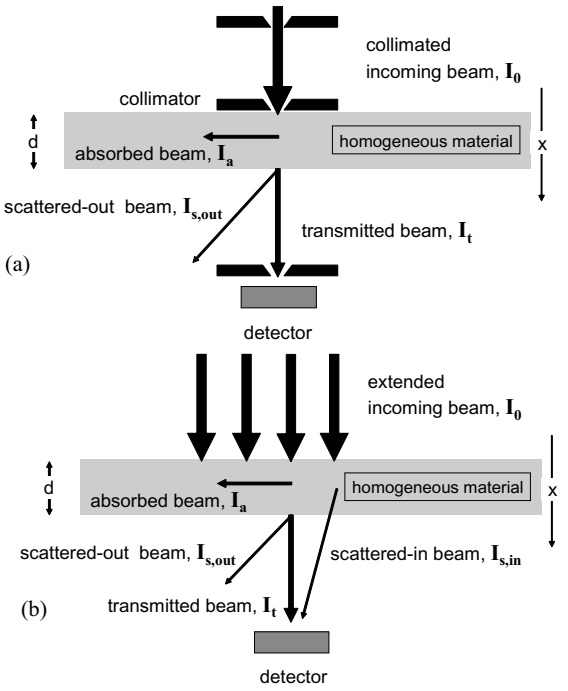


FIG. 4.3. Attenuation by a thin slab of material of, (a) a collimated beam and, (b) an extended beam.

since it has the same value for the solid, liquid or gaseous state of a given element. Equation 4.11 can be written as

$$\frac{I}{I_0} = \exp(-\mu_m d_m),$$

where the *mass-thickness* $d_m = \rho d$ is used (units: kg/m^2). Whenever a sample is composed of different elements of different attenuation coefficients μ_α, the effective mass attenuation coefficient of the mixture will be a function of the attenuation coefficients of these elements, μ_α, namely approximately:

$$\overline{\left(\frac{\mu}{\rho}\right)} = \sum_\alpha x_\alpha \left(\frac{\mu_\alpha}{\rho_\alpha}\right),$$

where x_α are the mass fractions in the mixture. This rule of mixture is considered quite precise whenever interactions between the atoms can be neglected, which is the case, e.g. for X-ray energies above 30 eV and sufficiently far away from the absorption edges, the error being in the range of a few per cent (Jackson and Hawkes, 1981).

The inverse of the attenuation coefficient has the unit of a length. It is the *mean path* after which a beam has been attenuated to $1/e \approx 37\%$ of its initial intensity.

If the incident beam is polychromatic, the attenuation behaviour will be a function of the energy-dependent attenuation coefficient $\mu(E)$. The transmitted beam will show a different spectral composition from that of the incident one, a phenomenon that is called *beam hardening* since rays with a smaller penetration depth ('softer' rays) are more attenuated than 'harder' ones. Some examples for beam hardening of X-rays can be found in Section 5.4.5.

4.1.4 *Refraction and reflection*

Refraction and reflection of X-rays and neutrons can be described using the same formalism as for visible light. Knowing the related properties is essential whenever designing optical components, see e.g. Section 7.2.1.

Reflection is governed by the equation relating the angles between the reflecting surface and the incoming and outgoing beams θ:

$$\theta_{\text{incident}} = \theta_{\text{reflected}},$$

while Snell's refraction law states that for refraction the following law holds:

$$n_1 \sin\theta_{\text{incident}} = n_2 \sin\theta_{\text{refracted}},$$

where n_1 and n_2 are the *refractive indices* of the respective media. For neutrons and X-rays $n < 1$ except for very few nuclei, see Fig. 14.1, unlike for visible light where $n > 1$ and the refraction direction is therefore different for X-rays and neutrons compared to light, see Fig. 7.21(b).

By defining a complex index of refraction $\hat{n}$ one describes refraction and absorption by one quantity and one writes:

$$\hat{n} = 1 - \delta + i\beta, \tag{4.12}$$

where δ is the refractive decrement describing the small deviation of the real part of n from unity and β is the absorptive part of n related to the linear absorption coefficient μ by:

$$\beta = \frac{\lambda}{4\pi}\mu. \tag{4.13}$$

Both δ and β are small for X-rays and neutrons, cf. eqn 7.7. Knowledge of the complex index of refraction n is sufficient to calculate the complex dielectric constant $\hat{\epsilon}$ and the reflectivity R of a material.

4.2 Generation of X-rays and their interaction with matter

4.2.1 *X-ray sources*

X-rays can be created in two ways: either by ejecting core electrons from atoms and letting other electrons fall into the holes created and emit X-ray photons, or by accelerating or decelerating free charged particles.

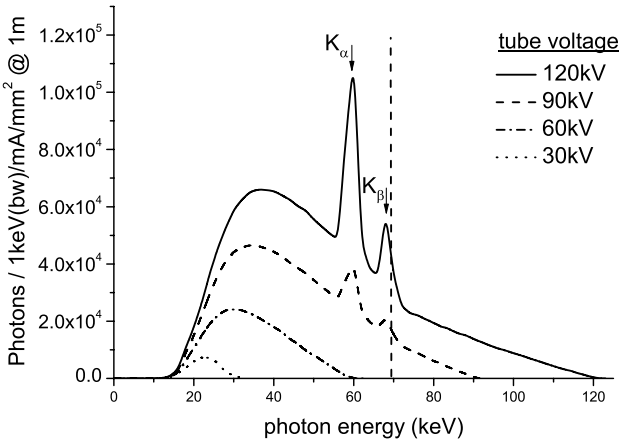


FIG. 4.4. Spectrum of a X-ray tube with a tungsten anode for 4 different tube voltages. Calculated with `XOP` (Sanchez del Rio and Dejus, 2004). The broken line denotes the absorption edge of tungsten at ≈69.5 keV.

4.2.1.1 *X-ray tubes* Traditional X-ray tubes exploit both mechanisms. By bombarding a metal target with electrons, holes in low-lying levels of the target atoms are created that are then filled by electrons. In some cases this gives rise to characteristic (fluorescence) radiation, in other cases de-excitation is radiationless. Other electrons are simply decelerated in one or more steps and release their kinetic energy as electromagnetic radiation giving rise to the typical continuous *bremsstrahlung*, see Section 4.3.4. X-ray tubes therefore have a spectrum that is the superposition of a broad distribution of energies with a cutoff at the energy of the bombarding electrons and some peaks related to the target material. Figure 4.4 shows the calculated spectrum for a tungsten target bombarded with electrons of different energies. Clearly, higher electron energies produce both a higher X-ray flux and a tail in the spectrum that extends to higher photon energies. Moreover, the fluorescence energy levels of tungsten are increasingly activated and contribute to the emitted radiation. The two K-Lines K_α and K_β can easily be distinguished.

4.2.1.2 *Synchrotrons and storage rings* Synchrotron radiation is created by accelerating free particles in high vacuum. By the laws of electrodynamics this acceleration leads to the emission of electromagnetic radiation (Jackson, 1975). Linear particle acceleration is one possibility, but apart from the very high electric fields one would need it is more practical to hold the charged particles on a closed trajectory in order to obtain a source of continuous radiation. Magnetic fields are used to force the particles onto the desired orbit and prevent them from flying in a straight line. The radial acceleration associated with the change of direction then generates radiation.

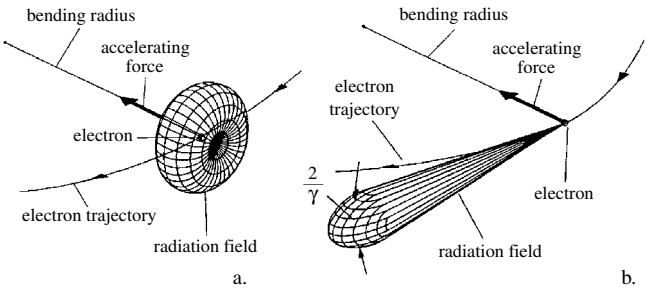


FIG. 4.5. Directionality of synchrotron radiation from (a) slow ($v \ll c$), (b) fast ($v \approx c$) electrons. Taken from Wille (2005).

In the easiest case a circular path is chosen. This corresponds to a uniform acceleration perpendicular to the direction of movement and the emission of radiation in the tangential direction. It can be shown that a charged particle, mass m, charge e, kinetic energy E, moving on a circle with radius R emits radiation of different wavelengths with a total energy Q per unit time t of (Duke, 2000, eqn 5.20):

$$\frac{dQ}{dt} = \frac{2}{3}\frac{c}{4\pi\epsilon_0}\frac{e^2}{(mc^2)^4}\frac{E^4}{R^2}.$$

In deriving this formula a particle velocity close to the speed of light c has been assumed, i.e. $v/c \approx 1$, which is always the case when X-rays are to be created. Obviously, in order to obtain a high radiation power, particles with a high charge and a low mass are ideal and therefore electrons or positrons are chosen. Moreover, the particle energy has to be high and the radius of orbit as low as possible. For this reason, in modern dedicated synchrotron facilities E ranges up to 8 GeV and a high effort is put into minimizing R. At these high relativistic velocities the emitted radiation has a very narrow spatial distribution in the forward direction of movement (see Fig. 4.5(b) as opposed to low energies, where radiation is emitted in all directions, Fig. 4.5(a)).

In the so-called *first generation* of synchrotron radiation sources X-rays were generated as a by-product. The main purpose of these facilities was the production of particle beams, e.g. for particle-physics experiments.

In the *second generation*, instead of a circular orbit a polyhedral path is chosen on which bending magnets change the particle's direction after which they fly on a straight section again before entering the next bending device. The advantage of such a design is that the individual radiation sources are more localized – allowing them to be used in beamlines placed around the loop – and the radii of the individual bends can be made small, producing X-rays of a higher energy and intensity. Moreover, on the section between the bending magnets there is space for other crucial devices needed for maintaining the beam

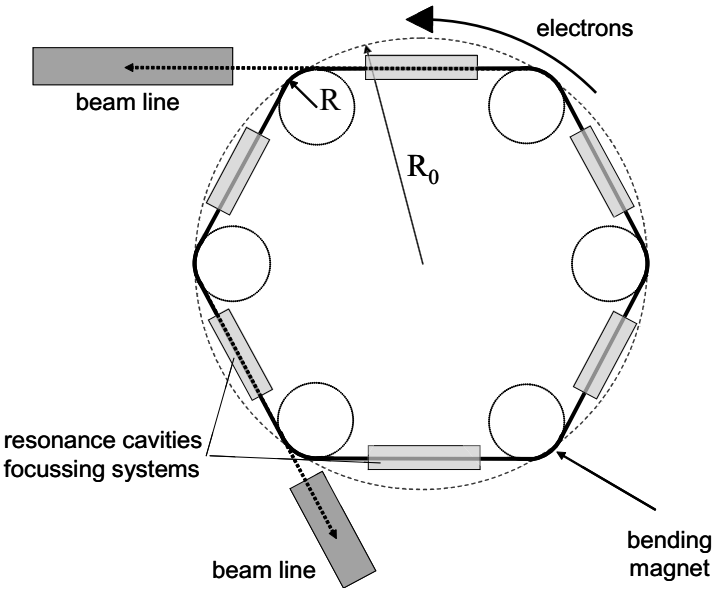


FIG. 4.6. Schematic layout of a second-generation synchrotron radiation source.

stability: accelerating radio-frequency cavities that compensate the loss in energy of the electrons and focusing magnetic lens systems that prevent the beam from defocusing. The schematic layout of a synchrotron radiation source is given in Fig. 4.6. Electrons circulate in a storage ring that consists of 6 bending magnets and 6 straight sections here and are kept on a constant orbit and at constant energy by means of dedicated systems between the bending devices. Radiation is created in the bends and is guided into beamlines and to the instruments there. Clearly, the radius of each bend is much smaller than in the case of one big circular orbit (compare R and R_0), and consequently, the radiation is harder and more intense,

In *third-generation* synchrotron sources radiation is created by so-called *insertion devices*, arrays of various magnets that undulate the electron beam many times such that radiation from each curve superimposes. Such *wigglers* or *undulators* can be equipped with superconducting magnets to keep the radius R as small as possible. Insertion devices are located in the straight sections between two bending magnets and give rise to a much more brilliant radiation than these.

The spectrum emitted by a bending magnet (BM) is a universal function of the reduced energy E/E_c of the electron, where the *characteristic photon energy* is defined by:

$$E_c = \frac{3\hbar c\gamma^3}{2R}, \qquad \gamma = \frac{1}{\sqrt{1-(v/c)^2}}.$$

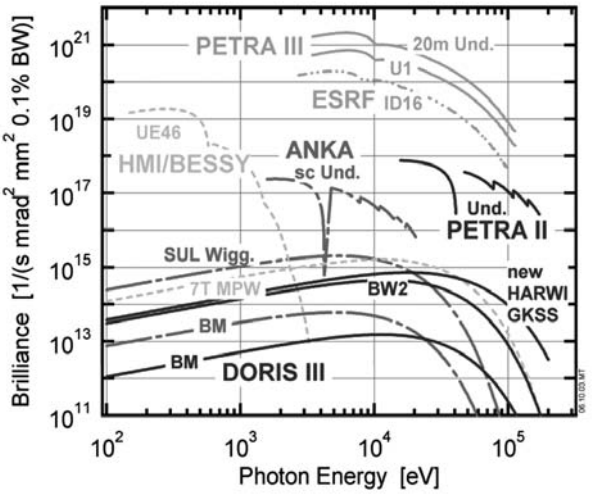


FIG. 4.7. X-ray spectrum emitted by various existing and planned facilities at different European synchrotron radiation sources (Source: HMI Berlin).

This energy divides the spectrum into two ranges of equal radiated power. The photon energy spectrum emitted by an electron revolving on a circular orbit can be calculated and will have a spectral distribution with a maximum value just below E_c (Duke, 2000). Instead of giving this spectrum, Fig. 4.7 shows the spectral brilliances (to be defined below) of a selection of some European synchrotron radiation sources. Obviously there is a large range of achievable spectra depending on the type of insertion device used and the electron energy.

Figure 4.6 only shows the actual storage ring. Electrons (or positrons) have to be created, accelerated and injected into this storage ring in separate facilities.

In order to characterize the quality of a synchrotron X-ray beam, specification of flux is not meaningful since it does not account for the collimation and spectral narrowness of a beam. Instead, the term *brilliance* is used. It specifies the number of photons N emitted per time t from an area A (this is flux) and divides this by the solid angle $d\Omega$ into which the radiation is emitted and then relates it to 10^{-3} of the spectral bandwidth $\Delta\lambda$ of the emitted radiation:

$$B = \frac{N}{Atd\Omega\ 10^{-3}\Delta\lambda}.$$

B is usually given as photons/(s mrad^2 mm^2 $10^{-3}\Delta\lambda$). *Brightness* or *spectral brightness* are synonyms for brilliance.

Nowadays, many synchrotron sources have been built worldwide with the only purpose to produce X-rays. The number and quality of sources and their use is rapidly growing. Figure 4.8 shows the development of brilliance for X-ray sources starting from the first X-ray tubes. The impressive and dynamic increase

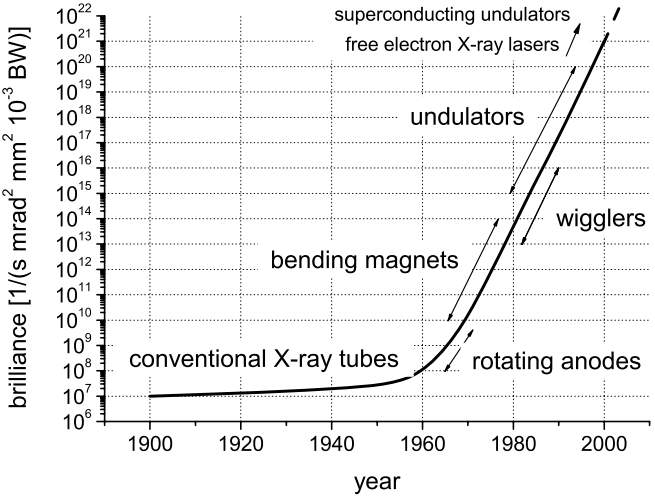


FIG. 4.8. Historical development of brilliance of X-ray sources (after Duke (2000, Fig. 1.9)).

is obvious. The development of increasingly powerful sources is ongoing. Free-electron lasers promise a further increase in brilliance and will also allow for imaging on very short time scales.

4.2.2 *X-ray interactions: overview*

X-rays are electromagnetic waves, i.e. oscillating electric and magnetic fields. Electrons respond to these fields by oscillating with the same frequency and by re-emitting radiation of the same wavelength and phase as the incident radiation but in directions different from the incident direction, thus scattering these X-rays.

For tomography or radiography in materials research, X-rays with energies between, say, 5–150 keV are most useful since engineering materials often contain metals and a certain penetration depth is required. Too soft X-rays do not penetrate deeply enough, too hard X-rays are not absorbed enough, leading to poor image contrast. This energy range is called the range of *hard X-rays*. Very soft X-rays, i.e. well below 1 keV as used in biology are less important. We shall therefore focus on effects that are important in this energy range.

The physics of photon interaction with atoms is a well-established field and shall be reviewed here only very briefly, see e.g. Jackson and Hawkes (1981). Any process that changes the direction of an incident photon or leads to its absorption can contribute to attenuation. A short summary of the relevant processes is given in Fig. 4.9. A parallel incident beam is partially transmitted without any interactions. Some photons are elastically scattered, some undergo inelas-

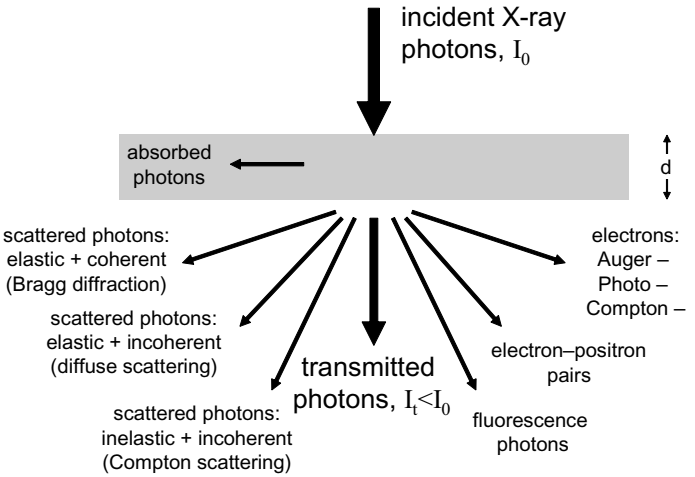


FIG. 4.9. Overview of X-ray interactions with matter. The direction of arrows is not directly related to direction of radiation.

tic collisions with the electrons in the sample. Elastic scattering from ordered crystalline solids can be coherent, while elastic scattering from a disordered assembly of atoms leads to a partial loss of coherence of the outgoing wave. The photoelectric effect can occur, after which the incident photon has disappeared and new photons or electrons leave the sample. Finally, a photon can create an electron–positron pair or be absorbed by a nucleus followed by the ejection of neutrons, but this requires photon energies above 2×511 keV or about 5 MeV, respectively, which is never the case in the field of interest of this book.

4.2.3 *Photoelectric effect*

The incident X-ray photon transfers its energy to an electron bound to an atom of the sample. The electron is ejected with the energy of the incident photon minus the binding energy of the electron in the atom. The excited atom relaxes as soon as another electron falls into the hole, either emitting characteristic fluorescence photons in all directions, or Auger electrons, the relative importance of each of these effects given by the fluorescence yield of the atom. High-Z atoms rather emit X-rays, whereas light atoms favour the Auger effect. The absorption cross-section of the atom depends on photon energy and shows rapid changes at energies corresponding to core electron states, the K, L, M-edges. The absorption edges contain some fine structure caused by the neighbourhood of the absorbing atom species that is exploited in spectroscopic methods such as XANES or EXAFS (see Section 8.2.2 for an imaging application).

The absorption cross-section well above an absorption edge is a function of both the atomic number Z of the atom and the energy (or wavelength) of the X-ray photon. Depending on the approximation used in the treatment of

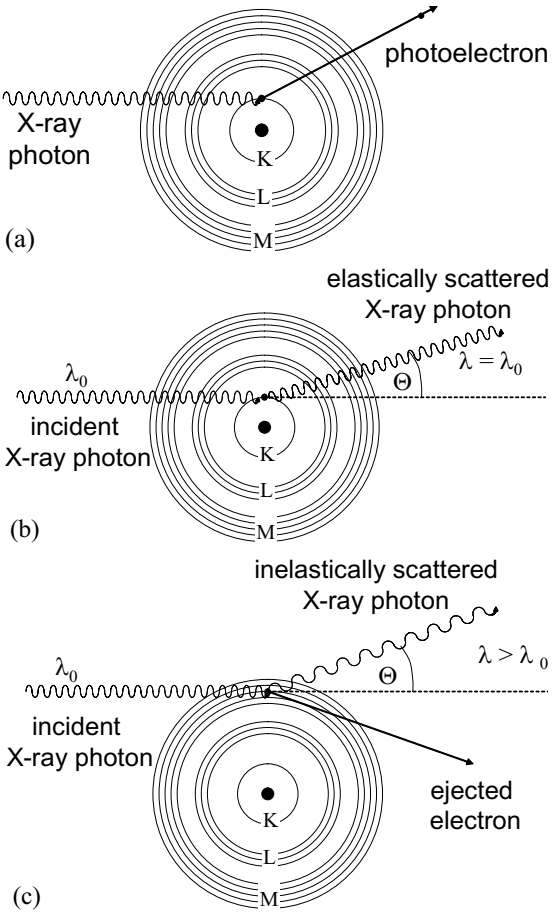


FIG. 4.10. Basic effects contributing to absorption of X-rays by matter in the energy range used for tomography. (a) photoelectric effect, (b) elastic scattering, (c) Compton effect.

the problem, different dependencies are obtained such as (Jackson and Hawkes, 1981):

$$\sigma_{\mathrm{pe}}(Z, E) \propto Z^5 E^{-3.5}. \tag{4.14}$$

Equations such as eqn 4.14 are called the *Bragg–Pierce law*. In the literature, the exact functional dependence of $\sigma_{\mathrm{pe}}(Z, E)$ varies slightly. One also finds other expressions such as (Jenkins and Snyder, 1996)

$$\sigma_{\mathrm{pe}}(Z, E) \propto AZ^4 E^{-3},$$

where $A = Z + N$ is the atomic mass number, or other laws $\sigma_{\mathrm{pe}} \propto Z^n$ with $n = 4 - 5$. There is no reason to believe that the Z dependence and the E

dependence of $\sigma_{\mathrm{pe}}(Z, E)$ can be exactly factorized for more than a fairly narrow energy range. For very high energies, e.g. above 511 keV, the E dependence flattens off to E^{-1} (Jackson and Hawkes, 1981). As it is difficult to provide analytical expressions for $\sigma_{\mathrm{pe}}(Z, E)$ this quantity is usually described numerically in terms of fitting procedures to experimental data (Ebel *et al.*, 2003), given as tabulated data (Hubbell and Selzer, 1996) or provided as an interactive web applet (Gullikson, 2004).

4.2.4 *Elastic scattering*

Whenever electromagnetic waves meet electrons in a sample they mainly interact via their oscillating electric field. Regardless of whether the electrons are either tightly bound to an atom or are free, they can elastically scatter the (X-ray) photon, a process that is called either *Rayleigh* or *Thomson scattering*, respectively. For low energies at which the photon momentum can be neglected the incident photon is scattered such that the wavevector just changes direction, not modulus, i.e. photon wavelength is conserved. Scattering from an atom containing many electrons is a co-operative effect of all electrons and the atom remains in the ground state after scattering The process is strongly forward-directed and the phase relationship between incoming and outgoing wave is conserved. If scattering takes place on an ordered array of scattering centres, Bragg diffraction is observed, since in certain directions the incident wave adds up in intensity by this coherent mechanism.

For a free electron the Thomson scattering cross-section does not depend on photon energy and is equal to $\sigma_T = 0.665$ barn (Jackson, 1975). For atoms with Z bound electrons the elastic scattering cross-section approximately scales with Z^2 for small angles, i.e. (Jackson and Hawkes, 1981):

$$\sigma_{\mathrm{el}}(Z) \propto Z^2,$$

and decreases with photon energy, as can be seen from Fig. 4.11.

4.2.5 *Inelastic scattering*

An incident photon with a non-negligible momentum interacts with an electron in the sample, either a bound electron or a fairly free band electron. Part of the kinetic energy of the incident photon is transferred to the electron, while a photon with lower energy is emitted, i.e. the process is inelastic from the viewpoint of the photon. As the emitted photon has lost any phase relationship with the incident one, this type of scattering is incoherent (Lifshin, 1999). The change of wavelength is given by the scattering angle θ:

$$\Delta\lambda = \frac{2\pi\hbar}{m_e c}(1 - \cos\theta),$$

where m_e is the electron mass. This formula can be derived by describing the electron/photon system as a classical elastic collision problem, thus demonstrating the semantic confusion around the word ‘inelastic’ mentioned on page 108.

The inelastic (Compton) scattering cross-section increases with the number of electrons, i.e. atomic number. Using the Klein–Nishina expression for the scattering cross-section from a single free electron $\sigma_{\mathrm{KN}}(E)$, one can write, in cases where the photon energy exceeds the electron binding energies:

$$\sigma_{\mathrm{Compt}}(Z, E) = Z\sigma_{\mathrm{KN}}(E).$$

In practice, this equation is not satisfied very well for the energies used for tomography (Jackson and Hawkes, 1981) and the cross-section is below this value. Figure 4.11 gives an impression of the energy dependence of σ_{Compt} for two chemical elements.

4.2.6 *Total interaction*

Summing up the relevant contributions to the total absorption cross-section – neglecting pair production and the photonuclear absorption – one obtains:

$$\sigma_{\mathrm{tot}} = \sigma_{\mathrm{pe}} + \sigma_{\mathrm{el}} + \sigma_{\mathrm{Compt}},$$

which is a measurable quantity via eqn 4.10. The relative importance of the various mechanisms strongly depends on energy and atomic number. For the energy range usually applicable in materials research the photoelectric effect dominates. Figure 4.11(a) gives the three contributions for aluminium. The K-edge of Al leads to the mentioned discontinuity at 1565 eV. Just for high energies ($E \geq 50\,\mathrm{keV}$) the Compton effect becomes dominant. For lead, K-, L- and M-absorption edges can be seen in the energy range given, see Fig. 4.11(b). The photoelectric effect is the dominant effect up to very high energies here. Figure 4.11(c) shows for which energies and elements the photoelectric or the Compton effect dominates.

4.2.7 *Refraction*

Refraction of X-rays can be characterized by two small quantities that describe their phase shift and their attenuation, see eqn 4.12. Applying the Born approximation one can obtain the useful formula for δ (Paganin, 2006):

$$\delta = \frac{ZN_V r_e}{2\pi}\lambda^2, \tag{4.15}$$

where $r_e \approx 2.82 \times 10^{-15}\,\mathrm{m}$ is the classical electron radius and ZN_V the number of electrons per unit volume. For typical X-ray wavelengths this formula yields values $\delta \ll 1$.

4.3 Generation of electrons and their interaction with matter

Electron tomography is carried out in transmission electron microscopes (TEMs). Details of the very elaborate image-formation process will not be treated in this book. The literature contains many books on TEM, ranging from elementary (Williams and Carter, 1996) books to monographs meant to serve the expert (Reimer, 1997).

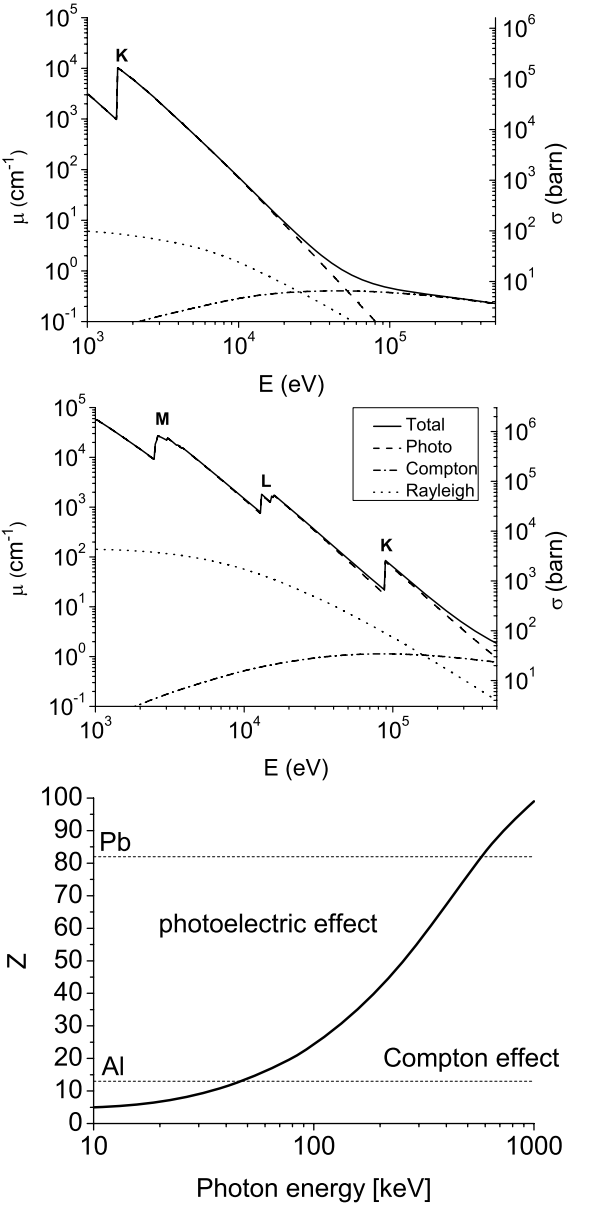


FIG. 4.11. X-ray absorption coefficient and absorption cross-section of (a) aluminium and, (b) lead. (c) Photon energy / atomic number map showing which of the two effects photoelectric effect/Compton effect dominates. Data determined using data base `DABAX` of the software package `XOP` (Sanchez del Rio and Dejus, 2004).

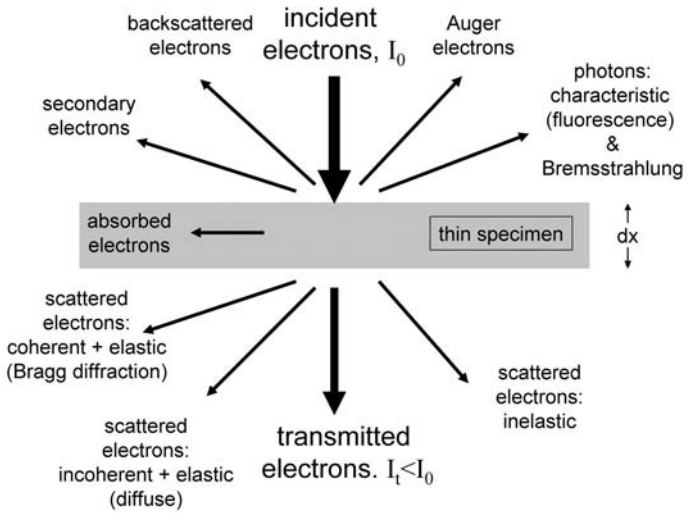


FIG. 4.12. Overview of electron interactions with a thin slab of matter. The direction of arrows is not exactly related to the direction of radiation.

4.3.1 *Electron sources*

High-quality imaging requires an electron beam that is bright, monochromatic with a low energy spread and geometrically well defined. The easiest way to generate electrons is by *thermoemission*. Heating to a high temperature gives electrons a high enough energy to overcome the work function of the surface and to leave a negatively charged cathode. Tungsten can be heated to 2400 °C and is therefore a preferred material. LaB_6 is another choice, although it cannot withstand such temperatures (just about 1400 °C), but its work function is much lower and therefore thermoemission is sufficient. After extraction from the cathode the electrons achieve their energy by acceleration to hundreds of keV, but the initially thermic energy spectrum is still superimposed on this energy, giving rise to energy spreads from 1.5 to 3 eV.

Another concept to generate electrons is *field emission*. High voltage applied to a very sharp tip ($R \approx 100$ nm) creates such high field gradients that electrons are removed even at room temperature. Field emission sources have a higher brightness, smaller size of the source (beams can be as narrow as 1 nm at the sample location) and a smaller spread in energy of the order of 0.3 eV.

A system of lenses shapes the electron beam and directs it to the specimen plane (Reimer, 1997). There, interactions between electrons and the atoms of the sample take place that eventually provide the information used for imaging.

4.3.2 *Electron interactions: overview*

Electrons interact with the electric charges in condensed matter – the positively charged nuclei and the electrons – primarily via electrostatic (Coulomb) inter-

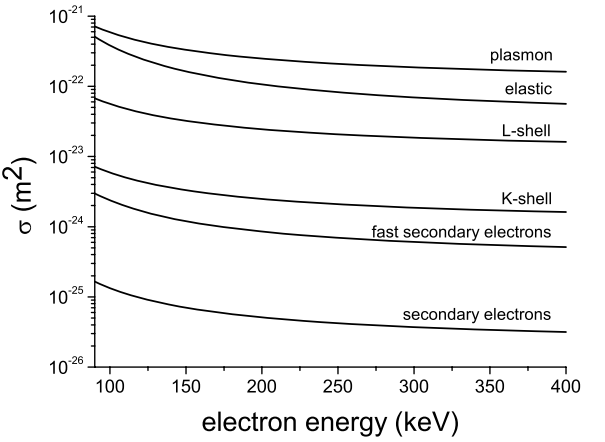


FIG. 4.13. Cross-sections for elastic (E) and various inelastic scattering processes in aluminium as a function of electron energy and assuming small scattering angles $\theta \approx 0$. 'P' stands for plasmon interaction, 'K' and 'L' for K- or L-shell ionization and (F)SE for the generation of (fast) secondary electrons (adapted from Williams and Carter (1996)).

actions. These interactions are very strong compared to those of X-rays and neutrons and the corresponding penetration depth is small. For this reason samples have to be very thin to allow for transmission by a significant fraction of the beam. Figure 4.12 summarizes the various mechanisms schematically. One can see that the incoming electrons are either transmitted, elastically scattered or scattered while losing energy that is transferred to secondary, backscattered or Auger electrons, or to characteristic or continuous X-ray photons. Figure 4.13 gives some interaction cross-sections for electrons of different energies via various interaction mechanisms that will be discussed in the following sections.

4.3.3 *Elastic scattering*

Interactions can be elastic in the sense that the scattered electron (hardly) changes its kinetic energy except for a small recoil effect. To give an example, a 100-keV electron merely transfers 1.9 eV of its kinetic energy to a copper atom when scattered under 90° due to its low mass (Reimer, 1997). In this treatment we neglect radiation losses that occur due to the deflection of the charged particle (see next section). An electron that impinges on a neutral atom will experience Coulomb interactions that are negligible at larger distances and become increasingly attractive as it penetrates the cloud of bound electrons and gets closer to the positively charged nucleus. The simplest case is that of the interaction between an individual electron and an atom, see Fig. 4.14(a). Far away from the neutral atom the interaction forces are very weak. In cases where the electron

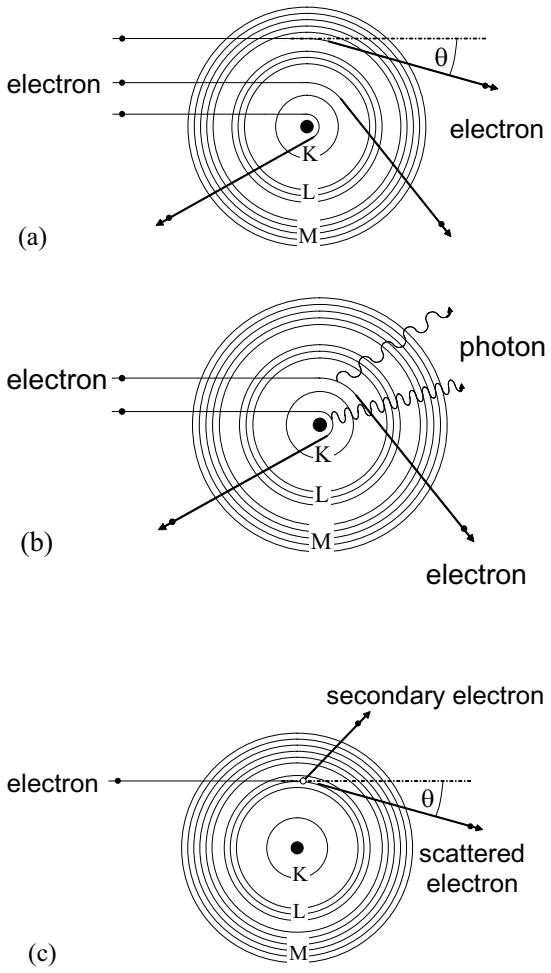


FIG. 4.14. Interactions between an electron and an atom. (a) elastic Coulomb scattering (atom and nucleus), (b) inelastic scattering associated with emission of bremsstrahlung, (c) inelastic scattering associated with generation of excitations.

enters the system of electrons, the positive nucleus is no longer screened completely and there is a net attractive force acting on the electron. Depending on the scattering parameters the electron can be deflected by any angle θ between 0° and 180°. Ignoring screening by the electron shell of the atom we can apply the well-known non-relativistic *Rutherford formula* for elastic Coulomb scattering:

$$\sigma_{\text{diff}} = \frac{Z^2 e^4}{16E^2 \sin^4(\theta/2)}. \tag{4.16}$$

An incoming electron will be deflected by a certain angle and also will be changed in phase by an θ-dependent shift. The strength of scattering obviously increases markedly with atomic number Z and decreases with electron energy E. Moreover, the scattered intensity is strong for $\theta = 0°$ (forward scattering), weak for $\theta = 180°$ (backward scattering). Screening can be included by defining a screening parameter and a relativistic counterpart of eqn 4.16 is also available. Elastic scattering is a major source of contrast in TEM (Williams and Carter, 1996). High-angle scattering is especially important since it allows one to distinguish between various elements. This Z contrast is exploited in electron tomography when the HAADF STEM mode is chosen, see Chapter 11.

4.3.4 *Inelastic scattering*

Whenever a charged particle changes its direction of flight it emits electromagnetic radiation. This was the basis for the generation of synchrotron radiation as discussed in Section 4.2.1.2. Electrons deflected by a positive charge therefore also emit electromagnetic radiation and lose kinetic energy. For small scattering angles and/or electron energies this loss is negligible and scattering can be approximately called 'elastic', but for larger energies and larger scattering angles (corresponding to a smaller radius of the orbit around the nucleus) radiation losses increase. The result is that scattering becomes inelastic, see Fig. 4.14(b). This is the physical basis for the generation of bremsstrahlung in a X-ray tube.

Another inelastic scattering mechanism is depicted in Fig. 4.14(c): an incoming electron interacts with a bound electron of a particular shell i and transfers a part of its kinetic energy to it. The shell electron is either transferred to a higher (empty) state and is completely removed from the atom that in turn is ionized. For ionization the incoming electron must have more kinetic energy than the ionization energy for the ith shell, E_i. As with the photoelectric effect in Fig. 4.10(a) various secondary effects can occur after ionization: the hole can be filled by a third electron giving rise to characteristic (fluorescence) radiation or a cascade of Auger electrons is created (not shown in Fig. 4.14(c)). The X-rays are emitted uniformly in all directions and give rise to the characteristic lines in the spectrum of an X-ray tube, see Fig. 4.4.

An electron can interact with electronic states of the target material (intra- or interband transitions), or with charge oscillations (plasmons) in the material. Typical interaction energies range from 1 to 50 eV. Moreover, inelastic interactions with phonons (lattice vibrations) of a solid target or molecular excitations are possible. In this case, the energies range well below 1 eV (Reimer, 1997).

The electrons generated in the various inelastic processes leave the specimen as *secondary electrons* that are very useful for some electron microscopical imaging modes. Sometimes, an incident electron can be backscattered after having undergone multiple scattering events that also leads to valuable imaging information.

4.4 Generation of neutrons and their interaction with matter

4.4.1 *Neutron sources*

Neutrons are abundant in nature since all atoms, except hydrogen ^{1}H, contain neutrons. However, as the binding energy of the neutrons within the nucleus is high – some MeV per nucleon – ejection of neutrons from nuclei requires high energies. In practice, one has to bombard suitable nuclei with nucleons or ions and initiate nuclear reactions. As a result, neutrons and other particles will be ejected as free particles.

4.4.1.1 *Radioactive sources* The easiest way to generate neutrons is to use a radioactive source such as radium or polonium that emits α-particles (^{4}He-nuclei) of sufficient energy and to let these particles hit other nuclei. One particularly useful reaction is:

$$^4\mathrm{He} + {}^9\mathrm{Be} \rightarrow {}^{12}\mathrm{C} + \mathrm{n} + 5.7\,\mathrm{MeV},$$

where the kinetic energy released is distributed over the reaction products. One can mix beryllium with polonium or radium powder and obtain an isotropically radiating neutron source. The production rate, however, is quite low, typically of the order of 1×10^7 n/s for a sample of 1 g mass (Dobrzynski and Blinowski, 1994).

4.4.1.2 *Accelerator-based sources* Nuclear reactions can be initiated by accelerating protons or deuterons to a few hundred keV, e.g. by using a van de Graaff generator or even an electrostatic accelerator and bombarding light atoms with them. Two such fusion reactions include (Dobrzynski and Blinowski, 1994, Carlisle, 2003):

$$\begin{aligned} {}^2\mathrm{H} + {}^3\mathrm{H} &\rightarrow {}^4\mathrm{He} + \mathrm{n} + 17.6\,\mathrm{MeV} \\ {}^2\mathrm{H} + {}^9\mathrm{Be} &\rightarrow {}^{10}\mathrm{B} + \mathrm{n} + 4.3\,\mathrm{MeV}. \end{aligned} \tag{4.17}$$

The second reaction is especially suitable for accelerator-based neutron sources.

4.4.1.3 *Reactors* Most neutron sources used for imaging are dedicated research reactors optimized to deliver neutrons. Fission of the nuclear fuel – mostly ^{235}U – takes place when ^{235}U nuclei are hit by slow neutrons. There are many possible reactions of this type, e.g.:

$$\mathrm{n}_{\mathrm{slow}} + {}^{235}\mathrm{U} \rightarrow {}^{144}\mathrm{Ba} + {}^{89}\mathrm{Kr} + 3\mathrm{n}_{\mathrm{fast}} + \gamma + 200\,\mathrm{MeV}.$$

Here, 'slow' means a few meV, 'fast' a few MeV. In order to obtain a chain reaction the fast ejectiles must be thermalized by a moderator – usually light or heavy water, or graphite – to be used as projectiles in the next reaction cycle. Thermalization is a series of scattering events of the fast neutrons with the light nuclei of the moderator that finally lead to an adjustment of neutron energies to the range of the moderator temperature. Research reactors produce up to 5×10^{18} neutrons per second (Carlisle, 2003), with core fluxes up to 2×10^{15} n/(cm^{2}s)

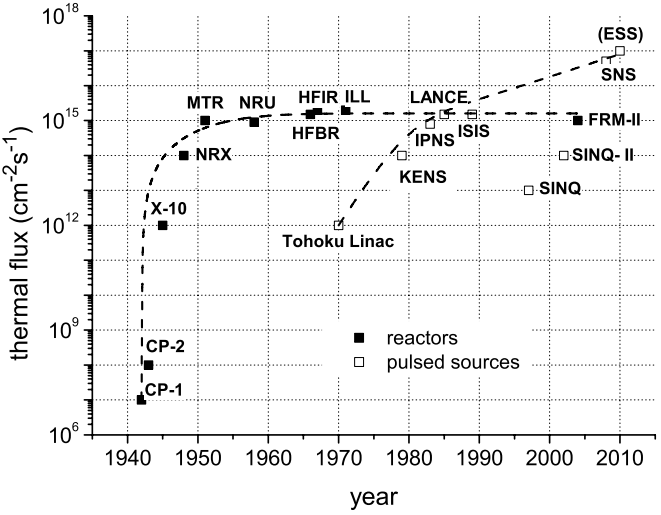


FIG. 4.15. Historical development of the flux of neutron sources. Adapted from Sköld and Price (1987) and Carlisle (2003).

but the important quantity is flux in the reactor channels (beam tubes) leading to the instruments that ranges up to 1.2×10^{11} n/(cm^{2}s) for the most powerful reactors (Dobrzynski and Blinowski, 1994). Appendix A contains a list of reactors that are being used for neutron imaging. It is interesting to note that reactor development has not produced any marked improvement of flux since 1970, in contrast to the development of synchrotron sources, as can be seen by comparing Fig. 4.15 with Fig. 4.8. The cooling problem is so serious and complicated that a further increase seems hardly viable economically and from the viewpoint of safety.

The neutrons coming out of the reactor core are directed to the measurement stations through flight tubes. The extracted beams contain a spectrum of neutrons energies, i.e. can be called 'polychromatic'. The moderation process gives the neutrons an energy in the range corresponding to the temperature of the moderator, but the spectrum is not truly Maxwellian since absorption depends on energy and preferentially removes slower neutrons. Moreover, if the exit channels permit fast neutrons to leave the reactor, some of these high-energy particles can be found too. Actually, some flight tubes have been constructed so as to allow fast fission neutrons to fly out of the reactor and to be used, e.g. for tomography, but usually the beam tubes are oriented tangentially to avoid too many fast neutrons (and γ-rays) in the beam. Using 'cold sources' – tanks containing liquid hydrogen or deuterium – cold neutrons can be produced. These are most conveniently transported to the instruments through neutron guides that consist of

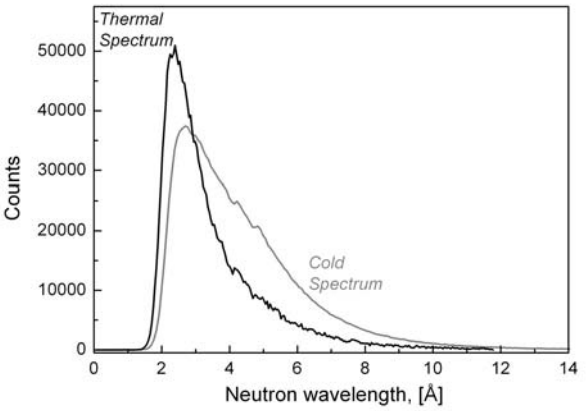


FIG. 4.16. Spectral distribution of neutron energies at the research reactor FRM-I in Munich. Cold and thermal spectra are shown. Courtesy of B. Schillinger.

evacuated glass tubes coated with a reflecting layer such that neutrons are kept in the guide. If such guides are bent with a large radius (>1000 m) high-energy neutrons and γ-rays originating from the reactor core can be eliminated since these fail to be reflected in the outer curve of the guide. This yields neutron beams with a low background that can be very useful for neutron tomography. Typical neutron spectra are shown in Fig. 4.16. The thermal and cold spectra look alike but are shifted in energy corresponding to the temperature of the moderator.

4.4.1.4 *Spallation sources* Spallation is a technology where neutrons are produced by bombarding a target made of a heavy material such as lead, tungsten or ^{238}U with high-energy protons. Proton energies range up to 1 GeV, time-averaged currents are 1.4 mA, and peak currents 50 A during 1 μs (data applies to the new Spallation Neutron Source SNS in Oak Ridge, USA). The proton beam is usually pulsed at repetition rates of a few tens of Hz and the corresponding neutron pulses therefore have a similar time structure. Each proton can produce more than 30 neutrons. Integrated over time the neutron production rate can be as high as 4×10^{16} n/s. Spallation technology has not been brought to an inherent technological limit yet and therefore the future of neutron production for research seems to lie in such sources (see Fig. 4.8).

4.4.1.5 *Preparation of neutron beams* Neutron imaging may require a further preparation of the beam as it comes out of a flight tube or neutron guide (Dobrzynski and Blinowski, 1994).

TABLE 4.1. Neutron energies in various ranges. To convert from one unit to the other use $E = k_B T$, $\lambda = h/\sqrt{2mE}$, $v = \sqrt{2E/m}$. The values listed are a rough estimate only to give an impression of the orders of the magnitude and should not be taken too literally.

range	temperature T (K)	energy E(meV)	wavelength λ(nm)	velocity v (m/s)
'cold'	22	2	0.66	600
'thermal'	300	26	0.18	2200
'epithermal'	3000	260	0.06	7000

- Collimation and focusing: the neutron beam can be made 'more parallel' or deliberately divergent or convergent by using collimators – usually spherical diaphragms – or more elaborate neutron optical components. Some of these techniques will be described in Chapter 13.
- Monochromatization: mechanical velocity selectors, chopper systems, or single crystals can be used to remove all neutrons from a beam that do not have the desired velocity (or wavelength) range.
- Polarization: Polarizing filters, magnetized mirrors or magnetized single-crystal reflectors remove one of the two possible polarization states of a neutron in an external magnetic field. Polarized neutron beams can be used for imaging of magnetic structures.

4.4.2 *Neutron interactions: overview*

Neutrons are massive particles ($m \approx 1.67 \times 10^{-27}\,\mathrm{kg} = 1.008\,\mathrm{amu}$) without any electric charge, but they possess spin 1/2 and a magnetic moment $\mu_n \approx -1.91\mu_N$, where μ_N is the nuclear magneton. In addition, they have a kinetic energy when they hit the sample in an imaging setup. Typical energy ranges are shown in Tab. 4.1. These properties determine the interaction of neutrons with matter. There are two distinct ways of interaction:

- via nuclear interaction with the nucleus of a target atom. These are attractive interactions with a very short range, i.e. of the order of the diameter of the nucleus $\approx 1 \times 10^{-14}\,\mathrm{m}$.
- via magnetic interaction with magnetic moments in the target atom that have a much longer range.

4.4.3 *Nuclear scattering*

4.4.3.1 *Scattering by individual nuclei* Neutrons used for imaging have wavelengths that are 5 orders of magnitude larger than the diameter of the nuclear potential. It is therefore justified to simplify the scattering problem and to describe the nucleus by a simple potential well with a radius R and depth V, both quantities depending on the nucleus considered. The nuclear radius can be approximated by $R = 1.3 \times 10^{-15}(Z+N)^{1/3}\,\mathrm{m}$, where Z is the atomic number and N the

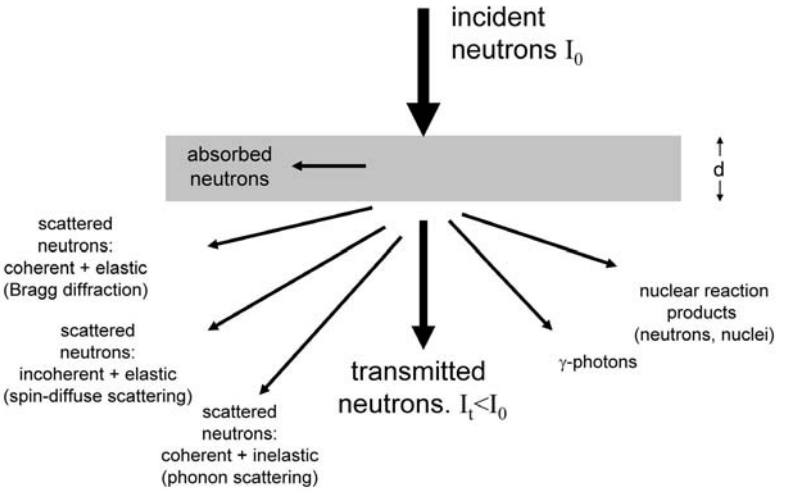


FIG. 4.17. Overview of neutron interactions with matter. The direction of arrows is not directly related to direction of radiation.

number of neutrons in the nucleus. For thermal neutrons $kR = \frac{2\pi}{\lambda}R = 5 \times 10^{-4}$ (for ^{27}Al) and in eqn 4.4 the case of pure s-wave scattering applies. The depth of the potential well V determines the strength of interaction. V depends on the nuclear configuration in a complicated way and is not a smooth function of either the number of protons or neutrons in the nucleus. It has values of the order of MeV, more than six orders of magnitude more than the kinetic energy of the incident neutron. The interaction problem can be further simplified by treating separately the finite probability of inelastic collisions during which energy is transferred from the neutron to the nucleus, e.g. leading to nuclear excitations or reactions.

This situation implies that we can describe the elastic interaction of a single neutron with a nucleus by a single complex function, the scattering amplitude f_k. As k is so small the quantity $b = -\lim_{k\to 0} f_k$ (the nuclear scattering length) allows us to characterize the interaction by a single number quite precisely. From eqn 4.2 and eqn 4.3 it follows that $\sigma_{\text{diff}} = b^2$ and $\sigma_s = 4\pi b^2$. b is mostly positive, but can have negative values in cases where the nuclear potential is shallow, i.e. V is small. b can even be complex when low-lying resonances of the neutron/nucleus system are involved.

4.4.3.2 *Scattering from an ensemble of nuclei* As mentioned in Section 4.1.2 scattering can be coherent and incoherent. Possible reasons are – besides the obvious case that one is dealing with a mixture of different elements – that the element consists of different isotopes that have different scattering properties. This we call *isotope incoherence*. Another reason is the spin dependence of nuclear scattering. It can vary with the relative orientation of nuclear and neutron spin

TABLE 4.2. Scattering properties of thermal neutrons for various solid elements and nuclei given by Rauch and Waschkowski (2003); similar but not identical values are provided by Sears (1987). Value for μ corresponds to solid element and is calculated twice: from σ_{abs} ('absorption coefficient') and from $\sigma_{\mathrm{coh}}+\sigma_{\mathrm{incoh}}+\sigma_{\mathrm{abs}}$ ('attenuation coefficient'), both times using eqn 4.8.

element	b_+	b_-	b_{coh}	b_{incoh}	σ_{coh}	σ_{incoh}	σ_{abs}	μ_{abs}	μ
	1×10^{-15} m				1×10^{-28} m^2			cm^{-1}	
nat. H			−3.74		1.756	80.26	0.33	-	-
^{1}H	10.82	−47.42	−3.74	25.12	1.758	80.27	0.33	-	-
^{2}H	9.53	0.98	6.67	4.03	5.59	2.05	5×10^{-4}	-	-
nat. B			5.3		3.54	1.7	767	105	106
^{27}Al	3.67	3.15	3.45	0.26	1.5	0.008	0.23	0.013	0.1
nat. Si			4.15		2.16	0.004	0.17	0.008	0.12
nat. Fe			9.45		11.22	0.4	2.56	0.21	1.2
nat. Ni			10.3		13.3	5.2	4.5	0.41	2.1
^{58}Ni			14.4	0	26.1	0	4.6	0.42	2.8
^{60}Ni			2.8	0	1	0	2.9	0.26	0.35
nat. Pb			9.4		11.11	0.003	0.17	0.005	0.38

that can be parallel or antiparallel – giving rise to two spin-dependent scattering lengths b_+ and b_- – and causing what is called *spin incoherence*. Whenever these two quantities are different and both nuclear spin directions are represented in a sample – which is usually the case – scattering will be partially incoherent.

Under the assumptions that nuclear spins and isotopes are randomly distributed in a solid, one can break up both scattering length and total cross-section into two parts: a coherent and an incoherent part:

$$b = b_{\mathrm{coh}} + b_{\mathrm{incoh}}; \quad \sigma = \sigma_{\mathrm{coh}} + \sigma_{\mathrm{incoh}}.$$

Let us consider natural hydrogen that consists of 99.985% ^{1}H (spin 1/2) and 0.015% ^{2}H (spin 1), each one having two possible spin states, i.e. we can specify four scattering lengths (2^{nd} and 3^{rd} column of Tab. 4.2). The equations (Dobrzynski and Blinowski, 1994)

$$b_{\mathrm{coh}} = \sum_i p_i b_i, \quad b_{\mathrm{incoh}} = \sqrt{\sum_i p_i b_i^2 - b_{\mathrm{coh}}^2},$$

where the index i sums up the different nuclear spin orientations and/or isotopes:

$$p_{i+} = x_i \frac{I+1}{2I+1}, \quad p_{i-} = x_i \frac{I}{2I+1},$$

where I is the spin quantum number and x the relative abundance of the particular isotope. This allows us to determine coherent and incoherent scattering lengths for each of the two isotopes separately, or for natural hydrogen (4th and

5th columns in Tab. 4.2). The corresponding cross-sections are calculated by $\sigma_{\text{(in)coh}} = 4\pi b^2_{\text{(in)coh}}$ and are given in the 6th and 7th column. Table 4.2 summarizes data for some more selected elements. Hydrogen (^{1}H) is an example for strong spin incoherence: more than 97% of the total scattering cross-section is due to spin-incoherent scattering. The two nickel isotopes displayed do not have a nuclear spin and therefore do not scatter incoherently, but as natural nickel contains many isotopes it does have an isotope-incoherent scattering component.

4.4.4 *Magnetic interactions*

As the neutron carries a magnetic moment it interacts with magnetic fields. Interactions with free fields will be briefly described in Section 14.3. In magnetic samples the individual ions contain unpaired electrons that give rise to a macroscopic magnetization. Neutrons interact with these unpaired electrons distributed around the ion with a density $\rho(\boldsymbol{r})$. One cannot expect a description of this scattering as simple as for scattering by the point-like nuclei. The differential cross-section can be written as

$$\sigma_{\text{diff}}^{\text{mag}} \propto \{[\mathcal{F}\rho](\Delta\boldsymbol{k})\}^2 ,$$

where $\mathcal{F}$ is the Fourier transform operator and $\Delta\boldsymbol{k}$ the scattering vector. The term in the curly brackets is called the *magnetic form factor* and can be either calculated, if the electronic structure of the material is known, or measured by neutron scattering.

4.4.5 *Nuclear absorption*

Whenever an incoming neutron reacts with a nucleus it 'disappears' from the viewpoint of the detector. There are many possible reactions, all leading to the emission of either secondary neutrons, protons, α-particles, photons or fission products. Absorption is characterized by the cross-section σ_{abs} that ranges from 0 for ^{3}H and ^{4}He to 254 kbarn for ^{156}Gd (Sears, 1987). Table 4.2 contains data for some other elements.

4.4.6 *Total attenuation*

Both elastic and incoherent scattering will remove neutrons from the incident trajectory and reduce the flux behind the nucleus. Therefore, it weakens the beam which makes the mechanism as important for imaging as the inelastic processes such as neutron absorptions. Scattering often dominates the attenuation coefficient, as can be seen in Tab. 4.2 by comparing μ_{abs} and μ. It depends somewhat on the experimental setup how scattering processes contribute to attenuation. For a point detector placed far away from the sample all processes including scattering and absorption will weaken the transmitted beam, an area detector close to the sample will still capture some of the scattered neutrons, see Fig. 4.3. Therefore, Tab. 4.2 provides two linear attenuation coefficients, one including just nuclear absorption, the other including all scattering processes. The measured absorption coefficient will be between these two values, presumably closer to the latter one.

TABLE 4.3. Main features of the interaction between X-rays, electrons and neutrons with solid matter.

property	X-rays (100 keV)	electrons (100 keV)	neutrons (thermal)
interaction mechanism	electromagnetic with electrons	Coulomb with electrons+nuclei	nuclear: nuclei, magnetic: electrons
interaction strength	medium	very strong	weak
penetration depth	millimetres	nanometres	centimetres
function $\mu(Z)$	regular	regular	irregular
sensitivity to low Z	no	no	yes
preparation effort	small	high	very small
sample thickness	1 to 10 mm	10 to 100 nm	10 to 500 mm

4.4.7 *Refraction*

Refraction is governed by the deviation δ of the real part of the refractive index from unity and the small imaginary part β. One can approximately write (Werner and Klein, 1987):

$$\delta = \frac{N_V \lambda^2}{2\pi} b_{\mathrm{coh}}; \quad \beta = \frac{N_V \lambda}{4\pi}(\sigma_{\mathrm{incoh}} + \sigma_{\mathrm{abs}}), \tag{4.18}$$

and thus relate the complex index of refraction to the coherent scattering length and the total cross-section. Values for the chemical elements and thermal neutrons are given in Fig. 14.1, Most values are positive, the largest being that of ^{58}Ni that is therefore a good material for the reflective coating of neutron guides since it has the highest angle of acceptance. Negative values of δ are rare, important examples being Ti, V and Mn.

4.5 Comparison of interactions

We have seen that X-rays, electrons and neutrons interact in quite different ways with matter. Table 4.3 gives an overview of the main differences. The differences in interaction strength explain why electrons are used for imaging the thinnest samples with the highest resolution, whereas neutrons image the thickest specimens at low resolutions, and X-rays take an intermediate position. The specific interaction of neutrons makes them useful for imaging light elements and magnetic structures.

This point is further highlighted by a direct comparison between the attenuation coefficients of the elements for X-rays and thermal neutrons, as shown in Fig. 4.18. The very high absorption of neutrons by light elements such as hydrogen and boron explains why they are used whenever an imaging application focuses on such elements. On the other hand, neutrons are not strongly absorbed

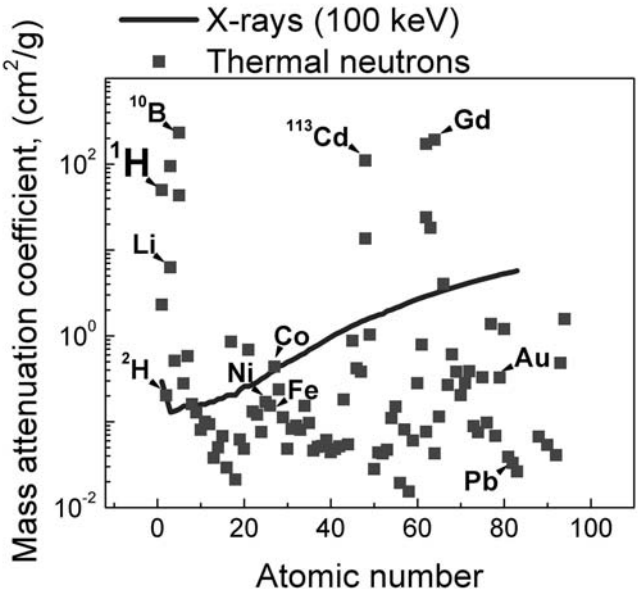


FIG. 4.18. Comparison of absorption coefficients between various elements for thermal neutrons and 100-kV X-ray radiation (Kardjilov *et al.*, 2006).

by heavy elements such as lead or gold that block X-rays efficiently. Aluminium has a very low absorption coefficient, which makes it a good first choice for windows, sample holders, *etc.* The usefulness of these properties will become evident in Chapter 13 where applications for neutron tomography are discussed.

4.6 Summary

X-rays, neutrons, and electrons are very different in the way they are produced as well as in the way they interact with matter. Each type of radiation has its own characteristics that determines the field of use of these rays for imaging. Electrons are most suitable for very small samples and very high resolutions, neutrons provide a tool for imaging thick samples at low resolutions. X-rays fill the intermediate range. More details about the actual use of these rays will be found in the following chapters.

4.7 References

Carlisle, C.J. (2003). The production of neutrons. In *Neutron data booklet* (ed. A.-J. Dianoux and G. Lander), pp. 3.1–1. ocp science, Philadelphia. `http://www.ill.fr/pages/menu_g/docs/NeutronDataBooklet.pdf`.

Cohen-Tannoudji, C., Diu, B., and Laloë, F. (1977). *Quantum mechanics*, Volume 2. John Wiley, New York.

Dobrzynski, L. and Blinowski, K. (1994). *Neutrons and solid state physics*. Ellis Horwood, New York.

Duke, P.J. (2000). *Synchrotron radiation*. Oxford University Press, Oxford.

Ebel, H., Svagera, R., Ebel, M.F., Shaltout, A., and Hubbell, J.H. (2003). Numerical description of photoelectric absorption coefficients for fundamental parameter programs. *X-ray Spectrometry*, **32**, 442.

Foderaro, A. (1971). *Neutron interaction theory*. MIT Press, Cambridge.

Gullikson, E.M. (2004). X-ray interactions with matter. `http://www-cxro.lbl.gov/optical_constants/`. [Online; accessed 11 April 2007].

Hubbell, J.H. and Selzer, S.M. (1996). Tables of X-ray mass attenuation coefficients and mass energy-absorption coefficients. `http://www.physics.nist.gov/PhysRefData/XrayMassCoef/cover.html`. [Online; accessed 11 April 2007].

Jackson, D.F. and Hawkes, D.J. (1981). X-ray attenuation coefficients of elements and mixtures. *Physics Reports*, **70**, 169.

Jackson, J.D. (1975). *Classical electrodynamics* (2^{nd} edn). John Wiley, New York.

Jenkins, R. and Snyder, R.L. (1996). *Introduction to X-ray powder diffractometry*. John Wiley, New York.

Kardjilov, N., Fiori, F., Giunta, G., Hilger, A., Rustichelli, F., Strobl, M., Banhart, J., and Triolo, R. (2006). Neutron tomography for archaeological investigations. *Journal of Neutron Research*, **14**, 29.

Lifshin, E. (1999). *X-ray characterisation of materials*. Wiley-VCH, Weinheim.

McDaniel, E.W. (1986). *Atomic collisions*. John Wiley, New York.

Paganin, D.M. (2006). *Coherent X-ray optics*. Oxford University Press, Oxford.

Rauch, H. and Waschkowski, W. (2003). Neutron scattering lengths. In *Neutron data booklet* (ed. A.-J. Dianoux and G. Lander), pp. 1.1–1. ocp science, Philadelphia. `http://www.ill.fr/pages/menu_g/docs/NeutronDataBooklet.pdf`.

Reimer, L. (1997). *Transmission electron microscopy — Physics of image formation and microanalysis* (4^{th} edn). Springer, Berlin.

Sanchez del Rio, M. and Dejus, R.J. (2004). X-ray oriented programs (XOP), Version 2.11. `http://www.esrf.eu/computing/scientific/xop2.1/`. [Online; accessed 12 February 2007].

Sears, V.F. (1987). Neutron scattering lengths and cross sections. In *Neutron scattering* (ed. K. Sköld and D. Price), pp. 521. Academic Press, Orlando.

Sköld, K. and Price, D.L. (1987). Introduction to neutron scattering. In *Neutron scattering* (ed. K. Sköld and D. Price), pp. 1. Academic Press, Orlando.

Werner, S.A. and Klein, A.G. (1987). Neutron optics. In *Neutron scattering* (ed. K. Sköld and D. Price), pp. 259. Academic Press, Orlando.

Wille, K. (2005). *The physics of accelerators*. Oxford University Press, Oxford.

Williams, D.B. and Carter, C.B. (1996). *Transmission electron microscopy — A textbook for materials science*. Plenum Press, New York.

Wu, T.-Y. and Ohmura, T. (1962). *Quantum theory of scattering*. Prentice-Hall, London.

PART II

SYNCHROTRON X-RAY TOMOGRAPHY

Synchrotron radiation provides the largest number of experimental options for tomography. One reason is the extraordinary brilliance of third-generation synchrotrons, many orders of magnitude higher than that of conventional X-ray tubes (see Fig. 4.8). This high brilliance combined with the continuous spectrum of the synchrotron radiation allows us to either select a suitable narrow energy band using a monochromator or to apply a white beam, depending on the purpose. Further unique properties of synchrotron radiation are the high collimation, the low emittance and the high level of polarization. One consequence of the high brilliance is the partial coherence of the X-ray beam that allows us to exploit the phase of the radiation.

This part of the book is organized as follows. In Chapter 5 we start with *non-magnifying absorption tomography*. Here, the image is formed by the different attenuation coefficients of the constituents in a sample. The parallel synchrotron beam projects an absorption map on a scintillator where it is converted to light, magnified and detected.

Next, tomography exploiting phase effects is described in Chapter 6. *Phase contrast* and *holotomography* measure the change in X-ray phase at the interface between two materials inside a sample and construct a three-dimensional image. These methods can be successful in cases where the difference in absorption contrast is too small for reliable imaging.

After this we turn back to absorption tomography in Chapter 7 but consider setups where the image is magnified using X-ray optics before conversion to light. This allows us to overcome the resolution limits of the scintillator and to achieve resolutions in the range of tens of nanometres instead of hundreds of nanometres.

Another development is that of *scanning tomography*, described in Chapter 8. Here, the sample is probed point- or linewise by very collimated or focused X-ray beams instead of illuminating all parts of a sample simultaneously. This allows us to use signals other than absorption or phase changes for constructing tomographic images, such as X-ray fluorescence, X-ray absorption spectroscopy or X-ray small-angle scattering.

A completely different kind of information is provided by the so-called *3D X-ray diffraction* method described in Chapter 9. Instead of the chemical information one is interested primarily in the crystallographical orientation of the

constituents of a sample – in three dimensions.

Chapter 10 gives information on one of the key issues of success of synchrotron tomography, the detection system. Scintillator and camera systems are reviewed and classified according to their use.

5

SYNCHROTRON X-RAY ABSORPTION TOMOGRAPHY

Astrid Haibel

The construction of third-generation synchrotron sources combined with the tremendous improvement of detector technology, data storage and processing capabilities since the 1990s has led to a boost of high-end synchrotron tomography in materials research with a wide range of different applications, e.g. the visualization and quantitative analysis of differently absorbing phases, microporosities, cracks, precipitates or grains in a specimen. In this chapter, the tomographic technique is described that exploits the different linear attenuation coefficients of the chemical elements present in a sample or, alternatively, the density differences for a given element. Thus, the image-forming signal is *absorption contrast*. In the following chapters other ways of forming images will be described.

5.1 Synchrotron radiation for tomography

The intensity I of a monochromatic X-ray beam after passing in the x-direction through a sample of thickness d characterized by a spatial distribution of the linear X-ray attenuation coefficient $\mu(x, y, z)$ is given by a generalization of eqn 4.11:

$$I(y, z) = I_0 \exp\left(-\int_0^d \mu(x, y, z)dx\right), \tag{5.1}$$

where I_0 is the intensity of the incoming beam. The absorption coefficient μ includes the effects of photoelectric attenuation, incoherent Compton scattering, and coherent Rayleigh scattering, as outlined in Section 4.2. As a result of this material-dependent absorption, synchrotron tomography with monochromatic radiation allows us to resolve and visualize the different components of a sample with a spatial resolution down to one micrometre in short times, i.e. in the range of minutes. In practice, this can hold even if the attenuation coefficients of the components are similar because there is always a contribution of edge enhancement in the signal, see Section 5.4.3). Due to the monochromaticity and parallelity of the synchrotron beam the images are virtually free of the artefacts known from laboratory tomography, such as e.g. beam hardening or excessive noise. In addition, it is possible to adjust the X-ray energy to the absorption-edge of one of the elements present in the sample, thus allowing for contrast enhancements by calculating the difference of the images taken below and above the absorption edge of one component. This method is called *absorption-edge*

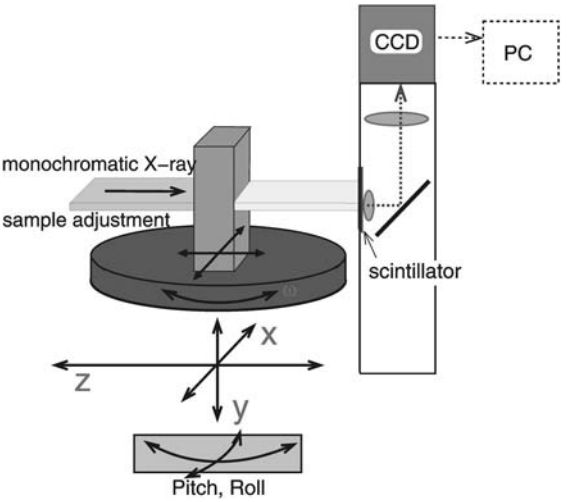


FIG. 5.1. Setup for synchrotron X-ray tomography. For a precise adjustment of the samples perpendicular to the beam several translation and tilt tables are necessary.

tomography (Bonse *et al.*, 1989, Thurner *et al.*, 2004). An example will be given in Section 7.1.4.3.

5.2 Principles of synchrotron X-ray tomography

Synchrotron absorption tomography is based on the detection of X-ray radiation attenuated by an object. The X-rays pass through a sample mounted on a rotation stage as sketched in Fig. 5.1. The attenuated X-ray beam is recorded by a detection system, in which it is converted into visible light by a scintillator. The resulting visible image is then projected onto a CCD chip, from which a digital image is read out into a computer where it is stored and further processed. Usually a 2D detector is used, providing faster image acquisition than line or even point detectors, which have to be used in a raster mode.

The first step in X-ray imaging is the preparation of a suitable beam. The radiation provided by the synchrotron can be used for tomography without any further treatment. These so-called *white beams* are polychromatic and very intense. They give rise to beam hardening, an artefact known from X-ray tubes. Therefore, their use is limited to a few cases, e.g. to fast tomography (Mathiesen and Arnberg, 2006), and usually monochromatic beams are preferred. For monochromatizing the incoming X-ray radiation single crystals as well as multilayer structures can be used. An incoming beam impinges on the crystal surface under a specific angle θ, see Fig. 4.2, and only a certain wavelength that interferes constructively will be reflected according to Bragg's equation, see eqn 4.6. By varying the angle of incidence of the beam, i.e. by changing the inclination of the

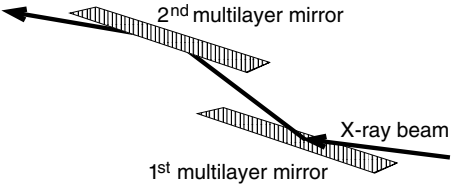


FIG. 5.2. Sketch of a double-multilayer monochromator.

monochromator crystals, the reflected wavelength changes. For monochromatizing synchrotron radiation two such crystals are usually combined as sketched in Fig. 5.2. Single-crystal monochromators select the wavelength – or energy $E = 2\pi\hbar c/\lambda$ – very precisely, i.e. they have an energy acceptance of the order of $\Delta E/\ E = 1 \times 10^{-4}$, limited by the mosaicity of the crystal. Correspondingly, the intensity of the monochromatic beam is much lower than that of the white beam.

To avoid an excessive loss of flux, in most tomographic applications multilayer monochromators are used. Multilayer monochromators consist of bilayers (e.g. of Si and W) sputter-deposited on silicon substrates, the bilayer spacing d being about 1.5 to 4 nm. Such multilayer monochromators have a much lower energy resolution ($\Delta E/E = 1 \times 10^{-2}$), but for many applications this resolution is adequate. The trade-off of higher flux for lower resolution is often a clear benefit. A certain drawback is that for high energies the angle of incidence for the multilayer monochromator becomes very small, which leads to a vertically small beam and possibly to inhomogeneous illumination of the multilayer mirrors whenever the angle is in the range of a few mrad. Hence, for high-energy applications *Laue transmission monochromators* were designed and constructed, see e.g. Grübel *et al.* 1996. Such monochromators consist of two vertically arranged thin single crystals that lead to interference effects when a beam is transmitted. The constructive interference of the transmitted beam is used to select a wavelength λ. By bending the crystals they deliver over ten times more flux than a flat crystal monochromator at high energies, with an increase in bandwidth to $\Delta E/E \simeq 1 \times 10^{-3}$. The beam is usually reduced to the required cross-section using a set of adjustable motor-driven slits.

After being attenuated by the sample the beam hits the scintillator where an image is created in the range of visible light. The choice of the scintillator material depends on the appropriate energy and the required spatial resolution. In order to obtain spatial resolutions below 5 µm, single-crystal scintillators are required. The image is magnified by an optical microscope system and then captured by a CCD camera. Scintillators and CCD cameras are described in detail in Chapter 10.

To obtain highly resolved, artefact-free tomograms the rotation axis of the sample and the camera have to be aligned accurately. This requires very precise mechanical elements and a good air conditioning to minimize thermal expansion effects. Unwanted movements of the axes (tilts, movements, drifts) should be

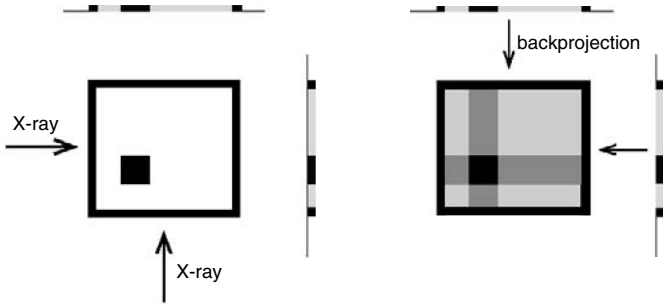


FIG. 5.3. Sketch of the backprojection method. (a) Two radiographic projections of the object are taken for two angular positions (0° and 90°. (b) From these two radiographic projections the image information can be reconstructed by backprojection.

smaller than the pixel resolution of the measurement.

The layout of the beamline used for obtaining most results presented in this chapter is described by Görner *et al.* (2001).

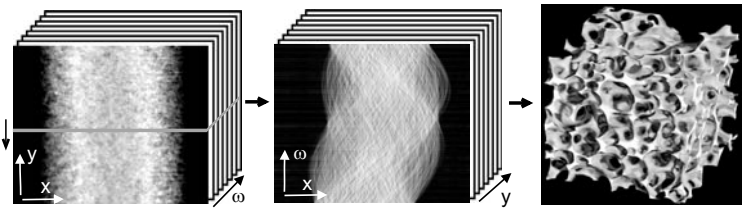
5.3 Data reconstruction

For each tomogram, radiographic projections are taken for many different angular positions. This results in a set of projections that are used for reconstructing a complete 3D image representing the local mass distribution in the sample. Mathematically, this is equivalent to inverting the Radon transform of the projection data. In practice, an explicit and computationally efficient, discretized version of the inverse Radon transform is used, called a *filtered backprojection algorithm* (Herman, 1980). A mathematical formulation of this algorithm is given in Chapter 2. Here, we shall just give a short description emphasizing the practical procedure using Fig. 5.3. The following steps are needed:

- The generation of sinograms, Fig. 5.4(b), from a set of N radiographic projections, Fig. 5.4(a), which is essentially a change of data representation,
- the filtering,
- the backprojection of the data: this procedure produces 3D data sets as displayed in Fig. 5.4(c).

One slice of a tomographic image is represented by one line in the radiograms for the entire set of projections, see Fig. 5.4(a). Such a set of lines, corresponding to one tomographic slice, is called a *sinogram* or Radon transform, see Fig. 5.4(b).

To filter out blurring that occurs as a result of backprojection, the data has to be treated by a filter function. The most frequently used filters are the truncated *Ram–Lak filter*, ramp filters, or other filter functions (Kak and Slaney, 1987, Horbelt *et al.*, 2002). In all cases a parallel-beam geometry is assumed. The number of projections has to be sufficiently high to ensure that the maximum possible



(a) (b) (c)

FIG. 5.4. Different stages of tomographic data reconstriction. (a) N projections each for a different angle stacked behind each other, (b) sinograms generated from a set of radiograms, (c) tomographic image after filtered back projection.

spatial resolution of the reconstructed image given by the optical components and the detector can be achieved (see Section 2.3.1).

5.4 Image artefacts

5.4.1 *Ring artefacts*

In synchrotron X-ray tomography *ring artefacts* are caused e.g. by imperfect detector elements as well as by defects or impurities on the scintillator crystals. They appear as a number of concentric rings centred around the rotation axis and superimposed on the tomographic image. As the grey levels in the reconstructed images are deteriorated by these ring artefacts, quantitative analysis of the image – e.g. segmentation – is complicated. Moreover, post-processing such as noise reduction or image segmentation does not yield satisfactory results in the presence of such artefacts. Figure 5.5(a) shows a tomographic image of a nickel-based superalloy containing a large number of ring artefacts (Link *et al.*, 2006). An attempt to separate the pores present in this sample from the surrounding matrix material by thresholding is difficult in the presence of these ring artefacts that have the same grey values as the pores themselves, see Fig. 5.5(b).

A common approach to reduce ring artefacts is known as *flat-field correction*. Thereby, images of the background without the sample are taken before and after acquisition of the actual tomographic projections and, if possible, in certain intervals also during data acquisition. The resulting flat-field includes the effects of the non-uniform response of the CCD camera, the non-uniform response of the scintillator screen, and the non-uniformities in the incident X-ray beam. By dividing the images by these flat-field images the artefacts are largely removed. However, ring artefacts will not be removed completely by this flat-field correction whenever some camera elements or parts of the scintillator have non-linear response functions. Alternatively, the effect of non-uniform response of different detector elements can be compensated by wobbling the sample or the detector during acquisition in defined horizontal and vertical steps (Davis and Elliott, 1997). As a consequence, the characteristics of the individual detector elements are averaged with neighbouring elements that eventually leads to

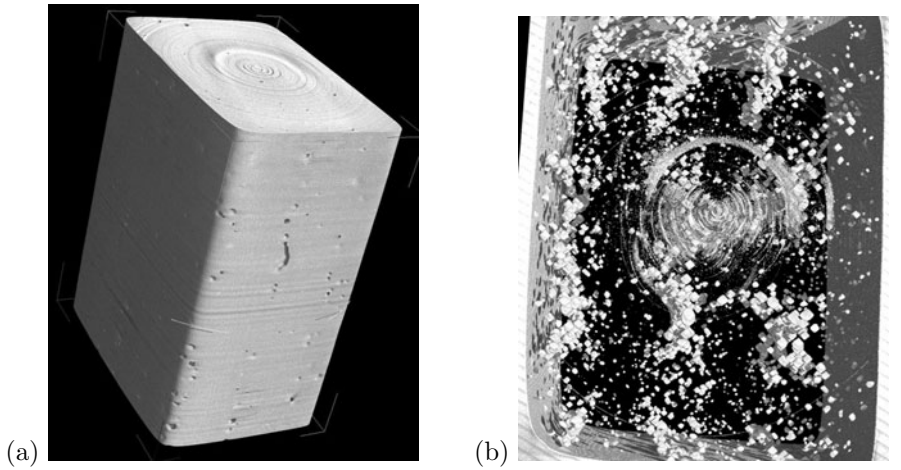


FIG. 5.5. Ring artefacts in a tomographic image (Link *et al.*, 2006). (a) The object is a needle made of a nickel-based superalloy. (b) Separation of the creep pores in the material, superimposed by ring artefacts. Sample width is ≈0.6 mm.

reduced ring artefacts. Finally, ring artefacts can also be reduced by different image-processing algorithms, e.g. size and shape filtering or sinogram processing (Sijbers and Postnov, 2004, Raven, 1998, Antoine *et al.*, 2002, Boin and Haibel, 2006).

5.4.2 *Image noise*

The quality of tomographic images can be degraded by image noise resulting from an unfavourable conversion rate of X-ray photons to visible light especially for higher X-ray energies, see Fig. 5.6(a). Median filters can be used to reduce such noise in the images. The use of median filtering in the reconstruction process allows for both noise reduction and reduction of edge enhancement, while preserving details of interest in the image (Marion, 1991). However, spatial resolution is reduced. A median filter inspects each pixel of an image, finds the surrounding pixels of it (e.g. the 3×3 neighbourhood), calculates its median, and replaces the pixel with the median of those values. By defining a threshold it is possible to replace only pixels with deviations above this threshold. Median filtering is especially useful when the noise occurs as isolated pixels with values that deviate markedly from those of other pixels, e.g. when high-energy photons cause spikes on the CCD camera.

5.4.3 *Edge artefacts*

Phase contrast occurring at the interface of two parts of an object with a different refraction index leads to edge enhancement in an image (Cloetens *et al.*, 1997, Raven, 1998). This effect is the basis of phase-contrast imaging methods (see next chapter) but is considered an adverse artefact for quantitative absorption

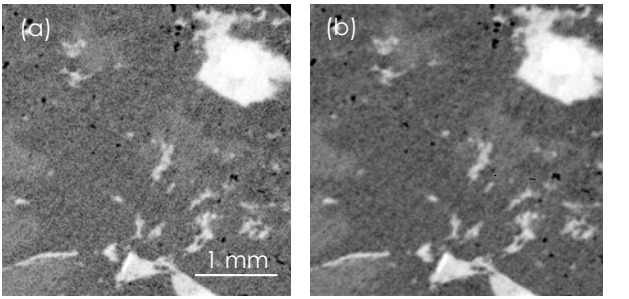


FIG. 5.6. 2D slice of a tomogram of granite rock. (a) Image noise. (b) Median–filtered image.

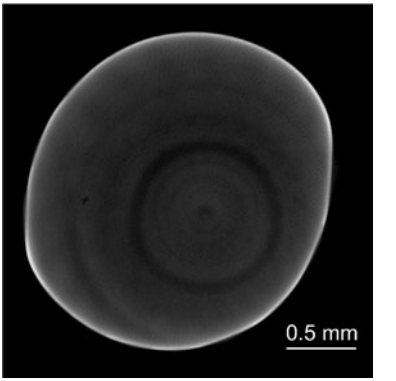


FIG. 5.7. Tomographic slice of a homogeneous iron sample showing a strong edge enhancement. X-ray energy was 60 keV.

tomography. These artefacts appear as aliasing patterns and as overshoots in the areas of sharp density transitions, e.g. from metal to air, see Fig. 5.7.

Due to edge enhancement a precise pre-processing of the data, e.g. the separation of the various components of a sample is difficult. To avoid this artefact, the distance between the sample and the detector has to be as small as possible. This reduces the angle of beam spread and a corresponding widening of the diffracted patterns. In practice, most high-resolution images contain some phase contrast as the distance between the sample and the detector is not exactly zero.

5.4.4 *Motion artefacts*

Motion artefacts occur either due to the movement of the sample during the measurement, e.g. tilting, shaking, or shrinking (see Fig. 5.8), or due to the imprecise movement of the step motors during data acquisition. Whereas the sample movement can be hardly corrected if it is inherent to the system investigated – e.g. when liquids are studied – the correction of the imprecise motion of the motors is possible.

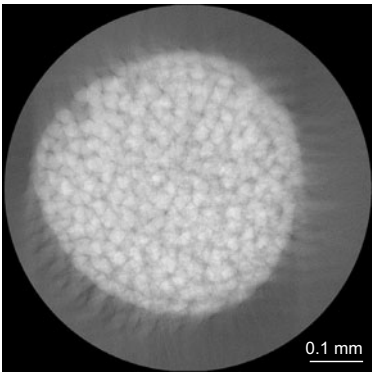


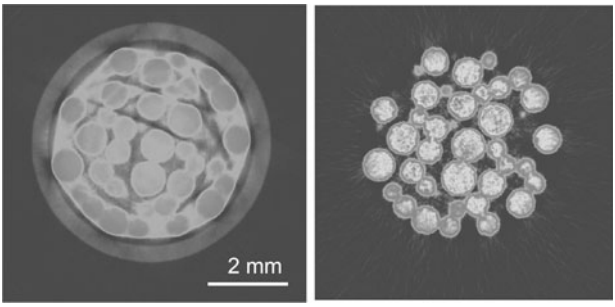
FIG. 5.8. 2D slice of a tomographic image of copper particles with motion artefacts.

FIG. 5.9. Deviations in the sinogram due to inexact translation motor movements during the measurement.

For taking flat-field images, the samples have to moved out of and into the beam during the measurement, whereas for minimizing ring artefacts the samples move in small, defined horizontal and vertical steps during acquisition that have to be counted back precisely after the measurement. Both require very exact motor positioning. Any major deviation will lead to artefacts. Such artefacts can be detected in sinograms as deviations from the ideal sinusoidal profile, see Fig. 5.9. Thus, these discrepancies can be corrected by a translational displacement of each sinogram line to the ideal sine surve (see e.g. Lu and Mackie 2002).

5.4.5 *Beam hardening*

Reconstruction algorithms are based on the assumption that the attenuation at a position of the sample is independent of the path on which the X-rays have reached this position. This is true only for monochromatic X-ray radiation. For polychromatic, e.g. white X-ray radiation, the X-ray spectrum is changed – it becomes 'harder', i.e. the average energy is shifted to higher values. This means that low-energy photons are attenuated more strongly than high-energy photons during propagation through a sample and the spectral properties of the transmitted X-ray beam change in addition to its intensity. This effect is called *beam hardening*. Figure 5.10(a) shows a tomographic slice of sintered copper particles containing such artefacts. It appears that inner regions are denser than outer areas (see Section 2.3.3 for a theoretical treatment of this effect). A well-



(a) (b)

FIG. 5.10. Tomographic slices of spherical sintered copper particles containing beam-hardening artefacts from the X-ray tube used; (a) without correction and (b) with correction (Noethe *et al.*, 2003).

known method to correct such artefacts is to measure a suitable reference object, to find a correction function, and to use it to transform the polychromatic to monochromatic radiographic images (*linearization technique*), see for example (de Casteele *et al.*, 2004).

5.4.6 *Metal artefacts*

These artefacts arise in tomographic images when highly absorbing objects in the samples avoid or reduce strongly the transmission of the X-ray beam along the beam trajectories in which the objects are located. Due to the limited dynamic range of the detector system the information on other objects in this trajectory is lost. Figure 5.11 shows a tomographic slice and a typical star-shaped metal artefact. These artefacts become visible as very bright sinusoidal lines in the corresponding sinograms. To compensate these artefacts the sharp edges of the lines have to be attenuated by an adequate algorithm, e.g. by linear interpolation, functional adaption, *etc.* (Haimerl, 2003, Boin, 2005).

5.4.7 *Centring errors of the rotation axis*

A deviation of the rotation axis during the measurements creates semi-circular artefacts for a 180° tomographic scan, as shown in Fig. 5.12(a). For a 360° scan the artefact becomes a full circle. These artefacts can be easily corrected by an adjustment of the parameter of the flawed rotation axis during reconstruction. Figure 5.12 shows the result of this correction (see e.g. European Synchrotron Radiation Facility 2007).

5.5 Applications and 3D image analysis

5.5.1 *Metal-foam stabilization by silicon carbide particles*

Solid metallic foams are produced by dispersing a large number of gas bubbles into a metallic melt, allowing them to arrange, and freezing the structure by solidification (Banhart, 2000). To guarantee good foam quality, sufficient film

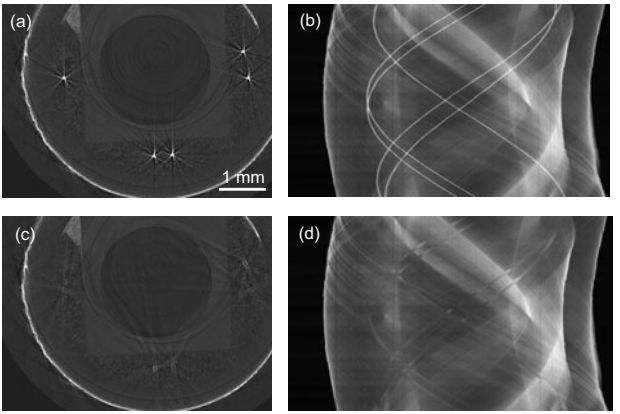


FIG. 5.11. (a) Tomographic slice with metal artefacts. (b) Sinogram of the data. (c) Corrected image. (d) Sinogram without the bright metal-artefact lines (Boin, 2005).

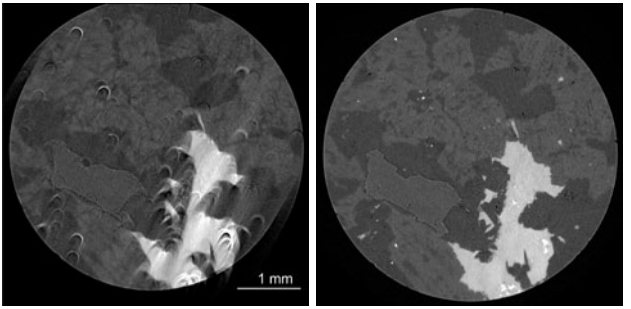


FIG. 5.12. Tomographic slice of a granite rock sample with a diameter of 5 mm. (a) The rotation axis of the image deviates from the correct axis by 30 pixels. (b) Corrected image.

stability during the foaming process is necessary. In contrast to aqueous foams that are stabilized by surface active agents, the liquid state of metallic foams is stabilized by admixing small non-soluble particles. Although this has been common knowledge for many years the way in which these solid constituents act in a liquid metal foam is still in dispute.

Synchrotron tomography as a non-destructive imaging technique permits us to analyse the structure of such foams even in the liquid state. The three-dimensional distribution of the foam-stabilizing particles before and after foaming can be determined with a spatial resolution of some micrometres and possible particle rearrangements during foaming and solidification can be detected (Haibel *et al.*, 2006).

A model that has recently become popular in metal-foam literature (Körner

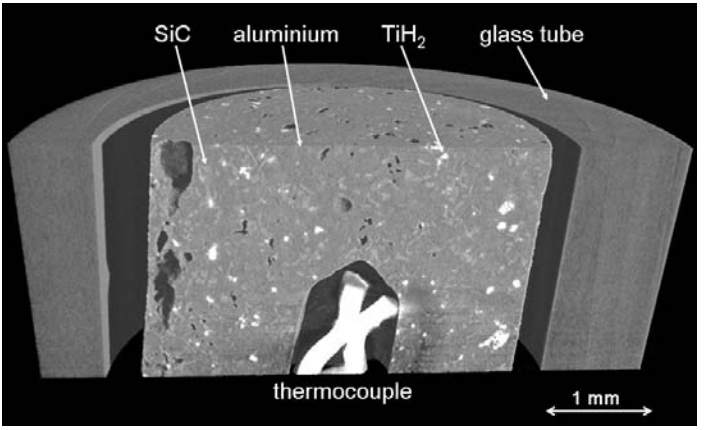


FIG. 5.13. Tomogram of precursor inside a glass tube prior to foaming.

et al., 2005, Wübben and Odenbach, 2005) assumes that opposing particle-covered liquid/gas interfaces are mechanically connected by particles bridges across films, thus providing repulsive mechanical forces keeping the two liquid/gas interfaces apart (Kaptay, 2004). A further approach postulates that the solid particles in the liquid increase viscosity so much that the melt is largely immobilized (Banhart, 2000).

To analyse the stabilization mechanism an aluminium alloy foam Al-10 wt.%Si-1 wt.%Mg containing 10 vol.% silicon carbide (SiC, mean particle size 13 µm) was investigated. A precursor material was manufactured by admixing 0.5 wt.% titanium hydride powder (TiH_2, mean particle size 25 µm) acting as blowing agent to a molten metal matrix composite after which the mixture was solidified. For foaming, the precursor samples were heated up to about 570 °C. At this temperature hydrogen gas is released into the melt and blows the metal foam. As the absorption coefficients of the alloy and SiC particles are very similar ($\mu_{\text{alloy}} = 5.0\,\text{cm}^{-1}$, $\mu_{\text{SiC}} = 5.7\,\text{cm}^{-1}$ at $E = 25\,\text{keV}$) diffraction effects at the metal/particle boundaries contribute significantly to contrast and allow us to distinguish between the two components, see Chapter 6. Figure 5.13 shows a tomogram of a foam precursor containing homogeneously distributed SiC particles. The precursor is situated inside a glass tube and is in contact with a thermocouple allowing for a precise temperature control during foaming.

Figure 5.14(a) shows a plane extracted from the 3D tomographic data set after foaming. The foaming process obviously leads to a redistribution of particles. SiC particles are now found to decorate the surfaces of the cell walls. The magnified inset shows some individual particles in more detail. To inspect the connectivity of the SiC particles, i.e. to check whether there are bridges or even networks in the metallic matrix, a connectivity analysis was carried out. For this, objects consisting of connected SiC particles were identified by an algorithm – described in Section 3.4.2 – which isolates all topologically connected regions of

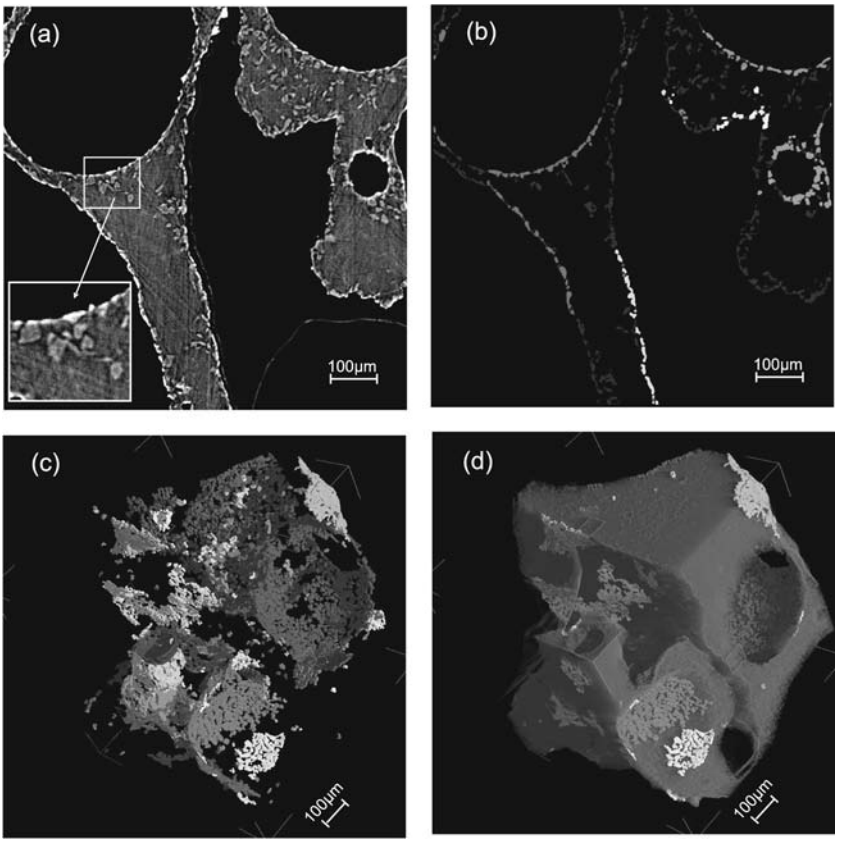


FIG. 5.14. Images of solid aluminium foam: (a) tomographic image of plane through foam, (b) same plane image processed, large groups of connected particles are colour-coded, small groups are shown in grey, (c) 3D visualization of connected particle clusters, (d) same as (e) but showing aluminium matrix, (pixel size 1.5 µm, box length 600 pixels). Reprinted with permission from Haibel *et al.* (2006). Copyright 2006, American Institute of Physics. This image is reproduced in **colour** in the central part of the book.

an image and labels them with a number or, for the purpose of visualization, by a colour. The result is displayed in Fig. 5.14(b). The particles on one surface are rather densely packed, but the layers covering opposite interfaces are largely unconnected and do not form bridges between two particle layers that therefore appear in different colours. Figures 5.14(c,d) display colour-coded 3D visualizations of the same sample and confirm the interpretation of Fig. 5.14(b).

The analysis shows that particles do not interact across liquid-metal films. Instead, liquid-metal foams must be stabilized by interactions between particles attached to one side of a film. In aqueous foams such interactions can be both

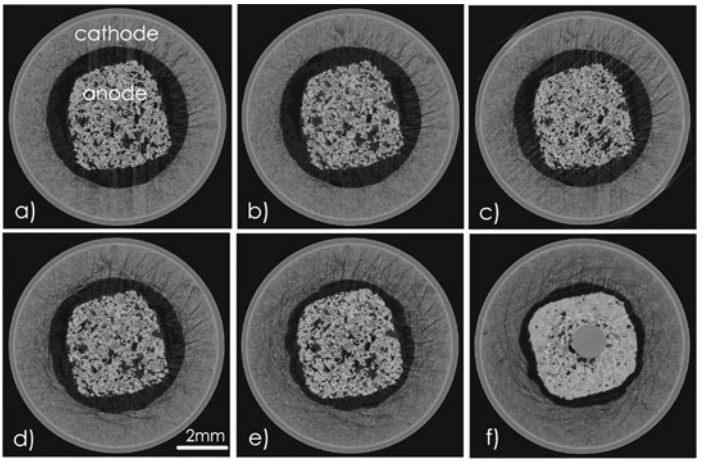


FIG. 5.15. Different stages of discharge of a Varta alkaline battery. (a) Fully charged state, voltage $U = 1.59\,\mathrm{V}$. (b) State after 15 min of discharge $U = 1.47\,\mathrm{V}$. (c) 45 min, $U = 1.42\,\mathrm{V}$. (d) 90 min, $U = 1.37\,\mathrm{V}$. e) 3 h, $U = 1.10\,\mathrm{V}$. (f) A second, structurally identical battery in the almost discharged state after 9 h, $U = 0.5\,\mathrm{V}$, here the arrester is visible in the middle of the image. Discharge was carried out over a resistor of $10\,\Omega$. This image is reproduced in **colour** in the central part of the book (Manke *et al.*, 2007).

repulsive and attractive (Kralchevski and Nagayama, 2000), the origin of these interactions, however, is still under debate.

5.5.2 *Discharge processes in alkaline cells*

Alkaline batteries have been in use for about 100 years. They perform very well, are inexpensive and very reliable in both industrial applications and domestic appliances (Dell, 2000, Köhler *et al.*, 2004, Souza *et al.*, 2001, Ohms *et al.*, 2002). By *in situ* tomography it is possible to visualize and analyse the discharge process inside a battery. The observation of zinc particles located in the anode material of an alkaline cell during discharge is of special interest. The batteries were mounted on a sample holder in the synchrotron beam throughout the entire experiment, i.e. during tomographic image acquisition as well as during discharge. In the experiment shown here the same part of one battery was observed in 5 different stages of discharge. Figure 5.15 shows tomographic cross-sections through this battery for these stages and a further slice of a fully discharged battery.

The batteries consist of zinc gel as anode (inner granular structure, 67% zinc and 33% electrolyte) and manganese oxide as cathode (outer green ring) isolated by a separation fleece (blue ring between anode and cathode) and covered by a steel tube (outer red ring). The electrolyte is made of 36% potassium hydroxide, 60% water and 4% zinc oxide. The discharge times between the individual tomographic measurements ranged from 15 min to 3 h. The last slice shows a

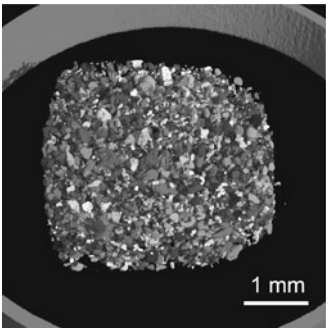


FIG. 5.16. Labelled and colour-coded zinc particles. They act as the anode in alkaline batteries. This image is reproduced in **colour** in the central part of the book.

battery in an almost discharged state after 9 h. In this image the arrester is also visible (middle part) because the tomographic data were taken from a different position in the battery (the arrester merely extends from the upper end to the middle part of the battery). In the first stage of discharge the number of small particles in the middle of the pictures in Fig. 5.15 decreases in favour of the surrounding areas, i.e. the zinc particles dissolve from the outside and zinc oxide is created. Furthermore, the outer ring in the figures increases, i.e. the manganese oxide layer expands. As a consequence, the separation fleece is compressed. The chemical reaction at the cathode and at the anode is given by (Souza *et al.*, 2001):

$$\begin{aligned} &\text{anode:} && \mathrm{Zn} + 2\mathrm{OH}^- && \longrightarrow && \mathrm{ZnO} + \mathrm{H_2O} + 2\mathrm{e}^- \\ &\text{cathode:} && 2\mathrm{MnO_2} + 2\mathrm{H_2O} + 2\mathrm{e}^- && \longrightarrow && 2\mathrm{MnOOH} + 2\mathrm{OH}^-. \end{aligned}$$

For the quantitative 3D image analysis of the zinc-particle dissolution process these particles were separated from the other components by choosing suitable grey-value thresholds and creating Boolean images. As the high number and dense packing of zinc particles in the battery inhibits reliable spatial separation of each particle in the Boolean records, a watershed algorithm combined with a Euclidian distance transformation was used to separate the particles in the Boolean image (Lohmann, 1998, Ohser and Mücklich, 2000). The particles identified in this way are displayed in a colour code for the fully charged state of the battery in Fig. 5.16. Based on this particle-separation algorithm the decrease of the zinc particle fraction for each discharge step and the shape of each of these particles could be quantified. In Figs. 5.17(a) and (b) these results are shown. The diagrams show that during discharge the mean diameter as well as the shape of the particles are almost unchanged, just the particle number decreases. This confirms the visual impression of Fig. 5.15 that indicates that particle dissolution takes place from the outside of the zinc gel, i.e. only the outer particles disappear, whereas the inner particles remain unchanged.

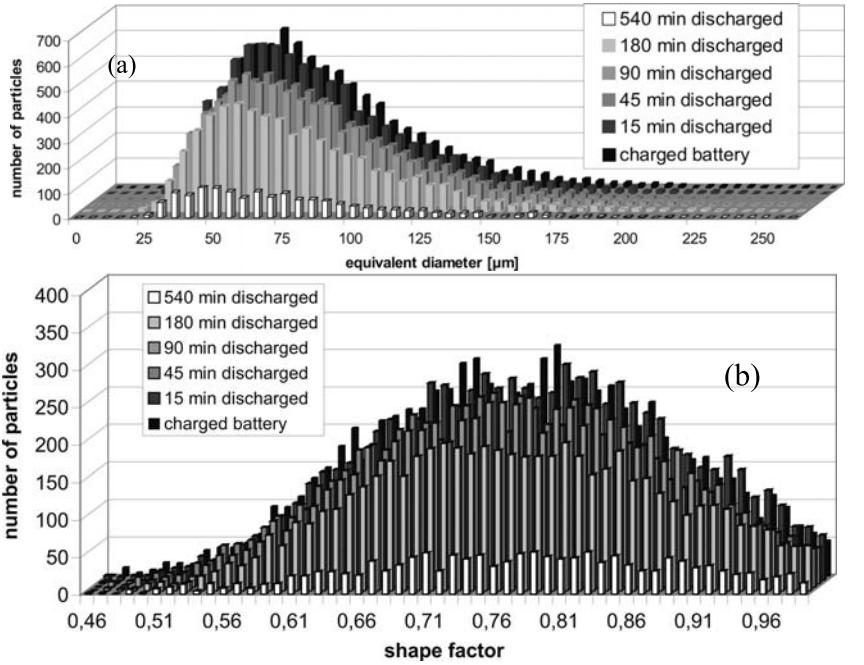


FIG. 5.17. (a) Zinc-particle number versus, (a) diameter, (b) particle shape factor for each discharge step.

5.5.3 *Ceramic foams as artificial bone marrow*

Ceramic foams have a structure similar to human bones. Therefore, an obvious idea was to use such foams as a 3D matrix for the cultivation of stem cells. The 3D porous structure of these ceramic foams provides a better cultivation environment than flat cell-cultivation dishes (Bagley *et al.*, 1999, Banu *et al.*, 2001). To guarantee an adequate supply of the cells with nutrients, special attention to the size of the pores and their interconnection is required. A three-dimensional investigation of foamed ceramic samples was carried out using synchrotron tomography.

The foams were prepared using ceramic powder (Al_2O_3), a chemically modified protein acting as foaming agent, a dispersing agent and demineralized water (Garrn *et al.*, 2004, Franzen and Kinsella, 1974, Baniel *et al.*, 1992). From food chemistry it is known that an acetylation or a glycation of proteins with sugar leads to enhanced foaming and emulsification. All components were mixed and foamed by a ball mill, after which the ceramic foam precursor material was processed in either a microwave furnace or a drying chamber and was then sintered in a high-temperature furnace. Tomographic images of four selected ceramic foams prepared in four different ways are displayed in Fig. 5.18.

Only the original components were used for making the foam shown in

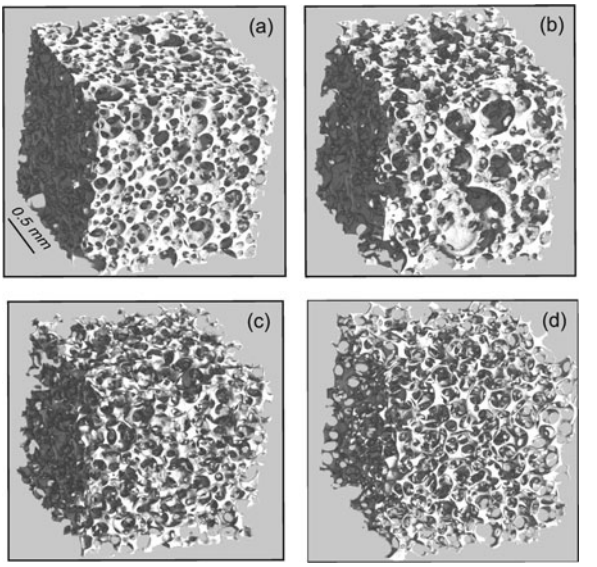


FIG. 5.18. Ceramic foam structure after sintering for four differently prepared samples. Edge length of each of the cubes is 1.44 mm. (a) Only the original components were used. (b) A dispersing agent was added to the original components. (c) In addition, the protein was modified by glycation with sugars. (d) The foam precursor was processed in a drying chamber.

Fig. 5.18(a). The precursor material was processed by microwave radiation before sintering. Here, the mean pore size and the interconnections are much smaller compared to the other foams. For the foam in Fig. 5.18(b), a dispersing agent was added to the precursor material. The effect is a broader pore-size distribution, but the interconnections are still small. By using a glycated protein for foaming, see Fig. 5.18(c), the mean pore size could be enlarged compared to the foams (a) and (b) and the interconnections of the pores are increased. The foam displayed in Fig. 5.18(d) was also prepared with glycated protein but was processed in a drying chamber instead of in a microwave furnace. In this case, the mean pore size is increased too and the pore interconnections are very large.

For the quantitative 3D image analysis of the tomographic images the interconnected pores were identified and separated by two transformations, a Euclidean distance transformation and a watershed transformation (see Fig. 5.19) and the pore-size distributions and pore shapes were analysed (Haibel *et al.*, 2005). The quantitative pore-size analysis is given in Fig. 5.20. By preparing the samples using a dispersing agent and chemically modified proteins as well as by sample processing in the drying chamber the mean pore size, the pore-size distribution and the porosity increase, whereas the number of pores decreases. Based on the knowledge on the influence of the various components and the production route on shape and size of the pores it will be possible to adapt the

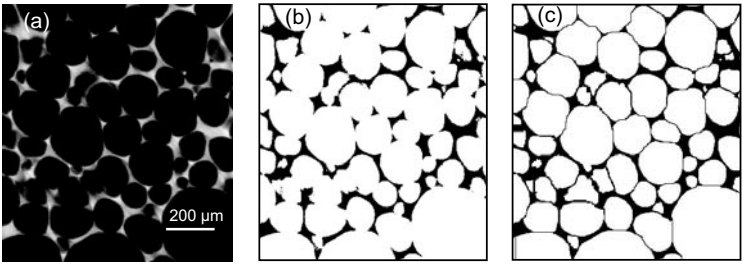


FIG. 5.19. 2D slice of a tomographic foam image, (a) original grey-value image, (b) Boolean image, (c) final image, Boolean data set masked with watershed transform.

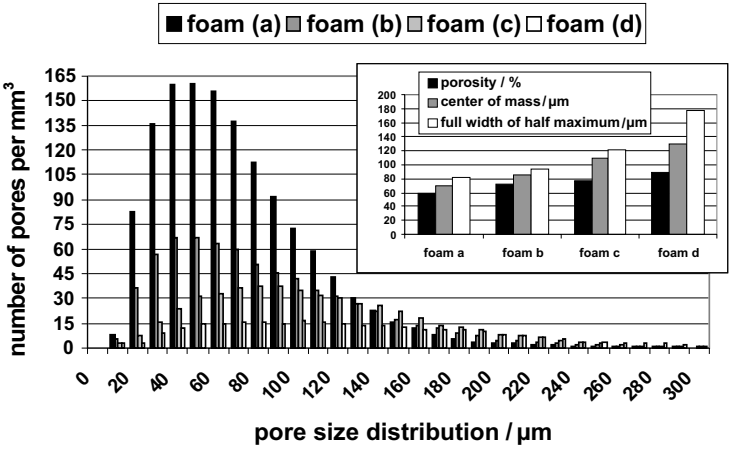


FIG. 5.20. Pore-size distribution of the four ceramic foams in Fig. 5.18. In the inset the porosity, the average diameter and the full width of half-maximum for the four foams are given.

foam structure individually to the requirement of the different specific stem-cell lines.

5.6 Summary

Microtomography provides three-dimensional images of a wide range of samples non-destructively and with a spatial resolution in the micrometre or even sub-micrometre range. Of all the different techniques absorption tomography is the most widely used. Due to the recent improvements in beam quality of modern synchrotron sources, in optics and in computer power this method is experiencing a significant boost in materials research at present, implying that the number of scientists 'discovering' tomography is increasing continuously. New developments in efficient magnifying techniques especially in the hard X-ray range will lead to

a further increase of the spatial resolution down to the nanometre scale, whereas new detector systems allows us to measure tomograms in the order of few seconds, which is mandatory for fast *in situ* or even real-time investigations.

5.7 References

Antoine, C., Nygård, Per, Gregersen, Ø., Holmstad, R., Weitkamp, T., and Rau, C. (2002). 3D images of paper obtained by phase-contrast X-ray microtomography: image quality and binarisation. *Nuclear Instruments and Methods in Physics Research A*, **490**, 392.

Bagley, J., Rosenzweig, M., Marks, D. F., and Pykett, M. J. (1999). Extended culture of multipotent hematopoietic progenitors without cytokine augmentation in a novel three-dimensional device - a case study on ex vivo erythropoiesis. *Experimental Hematology*, **27**, 496.

Banhart, J. (2000). Manufacturing routes for metal foams. *Journal of Metals*, **52**, 22.

Baniel, A., Caer, D., Colas, B., and Guerguen, J. (1992). Functional properties of glycosylated derivatives of the 11s storage protein from pea (pisum sativum l.). *Journal of Agricultural and Food Chemistry*, **40**, 200.

Banu, N., Rosenzweig, M., Kim, H., Bagely, J., and Pykett, M. (2001). Cytokine-augmented culture of haematopoietic progenitor cells in a novel three-dimensional cell growth matrix. *Cytokine*, **13**, 349.

Boin, M. (2005). Kompensation von Artefakten bei der Rekonstruktion tomographischer Datensätze. Master's thesis, Berufsakademie für Wirtschaft Berlin.

Boin, M. and Haibel, A. (2006). Compensation of ring artefacts in synchrotron tomographic images. *Optics Express*, **14**, 12071.

Bonse, U., Nusshardt, R., Busch, F., Pahl, R., Johnson, Q., Kinney, J., Saroyan, R., and Nichols, M. (1989). Optimization of CCD based energy modulated X-ray microtomography. *Review of Scientific Instruments*, **60**, 2478.

Cloetens, P., Pateyron-Salomé, M., Buffière, J.Y., Peix, G., Baruchel, J., Peyrin, F., and Schlenker, M. (1997). Observation of microstructure and damage in materials by phase sensitive radiography and tomography. *Journal of Applied Physics*, **81**, 5878.

Davis, R. and Elliott, J.C. (1997). X-ray microtomography scanner using time-delay integration for elimination of ring artefacts in the reconstructed image. *Nuclear Instruments and Methods in Physics Research A*, **394**, 157.

de Casteele, E. Van, Dyck, D. Van, Sijbers, J., and Raman, E. (2004). A model-based correction method for beam hardening artefacts in X-ray microtomography. *Journal of X-Ray Science and Technology*, **12**, 43.

Dell, R.M. (2000). Batteries, fifty years of materials development. *Solid State Ionics*, **134**, 139.

European Synchrotron Radiation Facility (2007). Scientific Software. `http://www.esrf.eu/UsersAndScience/Experiments/TBS/SciSoft/`. [Online; accessed 12 February 2007].

Franzen, K.L. and Kinsella, J.E. (1974). Parameters affecting the binding of volatile flavor compounds in model food systems. I. proteins. *Journal of Agricultural and Food Chemistry*, **22**, 675.

Garrn, I., Reetz, C., Brandes, N., Kroh, L.W., and Schubert, H. (2004). Clot forming: The use of proteins as binders for producing ceramic foams. *Journal of the European Ceramic Society*, **24**, 579.

Görner, W., Hentschel, M.P., Müller, B.R., Riesemeier, H., Krumrey, M., Ulm, G., Diete, W., Klein, U., and Frahm, R. (2001). BAMline: the first hard X-ray beamline at BESSY II. *Nuclear Instruments and Methods in Physics Research A*, **467**, 703.

Grübel, G., Abernathy, D., Vignaud, G., Sanchez del Rio, M., and Freund, A. (1996). A diamond double-crystal transmission monochromator for the TROIKA II station at ESRF. *Review of Scientific Instruments*, **67**, 9.

Haibel, A., Freidank, H., Berthold, A., Banhart, J., and Schubert, H. (2005). *Towards artificial bone marrow: 3D matrices to cultivate stem cells*. Berliner Elektronensynchrotron (BESSY), Berlin. Highlight Report, p. 28.

Haibel, A., Rack, A., and Banhart, J. (2006). Why are metal foams stable? *Applied Physics Letters*, **89**, 154102.

Haimerl, M. (2003). *Modalitätsspezifische Anpassung medizinischer Bilddaten*. Ph. D. thesis, Universität Karlsruhe.

Herman, G.T. (1980). *Image reconstruction from projections - The fundamentals of computed tomography*. Academic Press, New York.

Horbelt, S., Liebling, M., and Unser, M. (2002). Filter design for filtered back-projection guided by the interpolation model. In *Medical Imaging 2002: Image Processing* (ed. M. Sonka and J. Fitzpatrick), pp. 806. The International Society for Optical Engineering (SPIE). Proceedings of the SPIE, Vol. 4684.

Kak, A.C. and Slaney, M. (1987). *Principals of computerized tomographic imaging*. IEEE Press, New York.

Kaptay, G. (2004). Interfacial criteria for stabilization of liquid foams by solid particles. *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, **230**, 67.

Köhler, U., Antonius, C., and Bäuerlein, P. (2004). Advances in alkaline batteries. *Journal of Power Sources*, **127**, 45.

Körner, C., Arnold, M., and Singer, R.F. (2005). Metal foam stabilization by oxide network particles. *Materials Science and Engineering A*, **396**, 28.

Kralchevski, P.A. and Nagayama, K. (2000). Capillary interactions between particles bound to interfaces, liquid films and biomembranes. *Advances in Colloid and Interface Science*, **85**, 145.

Link, T., Zabler, S., Epishin, A., Haibel, A., Bansal, M., and Thibault, X. (2006). Synchrotron tomography of porosity in single-crystal nickel-base superalloys. *Materials Science and Engineering A*, **425**, 47.

Lohmann, G. (1998). *Volumetric image analysis*. John Wiley & Sons and G.B. Teubner, Chichester and Stuttgart.

Lu, W. and Mackie, T.R. (2002). Tomographic motion detection and correction directly in sinogram space. *Physics in Medicine and Biology*, **47**, 1267.

Manke, I., Haibel, A., Rack, A., Kardjilov, N., Hilger, A., Melzer, A., Zabler, S., H.Riesemeier, and Banhart, J. (2007). In situ investigation of the discharge of alkaline Zn-MnO_2 batteries with synchrotron X-ray and neutron tomography. *Applied Physics Letters*, **90**, 214102.

Marion, A. (1991). *An introduction to image processing.* Chapman and Hall, London.

Mathiesen, R.H. and Arnberg, L. (2006). X-ray monitoring of solidification phenomena in Al-Cu alloys. *Materials Science Forum*, **508**, 69.

Noethe, M., Pischang, K., Ponizil, P., Kieback, B., and Ohser, J. (2003). Investigation of sintering processes by microfocus computer tomography (μ-CT). In *DGZfP-proceedings, CT-IT*, Volume BB 84-CD, pp. 273.

Ohms, D., Kohlhase, K., Benczur-Ürmössy, G., and Schädlich, G. (2002). New developement on high power alkaline batteries for industrial applications. *Journal of Power Sources*, **105**, 127.

Ohser, J. and Mücklich, F. (2000). *Statistical analysis of microstructures in materials science.* John Wiley & Sons, Chichester.

Raven, C. (1998). Numerical removal of ring artifacts in microtomography. *Review of Scientific Instruments*, **69**, 2978.

Sijbers, J. and Postnov, A. (2004). Reduction of ring artifacts in high resolution micro-CT reconstructions. *Physics in Medicine and Biology*, **49**, 247.

Souza, C.C.B.M. de, Oliveira, D.C. de, and Tenorio, J.A.S. (2001). Characterization of used alkaline batteries powder and analysis of zinc recovery by acid leaching. *Journal of Power Sources*, **103**, 120.

Thurner, P., Müller, B., Sennhauser, U., Hubbell, J., and Müller, R. (2004). Tomography studies of biological cells on polymer scaffolds. *Journal of Physics: Condensed Matter*, **16**, S3499.

Wübben, T. and Odenbach, S. (2005). Stabilisation of liquid metallic foams by solid particles. *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, **266**, 207.

6

PHASE-CONTRAST AND HOLOGRAPHIC TOMOGRAPHY

Andrew G. Peele and Keith A. Nugent

The coherent output of X-ray sources, from the laboratory tube through to modern-day third-generation synchrotron sources, has been increasing at a phenomenal rate. While, the advent of X-ray free-electron lasers over the next few years will see the culmination of this progress with the availability of truly coherent sources, the coherence of the light from a modern insertion device beamline is already such that coherence effects are impossible to ignore. Soon after modern insertion devices were developed, it was noticed that small imperfections in the optics and windows of the beamline produced significant modulations in the observed intensity (Snigirev *et al.*, 1995). These often arose from small thickness variations in the windows used in the beamline; unexpectedly, X-ray phase-contrast images of the beamline windows were being acquired.

Further analysis showed that, in fact, the coherence requirements for phase-contrast imaging are very modest and it was subsequently shown that phase-contrast was possible with conventional laboratory X-ray sources (Wilkins *et al.*, 1996) and even in highly non-coherent environments such as neutron radiography (Allman *et al.*, 2000). Phase-contrast X-ray imaging has become a standard method for synchrotron-based radiography. This mechanism of phase-contrast is now used for imaging in both two and three dimensions (Cloetens *et al.*, 1999).

In this chapter we briefly review how X-ray phase contrast can be applied to tomographic imaging.

6.1 Propagation of light and phase contrast

Any detector ultimately senses the deposition of energy on its surface. In particular, an imaging detector, such as a CCD camera, measures the spatial distribution of energy. The key to phase-contrast imaging is to convert variations in phase into variations in intensity.

It is important to differentiate between phase contrast (or visualization) and phase retrieval. In the former, intensity becomes modulated by quantities related to the phase. Such images can be used in a qualitative manner to reveal shape information about a sample. On the other hand, with phase retrieval the modulated intensity images are inverted to quantify the phase. The same distinction also holds in phase-contrast tomography. Using phase-contrast tomography it is possible to visualize the three-dimensional distribution of some aspect of the phase distribution, such as, for example, its gradient. In phase-retrieval-based

tomography that raw phase-contrast data can be used as the basis for calculation that allows one to quantitatively map the complex refractive index of the sample in three dimensions.

A coherent wave is characterized by both its phase and amplitude distribution, and these can vary independently of each other. Due to the manner in which phase information becomes encoded in the intensity distribution of an image, the minimum requirement for quantitative phase retrieval frequently requires there to be two intensity measurements taken under different conditions. In the case of tomography, this implies two intensity measurements are needed for each tomographic projection. This requirement can be relaxed if there is additional *a-priori* information available about the sample.

The conversion of phase information into intensity contrast falls into three major classes:

1. *Interferometry*: Interferometry involves the coherent superposition of a wave that has passed through the sample onto a reference wave, resulting in interference fringes. This is a technique with a long history, and, in the absence of amplitude modulations, records a signal that is proportional to the sine of the phase shift. Interferometry is rarely used for phase visualization. However, related techniques, such as Zernike phase contrast and differential interference contrast, are interferometric in nature and are often used in phase-visualization applications.
2. *Diffraction-enhanced imaging*: Diffraction-enhanced imaging involves optical elements that, in the absence of amplitude modulations, are configured so as to reveal the phase gradient. Schlieren imaging is a familiar optical analogue. These techniques show exquisite sensitivity to phase gradients and are superb tools for phase-contrast imaging, but are of limited value for phase retrieval.
3. *Propagation-based contrast*: If a wave has a phase variation imprinted on it, areas in which the phase is curved in some way will tend to either focus or defocus the light as it propagates, yielding regions of either higher or lower intensity. This is the source of phase visualization that was observed using third-generation synchrotrons and revealing phase in this way does not require the introduction of any optical elements. For small propagation distances the signal, in the absence of amplitude modulation, yields the Laplacian of the phase.

In this chapter, we discuss these techniques, as well as some others, in the context of their application to tomographic imaging. Interferometry is discussed in Section 6.2, and its cousin, Zernike phase contrast, is discussed in Section 6.4. Phase-gradient sensitive methods are discussed in Sections 6.5. Section 6.6 describes approaches that arise from propagation of the beam.

Towards the end of the chapter we will consider a method based on far-field diffraction that can yield very high resolution imaging (Section 6.7). Phase retrieval also allows for the possibility of broadband radiation and consequently

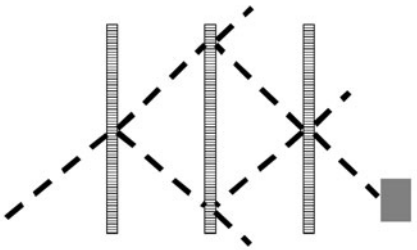


FIG. 6.1. Triple Laue case interferometer. Three wafers act, from left to right, as beam splitter, mirror and analyser, thus allowing an incident X-ray beam (at left) to follow two optical paths and recombine at a detector (at right).

fast data collection. We consider the practical aspects of this work in Section 6.8.

There are many additional approaches to visualizing and retrieving phase that will not be discussed here. In this chapter we explicitly limit the discussion to methods of phase visualization and phase retrieval that have been implemented in tomographic applications. For a broader view of the field of phase imaging, the reader is invited to pursue the many excellent review articles discussing phase methods (Paganin and Nugent, 2001, Momose, 2003, Paganin, 2006).

6.2 Interferometry for phase tomography

Interferometry for phase measurement has a long and distinguished history. A wave that has passed through the object is superimposed on a known reference wave to reveal the phase difference between the two waves via interference fringes. If the wave incident on the sample and the reference wave are both planar then, in the absence of a phase perturbation introduced by a sample, the fringes will be linear. The sample phase can be recovered by observing the deviation from linearity.

Interferometry will only yield the phase modulo 2π and to resolve the true phase requires a method known as phase unwrapping in which the ambiguities in the phase are resolved using additional information, or additional assumptions about the data, such as its continuity. In the presence of amplitude modulation of the projection, the analysis method is further complicated.

Many different schemes have now been implemented for interferometry – for a review see Momose 2003. We discuss here X-ray interferometric phase tomography as has been developed by Momose and colleagues and a recent demonstration based on the use of gratings.

6.2.1 *Bonse–Hart interferometry*

In this work, a Bonse–Hart interferometer (sometimes called an LLL interferometer for the three Laue geometry diffracting crystals employed) cut from a single silicon crystal is used to split and recombine an incident X-ray beam in a division-of-amplitude interferometer. A schematic of such an interferometer is

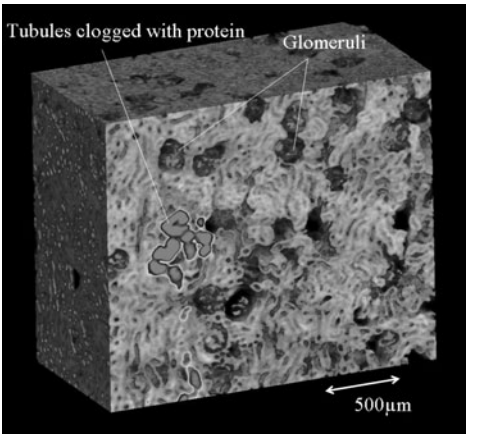


FIG. 6.2. Phase reconstruction of a section of rat kidney obtained at 0.1 nm using an LLL interferometer (Fig. 8 from Momose 2003, reused with permission).

shown in Fig. 6.1. In practice, the requirement that the wafers be cut from a single Si ingot limits the sample dimension to less than about 7 cm (Momose, 2003).

The interference patterns obtained from the interferometer can be interpreted using either a method in which a single fringe pattern is analysed using filtering in the Fourier space of the interferogram (Takeda *et al.*, 1982), or by the phase-shifting method, in which several interferograms are obtained by varying the phase in the reference arm.

Interferometry is not practical where there are large phase shifts and large phase gradients and the sample is typically immersed in a medium to remove the large phase shift at the sample/air interface. Conversely, the technique is extremely sensitive to small phase variations and can resolve sample density variations that are potentially well below the capacity of other methods.

In an example of this technique (Takeda *et al.*, 2000) a section of a rat's kidney was imaged with 0.1-nm wavelength X-rays using an LLL X-ray interferometer. The analyser crystal was particularly thin (40 μm) so as to reduce blurring due to refraction effects through the sample. The image in Fig. 6.2 maps the refractive index difference between the tissue and the formalin in which it was immersed. Tubules and other structure as well as tubules blocked with protein in the kidney can clearly be observed. The sensitivity of the method to density variation in the sample was estimated to be 1.2 mg/cm^3. A similar technique has been used for neutrons, see Chapter 14.2.1.

6.3 Differential interferometry

A recent scheme that performs interferometry using microfabricated gratings, and is particularly suited to high-energy X-rays, has also been demonstrated

(David *et al.*, 2002). The first (phase) grating produces a series of diffraction fringes that in the absence of a sample are detected via the second (amplitude) grating. When a sample is present the deviations to the beam produce a change in the detected intensity. This modulation can be used to produce a phase image of the sample. The scheme has been extended to tomography (Weitkamp *et al.*, 2005) and, by the use of a third grating, to laboratory sources (Pfeiffer *et al.*, 2006). The gratings used in this work have an extremely large spacing in comparison to crystal-based interferometry. Indeed, the work is perhaps even more closely related to the work in Talbot interferometry, a method that can be understood using geometric optics. See Section 14.2.3 for an adaption of this method to neutron imaging.

6.4 Zernike phase contrast for tomography

This form of contrast is discussed in many undergraduate text books (Hecht, 2001) and so will not be reviewed here. The method converts small phase shifts into an intensity modulation. More general expressions, for instance when the phase is not small (Paganin, 2006), also result in the conclusion that the re-imaged intensity becomes modulated by an expression related to the phase.

Zernike phase contrast in the context of tomography appears to have had a very limited discussion in the publication record. The first Zernike phase-contrast images in X-ray microscopy were discussed in the 1980s (Schmahl *et al.*, 1988). However, there is little reporting of the tomographic application of this modality in the refereed literature. Nevertheless, X-ray Zernike phase-contrast tomography is now available commercially and an example (Tkachuk *et al.*, 2006) of virtual delayering of a Pentium-4 microprocessor is shown in Fig. 6.3. The reconstruction was created from phase-contrast images obtained at a wavelength of 0.15 nm using the Zernike phase mode of a commercial X-ray computed tomography microscope. In the case of the example shown in Fig. 6.3, the reconstructed resolution is better than the 90 nm of the smallest features on the integrated circuit.

6.5 Diffraction-enhanced imaging

Diffraction-enhanced imaging (DEI) senses the refractive deviation from the object using crystal optics (Forster *et al.*, 1980). The conceptual scheme underlying DEI is shown in Fig. 6.4. A beam collimated by a crystal is passed through a sample. Phase modulations in the sample refract the beam, which then passes through an analyser crystal. The transmission of the analyser crystal is very sensitive to the incident angle via the rocking curve of the crystal and so the transmission through the analyser crystal converts the direction of propagation into an intensity modulation. The resulting image renders visible the phase-gradient component in one direction. The method is conceptually similar to Schlieren imaging (Settles, 2001), a method commonly used in, for example, visualizing fluid flow.

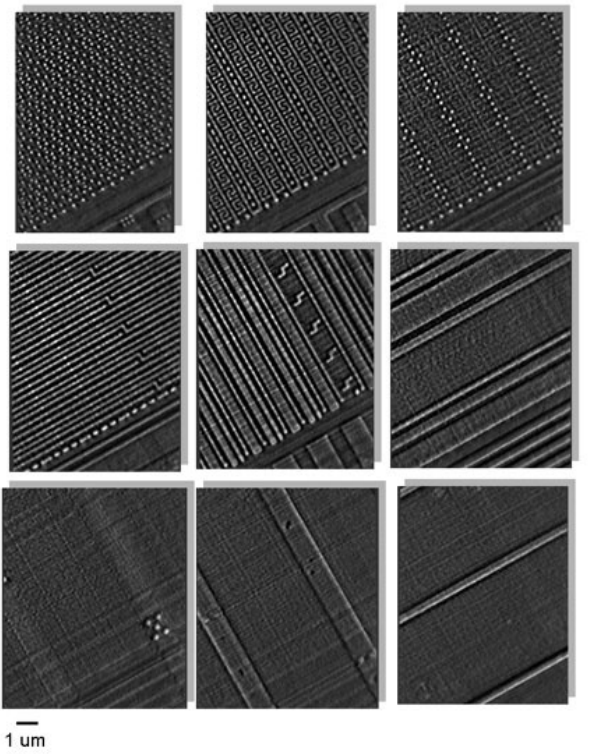


FIG. 6.3. Reconstructed layers of a Pentium 4 CPU obtained at 0.15 nm wavelength in Zernike phase mode (Fig. 6 from Tkachuk *et al.* 2006, reused with permission).

If the object introduces amplitude variations in the transmitted field then phase and absorption contrast are mixed in a complex way. This method is therefore very difficult for the production of quantitative phase images, though some methods have been developed (Nesterets *et al.*, 2004, Guigay *et al.*, 2007).

The first combination of an analyser crystal in DEI with tomography was by Dilmanian (Dilmanian *et al.*, 2000) to reconstruct refraction information from an object. Recent emphasis has been on the implementation of DEI-based phase-contrast tomography with particular emphasis on biological samples due to its exceptional sensitivity (Fiedler *et al.*, 2004, Zhu *et al.*, 2005, Pagot *et al.*, 2005).

In a recent application (Modregger *et al.*, 2006) reflections from asymmetrically cut crystals were used to produce an expanding beam with wavelength 0.15 nm that magnified the image in two dimensions. The resulting phase-contrast projected images of a spider's leg were used to obtain the reconstruction shown in Fig. 6.5. The authors also obtain a phase-sensitivity figure of

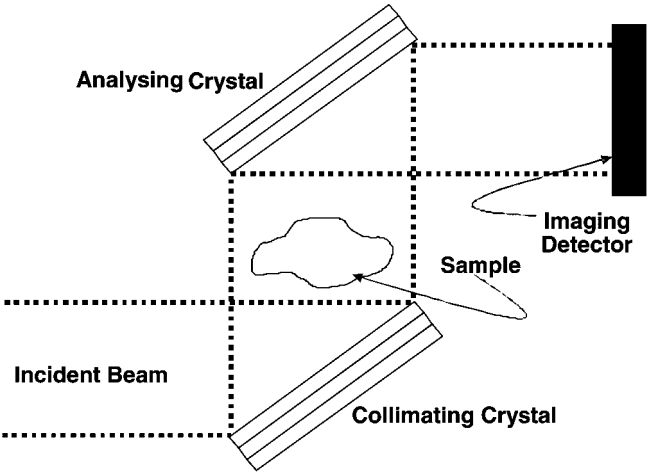


FIG. 6.4. Principle of DEI. The first crystal monochromates and collimates the beam via the Bragg condition. The collimated beam incident on the sample is refracted. Only rays satisfying the Bragg condition at the second crystal are passed to the detector. By rocking the analyser a spatially resolved map of phase gradients from the sample is obtained.

$$\frac{\Delta\phi}{\Delta x} = 0.06\,\frac{\text{rad}}{\mu\text{m}},$$

compared to an estimated

$$\frac{\Delta\phi}{\Delta x} = 1\,\frac{\text{rad}}{\mu\text{m}}$$

for inline holography. Given the strong anisotropy in the contrast distribution obtained with DEI it is necessary to process the data so that, in general, the quantification of reconstructed images is not straightforward and authors have discussed techniques such as changing the scan direction to provide the basis for good three-dimensional reconstruction (Zhu *et al.*, 2005, Maksimenko *et al.*, 2005, Wang *et al.*, 2006).

6.6 Propagation-based phase tomography

Propagation will yield phase contrast even without additional optical elements. When the propagation distance is small then tomographic methods may be directly applied to the data to recover a three-dimensional image of the Laplacian of the phase, though absorption may render the image interpretation more difficult. See Section 14.2.2 for an application to neutron imaging.

The first intentional use of propagation-based phase-contrast tomography was reported in 1996 for the reconstruction of a boron fibre (Raven *et al.*, 1996). While the edge-enhancement effect of propagation-based phase contrast is useful

FIG. 6.5. Rendered surface representation of part of a spider leg obtained using DEI-based contrast at 0.15 nm wavelength. The cube shown has a side length of 100 μm (reprinted Fig. 11 with permission from Modregger *et al.* 2006. Copyright 2006 by the American Physical Society).

for visualizing sharp phase gradients it can be seen that in reconstruction it reveals an edge-enhanced version of the object (a higher-resolution version from a more recent experiment is shown in Fig. 6.7(b) below). A spectacular application of this technique has recently been demonstrated in application to paleontology (Tafforeau *et al.*, 2006). Propagation-based phase contrast at 20 keV and a propagation distance of 150 mm was used to moderately enhance the features of a beetle encased in amber during the Cretaceous period, as shown in Fig. 6.6. After performing a tomographic reconstruction image processing was used to segment the beetle image from the amber. In this case the phase contrast operated as an enhancement to the absorption contrast. Perhaps even more striking are examples where there is little or no absorption contrast and propagation can be used to give a well-defined image (Baruchel *et al.*, 2006). We turn now to examples of where the phase-retrieval step has been implemented to produce representative maps of the sample.

6.6.1 *Transport of intensity approaches*

6.6.1.1 *General approach* The transport of intensity approach is based on an application of the conservation of energy. Consequently, it can be shown that as a beam of light passes from one plane to another it obeys the equation:

$$\frac{\partial I(\boldsymbol{r})}{\partial z} = -\frac{\lambda}{2\pi}\nabla[I(\boldsymbol{r})\nabla\Phi(\boldsymbol{r})], \tag{6.1}$$

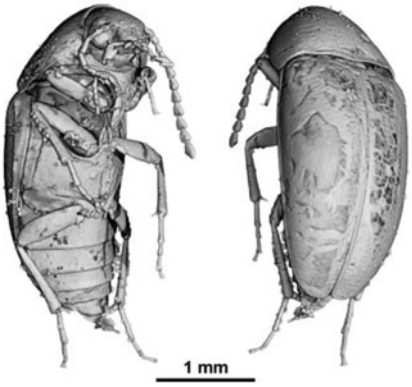


FIG. 6.6. Rendered surface representations of a Cretaceous beetle. The sample was encased in an opaque amber and the image shown was obtained using phase-contrast enhancement and by virtually removing the beetle from the amber in the reconstructed data volume (Fig. 5 of Tafforeau *et al.* 2006, Copyright 2006 by Springer-Verlag, reused with kind permission of Springer Science and Business Media).

where λ is the wavelength of the incident radiation, $\boldsymbol{r}$ is the position in a plane normal to the optical axis, z is distance along the optical axis and I and Φ are the intensity and phase, respectively.

In practice, measurements, are made over two planes separated by a small distance Δz. One can then form:

$$I(\boldsymbol{r}) \cong \frac{1}{2}[I_+(\boldsymbol{r}) + I_-(\boldsymbol{r})],$$
$$\frac{\partial I(\boldsymbol{r})}{\partial z} \cong \frac{1}{\Delta z}[I_+(\boldsymbol{r}) + I_-(\boldsymbol{r})],$$

where I_+ and I_- are the intensities recorded at the two positions. These two quantities are inserted into eqn 6.1 and it is then solved for the phase distribution, yielding a measurement of both the phase and amplitude of the wave. This method has been applied extensively to phase imaging, and also to the recovery of the complex refractive index in the X-ray regime. An example of such an approach was shown by McMahon *et al.* 2003. A boron fibre with a tungsten core was imaged using 35-keV X-rays as shown in Fig. 6.7. Figure 6.7(b) demonstrates the edge-enhanced reconstruction typical of this approach. For complicated objects when the phase contrast is strong the interpretation of such reconstructions can be difficult. If however, the phase-contrast image is used as the basis for phase retrieval then a projection of the real part of the refractive index of the sample can be obtained and reconstructed in the usual way, as shown in Figs. 6.7(a) and (c). While in Fig. 6.7(d) the imaginary part of the refractive index (representing absorption in the sample) is also accurately retrieved.

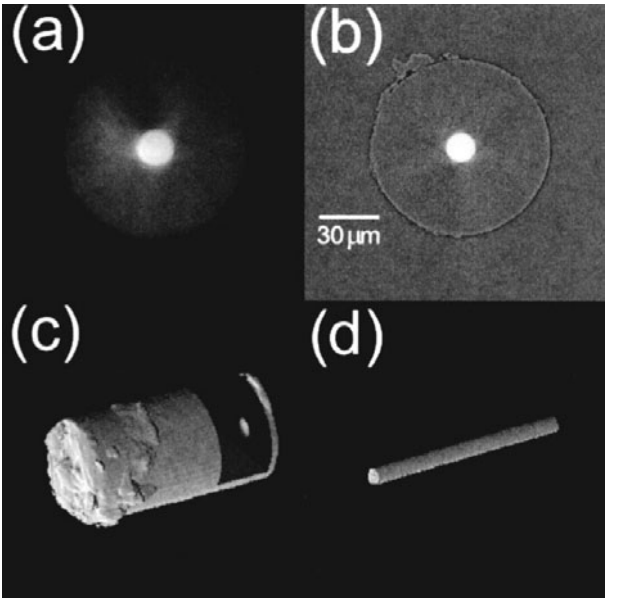


FIG. 6.7. (a) Phase reconstruction of a boron fibre with a tungsten core obtained using 35-keV X-rays and propagation distances of 150 and 200 mm. (b) Reconstruction from the phase-contrast images showing the edge-enhancement effect. (c) Rendered surface representation of the phase reconstruction. (d) Rendered surface representation of the imaginary part of the reconstructed refractive index (Reprinted Fig. 1 reused with permission from McMahon *et al.* 2003. Copyright 2003, American Institute of Physics).

6.6.1.2 *Homogeneous sample* As discussed in Section 6.1, in general a complex field will require two data sets to be acquired in order to be completely recovered, and the full implementation of the transport of intensity equation is an example of this.

In many cases, it is possible to use additional *a-priori* information such as knowledge of the composition of the sample, which in turn constrains the relationship between the real and imaginary parts of the refractive index. That is, the absorption and phase are not independent. The homogeneous sample approximation, then, assumes that the object consists of a single material with a known complex refractive index. Under this assumption, and also assuming that the object does not interact very strongly with the radiation, one may recover the physical thickness of the sample in one direction (its projection) from one image using the result:

$$t(\boldsymbol{r}) = -\frac{\lambda}{4\pi\beta} \ln \mathcal{F}^{-1}[\frac{\beta}{\beta + \lambda z \delta \left|\boldsymbol{q}\right|^2} \frac{\hat{I}_z(\boldsymbol{q})}{I_0}],$$

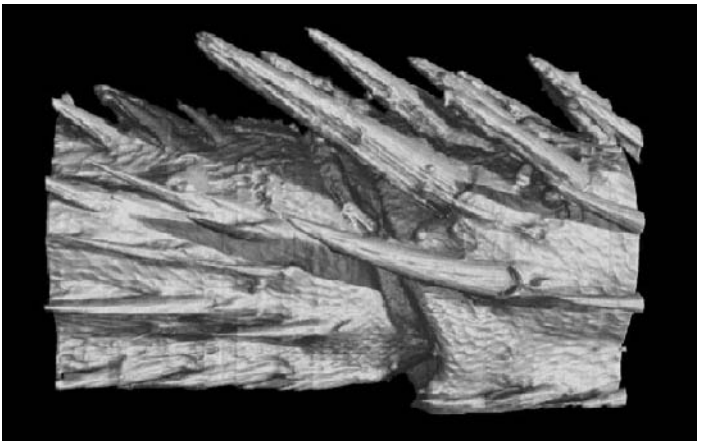


FIG. 6.8. Rendered surface representation of part of a fly's leg obtained in point-projection mode imaging using a laboratory source using the homogeneous-object single-plane retrieval method. The diameter of the leg is ∼100 μm (Fig. 11 of Mayo *et al.* 2003, reused with permission).

where λ is the wavelength of the incident radiation, $\boldsymbol{r}$ is the position in a plane normal to the optical axis, β and δ are the imaginary and real parts of the refractive index, respectively, z is the propagation distance, I_0 is the incident intensity, $\mathcal{F}$ is the Fourier transform operator and $\hat{I}_z(\boldsymbol{q}) = \mathcal{F}[I(\boldsymbol{r})]$, where I is the measured intensity and $\boldsymbol{q}$ is the Fourier conjugate variable to $\boldsymbol{r}$. This is the expression for a homogeneous object as first proposed by Paganin *et al.* 2002. We note that a very similar idea, based on the phase–attenuation duality of the interactions of high-energy X-rays with matter, has also been proposed for medical phase-contrast imaging (Wu *et al.*, 2005).

This approach does not require careful alignment of the two images as required for the general transport of intensity approach, and is rather more stable in the presence of noise. This method has been applied using both laboratory sources (Mayo *et al.*, 2003) and using synchrotron sources (Peele *et al.*, 2005).

The homogeneous sample solution to the phase has been used by Mayo *et al.* 2003 to produce tomographic reconstructions of part of an insect using a laboratory source as shown in Fig. 6.8. The sample was imaged in point-projection geometry with an source to object distance of 6.9 mm and a source to detector distance of 250 mm. The homogeneous sample approximation removes the need for rescaling and alignment of multiple images that simplifies the analysis greatly and even makes the technique a worthwhile qualitative one for non-homogeneous samples (Mayo *et al.*, 2003).

6.6.2 *Holotomography*

Once X-rays have passed through a sample the laws governing their propagation are well known. Where the interaction with the object is weak, in the sense that

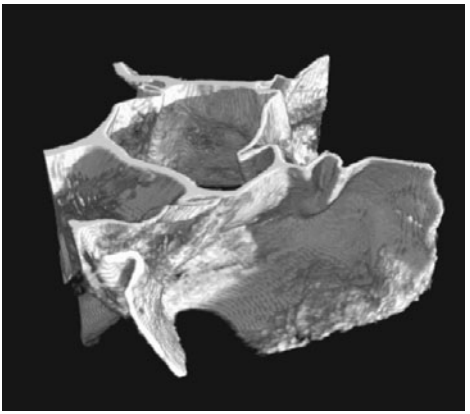


FIG. 6.9. Rendered surface representation of a piece of foam obtained using the holotomography method for 0.069-nm wavelength X-rays. The width of the image is ∼290 μm (Reused Fig. 3a with permission from Cloetens *et al.* 1999. Copyright 1999, American Institute of Physics).

absorption is small and the phase is either small or slowly varying, then it is possible to describe the intensity as a function of position in the following form (Guigay *et al.*, 1971, Guigay, 1977):

$$\hat{I}_z(\boldsymbol{q}) = I_0(\delta(\boldsymbol{q}) - 2\cos(\pi\lambda z\,|\boldsymbol{q}|^2)\mathcal{F}[\mu(\boldsymbol{r})] + 2\sin(\pi\lambda z\,|\boldsymbol{q}|^2)\mathcal{F}[\phi(\boldsymbol{r})], \qquad (6.2)$$

where $\delta(\boldsymbol{q})$ is the Dirac delta function, λ is the wavelength of the incident radiation, z is the propagation distance, $\mu(\boldsymbol{r})$ is the absorption coefficient, $\phi(\boldsymbol{r})$ is the phase shift produced by the sample, $\boldsymbol{r}$ is the position in a plane normal to the optical axis, I_0 is the incident intensity, $\hat{I}(\boldsymbol{q})$ is the Fourier transform of the intensity, $\mathcal{F}$ is the Fourier transform operator and $\boldsymbol{q}$ is the Fourier conjugate variable to $\boldsymbol{r}$.

Two independent measurements of the intensity at different propagation distances allow this pair of equations to be solved for both the absorption and the phase (Zabler *et al.*, 2005). This can be done for each projection then placed into a tomographic reconstruction algorithm. As with the transport of intensity approach a homogeneous sample solution is also possible (Guigay *et al.*, 1971, Guigay, 1977, Turner *et al.*, 2004). Unlike the transport of intensity this approach is not limited to short propagation distances. Consequently, the oscillations in the trigonometric terms in eqn 6.2 will manifest as changes and reversals in the contrast. At each reversal there is a point of zero contrast. The idea of holotomography is to measure data from several different propagated planes and combine the results so that the zeroes in contrast for certain spatial frequencies are 'filled in' by data from other planes. In the seminal example (Cloetens *et al.*, 1999) data was collected at a wavelength of 0.069 nm for 4 different propagation distances between 0.027 m and 0.901 m. The sample was a piece of polystyrene

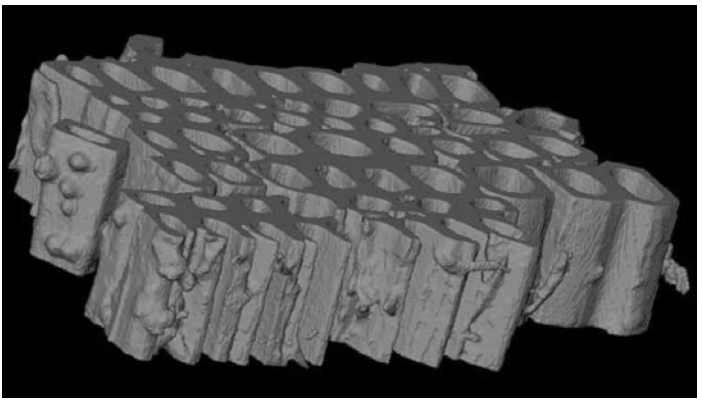


FIG. 6.10. Rendered surface representation of a 350-μm thick piece of wood obtained from a single-step phase-retrieval and reconstruction method using 25-keV X-rays and a propagation distance of 25 mm (Fig. 3b of Groso *et al.* 2006, reused with permission).

foam, which at the high energy used was essentially a phase object. The spatial details of the sample are well imaged as shown in Fig. 6.9 and the recovered value for the real part of the refractive index agreed with the expected value to within 5%.

6.6.3 *Integrated approach*

Certain of the Fourier space operations involved in the direct phase-retrieval operations described in Sections 6.6.1 and 6.6.2 can be combined with those involved in the Radon transform step in the tomographic reconstruction (Bronnikov, 1999, Bronnikov, 2002, Groso *et al.*, 2006, Gureyev *et al.*, 2006). This results in numerical efficiencies in the computation of the solution. This approach has been implemented as shown in Fig. 6.10 using a semi-empirical filter term as part of the reconstruction algorithm and applied for phase reconstruction using the transport of intensity approach to a 350-μm thick wood sample (Groso *et al.*, 2006). An integrated approach is also possible in the case of coherent diffractive imaging-based tomography and is described below. A 'one-step' solution has also now been presented for DEI (Huang *et al.*, 2006).

6.7 Coherent diffractive imaging

Coherent diffractive imaging is an emerging technique that is applicable to measurements in the far-field of the object. In this method, the far-field diffraction pattern is sampled at twice the Nyquist frequency corresponding to the known finite extent of the sample (Sayre, 1952) (hence the alternative appellation for the method of 'oversampling'). The finite extent of the sample can be determined by lower-resolution imaging methods or by constructing the autocorrelation function of the diffracted intensity. The knowledge that the sample has finite support

FIG. 6.11. Rendered surface representation of gold balls arranged in a pyramid shape obtained using coherent diffractive imaging at a wavelength of 1.65 nm and sampling the three-dimensional diffraction pattern. The base of the pyramid is 2.5 μm and its height 1.8 μm (Fig. 2b of Chapman *et al.* 2006, reused with permission).

(i.e. is of limited spatial extent) is sufficient when coupled with the measured far-field diffraction pattern to constrain the unknown phase in the diffraction pattern to a solution that is unique in almost all cases (Bates, 1982, Nugent *et al.*, 2003). Essentially, the problem becomes one of finding a global minimum in a space of possible phases of the wavefield in the far-field such that the corresponding reconstructed sample (given by the inverse Fourier transform of the complex wavefield at the detector) produces a diffraction pattern that most closely matches the measured data. The basic algorithm that is applied to obtain the solution is a variant of the Gerchberg–Saxton algorithm, which in turn can be shown to be equivalent to a method of steepest-descent least squares minimization (Fienup, 1982, Elser, 2003). The algorithm proceeds as follows:

1. Construct a guess at the detector wavefield by taking the square root of the intensity and a guess at the phase: $\Psi(r) = \sqrt{I(r)}e^{i\phi_{\text{guess}}}$.
2. Take the inverse Fourier transform of the guess in Step 1.
3. Apply the support constraint so that the result of Step 2 is unchanged in the region of the support and is set to zero outside that region.
4. Take the Fourier transform of the result of Step 3.
5. Replace the amplitude of the result of Step 4 with the square root of the measured intensity.
6. Repeat Steps 2 to 5 until the modulus of the result of Step 4 converges to an acceptably close approximation to the square root of the measured intensity.
7. Repeat Step 2 to obtain the exit-wave distribution.

Variations of this basic 'error-reduction' algorithm exist. For instance, Fienup's hybrid input-output scheme that results from a consideration of the nature of the operations described above and how they affect the solution path (Fienup, 1982).

FIG. 6.12. On the left is a rendered phase-retrieved image of the bee sting used for broad-beam imaging experiments. The top diagonal series on the right are reconstructed slices through the sting obtained using the single-plane transport of intensity phase-retrieval method at 13 keV. The lower diagonal series are approximately the same slices obtained using a broad spectrum between 10 keV and 30 keV and phase retrieved using an effective energy of 16.9 keV. The propagation distance between sample and detector was 970 mm. The average diameter of the sting is ∼100 μm (Figs. 5 and 6 of Peele *et al.* 2005, reused with permission. Copyright 2005, American Institute of Physics).

The first experimental demonstration of this approach was by Miao *et al.* 1999. A limited projection series data set was also reconstructed as a demonstration (Miao *et al.*, 2002). The approach taken was to recover the density distribution of the sample for each projection using the method of coherent diffractive imaging just outlined and then to perform the tomographic reconstruction. In a more recent result (Chapman *et al.*, 2006) an integrated approach was taken. In the far-field diffraction case a series of projections obtained by rotating the sample will result in a set of diffraction patterns that sample almost all of the reciprocal space volume of the three-dimensional object. Accordingly, the method described above can be applied to the three-dimensional data set and by applying three-dimensional Fourier transforms of the data. This approach was taken for a sample consisting of a number of 50 nm diameter gold balls arranged on a pyramid-shaped indentation. Diffraction patterns at a range of rotations were measured at a wavelength of 1.65 nm. The reconstruction of the sample shown in Fig. 6.11 displays a transverse resolution of about 10 nm.

6.8 Polychromatic phase imaging

The response of the real part of the refractive index to changes in wavelength for X-rays is such that for some phase-retrieval techniques (for instance propagation-based methods) a broadband incident spectrum can be used and an effective wavelength used in the analysis. This allows for extremely fast data collection times using synchrotron sources and for efficient use of laboratory sources

(Pogany *et al.*, 1997, Mayo *et al.*, 2002, Arhatari *et al.*, 2005, Peele *et al.*, 2005). In a recent example (Peele *et al.*, 2005) a bee sting was imaged using a broad spectrum from a bending magnet source. The effective energy used was 16.9 keV and significant flux was present between 10 to 30 keV. The net exposure time was reduced from 1 h for data collected at a monochromatic 13 keV to about 14 s. A rendered image of the bee sting and reconstructed phase slices for the monochromatic and polychromatic cases is shown in Fig. 6.12. It was not expected that the reconstructions be identical due to the different energies involved. Nevertheless, the similarities are clear and the benefit of vastly reduced exposure times is obvious.

6.9 Summary

We have explored several areas of phase visualization and retrieval that have been deployed towards the capture of tomographic data sets. While the experimental obstacles for techniques such as interferometry and diffraction-enhanced imaging are large – often requiring high-precision control (in the case of interferometry) or multiple scans (in the case of DEI), the benefits in the phase sensitivity of the methods for three-dimensional data sets are clear. The propagation-based approach has the benefit that little or no additional optics are required to complete the phase-imaging and retrieval system. Refinement of propagation-based methods (such as in holotomography) and advances in single-plane phase retrieval and single-step retrieval and reconstruction have been very successful and propagation-based methods now produce stunning three-dimensional images.

We note in closing that many papers published recently in the area of phase retrieval include in their introduction a comprehensive list of phase-retrieval techniques. The different approaches attempted in connection with tomography represent only a fraction of these methods. There will inevitably be an exploration of the wider family of phase-retrieval methods in relation to X-ray tomography. We mention differential interference contrast (Di Fabrizio *et al.*, 2002), Fourier transform holography (Eisebitt *et al.*, 2004), and Fresnel coherent diffractive imaging (Williams *et al.*, 2006) as a subset of methods that may find application here. In connection with exploring the wider field of phase retrieval should also come an evaluation of the strengths and weaknesses of the various methods in the context of tomography. Investigations along these lines will be crucial in advancing what is currently an exciting, albeit *ad hoc*, collection of results into a mature technique. Tomographic methods in phase retrieval are about a decade old. If we compare the current state of the field to the achievements of, say, electron microscopy ten years after it was initially developed, then the future for three-dimensional phase imaging is a bright one.

Acknowlegements

Both authors acknowledge the support of the Australian Research Council through its Fellowship schemes (Federation Fellowship for KAN and QEII Fellowship for AGP) and also through the Centre of Excellence scheme.

6.10 References

Allman, B.E., McMahon, P.J., Nugent, K.A., Paganin, D., Jacobson, D.L., Arif, M., and Werner, S.A. (2000). Imaging – phase radiography with neutrons. *Nature*, **408**, 158.

Arhatari, B.D., Nugent, K.A., Peele, A.G., and Thornton, J. (2005). Phase contrast radiography. II. imaging of complex objects. *Review of Scientific Instruments*, **76**, 6.

Baruchel, J., Buffiere, J.-Y., Cloetens, P., Di Michiel, M., Ferrie, E., Ludwig, W., Maire, E., and Salvo, L. (2006). Advances in synchrotron radiation microtomography. *Scripta Materialia*, **55**, 41.

Bates, R.H.T. (1982). Fourier phase problems are uniquely solvable in more than one dimension. I: Underlying theory. *Optik*, **61**, 247.

Bronnikov, A.V. (1999). Reconstruction formulas in phase-contrast tomography. *Optics Communications*, **171**, 239.

Bronnikov, A.V. (2002). Theory of quantitative phase-contrast computed tomography. *Journal of the Optical Society of America A*, **19**, 472.

Chapman, H.N., Barty, A., Marchesini, S., Noy, A., Hau-Riege, S.R., Cui, C., Howells, M.R., Rosen, R., He, H., Spence, J.C.H., Weierstall, U., Beetz, T., Jacobsen, C., and Shapiro, D. (2006). High-resolution ab initio three-dimensional X-ray diffraction microscopy. *Journal of the Optical Society of America A*, **23**, 1179.

Cloetens, P., Ludwig, W., Baruchel, J., Dyck, D. Van, Landuyt, J. Van, Guigay, J. P., and Schlenker, M. (1999). Holotomography: Quantitative phase tomography with micrometer resolution using hard synchrotron radiation X-rays. *Applied Physics Letters*, **75**, 2912.

David, C., Nohammer, B., Solak, H.H., and Ziegler, E. (2002). Differential X-ray phase contrast imaging using a shearing interferometer. *Applied Physics Letters*, **81**, 3287.

Di Fabrizio, E., Kaulich, B., Wilhein, T., and Susini, J. (2002). Differential interference contrast for X-ray microscopy: Fabrication and characterization of twin zone plate optics. *Surface Review and Letters*, **9**, 243.

Dilmanian, F.A., Zhong, Z., Ren, B., Wu, X.Y., Chapman, L.D., Orion, I., and Thomlinson, W.C. (2000). Computed tomography of X-ray index of refraction using the diffraction enhanced imaging method. *Physics in Medicine and Biology*, **45**, 933.

Eisebitt, S., Luning, J., Schlotter, W. F., Lorgen, M., Hellwig, O., Eberhardt, W., and Stohr, J. (2004). Lensless imaging of magnetic nanostructures by X-ray spectro-holography. *Nature*, **432**, 885.

Elser, V. (2003). Phase retrieval by iterative projections. *Journal of the Optical Society of America A*, **20**, 40.

Fiedler, S., Bravin, A., Keyrilainen, J., Fernandez, M., Suortti, P., Thomlinson, W., Tenhunen, M., Virkkunen, P., and Karjalainen-Lindsberg, M.L. (2004). Imaging lobular breast carcinoma: Comparison of synchrotron radiation DEI-CT technique with clinical CT, mammography and histology.

Physics in Medicine and Biology, **49**, 175.

Fienup, J.R. (1982). Phase retrieval algorithms: A comparison. *Applied Optics*, **21**, 2758.

Forster, E., Goetz, K., and Zaumseil, P. (1980). Double crystal diffractometry for the characterization of targets for laser fusion experiments. *Kristall und Technik – Crystal Research and Technology*, **15**, 937.

Groso, A., Abela, R., and Stampanoni, M. (2006). Implementation of a fast method for high resolution phase contrast tomography. *Optics Express*, **14**, 8103.

Guigay, J.-P. (1977). Fourier transform analysis of Fresnel diffraction patterns. *Optik*, **49**, 121.

Guigay, J.-P., Pagot, E., and Cloetens, P. (2007). Fourier optics approach to X-ray analyser-based imaging. *Optics Communications*, **270**, 180.

Guigay, J.-P., Wade, R.H., and Delpha, C. (1971). Optical diffraction of Lorentz microscope images. In *Proceedings of the 25th meeting of the Electron Microscopy and Analysis Group* (ed. W. Nixon), London, pp. 238. The Institute of Physics.

Gureyev, T.E., Paganin, D.M., Myers, G.R., Nesterets, Y.I., and Wilkins, S.W. (2006). Phase-and-amplitude computer tomography. *Applied Physics Letters*, **89**, 3.

Hecht, E. (2001). *Optics* (4^{th} edn). Addison Wesley, New York.

Huang, Z.F., Kang, K.J., and Li, Z. (2006). Refraction-angle resolution of diffraction enhanced imaging. *Physics in Medicine and Biology*, **51**, 3031.

Maksimenko, A., Ando, M., Hiroshi, S., and Yuasa, T. (2005). Computed tomographic reconstruction based on X-ray refraction contrast. *Applied Physics Letters*, **86**, 3.

Mayo, S.C., Davis, T.J., Gureyev, T.E., Miller, P.R., Paganin, D., Pogany, A., Stevenson, A.W., and Wilkins, S.W. (2003). X-ray phase-contrast microscopy and microtomography. *Optics Express*, **11**, 2289.

Mayo, S.C., Miller, P.R., Wilkins, S.W., Davis, T.J., Gao, D., Gureyev, T.E., Paganin, D., Parry, D.J., Pogany, A., and Stevenson, A.W. (2002). Quantitative X-ray projection microscopy: Phase-contrast and multi-spectral imaging. *Journal of Microscopy*, **207**, 79.

McMahon, P.J., Peele, A.G., Paterson, D., Nugent, K.A., Snigirev, A., Weitkamp, T., and Rau, C. (2003). X-ray tomographic imaging of the complex refractive index. *Applied Physics Letters*, **83**, 1480.

Miao, J.W., Charalambous, P., Kirz, J., and Sayre, D. (1999). Extending the methodology of X-ray crystallography to allow imaging of micrometre-sized non-crystalline specimens. *Nature*, **400**, 342.

Miao, J.W., Ishikawa, T., Johnson, B., Anderson, E.H., Lai, B., and Hodgson, K.O. (2002). High resolution 3D X-ray diffraction microscopy. *Physical Review Letters*, **89**, 4.

Modregger, P., Lubbert, D., Schafer, P., and Köhler, R. (2006). Magnified X-ray phase imaging using asymmetric Bragg reflection: Experiment and theory.

Physical Review B, **74**, 10.

Momose, A. (2003). Phase-sensitive imaging and phase tomography using X-ray interferometers. *Optics Express*, **11**, 2303.

Nesterets, Y.I., Gureyev, T.E., Paganin, D., Pavlov, K.M., and Wilkins, S.W. (2004). Quantitative diffraction-enhanced X-ray imaging of weak objects. *Journal of Physics D: Applied Physics*, **37**, 1262.

Nugent, K.A., Peele, A.G., Chapman, H.N., and Mancuso, A.P. (2003). Unique phase recovery for nonperiodic objects. *Physical Review Letters*, **91**, 4.

Paganin, D. (2006). *Coherent X-ray optics*. Oxford University Press, Oxford.

Paganin, D., Mayo, S.C., Gureyev, T.E., Miller, P.R., and Wilkins, S.W. (2002). Simultaneous phase and amplitude extraction from a single defocused image of a homogeneous object. *Journal of Microscopy*, **206**, 33.

Paganin, D. and Nugent, K.A. (2001). Noninterferometric phase determination. In *Advances in imaging and electron physics* (ed. P. Hawkes, B. Kazan, and T. Mulvey), Volume 118, pp. 85. Academic Press Inc., San Diego.

Pagot, E., Fiedler, S., Cloetens, P., Bravin, A., Coan, P., Fezzaa, K., Baruchel, J., and Hartwig, J. (2005). Qualitative comparison between two phase contrast techniques: diffraction enhanced imaging and phase propagation imaging. *Physics in Medicine and Biology*, **50**, 709.

Peele, A.G., Carlo, F. De, McMahon, P.J., Dhal, B.B., and Nugent, K.A. (2005). X-ray phase contrast tomography with a bending magnet source. *Review of Scientific Instruments*, **76**, 5.

Pfeiffer, F., Weitkamp, T., Bunk, O., and David, C. (2006). Phase retrieval and differential phase-contrast imaging with low-brilliance X-ray sources. *Nature Physics*, **2**, 258.

Pogany, A., Gao, D., and Wilkins, S.W. (1997). Contrast and resolution in imaging with a microfocus X-ray source. *Review of Scientific Instruments*, **68**, 2774.

Raven, C., Snigirev, A., Snigireva, I., Spanne, P., Souvorov, A., and Kohn, V. (1996). Phase-contrast microtomography with coherent high-energy synchrotron X-rays. *Applied Physics Letters*, **69**, 1826.

Sayre, D. (1952). Some implications of a theory due to Shannon. *Acta Crystallographica*, **5**, 843.

Schmahl, G., Rudolph, D., and Guttmann, P. (1988). Phase contrast X-ray microscopy - experiments at the BESSY storage ring. In *X-ray microscopy II* (ed. D. Sayre, M. Howells, J. Kirz, and H. Rarback), Volume 56 of *Springer Series in Optical Sciences*, pp. 228. Springer-Verlag, Berlin.

Settles, G. (2001). *Schlieren and shadowgraph techniques: Visualizing phenomena in transparent media*. Springer-Verlag, Berlin.

Snigirev, A., Snigireva, I., Kohn, V., Kuznetsov, S., and Schelokov, I. (1995). On the possibilities of X-ray phase contrast microimaging by coherent high-energy synchrotron radiation. *Review of Scientific Instruments*, **66**, 5486.

Tafforeau, P., Boistel, R., Boller, E., Bravin, A., Brunet, M., Chaimanee, Y., Cloetens, P., Feist, M., Hoszowska, J., Jaeger, J.J., Kay, R.F., Lazzari, V.,

Marivaux, L., Nel, A., Nemoz, C., Thibault, X., Vignaud, P., and Zabler, S. (2006). Applications of X-ray synchrotron microtomography for non-destructive 3D studies of paleontological specimens. *Applied Physics A: Materials Science and Processing*, **83**, 195.

Takeda, M., Ina, H., and Kobayashi, S. (1982). Fourier-transform method of fringe-pattern analysis for computer-based topography and interferometry. *Journal of the Optical Society of America*, **72**, 156.

Takeda, T., Momose, A., Hirano, K., Haraoka, S., Watanabe, T., and Itai, Y. (2000). Human carcinoma: Early experience with phase-contrast X-ray CT with synchrotron radiation - comparative specimen study with optical microscopy. *Radiology*, **214**, 298.

Tkachuk, A., Feser, M., Cui, H., Duewer, F., Chang, H., and Yun, W. (2006). High-resolution X-ray tomography using laboratory sources. In *Developments in X-ray tomography V* (ed. U. Bonse), pp. 63181D. The International Society for Optical Engineering (SPIE). Proceedings of the SPIE, Vol. 6318.

Turner, L.D., Dhal, B.B., Hayes, J.P., Mancuso, A.P., Nugent, K.A., Paterson, D., Scholten, R.E., Tran, C.Q., and Peele, A.G. (2004). X-ray phase imaging: Demonstration of extended conditions with homogeneous objects. *Optics Express*, **12**, 2960.

Wang, J.Y., Zhu, P.P., Yuan, Q.X., Huang, W.X., Shu, H., Chen, B., Hu, T.D., and Wu, Z.Y. (2006). Reconstruction of the refractive index gradient by X-ray diffraction enhanced computed tomography. *Physics in Medicine and Biology*, **51**, 3391.

Weitkamp, T., Diaz, A., David, C., Pfeiffer, F., Stampanoni, M., Cloetens, P., and Ziegler, E. (2005). X-ray phase imaging with a grating interferometer. *Optics Express*, **13**, 6296.

Wilkins, S.W., Gureyev, T.E., Gao, D., Pogany, A., and Stevenson, A.W. (1996). Phase-contrast imaging using polychromatic hard X-rays. *Nature*, **384**, 335.

Williams, G.J., Quiney, H.M., Dhal, B.B., Tran, C.Q., Nugent, K.A., Peele, A.G., Paterson, D., and de Jonge, M.D. (2006). Fresnel coherent diffractive imaging. *Physical Review Letters*, **97**, 4.

Wu, X.Z., Liu, H., and Yan, A.M. (2005). X-ray phase-attenuation duality and phase retrieval. *Optics Letters*, **30**, 379.

Zabler, S., Cloetens, P., Guigay, J.P., Baruchel, J., and Schlenker, M. (2005). Optimization of phase contrast imaging using hard X-rays. *Review of Scientific Instruments*, **76**, 7.

Zhu, P.P., Wang, J.Y., Yuan, Q.X., Huang, W.X., Shu, H., Gao, B., Hu, T.D., and Wu, Z.Y. (2005). Computed tomography algorithm based on diffraction-enhanced imaging setup. *Applied Physics Letters*, **87**, 3.

7

TOMOGRAPHY USING MAGNIFYING OPTICS

Yoshio Suzuki and Hiroyuki Toda (7.1)
Christian Schroer (7.2)

The main limitation of spatial resolution for conventional X-ray tomography as described in Chapter 5 is usually determined by the spatial resolution of the image detector. Although high-resolution photoresists for X-ray lithography have spatial resolutions better than 100 nm these cannot be applied to tomography due to their non-linearity and because real-time processing is needed for tomography. Scintillator materials that have these properties in the hard X-ray range are very difficult to develop for spatial resolutions well below a micrometre, see Chapter 10.

Conventional tomographic reconstruction is usually based on a backprojection algorithm assuming the geometrical optics approximation. This assumption is generally valid for medical CT scanners because of the short wavelength used ($\sim$0.02 nm) and the low spatial resolution ($\sim$ mm) required. However, the geometrical approximation becomes invalid for high-resolution tomography because Fresnel diffraction of X-rays can no longer be neglected. Fresnel diffraction patterns scale approximately as $(\lambda L)^{\frac{1}{2}}$, where λ is the X-ray wavelength. L represents the distance between sample and detector in the case of contact microscopy and corresponds to the distance from the source point to the sample in the case of projection imaging with a spherical wave emerging from a small point source, i.e. cone-beam tomography with high magnification. When $\lambda = 0.1$ nm and $L = 1$ cm the size of interference fringes becomes 1 µm. Therefore, high-resolution tomography is difficult, if simple projection imaging is used. Tomographic microscopy in combination with holographic image reconstruction exploits Fresnel diffraction effects to reconstruct the complex refractive index of the object.

Based on hard X-ray optics two types of full-field microscopes are currently used, shown schematically in Fig. 7.1. The first scheme (a) requires the generation of a small secondary (point) source that is used to illuminate the object with a divergent beam, projecting its magnified transmission image onto a two-dimensional position-sensitive detector. The second scheme (b) is based on magnifying imaging of the object onto the two-dimensional detector using appropriate X-ray optics. In both cases, high-resolution X-ray optics are needed to improve the spatial resolution to the submicrometre level.

The use of X-ray imaging optics for X-ray optical magnification – scheme (b) – has been shown to be a practical way to overcome this limitation in spatial resolution (Lengeler *et al.*, 1999*a*, Takeuchi *et al.*, 2001, Takeuchi *et al.*, 2002).

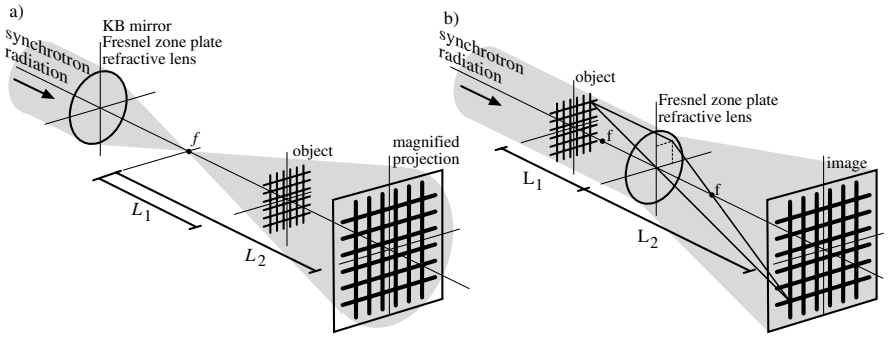


FIG. 7.1. Schematic diagram of (a) hard X-ray magnifying projection and (b) full-field magnifying imaging microscopy.

By combining X-ray microscopy optics with tomographic imaging, the effects of Fresnel diffraction and refraction phenomena at the interface can be largely eliminated. In this case spatial resolution is limited by the resolution of the optic, its depth of focus, and by wave-propagation effects inside the sample. Image blurring caused by beam deflection inside the object is now expressed as $(\lambda d)^{\frac{1}{2}}$, where d represents the thickness of the specimen. In addition, the thickness of the specimen is limited by the depth of focus of the objective lens, and the tomographic reconstruction based on the conventional backprojection algorithm is only valid within this limitation.

A number of different optical systems have been tested for imaging and at least three such systems have been shown to work. Mirrors can be used to focus X-rays if they are appropriately shaped with sufficient precision. Refractive lenses for X-rays can also be used for this purpose, although the refractive index of X-rays is close to unity (Lengeler *et al.*, 1999*a*), and tomographic imaging has been successfully demonstrated (Schroer *et al.*, 2002).

Fresnel zone-plate (FZP) optics are widely used in X-ray microscopy, and spatial resolutions better than 100 nm have already been achieved in the hard X-ray region with FZP objectives (Takano *et al.*, 2003, Suzuki *et al.*, 2005). The imaging properties of such optics make them suitable for tomographic microscopy, because the offaxis aberrations of FZP optics are usually negligible in the hard X-ray region. Although these FZP optics are well suited in soft X-ray microscopy (Weiss *et al.*, 2000), they tend to lose their efficiency at higher X-ray energies ($E > 2\,\mathrm{keV}$). In recent years, the energy range for FZPs has been steadily extended towards harder X-rays ($E \approx 10\,\mathrm{keV}$), reaching into the lower range of operation of reflective and refractive X-ray optics, such as mirrors and refractive lenses – the other type of optics used for magnifying X-ray images. These systems are the choice for energies higher than 10 keV and complement FZP optics.

The principles of FZP optics and imaging properties of FZP microscopy are

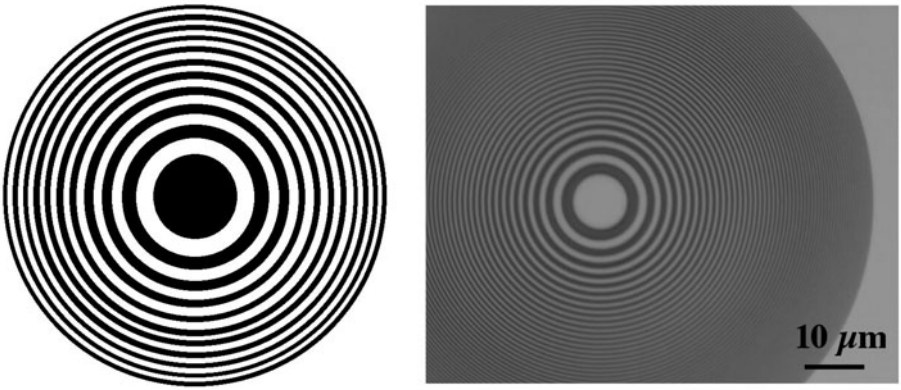


FIG. 7.2. Fresnel zone plate, schematic drawing and optical micrograph. The FZP shown in the optical micrograph has a diameter of 100 μm and a width of the outermost zone of 250 nm. This FZP was fabricated by electron beam lithography and is used at the SPring-8 synchrotron radiation beamlines as an objective lens for the X-ray microscope and X-ray microfocusing device (Suzuki *et al.*, 2001, Takeuchi *et al.*, 2002).

described in the context of tomographic application in Section 7.1. The limitation of zone-plate optics is discussed, and some typical examples of tomographic imaging experiment are shown.

In Section 7.2 a brief introduction to reflective and refractive X-ray optics is given since they are used for X-ray microscopy in the hard X-ray range relevant for materials research.

A further possibility to magnify images is by using asymmetric Bragg reflection (Schäfer and Köhler, 2003). This method will not be discussed here since it is uncertain whether it can be applied to problems of materials research and seems to have more application potential in the area of biology, hence being beyond the scopee of this book.

7.1 Fresnel zone-plate microscopy and microtomography

7.1.1 *Fresnel zone-plate for hard X-rays*

7.1.1.1 *Principle* A Fresnel zone plate (FZP) is a concentric transmission grating with a radially decreasing grating period, as shown in Fig. 7.2 schematically. To explain the principle of zone-plate optics it is useful to start from an equidistant grating. As shown in Fig. 7.3, the beam deflection angle at a grating with a period of d is described by $\sin\theta = \nu\lambda/d$, where $\nu = 0, \pm1, \pm2, \pm3, \ldots$. The integer ν is called the *order of diffraction*. $\nu = 0$ characterizes the undiffracted beam. When the grating period is changed to $d = \lambda/\sin\Theta$, as shown in Fig. 7.4, all the diffracted rays are focused at a point, where Θ is determined by $\tan\Theta = r/f$, and r is the radial position on the circular grating, and f is a specific distance

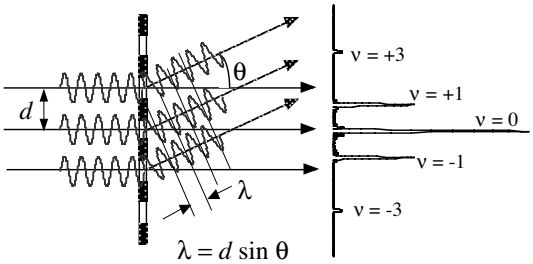


FIG. 7.3. Equidistant grating and diffraction pattern at grating.

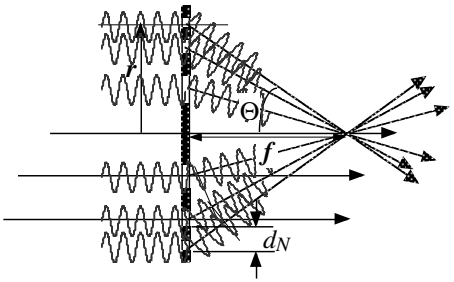


FIG. 7.4. Diffraction at Fresnel zone plate (schematic diagram).

from the zone plate. Thus, the zone plate focuses X-rays as a convex lens focuses visible light and it simultaneously defocuses X-rays as a concave lens does with visible light. The zeroth order and higher-order diffractions usually exist simultaneously. Therefore, a kind of spatial filtering is needed to ensure the proper function of the lens.

The most suitable FZP structure can be derived from optical path difference equations. Let us consider the zone plate that consists of alternating transparent and fully opaque zones. Fully constructive interference occurs, i.e. all the wavelets through each zone are in phase at the focal point, when the optical path difference between neighbouring transparent zones is one wavelength. Then, the νth zone boundary of a Fresnel zone plate that focuses the spherical wave emitted from point A to the point B is characterized by the following equation,

$$(R_a + R_b) = \frac{\nu\lambda}{2} + (a + b),$$

as shown in Fig. 7.5, and

$$(a^2 + r_\nu^2)^{\frac{1}{2}} + (b^2 + r_\nu^2)^{\frac{1}{2}} - (a + b) = \frac{\nu\lambda}{2}, \tag{7.1}$$

where a is the distance from point A to the FZP, b is the distance between B and FZP, r_ν is the boundary of the νth zone, and ν is an integer. Using a Taylor

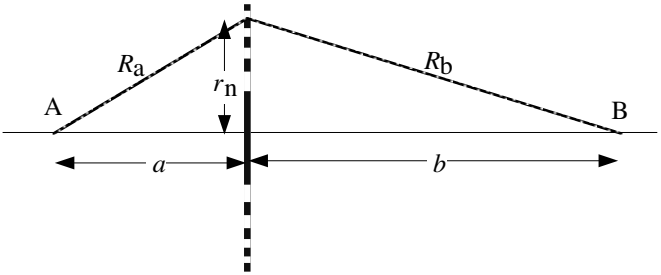


FIG. 7.5. Optical-path description of Fresnel zone plate.

expansion and omitting higher-order terms eqn 7.1 can be rewritten because $a \gg r_\nu$ and $b \gg r_\nu$:

$$a + \frac{r_\nu^2}{2a} + b + \frac{r_\nu^2}{2b} - (a+b) = \frac{\nu\lambda}{2}. \tag{7.2}$$

Then,

$$\frac{r_\nu^2}{a} + \frac{r_\nu^2}{b} = \nu\lambda. \tag{7.3}$$

By using $1/a + 1/b = 1/f$ – the well-known *Newton lens equation* – eqn 7.3 is rewritten as

$$r_\nu^2 = \nu\lambda f. \tag{7.4}$$

Thus, a FZP whose zone boundary is defined by the above equation works as a lens with a focal length f. The focal length f varies inversely with the X-ray wavelength. Therefore, the use of monochromatic X-rays is mandatory for FZP microscopes.

The spatial resolution of optical imaging is essentially determined by the diffraction of light. The diffraction-limited resolution Δ for a microscope with an objective lens of circular aperture (axisymmetric optics) and incoherent illumination is expressed by the well-known *Rayleigh criterion* as

$$\Delta = 0.61\frac{\lambda}{\mathrm{NA}}, \tag{7.5}$$

where NA is the so-called *numerical aperture* of the objective lens, defined as $\mathrm{NA} = n\sin\theta \approx \sin\theta$, since the index of refraction $n \approx 1$. Here, θ is the angle between an outermost 'marginal' ray and the optical axis, as shown in Fig. 7.6. It is also apparent from the diffraction of the linear grating that $\sin\theta$ is equal to λ/d, where d_N is the period of the outermost zone of the FZP, see Fig. 7.4. Then,

$$\Delta = 0.61 d_N. \tag{7.6}$$

The formula $\Delta = 1.22 dr_N$ is frequently used instead of eqn 7.6, where $dr_N = dN/2$ is the half-pitch of the outermost Nth zone period (or width of outermost zone).

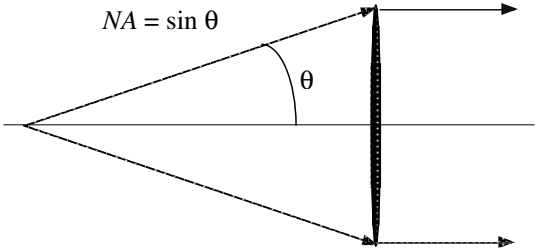


FIG. 7.6. Numerical aperture of objective lens ($n \approx 1$).

The diffraction efficiency is an important property of diffraction-based optics. From simple calculations it is known that the νth-order diffraction efficiency for alternating transparent and opaque zones, i.e. 'black and white' zones, with a 1:1 zone width ratio is $1/(\nu\pi)^2$ ($\approx 10\%$ for first-order diffraction), where $\nu = \pm 1, \pm 3, \pm 5, \ldots$ (even-order diffraction does not exit for such 1:1 gratings). More efficient gratings are realized by modulating the phase of the beam instead of merely stopping the beam with a completely opaque zone. The maximum diffraction efficiency is attained for a phase shift of a half-wavelength. The efficiency of the νth-order diffraction for an ideal phase-modulated phase grating of 1:1 zone ratio is $4/(\nu\pi)^2$, where $\nu = \pm 1, \pm 3, \pm 5, \ldots$), leading to an efficiency of about 40% for first-order diffraction. However, in the X-ray region, pure-phase materials do not exist. All media show both phase and absorption effects and the efficiency is lower in practice.

7.1.1.2 *Fabrication of zone plates* Most zone plates for X-ray microscopes are fabricated by electron-beam lithography, a technology developed in the semiconductor industry. Recent large-scale integration technology makes it possible to fabricate sub-100 nm microstructures on silicon surfaces. This is a key technology also in X-ray microscopy. A schematic drawing of a lens cross-section is shown in Fig. 7.7(a) together with a SEM micrograph in Fig. 7.7(b). The zone-plate pattern is drawn by an electron beam onto the resist on a thin membrane supported by a silicon wafer. The thin membrane is usually a few μm thick and consists of silicon nitride or silicon carbide. The pattern transfer from the photoresist to the zone material is done by dry etching or wet electroplating. Finally, the silicon wafer in the pattering area is removed by chemical etching. The difficulty in the fabrication of X-ray zone plates is the high aspect ratio (zone height/zone width) of the zone structures at the marginal zone area.

For X-rays, the phase shift for the materials used dominates over the absorptive contribution, i.e. $\beta/\delta \ll 1$ in eqn 4.12. Therefore, most of the FZPs are designed as phase-modulation plates. Assuming the free-electron approximation the deviation of the real part of the index of refraction n from unity can be written as:

$$\delta = 2.7 \times 10^{-6} \frac{Z}{A} \rho \ [\mathrm{g/cm^3}] \ \lambda[\text{Å}]^2, \tag{7.7}$$

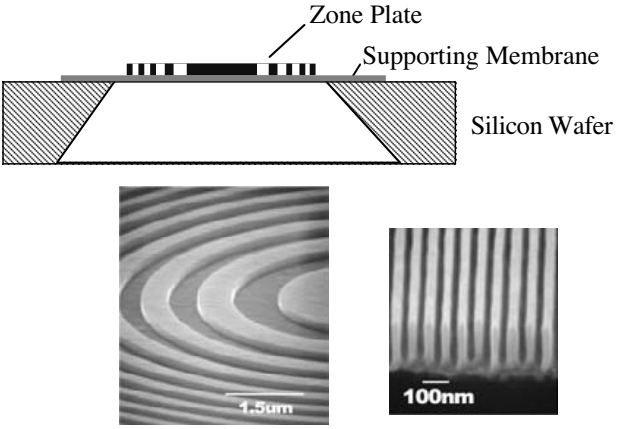


FIG. 7.7. Structure of Fresnel zone plate shown as (a) cross-sectional view, (b) SEM micrograph. Courtesy of NTT Advanced Technology.

where ρ is the density of the material, Z the atomic number and A the atomic mass number (to be inserted in the units specified in brackets [...]), see also eqn 4.15. The optimum thickness t of the zone plate is equivalent to the optical path difference of $\lambda/2$, that is calculated by $t\delta = \lambda/2$. Accordingly, highly dense materials are preferable for zone plates. Considering the manufacturing process of the microstructures, gold or tantalum is usually chosen as a zone material for hard X-ray FZPs. The optimized thickness t and δ e.g. for tantalum zone plates, is 2.9 μm and 1.71×10^{-5} at an X-ray wavelength of 0.1 nm, respectively. The aspect ratio should be 29 for the FZP having 100 nm zone width. This value is still difficult to achieve even by the present nanofabrication technologies.

The number of zones N is usually 100 to 1000. It is known that when the number of zones is greater than 100 the imaging property of an FZP is nearly equal to that of conventional lenses (Michette, 1986). Although larger N are preferable from the point of view of numerical aperture and field of view, there is a limitation of N given by the aberration of the optical system. This limitation is discussed below.

7.1.2 *Optical system and imaging properties of FZP microscope*

7.1.2.1 *Order selection by spatial filter and illuminating optics for FZP objective*

As described in the previous section the FZP cannot be used as a conventional refractive lens because of the presence of undesirable diffractions. A simple way of diffraction-order selection is shown in Fig. 7.8. Object and objective are illuminated by a planar wave with a cross-section limited by a diaphragm and only one half of the FZP objective is illuminated. In this case positive and negative diffraction orders are well separated in the image plane and the direct beam appears at the boundary between positive and negative diffraction patterns. The optical system shown in Fig. 7.8 is called *offaxis illumination*. The configuration

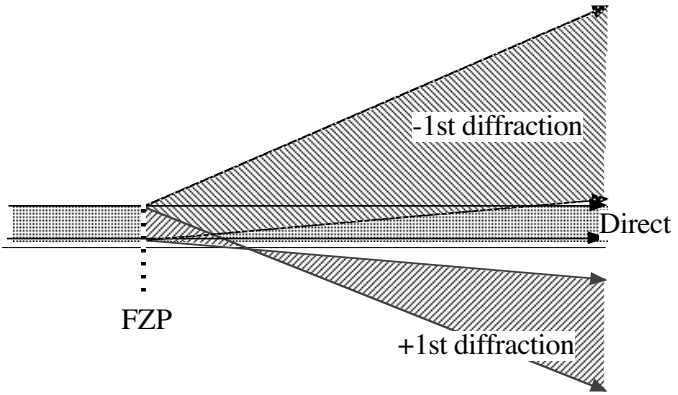


FIG. 7.8. Offaxis illumination: order selection for plane-wave illumination. Only half of the FZP aperture is illuminated.

shown in the figure cannot separate different orders of diffraction, which would require additional spatial filters. However, higher-order diffraction has a much lower efficiency and a larger magnification than low-order diffraction in the image plane and can therefore usually be ignored. A serious problem of offaxial illumination is the non-uniformity of imaging, i.e. a strong edge enhancement in the outer region of the image field. Therefore, spatial filtering in combination with condenser optics is more commonly used in imaging microscopy with FZP objectives. There are actually two types of illumination optics for imaging microscopy, namely *critical illumination* and *Köhler illumination*. The schematic diagrams of critical illumination and Köhler illumination are shown in Fig. 7.9(a) and Fig. 7.9(b), respectively. Critical illumination is not suitable for synchrotron radiation facilities because these sources are very small and have a small divergence. Using critical illumination a demagnified image of the source is formed at the object plane by the condenser lens and the typical field of view becomes only a few μm. In contrast, the condenser optics for Köhler illumination does not generate a demagnified image. The condenser lens optics produce a constant angle deflection and the deflected beams are merged together in the field of view of the objective lens. Köhler illumination results in a hollow-cone beam behind the objective, as shown in Fig. 7.9(b). This condenser optics is realized by using a constant-period zone plate with a centre beam stop, i.e. not a FZP, or by using a double rotating mirror (Nieman *et al.*, 2000). It is important that diffraction-order selection is naturally achieved when hollow-cone illumination is used. Only the first-order diffraction appears in the centre circle defined by the hollow-cone beam.

7.1.2.2 *Spatial resolution of FZP microscopy with condenser optics* In the previous section the spatial resolution limit imposed by diffraction theory was discussed. However, when condenser lens optics are used the spatial resolution is

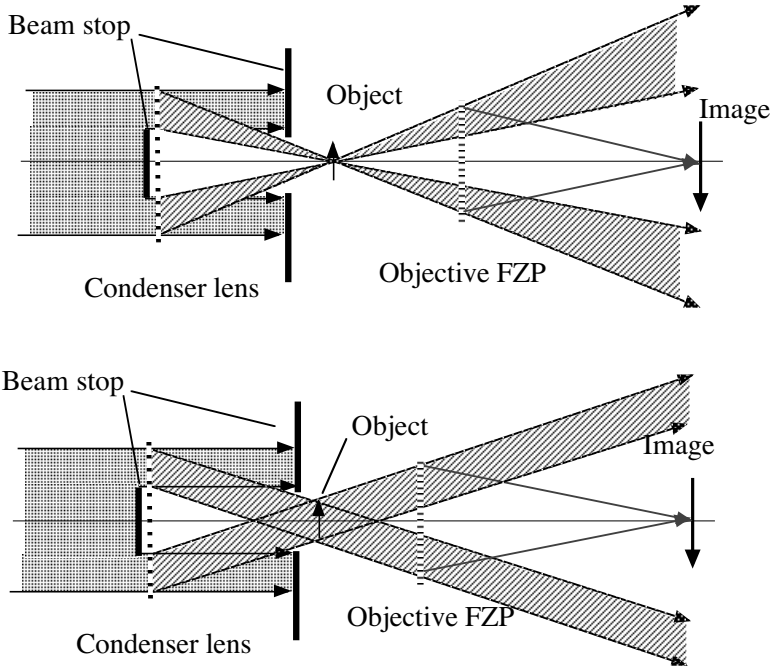


FIG. 7.9. (a) Critical illumination with condenser zone plate and objective zone plate. Hollow-cone illumination is used for diffraction-order selection. (b) Köhler's illumination with condenser zone plate. Hollow-cone illumination is also achieved by combination with a centre beam stop and an aperture in front of the object.

determined not only by the numerical aperture NA of the objective but also by the convergent angle of the illuminating beam (Born and Wolf, 1980). This is called *matching of aperture*. When the convergent angle is equivalent to the NA of the objective lens, spatial resolution is equal to that defined by eqn 7.5. The best resolution is achieved when the NA of the condenser is 1.5 times that of the objective lens. Under these conditions the spatial resolution becomes $\approx 0.57\lambda/\mathrm{NA}$. When no condenser optics is used, i.e. in the case of planar-wave illumination, the spatial resolution is written as $0.82\lambda/\mathrm{NA}$ instead of $0.61\lambda/\mathrm{NA}$.

7.1.2.3 *Depth of focus* One of the most important features of FZP optics for imaging is the depth of focus because in tomography the sample thickness should be nearly equal to the field of view. Although the depth of focus in hard X-ray microscopes is generally much greater than in optical microscopes the restrictions given by the depth of focus are sometimes not negligible. The depth of focus can be evaluated by geometrical optics as shown in Fig. 7.10. When a point source is placed at a distance D from the object plane the image size in the object plane

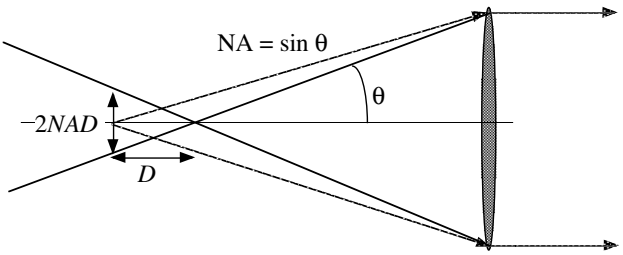


FIG. 7.10. Depth of focus. The full line represents the primary focus, the broken line is the marginal path of the defocused rays.

is expressed as $2D\times$NA. The amount of defocusing, $2D\times$NA, should be less than or equal to the diffraction-limited resolution in order to neglect the influence on spatial resolution. By taking both sides of the object plane into account the depth of focus, $2D$, is expressed by

$$d_l = 2D = 0.61\frac{\lambda}{\mathrm{NA}^2} = \frac{\Delta}{\mathrm{NA}}. \tag{7.8}$$

In other words, the NA of the objective determines the ratio between spatial resolution and field of view. As an example, when $\lambda = 0.1$ nm, and NA$= 1 \times 10^{-3}$, the diffraction-limited resolution is 61 nm and the limit of sample size is 61 μm in diameter. This is a critical limitation of imaging microtomography based on the conventional backprojection algorithm. It should be noted that the restriction of the depth of focus expressed by eqn 7.8 is essentially the same as the ultimate limitation on spatial resolution for finite sample thickness, $\Delta \approx (\lambda L)^{\frac{1}{2}}$, where L is the thickness of the sample, as already mentioned in the introduction to this chapter.

7.1.2.4 *Chromatic and spherical aberration* In the above discussion a perfectly monochromatic X-ray beam was assumed. Any real X-ray beam has a finite bandwidth, even if a crystal monochromator is used. The influence of the finite bandwidth of the incident X-ray beam is called *chromatic aberration*. The defocusing effect caused by chromatic aberration is evaluated in a similar manner to that of the depth of focus:

$$\Delta f \times \mathrm{NA} < 0.61\frac{\lambda}{\mathrm{NA}}. \tag{7.9}$$

Here, Δf is a displacement of focus caused by chromatic aberration. The focal length is defined by $r_\nu^2 = \nu\lambda f$. r_ν is constant for a given FZP, and the focal length f is proportional to $1/\lambda$. Then, the displacement of focus Δf is written as

$$\frac{\Delta f}{f} = \frac{\Delta\lambda}{\lambda},$$

where $\Delta\lambda$ is the bandwidth of the incident radiation. Equation 7.9 can be rewritten as

$$\frac{\Delta\lambda}{\lambda} f \times \mathrm{NA} < 0.61 \frac{\lambda}{\mathrm{NA}}.$$

By using an approximation for NA$\approx r_N/f$ and $r_N^2 = N\lambda f$, where N is the zone number of the outermost zone a useful formula is obtained:

$$\frac{\Delta\lambda}{\lambda} < \frac{0.61}{N}.$$

This formula gives a tolerance for the bandwidth of incident X-ray beams in a FZP microscope.

Spherical aberration comes from the approximation used to derive eqn 7.2. By including the second-order term the equation is approximately rewritten as

$$\frac{r_\nu^2}{2a} - \frac{1}{8}\frac{r_\nu^4}{a^3} + \frac{r_\nu^2}{2b} - \frac{1}{8}\frac{r_\nu^4}{b^3} = \frac{\nu\lambda}{2}. \tag{7.10}$$

For an imaging microscope $a \ll b$, and $a \approx f$. Then, eqn 7.10 can be written as

$$\frac{r_\nu^2}{2a} + \frac{r_\nu^2}{2b} - \frac{1}{8}\frac{r_\nu^4}{f^3} = \frac{\nu\lambda}{2}. \tag{7.11}$$

In order to evaluate the wave-front aberration Rayleigh's quarter-wavelength criterion can be used (Born and Wolf, 1980). The $\lambda/4$ rule states that when the optical path difference from the primary path is within a quarter of a wavelength, the influence of wavefront aberration on spatial resolution may be ignored. Therefore, by comparing eqn 7.2 and eqn 7.11, the $\lambda/4$ rule for spherical aberration is expressed by

$$\frac{1}{8}\frac{r_N^4}{f^3} < \frac{\lambda}{4}.$$

Using the Fresnel zone-plate formula eqn 7.4, eqn 7.6 can be rewritten as

$$\frac{1}{2}N^2 < \frac{f}{\lambda}.$$

This formula also gives a limitation on the total number of zones N as N is limited by chromatic aberration. Spherical aberration can be perfectly eliminated by using eqn 7.1. However, the FZP designated by eqn 7.1 has to be used at a fixed wavelength and fixed optical distances a and b. This lack of flexibility is a disadvantage. Therefore, it is better to use the zone plate with a limited number of zones defined by the above formula to make spherical aberration negligible.

7.1.2.5 *Offaxis aberration theory based on wave optics* In this section the offaxis imaging properties of FZP microscopes are discussed. Using Fig. 7.11 the optical path equation is written as

$$\begin{aligned} &\left((r_a - r_\nu\cos\phi)^2 + (r_\nu\sin\phi)^2 + a^2\right)^{\frac{1}{2}} \\ &+ \left((r_b + r_\nu\cos\phi)^2 + (r_\nu\sin\phi)^2 + b^2\right)^{\frac{1}{2}} \\ &= \frac{\nu\lambda}{2} + (a^2 + r_a^2)^{\frac{1}{2}} + (b^2 + r_b^2)^{\frac{1}{2}}, \end{aligned} \tag{7.12}$$

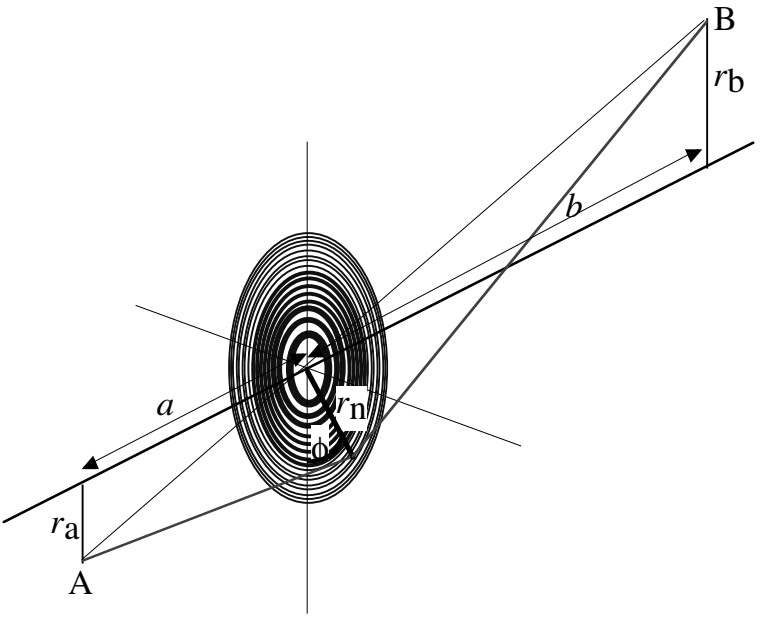


FIG. 7.11. Schematic diagram of offaxis optical path for FZP imaging optics.

where $r_a/a = r_b/b$, as shown in the figure, and $r_b/r_a = b/a$ represents the geometrical magnification of microscope. ϕ is defined in Fig. 7.11. The influence to spatial resolution is evaluated in a similar manner as that of spherical aberration. By using a second-order approximation, eqn 7.12 is rewritten as

$$\begin{aligned} &\frac{1}{2a}\left(r_a^2 - 2r_a r_\nu \cos\phi + r_\nu^2\right) - \frac{1}{8a^3}\left(r_a^2 - 2r_a r_\nu \cos\phi + r_\nu^2\right)^2 \\ &+\frac{1}{2b}\left(r_b^2 + 2r_b r_\nu \cos\phi + r_\nu^2\right) - \frac{1}{8b^3}\left(r_b^2 + 2r_b r_\nu \cos\phi + r_\nu^2\right)^2 \\ &\qquad = \frac{\nu\lambda}{2} + \frac{1}{2}\frac{r_a^2}{a} - \frac{1}{8}\frac{r_a^4}{a^3} + \frac{1}{2}\frac{r_b^2}{b} - \frac{1}{8}\frac{r_b^4}{b^3} \end{aligned} \tag{7.13}$$

The wave-front aberrations are analysed by the optical path given by eqn 7.13, while Rayleigh's quarter-wavelength rule is expressed as

$$\begin{aligned} &\left|\left(-\frac{r_a r_\nu \cos\phi}{a} + \frac{r_b r_\nu \cos\phi}{b}\right) + \frac{1}{2}\left(\frac{r_\nu^2}{a} + \frac{r_\nu^2}{b} - \nu\lambda\right)\right. \\ &\left. -\frac{1}{8a^3}\left(\left(r_a^2 - 2r_a r_\nu \cos\phi + r_\nu^2\right)^2 - r_a^4\right) + \frac{1}{8b^3}\left(\left(r_b^2 + 2r_b r_\nu \cos\phi + r_\nu^2\right)^2 - r_b^4\right)\right| \\ &\qquad < \frac{\lambda}{4}. \end{aligned}$$

The first term represents the definition of magnification $r_a/a = r_b/b$, the second term corresponds to the basic zone-plate equation, $r_\nu^2 = \nu\lambda f$. The residual aberration is rewritten to

$$\frac{1}{8}\left|\frac{1}{a^3}\left(-4r_a^3 r_\nu\cos\phi + 2r_a^2 r_\nu^2 + (2r_a r_\nu\cos\phi)^2 - 4r_a r_\nu^3\cos\phi + r_\nu^4\right)\right.$$
$$\left. + \frac{1}{b^3}\left(4r_b^3 r_\nu\cos\phi + 2r_b^2 r_\nu^2 + (2r_b r_\nu\cos\phi)^2 + 4r_b r_\nu^3\cos\phi + r_\nu^4\right)\right| \qquad (7.14)$$
$$< \frac{\lambda}{4}.$$

The aberration that includes the first order of r_N can also be ignored, because the residual aberration is written as

$$\frac{1}{8}\left|\frac{4r_a^3 r_N\cos\phi}{a^3} - \frac{4r_b^3 r_N\cos\phi}{b^3}\right| < \frac{\lambda}{4}.$$

This term is always zero, because $r_a/a = r_b/b$. The residual offaxis aberration that contains the square of r_N is extracted from eqn 7.14 as

$$\frac{1}{8}\left|\frac{1}{a^3}\left((2r_a r_N\cos\phi)^2 + 2r_a^2 r_N^2\right) + \frac{1}{b^3}\left((2r_b r_N\cos\phi)^2 + 2r_b^2 r_N^2\right)\right| < \frac{\lambda}{4}.$$

By using $-1 \le \cos\phi \le +1$, and $a \ll b$ and $a \approx f$, the following expression is derived.

$$\frac{r_a^2 r_N^2}{f^3} < \frac{\lambda}{3}. \qquad (7.15)$$

Concerning the third order of r_N, the following formula is obtained from eqn 7.14.

$$\frac{1}{8}\left|-\frac{4r_a r_N^3\cos\phi}{a^3} + \frac{4r_b r_N^3\cos\phi}{b^3}\right| < \frac{\lambda}{4}.$$

This expression may also be reduced to

$$\frac{r_a r_N^3}{f^3} < \frac{\lambda}{2}. \qquad (7.16)$$

Concerning the fourth-order term of r_N, the residual aberration is reduced to

$$\frac{1}{8}\left|\frac{r_N^4}{a^3} + \frac{r_N^4}{b^3}\right| < \frac{\lambda}{4}.$$

This formula is also approximated as

$$\frac{1}{8}\frac{r_N^4}{f^3} < \frac{\lambda}{4}.$$

This is the spherical aberration discussed in the previous section. These formulae give a restriction on the maximum field of view for a given FZP to achieve a

diffraction-limited resolution. By comparing eqn 7.15 and eqn 7.16 with eqn 7.8, it is apparent that the above conditions are generally well satisfied when the field of view is less than the depth of focus. Therefore, the Fresnel zone plate can be used as an objective lens for X-ray tomographic microscopes without considering the offaxis aberrations.

7.1.3 *Microtomography with Fresnel zone-plate objectives*

7.1.3.1 *Experimental setup* Figure 7.12(a) shows an experimental setup in a schematical way. It was installed at the beamline 47XU of the 8-GeV electron storage ring SPring-8. An image of the setup is shown in Fig. 7.12(b). The light source is an in-vacuum planar undulator with a permanent magnet array. The undulator radiation is monochromatized by a liquid-nitrogen-cooled silicon (111) double-crystal monochromator. An X-ray energy of 8 keV is selected in this experiment. The bandwidth of the X-ray beam is approximately $\Delta\lambda/\lambda \approx 1 \times 10^{-4}$, which is sufficient for the zone-plate objective.

The experimental setup is located 45 m from the light source. There is no optical element in the beamline except for the crystal monochromator. The beam cross-section at the experimental station is about 0.5 mm×1 mm (V×H). The condenser optics is a specially designed zone plate with a constant-grating pitch of 400 nm and a diameter of 500 μm. The zone material is 1.6-μm thick tantalum. This condenser zone plate is used in combination with a centre beam stop of 300 μm diameter in order to produce hollow-cone illumination. The circle of least confusion where the beam diameter is minimal is located at about 500 mm downstream of the condenser zone plate. By placing the object at the circle of least confusion a quasi-Köhler illumination condition is produced. In front of the object a pinhole with 100 μm diameter was placed as an order-selecting aperture for the condenser zone plate. The field of view is actually defined by this aperture. A beam diffuser made of charcoal powder (grain size of a few tens of μm, rotating at 600 rpm) is also placed upstream of the condenser zone plate for the purpose of reducing strong speckle noise due to the high coherence of the incident X-ray beam.

The objective zone plate is also made of tantalum with a thickness of 1 μm. The outermost zone width is 100 nm and the focal length is 100 mm at 8 keV. The diameter of the object zone plate is 155 μm, and the number of zones is 388. The condenser zone and objective zone plates were both fabricated at NTT-Advanced Technology by electron beam lithography and reactive ion etching. It is confirmed that the objective zone plate has a diffraction-limited resolution for hard X-rays. Although the numerical aperture (convergent angle) of the condenser optics is half that of the objective lens, the influence on spatial resolution is not serious in this case. The resolution limit is calculated as 150 nm for periodic patterns, i.e. 0.75λ/NA (Uesugi *et al.*, 2006). The first-order diffraction efficiency of the condenser zone plate and the objective zone plate is about 20% at 8 keV.

An X-ray imaging detector is used to measure the three-dimensional data in a single scan. The imaging detector consists of a fine powder phosphor screen

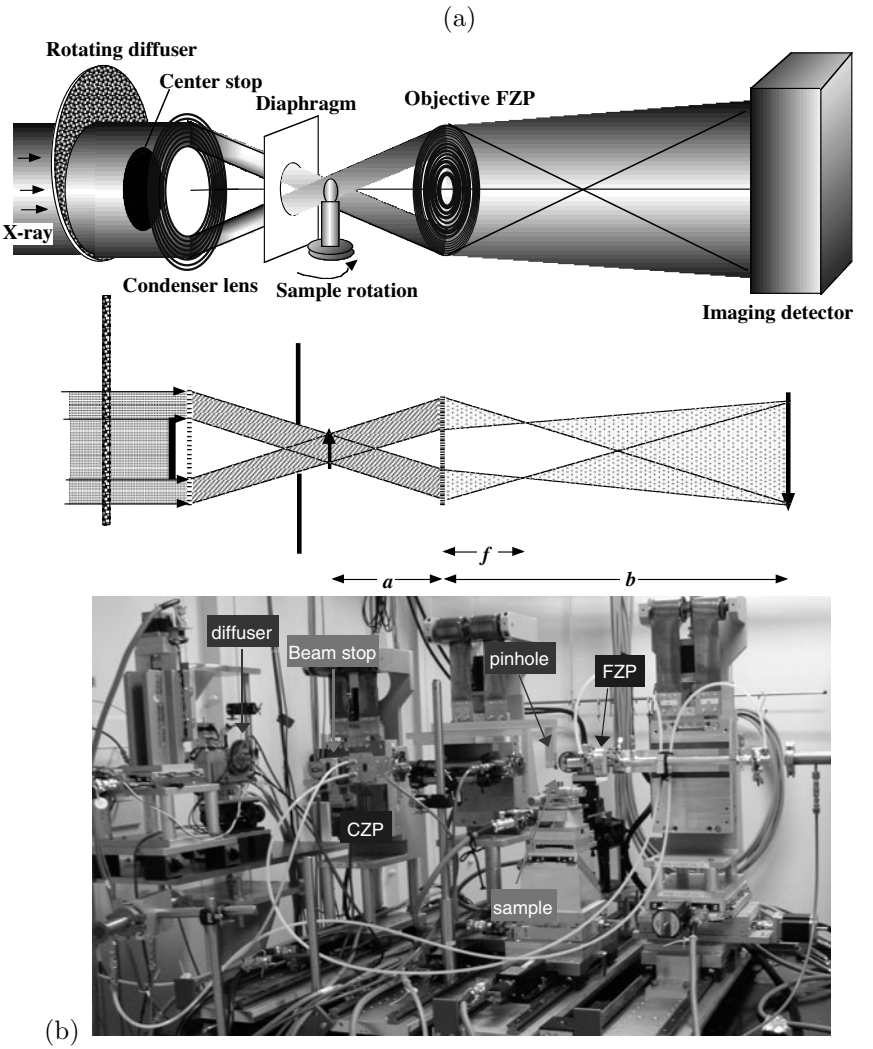


FIG. 7.12. (a) Schematic diagram of experimental setup for imaging microtomography (Uesugi *et al.*, 2006). (b) Experimental setup. Condenser optics, sample stage and objective FZP are shown in the figure. The imaging detector is located 7 m from the objective FZP, right-hand side out of the figure.

('Gadox' or 'P43', Gd_2O_2S:Tb, which was chosen for its good light conversion efficiency, see Chapter 10), a relay lens and a cooled CCD camera (C4880-41S, Hamamatsu Phtotonis KK, Japan). The field of view of the imaging detector is 5.48 mm×3.59 mm and the number of pixels is 2000×1312. The modulation transfer function of the detector is also evaluated, and the limiting resolution is measured to be 9.9 μm from the knife-edge test. The distance between sample

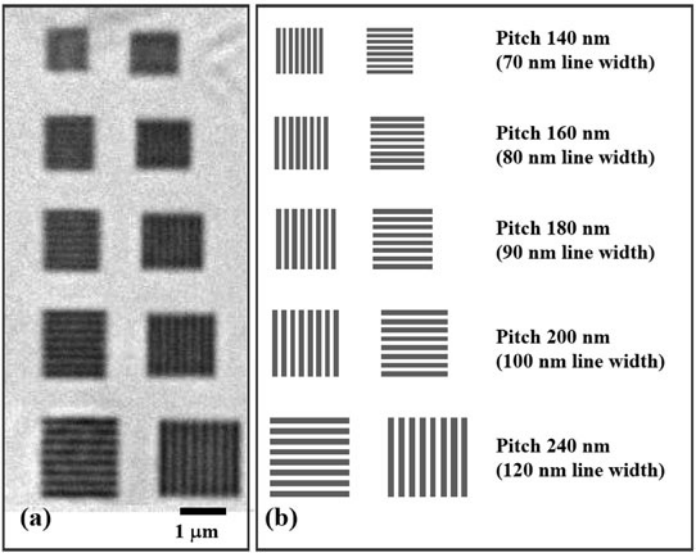


FIG. 7.13. Image of resolution test pattern taken by the imaging microscope and schematic drawing of the resolution test patterns. X-ray energy was 8 keV.

and detector is 7 m. The magnification of the X-ray optics is 68× at an X-ray energy of 8 keV. Therefore, the single pixel size is 40 nm at the object plane and the spatial resolution of the microscope is estimated as about 150 nm at 8 keV.

The spatial resolution of the microscope is verified experimentally by observing resolution test patterns. The results are shown in Fig. 7.13. The test chart consists of varieties of line and space patterns made of tantalum with 0.5 μm thickness. The 80-nm line and 80-nm space (160-nm pitch) pattern is resolved as shown in the figure, while the 70-nm line and 70-nm space pattern is hardly visible. This result agrees well with the theoretical estimation.

A high-precision rotating stage made at Kohzu Precision Co. is used for sample rotation in tomographic measurement. The wobbling of the rotation axis is measured as ±70 nm for a full 360° rotation. The conventional filtered-backprojection algorithm is used for image reconstruction, see Chapter 2. The sample-positioning error is nearly the same as the intrinsic spatial resolution of the microscope. No compensation of sample position and sample orientation is done in the image processing for tomographic reconstruction. Hence, the spatial resolution of the total system is predicted to be about 210 nm.

7.1.3.2 *Experimental results* Figure 7.14 shows a CT image of a test object, a synthetic multilayer deposited on an aluminium wire with 25 μm diameter (Tamura *et al.*, 2002). The X-ray energy is 8 keV, voxel size 40 nm and the number of projections is 1800 images in a 180° rotation. The exposure time for each projection image is 0.5 s, and the total scan time is 35 min. The measured

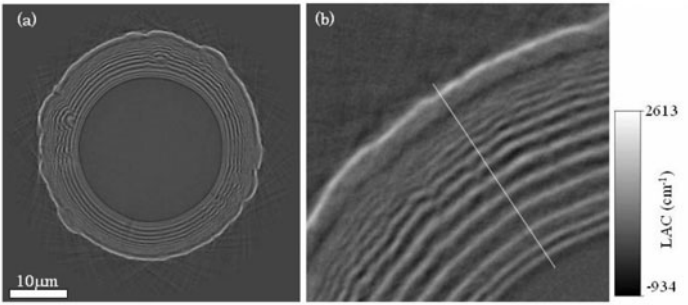


FIG. 7.14. Tomography image of test sample. Concentric multilayer structure is used as a resolution test object for imaging tomography. The outermost structure is 0.15 μm copper and 0.15 μm aluminium synthetic multilayer.

image indicates that the spatial resolution of the system is better than 0.3 μm. This result is consistent with the predicted value derived from the expected performance of the objective FZP, imaging detector and the wobbling error of the sample-rotation stage.

Figure 7.15 shows a CT image of an interplanetary dust particle (IDP) collected by NASA. The sample is mounted on a glass fibre of 5 μm diameter. The glass fibre is held in a glass capillary for easy manipulation. The dimension of the IDP is estimated as 8 μm×6 μm×5 μm by SEM. 1800 projections are acquired in a 180° rotation of the sample with an exposure time of 0.5 s for each projection at an X-ray energy of 8 keV. The transmittance of the sample is greater than 95%. Therefore, image contrast is not optimized for this measurement. However, the internal structure of the IDP can be observed in the measured CT images. It is shown that the IDP does not include any heavy elements. The image quality (signal-to-noise ratio) is not sufficient at present to reveal in detail the distribution of components within the IDP specimen, because the IDP consists of light elements, most likely carbon, silicon, aluminium, and their oxides.

7.1.4 *Applications to materials science and engineering*

Microtomography combined with hard X-ray imaging microscopy has been employed at SPring-8 since 2005 to observe nanoscopic features that have, to date, mainly been investigated by electron microscopy. As examples we present microstructural observations of aluminium alloys and an aluminium foam. Submicrometre-level microstructures could be observed and quantified in terms of volume fraction and orientation. It appears that this technique provides a unique potential to observe the 3D geometry and spatial distribution of nanoscopic features inside samples that are several orders of magnitude thicker than thin-foil specimens required for TEM observation.

Experiments were performed at the X-ray undulator beamline BL47XU at SPring-8. X-ray beams monochromatized to 9.8 keV by a liquid-nitrogen-cooled Si (111) double-crystal monochromator were used. The other details are similar

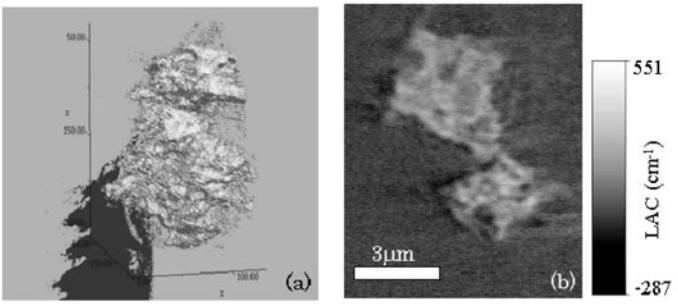


FIG. 7.15. Tomographic image of interplanetary dust particle (L2008D3 #17). (a) three-dimensional reconstruction from CT images. (b) sagital slice derived from three-dimensional reconstructed image.

to Fig. 7.12, except that no condenser zone plate was used for the study on Ag-Al alloys reported in Section 7.1.4.1 in this trial experiment.

7.1.4.1 *Precipitates in a Ag-Al alloy* The system Ag-Al was chosen because the large difference in atomic number between Al and Ag promised strong compositional contrast in X-ray imaging. The Al-10 wt.%Ag binary alloy was prepared by casting. The sample was observed in an overaged condition after soaking, hot rolling and solution heat treatment (400 °C for 3 h). The measured magnification of the X-ray microscope was 49.3× in this configuration. The voxel size was 88 nm, the field of view 35.3 μm×56.8 μm (vertical×horizontal). An X-ray radiograph of the test patterns showed that 180 nm and 240 nm pitches were barely resolved in the vertical and horizontal directions, respectively (Toda *et al.*, 2006*a*).

Figure 7.16 shows a 3D rendered perspective view of γ-Ag_2Al precipitates in the sample (Toda *et al.*, 2006*a*, Toda *et al.*, 2006*c*). It was confirmed that these precipitates could hardly be observed by conventional projection microtomography (without magnification) even when using the highest-resolution detector setup. The precipitates are coherent forms of equilibrium precipitates, representing the regular hexagonal shape of plate-like precipitates. These precipitates appear to be oriented along the four crystallographic planes (indicated by letters A to D in Fig. 7.16) with the reported orientation relationship $(0001)\gamma \parallel (111)\alpha$-Al, with edges parallel to the close-packed $[11\bar{2}0]\gamma \parallel [1\bar{1}0]\alpha$-Al directions (Moore and Howe, 2000). It is evident that the precipitate plates lie on all four {111} planes in the tomographic volumes, as indicated. Characteristic features of the precipitates are shown in Fig. 7.17 (Toda *et al.*, 2006*a*, Toda *et al.*, 2006*c*). Several precipitates show a plate intersection and indications of complex growth around the plate intersections presumably due to the overlap of strain fields between neighbouring plates. Between plates 1 and 3 a gap is visible, clearly confirmed in the virtual tomographic slice of the same region, see Fig. 7.17(b). This figure also demonstrates the presence of a Ag-depleted zone around each precipitate.

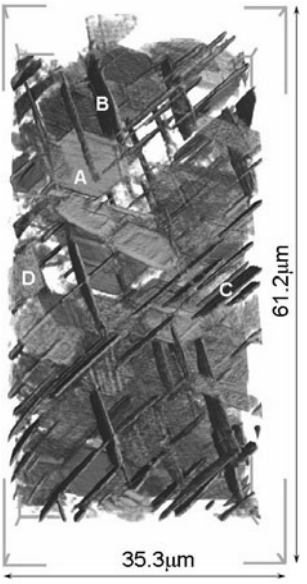


FIG. 7.16. 3D rendered perspective view of γ-Ag_2Al precipitates in a Al-10 wt.%Ag alloy. The aluminium matrix has been removed to show the internal spatial distribution of the precipitates that crystallographically lie on all four equivalent {111} planes of the fcc matrix.

Electron tomography has been used to study Ag-Al alloys in the earlier stages of precipitation and revealed lath-shaped precipitates similar to these in Fig. 7.17 but shorter and thinner (see Fig. 12.12).

7.1.4.2 *Dispersion particles in a Al-Mn engineering alloy* Figure 7.18 shows a tomogram of a Al-Mn alloy before and after hot rolling. The alloy was a newly developed experimental system. The measured magnification of the X-ray microscope was 56× here, the voxel size in the reconstructed slices was 48.9 nm. An X-ray radiograph showed that the test pattern with 160 nm pitch was resolved at an X-ray energy of 8 keV. Coarse particles of about 10 to 50 μm length are clearly observed in Fig. 7.18(a), showing the alloy just after the soaking treatment. These particles were crushed into particles shorter than 10 μm during hot rolling as shown in Fig. 7.18(b). The majority of these finer particles appear to be too small to be resolved by conventional projection microtomography even when using the highest-resolution detector setup available at SPring-8.

7.1.4.3 *Absorption-edge tomography on aluminium alloy foam* A closed-cell aluminium alloy Al-10%Zn-1.5%Ca-1.5%Ti-0.2%Mg (all in wt.%) foam with a typical cell size of 3 mm and a relative density of 0.291 Mg/m^3 was visualized in the same way (Toda *et al.*, 2006*b*). The foam was prepared through a special cast-

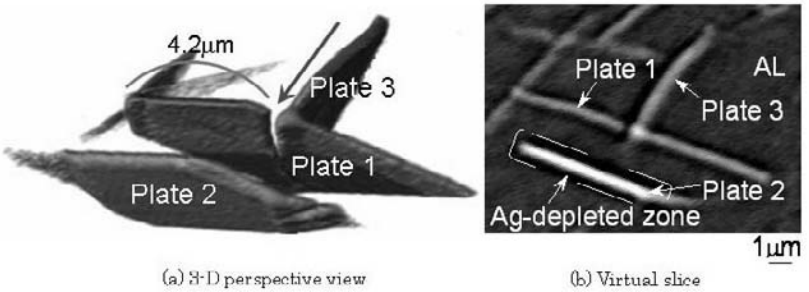


FIG. 7.17. Characteristic region in the tomographic volume of Fig. 7.16. (a) 3D perspective view of some precipitates extracted from the entire volume, demonstrating the existence of a gap in-between. (b) Virtual slice representing the existence of the gap together with a solute-depleted zone (in black) around each precipitate (shown in white). Aluminium matrix is shown in dark grey.

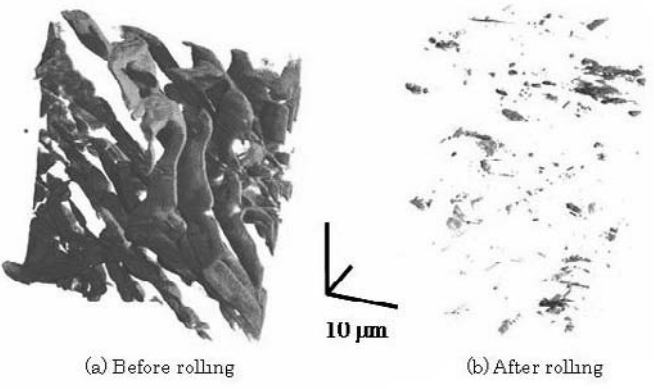


FIG. 7.18. 3D rendered perspective views of intermetallic particles in a new experimental Al-Mn engineering alloy. The matrix aluminium has been removed to show the internal spatial distribution of the particles.

ing process. Calcium and titanium in the chemical compositions originate from the thickening and foaming agents used in manufacture. The foam was subjected to a solution heat treatment at 723 K for 10 h. A set of 3D images was captured just below and above the K absorption edge of zinc located at 9.662 keV in order to visualize and distinguish the Zn-bearing and the Zn-free dispersion particles. Figure 7.19 proves the characteristic feature of these dispersion particles that an individual particle consists of two distinct regions with different zinc concentrations. The existence of such interfaces inevitably causes stress singularities along a nodal line between the interface and the free surface. Thereby, the premature fracture of such particles seems to be induced.

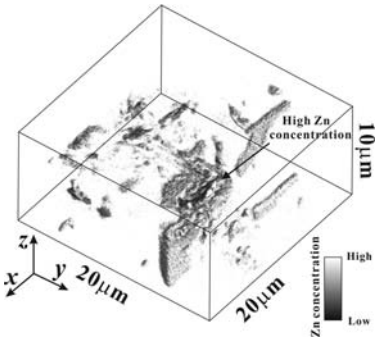


FIG. 7.19. A 3D rendered perspective view of Zn-bearing particles in a cell wall of a Al-10%Zn-1.5%Ca-1.5%Ti-0.2%Mg closed-cell aluminium foam. The Zn distribution is obtained by capturing images just below and above the absorption edge of zinc at 9.662 keV.

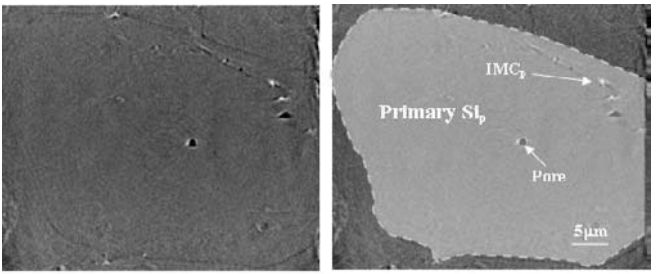


FIG. 7.20. Virtual slices representing internal defects in a coarse primary silicon particle embedded in a hypereutectic Al-Si alloy.

7.1.4.4 *Silion particles in Al-Si alloy* The last example is the observation of a primary silicon particle embedded in a hypereutectic Al-Si alloy. Figure 7.20 shows virtual slices, representing internal and surface defects such as intermetallic compound particles and micropores. It has been reported that the *in-situ* fracture strength of primary silicon particles is 200 to 300 MPa, which is significantly lower than that of much smaller eutectic silicon particles (500 to 900 MPa) (Nishido *et al.*, 2004). It can be inferred that various internal defects observed in Fig. 7.20 may be attributable to the low fracture strength of the primary silicon particles.

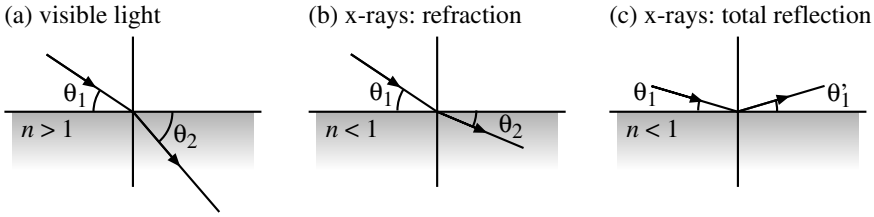


FIG. 7.21. Refraction of visible light (a) and hard X-rays (b), impinging from air into condensed matter. If the angle of incidence is smaller than the critical angle, external total reflection can occur for X-rays (c).

7.2 Hard X-ray microscopy and tomography based on refractive and reflective optics

7.2.1 *Refraction and total external reflection of hard X-rays*

7.2.1.1 *Refraction* Similar to visible light in glass, X-rays are refracted as they pass through matter. This refraction, however, is several orders of magnitude weaker than that for visible light. While this weak refraction is very advantageous for radiography and tomography, where the path of the X-rays through an object can be modelled as a straight line, it makes the manipulation of X-rays by optics difficult. Despite this fact, refractive (and reflective) optics have been successfully developed.

Refraction of hard X-rays in matter can be described by the complex index of refraction $\hat{n} = 1 - \delta + i\beta$, see eqn 4.12, where $\delta > 0$ describes a small deviation of the refractive index from 1 and the imaginary part β the attenuation of the X-rays inside the material (see Section 4.1.4). Figures 7.21(a) and (b) illustrate refraction of visible light and X-rays, respectively. The fact that the refractive index of hard X-rays in matter is < 1 means that the X-rays are refracted away from the surface normal ($\theta_2 < \theta_1$) when entering matter as opposed to visible light in glass that is refracted towards the surface normal ($\theta_2 > \theta_1$). As a consequence, focusing X-ray lenses must have concave surfaces as opposed to focusing lenses for visible light that are convex. In addition, the refraction of hard X-rays in matter is much weaker than that for visible light in glass. For example, for hard X-rays ($E = 10\,\mathrm{keV}$) in silicon $\delta = 4.87 \times 10^{-6}$ is about five orders of magnitude smaller than for visible light in glass, where $n \approx 1.5$. To obtain a focal length f of one metre for visible light, for example, a biconvex glass lens with one metre radius of curvature of both lens surfaces is needed. To reach the same focal length in the hard X-ray range with a biconcave lens made of Si, a radius of curvature of $R = 10\,\mu\mathrm{m}$ would be needed for both surfaces.

As such, a lens is difficult to realize, in particular with a large aperture, most X-ray lens designs are based on stacking a large number N of less strongly curved lenses behind each other as shown in Fig. 7.22(a). For most lens designs, the radius of curvature R is small against the aperture $2R_0$ (cf. inset in Fig. 7.22(a)). In that case, the spherical approximation does not hold, requir-

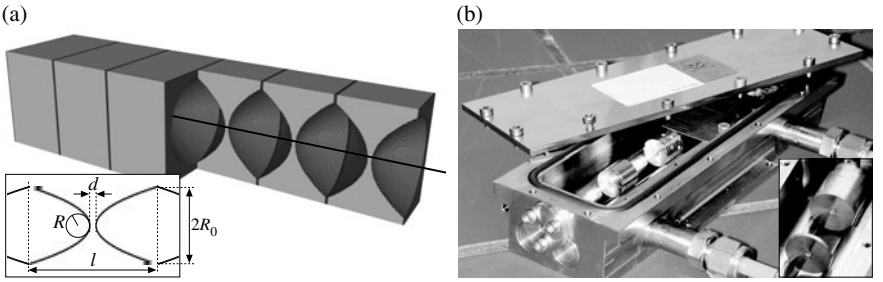


FIG. 7.22. (a) Schematic view of a paraboloid refractive X-ray lens composed of a number N of single biconcave lenses. (b) Parabolic refractive X-ray lens made of beryllium. Each individual lens is centred inside a coin-like holder (see inset). Many of these coins are stacked behind each other to form a compound lens.

ing a parabolic (aspherical) shape for high-quality imaging, such as hard X-ray microscopy (Lengeler *et al.*, 1999*a*, Lengeler *et al.*, 1999*b*, Lengeler *et al.*, 2005).

Attenuation of X-rays inside the lens (described by the imaginary part β of the refractive index $\hat{n}$ in eqn 4.12) is not negligible for any material. To minimize attenuation, X-ray lenses should be made of materials with low atomic number Z, such as beryllium ($Z = 4$). Figure 7.22(b) shows a paraboloid X-ray lens made of beryllium. The individual lenses, each mounted in the center of a coin-like holder, see inset in Fig. 7.22(b), are stacked behind each other on two polished parallel rods. The whole lens is kept in an inert atmosphere to avoid oxidation of the beryllium when irradiated. As the number of individual lenses can be chosen freely up to several hundred, the focal length f can be adjusted from several tens of m down to 0.3 m according to the requirements of the experiment.

In Section 7.2.2 a full-field microscope based on a parabolic refractive X-ray lens will be described as described in Fig. 7.1(b). Using this microscope, a magnified tomogram of a piece of a microprocessor is recorded at submicrometre resolution.

7.2.1.2 *Reflection* For small incidence angles of X-rays onto a surface $\theta_1 < \sqrt{2\delta}$ ($\sqrt{2\delta} = 0.18°$ for a Si surface and $E = 10\,\mathrm{keV}$ X-rays), hard X-rays cannot penetrate deeply into the material and are totally reflected as shown in Fig. 7.21(c). This effect can be exploited to build X-ray mirrors. The largest reflection angles are obtained for materials with high mass density ρ, such as gold or platinum, for which the critical angle of total reflection $\theta_c = \sqrt{2\delta}$ is relatively large, i.e. $\theta_{c(Au)} = 0.44°$ and $\theta_{c(Pt)} = 0.46°$ for $E = 10\,\mathrm{keV}$ photons.

To focus X-rays with total reflection mirrors to small dimensions, the figure error and roughness must be very small. To obtain satisfactory results the mirror figure must not deviate by more than a few nanometres from the ideal shape, e.g. elliptical, to focus from point to point in one dimension. To focus in two dimen-

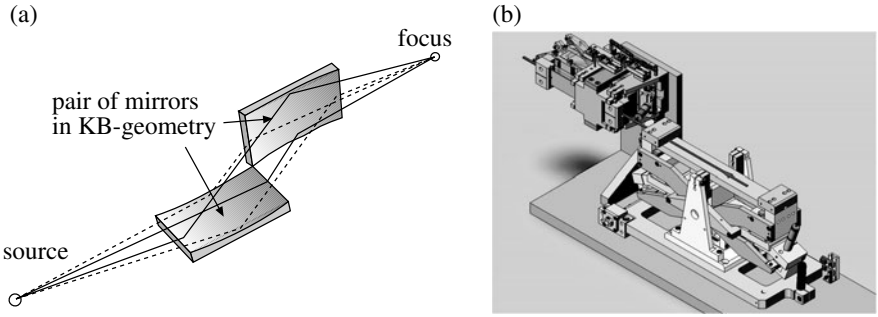


FIG. 7.23. (a) Schematic diagram of two-dimensional focusing using a pair of crossed meridionally curved mirrors (Kirkpatrick–Baez geometry). (b) Kirkpatrick–Baez mirror system designed at the European Synchrotron Radiation Facility (ESRF) (Hignette *et al.*, 2007).

sions with a single reflection, the mirror surface must be an ellipsoid of rotation. Due to the small reflection angle, two-dimensional focusing with a rotational ellipsoid is very difficult to realize, since such an ellipsoid is extremely elongated (needle-like). For such a mirror, the sagittal curvature would have to be several orders of magnitude larger than the meridional one. To avoid the difficulty of fabricating a double-curved mirror, two meridionally curved mirrors can be used in crossed geometry as proposed by Kirkpatrick and Baez (1948) and shown in Fig. 7.23(a). Such mirrors can be made with very high accuracy (Mimura *et al.*, 2005). The mirrors based on total external reflection are sometimes replaced by graded multilayer mirrors (Hignette *et al.*, 2005). These diffraction gratings allow for larger reflection angles and thus larger apertures.

In Section 7.2.3, a projection microscope based on a KB-mirror system is described, see Fig. 7.1(a). It can be used to record magnified tomograms.

7.2.2 *Tomography based on hard X-ray full-field microscopy*

Using the setup shown in Fig. 7.1(b), an X-ray optic can be used to image the sample onto a detector in a magnifying geometry. The sample is illuminated from the back by monochromatic hard X-rays. The optic is placed behind the sample at a distance L_1 slightly larger than the focal distance f. The image is formed at distance $L_2 = L_1 f/(L_1 - f)$ behind the lens on a high-resolution X-ray camera. The resulting magnification is $m = L_2/L_1 = f/(L_1 - f)$.

For incoherent illumination, the spatial resolution that can be achieved in such a microscope based on refractive lenses is:

$$\Delta = 0.75 \frac{\lambda L_1}{D_{\text{eff}}} = 0.75 \frac{\lambda}{2\text{NA}},$$

where λ is the X-ray wavelength, D_{eff} is the effective aperture and $\text{NA} = D_{\text{eff}}/2L_1$ is the numerical aperture of the objective lens in the microscope geometry. For

a Fresnel zone plate, D_{eff} is its diameter, for a refractive lens D_{eff} is slightly smaller than the geometric aperture $2R_0$, cf. Fig. 7.22(a), due to an increased attenuation of the X-rays in the outer (thicker) parts of the lens.

In order to illuminate the sample with incoherent radiation and to match the numerical aperture of the hard X-ray optic, a diffuser (rotating random scatterer) is placed into the illuminating beam slightly before the sample instead of a condenser lens. This simplified version of an X-ray full-field microscope without condenser lens is possible in the hard X-ray range, since the numerical aperture of the objective lens is typically smaller than 1 mrad. This small numerical aperture has the advantage of giving a large depth of field

$$d_l = 0.64\frac{\lambda}{\mathrm{NA}^2}$$

relative to the spatial resolution Δ. This is favourable to tomographic imaging, since it allows one to image thick objects in sharp projection (compare to eqn 7.8).

To acquire a tomogram a large number of magnified images of the sample are recorded, while rotating the sample in discrete steps around an axis perpendicular to the beam. Magnified tomography is illustrated using a piece of an AMD K6 microprocessor as a test object (Schroer *et al.*, 2002). The X-ray full-field microscope was setup at beamline ID22 of the ESRF. Using a parabolic refractive X-ray objective lens made of aluminium with a focal length of $f = 1048\,\mathrm{mm}$ at $E = 25\,\mathrm{keV}$, the sample was imaged with $20.9\times$ magnification onto a high-resolution X-ray camera ($L_1 = 1098\,\mathrm{mm}$, $L_2 = 23.02\,\mathrm{m}$). This microscope was used to record a tomogram of a $100 \times 150\,\mu\mathrm{m}^2$ fragment of an AMD K6 microprocessor that was prepared as a test object. 500 projections were recorded in an angular interval of 180°. The rotation axis was perpendicular to the plane of the microprocessor.

Figures 7.24(a) and (b) show two projections through the microprocessor. The dashed vertical line indicates the rotation axis. The silicon wafer (bottom of each image) on which the processor is built gives almost no attenuation contrast. The tungsten plug connects that make an electronic contact of the semiconductor devices with the first layer of aluminium interconnects is seen as a dark, structured horizontal line at the bottom. Within four layers of oxide the tungsten plug vias are visible as dark vertical bars, making the electrical connection between aluminium interconnects that lie on different oxide layers. The silicon, the oxides and the aluminium interconnects all attenuate the X-rays very weakly and produce hardly any absorption contrast above the noise level in this experiment. The contrast of these structures could be significantly increased by using phase-contrast techniques, see Chapter 6.

The microprocessor image data was reconstructed tomographically using filtered backprojection. Prior to backprojection, unwanted motion of the rotation axis was compensated for numerically. Tomographic reconstruction is shown in Fig. 7.25. The spatial resolution in the reconstruction is about 0.4 µm.

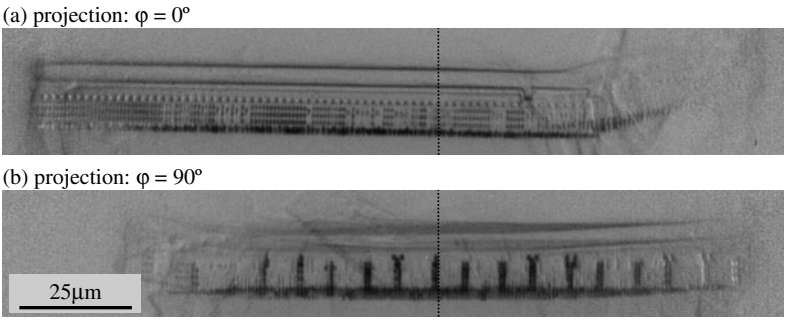


FIG. 7.24. (a) and (b) show two X-ray micrographs of a fragment of an AMD K6 microprocessor used as a test structure. They are part of a tomographic scan and were recorded perpendicularly to each other. The rotation axis is depicted as a dashed line in both images. Reused with permission from (Schroer *et al.*, 2002), Copyright 2006, American Institute of Physics.

In order to obtain high-resolution tomograms the projection images must be free of distortion and the setup must have sufficient mechanical stability. Most important is the mechanical stability of the sample stage that must allow for rotation of the sample with small eccentricity and wobble.

7.2.3 *Tomography using a hard X-ray microscope based on magnified projection imaging*

Following the scheme shown in Fig. 7.1(a), an X-ray optic can be used to generate a small focus that serves as a point source to illuminate an object with a diverging beam. The transmission image formed on the detector is magnified according to the ratio of source-to-object and source-to-detector distances, $m = L_2/L_1$. The size of the focus, the magnification ratio, and the detector pixel size determine the resolution that can be achieved in this setup.

As the pixel size of high-resolution detectors is limited from below to slightly below one micrometre, high spatial resolution requires a high magnification and a small focus. With current high-quality X-ray optics, such as Fresnel zone plates (Kang *et al.*, 2006), Kirkpatrick–Baez (KB) mirror systems (Mimura *et al.*, 2005, Hignette *et al.*, 2005), and refractive X-ray lenses (Schroer *et al.*, 2005), it is possible to generate a focus with a lateral size well below 100 nm and high X-ray intensity.

When the sample is moved into the focus, the same setup can be used to raster scan the sample with the small beam. Although scanning tomography is relatively slow, this scheme allows one to use X-ray analytical techniques, such as X-ray fluorescence, absorption spectroscopy, or scattering, as the contrast mechanism for tomographic imaging, revealing the distribution of chemical elements, of the chemical state of an atomic species, or of the local nanostructure inside a specimen. Chapter 8 describes these methods in more detail.

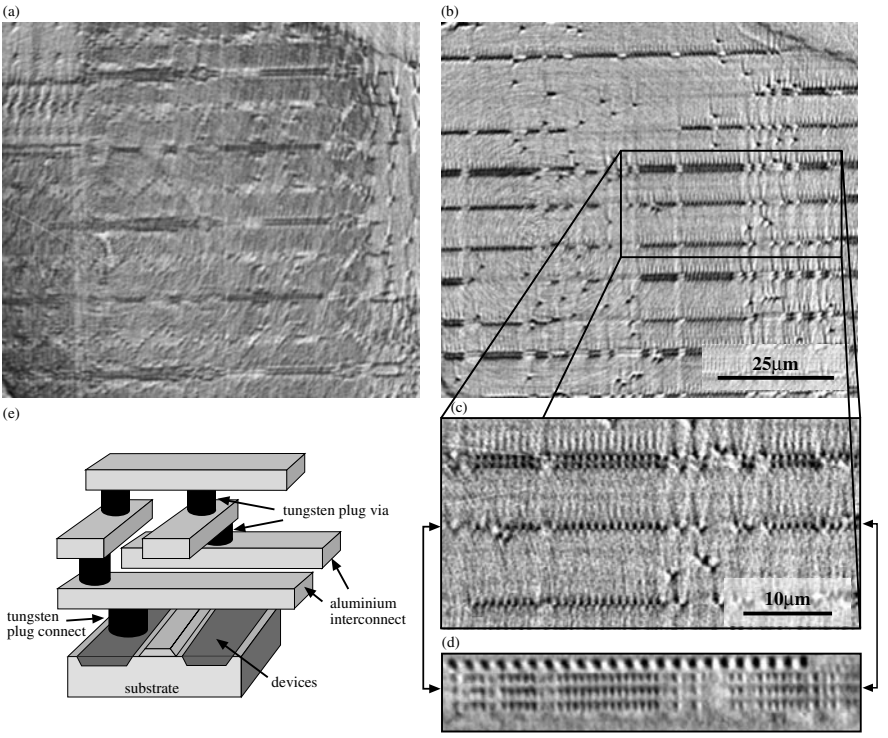


FIG. 7.25. (a) Tomographic (horizontal) slice through the device layer of the microprocessor, (b) (horizontal) slice through the second layer of tungsten plug vias. (c) is a magnified view of the rectangular region in (b), (d) shows a vertical slice through the row of plug vias marked by the arrows in (c). (e) Schematic sketch of the structure of a microprocessor. The devices are structured lithographically onto the substrate and connected by tungsten plug connects to the first layer of aluminium interconnects. Several layers of aluminium interconnects are connected by plug vias. (b–d) reused with permission from (Schroer *et al.*, 2002), Copyright 2006, American Institute of Physics.

7.3 Summary

Using magnifying optics the spatial resolution of tomographic imaging can be improved beyond the limit set by the currently available detector systems for non-magnifying projection tomography. Fresnel zone plates, refractive lenses, Kirkpatrick–Baez mirrors and asymmetric Bragg mirrors have been successfully used for imaging.

While such magnified tomographic imaging with hard X-rays has been shown to yield submicrometre resolution in a few demonstration experiments, it is still under development and not routinely available to users of synchrotron radiation

sources. New end stations implementing magnified imaging and tomography are under construction at existing beamlines, for example at the ESRF, and are planned at new sources such as PETRA III at DESY in Hamburg. At SPring-8, a high-resolution imaging microtomography system has been installed at beamline 47XU and has been opened to general users in autumn 2006.

7.4 References

Born, M. and Wolf, E. (1980). *Principles of optics*. Cambridge University Press, Cambridge.

Hignette, O., Cloetens, P., Morawe, C., Borel, C., and Ludwig, W. (2007). Nanofocusing at ESRF using graded multilayer mirrors. In *Synchrotron Radiation Instrumentation: Ninth International Conference on Synchrotron Radiation Instrumentation* (ed. J.-Y. Choi and S. Rah), Volume 879 of *AIP Conference Proceedings*, pp. 792. American Institute of Physics.

Hignette, O., Cloetens, P., Rostaing, G., Bernard, P., and Morawe, C. (2005). Efficient sub 100 nm focusing of hard X-rays. *Review of Scientific Instruments*, **76**, 063709.

Kang, H.C., Maser, J., Stephenson, G.B., Liu, C., Conley, R., Macrander, A.T., and Vogt, S. (2006). Nanometer linear focusing of hard X-rays by a multilayer Laue lens. *Physical Review Letters*, **96**, 127401.

Kirkpatrick, P. and Baez, A. (1948). Formation of optical images by X-rays. *Journal of the Optical Society of America*, **38**, 766.

Lengeler, B., Schroer, C.G., Kuhlmann, M., Benner, B., Günzler, T.F., Kurapova, O., Zontone, F., Snigirev, A., and Snigireva, I. (2005). Refractive X-ray lenses. *Journal of Physics D: Applied Physics*, **38**, A218.

Lengeler, B., Schroer, C.G., Richwin, M., Tümmler, J., Drakopoulos, M., Snigirev, A., and Snigireva, I. (1999*a*). A microscope for hard X-rays based on parabolic compound refractive lenses. *Applied Physics Letters*, **74**, 3924.

Lengeler, B., Schroer, C., Tümmler, J., Benner, B., Richwin, M., Snigirev, A., Snigireva, I., and Drakopoulos, M. (1999*b*). Imaging by parabolic refractive lenses in the hard X-ray range. *Journal of Synchrotron Radiation*, **6**, 1153.

Michette, A. (1986). *Optical system for soft X-rays*. Plenum Press, New York.

Mimura, H., Matsuyama, S., Yumoto, H., Yamamura, K., Sano, Y., Shibahara, M., Endo, K., Mori, Y., Nishino, Y., Tamasaku, K., Yabashi, M., Ishikawa, T., and Yamauchi, K. (2005). Hard X-ray diffraction-limited nanofocusing with Kirkpatrick-Baez mirrors. *Japanese Journal of Applied Physics*, **44**, L539.

Moore, K.T. and Howe, J.M. (2000). Characterization of gamma plate-shaped precipitates in an Al-4.2 at.% Ag alloy: Growth kinetics, solute field, composition, and modeling. *Acta Materialia*, **48**, 4083.

Nieman, B., Guttmann, P., Hambach, D., Schneider, G., Wei, D., and Schmahl, G. (2000). The condenser-monochromator with dynamical aperture synthesis for the TXM at an undulator beamline at BESSY II. In *X-ray microscopy* (ed. W. Meyer-Ilse, T. Warwick, D. Attwood, and E. Orlando), Volume

507, pp. 440. Proceedings of the VI International Conference, Berkeley, 2.–6.8.1999.

Nishido, S., Kaneso, M., Kobayashi, T., and Toda, H. (2004). Role of Si particle damage on fatigue characteristics of cast Al-Si alloys. *International Journal of Cast Metals Research*, **17**, 345.

Schäfer, P. and Köhler, R. (2003). Asymmetric Bragg reflection as X-ray magnifyer. *Journal of Physics D: Applied Physics*, **36**, A113.

Schroer, C.G., Kurapova, O., Patommel, J., Boye, P., Feldkamp, J., Lengeler, B., Burghammer, M., Riekel, C., Vincze, L., van der Hart, A., and Küchler, M. (2005). Hard X-ray nanoprobe based on refractive X-ray lenses. *Applied Physics Letters*, **87**, 124103.

Schroer, C.G., Meyer, J., Kuhlmann, M., Benner, B., Günzler, T.F., Lengeler, B., Rau, C., Weitkamp, T., A.Snigirev, and Snigireva, I. (2002). Nanotomography based on hard X-ray microscopy with refractive lenses. *Applied Physics Letters*, **81**, 1527.

Suzuki, Y., Takeuchi, A., Takano, H., Ohigashi, T., and Takenaka, H. (2001). Diffraction-limited microbeam with Fresnel zone plate optics in hard X-ray regions. *Japanese Journal of Applied Physics*, **40**, 1508.

Suzuki, Y., Takeuchi, A., Takano, H., and Takenaka, H. (2005). Performance test of Fresnel zone plate with 50 nm outermost zone width in hard X-ray region. *Japanese Journal of Applied Physics*, **44**, 1994.

Takano, H., Suzuki, Y., and Takeuchi, A. (2003). Sub-100 nm hard X-ray microbeam generation with Fresnel zone plate optics. *Japanese Journal of Applied Physics*, **43**, L132.

Takeuchi, A., Uesugi, K., Suzuki, Y., and Aoki, S. (2001). Hard X-ray microtomography using X-ray imaging optics. *Japanese Journal of Applied Physics*, **40**, 1499.

Takeuchi, A., Uesugi, K., Takano, H., and Suzuki, Y. (2002). Submicrometer-resolution three-dimensional imaging with hard X-ray imaging microtomography. *Review of Scientific Instruments*, **73**, 4246.

Tamura, S., Yasumoto, M., Kamijo, N., Suzuki, Y., Awaji, M., Takeuchi, A., Takano, H., and Handa, K. (2002). Development of multilayer Fresnel zone plate for high-energy synchrotron radiation X-rays by DC sputtering deposition. *Journal of Synchrotron Radiation*, **9**, 154.

Toda, H., Minami, K., Kobayashi, M., Uesugi, K., and Kobayashi, T. (2006*a*). Observation of precipitates in aluminum alloys by sub-micrometer resolution tomography using Fresnel zone plate. *Materials Science Forum*, **519-521**, 1361.

Toda, H., Ohgaki, T., Uesugi, K., Kobayashi, M., Kuroda, N., Kobayashi, T., Niinomi, M., Akahori, T., Makii, K., and Aruga, Y. (2006*b*). Quantitative assessment of microstructure and its effects on compression behaviour of aluminium foams via high-resolution synchrotron X-ray tomography. *Metallurgical and Materials Transactions A*, **37**, 1211.

Toda, H., Uesugi, K., Takeuchi, A., Minami, K., Kobayashi, M., and Kobayashi,

T. (2006*c*). Three-dimensional observation of nanoscopic precipitates in an aluminum alloy by microtomography with Fresnel zone plate optics. *Applied Physics Letters*, **89**, 143112.

Uesugi, K., Takeuchi, A., and Suzuki, Y. (2006). Development of microtomography system with Fresnel zone plate optics at SpRing8. In *Developments in X-ray tomography V* (ed. U. Bonse), Volume 6318, pp. 154. SPIE. Proceedings of SPIE International Conference, San Diego, 15.08.2006.

Weiss, D., Schneider, G., Nieman, B., Guttmann, P., Rudolph, D., and Schmahl, G. (2000). Computed tomography of cryogenic biological specimens based on X-ray microscopic images. *Ultramicroscopy*, **84**, 185.

8

SCANNING TOMOGRAPHY

Alexandre Simionovici, *Pierre Bleuet* and *Bruno Golosio* (8.1)
Christian Schroer (8.2)

The tomographic methods presented in the preceding three chapters are all based on the simultaneous acqusition of two-dimensional images of an object placed in an extended beam using appropriate 2D detectors. Simultaneous acquisition of many image pixels is valid under the assumption that the information contained in each detector pixel corresponds to a different area in the sample and that there is no cross-talk between the individual pixels. Absorption of parallel X-rays satisfies this requirement approximately as long as one neglegts scattering effects that change the direction of the penetrating beams. Such scattering can be treated as a perturbation and can be corrected partially, as will be shown in Chapter 13 for neutrons. In cases where the effect of interest changes the direction of the incoming radiation, the image pixels are no longer decoupled and one has to scan the sample sequentially using a small focused beam to avoid the information from different regions of the sample being convoluted irreversibly. Such scanning methods are described in this chapter. Three examples are treated: tomography using fluorescent X-ray radiation, tomography exploiting the absorption fine structure near the absorption edges and tomography utilizing small-angle scattering from mesoscopic structures within the sample.

8.1 Fluorescence tomography

8.1.1 *History*

X-ray computed fluorescence tomography (XCFT) is a 'first-generation' tomographic technique requiring a 'pencil beam', i.e. a narrow parallel or even focussed beam with a small rectangular footprint and a point detector. XCFT appeared in the wake of the invention of computer-assisted tomography by Sir Godfrey N. Hounsfield, who was awarded the Nobel prize in medicine in 1979 for this invention together with Allan M. Cormack. The principles of their invention guided an exceptional effort towards medical imaging over the next few decades. Many probes such as NMR, X-ray scattering, optical (diffuse, coherent and diffraction), MRI, PET, laser, electron beam, ultrasound, *etc.* found immediate applications and pioneered subdomains of the general computed tomography field. It gradually became clear that obtaining elemental sensitivity coupled to the already well-developed tomographic imaging would be an extraordinary probe of the structure and morphology of solid samples, with potential applications in the

fields of materials science, earth and planetary sciences or life and environmental sciences.

Two-dimensional X-ray fluorescence flat imaging has been applied since the 1960s with moderate resolution (mm) using setups of X-ray tubes and crossed slits, but no-one had tried to tackle the complexity of the 3D problem. Yet, only seven years after Hounsfield's Nobel prize, Boisseau (1986) had already defended his PhD thesis at the MIT, Cambridge, dedicated to fluorescence tomography using focused X-ray beams and applied to three-dimensional imaging.

The first paper in this field was published a year later by Boisseau and Grodzins (1987) in a study of biological specimens using synchrotron radiation delivered by the newly installed NSLS synchrotron. Cesareo and Mascarenhas (1989) developed the first confocal setup for fluorescence tomography using two collimators of 1 mm internal diameter. One was placed in the incident beam to delimit a $1\,\mathrm{mm}^2$ pencil beam and the other in front of the detector to constrain the solid angle of detection. The tomographed volume was $1\,\mathrm{mm}^3$ at the intersection of the two beams. In order to obtain quasi-monochromatic radiation, they used the secondary fluorescence of a foil excited by an X-ray tube. The foil was placed on top of the incident beam collimator to select only its fluorescent radiation. The choice of the foil was optimized with respect to the element to be imaged in the actual sample. In the meantime, this confocal technique has been significantly advanced using polycapillary optics to define the acceptance of the fluroescence detector (Vincze *et al.*, 2004, Vekemans *et al.*, 2004)

Iida *et al.* (1989) used the traditional translation-rotation scanning scheme in order to enhance the signal-to-noise ratio during scanning of trace-element distributions and obtained reconstructed elemental maps. Their idea was to propose a high count rate alternative to the traditional 2D scanning that required very thin slices in order to yield high-resolution maps. Hogan *et al.* (1991) further elaborated on the simple reconstruction algorithms used so far by introducing absorption corrections for both incident and fluorescent radiation. Their method required performing standard absorption tomography at both incident and fluorescent energies that can only be achieved using a synchrotron for tuning the incident radiation to the appropriate range. Their reconstruction made use of the filtered backprojection algorithm (FBP) customarily used in absorption tomography and reliably corrected for the self-absorption effect of all fluorescence lines (Kak and Slaney, 1987). As an average over all θ angles (the sample rotation angle is defined in Fig. 8.1) is performed for the absorption corrections their results are only applicable to low absorption cases and relatively large concentrations.

In the hope of applying it to imaging of clinical samples, Takeda *et al.* (1994) and Yuasa *et al.* (1997) further developed the XCFT method by applying it to imaging of low-Z phantoms into which high-Z trace elements such as iodine had been injected. This allowed the first demonstration of XCFT as an imaging method fit for medical diagnostics of organs with injected contrast agents. Finally, Rust and Weigelt (1998) applied the same XCFT method using two absorption tomograms at the incident and fluorescent energies for quantitatively

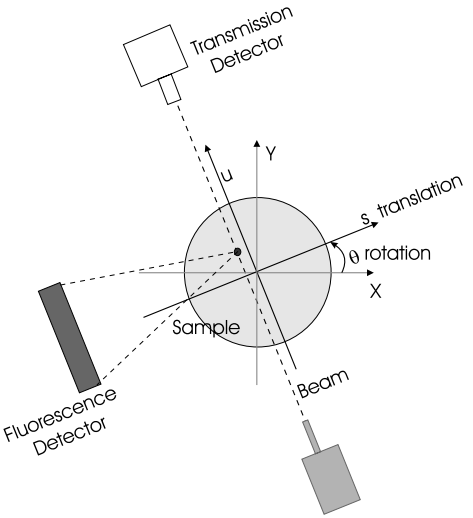


FIG. 8.1. Schematic setup for fluorescence tomography.

imaging human thyroid glands and related phantoms, with an average error of 20% for an iodine concentration in the mmol/l range. They showed that the maximum likelihood method with expectation maximization (MLEM) performed better than the algebraic reconstruction technique (ART) (Kak and Slaney, 1987), particularly when it came to reconstruction artefacts. However, when comparing reconstruction methods different conclusions can be drawn sometimes, as can be seen from Section 2.2.2 where ART is found to be equal or superior to MLEM. Such comparisons depend very much on the quality of the data and the particular algorithms used and should therefore be treated with care as stated in the first paragraph of Section 2.2.

With the advent of third-generation synchrotron sources, XCFT gained a new field of application, by making use of the monochromatic, high-intensity ($> 1 \times 10^9$ ph/s) focused beams of a few micrometres size. The group, based at ESRF, started developing XCFT on the ID22 microfluorescence/imaging/diffraction beamline (Somogyi *et al.*, 2005, European Synchrotron Radiation Facility, 2006) beginning in 1998, using the available microbeams with resolutions in the range of a few micrometres (Simionovici *et al.*, 2000, Simionovici *et al.*, 2002). This method was also implemented at other beamlines (Terzano *et al.*, 2005) and synchrotron radiation sources (McNear *et al.*, 2005).

8.1.2 *Description of fluorescence tomography signal and relevant parameters*

8.1.2.1 *Problems and limitations in XCFT* As for any first-generation tomographic system, XCFT scans are traditionally made using a pencil beam and scanning the sample in translation across the beam in N steps while collecting the emitted fluorescence on an energy-dispersive detector. These translational scans are done while rotating the sample in M angular steps between 0° and

360°, with the rotation axis perpendicular to both the incident beam and the translation direction. An XCFT image will therefore have $N \times M$ points with associated spectra corresponding to each of the translation/rotation scans. In absorption tomography, as the incident beam and detection directions are identical, angular scanning is only made on 180°. In XCFT, however, the non-symmetrical reflection geometry shown in Fig. 8.1 requires the placement of the detector on one side of the sample. This imposes that spectra obtained at 0° and 180° are highly different due to self-absorption in the sample. As this self-absorption is in general very strong and the dynamical range of the detector is limited, a 180° scan misses essential information. As an illustration: when taking a tomogram from a homogeneous sphere with a particle embedded in an offcentric position, the one of the two possible 180° scans where during rotation the particle would be on a semi-circle opposing the detector would not yield any signal in an extreme case while the second scan with the particle facing the detector could create a strong signal. Consequently, scans over 360° are necessary to expose all parts of the sample to all the positions with respect to the detector, thus ensuring that all emission and self-absorption positions of the sample are spanned.

One of the intrinsic limitations of fluorescence tomography is given by the angular distribution of the emitted radiation. Indeed, fluorescence emission covers isotropically the 4π solid angle around the emission point. As the emitted radiation is recorded by a detector located at some angle with respect to the incident beam and emission point, no possibility exists of expanding the incoming pencil beam into a flat or ribbon beam so as to encompass the whole sample and therefore allow us to only perform the rotational scans as in traditional second-generation tomography. Doing so would superimpose contributions from all the sample's pixels in the same detector in a non-separable way. Therefore, XCFT is and remains a first-generation tomography technique with its inherent speed limitations. Still, some hybrid setups proposed a few years back (Simionovici *et al.*, 2001) are still promising, such as symmetrically placing two identical detectors on both sides of the incident beam at 90° in the horizontal or vertical plane so as to scan only 180° instead of the required 360°, but they provide only a reduction of the measurement time to 50% in the best case. One could call such extensions '*1.5th-generation* setups', see Section 8.1.6.4.

The greatest limitation of XCFT lies in the maximum volume spanned by the fluorescence photons, which are subject to self-absorption once produced by the pencil beam in a particular pixel. Sample volume must necessarily be confined within a response volume, of the order of a few l^3, with l the attenuation length (the length that transmits $1/e$ of the incident intensity) of the lowest-energy fluorescence radiation in the particular matrix. As an example, the attenuation lengths of the K_α radiation of Ca and Mo in a quartz matrix (SiO_2) are 11.7 µm and 1080 µm, respectively, which effectively constrains the maximum analysable volume to a few tens of micrometres in each dimension if both elements are needed. Consequently, the response volume is a function of the atomic number of the analysed elements and in order to image samples with volumes given

by the practical problem to be solved one has to be prepared to restrict the analysis to higher-Z elements in many cases. This raises the important question of the absorption of the emitted fluorescence lines by the other lower-Z elements, considered now as a 'matrix'. A method has been developed to address the problem of low-Z elements that produce little or no fluorescence but that must be considered in order to achieve a quantitatively accurate high-Z imaging. This method makes use of Compton scattering and is presented in Section 8.1.2.3.

8.1.2.2 *Transmission tomography* The signal used in transmission tomography is collected by the beam monitors/detectors set before and after the sample. As is customary in first-generation tomography using pencil beams, scanning of the sample is done by a series of translations s across the beam axis followed by a rotation by an angle θ, after which the whole procedure is repeated until the whole 360° have been spanned. Usually, the X-ray source and the detectors are fixed, while the sample is rotated and translated. However, the reconstruction problem is more easily described in terms of a stationary sample and a moving source/detector system. In Fig. 8.1, the (x, y)-axes are fixed to the reference frame of the sample. The s- and u-axes are parallel to the translation direction and to the beam, respectively, and they are related to (x, y) by a rotation of the angle θ:

$$\begin{aligned} s &= x\cos\theta + y\sin\theta & \qquad x &= s\cos\theta - u\sin\theta \\ u &= -x\sin\theta + y\cos\theta & \qquad y &= s\sin\theta + u\cos\theta. \end{aligned}$$

In the beam referential, the intensity I of the transmitted beam across the sample is:

$$I(s,\theta) = I_0 \exp\left(-\int_{-u_{\min}}^{u_{\max}} \mu(s,u,E_0)du\right),$$

which is identical to eqn 5.1, except for the dependence on the energy of the primary beam E_0. I_0 is the intensity of the incident beam and μ is the linear absorption coefficient of the sample at co-ordinates (s, u), see Fig. 8.1. The projection function P is defined as:

$$P(s,\theta) = \ln\left(\frac{I_0}{I(s,\theta)}\right) = \int \mu(s,u,E_0)du = \mathcal{R}\mu,$$

where $\mathcal{R}$ was defined as the Radon transform operator (Kak and Slaney, 1987), see eqn 2.1. In (x, y) co-ordinates, if we define the following vectors $\boldsymbol{X} = (x, y)$ and $\boldsymbol{A} = (\cos\theta, \sin\theta)$, the projection function writes as:

$$P(s,\theta) = \int \mu(x,E_0)\delta(s - \boldsymbol{A}\cdot\boldsymbol{X})dx = \mathcal{R}\mu.$$

The reconstruction problem of transmission tomography consists of the inversion of the Radon transform operator. Given a set of measured projections $P(s, \theta)$, the reconstruction determines a distribution of the absorption coefficient μ whose

Radon transform $\mathcal{R}\mu$ is equal to the measured projections. Using various reconstruction algorithms it is possible to reconstruct the spatial distribution of the linear absorption coefficient at the energy E_0 of the primary beam.

8.1.2.3 *Compton/Rayleigh tomography* The signal produced on the energy-dispersive detector by the photons that come from elastic (Rayleigh) and inelastic (Compton) single-scattering events are the basis of what is called *Compton/Rayleigh tomography.* Traditionally, this type of measurement was associated with γ-rays as their penetration allows imaging low-Z elements inside massive samples (Schätzler, 1978). These two effects produce distinct lines, e.g. at 15 keV and 90° angle the Compton shift is 430 eV while the resolution of the SSD detector is 180 to 220 eV, but as they are quite close together, the two contributions will be summed as a single line, named 'scat'. As both these effects depend on the polarization of the incoming beam, in the following, the case of a polarized beam as is the case with all synchrotron X-ray beams, will be considered.

Referring to Fig. 8.1, the contribution of a small path du along the beam to the detected signal is given by

$$dN = I_0 \epsilon_D f(s,u) p_{\text{scat}}(s,u) g(s,u) du,$$

with I_0 the intensity of the incident beam, ϵ_D the detector efficiency and $f(s,u)$ the probability that a photon entering the sample reaches the point (s,u):

$$f(s,u) = \exp\left(- \int_0^u \mu(s,u',E_0) du' \right),$$

and p_{scat} is the scattering probability per unit path length per unit solid angle

$$p_{\text{scat}}(s,u) = \sum_i N_A \frac{\rho_i}{A_i} \frac{d\sigma_{\text{scat}}}{d\Omega} (Z_i, E_0, \theta, \phi),$$

where N_A is the Avogadro number, ρ_i, A_i and Z_i are the density, the atomic weight and the atomic number of the ith element, $d\sigma/d\Omega$ is the differential scattering cross-section of Compton plus Rayleigh scattering (see eqn 4.2) and θ and ϕ are the scattering polar (0–360°), respectively, azimuthal angles (±90°) of the detector with respect to the incident beam direction. $g(s,u)$ is the probability that a photon scattered at the point (s,u) reaches the detector

$$g(s,u) = \int_{\Omega_D} d\Omega \exp\left(- \int_{(s,u)\to\text{Det}} \mu(l,E') dl \right), \tag{8.1}$$

where Ω_D is the solid angle from the interaction point to the detector surface and $\mu(l,E')$ is the absorption coefficient at the energy of the scattered photon. The

radiation reaching the detector, normalized by the factor $\epsilon_D I_0$, can be obtained by an integration along the path du:

$$S_{\text{scat}}(s,\theta) = \int K(s,u) p_{\text{scat}}(s,u) du \quad \text{with } K(s,u) = f(s,u)g(s,u), \tag{8.2}$$

which can be written in the reference frame of the sample as:

$$S_{\text{scat}}(s,\theta) = \int dX K(\theta,X) p_{\text{scat}}(X) \delta(s - \boldsymbol{A}\cdot\boldsymbol{X}). \tag{8.3}$$

This equation is often referred to as the *generalized Radon transform* of p_{scat} and lacks a general analytical solution, although several algorithms to numerically solve it have been proposed (Hogan *et al.*, 1991, Brunetti and Golosio, 2001). At high energies and not too small angles the scattering cross-section is dominated by the Compton contribution, through the momentum transfer $q = E_0 \sin\frac{\theta}{2}$. For high values of q the free-electron approximation holds, and the Compton scattering probability per unit path length per unit solid angle can be written simply as the product of the electron density n_e and of the Klein–Nishina (KN) free-electron scattering cross-section, see Section 4.2.5. Expanding n_e one obtains:

$$p_{\text{Compt}}(E_0,\theta,\phi) = \sum_i N_A \, \rho_i \frac{Z_i}{A_i} \frac{d\sigma_{\text{KN}}}{d\Omega}.$$

The Z/A ratio is generally very close to $1/2$ except for H isotopes and very heavy elements, so we finally write:

$$p_{\text{Compt}}(E_0,\theta,\phi) \approx \frac{1}{2} N_A \, \rho \frac{d\sigma_{\text{KN}}}{d\Omega}.$$

This is an approximate relationship that can be readily used for estimating the spatial distribution of the mass density from the Compton tomography reconstruction. In the general case, an analytic solution to eqn 8.3 does not exist although some algorithms have been proposed for a numerical solution of this problem (Golosio *et al.*, 2004). A good approach, both in terms of accuracy of the results and convergence properties, is a variation of the *simultaneous algebraic reconstruction technique* ( ) algorithm (Andersen and Kak, 1984), first used for transmission tomography, modified in order to take into account absorption corrections (Grant *et al.*, 1995).

8.1.2.4 *Fluorescence tomography* Fluorescence tomography is based on the signal produced on an energy-sensitive detector, generally placed in the horizontal plane at 90° to the incident beam caused by photons coming from fluorescence emission. For an element m, the expression of the fluorescence signal is formally identical to that of the Compton/Rayleigh signal with the difference that the cross-section of scattering is replaced by the fluorescence one and the absorption coefficient at the scattered energy (close to the incident energy E_0) is replaced

by the one at the fluorescence emission line E', potentially at significantly lower energies. The fluorescence probability per unit path length and per unit solid angle is:

$$p_m(s,u) = \frac{1}{4\pi}\frac{N_A \rho_m}{A_m}\sigma_m(E_0), \tag{8.4}$$

where σ_m is the fluorescence cross-section of element m at the incident energy E_0. The expression of the fluorescence signal is similar to eqn 8.3:

$$S^m_{\text{fluo}}(s,\theta) = \int dX f(\theta, X) g_m(\theta, X) p_m(X) \delta(s - \boldsymbol{A}\cdot\boldsymbol{X}).$$

Here, g_m is written as in eqn 8.1, except that the energy E' is replaced by E_m, the energy of the fluorescence line of element m. Note that while $\mu(E_0)$ can be reconstructed from transmission tomography, $\mu(E')$ is a difficult quantity to evaluate. As a practical solution, the absorption at the fluorescent line energies is sometimes scaled as E^{-3}, independent of the composition under the assumption that the main contribution to the absorption is the photoelectric effect. Indeed, following the Bragg–Pierce law (see Section 4.2.3), one can write the following equation for the photoelectric absorption coefficient:

$$\frac{\mu(X,E')}{\mu_0(X,E_0)} = \left(\frac{E_0}{E}\right)^3.$$

This approximation is justified only if the scaling interval is not in the vicinity of the absorption edges of the sample elements or is much larger than these. A first attempt to solve the problem of estimating the absorption coefficient of a sample at all energies was made by Simionovici *et al.* (2000) and Chukalina *et al.* (2002), who iteratively summed all the contributions of the elements whose fluorescence was detected in the sample. The total absorption coefficient at energy E was given by:

$$\mu(x,y,E) = \sum_i \mu_i(E) c_i(x,y),$$

and was calculated over all the fluorescing elements i of concentrations c_i. The disadvantage of this approach is that it neglects the low-Z elements and thus overestimates the concentrations of the high-Z elements in the sample.

Schroer (2001) solved this problem by a self-consistent approach based on the symmetry of the measured fluorescence signal as a function of the translation step and rotation angle, so-called *sinograms*. He calculated the absorption coefficients at the energies of the fluorescence lines by extrapolating from the incident energy E_0 and factoring in the element densities calculated in the previous step:

$$\mu(x,y,E) = E^{-3}\sum_i a_i \rho_i(x,y), \tag{8.5}$$

where a_i are element-specific proportionality coefficients. His calculation starts with the highest Z in the sample and continues down to the lowest Z imaged.

The sinograms for all elements are affected by self-absorption, producing an asymmetric sine function with respect to the inversion $x \rightarrow -x$. By applying a correction to this sinogram, one restores the symmetry and thus corrects for absorption. This approach runs into difficulties in cases of strong self-absorption and/or weak signals. It is therefore desirable to collect additional information as is done by recording the Compton signal.

La Riviere and Billmire (2004) used a similar approach for estimating the unknown attenuation map at the energy of the K emission of a given element, expressed as a linear combination of empirically fit analytic expressions for X-ray attenuation coefficients as a function of energy and the element's own unknown distribution. An iterative scheme of estimating the distribution of the element of interest is established using a penalized-likelihood optimization method that is computationally fast and allows for an explicit modelling of the Poisson photon-counting statistics associated with XCFT. No special treatment of the elements for which no fluorescence lines are detected is performed. Again, as with Schroer's approach, there might be practical limitations associated with strong self-absorption, once more a motivation to seek for more information.

A method was developed that provides a numerical solution to the reconstruction problem by combining the information that comes from transmission, Compton/Rayleigh and fluorescence tomography (Golosio *et al.*, 2003*b*). This method uses so-called *optimal-estimate functions* to extrapolate the absorption coefficient $\mu(E')$ at all fluorescence line energies from $\mu(E_0)$ and $p_{\text{scat}}(E_0, \theta, \phi)$. However, for the fluorescence reconstruction it is only necessary to estimate $\mu(E')$ at the energies corresponding to elements not accessible in fluorescence ($Z \leq 15$). If the mass-density distribution of some elements is known one can separate the contribution of the known elements to $\mu(E_0)$ and $p_{\text{scat}}(E_0, \theta, \phi)$ from the contributions of the unknown elements. The reconstruction algorithm starts from an initial guess of the distributions of all the elements with $Z \geq 15$ that can be obtained by a reconstruction with no self-absorption corrections.

8.1.2.5 *Optimal estimate functions for global parameters* As the solution relies on Compton scattering measurements, the geometry adopted for the setup features a fluorescence detector, located in the vertical plane at 90° from the incident beam. The sample is placed on a rotation-translation stepper system for translation along the vertical direction perpendicular to the beam and rotation around a horizontal axis perpendicular both to the beam and to the translation direction.

One can represent the unknown elemental densities as a vector $\boldsymbol{\rho}$ of n by definition positive elements, with n the number of chemical elements present in the sample. From transmission and Compton tomography two global quantities were obtained that depend on the vector $\boldsymbol{\rho}$: the absorption coefficient at the incident energy $\mu(E_0)$ and the scattering probability $p_{\text{Compt}}(E_0, \theta, \phi)$. These two global quantities B and C can be written as two scalar products between ρ and vectors $\boldsymbol{\beta}$ and $\boldsymbol{\chi}$:

$$\begin{cases} \boldsymbol{\beta} \cdot \boldsymbol{\rho} = B \\ \boldsymbol{\chi} \cdot \boldsymbol{\rho} = C \end{cases} , \rho_i \geq 0 \quad \forall i. \tag{8.6}$$

It is assumed that this system of equations has at least one solution so there is a subset S with $l \leq i$ for which ρ_l satisfies the conditions. We want to evaluate an unknown quantity D that obeys a similar equation: $D = \boldsymbol{\delta} \cdot \boldsymbol{\rho}$. The quantities B and C are physical observables and the transmission coefficient and the scattering probability are all positive quantities. Given $\boldsymbol{\rho}$, B and C as non-negative quantities, D is non-negative as well and is bracketed from above and below. As this set is a closed one, it follows that it has a minimum and a maximum.

$$D_{\max} = \max_{\rho \in S}(\boldsymbol{\delta} \cdot \boldsymbol{\rho}) \quad \text{and } D_{\min} = \min_{\rho \in S}(\boldsymbol{\delta} \cdot \boldsymbol{\rho}).$$

It can be shown that there exists a vector $\boldsymbol{\rho}$ that has two non-zero elements such that we can either directly calculate D or at worst the values $D_{\max}$ and $D_{\min}$. In this last case, we decide to use an optimal value for D that minimizes the largest deviation as:

$$D_{\text{opt}} = \frac{1}{2}(D_{\max} + D_{\min}).$$

The values $D_{\max}$, $D_{\min}$ and D_{opt} are functions of the two measured quantities B and C and will be called *boundary functions* or *optimal-estimate functions*, respectively. It follows from the linearity of eqn 8.6 that D_{opt}/B, $D_{\min}/B$, $D_{\max}/B$ are functions of B/C only. As the elements with $Z \geq 15$ emit detectable fluorescence radiation and thus can be directly reconstructed by fluorescence tomography, we merely have to estimate the values of B and C for all elements undetected in fluorescence as an average of $Z \leq 15$. If, at a later time, one can reconstruct any of these elements individually, its values will be removed from the average values and a new average will be calculated.

An application of this formalism is the calculation of optimal estimates of $\mu(E')$ and ρ as a function of $\mu(E_0)$ and $p_{\text{scat}}(E_0, \theta = \pi/2, \phi = \pi/2)$ for the measurement of the elemental distribution in a Martian meteorite (see Section 8.1.7.1). The incident energy was 25 keV and the global estimates were made for $6 \leq Z \leq 15$ at the Ca K_α line ($E = 3.69$ keV). In Fig. 8.2 the dashed lines represent the corresponding boundary functions. As a comparison, the horizontal dash-dotted line represents the value of $\mu(E')/\mu(E_0)$ evaluated using the E^{-3} scaling law.

8.1.3 *Reconstruction algorithm*

The optimal-estimate functions can be used for extrapolating $\mu(E')$ from $\mu(E_0)$ and $p_{\text{scat}}(E_0, \theta, \phi)$. The original idea is to use these functions only for the elements that are not accessible to fluorescence measurements. As stated previously, when the spatial (mass)-density distribution of some elements is known, one can separate the contribution of the known elements to both $\mu(E_0)$ and $p_{\text{scat}}(E_0, \theta, \phi)$ from the contribution of the unknown elements. The reconstruction algorithm that has been used for fluorescence tomography is a modification

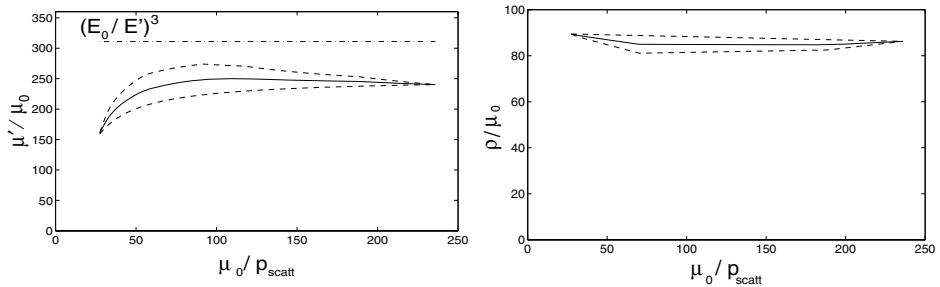


FIG. 8.2. Optimal estimate functions (solid lines) and boundaries (dashed line) for μ and ρ for $6 \leq Z \leq 15$ (Kak and Slaney, 1987, Schätzler, 1978), incident energy $E_0 = 25\,\text{keV}$, at the Ca K_α energy (3.69 keV).

of the simultaneous algebraic reconstruction technique (SART) algorithm (see Section 2.2.2) (Tanabe, 1971) that has been tailored in order to take into account the absorption corrections. Optimal-estimate functions have also been incorporated for extrapolating the absorption coefficient at all fluorescent line energies. This algorithm will be referred to as *integrated tomographic techniques* (ITT).

All spatial distributions are sampled on a square grid of $N \times N$ pixels, with N the number of translations (assuming a sufficient number of rotations) and the pixel size typically equal to the translation step of the tomographic scan. The reconstruction algorithm starts from an initial guess of the distributions of all the elements with $Z > 15$. This guess can be obtained by a reconstruction with no self-absorption corrections. For the coverage of the sampled volume, the beam is modelled as a strip having a width equal to the translation step of the actual measurement, which is itself chosen roughly equal to the transverse size of the focused beam. The weight of the contribution of each pixel k to the fluorescence signal p_k^m of element m at the angular step θ_i and translation position s_j is evaluated as:

$$w_{jk}^i = \frac{a_{jk}^i}{\tau},$$

where $k = 1 \ldots N^2, i = 1 \ldots$ number of rotations, $j = 1 \ldots$ number of translations, and a_{jk}^i is the area of the intersection of the pixel k with the beam of transverse size τ. For each element m, the path integral of eqn 8.4 along a ray is approximated by a discrete sum of pixels as:

$$S_j^{m,i} = \sum_k w_{jk}^i K_{jk}^{m,i} p_k^m, \tag{8.7}$$

where $K_{jk}^{m,i}$ is the absorption correction factor of eqn 8.2 for the pixel k of element m at the rotation step i and translation step j. Since the correction factor K depends on the absorption coefficient at the fluorescence energy $\mu^m(E')$, eqn 8.7 establishes a relationship between the two unknown distributions $\mu(E')$ and p_k^m,

while the other relationship is given by the optimal-estimate functions. An image guess at step s_j and angle θ_j is calculated as:

$$Q_j^{m,i} = \sum_k w_{jk}^i K_{jk}^{m,i} p_k^m. \tag{8.8}$$

The guess is then compared to the measured signal of eqn 8.7 and their difference is minimized. The ART algorithm updates all pixels intersected by the ray in order to correct the path integral. Pixels are corrected by the weight:

$$\Delta p_k^m = w_{jk}^i \frac{S_j^{m,i} - Q_j^{m,i}}{\sum_{k'} K_{jk'}^{m,i} (w_{jk'}^i)^2}. \tag{8.9}$$

For each angle θ the correction is always applied starting from the ray closest to the detector (position s_0) outwards because for this ray there is no sample absorption in the path from the fluorescing point to the detector. In this case, the correction factor K can be readily evaluated even if the absorption coefficients μ^m are not known. For each ray k, the values of p_k are updated then those of μ_k using the optimal-estimate functions. Each time a new ray is considered, the values of μ_k have already been updated on all pixels between the detector and the specific ray. One ART iteration is completed after having performed the correction in eqn 8.9 for all k pixels of the image.

The convergence rate of this iterative algorithm depends on several factors. The order in which the k pixels are updated is important as this strongly affects the convergence speed of the algorithm. In fact, the order of updating is important for the angular steps (Hamaker and Solomon, 1978) and here a random call-up scheme was used. Furthermore, in order to increase the convergence speed, an underrelaxation of the amplitude of the corrections was used, already pointed out by Herman *et al.* 1978. A linear multiplicative coefficient that varies between 0 and 1 was applied to the correction values as a function of the iteration number.

Application of the reconstruction algorithm requires several simplifications and assumptions to reduce the complexity of the problem and the necessary calculation time. The sample is divided into a series of volume pixels ('voxels') of width and height equal to the beam size. The depth co-ordinate is projected along the impinging pencil beam and a single value is recorded from the fluorescence simultaneously emitted by all the illuminated voxels in this path.

The following approximations are made throughout the calculation:

- the correction for the finite horizontal (vertical for horizontal scans) size of the beam is not applied, i.e. the sample is considered to be a two-dimensional slice with no concentration variations in the perpendicular dimension.
- the 'enhancement effects' due to secondary/tertiary fluorescence are not considered,

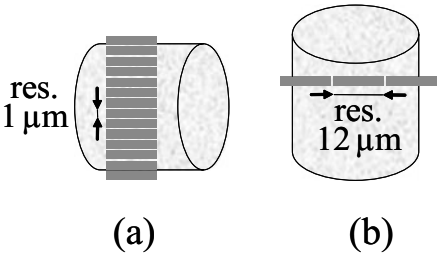


FIG. 8.3. Rectangular beam spot used in tomography with a horizontal (a) and vertical (b) rotation axis.

- secondary fluorescence excited by Compton or multiple scattering in the sample is neglected.
- secondary excitation of fluorescence by photoelectrons emitted from the sample is neglected.

These approximations reflect geometrical factors neglected in the calculation as well as second-order effects for the absolute quantification.

8.1.4 *3D fluorescence tomography helical scan*

After slice XCFT has been carried out the next step is three-dimensional quantitative elemental imaging as it is in transmission tomography. The quantitative part having been solved using the ITT algorithm, the 3D part comes under scrutiny. Several limitations exist for this measurement, the most important being the very long acquisition time. In a first generation of 3D fluorescence tomography a series of adjacent slices of small samples were acquired in view of later assembling them into a volume (Menez *et al.*, 2001). On a standard undulator beamline, using a single focusing element, as was the case in this experiment, the focused beam is a scaled down image of the electron beam footprint with the same aspect ratio. In this case this translates to beam spots of $1 \times 12\,\mu m^2$ (as seen in Fig. 8.3) that are not optimal for 3D imaging where a more square beam is necessary.

Indeed, the resolution in the slice plane, if a sufficient number of rotations are made, is at best equal to the translation step in the direction of the smallest beam size ($1 \times 1\,\mu m^2$), while the resolution in the perpendicular direction is equal to the beam size in the other direction. The other insurmountable obstacle was the time required by the communications protocol controlling the scan (also called 'overhead') imposed by the step by step acquisition, which was of the order of 0.5 s/point. This means that a 50% counting time for 1 s acquisition that was clearly unacceptable. The very long alignment time for a 3D scan in order to avoid precession of the sample as well as the padding space added to the scan on both sides of the sample when scanning in translation non-cylindrical samples were other sources of waste, which made a $100 \times 100 \times 100\,\mu m^3$ volume scan unfeasible.

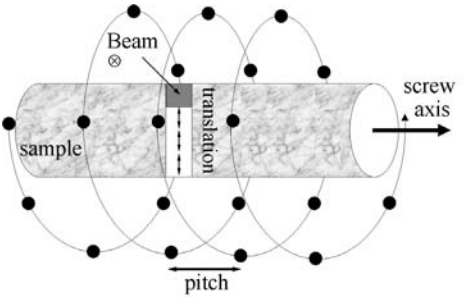


FIG. 8.4. Helical scanning geometry.

These problems were addressed for the first time in the PhD work of Golosio (2003) at the ESRF. He realized that spiral tomography routinely applied for medical 3D imaging in absorption tomography (Kalender *et al.*, 1989) was a readily transferable technique to fluorescence tomography. The only modifications to the previous setup was the change of the rotation drive into a helical one (see Fig. 8.4) composed of two movements: the previous rotation around the longitudinal axis perpendicular to the beam and to the main translation direction and a simultaneous linear translation along this longitudinal axis mimicking the spiral motion of a screw. For this screw the longitudinal translation interval used for a full 360° rotation is called the *pitch*. The other scan in translation is made as before in standard XCFT, here in the vertical direction for better resolution due to beam sizes of about $V \times H = 2 \times 3\,\mu\text{m}^2$.

This helical scan acquisition has certain advantages: compared to the volume obtained from adjacent slices at fixed resolution equal to the slice thickness it has the possibility to reconstruct volumes at any chosen resolution up to the best resolution. Furthermore, for equivalent doses (or collection time), the helical CT longitudinal resolution is superior to that in conventional CT. For optimal dose/resolution efficiency, helical CT needs about 4 to 5 slices to be reconstructed per helical full (360°) pitch. For the same resolution, conventional CT needs a 30% spatial overlap or a 60% higher X-ray dose or acquisition time (Wang and Vannier, 1994). As the screw movement is continuous, the intermediate longitudinal slices are reconstructed using a full-scan interpolation scheme (Crawford and King, 1990). This scheme uses a standard interpolation algorithm to calculate the specific slice value by using the closest parallel projections located at −180° and 180° from the desired angle value. Thus, the interpolation is spanning a full pitch length, and one is at worst half a pitch away from the actual measured parallel angles.

8.1.5 *Instrumental setup*

The general setup of the photon microprobe installed at the ID22 synchrotron beamline of the European Synchrotron Research Facility in Grenoble is presented in Fig. 8.5. Tomography can be performed by rotating around either of two axes:

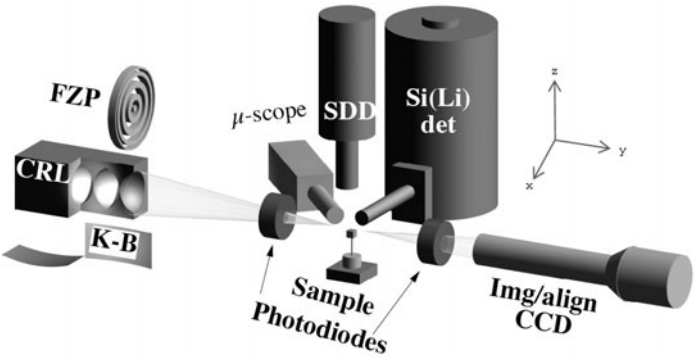


FIG. 8.5. Schematic setup of the ID22 microprobe at ESRF.

a vertical one perpendicular to the experimental table shown in Fig. 8.5 and a horizontal one. The horizontal axis is perpendicular to the beam direction and is especially used for fluorescence tomography. Indeed, the horizontal axis of rotation is associated to the vertical translation direction for which the fluorescence detector is located vertically above the sample (SDD detector in Fig. 8.5) in order to record an intense Compton signal. The two movements used for the tomographic scans are Z/X (vertical/horizontal translations, precision 0.1 μm) and R_X/R_Z (rotation around a horizontal/vertical axis, precision 0.001°. When necessary, the Z/X translations are replaced by piezo-drives of limited range (100 μm) but higher resolution (20 nm). The other movements (Y, R_Y, x) are used to align the sample rotation axis in the beam[3]. The sample is mounted on a Huber goniometer head that is pre-aligned on a Huber visible microscope setup in order to bring its rotation axis perpendicular to the beam and to reduce its precession at the incident beam position.

The focusing optics currently used is an achromatic system of elliptically bent mirrors used in the Kirkpatrick–Baez (KB) geometry (Kirkpatrick and Baez, 1948) which only in the past few years reached small enough slope errors[4] to produce micrometre-sized beams. By coating the two mirrors with multilayers, the incident Bragg angle becomes about ten times larger than the critical angle needed for total reflection on plane mirrors. Thus, the aperture of this kind of optics becomes very convenient and in conjunction with the large bandwidth of multilayers, the photon flux on the sample can be extremely high. In this case the system is no longer achromatic and this combination should be reserved for microfluorescence analysis of ultratrace elements, including fluorescence tomography. Recently, KB systems of very high flux ($> 5 \times 10^{11}$ ph/s) or very small

[3] x is a small drive in the same direction as X but located below the Z, X scanning movements that produce the translation and the screw movement. Its role is to align the desired part of the sample in the beam, after which the other drive X is used only to simulate the screw movement.

[4] First-order geometric errors in the profile of a bent mirror.

beam spots ($\approx 100\,\text{nm}$) have been commissioned at the ESRF. One such system is currently used at ID22 (Hignette *et al.*, 2003). Alternatively, a Fresnel zone-plate (FZP) lens can be used for focusing at energies below 10 keV or a set of compound refractive lenses (CRL) for high-energy focusing ($E > 20\,\text{keV}$).

The detectors used are energy-dispersive ones, either a 13-element Si(Li) (Gresham Scientific Instruments) for horizontal detection, or a silicon drift diode (Ketek GmbH) for vertical detection and high count rates. These detectors record the energy spectra of emitted photons and thus acquire the fluorescence and Compton signals necessary for the ITT algorithm. The other signal recorded is the transmission signal that is acquired by the beam intensity monitors placed before and after the sample in transmission. In Fig. 8.5 these monitors are a pair of Si PIN photodiodes, which are coupled to either a microamperemeter or a digital $I \rightarrow f$ (current to frequency) converter. The incident flux recording by the monitor upstream of the sample serves for normalization of all acquired signals to compensate for the long-term synchrotron current decay and to account for rapid fluctuations. When lower-energy incident beams are used the Si photodiodes cannot be used as they absorb a large part of the incident flux. In this case, ionization chambers filled with either argon or nitrogen gas are used for the incident flux monitor. Alternatively, hybrid monitors using an offaxis PIN diode that records the fluorescence emitted by a thin Al or capton foil are used. A calibrated and previously aligned video microscope of large aperture is used in order to observe the sample zone under the beam. The microscope is connected to a CCD camera that allows storing sample images for comparison and identification of their morphology.

Finally, an ESRF-built, high-resolution, Fast REad-out, LOw-Noise (FReLoN) 14-bit CCD camera of 2000×2000 pixels optically coupled to a LAG:Eu thin scintillator (see Chapter 10) is used in transmission behind the sample, for recording the 3D absorption tomography projections using incident unfocused beams of $1 \times 1\,\text{mm}^2$ cross-section.

8.1.6 *Time-saving strategies in fluorescence tomography*

The most important obstacle in the development of XCFT is the exceedingly long time taken by a 3D acquisition. Nowadays, acquiring routinely a $50 \times 50 \times 50\,\mu\text{m}^3$ volume of trace-element levels of a distribution of intermediate-Z atoms with a $1\,\mu\text{m}$ effective resolution would take about 18 h including sample alignment. Alignment is performed so as to reduce the precession of the sample around the rotation axis in order to minimize the time spent scanning a large volume corresponding to the convolution of the sample volume with the envelope of the displacements of all three drives used for the tomographic scans. This can be reduced by means of optimizing the scanning strategies as described in the paragraph below.

8.1.6.1 *Smart scans and optical shape-recognition scans* Whenever an object of non-circular cross-section for 2D slice tomography (non-cylindrical shape for 3D tomography) must be tomographed, one must scan relatively large empty

volumes where the beam does not intersect the sample. This is due to the fact that the scanning is performed by simple movements covering circles or cylinders owing to the rotation around the tomographic axis. In addition to this wasteful strategy, a possibly even larger volume must be scanned if the sample has a precession movement around this axis. As samples are hardly ever perfectly symmetrical nor optimally placed (at normal incidence) in the sample holder most of the time, one must accept a non-negligible amount of precession. This can be effectively reduced if one uses a goniometer head as a sample holder and reduces the precession at a specific position on the rotation axis where one wants to perform a slice tomography. It is always possible to centre that particular slice at minimum precession with respect to the centre of rotation by using the two manual translations and two manual rotations of a goniometer head[5]. If 3D tomography is desired, the centring works in as much as the volume to be scanned has a symmetry axis that can be rendered collinear with the rotation axis. However, even if a careful alignment has been performed on a special Huber microscope devoted to precession alignment, the sample holder of the beam setup is not necessarily calibrated to the Huber microscope so another alignment must be performed once the sample is in the microprobe setup. This is often time consuming as the beamline precision drives are slow and the axis alignment procedure takes up to several hours.

A first-order optimization of the scan would center the specific slice located close to the volume centre to be concentric with the rotation axis. For a slice tomography, if one accepts an amount of precession, this translates into expanding the radius of the circle to be scanned until the precessing slice is fully comprised inside the circle of rotation. This is called the 'first level of optimization' of the alignment. The translation scan, which is the one that can be optimized, is then constant and equal to the diameter of the circle of rotation, see Fig. 8.6(a).

A second level of optimization can be achieved if instead of a scan of constant length one estimates a shorter length. This can be done by performing a fit at 4 angular positions of the rotation angle, namely θ equal to 0°, 90°, 180° and 270°. A simple fitting function such as a sine wave can be applied to compensate for the precession of an object offcentre of rotation. In Fig. 8.6(b) the simple sine-wave fitting procedure presented can lead to time savings of about 30%, depending on the shape of the object and its offcentre position.

A procedure was devised whereby the translation scan is performed only on the effective transverse size of the sample at the specific angle of rotation θ, Fig. 8.6(c). The procedure was called the *smart scan* as it is semi-automatic in its action. Once a detector signal or function of detector signals is defined as a counter to optimize and a threshold value is selected, the smart scan uses this counter to monitor the intersection of the beam with the sample. As soon as the smart counter falls below the pre-defined threshold value, the procedure considers that the beam is outside the sample and stops the scan. A new scan is performed

[5]In this case a goniometer by Huber Diffraktionstechnik GmbH.

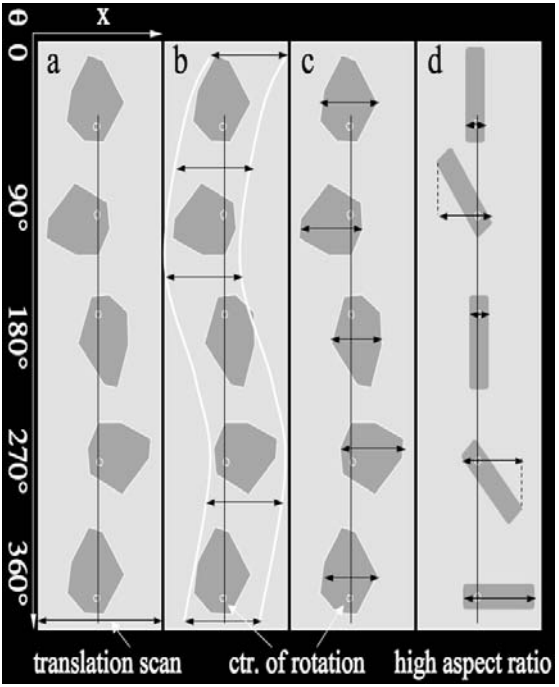


FIG. 8.6. Scanning strategies for fluorescence tomography.

at the same position of the last scan plus or minus a pre-defined number of steps used as a limit. If the smart counter does not go above the threshold value within the limit number of steps the sample position is considered lost and a new scan of a large range called 'scan max' is performed. The limits of this 'scan max' are chosen so as to encompass any possible precession and sample size and any rotation θ. The 'scan max' is longer but guaranteed to intersect the sample at some position. Once the smart counter has gone above the threshold value, the sample has been intersected so the procedure continues from the intersect position further. The θ angle is incremented and the smart scan continues, for an optimized translation length. Thus, only the sample transverse length is scanned for any θ position, within the chosen limit number of steps.

The procedure is straightforward if the smart counter is sufficiently sensitive and the threshold value chosen appropriately. Once these conditions are met, the procedure continues automatically until all the sample has been scanned. This guarantees a maximum time saving regardless of the amount of precession or sample shape. The only cost of this optimized procedure is the time spent in selecting a smart counter and choosing a threshold value. The smart counter cannot be simply chosen among the fluorescence counters of any element, as they are not guaranteed to be present all over the sample. A reasonable choice is a counter obtained by dividing the Compton counter by the incident flux

counter, so as to correct for the variations due to the synchrotron beam decay and fluctuations. The value of the threshold must be carefully chosen as this is the critical measure of the time saving of the procedure. If both smart counter and threshold value are well chosen the procedure is optimal and the time savings are maximum. If the threshold is too high the scan will stop before reaching the sample edges, thus rendering it useless. If the threshold is too low the scan will be overstepping the sample edges and thus be too long with only negligible time savings.

The smart scan procedure will obviously correct for cases where the sample is attached to the sample holder in a position that has a large amount of precession. The smart scan will also correct for translation scans of non-cylindrical samples, which have a high aspect ratio, as shown in Fig. 8.6(d). This is called the 'second level of optimization' and one cannot do better than this in a normal scan. The only time lost is during the selection of the threshold value and/or the smart counter. These procedures require several tests of the performance of the smart scan and can take up to a two or three hours if the sample is of low total density. If the sample has voids or cracks the smart scan will inevitably fail unless one chooses a large limit in the number of steps added to the optimal scan length. This guarantees that the scan will work in most cases unless the void size is larger than the chosen limit size. However, if one selects too large a limit size, the smart scan will take too long and be unavoidably time consuming. Therefore, work on implementing a scan of optimal size based on the optical images of the sample acquired by a microscope coupled to a CCD device has been recently started. The actual translation scan co-ordinates are obtained by an optical pattern-recognition algorithm that discretizes the sample shape. Direct scan limits can then be established using the sample transverse size at any height as directly seen in the digitized optical image. The only condition is that the microscope viewing direction has to be the same as that of the incident beam and the position of the focused beam must be indexed to the vertical and horizontal optical image limits.

8.1.6.2 *Fast continuous scans* The overhead time currently observed when performing scans using step by step motor drives is about 0.3 to 0.5 s per point. This is due to the beamline data acquisition and control program called `SPEC` (Certified Scientific Software) that deals with numerous (120 or more) drives, detectors and counters at the ID22 beamline. Therefore, a continuous scanning mode was implemented, the so-called '*zap mode*', which sets the specified drive into movement, accelerating it until reaching the required velocity and then acquiring the fluorescence signals 'on the fly' as the drive continues to move at constant velocity. The drive then slows down and returns to the beginning of the scan for a new pass. The acceleration and deceletarion phases of the movement are not used for fluorescence acquisition. Therefore, the scan is slightly larger than the required scanned dimension. The only overhead time is spent in these acceleration and deceleration stages, but these do not contribute significantly

to the total moving time. The constant speed of the drive can be changed in a range of settings between approximately 0.1 to 2 s dwell time for stepper drives or even more for piezo-drives. The total equivalent time savings provided by this technique is about 50% for dwell times of 1 s per point or less. Two-dimensional maps are acquired using only one of the drives in zap mode as the other drive moves only after each line is finished and thus does not require zapping.

8.1.6.3 *High count rate detectors* Practically all applications of synchrotron radiation using solid-state detectors – Li-drifted Si (Si(Li)) and high-purity Ge detectors (HpGe) – are subject to saturation effects (pile-up) in their spectra as soon as the fluorescence count rate is higher than 15 000 cps (counts per second). With third-generation synchrotron sources, incident beam intensities for focused beams, regardless of the focusing device, are in the 1×10^{10} ph/s range or more. Several microprobe beamlines have focused beam intensity as high as 5×10^{12} ph/s and must therefore use absorbers in the beam path in order to reduce the intensity to suitable count rates. Recently, driven by the need to accept high count rates a new class of detectors have appeared – the *silicon drift diode (SDD)*. These compact detectors feature a Peltier-effect cooling stage, thus avoiding the impractical liquid-nitrogen cooling, and have a small weight and footprint allowing integration in complex sample environments. Their dedicated electronics allows enormous count rates, in excess of 1 Mcps, with little or no pile-up. Furthermore, the energy resolution of some of these devices is similar to that of cryogenically cooled detectors and the position of their peaks does not vary with the count rate, as is generally the case with SDDs. One such example is the Vortex (SII Nano-Technology Inc.) which is the largest-area (50 mm^2) single-element detector capable of a resolution of 136 eV at the Mn K_α-edge (6.5390 keV) and count rates of about 600 kcps. Using such detectors will enable taking advantage of the full intensity of the beams while drastically reducing the acquisition time required to obtain a given level of statistics. Currently, for focused beam intensities comparable to those obtained at ESRF, ID22 (1×10^{12} ph/s), Vortex detectors could be used to acquire fluorescence signals in only 0.1 s from $Z \geq 15$ elemental concentrations of 1 ppm or more.

8.1.6.4 *'1.5th generation' setup* Due to the isotropic emission of the signal and to the fact that the whole chord of the sample intersected by the beam becomes a secondary source, in fluorescence tomography the acquisition is made using first-generation setups, i.e. using a pencil beam and a single-element detector. Furthermore, given the self-absorption of the fluorescence radiation emitted by an element from a sample point into all the pixels between the detector and the respective point, the rotation movement is performed over the entire 360° range. In transmission tomography, 180° acquisitions are sufficient, given the forward beam propagation and the detector position in the forward direction. In XCFT, the detector is placed at 90° with respect to the incident beam, in either the vertical or horizontal plane and records the signal from sample pixels located at different distances from it. Therefore, in order to collect the same signal as the

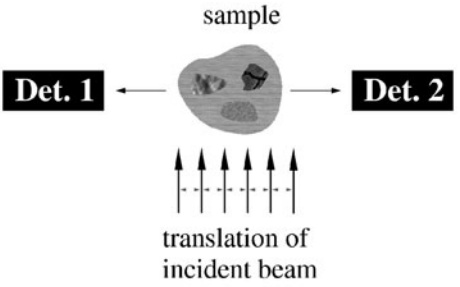


FIG. 8.7. Two-detector system for time savings in fluorescence tomography.

one obtained from a 360° scan, one can use two symmetrical detectors, placed on both sides of the beam, in either the horizontal or vertical plane and perform only a 180° scan. This requires precise alignment of the two detectors, placed at exactly 90° with respect to the beam but effectively halves the acquisition time previously necessary. The setup shown in Fig. 8.7 represents a time gain of a factor of two with respect to the first-generation setup and represents an intermediate step between the first and second tomography generations.

8.1.7 *Applications*

8.1.7.1 *2D slice tomography: study of micrometeorites* The planetary sciences were revolutionized a few years ago with the publication of the work of McKay *et al.* 1996 on biogenic traces on Mars meteorites. For the first time, such a claim was made based on morphological and chemical arguments collected from SEM/TEM microscopy scans of a Martian meteorite. For rare and valuable samples it is worthwhile to first subject them to a series of non-destructive measurements such as the ones using synchrotron radiation. Therefore, a measurement protocol for similar samples was perfected based on the study of traces of aqueous alteration of an asteroidal meteorite, i.e. the effect of water interacting with the material.

A measurement was performed on a microfragment of the Tatahouine meteorite, a diogenite fallen in 1931 and promptly stored at the Natural History Museum in Paris (Simionovici *et al.*, 2002). The fragment analysed by XCFT was recently retrieved from the original site and recovered in a weathered state from the soil and was compared to the pristine ones retrieved in 1931 and stored at the Museum. A series of SEM and TEM investigations was carried out on both pristine and weathered fragments of this meteorite in order to study the remnants of pleomorphic bacteria present in fractures and fault lines of the meteorite (Gillet *et al.*, 2000).

For the XCFT measurement, a monochromatic beam of 1 keV was obtained using composite refracting lenses, associated with a 10-μm pinhole to define a $2 \times 10\,\mu m^2$ beam spot. The grain was placed in a sealed thin ($\leq 10\,\mu m$ thick) quartz capillary in order to mount it on the goniometer head. This helped to develop the

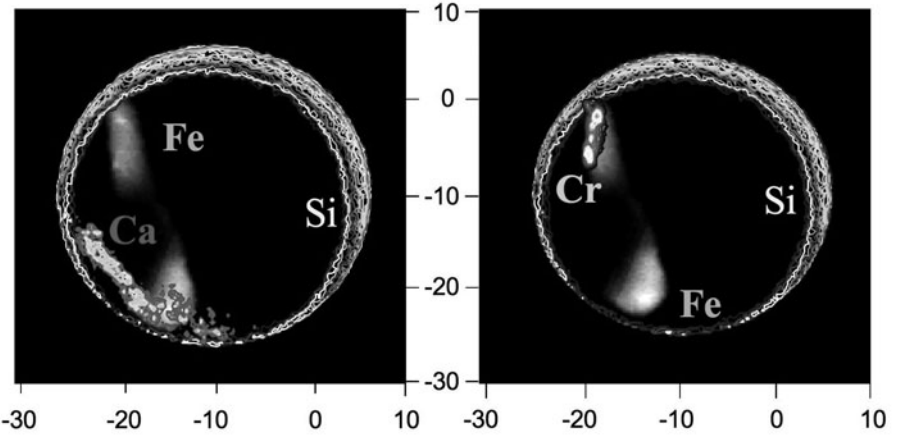


FIG. 8.8. Fluorescence tomograms of a micrometeorite grain inside a quartz capillary. Reconstructions (using ART) of the K_α lines for Si (capillary), Fe, Cr and Ca are shown. Resolution $\approx 2\,\mu$m. This image is reproduced in **colour** in the central part of the book.

methodology of analysing small, rare, unique samples, in quarantine conditions in *ad-hoc* sample holders guaranteeing the planetary protection requirements established by NASA. Tomography was done in the vertical scanning geometry, making use of the best spatial resolution available (2 μm). The detector used was a SDD diode (Röntec GmbH) of moderate resolution (190 eV at 5.9 keV) but capable of high count rates (20 kcps) without pile-up effects (see Section 8.1.6.3).

A biological study of this meteorite (Gillet *et al.*, 2000) revealed bacteriomorphs of sizes between 0.1 to 0.6 μm that were obtained by cultures of the soil surrounding the grains. These were observed in SEM and TEM and are presumed to be real bacteria or their remnants. It was postulated that these bacteria appear at the fracture sites of the grain, following fluid circulation of carbonates from the soil due to terrestrial weathering. Therefore, the study aimed at identifying and locating non-invasively carbonate phases specific to the grain. The analysis was done through the quartz capillary for all elements of $Z \geq 14$ with the exception of Si, which was the main constituent of the capillary walls. In Fig. 8.8, the distribution of Fe, Cr and Ca is shown, with a resolution of 2 μm obtained by ART reconstruction.

The carbonate phase is easily located and quantified as a slab of approximately $15 \times 5\,\mu$m on the outside of the mineral grain. After validation, this measurement protocol was applied to the study of a grain extracted from the Martian meteorite NWA817 and stored in a silica container (Golosio *et al.*, 2003*a*). Special attention was given to the determination of the Fe and Mn contents since they are key values to discuss the Martian origin.

The grain was placed inside a pure silica capillary having a 170-μm internal

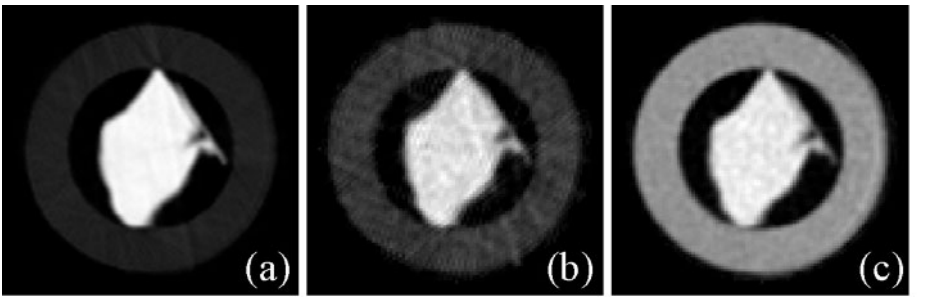


FIG. 8.9. NWA 817 grain reconstructed from (a) transmission, (b) Rayleigh scattering and (c) Compton scattering signals.

diameter and 45-μm thick walls. A monochromatic beam of 25 keV was used, with a beam size of $3 \times 5\mu m^2$ (vertical × horizontal).

A beam spot was obtained using compound refractive lenses for focusing coupled to a 5-μm pinhole to reduce the horizontal size. The measured flux at the sample was about 1×10^{10} ph/s in the focused beam. The sample was scanned with 120 angular steps of 3° and 100 translation steps for each angle. All distributions were reconstructed on 100×100 square grids of 3 μm pixel size.

The distribution of $\mu(E_0)$ was reconstructed from transmission tomography using the filtered backprojection algorithm and is presented in Fig. 8.9(a). The distributions of the scattering probabilities p_{Compt} and p_{Ray} have been reconstructed from the Compton and Rayleigh scattering signals using the SART algorithm, modified for absorption correction, see Figs. 8.9(b) and (c), respectively. The relative contrast between the pure silica capillary and the ferrous silicate grain increases from the Compton to the Rayleigh scattering signal and finally to the signal from transmission tomography, as contrast is proportional to Z, Z^2 and Z^4, respectively, see Section 4.2.2. Iron and manganese gave significant contributions to the fluorescent signal. Their reconstructions with and without self-absorption corrections are presented in Figs. 8.10(a) and (b) (Fe) and 8.10(c) and (d) (Mn). Self-absorption effects are clearly visible in the images without self-absorption corrections, leading to an underestimate of the grey level inside the sample. The average density of iron and manganese, calculated from the distributions shown in Figs. 8.10(b) and (d) are $\rho_{\mathrm{Fe}} = 1.45 \pm 0.07\,\mathrm{g/cm^3}$ and $0.058 \pm 0.017\,\mathrm{g/cm^3}$. The average total density, evaluated from the distributions of Fe, Mn and unmeasured fluorescences is $3.89 \pm 0.3\,\mathrm{g/cm^3}$, while the corresponding weight fractions of iron $\rho_{\mathrm{Fe}}/\rho_{\mathrm{tot}}$ and manganese $\rho_{\mathrm{Mn}}/\rho_{\mathrm{tot}}$ are approximately 37.3 ± 3 wt.% and 1.5 ± 0.5 wt.%, respectively. All these results were compared to the quantitative EDX study of thin sections of the same meteorite (Sautter *et al.*, 2002). Discrepancies between the Fe estimates and the average bulk chemical composition of the NWA817 olivine grains appear of the order of 1 wt.%. For Mn the discrepancies are of the order of 0.5 wt.% and they appear largely due to the deconvolution of the Mn/Fe overlap.

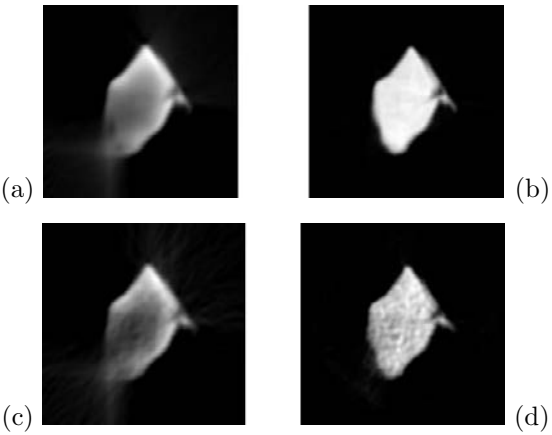


FIG. 8.10. Fe reconstructions (a) without self-absorption, (b) with self-absorption corrections. (c) and (d): same for Mn.

To conclude, a new technique has been developed that combines the information from transmission, fluorescence and scattering tomography. The studied reconstruction methods applied to submillimetre silicates are efficient to determine in any volume of the sample the average densities and weight fractions of the major and detectable elements with precisions better than 10 wt.%. Analysis of the minor elements remains semi-quantitative.

8.1.7.2 *3D study of fly ashes from municipal waste* A very active field in X-ray imaging has been that of environmental sciences that brought forward specific samples requiring compositional and morphological analyses for establishing scenarios of soil and river contamination. An example of this type of study was performed on fly ashes issued from municipal waste processed by an incinerator near Gøteborg in Sweden. The ashes emitted by this facility contained elements such as Cd that was released in a highly toxic state. These samples were previously studied by XRF and XAS spectroscopies, establishing compositional types as a function of their morphology (Camerani *et al.*, 2002).

Figure 8.11 illustrates the internal structure and elemental composition of a single fly ash particle. This particle was part of a protocol of measurements of the morphology and elemental distributions of fly ash particles previously analysed by slice-XCFT and XANES spectroscopy (Golosio *et al.*, 2003*a*, Golosio *et al.*, 2004, Pinzani *et al.*, 2004) in order to unravel its local chemistry. The particle has been scanned at 16 keV in 7 helical turns, 12 μm/turn helical pitch, 120 angular steps per turn of 3° and 80 vertical translation steps of 3 μm.

The dwell time was 1 s/step, which yielded a total measurement time of 18.7 h. 28 slices with 3 μm slice-to-slice separation have been reconstructed. The figure compares the distribution of the absorption coefficient, obtained by full-beam conventional absorption tomography performed with a CCD, to the maps of the

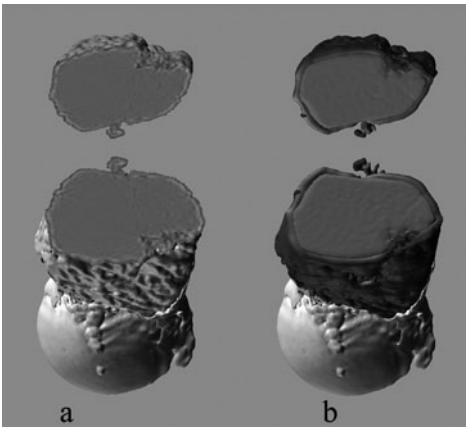


FIG. 8.11. Reconstruction of the density (a), and Mn (blue), Rb (red), Fe (blue) concentrations (b) in a fly ash. This image is reproduced in **colour** in the central part of the book.

concentrations of three elements, rubidium, iron, and manganese, obtained by the reconstruction of the helical scan. The combined helical tomography reveals an external shell of Mn, a Rb core and a fragmented Fe phase. These observations give important complementary information about the possible fly-ash formation mechanisms, e.g. volatilization, condensation, diffusion and solid-solution formation during combustion.

The possibility to study the 3D internal elemental composition and its relation with the sample structure by fluorescence tomography will open new possibilities in several research fields and is expected to improve the understanding of important chemical, physiological, and environmental processes. The potential of the technique in resolution and measurement time can be improved both by developing the reconstruction method and the experimental conditions. A significant reduction of the measurement time can be achieved by optimized setups including modern detectors and fast, smart scanning techniques as pointed out in Section 8.1.6.1. The spatial resolution of the method can be readily improved by using recently developed focusing devices providing spot-sizes in the 100 nm range (Hignette *et al.*, 2003). It is important that this method be generalized as it is the only one achieving non-destructive elemental quantification in the volume of massive and sometimes unique samples.

8.2 X-ray absorption and small-angle scattering tomographies

8.2.1 *Introduction*

A variety of X-ray analytical techniques are available today, such as X-ray fluorescence (XRF) analysis, X-ray absorption spectroscopy (XAS), or X-ray scattering, that yield the elemental composition, the chemical state of a given atomic species, or the (crystalline) nanostructure of the specimen, respectively. One of

the key strengths of hard X-ray scanning microscopy or tomography is that these analytical techniques can be exploited as contrast mechanisms, allowing one to obtain elemental, chemical and structural information from inside a specimen. One example of this has already been given in the previous section, where the element distribution inside a sample was determined by fluorescence tomography. Similar to X-ray fluorescence, other hard X-ray analytical techniques can be used as contrast mechanisms. In this section, we take advantage of absorption spectroscopic and scattering contrast in scanning tomography to obtain the distribution of the chemical oxidation state and the local nanostructure on a virtual section through the sample.

To generate a microbeam, the X-ray source is imaged onto the sample position in a reducing geometry with an appropriate X-ray optic, such as Fresnel zone plates, Kirkpatrick–Baez mirrors, or refractive X-ray lenses, see Chapter 7. This requires a highly brilliant X-ray source. Therefore, these techniques are most efficiently implemented at synchrotron radiation sources. X-ray microbeams with a lateral dimension in the micrometre range are routinely available at synchrotron radiation sources. Recently, hard X-ray nanobeams with lateral sizes down to 50 nm and below have been demonstrated (Mimura *et al.*, 2005, Hignette *et al.*, 2005, Schroer *et al.*, 2005, Kang *et al.*, 2006), and scanning microscopy techniques based on these small beams are currently under development.

8.2.2 *Tomographic imaging by absorption spectroscopy*

Absorption spectroscopy is a powerful tool to analyse the chemical state and local environment of a given element without the need for long-range (crystalline) order in the specimen. Classical absorption spectroscopy is performed on homogeneous samples of well-defined thickness s. The sample is illuminated by monochromatic X-rays, whose energy is tunable around an absorption edge of the element of interest. The experimental setup is shown in Fig. 8.12(a). Two ionization chambers I_0 and I_1 measure the intensity of the radiation before and behind the sample, respectively. From the ratio of I_1 and I_0 measured at a given energy E, the attenuation coefficient $\mu(E)$ can be extracted from Beer–Lambert's law, see eqn 4.11 and eqn 5.1:

$$\mu(E)s = -\log \frac{I_1}{I_0}. \tag{8.10}$$

An absorption spectrum is acquired by scanning the energy in an interval around an absorption edge of the element of interest. From these data, various pieces of information about the chemistry of the element of interest can be extracted. The exact position of the absorption edge yields the oxidation state of the element. The X-ray absorption near-edge structure (XANES) extending to about 50 eV above the edge probes the local projected density of states for the given element. The extended X-ray absorption fine structure (EXAFS) that extends to about 1 keV above the edge gives information about the local chemical neighbourhood of the element, yielding the interatomic distances with the neighbours

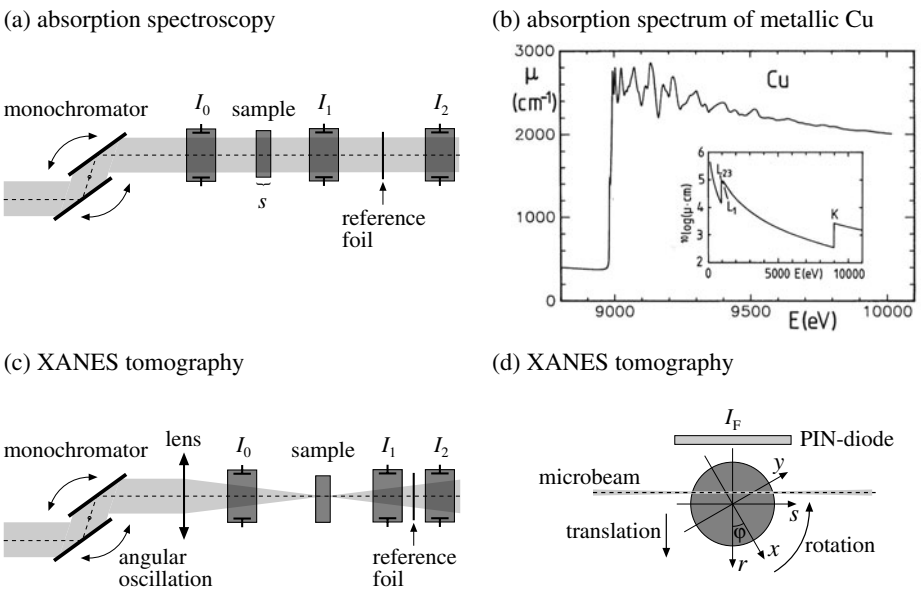


FIG. 8.12. (a) Experimental setup for absorption spectroscopy at a synchrotron radiation source (side view). (b) Absorption spectrum of metallic Cu. The inset shows the absorption spectrum of atomic copper. (c) Setup for XANES microtomography (side view). (d) Detail image of the sample region for XANES tomography as seen from the top. (c,d) reused with permission from Schroer *et al.* (2003), Copyright 2006, American Institute of Physics.

and co-ordination numbers (Koningsberger and Prins, 1988). As an example, the EXAFS spectrum, i.e. μ as a function of E, of metallic copper is shown in Fig. 8.12(b). For comparison, the inset in Fig. 8.12(b) shows the spectrum of copper atoms in the gas phase.

Most absorption spectroscopy experiments are carried out at synchrotron radation sources due to their high spectral brilliance at tunable X-ray energies. The radiation coming from the synchrotron radiation source is monochromatized using a double-crystal monochromator. Since it is difficult to measure the absolute X-ray energy with high accuracy, the absorption spectrum of a known reference sample is measured simultaneously using the ionization chambers I_1 and I_2 in Fig. 8.12(a). A PIN (positive intrinsic negative) diode facing the sample from the side records the fluorescence radiation from the sample, see Fig. 8.12(d).

Absorption spectroscopy is a standard tool at synchrotron radiation sources. It is mostly limited to the investigation of large homogeneous samples. By combining absorption spectroscopy with microbeam or imaging techniques, it is possible to locally analyse spatially inhomogeneous samples such as porous catalysts or biological specimens (Kinney *et al.*, 1986, Ade *et al.*, 1992, Yun *et al.*,

1998, Youn *et al.*, 2001, Grunwaldt *et al.*, 2006). While the analysis of thin samples by mapping is widespread, the investigation of the interior of bulk samples using tomographic techniques is a fairly new field. One approach to tomographic absorption spectroscopy is to acquire full transmission tomograms at a series of energies around the absorption edge of an element of interest (Rau *et al.*, 2002). While this approach yields three-dimensional tomographic reconstructions, the spectral information is usually limited to a few energies by the acquisition time. Alternatively, the sample can be scanned tomographically with a microbeam, acquiring at each position a full absorption spectrum. A major advantage of recording full spectra is that one can reliably fit reference spectra to the measured data and detect even subtle differences in the concentrations of different chemical compounds of the atomic species investigated.

The latter method is implemented at a synchrotron radiation source by combining a quick-scanning monochromator and data-acquisition system (Grunwaldt *et al.*, 2001, Richwin *et al.*, 2002) with a scanning microprobe setup. Similarly to fluorescence microtomography (cf. Chapter 8.1), the sample is scanned across the microbeam in translation and rotation, cf. Figs. 8.12(c) and (d). At each position of a scan, a full absorption spectrum is recorded.

From the transmission data, the integral of the attenuation coefficient along the microbeam is determined using Beer–Lambert's law

$$\int \mu\big(r(x,y), s(x,y), E\big) ds = -\log \frac{I_1}{I_0} \tag{8.11}$$

for each energy in the spectrum. At fixed energy, the tomographic transmission data can be reconstructed using standard tomographic techniques, such as filtered backprojection. This reconstruction is done for each energy, yielding a full absorption spectrum at each point of the reconstruction.

As an example, an experiment is described that was carried out at beamline 1-ID of the Advanced Photon Source (APS) in Argonne, Illinois. In this experiment, a XANES tomogram of a CuO/ZnO catalyst powder inside a glass reactor capillary was recorded around the copper K-edge (Schroer *et al.*, 2003). The microbeam was defined by focusing the synchrotron radiation onto the sample by a refractive X-ray lens made of beryllium. On a given virtual slice through the capillary, 101 projections with 90 translational steps of 10 µm each were recorded over a full rotation of the capillary. At each position of the scan, a full XANES spectrum was recorded in 1 s. For each energy in the spectrum, the attenuation coefficient $\mu(x, y, E)$ was reconstructed from this data, giving at each location in the reconstructed slice a full XANES spectrum. Figures 8.13(a) and (b) show reconstructed slices through the reactor capillary below and above the absorption edge, respectively. The reconstructed spectra are shown in Fig. 8.13(c) for a few locations marked by arrows in Fig. 8.13(b). Note that the spectrum in location (5) is featureless, indicating that there is no copper in the glass of the capillary.

Also shown in Fig. 8.13(c) are three reference spectra of metallic copper, monovalent copper in $Cu(I)_2O$, and bivalent copper in Cu(II)O. These spectra,

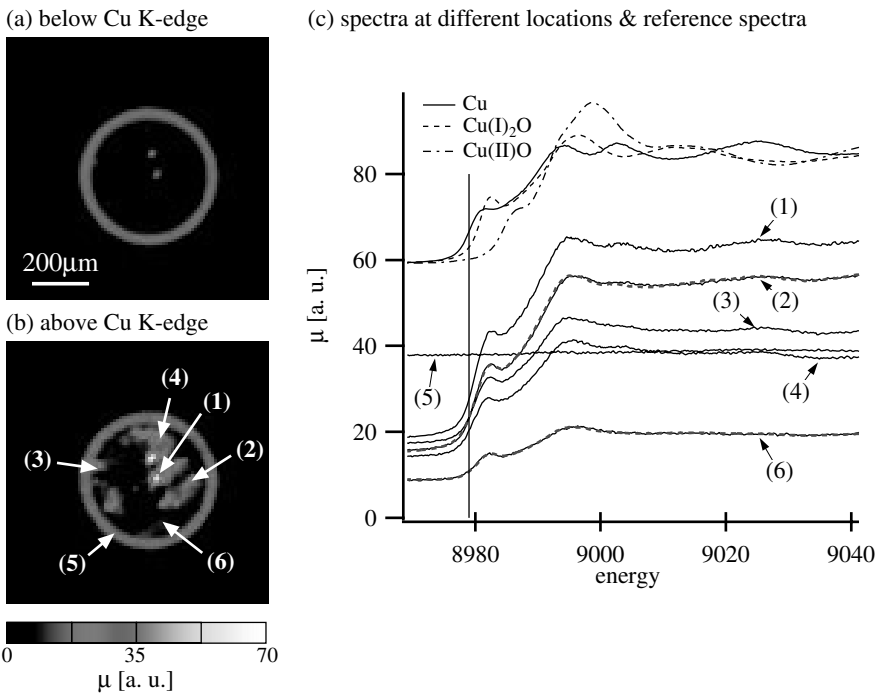


FIG. 8.13. Reconstructed tomogram of a capillary filled with a CuO/ZnO catalyst (a) below ($E = 8970\,\text{eV}$) and (b) above ($E = 8995\,\text{eV}$) the Cu K-edge. (c) reconstructed XANES spectra at different locations marked in (b) and (offset) reference spectra of metallic Cu, $Cu(I)_2O$, and Cu(II)O. The dashed red curves in (c) represent two fits using the reference spectra. This image is reproduced in **colour** in the central part of the book. Reused with permission from Schroer *et al.* (2003), Copyright 2006, American Institute of Physics.

together with a featureless spectrum, can be fitted to the spectrum at each location in the tomographic reconstruction, giving the relative concentration of each oxidation state at each location together with the attenuation of other components that do not contain copper. The resulting distributions are given in Figs. 8.14(a) through (d). The distribution of metallic and monovalent copper are very similar, while the concentration of bivalent copper is below the detection limit of the method. To show the differences in the distribution of metallic and monovalent copper, the difference map between the two species is given in Fig. 8.14(e). Lighter regions in Fig. 8.14(e) are those with a higher monovalent copper concentration while darker regions contain more metallic copper.

Following this first demonstration experiment the method can be developed further and put to use. For example, the scanned energy range could be extended to the EXAFS regime to record EXAFS tomograms. While a quick-scanning

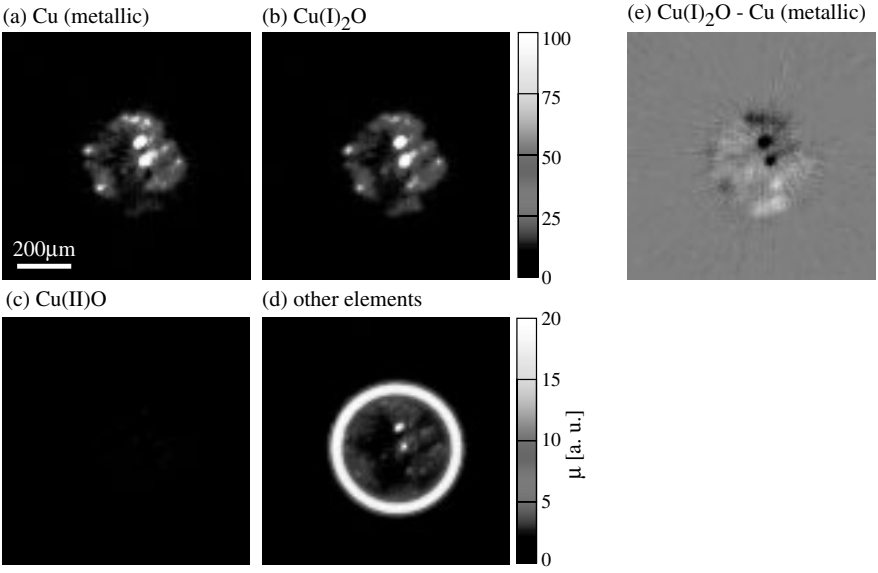


FIG. 8.14. Relative concentrations of (a) metallic and (b) monovalent Cu inside the glass capillary on the same scale. (c) difference image between (b) and (a). The shade of grey outside the specimen corresponds to zero difference. Darker areas depict a higher concentration of metallic Cu, while brighter regions signify a higher concentration of Cu(I)$_2$O. (d) attenuation generated by elements other than Cu. This image is reproduced in **colour** in the central part of the book. (a,b,d,e) reused with permission from Schroer *et al.* (2003), Copyright 2006, American Institute of Physics.

monochromator (QEXAFS monochromator) is capable of scanning such a large energy range (Richwin *et al.*, 2001), the chromatic aberration of refractive lenses is too strong to allow for the use of these optics in such an experiment. Achromatic optics, such as a Kirkpatrick–Baez mirror system will be needed to create a microbeam whose size and position is unaffected by large energy variations.

8.2.3 *Tomographic imaging by small-angle X-ray scattering*

X-ray scattering is the most prominent method for structure determination of (crystalline) materials on the atomic level. The scattering signal around the forward direction, the so-called small-angle X-ray scattering (SAXS) regime, reveals the structure of a specimen on scales of nanometres to several hundred nanometres. In this regime a lot of structural information can be obtained, also from non-crystalline and disordered samples. By combining small-angle scattering with scanning microscopy the variation in the local nanostructure of a heterogeneous specimen can be investigated. To obtain nanoscale information from a virtual section through a sample, scanning tomography can be combined with

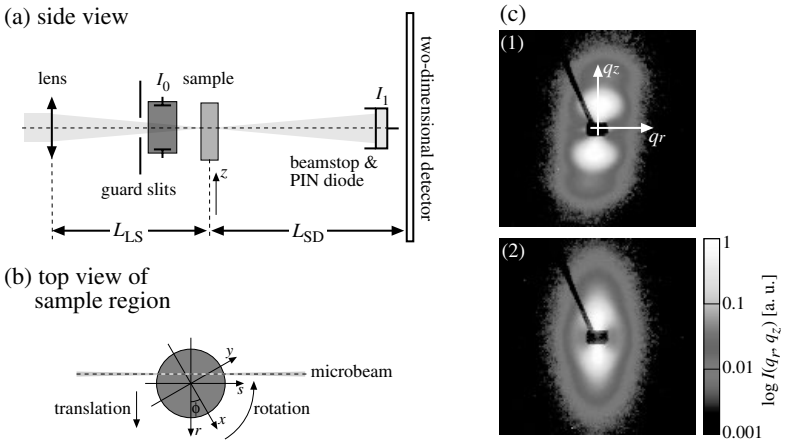


FIG. 8.15. (a) Side view of the setup for micro-small-angle scattering tomography. (b) Top view of the sample region. (c) Small-angle X-ray scattering patterns of an injection-moulded polyethylene rod (cylinder axis parallel to the rotation axis z), recorded with an X-ray microbeam impinging (1) peripherally and (2) centrally. Reused with permission from Schroer *et al.* (2006), Copyright 2006, American Institute of Physics.

small-angle scattering (Schroer *et al.*, 2006). This very recent technique opens new possibilities to investigate non-destructively the local nanostructure inside a specimen and is planned to be implemented at the micro-SAXS beamline at the future storage ring PETRA III at DESY in Hamburg.

The setup for small-angle scattering tomography is shown in Figs. 8.15(a,b). The sample is scanned through a monochromatic microbeam in translation and rotation as described in Sections 8.1 and 8.2.2. The transmitted signal I_1 is recorded by a PIN diode that serves as the beam stop for a diffraction camera that records the two-dimensional small-angle scattering pattern $I_{(q_r,q_z)}(r,\phi)$ from the sample, which in general depends on the sample position r, orientation ϕ, and the momentum transfer (q_r, q_z). The guard slit removes scattered radiation from all optical components upstream and the incident beam is monitored by an ionization chamber I_0 behind the guard slit.

As an example, a SAXS tomogram is shown of a polyethylene rod that was made by injection moulding and that was subsequently stretched at elevated temperatures. The details of this experiment can be found in Schroer *et al.* (2006) and Stribeck *et al.* (2006). The rod was scanned with a microbeam generated with parabolic refractive X-ray lenses made of beryllium at beamline BW4 at the synchrotron radiation source DORIS III at DESY in Hamburg. 101 projections with 69 translational steps each were recorded, acquiring at each position (r,ϕ) of the scan the incident flux I_0, the transmitted flux I_1, and the SAXS pattern $I_{(q_r,q_z)}$. Figure 8.15(c) shows two SAXS patterns of the polyethylene

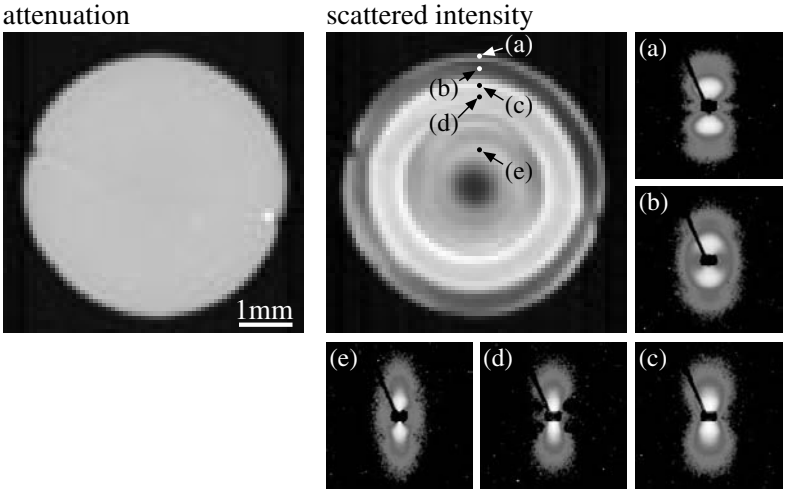


FIG. 8.16. Reconstructed attenuation coefficient μ and the integral small-angle scattering cross-section $p_{(0,q_z)}(x,y)$ along the q_z-direction parallel to the tomographic rotation axis. The cross-section was reconstructed on a grid of 80×80 pixels, given by the number of scanning positions in the projections recorded in this scanning technique. Reused with permission from Schroer *et al.* (2006), Copyright 2006, American Institute of Physics.

rod, recorded with the X-ray beam impinging (1) peripherally and (2) centrally.

The transmission signal I_1 measured by the PIN diode can be readily used to determine the attenuation coefficient $\mu(x,y)$. It can be easily reconstructed from the transmission tomographic data

$$I_1 = I_0 \exp\left(-\int_{\text{sample}} \mu[x(s',r), y(s',r)]\right) ds' \tag{8.12}$$

by filtered backprojection. Figure 8.16(left) shows the reconstructed attenuation coefficient of the polyethylene rod. It appears homogeneous, with a homogeneous mass density of $\rho = 0.88 \pm 0.04\,\text{g/cm}^3$. From attenuation contrast alone it would be concluded that the sample is homogeneous.

The reconstruction of the SAXS data is more involved. Assuming the first Born approximation, the SAXS pattern at a given position (r,ϕ) within the tomographic scan is given by:

$$I_{(q_r,q_z)}(r,\phi) = I_0 \int_{\text{sample}} f(\phi,s,r) p_{q_r,q_z,\phi}(x,y) g(\phi,s,r) ds, \tag{8.13}$$

where $f(\phi,s,r) = \exp\left(-\int_{-\infty}^{s} \mu(x(s',r), y(s',r)) ds'\right)$ is the attenuation of the incident beam as it propagates through the sample and

$$g(\phi, s, r) = \exp\left(-\int_{-\infty}^{s} ds'\ \mu[x(s', r), y(s', r)]\right)$$

is the attenuation of the scattered beam in the forward direction. Here, it was assumed that the scattering angles are small. f and g can be combined and taken out of the integral in eqn 8.13, which using eqn 8.12 can be simplified:

$$I_{(q_r,q_z)}(r, \phi) = I_1 \int_{\text{sample}} p_{q_r,q_z,\phi}(x, y)\ ds. \tag{8.14}$$

In general, $p_{q_r,q_z,\phi}$ depends on the rotation angle ϕ, such that the acquired tomographic projections do not provide sufficient information to solve this tomographic problem. If we restrict ourselves to the diffraction pattern along the rotation axis ($q_r = 0$), then $p_{0,q_z,\phi}$ is independent of ϕ and standard tomographic methods can be used to reconstruct the scattering cross-section for this restricted set of data. The integral of the cross-section along this line ($q_r = 0$) is shown in Fig. 8.16. As opposed to the attenuation coefficient, the reconstructed SAXS cross-sections vary strongly over the virtual section through the sample, clearly showing a core-shell structure of the rod.

If, in addition, the SAXS cross-section is rotationally symmetric with respect to the rotation axis, then the full cross-section p_{q_r,q_z} is independent of ϕ and can be reconstructed. In the given example scattering from the polyethylene rod has approximately this symmmetry, showing a strong fibre texture (cf. Fig. 8.15(c)) along the rotation axis. This allows one to reconstruct the full SAXS cross-sections at each point on the virtual slice through the sample. Figure 8.16 shows the SAXS cross-sections reconstructed at five different locations inside the sample labeled (a) through (e) in Fig. 8.16. The clear diffraction peaks at the periphery of the rod, Figs. 8.16(a,b), indicate an almost lattice-like arrangement of alternating crystalline and amorphous polyethylene lamellae. The bright shell in Fig. 8.16(right) is dominated by more or less diffuse scattering of uncorrelated domains of stacked crystalline and amorphous lamellae of varying thickness, see Figs. 8.16(c,d). In the centre, discrete scattering is observed again, see Fig. 8.16(e). A detailed analysis of the reconstructed diffraction patterns can be made using the multidimensional chord distribution function method (Stribeck *et al.*, 2006).

8.3 Summary

X-ray fluorescence, absorption at the K-edge and small-angle scattering provide signals carrying a different kind of information as compared with simple X-ray attenuation. By using intense focused and monochromatic synchrotron X-ray beams these signals can be used to measure three-dimensional tomographic images in analogy to the more traditional absorption tomography.

8.4 References

Ade, H., Zhang, X., Cameron, S., Costello, C., Kirz, J., and Williams, S. (1992). Chemical contrast in X-ray microscopy and spatially resolved XANES spectroscopy of organic specimens. *Science*, **258**, 972.

Andersen, A.H. and Kak, A.C. (1984). Simultaneous algebraic reconstruction technique (SART): a superior implementation of the ART algorithm. *Ultrasonic Imaging*, **6**, 81.

Boisseau, P. (1986). *Determination of three-dimensional trace element distributions by the use of monochromatic X-ray microbeams.* Ph. D. thesis, Massachussetts Institute of Technology (MIT), Cambridge.

Boisseau, P. and Grodzins, L. (1987). Fluorescence tomography using synchrotron radiation at the NSLS. *Hyperfine Interaction*, **33**, 283.

Brunetti, A. and Golosio, B. (2001). Software for X-ray fluorescence and scattering tomographic reconstruction. *Computer Physics Communications*, **141**, 412.

Camerani, M.C., Somogyi, A., Simionovici, A.S., Ansell, S., Steenari, B.-M., and Lindqvist, O. (2002). Direct determination of cadmium speciation in municipal solid waste fly ashes by synchrotron radiation induced X-ray fluorescence and X-ray absorption spectroscopy. *Environmental Science and Technology*, **36**, 3165.

Cesareo, R. and Mascarenhas, S. (1989). A new tomographic device based on the detection of fluorescent X-rays. *Nuclear Instruments and Methods in Physics Research A*, **277**, 669.

Chukalina, M., Simionovici, A., Snigirev, A., and Jeffries, T. (2002). Quantitative characterization of microsamples by X-ray fluorescence tomography. *X-ray Spectrometry*, **31**, 448.

Crawford, C.R. and King, K.F. (1990). Computed tomography scanning with simultaneous patient translation. *Medical Physics*, **17**, 9672.

European Synchrotron Radiation Facility (2006). ID22: Micro-fluorescence, imaging and diffraction. `http://www.esrf.fr/UsersAndScience/Experiments/Imaging/ID22/`. [Online; accessed 12 February 2007].

Gillet, Ph., Barrat, J.A., Heulin, Th., Achouak, W., Lesourd, M., Guyot, F., and Benzerara, K. (2000). Bacteria in the Tatahouine meteorite: nanometric-scale life in rocks. *Earth and Planetary Science Letters*, **175**, 161.

Golosio, B. (2003). *Structural and elemental microanalysis using X-ray tomographic techniques.* Ph. D. thesis, University of Cagliari.

Golosio, B., Brunetti, A., and Cesareo, R. (2004). Algorithmic techniques for quantitative Compton tomography. *Nuclear Instruments and Methods in Physics Research B*, **213**, 108.

Golosio, B., Simionovici, A., Somogyi, A., Camerani, C., and Steenari, B.M. (2003*a*). X-ray fluorescence tomography of individual waste fly ash particles. *Journal de Physique IV*, **104**, 647.

Golosio, B., Simionovici, A., Somogyi, A., Lemelle, L., Chukalina, M., and Brunetti, A. (2003*b*). Internal elemental microanalysis combining X-ray

fluorescence, Compton and transmission tomography. *Journal of Applied Physics*, **94**, 145.

Golosio, B., Somogyi, A., Simionovici, A., Bleuet, P., and Lemelle, L. (2004). Nondestructive quantitative 3D elemental microanalysis by combined helical X-ray microtomographies. *Applied Physics Letters*, **84**, 2199.

Grant, J.A., Morgan, M.J., Davis, J.R., and Wells, P. (1995). Reconstruction strategy suited to X-ray diffraction tomography. *Journal of the Optical Society of America A*, **12**, 291.

Grunwaldt, J.-D., Hannemann, S., Schroer, C.G., and Baiker, A. (2006). 2D-mapping of the catalyst structure inside a catalytic microreaktor at work: Partial oxidation of methane over Rh/Al_2O_3. *Journal of Physical Chemistry B*, **110**, 8674.

Grunwaldt, J.-D., Lützenkirchen-Hecht, D., Richwin, M., Grundmann, S., Clausen, B.S., and Frahm, R. (2001). Piezo X-ray absorption spectroscopy for the investigation of solid-state transformations in the millisecond range. *Journal of Physical Chemistry B*, **105**, 5161.

Hamaker, C. and Solomon, D.G. (1978). The angles between the null space of X-rays. *Journal of Mathematical Analysis and Applications*, **62**, 1.

Herman, G.T., Lent, A., and Lutz, P.H. (1978). Relaxation methods for image reconstruction. *Communications of the Association for Computing Machinery (ACM)*, **21**, 152.

Hignette, O., Cloetens, P., Lee, W.K., Ludwig, W., and Rostaing, G. (2003). Hard X-ray microscopy with reflecting mirrors: status and perspectives of the ESRF technology. *Journal de Physique IV*, **104**, 231.

Hignette, O., Cloetens, P., Rostaing, G., Bernard, P., and Morawe, C. (2005). Efficient sub 100 nm focusing of hard X-rays. *Review of Scientific Instruments*, **76**, 063709.

Hogan, J.P., Gonsalves, R.A., and Krieger, A.S. (1991). Fluorescent computer tomography: A model for correction of X-ray absorption. *IEEE Transactions on Nuclear Science*, **38**, 1721.

Iida, A., Takahashi, M., Sakurai, K., and Gohshi, Y. (1989). Synchrotron radiation X-ray fluorescence imaging by image reconstruction technique. *Review of Scientific Instruments*, **60**, 2458.

Kak, A.C and Slaney, M. (1987). *Principles of computerized tomography imaging.* IEEE Press, New York.

Kalender, W.A., Seissler, W., and Vock, P. (1989). Single-breath-hold volumetric CT by continuous patient translation and scanner rotation. *Radiology*, **173**, 414.

Kang, H.C., Maser, J., Stephenson, G.B., Liu, C., Conley, R., Macrander, A.T., and Vogt, S. (2006). Nanometer linear focusing of hard X-rays by a multilayer Laue lens. *Physical Review Letters*, **96**, 127401.

Kinney, J., Johnson, Q., Nichols, M., Bonse, U., and Nußhardt, R. (1986). Elemental and chemical-state imaging using synchrotron radiation. *Applied Optics*, **25**, 4583.

Kirkpatrick, P. and Baez, A. (1948). Formation of optical images by X-rays. *Journal of the Optical Society of America*, **38**, 766.

Koningsberger, D. C. and Prins, R. (1988). *X-ray absorption.* John Wiley & Sons, New York.

La Riviere, P.J. and Billmire, D.M. (2004). Penalized-likelihood image reconstruction for X-ray fluorescence computed tomography with unknown fluorescence attenuation maps. In *Developments in X-ray tomography IV* (ed. U. Bonse), pp. 243. The International Society for Optical Engineering (SPIE). Proceedings of the SPIE, Vol. 5535.

McKay, D.S., Gibson Jr., E.K., Thomas-Keprta, K.L., Vali, H., Romanek, C.S., Clemett, S.J., Chillier, X.D.F., Maechling, C.R., and Zare, R.N. (1996). Search for past life on Mars: Possible relic biogenic activity in Martian meteorite ALH84001. *Science*, **273**, 924.

McNear, D.H., Peltier, E., Everhart, J., Chaney, R.L., S.S., Newville, M., Rivers, M., and Sparks, D.L. (2005). Application of quantitative fluorescence and absorption-edge computed microtomography to image metal compartmentalization in *Alyssum murale*. *Environmental Science and Technology*, **39**, 2210–2218.

Menez, B., Simionovici, A., Philippot, P., Bohic, S., Gibert, F., and Chukalina, M. (2001). X-ray fluorescence micro-tomography of an individual fluid inclusion using a third generation synchrotron light source. *Nuclear Instruments and Methods in Physics Research B*, **181**, 749.

Mimura, H., Matsuyama, S., Yumoto, H., Yamamura, K., Sano, Y., Shibahara, M., Endo, K., Mori, Y., Nishino, Y., Tamasaku, K., Yabashi, M., Ishikawa, T., and Yamauchi, K. (2005). Hard X-ray diffraction-limited nanofocusing with Kirkpatrick-Baez mirrors. *Japanese Journal of Applied Physics*, **44**, L539.

Pinzani, M.C.C., Golosio, B., Somogyi, A., Simionovici, A., Steenari, B.M., and Panas, I. (2004). X-ray fluorescence tomography of individual MSW and biomass fly ash particles. *Analytical Chemistry*, **76**, 1586.

Rau, C., Somogyi, A., Bytchkov, A., and Simionovici, A. (2002). XANES microimaging and tomography. In *Developments in X-ray tomography III* (ed. U. Bonse), pp. 249. The International Society for Optical Engineering (SPIE). Proceedings of the SPIE, Vol. 4503.

Richwin, M., Zaeper, R., Lützenkirchen-Hecht, D., and Frahm, R. (2001). Piezo-QEXAFS: advances in time-resolved X-ray absorption spectroscopy. *Journal of Synchrotron Radiation*, **8**, 354.

Richwin, M., Zaeper, R., Lützenkirchen-Hecht, D., and Frahm, R. (2002). Piezo-XAFS-time-resolved X-ray absorption spectroscopy. *Review of Scientific Instruments*, **73**, 1668.

Rust, G.F. and Weigelt, J. (1998). X-ray fluorescent computer tomography with synchrotron radiation. *IEEE Transactions on Nuclear Science*, **45**, 75.

Sautter, V., Barrat, J.A., Jambon, A., Lorand, J.P., Gillet, Ph., Javoy, M., Joron, J.L., and Lesourd, M. (2002). A new Martian meteorite from Mo-

rocco: the nakhlite North West Africa 817. *Earth and Planetary Science Letters*, **195**, 223.

Schätzler, H.P. (1978). Basic aspect on the use of elastic and inelastic scattered gamma radiation for the determination of binary systems with effective atomic numbers of less than 10. *International Journal of Applied Radiation and Isotopes*, **30**, 115.

Schroer, C. (2001). Reconstructing X-ray fluorescence microtomograms. *Applied Physics Letters*, **79**, 1912.

Schroer, C.G., Kuhlmann, M., Günzler, T.F., Lengeler, B., Richwin, M., Griesebock, B., Lützenkirchen-Hecht, D., Frahm, R., Ziegler, E., Mashayekhi, A., Haeffner, D., Grunwaldt, J.-.D, and Baiker, A. (2003). Mapping the chemical states of an element inside a sample using tomographic X-ray absorption spectroscopy. *Applied Physics Letters*, **82**, 3360.

Schroer, C.G., Kuhlmann, M., Roth, S. V., Gehrke, R., Stribeck, N., Almendarez-Camarillo, A., and Lengeler, B. (2006). Mapping the local nanostructure inside a specimen by tomographic small angle X-ray scattering. *Applied Physics Letters*, **88**, 164102.

Schroer, C.G., Kurapova, O., Patommel, J., Boye, P., Feldkamp, J., Lengeler, B., Burghammer, M., Riekel, C., Vincze, L., van der Hart, A., and Küchler, M. (2005). Hard X-ray nanoprobe based on refractive X-ray lenses. *Applied Physics Letters*, **87**, 124103.

Simionovici, A., Chukalina, M., Drakopoulos, M., Snigireva, I., Snigirev, A., Schroer, C., Lengeler, B., Janssens, K., and Adams, F. (2000). 2D imaging by X-ray fluorescence microtomography. In *X-ray microscopy* (ed. W. Meyer-Ilse, T. Warwick, and D. Atwood), Melville, pp. 539. American Institute of Physics. AIP Conference Proceedings, Vol. 507.

Simionovici, A., Chukalina, M., Drakopoulos, M., Snigireva, I., Snigirev, A., Schroer, C., Lengeler, B., Janssens, K., and Adams, F. (2002). X-ray fluorescence microtomography: experiment and reconstruction. In *Developments in X-ray tomography II* (ed. U. Bonse), pp. 304. The International Society for Optical Engineering (SPIE). Proceedings of the SPIE, Vol. 3772.

Simionovici, A., Chukalina, M., Günzler, F., Schroer, C., Snigirev, A., Snigireva, I., Tümmler, J., and Weitkamp, T. (2001). X-ray microtome by fluorescence tomography. *Nuclear Instruments and Methods in Physics Research A*, **467-468**, 889.

Simionovici, A., Chukalina, M., Schroer, C., Drakopoulos, M., Snigirev, A., Snigireva, I., Lengeler, B., Janssens, K., and Adams, F. (2000). High resolution X-ray fluorescence tomography on homogenous samples. *IEEE Transactions on Nuclear Science*, **47**, 2736.

Simionovici, A., Chukalina, M., Vekemans, B., Lemelle, L., Gillet, Ph., Schroer, C., Lengeler, B., Schröder, W., and Jeffries, T. (2002). New results in X-ray fluorescence tomography. In *Developments in X-ray tomography III* (ed. U. Bonse), pp. 222. The International Society for Optical Engineering (SPIE). Proceedings of the SPIE, Vol. 4503.

Somogyi, A., Tucoulou, R., Martinez-Criado, G., Homs, A., Cauzid, J., Bleuet, P., and Simionovici, A. (2005). ID22: a multitechnique hard X-ray microprobe beamline at the european synchrotron radiation facility. *Journal of Synchrotron Radiation*, **12**, 208.

Stribeck, N., Almendarez-Camarillo, A., Nöchel, U., Schroer, C.G., Kuhlmann, M., Roth, S.V., Gehrke, R., and Bayer, R.K. (2006). Volume-resolved nanostructure survey of a polymer part by means of SAXS microtomography. *Macromolecular Chemistry and Physics*, **207**, 1139.

Takeda, T., Ito, K., Kishi, K., Maeda, T., Yuasa, T., Wu, J., Kazama, M., Hyodo, K., Akatsuka, T., and Itai, Y. (1994). Preliminary experiment of the fluorescent X-ray computed tomography with synchrotron radiation. *Medical Imaging Technology*, **12**, 537.

Tanabe, K. (1971). Projection method for solving a singular system of linear equations and its applications. *Numerische Mathematik*, **17**, 203.

Terzano, R., Spagnuolo, M., Medici, L., Vekemans, B., Vincze, L., Janssens, K., and Ruggiero, P. (2005). Copper stabilization by zeolite sythesis in polluted soils treated with coal fly ash. *Environmental Science and Technology*, **39**, 6280.

Vekemans, B., Vincze, L., Brenker, F. E., and Adams, F. (2004). Processing of three-dimensional microscopic X-ray fluorescence data. *Journal of Analytical Atomic Spectrometry*, **19**, 1302.

Vincze, L., Vekemans, B., Brenker, F. E., Falkenberg, G., Rickers, K., Somogyi, A., Kersten, M., and Adams, F. (2004). Three-dimensional trace element analysis by confocal X-ray microfluorescence imaging. *Analytical Chemistry*, **76**, 6786.

Wang, G. and Vannier, M.W. (1994). Longitudinal resolution in volumetric X-ray CT – Analytical comparison between conventional and helical CT. *Medical Physics*, **21**, 429.

Youn, H.S., Lee, D.-H., Kim, K., Choi, H.J., and Koo, Y.M. (2001). μ-XAFS beam line at Pohang Light Source. *Nuclear Instruments and Methods A*, **467-468**, 1557.

Yuasa, T., Akiba, M., Takeda, T., Kazama, M., Hoshino, A., Watanabe, Y., Hyodo, K., Dilmanian, F.A., Akatsuka, T., and Itai, Y. (1997). Reconstruction method for fluorescent X-ray computed tomography by least-squares method using singular value decomposition. *IEEE Transactions on Nuclear Science*, **44**, 54.

Yun, W., Pratt, S. T., Miller, R. M., Cai, Z., Hunter, D. B., Jarstfer, A. G., Kemner, K. M., Lai, B., Lee, H.-R., Legnini, D. G., Rodrigues, W., and Smith, C. I. (1998). X-ray imaging and microspectroscopy of plants and fungi. *Journal of Synchrotron Radiation*, **5**, 1390.

9

THREE-DIMENSIONAL X-RAY DIFFRACTION

Henning Friis Poulsen

In nature, most materials such as rocks, ice, sand and soil, appear as aggregates composed of a set of crystalline elements. Similarly, many man-made materials are polycrystalline, e.g. metals, alloys and ceramics. So are drugs, bones and trace particles relevant to environmental matters.

Remarkably, until recently no non-destructive methods existed for providing comprehensive three-dimensional information on the structure and dynamics of polycrystals on the scale of the individual microstructural elements (the grains, subgrains, particles or domains). X-ray and neutron diffraction have been confined to two limiting cases: powder diffraction, which averages over the microstructural elements, and diffraction on single crystals. Most real-world materials occur as heterogeneous aggregates with substantial internal structure, and thus fall between these two extremes. Absorption contrast tomography is not sensitive to crystalline lattice orientation, and as such cannot visualize the elements of the crystal structure within a monophase material. Local information has been provided by tools such as optical, electron, ion beam and scanning probe microscopy. However, these methods probe the near-surface regions only. Hence, the characterization is only two-dimensional and prohibits studies of the bulk dynamics.

Three-dimensional X-ray diffraction (3DXRD) is a novel technique, aiming at a fast and non-destructive characterization of the individual elements within millimetre- to centimetre-sized specimens. It is based on two principles: the use of highly penetrating hard X-rays from a synchrotron source (X-ray energies above 30 keV) and the application of 'tomographic' reconstruction algorithms for the analysis of the diffraction data. In favourable cases, the position, morphology, phase, and crystallographic orientation can be derived for hundreds of elements simultaneously and their elastic strains can be determined. Furthermore, the dynamics of the individual elements can be monitored during typical processes such as deformation or annealing. Hence, for the first time information on the interaction between microstructural elements can be obtained directly. The provision of such data is vital in order to extend beyond state-of-the-art structural models.

In this chapter the 3DXRD methodology will be outlined with the focus on characterizing grains and orientations. Notably, in a 3DXRD experiment one must prioritize between spatial, angular and time resolution. This has led to a variety of 3DXRD strategies, which in the following will be summarized as

four *standard modes of operation*. The first two modes enable fast measurements of the average characteristics of each grain (such as their centre-of-mass position, volume and/or average orientation), while the exact location of the grain boundaries is unknown. The latter two enable full 3D mappings of grains and orientations. For general reference and more information on e.g. strain determination, see Poulsen (2004).

In the following, the 3DXRD setup will be presented and 3DXRD strategies discussed. Then, the standard modes of operation are presented in more detail with emphasis on reconstruction principles and providing selected examples of use. Next, a novel method called *diffraction contrast tomography* is proposed – this is related to 3DXRD in the same way as bright-field and dark-field images are in transmission electron microscopy. At the end, the mathematics underlying the 3DXRD diffraction geometry and the representation of the crystallographic orientation are summarized in a self-contained section.

Alternative approaches to provision of 3D maps based on X-ray diffraction exist. These are based on inserting wires (Larson *et al.*, 2002), slits (Bunge *et al.*, 2003) or collimators (Wroblewski *et al.*, 1999) between the sample and the detector and scanning the sample with respect to these elements. Not surprisingly, such methods will be slower than the tomographic approach of 3DXRD, but they may be associated with other advantages, such as improved options for measuring the local elastic strain. In particular, we mention the technique of *differential-aperture X-ray microscopy* (Larson *et al.*, 2002). Using a polychromatic microbeam with energies of 8 to 20 keV and scanning a wire, they have demonstrated submicrometre resolution in 3D. For a comparison of methods, see e.g. Larson and Lengeler (2004).

9.1 Basic setup and strategy

The basic 3DXRD setup as sketched in Fig. 9.1 is quite similar to conventional tomography settings at synchrotrons. A (nearly) parallel and uniform monochromatic X-ray beam impinges on the sample. The sample is mounted on an ω-rotation stage, where ω is the rotation around an axis perpendicular to the incoming beam. As an option, x-, y-, and z-translations may be added as well as additional rotations.

Any part of the illuminated structure, which fulfils the Bragg condition eqn 4.6, will generate a diffracted beam. This beam is transmitted through the sample and is probed by a 2D detector. To probe the complete structure, and not just the part that happens to fulfil the Bragg condition, the sample is rotated. Hence, exposures are made for equiangular settings of ω with a step of $\Delta\omega$. The sample is continuosly rotated during each exposure by $\Delta\omega$. Essential to 3DXRD is the idea to mimic a 3D detector by positioning several 2D detectors at different distances L to the centre-of-rotation and exposing these either simultaneously (the detectors are semi-transparent to hard X-rays) or sequentially.

Presently, two types of detectors are used: *near-field detectors* (fluorescence screens combined with magnifying optics and a CCD camera) with a spatial

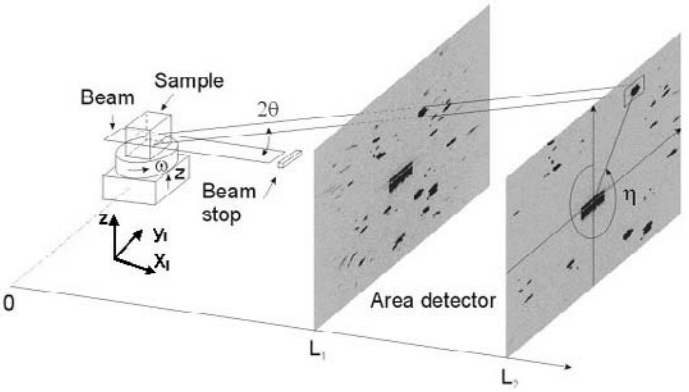


FIG. 9.1. Sketch of the 3DXRD principle for the case of the incoming monochromatic beam illuminating one horizontal layer in the sample. Two types of detectors are used: high spatial resolution detectors close to the sample and low-resolution detectors far away from the sample (not shown). The Bragg angle, 2θ, the rotation angle, ω, and the azimuthal angle, η, are indicated for the diffracted beam arising from one grain of a coarse-grained specimen, and for two settings of the high-resolution detector. The axes for the laboratory co-ordinate system are also shown. From Poulsen (2004) (with permission from Springer).

resolution of 3 to 10 µm in close proximity to the sample and *far-field detectors* (such as a flourescence screen connected by demagnifying optics to a CCD, e.g. a FReLoN camera, see Chapter 10) with a resolution of 100 to 200 µm. Typical L values are 2 to 10 mm and 20 to 50 cm, respectively. The former detector type provides information on position and orientation degrees of freedom, while the latter probes strain and orientation, as we will see later. With the near-field detector, images may be acquired at several – typically 3 – distances as illustrated in Fig. 9.1. This enables ray tracing of the diffracted beam, also known as *tracking*.

The incoming beam may illuminate the full sample or be focused in one direction to probe only a layer within the material (see Fig. 9.1). In the following, we shall refer to these as the 3D and 2D case, respectively. Notably, the 2D case is restricted to use at third-generation synchrotron sources due to the need for focusing the beam, while the 3D case also applies to second-generation sources and in favourable cases – large grains and a low requirement for spatial resolution – even to work with laboratory sources. In the 2D case, 3D information is generated simply by repeating data acquisition for a set of layers and stacking the resulting reconstructions. The data analysis for this case is presently more mature, but ultimately the 3D case promises a better time resolution. An example of a set of exposures obtained in the 2D case is provided in Fig. 9.2. Note that in the near-field images the spots are extended (comprising information on

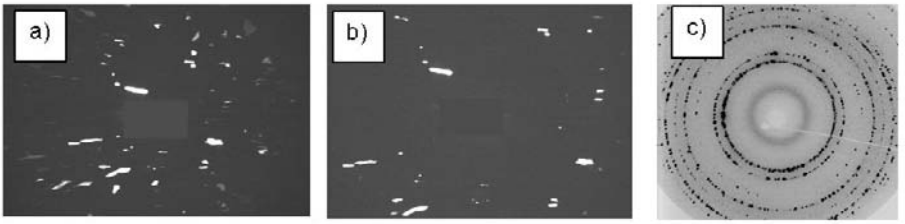


FIG. 9.2. Typical 3DXRD data from a coarse-grained undeformed polycrystal. (a) and (b) are 'near-field' exposures with the high spatial resolution detector (see Fig. 9.1) at distances of $L = 4\,\text{mm}$ and $L = 8\,\text{mm}$, respectively. (c) is a corresponding 'far-field' image made with the low-resolution detector at $L = 400\,\text{mm}$ and shown with inverted greyscales to improve visibility.

grain shape) and form no clear pattern on the detector as their position is determined by Bragg's law as well as by the positions of the associated grains within the sample. In contrast, in the far-field image the spot sizes are instrumentally limited. The position of grains within the sample now makes a negligible contribution to the positions of the spots. Hence, the spots are positioned on a set of Debye–Scherrer rings, similar to the case of conventional powder diffraction.

A vital difference between the two types of detectors relates to data-acquisition time: with the current technology the far-field detectors are much more efficient then the near-field ones and hence provide a time resolution that is several orders of magnitude better. This makes it necessary to carefully consider the priority of time and space resolution in a 3DXRD experiment. In practice, we may distinguish between four *modes of operation*, as sketched in Fig. 9.3. Modes I and II aim at fast data acquisition with limited spatial information. In Mode II for each grain in the illuminated part of the specimen, one can determine its centre-of-mass position, volume, average orientation and average strain tensor. Mode I is an option for very fast measurements, where only a subset of all grains is characterized and no spatial information is available. Also, the orientation and strain characterization will not be complete. Experimentally, the main difference between the modes I and II is the ω-range covered. Modes III and IV aim at the generation of complete 3D maps. In the case of an undeformed material the orientation is constant within each grain and the aim is to provide a *grain map*. In the case of a deformed material the orientation varies locally. This makes it relevant to measure *orientation maps*, where each volume pixel (voxel) in the sample is associated with its own orientation. In both cases, evidently a detector with a high spatial resolution is needed.

In many experiments, combinations of these modes are relevant. As an example, one may wish to start by mapping extended parts of a sample in order to identify subvolumes of particular interest. Then, one focuses on such parts and performs a fast centre-of-mass study on these parts during *in-situ* processing. This procedure is then complemented by mapping extended parts again after

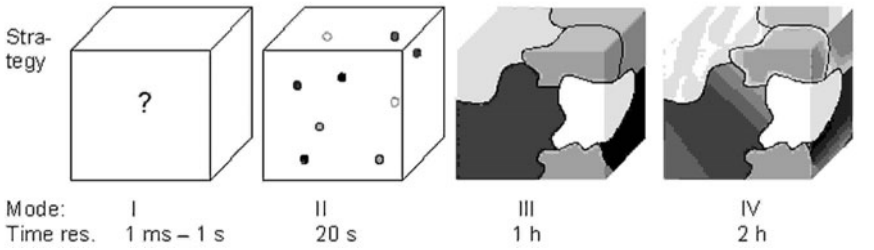


FIG. 9.3. Sketch of four typical 3DXRD modes of operation, listed from right to left in the direction of decreasing spatial information: Mode IV: 3D orientation map of a deformed sample, Mode III: 3D grain map for undeformed specimens, Mode II: only centre-of-mass position, volume and average orientation for each grain, Mode I: not all grains within the illuminated volume are probed. Furthermore, the information about orientation and position of those probed is not complete. In all cases, orientations are represented by a greyscale. Typical data acquisition times for characterization of 100 grains are indicated.

the processing is completed. Also, centre-of-mass-type information from mode II is often needed to initialize reconstruction algorithms for grain and orientation mapping in modes III and IV, as explained in more detail below.

9.2 Indexing and characterization of average properties of each grain

9.2.1 *Polycrystal indexing*

A very important simplification arises in the case where the diffraction pattern is composed of a set of (primarily) non-overlapping diffraction spots. Three examples of such data sets are shown in Fig. 9.2. In this case, the diffraction spots can be sorted with respect to their element of origin, say the grain, by means of a polycrystalline indexing scheme (Lauridsen *et al.*, 2001). Based on this, a comprehensive structural analysis of the average properties of each grain is readily available (Poulsen *et al.*, 2001).

The starting point for such schemes is to apply image analysis routines that identify the spots and determine their centre-of-mass positions in $(y_{\mathrm{det}}, z_{\mathrm{det}}, \omega)$-space, where y_{det} and z_{det} are the detector co-ordinates. Next, the directions of the corresponding diffracted X-rays are determined. In the case of near-field data – that is the tracking setup illustrated in Fig. 9.1 – ray tracing is used. Having identified a set of spots at different detector-to-sample distances that belong to the same reflection, the best fit to a straight line through the centres-of-mass of these spots is determined. In the case of far-field data one may simply assume that all grains are positioned at the centre-of-rotation and for any spot draw the line from its centre on the detector to the rotation axis. In both cases, once

the direction of the ray is known, the corresponding scattering vector is readily determined (see Section 9.7).

The sorting can be based on three principles:

- *Orientation.* As explained in detail in Section 9.7, for each scattering vector, we can determine a line in 3D orientation space: the projection line. By definition, the orientation of the associated grain is on this line. Furthermore, the projection lines for all the scattering vectors belonging to a given grain will intersect in one point. Hence, determining orientations becomes a question of finding intersections. This is an exercise in discrete tomography. Combinatorial methods (where one tests whether a set of spots forms a pattern consistent with the crystallographic symmetry) may be used, but only when the number of grains is small. Instead, the intersections can be found by scanning orientation space and for each orientation compare a simulated diffraction pattern with the one observed experimentally (Lauridsen *et al.*, 2001).
- *Grain volume.* The integrated intensity of any diffraction spot is proportional to the volume of the associated grain (in the 2D case: to the area of the illuminated section of the grain). Having normalized intensities appropriately, as discussed by Poulsen (2004) and Lauridsen *et al.* (2000), all diffraction spots associated with a given grain should exhibit identical integrated intensities within the limits given by experimental noise.
- *Position.* If ray tracing is used, the diffraction spots can be sorted according to position as well. To do so, the fitted X-ray line is extrapolated to the sample. In this way, in the 2D case, the (x, y) centre-of-mass of the associated illuminated grain section is determined. In the 3D case, the sorting criterion is that the extrapolated lines all intersect in one point, namely the (x, y, z) centre-of-mass of the associated grain.

Sorting based on orientation – and to a lesser extent position – has been implemented in the program `Graindex` (Lauridsen *et al.*, 2001) which assumes a single-phase sample with a known crystallographic space group.

Based on the output of `Graindex` a comprehensive characterization of each grain immediately follows. If only the far-field detector is used the grain volume, average orientation and the average values of the components of the elastic strain tensor (Martins *et al.*, 2004) can be determined. The time resolution is often $\approx 10\,\mathrm{s}$, limited by the readout time of the detector. The accuracy is of order 0.3° for orientations and $\Delta\varepsilon/\varepsilon = 1 \times 10^{-4}$ for strain. Furthermore, it is possible to perform a full crystallographic analysis of each grain, e.g. a structural refinement (Schmidt *et al.*, 2003, Vaughan *et al.*, 2004).

In a variant of mode II the centre-of-mass position of each grain is determined in addition at the expense of time resolution. This can be done by tracking using a detector with a high spatial resolution as outlined above. To determine positions with high accuracy, a non-linear least square fit to the centre-of-mass positions of the spots on the detector is performed. The accuracy is currently $\approx 5\,\mu\mathrm{m}$.

Alternatively, for large grains with a size comparable to the resolution of a far-field detector, the same type of fit can also be applied to spot positions on such a detector. This alternative allows for faster data acquisition than tracking but the resulting centre-of-mass positions of the grains are less accurate.

Notably, the combined knowledge of the centre-of-mass position and volume of the grains can be used for a primitive tessellation that can be of sufficient quality to identify which grains are likely to be neighbours.

A large number of problems within polycrystal and powder research can be tailored to apply to mode II analysis. The main limitation is spot overlap on the detector. The probability of such overlap is determined for each of the grains, the texture (the probability of grains exhibiting similar orientations), and the orientation spread of each grain. Simulations show that for samples with a weak texture and grains with a near-perfect lattice – that is with an orientation spread of order 0.1° or less – several thousand grains can be indexed simultaneously (Schmidt *et al.*, 2003). On the other hand, plastic deformation introduces orientation spread within the grains, which in practice prohibits indexing of embedded grains in materials that are deformed by more than 20%.

9.2.2 *A statistical description of dynamics*

For very fast data acquisitions mode I is relevant. In this case one simply repeats acquisitions with the far-field detector, oscillating around a given ω-setting. Assuming grains do not rotate, one can then monitor the change in volume of the grains giving rise to diffraction spots at this ω-setting. Several hundred so-called *growth curves* can be obtained simultaneously in this way (Lauridsen *et al.*, 2003). To ensure that the integrated intensities monitored are truly proportional to corresponding grain volumes it is required to validate the spots at regular intervals, e.g. by comparing intensities obtained at a nominal setting of beam size and rotation range, $\Delta\omega$, and a larger test setting of beam size and $\Delta\omega$. For more information on deriving growth curves see Lauridsen *et al.* (2000) and Lauridsen *et al.* (2003). Notably, one only probes a fraction of all the grains in the illuminated volume in this way, so the analysis is statistical in nature. Another limitation is that only components of the orientation and strain tensor are measured.

9.2.3 *Applications*

9.2.3.1 *Application I: nucleation and growth studies* Traditionally, nucleation and growth phenomena have been analysed using ensemble-average properties, such as the volume fraction of transformed material. However, the predictive power of average properties is limited by the neglect of heterogeneities. For example, nucleation may take place preferentially at specific sites, and the growth rate of nuclei may depend strongly on orientation, size, stoichiometry, or relationships with neighbouring volumes. 3DXRD is an ideal tool to study the effect of heterogeneities, and as such it has been used for a series of studies related to recrystallization (Lauridsen *et al.*, 2000, Lauridsen *et al.*, 2003), solidification (Iqbal

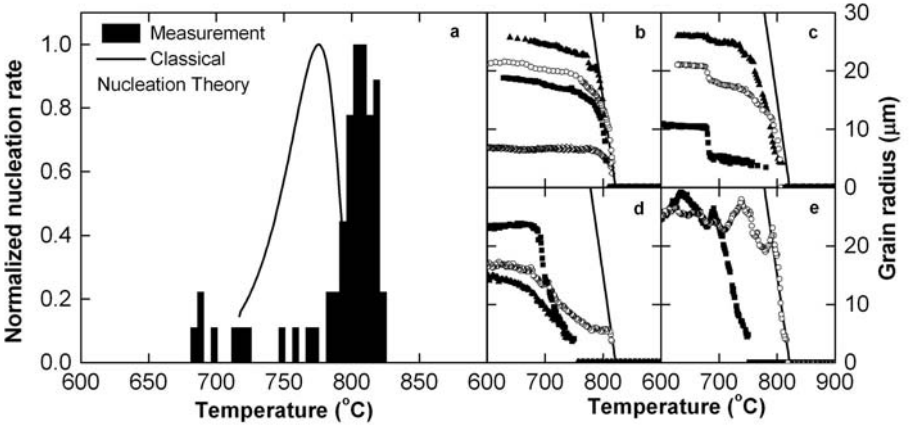


FIG. 9.4. (a) Nucleation of ferrite grains during cooling of a carbon steel. The normalized observed nucleation rate (bars) is compared to a theoretical prediction (line) based on a classical model (Lange *et al.*, 1988). (b)–(e) Examples of growth curves for individual ferrite grains, each being represented by one of the curves: (b) comprises ferrite grains that started to grow according to Zener theory, (c) ferrite grains that continued to grow with the same crystallographic orientation during pearlite formation at $\approx 685\,^{\circ}\mathrm{C}$ as part of a pearlite colony, (d) ferrite grains that exhibited retarded growth due to soft impingement and (e) ferrite grains that exhibit complex growth, perhaps due to a combination of hard and soft impingement. The lines are predictions based on the parabolic growth model of Zener. From Offerman *et al.* (2002) (with permission from Science).

et al., 2005), and phase transformations in steel, ceramics and ferroelectrics. In all cases it was demonstrated that the ensemble-average 'Avrami-type' models are at best gross simplifications.

As an example of a 3DXRD study in mode I, Offerman *et al.* (2002) have studied the phase transformation from the high-temperature austenite to the medium-temperature ferrite phase in carbon steel *in-situ*. The results shown in Fig. 9.4 relate to a study where the steel was cooled at a constant rate from $900\,^{\circ}\mathrm{C}$ to $600\,^{\circ}\mathrm{C}$ over 1 h. During this process nucleation and growth of ferrite grains was inferred from the appearance and evolution of about 60 distinct diffraction spots originating from the ferrite. By simply counting the number of new spots appearing in a given time interval the nucleation rate is determined as a function of temperature – a parameter that cannot be determined with high fidelity in any other way. In Fig. 9.4(a) the nucleation rate is compared to a prediction from a classical nucleation model based on two assumptions: that the driving force for nucleation – the decrease in Gibbs free energy – can be determined from thermodynamic reference data, and that the nucleus is shaped as

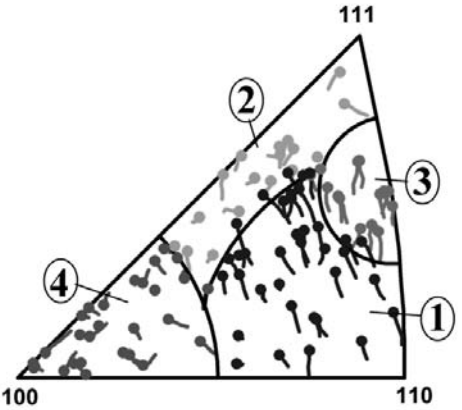


FIG. 9.5. Rotation of the tensile axis of 95 grains within an aluminium specimen, as expressed in a stereographic triangle. The curves are the observed paths for the average orientation of each grain during tensile deformation from 0 to 6%. The final orientation is marked by a circle. The rotation behaviour is interpreted in terms of four regions marked by solid lines. From Winther *et al.* (2004) (with permission from Elsevier).

a pillbox with coherent and semi-coherent interfaces (Lange *et al.*, 1988). From analysis of the data shown in the figure it is found that the predicted and measured activation energies differ by more than two orders of magnitude. Relevant revisions to the model are currently being discussed.

Shown in Figs. 9.4(b)–(e) are examples of resulting growth curves for the nuclei. It appears that the growth of some grains – those in (b) – followed closely the predictions of Zener, while others displayed large discrepancies. The latter was explained by several effects including local enrichment in the carbon content of the austenite.

9.2.3.2 *Application II: plastic deformation* Polycrystal deformation is a topic of prime interest to both metallurgists and geoscientists. However, despite 70 years of effort, there is no consensus on how to approach the modelling. In particular, it is unclear to what extent deformation behaviour is determined by the initial grain orientation, the grain–grain interactions or by the emerging dislocation structures. 3DXRD provides *in-situ* methods for addressing these issues, for example through real-time observations of grain rotations in bulk materials during the deformation process (Margulies *et al.*, 2001, Poulsen *et al.*, 2003). As an example of 3DXRD work in mode II, Fig. 9.5 provides a mapping of the rotation path of 95 grains in a 4-mm thick polycrystalline aluminium sample strained in tension. These data provide a detailed combinatorial database for critical evaluation of models for polycrystalline deformation. In major parts of the orientation triangle, the paths exhibit a clear dependence on initial orientation.

It is found that the grain responses do not match predictions from classical Taylor, Sachs or self-consistent models, but can be described by subdividing orientation space into four regions with distinctly different rotation behaviour (Winther *et al.*, 2004).

9.2.3.3 *Application III: studies of subgrains and nanocrystalline materials*
3DXRD studies are also relevant for crystalline structures with a grain size smaller than the spatial resolution. Evidently, in such cases the generation of a grain map is not possible, but by focusing the beam to say $10 \times 10\,\mu m^2$, diffraction spots from individual grains as small as 70 nm can be detected. By tracing the integrated intensities and positions of such spots as a function of time, one can infer changes in the volume, orientation and strain of the grains of origin. The main limitation for such studies is spot overlap. To overcome this problem two approaches have been pursued. The first approach is based on reducing the number of illuminated grains by investigating foils (Gundlach *et al.*, 2004). Provided the foil thickness is at least 10 times the grain size, the grains at the centre of the foil may be considered bulk grains for many annealing processes. This methodology has been applied to a series of mode I and II coarsening studies of subgrains in aluminium deformed to both medium and very high strains.

The second approach is to increase the angular resolution of the instrument by 2 orders of magnitude – to 0.004° in η and ω and 0.0005° in 2θ. Implemented at sector 1-ID at the Advanced Photon Source (APS) in Chicago, this improvement is achieved at the expense of flux by using a six-bounce monochromator – comprised of 6 perfect crystals – and a detector positioned at a distance of 4 m from the sample (Jakobsen *et al.*, 2006). This setup is designed specifically for studies of dislocation patterning. When deforming metals, typically the dislocations assemble into walls separating nearly perfect dislocation-free crystals – the subgrains – with dimensions of 100 nm to 2 μm. With the setup at APS individual spots from millions of illuminated subgrains in a 300-μm thick Cu sample can be resolved. Full 3D reciprocal space maps are gathered continuously while deforming the specimen. Uniquely, from their time dependence fundamental questions can be addressed in a direct way, such as 'how and when do the dislocation structures form?' and 'how do the structures subdivide as a function of increasing strain?' In the case of tensile deformation of Cu, it is found that the dislocation structures form just after the plastic onset and that the structures display intermittent dynamics – a surprising fact not predicted by theory (Jakobsen *et al.*, 2006).

9.3 Mapping of grains within undeformed specimens

Mapping grains in undeformed specimens can naturally be divided into two parts: identification of the number of grains and their orientations, and determination of the spatial position of grain boundaries, i.e. their morphologies. Provided spot overlap is not too severe an indexing program like `Graindex` can handle the first part.

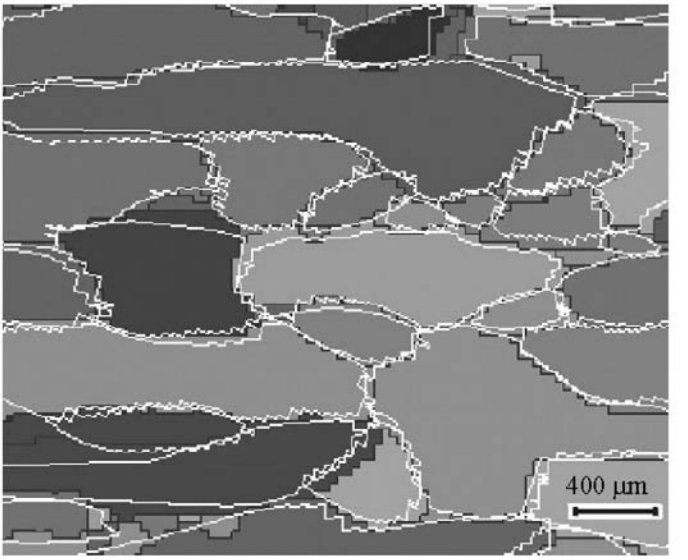


FIG. 9.6. Validation of the simple backprojection algorithm for the generation of grain maps. Greyscales and black outlines mark the grain orientation and grain boundaries on the surface of an aluminium polycrystal as determined by electron backscatter diffraction (EBSD). Superposed as white lines are the grain boundaries resulting from the synchrotron experiment. The average deviation between the boundary positions in the 3DXRD and the EBSD maps is 24 µm, comparable to the point-spread function of the detector available at that time. From Poulsen *et al.* (2001) (with permission from IUC).

Mapping is conceptually different for the 2D and 3D case. In the 3D case, the intensity acquired in a given pixel on the detector is a line integral: a sum of contributions from positions forming a line in the grain. Applying reconstruction principles to solve this inverse problem is clearly relevant, cf. Chapter 2.

In the 2D case with ideal optics and an ideal detector, any diffraction spot will be a projection of the boundary of the associated illuminated grain section. The projection implies that the contour of the grain boundary is demagnified in the direction along the beam by a factor of $\sin(2\theta)\cos(\eta)$. Hence, in principle the shape of any grain can be retrieved by backprojecting the periphery of one selected diffraction spot and correcting for the demagnification. This relatively straightforward approach has been described by Poulsen *et al.* (2001), which also includes verification by means of comparison to an electron backscatter diffraction (EBSD) map. This comparison is reproduced here as Fig. 9.6. Unfortunately, the algorithm is associated with a spatial resolution along the direction of the incident beam that is one order of magnitude worse than in the transverse direction.

This problem can be overcome by applying a reconstruction approach. During the last 3 years a number of reconstruction approaches have been suggested (Poulsen and Fu, 2003, Markussen *et al.*, 2004, Schmidt, 2007, Alpers *et al.*, 2006, Rodek *et al.*, 2007, Suter *et al.*, 2006). It is found that the task of performing such reconstructions is inherently different from absorption contrast tomography, as the number of projections in the latter case can be determined at will by adjusting the angular step. In contrast, in 3DXRD the number and directions of projections giving rise to non-zero intensities are given by crystallography. In practice, only 5 to 20 projections may be of sufficient quality. This effectively rules out the use of methods such as a 3DXRD version of the filtered backprojection algorithm.

In the following, a summary of three 2D approaches is given, all of which have been implemented in code optimized for data-processing speed. To enable quantitative comparisons, extensive simulations have been performed on phantoms. At the end, two applications are summarized. Corresponding work on algorithms for the 3D case is currently underway, see e.g. Markussen *et al.* (2004).

9.3.1 *Forward projection*

In their most basic implementation forward projection programs simply scan over the grid of pixels in the sample and for each pixel scan over orientation space to test which orientations match the diffraction patterns. Cross-talk between pixels in the layer is neglected. Hence, one can only infer the set of possible orientations. In practice, the constraint that all pixels belonging to a given grain must have the same orientation often implies that there is one and only one (possible) orientation for each pixel. If that is not the case – e.g. in cases with substantial spot overlap – the ambiguous result may be used as input for the inverse methods presented below.

Two forward projection programs have been developed by Suter *et al.* (2006) and Schmidt (2007), respectively. The latter program is called `Grainsweeper`. In this, the search for possible orientations is performed by a discrete tomographic approach. This algorithm has been applied to a number of real data sets. Typically, the reconstruction of a layer in the sample with say 100×100 pixels takes a few minutes.

9.3.2 *Algebraic solution*

An algebraic reconstruction method has been presented in detail by Poulsen and Fu (2003). It assumes that the orientations and approximate centre-of-mass positions of all grain sections are known a priori, e.g. determined by `Graindex` or `Grainsweeper`. The method attempts to reconstruct the boundary of each grain separately. For a specific grain this is done by associating a 'grain density' to each pixel in the layer. For pixels fully outside the grain the density should be 0, and for pixels fully inside it should be equal to a materials' constant ϱ_0. Pixels at the boundary may be partly inside the grain and should therefore have intermediate values.

For each of the diffraction spots associated with the grain of interest, an area-of-interest (AOI) on the detector is identified. The background-corrected pixel intensities within the AOI are normalized to well-known geometric factors (the structure factor and the Lorenz factor) for the reflection. For each reflection r the normalized intensities are saved in an array b^r. Under the basic assumption of kinematic scattering this enables a formulation in terms of a set of linear equations:

$$\underline{\underline{\boldsymbol{A}}}\boldsymbol{x} = \boldsymbol{b}.$$

Here, the unknown $\boldsymbol{x}$ comprises the pixel densities, and the information on the experimental setup and diffraction geometry is stored in the matrix $\underline{\underline{\boldsymbol{A}}}$. The compound vector $\boldsymbol{b}$ comprises the arrays b^r stacked on top of each other and thus forming a larger vector. The solution is constrained by the requirement stated above

$$0 \leq x_i \leq \rho_0; \forall i.$$

In general, $\underline{\underline{\boldsymbol{A}}}$ will be huge and very sparse with a rank that is a priori unknown. These facts in combination with the need for handling the constraint on density make an iterative solution attractive. So far, focus has been on adapting *algebraic reconstruction techniques* (ART, see Chapter 2) for this 3DXRD application – so-called 2D-ART.

Once the solution has converged, the grain boundary can be defined by setting a threshold. A full grain map may be obtained by superposing the solutions – the boundaries – of the individual grains. However, such a map will not be space filling as boundaries from neighboring grains may overlap or leave 'voids' in the map. To avoid this, one may choose for each pixel the grain where the reconstruction exhibited the largest density.

Simulations have been performed in order to estimate the spatial resolution of 2D-ART in the absence of experimental errors. The most critical parameter is found to be the number of reflections available for the analysis. Remarkably, reconstructions of reasonable quality are possible with as little as 5 reflections. The main disadvantage of 2D-ART is the fact that the grain map is pieced together from independent solutions. This gives rise to 'discretization errors' in the map. In principle, this problem could be avoided by simultaneous ART-type reconstruction of all grains. Unfortunately, $\underline{\underline{\boldsymbol{A}}}$ becomes extremely large. Novel discrete versions of ART are promising candidates for such reconstructions and are being tested at the time of writing.

The first experimental data for 2D-ART were presented by Poulsen and Fu (2003). The grain map resulting from an independent reconstruction of 27 grains based on 5 reflections from each is reproduced in Fig. 9.7. From the overlap between grains and voids in the figure the spatial resolution is estimated to be $\approx 5\,\mu$m. The limitation is identified to be the instrumental point-spread function. A few grains are obviously missing from the map as for the particular geometry used the number of independent reflections were too few to enable reconstruction.

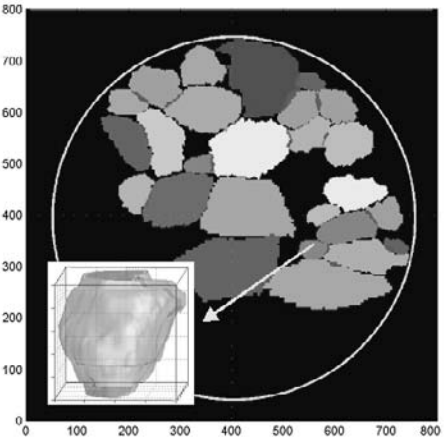


FIG. 9.7. First reported partial grain map of one layer using 2D-ART. The map is in units of μm with the surface of the aluminium polycrystal indicated by a solid circle. The grain sections were reconstructed independently. This data set was acquired in 10 min. By stacking layers, 3D maps are obtained, an example of which is shown in the insert. Orientations are represented by a greylevel scheme. Courtesy of X. Fu.

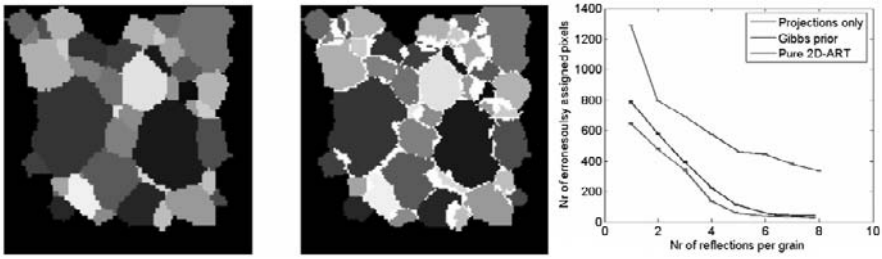
9.3.3 *Monte-Carlo-based reconstruction*

Stochastic approaches are attractive as they easily enable genuine simultaneous reconstructions of all grains rather than grain-by-grain reconstruction as 2D-ART above. On the other hand, based on experience in other fields one may fear that they would be slow and could 'get stuck' in local minima if the configuration space is too large.

Two approaches have been demonstrated:

- A *restoration approach*, where a coarse or ambiguous grain map is first generated by a forward projection algorithm (Suter *et al.*, 2006) or by ART (Alpers *et al.*, 2006). A Monte-Carlo-based routine is then used to 'restore' the correct pixel affiliations in the near-grain-boundary regions.
- An *indexing* approach. In this approach there is no need for an initial grain map. Instead, the grain map is reconstructed based only on the output from `Graindex` – that is the number of grains and their orientations (Alpers *et al.*, 2006).

The work by Alpers *et al.* (2006) comprises a comparison of algorithms based on simulations. The phantom used is shown in Fig. 9.8(a). With on average 8 available diffraction spots per grain the corresponding 2D-ART reconstruction – shown in Fig. 9.8(b) – was associated with ambiguous areas, appearing at the grain boundaries. Subsequent restorations of the 1490 pixels in these white areas converged rapidly. Based on 1000 Monte Carlo cycles the computation time was



(a) (b) (c)

FIG. 9.8. Reconstruction of a 124×124 grain map with 44 undeformed grains. (a) original image. (b) reconstruction using 2D-ART. The white areas indicate areas of ambiguity. (c) Comparison of three reconstruction algorithms: 2D-ART, 2D-ART plus stochastic restorations of ambiguous areas and `Graindex` plus stochastic restorations. Shown is the quality of the reconstruction as a function of the number of useful spots per grain. From Alpers *et al.* (2006) (with permission from the IUC).

10 s. As a result all pixels except for 43 were correctly restored. Applying the indexing approach to the same data set resulted in longer running times. Stopping the algorithm somewhat arbitrarily after one minute, the number of incorrectly assigned pixels was 120. In Fig. 9.8(c) a comparison of the performance of the three methods as a function of number of spots available is given.

9.3.4 *Applications*

9.3.4.1 *Application I: recrystallization* The first 3D movies of grain coarsening were related to the observation of an individual grain (nucleus) during the early stage of recrystallization (Schmidt *et al.*, 2004). More specifically, nucleation in an aluminium single crystal deformed to 42% was stimulated by hardness indents. The crystal was subject to annealing for 30 h at temperatures between 270 °C and 310 °C while being characterized *in-situ*. The mapping comprised 50 z-layers with a spacing of 6 µm. One emerging nucleus was identified at an early stage of the growth. Exposures were made around the ω setting of a specific reflection belonging to this grain. The nucleus was mapped continuously with a time resolution of at most 10 min. Snapshots of the resulting 3D movie are shown in Fig. 9.9. The data quality is sufficiently high so that two general features of growth are revealed:

- Growth is very heterogeneous, so the shape is at times remarkably irregular.
- Growth does not occur smoothly with time, but is often jerky.

These unique observations were explained by the heterogeneous nature of the deformed microstructure.

9.3.4.2 *Application II: grain growth* Recently, Schmidt and co-workers performed the first grain-growth study in which the evolution of the morphology of

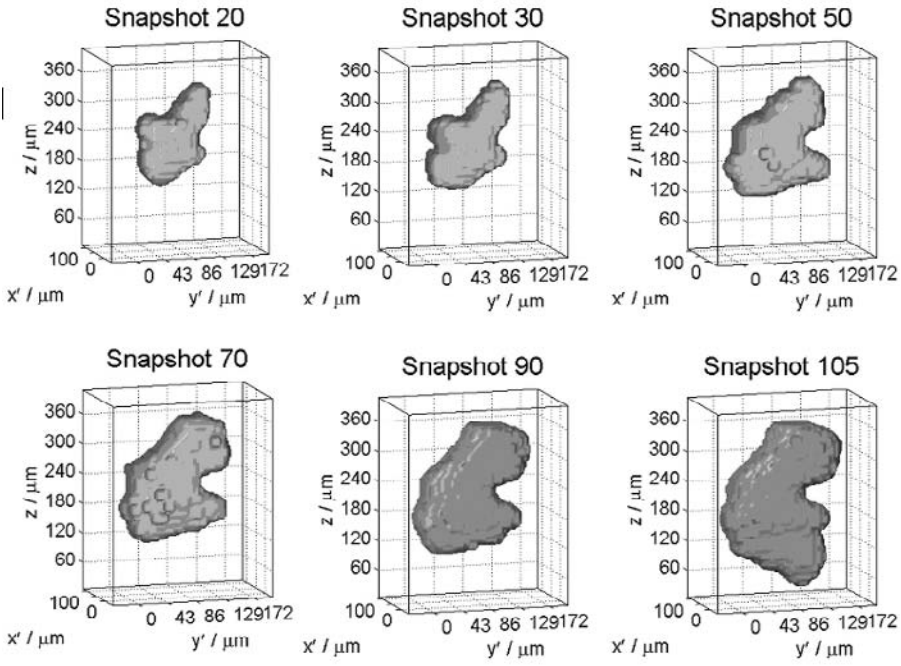


FIG. 9.9. Storyboard visualising the growth of one emerging nucleus in an aluminium single crystal deformed by 42% as a function of annealing. Selected snapshots are shown. From Schmidt *et al.* (2004) (with permission from Science).

several hundred grains was monitored simultaneously. A cylindrical aluminium sample with a diameter of 700 µm was mapped layer by layer within a total height of 350 µm and with a resolution of ≈ 10 µm. The sample was annealed in an external furnace. The grain volumes were reconstructed using the `Grainsweeper` algorithm. In total, five annealing steps were made at annealing temperatures between 400 °C and 450 °C. Initially, 491 grains were present, while only 49 grains remained after the final annealing step. At the time of writing, the data analysis is still in progress.

9.4 Mapping of orientations within deformed specimens

For deformed specimens it is desirable to generate orientation maps in which each voxel in the specimen is associated with one and only one orientation. This is at times an approximation as each voxel may actually comprise several orientations. In such cases the idea is to reconstruct the most prominent orientation. In 2D, such maps may be familiar as they are equivalent to those routinely generated by EBSD.

Mathematically speaking, the task in this case is to reconstruct a 3D vector-field $\boldsymbol{r}(\boldsymbol{x})$, where $\boldsymbol{r}$ symbolizes orientations and $\boldsymbol{x}$ spatial position. Notably, this particular vector field is not continuous, neither at the grain boundaries nor within the grains. Furthermore, the intensity acquired in a given detector pixel is not the sum over some projection of $\boldsymbol{r}$ along a given direction in direct space. Instead, each projection $\boldsymbol{x}$ contributes if and only if $\boldsymbol{r}(\boldsymbol{x})$ is directed along a certain line – the projection line.

To the knowledge of the author no transform or algebraic reconstruction algorithm is readily available for such fields. As we shall see shortly, a stochastic approach to reconstruction is possible. However, for such an approach it is a concern that the configuration space is much larger for the deformed case than for the undeformed.

Fortunately, the microstructure exhibits additional properties that may be exploited in the reconstruction. This is true in particular for *moderately deformed* specimens: that is specimens where the degree of deformation applied is so low that it still is possible to distinguish the boundaries of the (distorted) grains, although perhaps with some ambiguity. In particular, grains – whether distorted or not – exhibit the following properties:

- they are discrete, simply connected 3D space-filling objects,
- their boundaries are smooth,
- their morphologies are not random.

It follows from this list that tools from the new discipline of discrete tomography should be well suited for the reconstruction of grain maps (see Section 2.2.3).

9.4.1 *Discrete tomography algorithm for moderately deformed specimens*

Recently, an algorithm for moderately deformed specimens was presented by Rodek *et al.* (2007). This is a Monte-Carlo-based approach where the discrete nature of the microstructure is exploited. The routine assumes that prior information – obtained e.g. from `Grainsweeper` – makes it possible to identify the number of grains, and for each determine an approximate average orientation and one pixel in the layer that belongs to this grain. These points act as seeds for the algorithm.

The algorithm associates each pixel in the illuminated layer by both an orientation and a label that identifies to which grain it belongs. Initially, except for the seeds all orientations are unknown and no labels are assigned. Then two maps – an orientation map and a grain map – are constructed simultaneously according to three principles:

- a pixel that has not yet been labelled can be given the label and orientation of an adjacent labelled pixel,
- a labelled pixel may retain its association with a grain, while its orientation is replaced by one of its nearest neighbours in orientation space,
- A labelled pixel at a grain boundary may be given the same label and orientation as a neighbouring pixel associated with an adjacent grain.

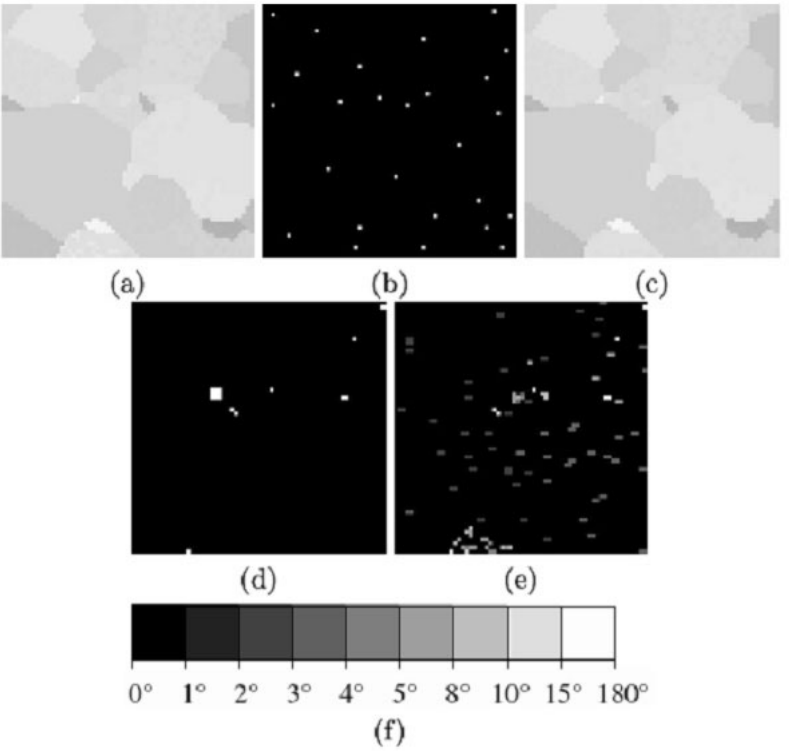


FIG. 9.10. Reconstruction of a test case representing a deformed microstructure. (a): Reference orientation map; (b): seeds; (c): reconstructed orientation map; (d): difference of the reference and the reconstructed grain maps, where black pixels denote identical grain labels, white pixels represent mismatching ones; (e): difference of the reference and the reconstructed orientation maps, where the brightness of the pixels is determined by the misorientation of corresponding orientation pairs, as shown in the greyscale key at the bottom of the figure. From Rodek *et al.* (2007) (with permission from the IUC).

The choice of whether to accept a suggested update of the maps or not is determined by the *Metropolis criterion* using an object function comprising three terms. One of these measures the correspondence between experimental data and the forward projection of the current orientation map. The other two are optional: they enforce a certain degree of homogeneity of orientations within a grain and penalize 'unphysical' grain boundaries. In Rodek *et al.* (2007) this approach is tested on a set of 64×64 pixel phantoms representing microstructures typical of various degrees of deformation. The results are encouraging. In general, the reconstructions converged within 600 Monte Carlo cycles, corresponding to about 8 min of computation time on a modern personal computer (year 2006).

The reconstruction of a microstructure with 22 grains and an orientation spread within individual grains of up to 22° is shown in Fig. 9.10. In this reproduction of the maps (a) and (c) it is difficult to see the orientation gradient within the grains since the variations in greylevel are quite small. At low noise levels the reconstructions based on a comparison with the experimental data were nearly perfect. At high noise levels the performance was also surprisingly good, in particular when including the additional two terms in the object function. At the time of writing, analysis on real data is pending.

9.4.2 *Heavily deformed materials*

Heavily deformed materials are a major challenge to 3DXRD. With increasing deformation the size of the subgrains become smaller, and at the same time the misorientation between neighbouring subgrains increases. Given a fixed spatial resolution this implies that above a certain degree of deformation the basic assumption of associating each voxel in the sample with one and only one orientation fails.

To overcome this problem it may be necessary to operate in the full 6D position–orientation space where each position is associated with an orientation distribution. The geometry of this space and the associated 3DXRD projection surfaces have been laid out by Poulsen (2003). Operating in this space is not trivial. In particular, both the projection surfaces and the space itself are curved. To the knowledge of the author this rules out the use of transform methods based on Fourier transforms such as, e.g. equivalents of the filtered backprojection. On the other hand, algebraic and stochastic approaches are possible, e.g. the ART routine can be generalized to the 6D case (Poulsen, 2003).

An obvious obstacle is the size of the 6D space: to represent orientation elements of 113° over say $100 \times 100 \times 10$ positions in a sample requires a discretization into $\approx 10^{10}$ 6D voxels. Fortunately, the vast majority of this space will be empty, a fact that should be used in the analysis. As such, reconstructions based on wavelet theory are currently being investigated.

9.5 Combining 3DXRD and tomography

The 3DXRD setup sketched in Fig. 9.1 is very similar to the setup typically used for parallel-beam tomography (see Chapters 5 and 6). Furthermore, it appears that the specifications of the two methods in terms of space and time resolution are not that different. Hence, it is relevant to attempt probing the same specimen with both techniques either simultaneously or consecutively. This combination is most interesting as diffraction and tomography utilize different contrast mechanisms. While 3DXRD is sensitive to variations in phase, orientation and elastic strain of crystalline materials, tomography is sensitive to density variations, regardless of whether the material is crystalline or not.

The possible applications of combined 3DXRD and tomography data sets are numerous:

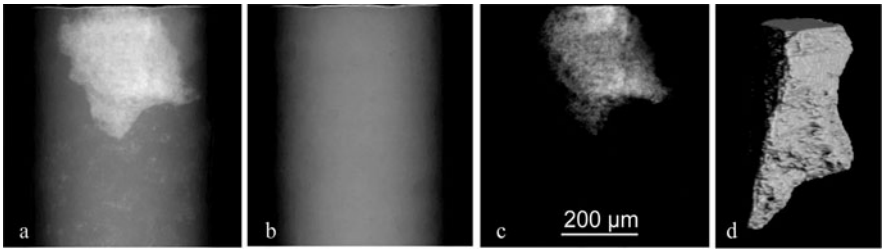


FIG. 9.11. Diffraction-contrast tomography: (a) experimental projection with one grain fulfilling the Bragg condition, (b) calculated absorption background, (c) extracted grain projection, in all cases with image brightness proportional to attenuation. (d) reconstruction of 3D grain shape by means of algebraic reconstruction techniques (different scale). Courtesy of W. Ludwig.

- The basic 3DXRD algorithms developed so far all apply to monophase materials only. A combined data set enables the methodology to be extended to multiphase materials by first segmenting the volume to be reconstructed based on density and then applying the 3DXRD method to each phase one at a time.
- On the macroscopic scale the possibility of combining tomography and strain scanning has been realized by a number of authors (Preuss *et al.*, 2002, Pyzalla *et al.*, 2005). On the mesoscale the combination of tomography and 3DXRD is equally appealing. In particular, for fatigue and damage-control studies one may investigate the interaction between propagating voids or cracks, the associated elastic strain fields and the grain or dislocation structures in the material.
- Recently, a method has been demonstrated for the characterization of 3D plastic strain fields using X-ray tomography and embedded marker particles (Nielsen *et al.*, 2003). Strains in the range of 1% upwards can be mapped with a spatial resolution of $\approx 20\,\mu m$ within millimetre-sized specimens. Combing plastic strain mapping with grain mapping by 3DXRD would be a very powerful tool for studies of polycrystal deformation.

Conventional wisdom tells us that attenuation and/or phase-contrast tomography is not sensitive to orientations and as such cannot be used to map grains in a monophase material. Recently, two methods that overcome this basic limitation have been demonstrated by Ludwig and collaborators from ESRF and Risø National Laboratory.

- *Diffraction-contrast tomography (DCT).* Figure 9.11 illustrates the principle. Assume a grain is in the Bragg condition. Then there will be two contributions to attenuation of the direct beam: absorption in the sample and the missing intensity due to diffraction. The latter additional contrast is clearly visible if the grains are nearly perfect, i.e. their mosaic spread

is well below 1°. Performing a conventional reconstruction first, a density map of the sample is obtained. Based on this, one can simulate what the absorption data should look like for a non-crystalline sample. Subtracting the actual images from the simulated ones, one may sort the residual 'grain contrast' images, based on a combination of real space (grain shape and position) and orientation space scanning procedures that compare simulated and measured angles. Having indexed a grain in this way the 3D shape can be reconstructed based on the 3D-ART approach.
DCT and 3DXRD may be seen as the equivalent of 'bright-field' and 'dark-field' imaging in TEM. Combined data sets have been acquired in several experiments at ID19 at ESRF. The results demonstrate that this combined method produces grain maps of superior resolution to those based on 3DXRD only. The drawback is that the number of projections required needs to be much higher than for 3DXRD, of the order of a few thousand images.

- *Topo-tomography* is an alternative tomographic scanning procedure that allows reconstructing a single grain using a setup with additional tilt stages (Ludwig *et al.*, 2001). For a selected grain the scattering vector of a suited reflection is aligned parallel to the rotation axis of the tomographic setup. This particular setting assures that the diffraction condition is maintained while the sample is turned 360° around the rotation axis. Tomographic reconstruction can be performed from data acquired in direct and/or in diffracted beams by means of a standard (cone beam) filtered backprojection algorithm. In comparison to mode III reconstructions by 3DXRD above, the spatial resolution can be substantially better. The disadvantage is the restriction to one grain only.

9.6 3DXRD microscopes

Presently, one dedicated 3DXRD microscope, commissioned in 2000 at the Materials Science beamline of the ESRF (European Synchrotron Radiation Facility, 2006, Risø National Laboratory, 2006) exists. The optics enable us to work in the energy range 30 to 90 keV with a choice between parallel beam, vertical focusing only and with both vertical and horizontal focusing. The energy bandwidth can be varied from 1×10^{-4} to 1×10^{-2}, with the most popular setting being $\approx 1 \times 10^{-3}$. The microscope can carry loads of up to 200 kg. It is equipped with a 25-kN stress rig and several furnaces with operational temperatures of up to 1300 °C. In addition, 3DXRD experiments have been performed at 'Sector 1' at the APS (Jakobsen *et al.*, 2006). Currently, this setup is being converted into a dedicated experimental station.

9.7 Geometric principles

For reference purposes we provide key equations relating 3DXRD diffraction events with observations. Next, the geometry of orientation space is laid out and it is explained what relationship exists between orientations and observations. A

more detailed description has been given by Poulsen (2004). We will assume the material to be monophase with a known atomic structure. Furthermore, we will neglect elastic strain. Most of the algorithms presented above in Sections 9.2–9.5 are not sensitive to strain in any case.

9.7.1 *Diffraction geometry*

The geometry is similar to the 'rotation method' for single-crystal diffraction (Warren, 1990, Als-Nielsen and McMorow, 1999, Giacovazzo *et al.*, 2002). Assume a diffraction event takes place at position (x, y, z) in the sample co-ordinate system. Let the corresponding co-ordinates in the laboratory system be (x_l, y_l, z), see Fig. 9.1. Then

$$\begin{pmatrix} x_l \\ y_l \\ z \end{pmatrix} = \underline{\underline{\boldsymbol{\Omega}}} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} \cos(\omega) & -\sin(\omega) & 0 \\ \sin(\omega) & \cos(\omega) & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}. \tag{9.1}$$

The direction of the diffracted ray is determined by the Bragg angle 2θ and the azimuthal angle η, both defined in Fig. 9.1. Assuming the detector plane to be perpendicular to the incoming beam, the intersection between the ray and the detector plane is $(L, y_{\mathrm{det}}, z_{\mathrm{det}})$. Here, by definition $(y_{\mathrm{det}}, z_{\mathrm{det}}) = (0, 0)$ for the intersection of the detector plane with the incident ray passing through $(x_l, y_l, z) = (0, 0, 0)$.

The scattering vector associated with the diffraction event is denoted $\boldsymbol{G}$. To describe its relationship with reciprocal space three Cartesian co-ordinate systems are introduced: the laboratory system, the sample system, and the Cartesian grain system. These are identified by subscripts l, s and c, respectively. Hence, the scattering vector transforms as $\boldsymbol{G_l} = \underline{\underline{\boldsymbol{\Omega}}}\boldsymbol{G}_s$ with $\underline{\underline{\boldsymbol{\Omega}}}$ given by eqn 9.1. The crystallographic orientation of a grain with respect to the sample is represented by the matrix $\underline{\underline{\boldsymbol{g}}}$,

$$\boldsymbol{G}_c = \underline{\underline{\boldsymbol{g}}}\boldsymbol{G}_s, \tag{9.2}$$

where the index c refers to a Cartesian grain system $(\hat{x}_c, \hat{y}_c, \hat{z}_c)$. This is fixed with respect to the reciprocal lattice $(\boldsymbol{a}^*, \boldsymbol{b}^*, \boldsymbol{c}^*)$ in the grain. We use the convention that $\hat{x}_c$ is parallel to $\boldsymbol{a}^*$, $\hat{y}_c$ is in the plane of $\boldsymbol{a}^*$ and $\boldsymbol{b}^*$, and $\hat{z}_c$ is perpendicular to that plane. Let $\boldsymbol{G}$ be represented in the reciprocal lattice system by the Miller indices $\boldsymbol{G}_{hkl} = (h, k, l)^t$. The correspondence between the Cartesian grain system and reciprocal space is then given by the $\underline{\underline{\boldsymbol{B}}}$ matrix: $\boldsymbol{G}_c = \underline{\underline{\boldsymbol{B}}}\boldsymbol{G}_{hkl}$, with

$$\underline{\underline{\boldsymbol{B}}} = \begin{pmatrix} a^* & b^*\cos(\gamma^*) & c^*\cos(\beta^*) \\ 0 & b^*\sin(\gamma^*) & -c^*\sin(\beta^*)\cos(\alpha) \\ 0 & 0 & c^*\sin(\beta^*)\sin(\alpha) \end{pmatrix}$$

and

$$\cos(\alpha) = \frac{\cos(\beta^*)\cos(\gamma^*) - \cos(\alpha^*)}{\sin(\beta^*)\sin(\gamma^*)}.$$

Here, $(a^*, b^*, c^*, \alpha^*, \beta^*, \gamma^*)$ symbolize the lattice parameters in reciprocal space.

With these definitions the normalized scattering vectors in the laboratory system, $\boldsymbol{G}_l/\|\boldsymbol{G}_l\|$, and in the sample system, $\boldsymbol{y}$, can be expressed as

$$\frac{\boldsymbol{G}_l}{\|\boldsymbol{G}_l\|} = \underline{\underline{\boldsymbol{\Omega}}}\boldsymbol{y} = \underline{\underline{\boldsymbol{\Omega}}}\left(\underline{\underline{\boldsymbol{g}}}^{-1}\boldsymbol{h}\right) = \cos(\theta)\begin{pmatrix}-\tan(\theta)\\ -\sin(\eta)\\ \cos(\eta)\end{pmatrix}. \tag{9.3}$$

The unit vector $\boldsymbol{h}$ is defined as

$$\boldsymbol{h} = \begin{pmatrix}h_1\\ h_2\\ h_3\end{pmatrix} = \frac{\underline{\underline{\boldsymbol{B}}}\boldsymbol{G}_{hkl}}{\left\|\underline{\underline{\boldsymbol{B}}}\boldsymbol{G}_{hkl}\right\|}.$$

With these definitions

$$\begin{aligned} y_{\mathrm{det}} &= y_l + \frac{(L-x_l)\tan(2\theta)}{\cos(\theta)}\underline{\underline{\boldsymbol{\Omega}}}\left(\underline{\underline{\boldsymbol{g}}}^{-1}\right)h_2, \\ z_{\mathrm{det}} &= z + \frac{(L-x_l)\tan(2\theta)}{\cos(\theta)}\underline{\underline{\boldsymbol{\Omega}}}\left(\underline{\underline{\boldsymbol{g}}}^{-1}\right)h_3. \end{aligned} \tag{9.4}$$

The equations above comprise the desired relationships between the position (x, y, z), the orientation matrix $\underline{\underline{\boldsymbol{g}}}$ and the observables: the intensities in the detector pixels parameterized by $(y_{det}, z_{det}, \omega, L)$.

9.7.2 *Representation of crystallographic orientation*

Crystallographic orientations can be expressed in numerous ways, as described in detail in the literature on texture (Kocks *et al.*, 1998, Morawiec, 2004). For algebra, the natural choice is the 3×3 orthogonal matrix $\underline{\underline{\boldsymbol{g}}}$ as defined by eqn 9.2. For sampling, a representation by three parameters is preferable. Even more so for visualization, as this enables the use of colour schemes. Traditionally, orientations are parameterized by a set of *Euler angles* $(\varphi_1, \phi, \varphi_2)$, expressing subsequent rotations around three axes (Bunge, 1969). However, this representation is badly suited for numerical work, as it is non-linear with mathematical singularities. Instead, it is recommended to use *Rodrigues vectors* (Frank, 1988) – our choice in the following – or *unit quaternions* (Morawiec, 2004). The Rodrigues representation is based on the fact that any rotation can be represented in a unique way by a rotation axis $\boldsymbol{n}$ and a rotation angle ϕ, defined on [0,π]. The Rodrigues vector is defined as (see Fig. 9.12)

$$\boldsymbol{r} = \tan(\phi/2)\boldsymbol{n}.$$

The vector $\boldsymbol{r}$ can be treated as a vector in a three-dimensional Euclidian space, with the exception of points with a rotation angle of π, which are represented by two opposite points in infinity. The axes of this so-called Rodrigues space are collinear with those of the sample system in the sense, that a vector $\boldsymbol{r} = (r_1, 0, 0)$ describes a rotation around the sample x-axis.

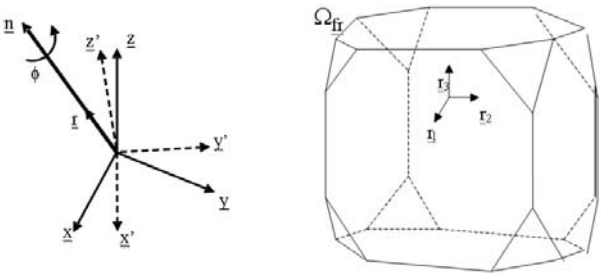


FIG. 9.12. (a): Definition of the Rodrigues vector $\boldsymbol{r}$. The co-ordinate system (x, y, z) is rotated around $\boldsymbol{n}$ by angle ϕ into (x', y', z'). $\boldsymbol{r}$ is parallel to $\boldsymbol{n}$: $\boldsymbol{r} = \boldsymbol{n}\tan(\phi/2)$. (b): The fundamental zone in Rodrigues space for face-centred cubic systems

From a single diffraction event, only the direction of the scattering vector is probed. Measurements will be invariant to a rotation of the sample around the vector. The set of orientations, which for given $\boldsymbol{y}$ and $\boldsymbol{h}$ fulfil $\boldsymbol{y} = \underline{\underline{\boldsymbol{g}}}^{-1}\boldsymbol{h}$ (cf. eqn 9.3) constitutes a curve in orientation space, the so-called projection line for pole-figure inversion. In Rodrigues space, projection lines are straight. Specifically, for a given set of vectors $\boldsymbol{h}$ and $\boldsymbol{y}$, the projection line is given by

$$\boldsymbol{r} = \boldsymbol{r}^0 + t\frac{\boldsymbol{h} + \boldsymbol{y}}{1 + \boldsymbol{h} \cdot \boldsymbol{y}}; \quad -\infty < t < \infty$$

and

$$\boldsymbol{r}^0 = \frac{\boldsymbol{h} \times \boldsymbol{y}}{1 + \boldsymbol{h} \cdot \boldsymbol{y}} = \tan(\frac{\phi_0}{2})\boldsymbol{n},$$

where $\boldsymbol{r}^0$ is the rotation from $\boldsymbol{h}$ into $\boldsymbol{y}$ with the minimum rotation angle ϕ_0.

Operating with orientations is complicated by the fact that the crystalline lattice maps onto itself when certain symmetry operations are applied as specified by crystallography (Morawiec, 2004). This crystal symmetry implies that Rodrigues space is tessellated into N regions of different size, N being the number of symmetry operations. A given set of data will be consistent with N orientations, one belonging to each of these regions. One of the regions is the fundamental region Ω_{fr}: a polyhedron centred around the origin and with its faces aligned with the symmetry axes. As an example, for cubic crystal symmetry $N = 24$ and the fundamental region is a truncated cube, illustrated in Fig. 9.12. For a detailed calculation of the shape of the fundamental region and the boundary conditions for various crystal symmetries see Morawiec and Field (1996).

Confining the solution to Ω_{fr} the projection line extending through the full Rodrigues space is transformed into a set of projection lines given by the number of reflections belonging to the $\{hkl\}$-family in question. As an example, for a $\{200\}$ reflection of a face-centred cubic material, there are 6 projection lines corresponding to reflections: (200), (–200), (020), (0–20), (002) and (00–2). Only one of these will pass through the point marking the actual orientation.

9.8 Summary and outlook

At the time of writing, the ID11 beamline at ESRF is undergoing a major refurbishment. A new generation of hard X-ray optics, an advanced feedback system and temperature control will be installed in order to generate a stable, fully energy and bandwidth tuneable beam, with dimensions as small as 50 to 100 nm. At the same time, work is ongoing to develop a new type of structured fluorescence screen for the high-resolution detectors. This detector is specified to have a resolution of 1 μm and an efficiency in excess of 25%. For comparison, present detectors have a resolution of 5 μm and an efficiency of a few per cent.

For work in mode I and II these developments are predicted to enable *in-situ* studies of crystalline structures as small as 5 nm. For work in modes III and IV the aim is the generation of 3D maps with a spatial resolution of 1 μm or alternatively complete maps of specimens 1 mm^3 large with a resolution of 5 μm in a few minutes.

9.9 References

Alpers, A., Poulsen, H.F., Knudsen, E., and Herman, G.T. (2006). A discrete tomography algorithm for improving the quality of three-dimensional X-ray diffraction grain maps. *Journal of Applied Crystallography*, **39**, 582.

Als-Nielsen, J. and McMorow, D. (1999). *Elements of X-ray physics*. Wiley, New York.

Bunge, H.J. (1969). *Matematische Methoden der Texturanalyse*. Akademie Verlag, Berlin.

Bunge, H.J., Wcislak, L., Klein, H., Garbe, U., and Schneider, J.R. (2003). Texture and microstructure imaging in six dimensions with high-energy synchrotron radiation. *Journal of Applied Crystallography*, **36**, 1240.

European Synchrotron Radiation Facility (2006). Facility websites. `http://www.esrf.fr`. [Online; accessed 05 January 2007].

Frank, F.C. (1988). Orientation mapping. *Metallurgical Transactions A*, **19**, 403.

Giacovazzo, C., Monaco, H.L., Viterbo, D., Scordari, F., Gilli, G., Zanotti, G., and Catti, M. (2002). *IUCr texts on crystallography 2: Fundamentals of crystallography*. Oxford University Press, Oxford.

Gundlach, C., Pantleon, W., Lauridsen, E.M., Margulies, L., Doherty, R., and Poulsen, H.F. (2004). Direct observation of sub-grain evolution during recovery of cold-rolled aluminium. *Scripta Materialia*, **50**, 477.

Iqbal, N., van Dijk, N.H., Offerman, S.E., Moret, M.P., Katgerman, L., and Kearley, G.J. (2005). Real-time observation of grain nucleation and growth during solidification of aluminium alloys. *Acta Materialia*, **53**, 2875.

Jakobsen, B., Poulsen, H.F., Lienert, U., Almer, J., Shastri, S.D., Sørensen, H.O., Gundlach, C., and Pantleon, W. (2006). Formation and subdivision of deformation structures during plastic deformation. *Science*, **312**, 889.

Kocks, U.F., Tome, C.N., and Wenk, H.R. (1998). *Texture and anisotropy*. Cambridge University Press, Cambridge.

Lange, W.F., Enornoto, M., and Aaronson, H.I. (1988). The kinetics of ferrite nucleation at austenite grain-boundaries in Fe-C alloys. *Metallurgical Transactions A*, **19**, 427.

Larson, B.C. and Lengeler, B. (2004). *High-resolution three-dimensional X-ray microscopy*, Volume 29(3) of *MRS Bulletin*. Materials Research Society, Warrendale.

Larson, B.C., Yang, W., Ice, G.E., Budai, J.D., and Tischler, T.Z. (2002). Three-dimensional X-ray structural microscopy with submicrometre resolution. *Nature*, **415**, 887.

Lauridsen, E.M., Juul Jensen, D., Poulsen, H. F., and Lienert, U. (2000). Kinetics of individual grains during recrystallization. *Scripta Materialia*, **43**, 561.

Lauridsen, E.M., Poulsen, H.F., Nielsen, S.F., and Juul Jensen, D. (2003). Recrystallization kinetics of individual bulk grains in 90% cold-rolled aluminium. *Acta Materialia*, **51**, 4423.

Lauridsen, E.M., Schmidt, S., Suter, R.M., and Poulsen, H.F. (2001). A method for structural characterization of grains in powders or polycrystals. *Journal of Applied Crystallography*, **34**, 744.

Ludwig, W., Cloetens, P., Härtwig, J., Baruchel, J., Hamelin, B., and Bastie, P. (2001). Three-dimensional imaging of crystal defects by 'topo-tomography'. *Journal of Applied Crystallography*, **34**, 602.

Margulies, L., Winther, G., and Poulsen, H. F. (2001). In situ measurement of grain rotation during deformation of polycrystals. *Science*, **291**, 2392.

Markussen, T., Fu, X., Margulies, L., Lauridsen, E.M., Nielsen, S.F., Schmidt, S., and Poulsen, H.F. (2004). An algebraic algorithm for generation of three-dimensional grain maps based on diffraction with a wide beam of hard X-rays. *Journal of Applied Crystallography*, **37**, 96.

Martins, R.V., Margulies, L., Schmidt, S., Poulsen, H.F., and Leffers, T. (2004). Simultaneous measurement of the strain tensor of 10 individual grains embedded in an Al tensile sample. *Materials Science and Engineering A*, **387-389**, 84.

Morawiec, A. (2004). *Orientation and rotations - Computations in crystallographic textures*. Springer, Berlin.

Morawiec, A. and Field, D.P. (1996). Rodrigues parameterization for orientation and misorientation distributions. *Philosophical Magazine A*, **73**, 1113.

Nielsen, S.F., Poulsen, H.F., Beckmann, F., Thorning, C., and Wert, J.A. (2003). Measurements of plastic displacement gradient components in three dimensions using marker particles and synchrotron X-ray absorption microtomography. *Acta Materialia*, **51**, 2407.

Offerman, S.E., van Dijk, N.H., Sietsma, J., Grigull, S., Lauridsen, E.M., Margulies, L., Poulsen, H.F., Rekveldt, M.T., and van der Zwaag, S. (2002). Grain nucleation and growth during phase transformations. *Science*, **298**, 1003.

Poulsen, H.F. (2003). A six-dimensional approach to microstructure analysis.

Philosophical Magazine, **83**, 2761.

Poulsen, H.F. (2004). *Three-dimensional X-ray diffraction microscopy*. Springer, Berlin.

Poulsen, H.F. and Fu, X. (2003). Generation of grain boundary maps by an algebraic reconstruction technique. *Journal of Applied Crystallography*, **36**, 1062.

Poulsen, H.F., Margulies, L., Schmidt, S., and Winther, G. (2003). Lattice rotations of individual bulk grains part I: 3D X-ray characterization. *Acta Materialia*, **51**, 3821.

Poulsen, H.F., Nielsen, S.F., Lauridsen, E.M., Schmidt, S., Suter, R.M., Lienert, U., Margulies, L., Lorentzen, T., and Jensen, D.J. (2001). Three-dimensional maps of grain boundaries and the stress state of individual grains in polycrystals and powders. *Journal of Applied Crystallography*, **34**, 751.

Preuss, M., Withers, P.J., Maire, E., and Buffiere, J.Y. (2002). SiC single fibre full-fragmentation during straining in a Ti6Al4V matrix studied by synchrotron X-rays. *Acta Materialia*, **50**, 3177.

Pyzalla, A., Camin, B., Buslaps, T., Michiel, M. Di, Kaminski, H., Kottar, A., Pernack, A., and Reimers, W. (2005). Simultaneous tomography and diffraction analysis of creep damage. *Science*, **308**, 92.

Risø National Laboratory (2006). Facility websites. `http://www.risoe.dk/afm/synch`. [Online; accessed 05 January 2007].

Rodek, L., Poulsen, H.F., Knudsen, E., and Herman, G.T. (2007). A stochastic algorithm for reconstruction of grain maps of moderately deformed specimens based on X-ray diffraction. in print.

Schmidt, S. (2007). The Grainsweeper. preprint.

Schmidt, S., Nielsen, S.F., Gundlach, C., Margulies, L., Huang, X., and Jensen, D.J. (2004). Watching the growth of bulk grains during recrystallization of deformed metals. *Science*, **305**, 229.

Schmidt, S., Poulsen, H.F., and Vaughan, G.B.M. (2003). Structural refinements of the individual grains within polycrystals and powders. *Journal of Applied Crystallography*, **36**, 326.

Suter, R.M., Hennessy, D., Xiao, C., and Lienert, U. (2006). Forward modeling method for microstructure reconstruction using X-ray diffraction microscopy: Single crystal verification. *Review of Scientific Instruments*, **77**, 123905.

Vaughan, G.B.M., Schmidt, S., and Poulsen, H.F. (2004). Multicrystal approach to crystal structure solution and refinement. *Zeitschrift für Kristallographie*, **219**, 813.

Warren, B.E. (1990). *X-ray diffraction*. Tracts in Modern Physics. Springer, Dover, New York.

Winther, G., Margulies, L., Schmidt, S., and Poulsen, H. F. (2004). Lattice rotations of individual bulk grains part II: Correlation with initial orientation and model comparison. *Acta Materialia*, **52**, 2863.

Wroblewski, T., Clauss, O., Crostack, H.-A., Ertel, A., Fandrich, F., Genzel,

Ch., Hradil, K., Ternes, W., and Woldt, E. (1999). A new diffractometer for materials science and imaging at HASYLAB beamline G3. *Nuclear Instruments and Methods in Physics Research A*, **428**, 570.

10

DETECTORS FOR SYNCHROTRON TOMOGRAPHY

Heinz Graafsma and Thierry Martin

Two-dimensional X-ray detectors have become a standard at synchrotron sources for practically all applications including tomography. Consequently, a lot of development to optimize these detectors has been, and is still being, done. One can divide area detectors into two main classes: in *direct detection* the X-ray photons are directly converted to an electrical signal via electron–hole pair creation, either in a semiconductor or in a gas. In *indirect detection* the X-ray photons are first converted into visible light that is subsequently converted to an electronic signal in the semiconductor.

Direct detection has certain advantages. It avoids the conversion step into visible light that can lead to degradations of the recorded image, as will be explained in detail in this chapter. It also allows for single-photon counting, where each individual X-ray photon is detected and processed, resulting in a virtually noise-free detector that is sensitive to the energy of the incident X-rays. Such detectors are under development and require quite sophisticated electronics.

A disadvantage of direct detection is that commercial sensors such as charge-coulpled devices (CCD), being mainly developed for visible-light detection, cannot be used because of their low sensitivity in the X-ray regime. Other disadvantages of photon counting are the need for sophisticated electronics and the fact that time is needed to process each photon, which limits the incident flux that can be detected. The alternative is to convert the X-ray photons to visible light and to use integrating detectors such as CCDs. So far, the most effective, low-cost and low-risk solution for tomography has been indirect detection, where a scintillator converter screen is optically coupled to an integrating detector, either a commercial CCD camera, or a specially developed amorphous photodiode array. In this chapter we will explain this technology and discuss the various components. We will start with a survey of the various existing types of X-ray conversion screens. This is followed by a discussion of the optical relays and the light sensors required to convert the light into a digital image.

A large number of beamlines at synchrotron radiation sources use tomography X-ray imaging techniques in an increasingly wide range of research, including materials science, medical, geophysical, environmental or paleontological studies. Consequently, both the input field of view (FOV) and the required spatial resolution span a wide range. For instance, the largest FOVs for tomography at synchrotron radiation sources is at medical beamlines and can be as large as $300\,\text{mm}\times20\,\text{mm}$. The currently smallest FOV is $300\times300\,\mu\text{m}^2$ with

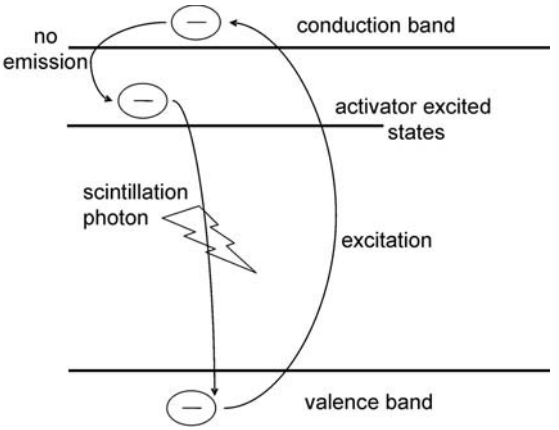


FIG. 10.1. The electronic band structure of an activated inorganic crystal scintillator. When an X-ray photon interacts with the scintillator material, electrons move from the valence band to the conduction band. It can fall down into one of the excited states of the activator without emission. When it is transmitted from the activator energy band back to the ground state, a photon is emitted. These de-excitations are in the visible-light band.

2048×2048 pixels each of 0.14 μm (Koch *et al.*, 1999*a*). This large range of three orders of magnitude necessarily implies the use of different technologies and materials but they are almost all based on the principle of an X-ray converter screen optically relayed by either lens coupling or tapered fibre-optic coupling to commercial CCD cameras. The exception is the large panel imagers developed for the medical imaging market, which will be briefly discussed at the end of the chapter.

10.1 Scintillation mechanism

Detectors used for tomography at synchrotron beamlines use a converter screen to convert X-ray photons to visible-light photons that are subsequently detected by a photodetector, mostly of the CCD type. This conversion is done via a scintillation or luminescence process, and therefore these converter screens are often called *scintillation screens*. The converter screens used are made of inorganic scintillators that have a high density and a high effective atomic number, resulting in high stopping power and high light yield for X-rays photons. In the so-called *intrinsic scintillators* the luminescence is produced by part of the crystal lattice itself. These scintillation materials have an energy-band diagram similar to that of semiconductors, with a valence band separated by an energy gap from the conduction band. When an X-ray photon is absorbed energy is deposited in the crystal and electrons in the valence band can be excited to the conduction band, leaving holes in the valence band. When an excited electron falls back to the valence band and recombines with a hole the excess energy is dissipated by

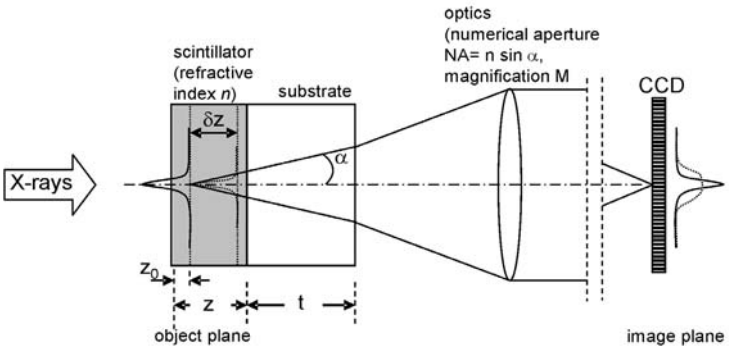


FIG. 10.2. X-ray imaging with scintillators and lens coupling. Identical visible-light images are created by the X-ray beam in different planes of the scintillator. An image in plane z_0 is focused onto the CCD. An image in plane $Z_0 + \delta z$ is out of focus at the CCD.

emission of photons with an energy equal to the width of the bandgap. For intrinsic scintillators the efficiency of the de-excitation process is, however, low and the bandgaps are, in general, too large to emit light in the visible range that is needed for detection by common CCDs. To overcome this limitation scintillators are doped with typical fluorescent ions such as europium (Eu), terbium (Tb) or cerium (Ce). Inorganic scintillators doped with rare-earth elements are thus suitable because they exhibit a relatively large bandgap, good light yield, fast decay time and a minimal self-absorption. Doping with these *activators* makes the scintillation process more efficient and creates energy levels inside the energy gap producing scintillation light in the visible region (see Fig. 10.1). These *doped scintillators* are the ones used to construct the 2D detectors for tomography at synchrotron beamlines.

10.2 Spatial resolution and detective quantum efficiency

Two important characteristics of an imaging system, in particular for tomography, are the *spatial resolution* and the *detective quantum efficiency* (DQE). We will first discuss the various parameters that determine the spatial resolution, secondly we will analyse the contributions of the various components to the DQE. It will be shown that these are often two competing requirements.

A schematic layout of the most commonly used detector principle for tomography at synchrotron beamlines is given in Fig. 10.2. In this case, the converter screen (or scintillator) is mounted on a substrate, which will be explained in more detail in Section 10.4. In such a system the spatial resolution is determined by the X-ray interaction in the scintillator, by the X-ray interaction in the substrate, and by the performance of the optics. We will first discuss the X-ray interaction in the scintillator and the substrate.

As pointed out in Chapter 4 there are various kinds of X-ray interactions with solids, namely the photoelectric effect, (elastic) Rayleigh scattering and (inelastic) Compton scattering. For high energies above 150 keV and a yttrium aluminium garnet (YAG) crystal the Compton effect becomes dominant. The scattered photon, the energy and direction of which are governed by the Compton scattering distribution, can escape, be scattered again or be absorbed. The resulting electron range (the distance at which half of the particles have been absorbed) which is very short for low-energy photoelectrons becomes an appreciable fraction of the scintillator thickness at higher Compton electron energies. The assumption of total electron energy absorption at the interaction site is less accurate. An increasing part of the electron path takes place outside the interaction volume, resulting in the loss of energy absorbed in the 'phosphor' and/or a decrease in the spatial resolution. As an example, the average path length travelled by an electron in a YAG crystal is $\approx 2.8\,\mu m$ at 20 keV and $\approx 44\,\mu m$ at 100 keV (National Institute of Standards and Technology, 2006). This shows that the effect can be an important source of degradation whenever the Compton effect is dominant. Compton scattering is negligible as a source of background noise, however, in imaging applications for X-ray energies below 100 keV (Hoheisel *et al.*, 2004).

The predominant physical process that reduces the spatial resolution at low X-ray energies is the fluorescence generated by photoelectric absorption. Monte Carlo simulations of photon and electron transport have been made for a YAG crystal (Koch *et al.*, 1998). The *Integral Tiger Series* (ITS) code uses a Monte Carlo simulation of photon and electron transport in matter to derive the radial absorbed dose distributions. The calculation used a narrow and parallel pencil beam of 14-keV X-rays on a 5-μm thick YAG:Ce scintillator supported by a 100-μm thick undoped YAG substrate. The simulation, which did not take into account the scattering of the emitted visible light, shows a very high absorbed dose in 100 nm full width at half-maximum (FWHM) due to the short path of the Auger electrons. The energy deposited outside these 100 nm is due to secondary electrons and characteristic fluorescence X-rays. On the other hand, calculations at 30 and 100 keV have shown a point-spread function (PSF) tail of the radial energy distribution that degrades the spatial resolution. The small FWHM of the PSF at low energies indicates that the spatial resolution is in this case not limited by the scintillator but by the resolving power of the optics, i.e. in particular of the microscope objective, which will be discussed later.

Another source of tails in the PSF is the undesired luminescence from the substrate, e.g. by Ce impurities in undoped YAG. Light emission from substrates has been measured (Martin *et al.*, 2005) and corresponds in the worst case to 20% of the luminescence of the YAG:Ce epitaxial layer. This becomes even more important as thinner layers of scintillator are used. With the following light yields[6]

[6]Number of emitted visible photons emitted in 4π steradian per unit of absorbed X-ray energy.

measured by Martin *et al.* (2005) – 11 ph/keV for LAG:Eu and 1.2 ph/keV for undoped YAG (which corresponds to 6% of the typical value of doped YAG:Ce) – a 170-μm undoped YAG substrate emits more light than a 5-μm LAG:Eu layer for X-ray energies above 20 keV. This means that at higher energies an undoped YAG substrate does not allow the use of very thin LAG:Eu layers. Therefore, LAG and YAG scintillators layers grown by liquid phase epitaxy on a YAG substrate are not ideal for high spatial resolution at high energies. In the case of LAG:Eu layers on a YAG substrate one could use colour filters to suppress about half of the substrate's luminescence signal. This solution is not possible for the YAG:Ce layer because the YAG substrate luminescence wavelength is too close to the YAG:Ce layer emission line. Other solutions are to use purer YAG substrates or some other deposition techniques such as the sol-gel method, which permits a larger choice of substrate materials.

Another component that determines spatial resolution is the optical coupling between the converter screen and the CCD. The CCD camera is focused onto the object plane inside the scintillator located at the distance z_0 from the surface (see Fig. 10.2). All other planes in front of and behind the object plane are out of the focus but nevertheless contribute to the total light projected onto the CCD. The image resolution is determined by the defect of focus (δz) of the image distribution outside the object plane. Other degradations of the image are due to diffraction and spherical aberrations arising from the thickness of the scintillator (z) and the substrate (t). The effects affecting spatial resolution are the following:

- defect of focus: $R \approx \delta z \cdot \mathrm{NA}$,
- diffraction: $R \approx \lambda/\mathrm{NA}$,
- spherical aberration: $R \approx t \cdot \mathrm{NA}^3$,

where δz is the defect of focus, λ is the wavelength of light and NA is the numerical aperture. The spatial resolution R determined by the first two effects can be calculated by numerical simulations (Koch *et al.*, 1998) and can be fitted by:

$$R_{\mathrm{fw50\%int}} = \sqrt{(\frac{p}{\mathrm{NA}})^2 + (q \cdot z \cdot \mathrm{NA})^2},$$

where R and z are in micrometres. The coefficients $p = 0.18$ and $q = 0.075$ (Koch *et al.*, 1998) are calculated for a 50% integrated line-spread function (LSF). Here, the first term is due to diffraction and the second to the defect of focus. The spatial resolution is given in Fig. 10.3 as a function of numerical aperture for different screen thicknesses.

The other parameter, besides spatial resolution, which is important for imaging systems is the detective quantum efficiency, DQE. This parameter relates the output to the input signal-to-noise ratio:

$$\mathrm{DQE} \equiv \frac{\mathrm{SNR}_{\mathrm{out}}^2}{\mathrm{SNR}_{\mathrm{in}}^2} \approx \eta_{\mathrm{abs}} \cdot \left(1 + \frac{1 + \frac{1}{\eta_{\nu/e}}}{\eta_{\mathrm{coll}} \cdot \left(\frac{E_X}{E_\nu}\right) \cdot \eta_{x/\nu}}\right)^{-1}, \qquad (10.1)$$

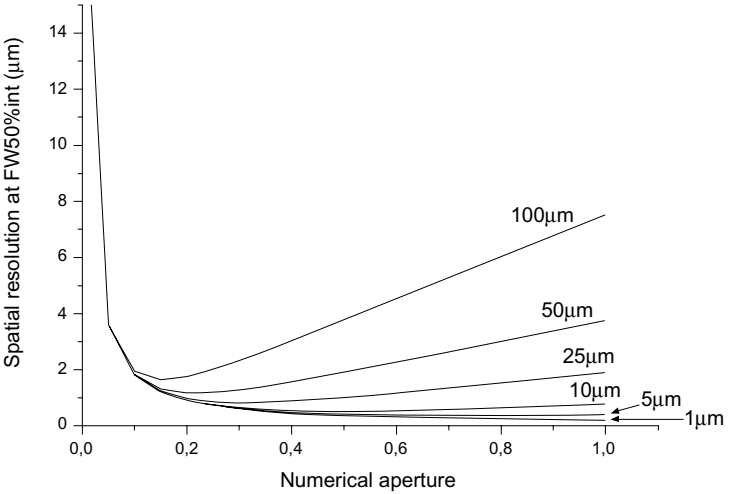


FIG. 10.3. Spatial resolution versus numerical aperture (NA) of an optical system for different thicknesses of the single-crystal scintillator. The parameters used for the YAG:Ce scintillator are: refractive index $n = 1.95$ and X-ray wavelength λ=550 nm. One way to increase the spatial resolution is to reduce the thickness, which in turn reduces the absorption efficiency.

TABLE 10.1. Overview of commercially available CCD cameras.

CCD manufacturer	pixel size	array size	URL
Apogee	9 µm	1536 × 1024	`www.ccd.com`
Frelon	14 µm	2048 × 2048	`www.esrf.fr`
PCO	7.4 µm	2048 × 2048	`www.pco.de`
	7.4 µm	1600 × 1200	
Roper	6.8 µm	1317 × 1035	`www.roperscientific.com`
	6.45 µm	1392 × 1040	
Photonic	6.45 µm	1392 × 1040	`www.photonic-science.co.uk`
ATMEL	14 µm	2048 × 2048	`www.atmel.com`
Pixel Vision	14 µm	2048 × 2048	`www.pvinc.com`
Dalsa	12 µm	1024 × 1024	`www.dalsa.com`
	14 µm	1024 × 1024	

where $\mathrm{SNR}_{\mathrm{out}}$ and $\mathrm{SNR}_{\mathrm{in}}$ are the input and output signal-to-noise ratios of the detector, η_{abs} is the absorption efficiency for X-rays in the scintillator, $\eta_{\nu/e}$ is the quantum efficiency of the CCD, η_{coll} is the collection efficiency of light by an objective with numerical aperture NA in combination with a scintillator with a refractive index n, E_X is the X-ray energy, E_ν is the photon energy of the visible-light photons and $\eta_{x/\nu}$ is the conversion efficiency of X-rays to visible-

light photons (light yield), meaning the fraction of the incident X-ray energy that appears as scintillation light. The DQE, ranging between 0 and 1, is reduced if the imaging system degrades the spatial resolution, if the detector adds noise to the final image, or if not all incoming X-rays are absorbed. In order to improve the absorption of thin scintillators they are often used with X-rays above the K-absorption edge (17.04 keV for YAG:Ce, 50.24 keV for GGG and 63.3 keV for LAG:Eu). While this improves the absorption efficiency, it also increases X-ray fluorescence, which leads to a loss of spatial resolution as explained above. For systems with sufficiently high NA the DQE is mainly determined by the absorption efficiency. As shown in Tab. 10.2, the quantum efficiency of the CCD camera and the type of scintillator have an impact on the DQE at low NA. For example, GGG:Eu and LAG:Eu layers have the same absorption efficiency but different light yields, and we note a DQE improvement for NA=0.3 of 20% and 23% with the FReLoN and Dalsa, respectively. Table 10.1 lists these two and other CCD cameras and specifies their resolution. At higher numerical aperture, the improvement is less significant: 8% and 12.5% for FReLoN and Dalsa cameras, respectively. A promising new camera for fast microtomography is the new fast Dalsa 1M60P that uses a FTT1010M CCD chip and is more sensitive than the 1M60 type. With a LAG:Eu converter layer and a NA=0.3 it has a DQE of 0.027, which is an improvement of 58% over the DALSA 1M60. As stated before, at high numerical apertures the DQE is mainly determined by the absorption efficiency of the converter screen and less by the choice of the CCD camera. For example, although the KAF-4320E Kodak CCD chip has a high quantum efficiency (70% at 580 nm) that is 150% more sensitive than the ATMEL (see Tab. 10.1) CCD chip, the improvement of the DQE at NA=0.7 is weak, 0.09 instead of 0.076 with LAG:Eu layer.

10.3 Powder screens

In the previous section we analysed the characteristics of the detector that determine the overall performance, in particular the spatial resolution and the detective quantum efficiency. It was shown that the converter or scintillator screen is of crucial importance for high-resolution tomography. Converter screens can be classified into two categories: powder converter screens and crystal converter screens. In this section we will discuss powder screens, in the next crystal screens.

In powder converter screens a small-grained 'phosphor' powder is mixed with a binding agent and is processed to a film. The two main powder screens used for X-ray detection at synchrotrons are Gd_2O_2S:Tb, called *P43* or *Gadox* and $Y_3Al_5O_{12}$:Ce called *P46*. Although the packing density of the powders is half that of the solid material Gadox is, nevertheless, very attractive for its high density and its X-ray to visible-light yield (≈60 photons/keV). A disadvantage of P43 (Gadox) is its long light decay time of a few hundred microseconds, which severely limits its application in fast imaging applications. In this case, P46 is preferred for its short decay time. Screens based on powder phosphors are commercially available down to a grain size of approximately 1 µm. Their resolution in terms

TABLE 10.2. Characteristics of some scintillator materials. The absorption efficiency and DQE are calculated at 20 keV with a 5-μm thick scintillator layer using eqn 10.1. The resolution is deduced from experiment. Cloetens *et al.* (1999) has measured a width of the LSF of 2 μm with NA=0.3 and 25 μm YAG:Ce. A spatial resolution of 0.8 μm has been measured by Wang *et al.* (2001) with a Nikon plane achromat objective with 40× magnification, NA=0.6 and 5 μm YAG:Ce layer. Stampanoni *et al.* (2002) have measured a spatial resolution of 1.04 μm with a 1.8-μm thick YAG:Ce scintillator and 20×, NA=0.7 Olympus optic.

optic	scintillator	absorption	DQE FReLoN	DQE Dalsa	spatial resolution	Ref.
10×	YAG:Ce	0.07	0.021	0.013	2 μm	(1)
NA=0.3	LAG:Eu	0.11	0.031	0.017		
	GGG:Eu	0.11	0.037	0.021		
20×	YAG:Ce	0.07	0.049	0.038	1.04 μm	(2)
NA=0.7	LAG:Eu	0.11	0.076	0.056		
	GGG:Eu	0.11	0.082	0.063		

(1) Cloetens *et al.* 1999
(2) Stampanoni *et al.* 2002

of FWHM of the line-spread function is, however, approximately equal to their thickness (Swank, 1973).

Homogeneous screens of 2 to 3 μm thickness and similar resolution can be deposited. Smaller grain sizes are manufactured with 100 nm diameter but these screens show powder agglomerates as well as a poor X-ray to light conversion yield (Koch *et al.*, 1999*b*).

10.4 Crystal converter screens

In the previous section it was mentioned that the spatial resolution of powder converter screens is equal to the thickness of the screen, giving an unfavourable compromise between X-ray stopping power and spatial resolution. To overcome this limitation, thin single-crystal scintillators and single-crystal films (SCF) have been developed and are now commercially available in a thickness ranging from 1 μm to 50 μm. Due to the better lateral spatial resolution compared to powder screens, thin-film scintillators have become the main converter screen technology for high-resolution X-ray imaging. There are many methods for deposition of luminescent films, including pulsed laser deposition, sputtering, liquid phase epitaxy and the sol-gel process. The latter technique is suitable for the deposition of amorphous and crystalline thin films through dip- and spin-coating. However, a detailed treatment of the various thin-film growth techniques is outside the scope of this book. In Section 10.4.1 we will discuss the essential properties of crystal converter screens. In Section 10.4.2 the most commonly used materials will be discussed.

10.4.1 *Essential properties of crystal converter screens*

A number of properties of crystal converter screens are of essential importance for high-resolution X-ray imaging, namely X-ray absorption, visible-light emission, optical and engineering properties. We will discuss them separately.

- X-ray absorption can be maximized by a combination of high-density material ($>5\,\mathrm{g/cm^3}$) with a large atomic number (>50). The X-ray absorption efficiency in a scintillator is proportional to ρZ^n with ρ the density, Z the effective atomic number and n varying between 4 (low energies $<500\,\mathrm{keV}$) and 5 (high energy), see Section 4.2.3.
- The visible-light emission has various aspects. First, one wants a high light yield (>15 visible photons/keV). The emission efficiency is defined as the ratio of the output light energy to absorbed radiation energy. Secondly, the emission wavelength has to be well matched to the peak sensitivity of the CCD chip used, which is generally at wavelengths between 550 and 650 nm. In most cases the emission peak wavelength is determined by the activator material and one thus has some flexibility by using different doping materials. Thirdly, one desires rapid decay and short afterglow (2 ms) of the crystal converter screen. Decay and afterglow produce the same phenomenon – image lag – but have a different origin. Decay reflects the time that it takes for the excited energy states to be de-excited. It is determined by the scintillation material and the type of activator. Afterglow is due to another mechanism. When electrons are excited by radiation, some excited electrons are activated to an energy level from which the direct transition to the valence band is forbidden. These electrons have to be further excited to a higher energy state before they can return to the valence band. Such a mechanism can cause severe lag in the light emission. Afterglow can be caused by the material, by imperfect processing, unintentional doping, *etc.* Both decay and afterglow are harmful to synchrotron tomography applications in cases where high frame rates are essential and scintillator afterglow limits the useful dynamic range of the detector. Furthermore, the light output should be linear with the incident X-ray flux.
- The optical properties required are a high transmittance of the emitted visible light with no scattering.
- Finally, the crystal converter screen should have a good mechanical strength, preferably be non-toxic, be easy to machine and should be radiation hard, meaning no degradation caused by X-ray irradiation.

An overview of the key parameters for different crystal converter screens is given in Tab. 10.3.

10.4.2 *Bulk converter screens*

Crystal converter screens are either used in bulk crystalline form or in thin-film-on-substrate form. We will first describe the most important bulk crystals used, followed by the most suitable thin-film converter screens.

TABLE 10.3. Characteristic data of scintillators used for X-ray imaging.

material	name	density (g/cm^3)	Z_{eff}	λ (nm)	LY (ph/keV)	refractive index	type
Gd_2O_2S:Tb	P43	7.3	59.5	545	60		powder
$Y_3Al_5O_{12}$:Ce	P46	4.55	32	530			powder
CsI(Tl)		4.53	54	550	65	1.8	crystal
$Bi_4Ge_3O_{12}$	BGO	7.13	75	480	8.2	2.15	crystal
$CdWO_4$		7.90	64	530	15	2.25	crystal
$Lu_3Al_5O_{12}$	LAG:Ce	6.73	63	550			crystal
$Y_3Al_5O_{12}$:Ce	YAG:Ce	4.55	32	550	40	1.82	crystal
Lu_2SiO_5:Ce	LSO:Ce	7.4	66	420	25	1.82	crystal
$Y_3Al_5O_{12}$:Ce	YAG:Ce	4.55	32	550	20	1.82	SCF
$Lu_3Al_5O_{12}$:Eu	LAG:Eu	6.73	63	595	11		SCF
$Gd_3Ga_5O_{12}$:Eu	GGG:Eu	7.1	53	595	44	1.96	SCF
Lu_2O_3:Eu^{3+}		8.4	68.8	611	20	1.88	polycrystal
Gd_2O_3:Eu^{3+}		7.1	61	611	19	1.82	polycrystal

Application of crystal converter screens in X-ray medical imaging has resulted in rapid development of high-density bulk single-crystal scintillators. Polishing and classical cut technologies are used to prepare single-crystal scintillators for X-ray tomography applications.

10.4.2.1 *Bismuth germanate* One system is bismuth germanate ($Bi_4Ge_3O_{12}$, or commonly abbreviated as *BGO*). BGO has a high density and a large atomic number. Unfortunately, the poor light yield, difficulties in obtaining a smooth surface as well as its high refractive index are disadvantages for high-resolution imaging (Tseng *et al.*, 2000). Therefore, BGO seems to be more suited for high gamma energy counting probes than for X-ray tomography.

10.4.2.2 *Cadmium tungstate* Another system is cadmium tungstate ($CdWO_4$, sometimes abbreviated as *CWO*) which is a suitable material for various applications due to its high X-ray stopping power, its high light yield with little afterglow, emission near 470 nm well matched to the sensitivity of CCD detectors and the fact that it is non-hygroscopic. However, machining and thinning of $CdWO_4$ single-crystals is not easy due to its intrinsic (010) cleavage plane. As a result it is not possible to find commercial $CdWO_4$ scintillators much thinner than 50 μm. Lee *et al.* (1997) have used a 60-μm thick $CdWO_4$ single crystal and optics with a numerical aperture of 0.5. The spatial resolution obtained was about 1.5 μm at 20 keV. Jung *et al.* (2002) have used a $CdWO_4$ single-crystal scintillator cleaved to a thickness less than 100 μm in a pink beam (non-monochromatic X-ray beam with 6 to 30 keV energy). The spatial resolution was 2.9 μm.

To obtain a better spatial resolution, the scintillator must be thinner. A sol-gel processing technique has been applied to produce a thin film of cadmium tungstate (Lennstrom *et al.*, 2003). The resulting film consists of crystal grains of approximately 1 μm in diameter, which is a limitation for high-resolution imaging.

10.4.2.3 *Lutetium oxyorthosilicate* A promising bulk-crystal scintillator is cerium-doped, high-density, high-Z lutetium oxyorthosilicate (LSO). Although the absorption efficiency of this scintillator is lower than that of BGO it appears to have a better light yield than BGO and a shorter decay time than Eu-doped scintillators. LSO is available from CTI/USA[7]. Hamamatsu[8] used this scintillator in a high-resolution X-ray imaging camera installed at SPring8. The single-crystal scintillator is 10 μm thick and is glued to an amorphous carbon plate. Uesugi *et al.* (2001) have achieved 1 μm spatial resolution. Spatial resolution is limited by the 10 μm thickness of the crystal, which cannot be reduced any further by mechanical polishing.

10.4.3 *Composed converter screens*

In order to get even thinner converter screens one has to deposit a single-crystal or a polycrystalline film on a substrate. There are various different methods to grow thin films on a substrate, and the growth mechanism can have a substantial effect on the performance of the thin-film converter screens. However, this subject is outside the scope of this book and we will only treat the physical properties of the most commonly used films.

10.4.3.1 *YAG:Ce on YAG* Yttrium aluminium garnet (YAG:Ce) has promising scintillation properties when doped with cerium. The spectrum is unusual for a cerium-doped scintillator and the peak in the emission spectrum occurs at 550 nm. This longer wavelength is a good match to the quantum efficiency of CCD cameras.

10.4.3.2 *LAG:Eu on YAG* Lutetium aluminium garnet (LAG:Eu) has an absorption efficiency 8 times higher at 15 keV and 2 times higher at 40 keV than YAG:Ce. However, the growth process for LAG on undoped YAG needs to accommodate the lattice mismatch between LAG and YAG and can be achieved by partial substitution of aluminium by scandium in LAG. Unfortunately, LAG doped by Eu and Sc has a considerably lower light yield compared to an epitaxial layer of YAG:Ce. The principal cause of this smaller light yield is the loss of excitation energy on exciting the luminescent centers formed by Sc (Martin *et al.*, 2005). Even though YAG:Ce has a faster decay than LAG:Eu, which is of importance for fast imaging applications, LAG:Eu films are preferred since they have a higher X-ray stopping power and a lower light output, resulting in a net gain of a factor 4 for a given exposure time above the absorption edge of lutetium (63 keV). Opportunities for improving X-ray detectors in terms of spatial resolution are somewhat limited as they are already near the theoretical limits of 0.4 μm. The best performance measured is 0.5 μm FWHM obtained with 1-μm LAG:Eu ($Lu_3Al_5O_{12}$:Eu) thick film and NA=0.95 (Koch *et al.*, 1999*a*). The only

[7] www.siemens.com.

[8] www.hamamatsu.com.

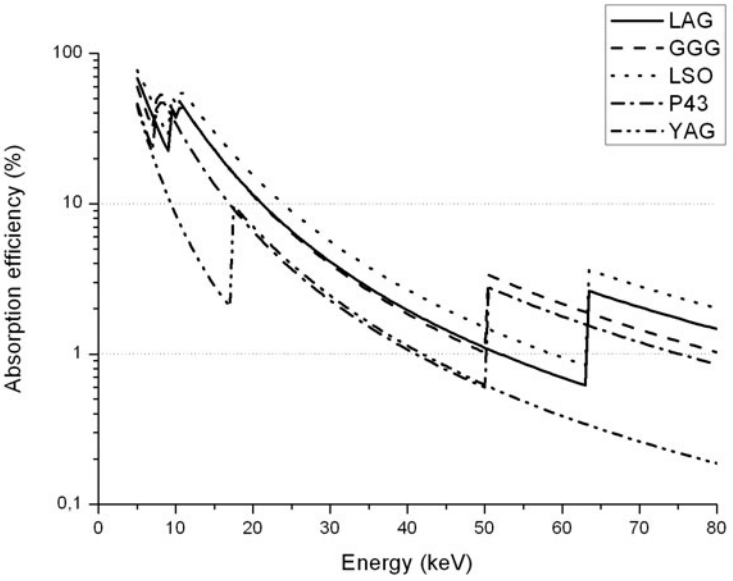


FIG. 10.4. Absorption efficiency of more current scintillators used for X-ray imaging on synchrotron radiation sources for 5-μm thick layers: $Lu_3Al_5O_{12}$ (LAG), $Gd_3Ga_5O_{12}$ (GGG), Lu_2SiO_5(LSO), $Y_3Al_5O_{12}$ (YAG) and Gd_2O_2S:Tb (P43).

possible improvements are in X-ray stopping power, light yield (and thus DQE) and decay time.

10.4.3.3 *GGG:Eu* Gadolinium gallium garnet (GGG:Eu) shows excellent scintillator properties compared to YAG:Ce and LAG:Eu. Its absorption efficiency is comparable to LAG:Eu material (see Fig. 10.4) but its light yield is better than YAG:Ce and more than 2 times better than LAG:Eu. In addition, the afterglow is fast compared to LAG:Eu and allows resolution of a dynamic range between 16 and 17 bit in a successive image mode (Martin *et al.*, 2005) which is a significant improvement over the 10-bit dynamic range obtained with a YAG:Ce layer (Koch *et al.*, 1999*b*).

10.4.4 *Polycrystalline scintillators*

Both lutetium oxide (Lu_2O_3:Eu^{3+}) and gadolinium oxide (Gd_2O_3:Eu^{3+}) doped with europium represent a dense version of Y_2O_3:Eu^{3+} phosphor. Lu_2O_3:Eu^{3+} has an absorption efficiency that is ≈2 times higher at 15 keV, a light yield ≈2 times higher than LAG:Eu and its emission line shows a strong peak at 610 nm, which is well adapted to a CCD chip. Lu_2O_3:Eu^{3+} and Gd_2O_3:Eu^{3+} have been prepared by sol-gel processing. This technique is suitable for the deposition of amorphous and crystalline thin films through dip- or spin-coating. Researchers at

the University of Lyons (Garcia-Murillo *et al.*, 2003) have studied dip-coating for X-ray imaging applications. Multicoating (up to 50 coatings) and heat treatments are required to obtain 800-nm thick films. The layers are finally annealed for 1 h at 1000 °C for high densification and crystallization of the materials. A definite advantage of the sol-gel method as compared to liquid phase epitaxy is a larger choice of substrate material for crystal growth. Deposition on silicon and silica has been realized by Garcia-Murillo *et al.* (2002). However, this technique is limited in terms of thickness (1 μm) due to the mechanical stresses between layers. Dujardin *et al.* (2005) have also deposited the same material by pulsed laser deposition (PLD) and they have shown the possibility of depositing a thicker layer by PLD, but further development is necessary to obtain a homogeneous layer.

All the described screens have been used in tomographic applications at synchrotron beamlines. The commercially available inorganic scintillators, YAG:Ce, $CdWO_4$ and LAG:Eu were successfully used by Stampanoni *et al.* (2002), Carlo *et al.* (2001), Koch *et al.* (1998) and Beckmann (2001). However, for imaging experiments with higher X-ray energies LSO:Ce was preferred by Yagi *et al.* (2004*a*). Finally, GGG:Eu seems to be preferable over YAG:Ce and LAG:Eu for various reasons: the light yield of LAG:Eu is low and the undoped YAG substrates for the YAG:Ce and LAG:Eu layers emit undesired luminescence that degrades the spatial resolution (Martin *et al.*, 2005).

10.5 Optical coupling

In Sections 10.3 and 10.4 we discussed the screens that convert the incident X-rays to visible-light photons. In this section, we will discuss the optical coupling between these converter screens and the photodetector. Two ways to achieve this optical coupling will be discussed: lens coupling and fibre-optic (FO) coupling. For the latter, a taper contains millions of individual fibres (core + cladding) that transmit an element of the image created on the powder screen. Fibres are clad to guide the light by internal reflection. We will start with a comparison of the efficiency between lens and FO coupling for various magnifications and demagnifications. After that, a number of important issues related to lens coupling are treated.

Lens coupling is the simplest and most convenient way and works well for systems with a high spatial resolution. For powder converter screens (Section 10.3) the emitted light is that of a so-called Lambertian source, see Fig. 10.11(a), and the transmission efficiency through the lens coupling is given by (Liu *et al.*, 1994):

$$\eta = \frac{T_L M^2}{M^2 + 4f^2(1+M)^2},$$

where f is the ratio of the focal length to the effective lens diameter, T_L is the transmission factor of the lens and M is the magnification ratio, i.e.

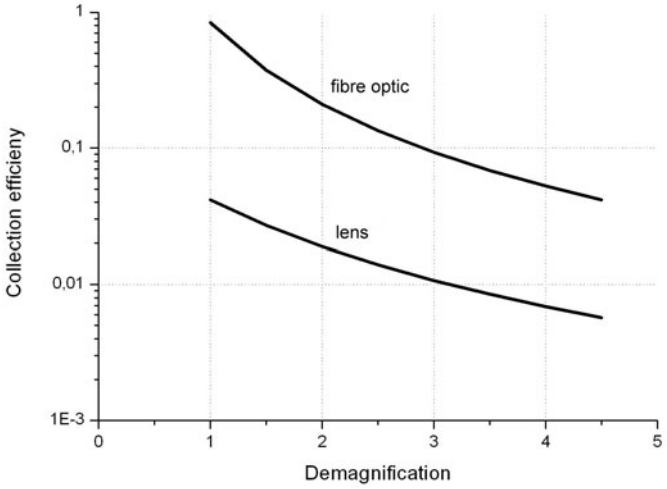


FIG. 10.5. Effect of magnification factor on the collection efficiency for lens coupling.

$$M = \frac{\text{image size}}{\text{object size}}.$$

For a crystal converter screen (Section 10.4) the source is non-Lambertian, see Fig. 10.11(b), and the transmission efficiency is:

$$\eta = \frac{T_L M^2}{16 n^2 f^2 (1+M)^2},$$

where n is the optical refractive index of the crystal converter screen. Collection efficiencies (Liu *et al.*, 1994) of lens-coupled systems are plotted in Fig. 10.5. FO coupling on the other hand is preferred when large fields of view are needed. In this case, powder converter screens are mostly used and the transmission of a the tapered fibre optic can be expressed by (Hejazi and Trauernicht, 1997):

$$\eta = \left(\frac{1}{m}\right)^2 \left(\frac{(n_2^2 - n_3^2)^{1/2}}{n_1}\right)^2 T_F (1 - L_R) F_c,$$

where m is the demagnification factor, n_2 is the core index, n_3 is the cladding index, n_1 is the external index, T_F is the transmission of the fibre core, L_R is the loss at the surface due to Fresnel reflection and F_c is the fill factor of the fibre core.

In Tab. 10.4 the collection efficiency for both lens coupling and fibre-optic coupling are shown for comparison. Lens coupling is more efficient than FO coupling when the object is magnified because in that case the solid angle is large enough to collect the light coming from the luminescent screen. This is the

TABLE 10.4. Efficiency of f/1.2 lens and fibre-optic coupling at various magnifications calculated for a Lambertian source, i.e. for a powder screen. Calculated with $T_L = 1$ and $T_F = 1$.

magnification	lens coupling efficiency %	fibre-optic efficiency %
20×	18	-
10×	17	-
4×	13.7	-
1×	4.1	84
0.5×	1.9	21
0.25×	0.7	6

basis of normal microscopy and it is also used for synchrotron microtomography. The situation is different for demagnification, where the efficiency of lens coupling drops with increasing demagnification. This is because luminescent screens give off light distributed over an angle of 180° (half-space, Lambert distribution), and lenses transmit only a relatively small fraction of the emitted light. In this case, fibre-optic coupling is more efficient. As an example, a non-magnifying (1:1) optic of the relative aperture f/1.2 has an efficiency of 4.1%, while a high-quality non-magnifying FO shows an efficiency of up to 84%. Fibre-optic tapers can thus boost the light collected by a CCD by 20 times compared to a f/1.2 lens and by 14 times compared to a f/1.0 lens.

10.5.1 *Lens coupling: finite-focused versus infinity-focused systems*

Lens coupling is the preferred optical relay for the magnifying used in microtomography. For high-resolution detectors based on microscope objectives one can adopt two strategies depending on whether one uses finite-focused or infinity-focused optical systems. A finite optical system, or *fixed tube length*, is an optical system in which the image is formed by the objective alone; the distance between the microscope objective shoulder and the image seat in the photo-eyepiece is fixed. This distance is named the *tube length of the microscope* and the tube length is standardized to 160 mm. The situation is different for the infinity-corrected optical systems, where the image is formed by a combination of objective and tube lenses with the microscope objective producing a flux of parallel visible light, imaged at infinity. The tube lens forms an intermediate image in the tube. The magnification produced by an infinity-corrected system is calculated by the ratio of objective focal length and tube lens focal length. Finite optical microscopes are less and less commercially available. Spindler and Hoyer[9] sells refractive objectives and Ealing[10] offers reflecting objectives, both with a finite tube length of 160 mm. The situation is different for infinity-corrected optical systems that are sold by different companies that do not use the same tube

[9]Linos, Koenigsalle 23, D-37081 Goettingen; www.linos.com.

[10]Ealing, 3845 Atherton Road, Suite #1, Rocklin, CA 95765; www.ealingcatalog.com.

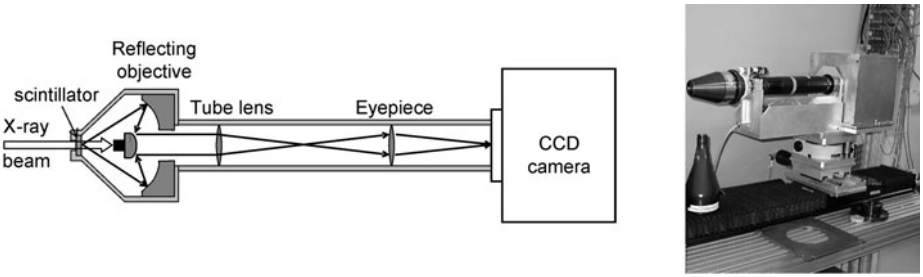


FIG. 10.6. Inline detector using reflective optics. Courtesy: ESRF, Instrument Support Group.

lens focal length. Objectives designed for infinity are usually not interchangeable with finite optical tube length and vice versa because the infinity lenses suffer from enhanced spherical aberration when used on a finite system. It is possible to use finite objectives with an infinity-corrected microscope but the magnification will be decreased and the spatial resolution degraded. Equally, the mixing of objectives and tube lenses from different manufacturers will result in a change of magnification and a loss of resolution.

10.5.1.1 *Lens coupling: radiation damage* Recent research trends at synchrotrons show a move to use fast microtomography at increasingly higher X-ray energies, which poses a severe risk for the optical elements used in the microscope. At high energies, most of the X-ray photons will be transmitted by the converter screen and impinge on the first lens (see Fig. 10.2). Experiments at 65 keV, have resulted in severe radiation damage of a refractive objective already after 70 s. Therefore, a new design based on reflected light has been considered to avoid darkening of lenses by the X-rays (Ham *et al.*, 2002). Two ways are possible.

The first uses a mirror between the scintillator and the microscope objective that reflects the visible light emitted by the luminescent screen at 90° to the X-ray beam and both prevents radiation damage of the optical components and facilitates shielding of the CCD camera. The mirror has to be placed sufficiently far away from the scintillator screen to avoid backscattered X-rays falling onto the screen. This setup requires a high mechanical precision of the mirror support that must be fixed at 45° to preserve a constant distance between the plane of the first lens and the plane object. An angular misalignment of the mirror will be a source of resolution loss as a result of a defocalization. The plane mirror is therefore only conceivable with low magnification.

The second solution to avoid radiation damage and consequent darkening of the optics is to use reflective optics (Di Michiel *et al.*, 2005). Reflecting objectives for microscopes are generally simpler than refractive objectives since they consist of only two mirrors in the Schwarzschild configuration. The system is depicted in Fig. 10.6. It can be understood as an inverse Cassegrain system, such as a telescope. The system consists of a concave primary mirror and a con-

vex secondary mirror. Commercial reflecting objectives have a magnification not lower than 15×. The present application requires a magnification of 5×. Therefore, a specific development was made by Nachet[11] and the ESRF to produce a 5×, NA=0.2 objective. The radius of the primary mirror is 135 mm, the radius of the secondary mirror is 50 mm and the two mirrors are separated by a distance of 85 mm. The long working distance of 90 mm has allowed a beam stopper to be fixed in front of the secondary mirror and thus protect the CCD and photo-eyepiece against radiation damage. In this design, the ESRF uses the infinity-corrected system. This solution has allowed to be maintained compatible with Olympus optics, and intermediate optical components can also be accommodated. No additional optics are needed to correct the image in the 'infinity space' between the objective and the tube lens. Advantages compared to refractive optics are: no scattering from the 45° plane mirror in the folded detector system, and no damage to the optical components due to the protection by the beam stop. The disadvantages are the dimensions (96 mm diameter × 140 mm length), weight and cost.

10.5.2 *Fibre-optical coupling*

Fibre-optical coupling is largely preferred for demagnifying systems when a large FOV is needed. However, FO coupling has its own technical challenges. The tapered fibre optic must be attached to the CCD sensor with extreme precision. The optical coupling is usually a permanent bond with a thin layer of gels or epoxy optical cement that must withstand the heat cycle between room temperature and the CCD operation range of −40°C to −20°C. In addition to the hazardous operation of optical coupling, the general uniformity is degraded by spot blemishes (burned or broken fibres) (Ponchut, 2006) and line blemishes, called chickenwire (Gruner *et al.*, 2002). Geometrical distortions owing to gaps and irregularities as an example of a lateral displacement of straight line for a multitaper CCD system (shear distortion at the taper's edge) are an inherent disavantage of a FO. The usable demagnification ratio is limited to less than 3 because the transmission efficiency decreases with increasing demagnification (see Fig. 10.7). Applications that require the largest FOV consist of a modular array of CCDs each with a tapered fibre optic. Figure 10.8 shows the CCD camera and its taper used on the medical beamline at the ESRF.

10.6 Readout based on CCD cameras

In the previous sections we have discussed the converter screens that convert X-ray photons to visible-light photons and the optical coupling that relays these optical photons to the photodetector. By far the most commonly used photodetector for tomography at synchrotron beamlines is the charge-coupled device (CCD) detector. In this section we will discuss some of the aspects of the CCD camera that are of importance to tomography experiments. Whereas spatial resolution and DQE are determined mainly by the converter screen and the optical

[11] NACHET, 7 rue Ernst Chaput, 21059 DIJON, France; www.nachet.com.

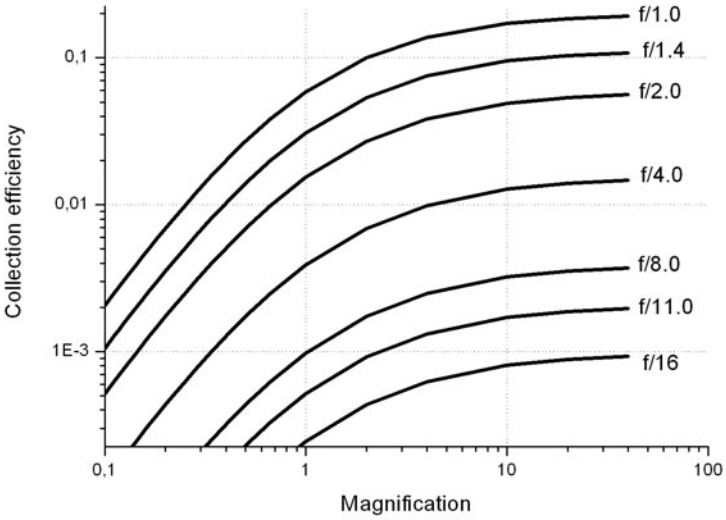


FIG. 10.7. Effect of demagnification factor on the collection efficiency. $n_1 = 1$, $n_2 = 1.8$, $n_3 = 1.5$, $T_F = 1.0$, $F_c = 0.85$, $L_R = 0.8$, $T_L = 1.0$ and $f = 1.2$.

coupling used, the frame rate and the dynamic range are mainly determined by the performance of the analogue-to-digital converter inside the CCD camera. Various sizes of CCD chips are commercially available, ranging from 1k×1k to 4k×4k. At low frame rates, e.g. 1 frame per second (fps), these cameras allow for very large dynamic range images to be collected. High frame rates (25 fps) needed for fast and time-resolved tomography experiments are possible at the expense of dynamic range (10 to 12 bits). Figure 10.9 shows the present situation of scientific cameras used for tomography at synchrotrons.

10.6.1 *Categories of CCD cameras*

CCD cameras used for imaging experiments at synchrotrons can be roughly divided into three categories. The first category is the *slow-scan camera*, which approaches 16 true bit depth. This large dynamic range is achieved by slow and accurate analogue-to-digital converters in order to reduce readout noise combined with cooling of the CCD chip, which reduces the dark current (see below). Because of the relatively long readout times these systems are suitable for experiments with long exposure times (Wang *et al.*, 2001). The second category are the *fast 12*-bit *cameras*. A large range of 12-bit CCD cameras with different pixel resolutions is commercially available, and they are used for many different applications (Jung *et al.*, 2002). These cameras offer 30 to 60 fps and are used for fast tomography applications whenever speed is more important than dynamic range. The third category, which is the most popular for tomography applications at synchrotrons, is *intermediate* between fast and slow cameras and

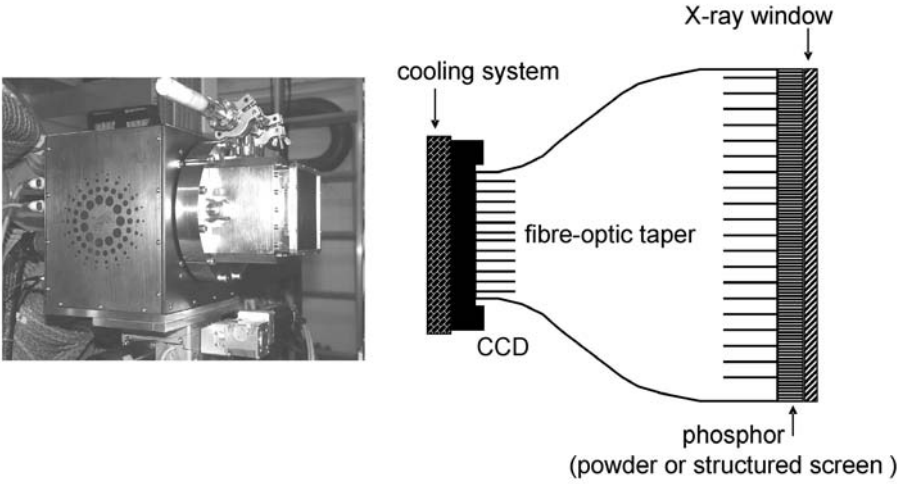


FIG. 10.8. CCD camera with a taper coupling. Courtesy: ESRF, Instrument Support Group.

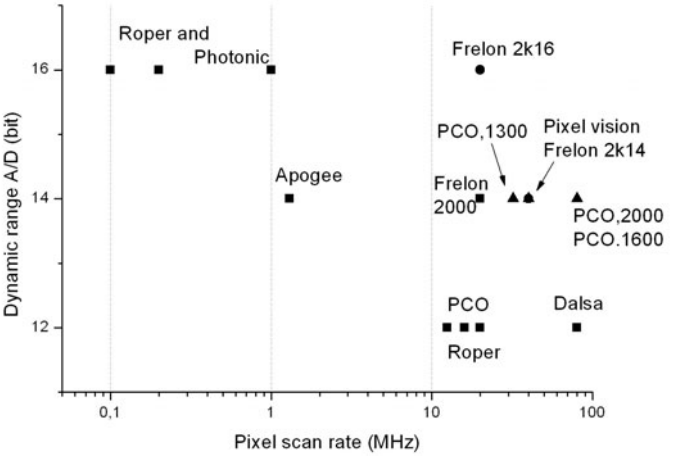


FIG. 10.9. CCD cameras used for tomography applications with synchrotron radiation.

represents a compromise between frame rate and dynamic range. Examples are the FReLoN2k (Bravin *et al.*, 2003) and the pixel vision camera (Stampanoni *et al.*, 2002), both based on the ATMEL TH7899M CCD chip. Another example is the Apogee system, see Tab. 10.1, that uses the Kodak KAF-1602 E chip (Ham *et al.*, 2002, Beckmann, 2001).

10.6.2 *Noise factors in CCD cameras*

In general, one can categorize noise as either *external* or *internal*. External noise includes radiated electromagnetic interference and 50 or 60 Hz noise from the power supply. Internal noise, associated with the acquisition of an image, comprises readout noise, dark-current noise and photon shot noise. The *readout noise* depends on the readout speed (the faster the readout rate, the higher the noise) and the electronics controlling the CCD. This is why high-speed cameras have a reduced dynamic range. The readout electronics inside the FReLoN camera is specifically designed to reduce readout noise to a minimum, so as to allow for fast frame rates with high (or medium) dynamic range. CCD cameras are integrating detectors that also integrate the thermally generated noise, known as *dark-current noise* or dark noise. Inside the silicon of the CCD chip thermal energy will create electron–hole pairs. The electrons will be captured in the pixel wells and contribute to the signal. The dark current is highly temperature dependent, hence the CCDs are cooled to $-40°$C to $-20°$C to reduce the dark current, thus allowing for long exposure times. Cooling is achieved by thermoelectrical cooling (Peltier elements), forced air, water or liquid-nitrogen cooling. Since the accumulated dark noise increases linearly with time this factor is less important for fast tomography applications. The *photon shot noise* is a fundamental property of the quantum nature of light. According to statistical theory the number of visible photons collected by the CCD will exhibit a Poisson distribution. The charge induced by photon shot noise will also be Poisson distributed. The magnitude of the noise equals the square root of the number of corresponding signal photons.

10.6.3 *Readout schemes of CCD cameras*

The CCD detector allows for a certain flexibility for the readout of the pixels. For instance pixels can be binned together within rows or columns, increasing the frame rate at the cost of spatial resolution. Selected readout of a region of interest also increases the frame rate, but at the cost of image size (or FOV). Depending on the application one of the following general readout schemes is used: full-frame mode, frame-transfer mode, or pipeline mode.

The *full-frame mode* is the standard mode of operation of CCD detectors. In this mode the full chip is used to record the image. During readout the image is shifted down row-by-row to a horizontal register that is subsequently read out and digitized pixel-by-pixel through a single readout port. One can speed up the readout by using two or four readout ports (one at each corner). The speed-limiting step is, in general, the readout and digitization of the pixels, and not the shifting of the charge inside the pixel towards the readout node, which can be very fast. A disadvantage of the full-frame mode is the fact that during readout the CCD is still sensitive, while the image is shifted down, resulting in smearing of the image. Therefore, a shutter has to be placed in the X-ray path (X-ray shutter) or in the optical path, between the luminescent screen and the CCD, in order to stop image acquisition during readout.

An alternative mode of operation that circumvents this problem is the *frame-transfer mode*. In this mode, only half of the CCD chip is used to record the image, the other half of the chip is used as storage. While the image area is exposed to light, the storage area can be read. Once the reading is finished, the data in the image area is quickly shifted into the storage area for readout. Since this only involves shifting of charges and no digitization, this can be done very fast and no additional shutter is needed. In this mode, the CCD camera can run continuously (no dead time) with high frame rates. The disadvantage is that only half the CCD chip is used for recording data, thereby reducing the FOV.

Finally, there is the *kinetic pipeline mode* that is used for fast tomography. This can be considered an extreme case of the frame-transfer mode. This time, only a few lines are exposed to light, while the rest of the chip is masked. The exposed lines are shifted to the two nearest readout nodes and are read out in synchronization with the sample movement. In order to accelerate the image-acquisition frequency, the size of the exposed area should be as small as possible. The exposed area has to be equal to 2^n pixels, where n is between 0 and 11 for a CCD with 2048 lines. A minimum dead time of 850 µs between the acquisitions of two lines can be achieved in this mode, thus allowing very fast tomography experiments.

10.7 Review of potential solutions for large field of view detector

The main part of this chapter has dealt with very high spatial resolution tomography with a small FOV. Another important field is tomography of larger objects, needing a detector with a much larger FOV and smaller spatial resolution. In this section, we will discuss two systems that show good potential to solve the needs for this particular field. The first system is a large flat panel based on amorphous silicon (a-Si) and developed for medical imaging. The second system is a flat panel based on complementary metal-oxide semiconductor (CMOS) technology.

10.7.1 *a-Si-based flat panels*

The medical imaging industry has recently introduced large fully digital flat-panel imagers that will gradually replace the traditional X-ray film cassettes in radiographic tables for general radiographic applications in vascular, thoracic and skeletal diagnostics. These flat-panel X-ray detectors are based on large-area solid-state integrated circuit technology in amorphous silicon. The integrated circuits consist of a photodiode array and an active TFT (thin-film transistor) matrix readout. A converter layer is deposited on top of the photodiode array. This layer can either be a direct converter such as amorphous Se or a scintillator converting X-rays to visible light such as cesium iodide (CsI). Trixell[12] produces flat-panel imagers based on a-Si and using a CsI scintillator. A schematic layout of their system is given in Fig. 10.10. Their Pixium 4700 detector has an active

[12]TRIXELL, 460 rue du Pommarin, 38430 Moirans, France; www.trixell.com.

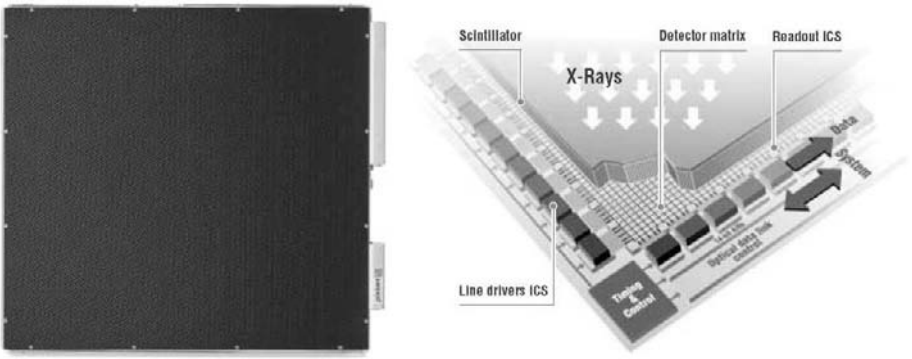


FIG. 10.10. Digital flat-panel detector for real-time X-ray imaging, TriXell. Courtesy of Trixell/Thales.

matrix of 30×40 cm, containing a total of 2480×1910 pixels each of 154 μm. The data are digitized by 14-bit analogue-to-digital converters. At full resolution using all 5 megapixels a frame rate of 7.5 fps is obtained, 60 fps is possible with 4×2 binning. The CsI converter layer consists of many thin, rod-shaped, CsI crystal needles aligned parallel to one another, see Fig. 10.11(c). When an X-ray photon is absorbed by a CsI crystal the scintillator produces light that is reflected within the needle and transmitted out of one end of the needle without lateral diffusion. The photodiodes in the a-Si panel detect the scintillator light and convert it into an electrical signal. The primary benefit of CsI technology is the excellent DQE. The advent of these flat-panel digital imagers is revolutionary for medical imaging, but also offers opportunities for large FOV applications in the 30 to 100 keV energy range at synchrotron radiation sources. One disadvantage of the a-Si flat panels is the limitation in spatial resolution that amounts to 3.2 LP/mm for the pixium 4700 owing to the properties of the a-Si semiconductor material.

10.7.2 *CMOS photodiode arrays*

The second system we discuss here is produced by Hamamatsu Corporation. Their flat panels are based on the CMOS technology and also consist of a photodiode array but with smaller pixel size than the TFT arrays presented above. These imagers are very well suited for mammography and non-destructive analysis, but are certainly also interesting for many synchrotron radiation experiments (Yagi *et al.*, 2004*b*). The C7930DP has 4416×3520 pixels at a 50 μm pitch (220.8 mm×176 mm) and can be run at frame rates of 1 fps (no binning) to 3.5 fps (2×2 binning), with 14-bit dynamic range (of the ADC). The spatial resolution obtained with this system is 10 LP/mm. The C7930DP is designed with a scintillator as the converter layer. Two types of scintillating materials are available: CsI and Gadox (Gd_2O_2S:Tb) which are deposited on aluminium, amorphous carbon and fibre-optic plates. As for the Trixel flat-panel imager the CSI scintillator is

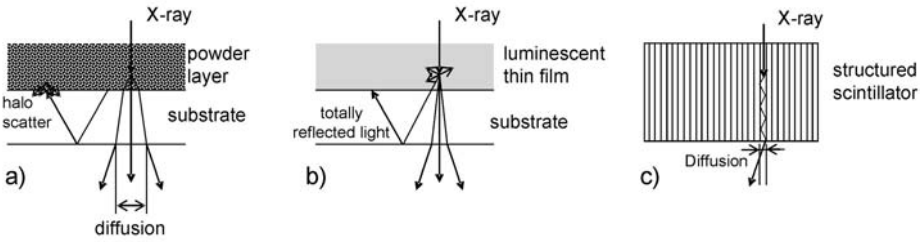


FIG. 10.11. Difference between powder screen (a), transparent thin layer (b) and structured scintillator (c). In a powder screen, the light emitted leaves the screen by scattering in all directions. Scattering by neighbouring particles produces the familiar Lambertian angular distribution, a thicker screen involves a greater spreading. In structured scintillators, light emitted is propagated within the needle without significant spreading. In a transparent screen only the part of the light that is emitted at angles below the critical angle leaves the screen. The rest of the light is guided to the side of the screen by multiple reflection.

grown in parallel needles with a 6 to 7 μm diameter, resulting in the excellent spatial resolution, see Fig. 10.11(c).

10.8 Summary

In this chapter we have discussed the components of the detectors used for tomography at synchrotron beamlines. The main application is microtomography requiring detectors with high spatial resolution in the micrometre range. This is achieved by using very thin converter screens that are coupled via microscope optics to a CCD detector. We have shown the limitations of the various components of the system. The achieved spatial resolution of 0.5 μm is close to the theoretical limit of such a system. Therefore, the main gain that can still be achieved is in the quantum efficiency by using converter screens with higher X-ray stopping powers (higher-Z elements) and in the speed by using faster CCD detectors (or detection schemes). We have also presented the large field-of-view detectors developed by the medical imaging industry. These detectors will certainly be of interest for synchrotron experiments when the highest spatial resolution is not needed, but when speed and size are of importance.

10.9 References

Beckmann, F. (2001). Microtomography using synchrotron radiation as user experiment at beamlines BW2 and BW5 of HASYLAB at DESY. In *Developments in X-ray tomography III* (ed. U. Bonse), pp. 34. The International Society for Optical Engineering (SPIE). Proceedings of the SPIE, Vol. 4503.

Bravin, A., Fielder, S., Coan, P., Labiche, J.-C., Ponchut, C., Peterzol, A., and Thomlinson, W. (2003). Comparison between a position sensitive germanium detector and a taper optics CCD 'FRELON' camera for diffraction enhanced imaging. *Nuclear Instruments and Methods in Physics Research A*, **510**, 35.

Carlo, F. De, Albee, P., Chu, Y.S., Mancini, D.C., Tieman, B., and Wang, S. Y. (2001). High-throughtput real-time X-ray microtomography at the Advanced Photon Source. In *Developments in X-ray tomography III* (ed. U. Bonse), pp. 1. The International Society for Optical Engineering (SPIE). Proceedings of the SPIE, Vol. 4503.

Cloetens, P., Ludwig, W., Baruchel, J., Dyck, D. Van, Landuyt, J. Van, Guigay, J.P., and Schlenker, M. (1999). Holotomography: Quantitative phase tomography with micrometer resolution using hard synchrotron radiation X-rays. *Applied Physics Letters*, **75**, 2912.

Di Michiel, M., Merino, J.M., Fernandez-Carreiras, D., Buslaps, T., Honkimaki, V., Falus, P., Martin, T., and Svensson, O. (2005). Fast microtomography using high energy synchrotron radiation. *Review of Scientific Instruments*, **76**, 043702.

Dujardin, C., Luyer, C. Le, Martinet, C., Garapon, C., Mugnier, J., Murrillo, A.G., Pedrini, C., and Martin, T. (2005). Thin scintillating films of sesquioxides doped with Eu3^{+}. *Nuclear Instruments and Methods in Physics Research A*, **537**, 237.

Garcia-Murillo, A., Luyer, C. Le, Dujardin, C., Martin, T., Garapon, C., Pedrini, C., and Mugnier, J. (2002). Elaboration and scintillation properties of Eu^{3+} doped Gd_2O_3 and Lu_2O_3 sol-gel films. *Nuclear Instruments and Methods in Physics Research A*, **486**, 181.

Garcia-Murillo, A., Luyer, C. Le, Dujardin, C., Pedrini, C., and Mugnier, J. (2003). Rare-earth actived sol-gel films for scintillators applications. *J. Sol-Gel Science and Technology*, **26**, 957.

Gruner, S.M., Tate, M. W., and Eikenberry, E. F. (2002). Charge-coupled device area X-ray detectors. *Review of Scientific Instruments*, **73**, 2815.

Ham, K., Jin, H., Butler, L. G., and Kurtz, R.L. (2002). A microtomography beamine at the Louisana State University Center for Advanced Microstructures and Devices synchrotron. *Review of Scientific Instruments*, **73**, 1521.

Hejazi, S. and Trauernicht, D. (1997). System considerations in CCD-based X-ray imaging for digital chest radiography and digital mammography. *Medical Physics*, **24**, 287.

Hoheisel, M., Giersch, J., and Bernhardt, P. (2004). Intrinsic spatial resolution of semiconductor X-ray detectors: a simulation study. *Nuclear Instruments*

and Methods in Physics Research A, **531**, 75.

Jung, H., Kim, H., Hong, S. Hong. J., Jeong, H., Je, J. Ho, Kim, B., and Yoo, H. (2002). Computed microtomography μ-CT with unmonochromatized synchrotron X-rays for cancerous human breast tissue and mouse vertebra. *IEEE Transactions on Nuclear Science*, **49**, 2262.

Koch, A., Cloetens, P., Ludwig, W., Labiche, J.C., and Ferrand, B. (1999*a*). Reading thin-film with optical microscopes for X-ray imaging. In *Proceedings of the Fifth International Conference on Inorganic Scintillators and Their Applications SCINT99* (ed. V. Mikhailin), pp. 157.

Koch, A., Perrin, F., Heurtier, P., Ferrand, B., Chambaz, B., Ludwig, W., and Couchaud, M. (1999*b*). X-ray camera for computed microtomography of biological samples with micrometer resolution using $Lu_3Al_5O_{12}$ and $Y_3Al_5O_{12}$ scintillators. In *Medical imaging 1999: Physics of medical imaging* (ed. J. Boone and J. Dobbins), pp. 170. The International Society for Optical Engineering (SPIE). Proceedings of the SPIE, Vol. 3659.

Koch, A., Raven, C., Spanne, P., and Snigirev, A. (1998). X-ray imaging with submicrometer resolution employing transparent luminescent screens. *Journal of the Optical Society of America A*, **15**, 1940.

Lee, H.R., Lai, B., Yun, W., Mancini, D.C., and Cai, Z. (1997). X-ray microtomography as a fast three-dimensional imaging technology using a CCD camera coupled with a $CdWO_4$ single crystal scintillator. In *Developments in X-ray tomography* (ed. U. Bonse), pp. 257. The International Society for Optical Engineering (SPIE). Proceedings of the SPIE, Vol. 3149.

Lennstrom, K., Limmer, S.J., and Cao, G. (2003). Synthesis of cadmium tungstate films via sol-gel processing. *Thin Solid Films*, **434**, 55.

Liu, H., Karellas, A., Harris, L. J., and D'Orsi, C. J. (1994). Methods to calculate the lens efficiency in optically coupled CCD X-ray imaging systems. *Medical Physics*, **21**, 1193.

Martin, T., Couchaud, M., Ferrand, B., Caillet, A., Pelenc, D., Chambaz, B., and Passero, A. (2005). Thin-film scintillators for computed microtomography with micrometer resolution using $Gd_3Ga_5O_{12}$ scintillators. In *International Conference on Inorganic Scintillators and Their Industrial Applications SCINT2005* (ed. A. Gektin and B. Grinyov), pp. 459.

National Institute of Standards and Technology (2006). Stopping-power and range tables for electrons. `http://physics.nist.gov/PhysRefData/Star/Text/ESTAR.html`. [Online; accessed 16 May 2007].

Ponchut, C. (2006). Characterization of X-ray area detectors for synchrotron beamlines. *Journal of Synchrotron Radiation*, **13**, 195.

Stampanoni, M., Borchert, G., Wyss, P., Abela, R., Patterson, B., Hunt, S., Vermeulen, D., and Rüegsegger, P. (2002). High resolution X-ray detector for synchrotron-based microtomography. *Nuclear Instruments and Methods in Physics Research A*, **491**, 291.

Swank, R. K. (1973). Calculation of modulation transfer function of X-ray fluorescent screens. *Applied Optics*, **12**, 1865.

Tseng, P., Hwang, C.S., and Bong, Y.F. (2000). Phasecontrast X-ray microscopy. Annual report, National Synchrotron Radiation Research Center, Taiwan.

Uesugi, K., Suzuki, Y., Yagi, N., Tsuchiyama, A., and Nakano, T. (2001). Development of sub-micrometer resolution X-ray CT system at Spring-8. In *Developments in X-ray tomography III* (ed. U. Bonse), pp. 291. The International Society for Optical Engineering (SPIE). Proceedings of the SPIE, Vol. 4503.

Wang, Y., Carlo, F. De, Mancini, D.C., McNulty, I., Tieman, B., Bresnahan, J., Foster, I., Insley, J., Lane, P., Laszewski, G., Kesselman, C., Su, M., and Thiebaux, M. (2001). A high-throughput X-ray microtomography system at the Advanced Photon Source. *Review of Scientific Instruments*, **72**, 2062.

Yagi, N., Inoue, K., and Oka, T. (2004*a*). CCD-based X-ray area detector for time-resolved diffraction experiments. *Journal of Synchrotron Radiation*, **11**, 456.

Yagi, N., Yamamoto, M., Uesugi, K., and Inoue, K. (2004*b*). A large-area CMOS imager as an X-ray detector for synchrotron radiation experiments. *Journal of Synchrotron Radiation*, **11**, 347.

PART III

ELECTRON TOMOGRAPHY

Transmission electron microscopy is an established method in materials research. Commercial TEMs are being improved constantly and every new generation of microscopes provides new experimental options, such as improved resolution or novel analytical tools. Transmission microscopy implies that features in three dimensions are projected on a plane and therefore information is lost. This limitation can now be overcome by applying tomographic techniques. Due to the extremely high magnification in such instruments special care had to be taken to develop sample holders that allow us to tilt samples in a stable and reproducible way. Nowadays, commercial solutions are available and the field of electron tomography is developing quickly.

In this part of the book, the fundamentals of electron tomagraphy will be reviewed, after which a wide range of applications will be presented. Due to the importance of transmission electron microscopy for the semiconductor industry these applications are given ample coverage.

This page intentionally left blank

11

FUNDAMENTALS OF ELECTRON TOMOGRAPHY

Paul A. Midgley and *Matthew Weyland*

In this chapter, we will review the fundamentals of electron tomography and highlight aspects that are unique to using electrons. Each imaging mode used for electron tomography will be discussed and we will outline their benefits and drawbacks with examples of the diverse range of materials now being studied. First, we review some of the fundamental aspects of electron tomography. Many of the theoretical concepts described in earlier chapters also apply here and so we restrict ourselves to a discussion of the theory that is of particular importance to electron tomography. In the next chapter we shall present some applications of electron tomography.

11.1 Introduction

Electron tomography, in its many forms, has been of great use in structural biology for three decades, but it is only in the last 5 to 10 years that electron tomography has become a valuable tool in materials science. In biology, the need is to understand cellular structures, viruses and bacteria by visualizing their complex 3D structure and recognizing how functional groups (e.g. proteins), determined perhaps by X-ray crystallography, fit into the overall structural envelope reconstructed by electron tomography. In materials science, there is a need to understand the functional and mechanical behaviour of materials in three dimensions at the micro- and nanoscale. This need comes from many quarters: from the burgeoning field of nanotechnology, where the hope is to assemble devices from the bottom-up, from the more 'traditional' semiconductor industry, where ever more complex 3D arrays of components are integrated onto silicon wafers, and from the chemical industry, where a better knowledge of the fine-scale 3D structure of polymers and catalysts is needed to enhance their properties. Of course, electron microscopy is just one of a raft of imaging techniques available to the materials scientist and whilst it may provide excellent 3D spatial resolution from a wide variety of materials, it cannot be used to examine large volumes of material and therefore must be viewed as complementary to other tomographic techniques used across the length scales. Indeed, there are a growing number of materials systems for which, to understand their full functionality, knowledge of the structure is required from nanometres to millimetres.

The transmission electron microscope (TEM), or scanning transmission electron microscope (STEM), is ideally suited as a tool for high-resolution tomographic analysis of materials systems. In the microscope, the electron beam prop-

agates through the whole structure and one can think of the TEM as a 'structure projector' (Hawkes, 1992). The parallel nature of the illumination simplifies the geometry and the interpretation of the transmitted image (or 'projection'). In some cases a single image is sufficient to understand a structural problem but, increasingly, the thickness of the sample through which the electron beam is transmitted (typically 10 to 100 nm) contains overlapping structural details that can be resolved only through 3D tomographic reconstructions.

Although the first experimental demonstrations of electron tomography were made in 1968 (Rosier and Klug, 1968, Hoppe *et al.*, 1968), it is with the advent of computer-controlled electron microscopes, to aid acquisition and minimize dose, and the increasing processing power of modern computers, that electron tomography has become a technique used routinely. The first application of electron tomography in materials science (Spontak *et al.*, 1988), was to visualize 3D polymer microstructures formed from several distinct monomer units with mutual immiscibilities; structures ranged from simple lamellae to complex gyroid networks. Many other materials systems have now been studied by electron tomography ranging from porous zeolites and similar silica-based catalyst supports to complex polymers, metals, ceramics and semiconductors (Midgley and Weyland, 2003, Weyland and Midgley, 2003) and references therein). In addition, over the past few years it has become apparent that, unlike in the life sciences where, with rare exceptions, only the bright-field (BF) signal is used to form images suitable for tomography, in materials science, many new imaging modes for electron tomography are being developed, leading to the visualization of 3D composition, electrostatic potential, magnetic fields, dislocation networks and so on.

11.2 Tomography using the electron microscope

11.2.1 *The projection requirement*

It is often stated that an image signal used for tomographic reconstructions must conform to the 'projection requirement' – not to be confused with the projection theorem, eqn 2.9 – which states that the projected signal must be a monotonic function of a projected physical quantity (Hawkes, 1992). This is exemplified in conventional X-ray absorption or neutron tomography, described in Chapter 5 or Chapter 13, in which image contrast is generated by changes in the absorption coefficient and the transmitted X-ray signal varies monotonically with thickness. There are indications, however, that the projection requirement is not necessary for tomographic reconstruction, i.e. even when this requirement does not hold, proper tomographic 3D images can be obtained by using appropriate mathematical algorithms. However, an imaging signal satisfying the projection requirement is conceptionally easier to process and one therefore prefers such signals in practice. In conventional electron microscopy, the depth of focus is sufficiently large that to a good approximation the recorded image can be regarded as a projection; this may break down under some circumstances, see Section 11.7.3. In addition, there are many competing contrast mechanisms all of which, to

varying approximations, obey the projection requirement. For amorphous materials, such as many biological structures and polymers, conventional BF TEM contrast arises from changes in specimen density or thickness (much like the X-ray case). For very weakly scattering specimens (i.e. weak phase-object approximation), phase-contrast images can be used for tomography, rather like the phase-contrast images obtained from an X-ray synchrotron source. Phase contrast is normally generated in the TEM by a small defocus, using the objective lens to transfer phase information to variations in amplitude. This method is used with great success in life science cryomicroscopy and cryotomography in which frozen-hydrated specimens are examined in a 'natural state'. The contrast transfer function (CTF) of the objective lens should be taken into account when interpreting such phase images (Fernandez *et al.*, 2006).

For crystalline materials, complex dynamical interactions between the electron beam and the crystal potential ensure that, in general, there is no simple relationship between the image intensity and the physical properties of the specimen. In some circumstances this barrier can be overcome and BF tomography has been used to reconstruct crystalline structures, see Section 12.1, but tomograms often show artefacts and poor resolution of internal structure. However, there is now a growing number of alternative imaging and contrast modes that satisfy the projection requirement and these are discussed in more detail later.

11.2.2 *Acquisition*

More than many other tomographic experiments, the number of projections that can be acquired in the electron microscope is severely limited for two principal reasons: beam damageand constrained geometry. Firstly, a large number of samples examined in the microscope will be beam-sensitive (Egerton *et al.*, 2004). The electron flux can be many orders of magnitude greater than that used in X-ray or neutron tomography and so damage can be faster and more severe, see also the discussion in Chapter 4. Organic materials, hydrated specimens and insulating specimens may all be damaged under the beam from the effects of local beam heating, charging and ionization damage (Hobbs, 1979). Damage in such materials often arises via inelastic scattering events that provide sufficient energy to break interatomic bonds. The cross-section for all inelastic scattering events falls with increasing beam energy and therefore damage can be reduced by increasing the microscope accelerating voltage. Semiconductors and metals can also suffer beam damage, such as compositional diffusion, the movement of defects under the beam and 'knock-on' damage, caused when an incident electron imparts sufficient energy in an elastic collision to knock an atom from its equilibrium position. There is a displacement threshold energy above which this interaction will occur and therefore to avoid such damage the microscope should be operated below that threshold. In the life sciences in particular, damage is a major problem and techniques of 'dose-fractionation' have been developed to maximize the overall 3D signal-to-noise ratio from a restricted number of images.

Although electron microscopes are high-vacuum instruments, and specimens

are prepared with care, in practice, some degree of contamination will occur during the acquisition of a tilt series of images. Temperature gradients at the specimen surface brought about by specimen–beam interactions drive the diffusion of mobile hydrocarbons under the electron beam that then crack to leave unwanted carbon deposits (Egerton *et al.*, 2004). Such contamination can be minimized by bathing the specimen in an argon/oxygen plasma for a short time ($\approx$ few minutes). The plasma encourages carbon-based contaminants to form gaseous oxides that are pumped away. This is clearly not suitable for many organic specimens as they will be damaged by the plasma. An alternative is to 'flood' the specimen, which involves the illumination by the electron beam of a wide area using parallel illumination and a high electron flux for ≈ 30 min. This leads to an even cracking of hydrocarbons across the entire illuminated area and produces a very thin surface carbon layer that should not degrade the image quality substantially.

Secondly, unlike many other tomography methods, electron tomography is undertaken in an instrument with a highly restricted working space. The sample is tilted about an axis within the polepiece gap of the objective lens. In order to achieve high spatial resolution, the aberrations of the objective lens must be kept to a minimum. To do so, the polepiece gap must be as small as possible. There is therefore a conflict between having room within the polepiece gap in order to tilt samples and the need to reduce the gap to obtain high-resolution images. Very recently, aberration correctors have been introduced on some instruments that allow the gap to be large, whilst still maintaining a high spatial resolution. Modern tomography specimen holders, with slender profiles, allow high tilt even within a narrow polepiece. Nevertheless, with conventional specimens, that are slab-like in nature, the projected specimen thickness will increase dramatically at high tilts; for example the projected thickness at 70° tilt is three times that at zero tilt. Large specimen thicknesses will tend to lead to blurring, through chromatic aberration in TEM or beam broadening in STEM, and a general degradation of image contrast, brought about by increasing 'absorption'. In electron microscopy, 'absorption' normally refers to electrons, scattered to very high angles, which are lost outside the imaging system. Thus, in almost all electron tomography experiments there is a 'missing wedge' of information that is centred about the optic axis, see Fig. 11.1. The consequences of this missing information will be discussed later.

In materials science the drive towards higher spatial resolution has led to the use of objective lenses with small (2 to 5 mm) polepiece gaps but still with sufficient space to enable, for example, the integration of an energy-dispersive X-ray (EDX) detector. Within those constraints, and the standard TEM specimen diameter of 3 mm, modern high-tilt, single-tilt and dual-axis holders have been designed to tilt to 70° and above (Weyland *et al.*, 2004). An example of a single-tilt high-tilt holder is shown in Fig. 11.2. An alternative solution to the restricted tilt range is to manufacture a fine 'needle' or 'pillar' specimen and to tilt the specimen about its own axis with an internal tilt mechanism

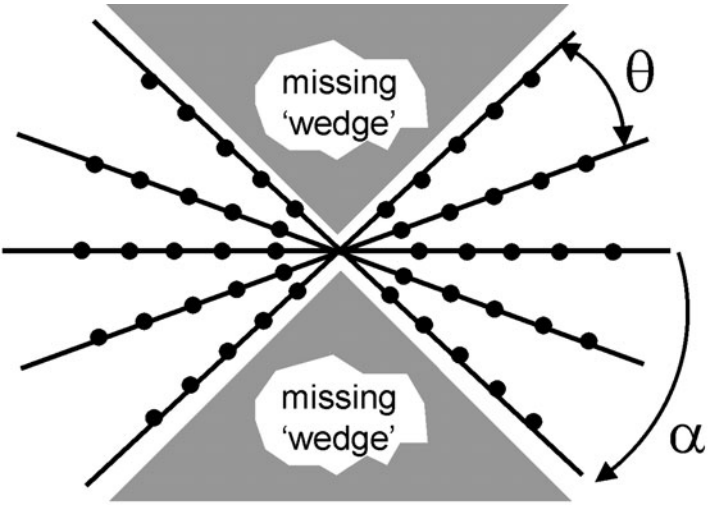


FIG. 11.1. A representation in Fourier space of the sampling of an object by a single tilt series of images. The radial lines in the figure represent Fourier planes. Although each plane might have an even sampling of data points (solid circles), the ensemble of images leads to an undersampling of high-frequency information and a consequent blurring of the reconstructed object. θ is the angular sampling and α is the maximum tilt angle.

(Koguchi *et al.*, 2001). This allows complete 360° tilt of the specimen without shadowing and without a significant increase in projected thickness at certain tilts. The ensemble of images acquired from such a specimen will not suffer from the missing wedge of information described before and hence will, in principle, have reduced reconstruction artefacts. The needle/pillar specimen can be prepared from almost any material in a site-specific fashion by focused ion beam milling (FIB) using a beam of gallium ions. However, significant damage, and often complete amorphization to a depth of 20 nm or more below the surface, can result from high-energy ion milling (Ishitani and Yaguchi, 1996). Modern FIB methods, using lower-energy gallium ions, can reduce the damaged layer significantly.

Despite the high mechanical precision to which microscope goniometers are now made, it is inevitable that the object of interest will shift, both laterally and vertically, in the field of view over the range of a tomographic tilt series. To combat this, either the object has to be shifted manually, or the microscope is pre-calibrated in order to compensate for the shift at each tilt. Further, a computer can carry out repositioning and refocusing tasks with a smaller number of images and in a shorter time than a human operator, leading to a reduction in total dose (Dierksen *et al.*, 1993). Automated acquisition schemes usually make use of filtered cross-correlation algorithms to determine the shift between successive images. Focus can be corrected in BF TEM by measuring the image shift

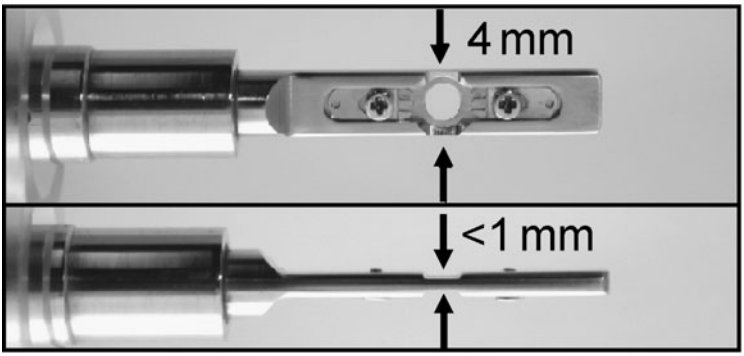


FIG. 11.2. Fischione model 2020 room-temperature single-tilt axis high-tilt sample holder for analytical TEM polepieces. The holder has a narrow profile, reducing shadowing to allow specimen tilt to $\pm 70^\circ$ or higher over most of the specimen area. Copyright 2006 by Elsevier.

induced by a small beam tilt (Koster *et al.*, 1987). In HAADF STEM imaging, focus is determined by maximizing the image contrast. To further reduce dose, the focus and shift correction can be determined in a region outside the volume of interest (Dierksen *et al.*, 1993), akin to the 'low-dose' methodology of cryomicroscopy. The goniometer can be pre-calibrated by acquiring a tilt series from a flat specimen with high contrast features, such as gold on a carbon film, and recording image movement as a function of tilt (Ziese *et al.*, 2002). Both the calibration and the actual tilt series should be acquired close to the eucentric height of the goniometer at which unwanted movement will be a minimum. With a pre-calibrated goniometer, a 140-image bright-field TEM tilt series can be acquired using this approach in under one hour.

The image may also shift with tilt because of an offset between the microscope's optic axis, defined by the objective lens and deflector alignment, and the eucentric position defined by the position of the goniometer. Typically a few micrometres in size, this can have a significant influence on the apparent movement of an image through a tilt series. If this offset is known, it can be corrected by beam-shifting the optic axis onto the eucentric axis. Several automation schemes now calibrate this offset and build in compensating beam shifts during the acquisition process (Mastronarde, 2005).

11.3 Alignment and reconstruction

11.3.1 *Alignment of tilt series*

The key to high-quality tomographic reconstructions is a very well-aligned data set. Two conventional approaches are used (i) tracking fiducial markers and (ii) cross-correlation. The fiducial technique determines both spatial alignment and the direction of the tilt axis, plus any secondary distortions caused by optical effects (Lawrence, 1992). Cross-correlation alignment makes use of the information

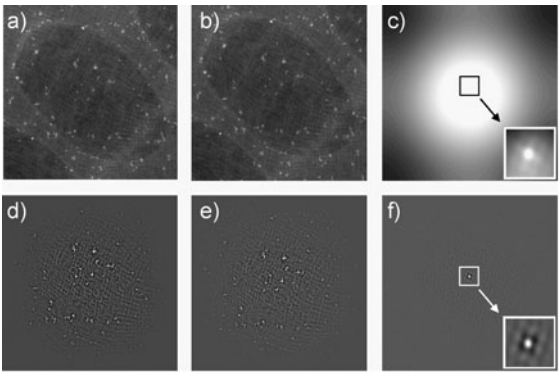


FIG. 11.3. Example of cross-correlation to determine the relative shift between two images recorded 2° apart. (a),(b) Two images of a nanostructured block--copolymer from a HAADF STEM tilt series. (c) cross-correlation image between the two images, showing a broad intensity peak near the image centre. This corresponds to the shift between the two images. (d)–(f) are the same series of image, but show the improvement in the width of the cross-correlation peak on application of a highpass filter to the original images.

in the whole image (rather than a few selected points in the fiducial technique) and makes no assumptions about the shape of the support film. It also avoids possible reconstruction problems associated with high-contrast objects, such as colloidal gold, which can mask details in the reconstruction. However, cross-correlation alignment does not automatically determine the tilt-axis direction.

11.3.2 *Alignment by cross-correlation*

Given a finely spaced angular increment, any two projections in a tilt series, will share many common image features that can be found through the use of a cross-correlation algorithm (Frank and McEwen, 1992). The accuracy of the spatial alignment is affected strongly by the contrast and SNR of the images. When dealing with low-contrast and/or noisy images, the cross-correlation peak may be indistinct and, to improve the correlation, images are filtered (using, for example, bandpass and Sobel filters) to enhance features (Russ, 1995). Figure 11.3 demonstrates the improvement in correlation achieved by filtering images. Similar image filtering is also used during the tilt-series acquisition to improve tracking and refocusing.

The change in the projected shape of an object after tilt means that cross-correlation must be carried out sequentially. A consequence of this is that small, often subpixel, misalignments can occur between each projection (Frank and McEwen, 1992) that, over the tilt series, can build up to be a significant source of error. To minimize this effect, a tilt series is split into two, each using the zero-tilt image as its first reference, and all alignment steps are applied from zero out to the largest tilt in both directions. In addition, the cross-correlation function

between consecutive images will show a broadened correlation peak, dependent on the angle θ between them; applying a linear 'stretch' to the images of $1/\cos\theta$ perpendicular to the tilt axis helps to restore the spatial relationship between successive projections (Guckenberger, 1982). In practice, it is found that more than one cross-correlation alignment pass is necessary because of the nature of the rotational alignment, especially without fiducial markers, and the use of apodizing filters that subtly alter the correlation, especially for large shifts.

11.3.3 *Alignment by tracking of fiducial markers*

As an object is projected throughout a tilt series, its position and apparent shape will change depending on the tilt of the sample with respect to the beam and axis about which the sample is tilted. The task of alignment can be simplified by dispersing fiducial markers, often spherical colloidal gold particles, onto the film that can be used as an alignment reference. This approach, first demonstrated by Hart in 1968, is still the principal method for tilt-series alignment of biological specimens, e.g. Berriman *et al.* 1984. The movement of these particles in each projection is recorded and the tilt-axis direction, relative lateral shift, magnification change and image rotation are determined by comparison to a reference projection. The number of fiducial markers required will depend on the number of images, whether different tilt axes are used and on the accuracy of the goniometer tilt readings; typically 15 to 20 markers are sufficient. Automated approaches to marker selection can be unreliable and often need manual adjustment (Fung *et al.*, 1996).

11.3.4 *Tilt-axis alignment without fiducial markers*

The correction of an inplane rotation can be carried out without using fiducial markers by a number of different techniques (Frank, 1981); an out-of-plane rotation (or tilt), as occurs during a tomographic tilt series, is a more difficult problem to solve. Three approaches are presented here: common lines, series summation and arc minimization.

11.3.4.1 *Common lines* In the direction parallel to the tilt axis, there is always a set of features that are common between two images of the same object, recorded at different tilts. In the perpendicular direction, the correlation peak will be lengthened due to the foreshortening of image features and an estimate of the tilt axis can be made from the direction of this spread. The common-lines approach to axis alignment is based on measuring the trend in this spread throughout the tilt series. However, this spread is small between images that are close in tilt, and thus difficult to measure with certainty, and between images that are far apart, this spread can be masked by changes in object shape and/or poor noise statistics. There is an equivalent line in each image of the tilt series. Searching for this line, in Radon (or sinogram) space, is the basis behind the most recent implementation of the common-lines approach (Liu *et al.*, 1995).

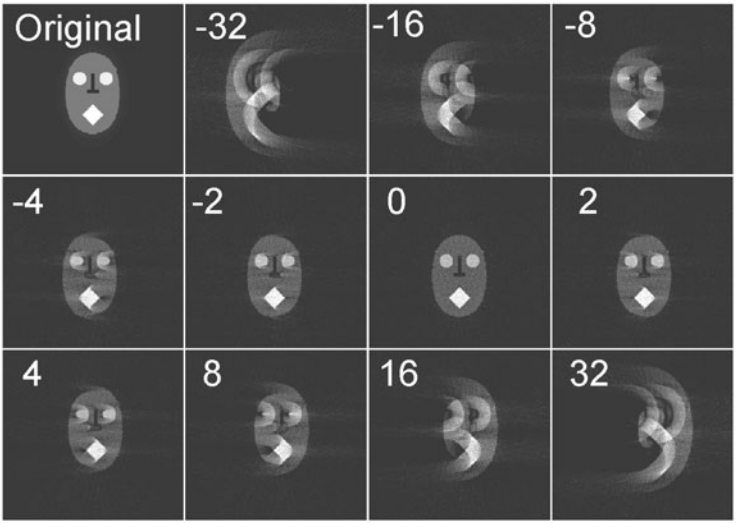


FIG. 11.4. The effect of tilt-axis misalignment on the reconstruction of a 'head phantasm' test object. Distinctive 'arcing' is observed whose sense of curvature is dependent on the direction of misalignment and the degree of which is directly related to the magnitude of misalignment. Here, the misalignment deliberately introduced is quoted in pixels perpendicular to the tilt axis. Copyright 2003 by Elsevier.

11.3.4.2 *Series summation* For the single tilt-axis geometry, movement of objects through the tilt series should follow a path that is perpendicular to that tilt axis and perpendicular to the 'common line'. If the spatial (x, y) alignment is close to optimal then by summing over all, or some, of the tilt series, the movement of any objects through the series will be highlighted (Renken and McEwen, 2003). Once the tilt-axis direction is determined the whole data set can be rotated to place the axis parallel to the image x-axis. This approach may be less effective for images with low contrast and/or large extended objects, but it is often used as a 'first guess' before being refined with slower but more accurate methods.

11.3.4.3 *Arc minimization* If the tilt axis is misaligned, the reconstruction will be smeared out into 'arcs'. The size of the arcs depends on the magnitude of the misalignment and the sign of curvature depends on the direction of the misalignment away from the correct axis. This misalignment is demonstrated for a test object in Fig. 11.4. These distinctive distortions provide a means to determine the axis direction by manually aligning, in real-time, 2D reconstruction slices from the data set. Ideally, this would be carried out by rotating the tilt series before reconstructing the new slice. For small angles this can be approximated by a shift of the projections in a single slice in the direction perpendicular to the

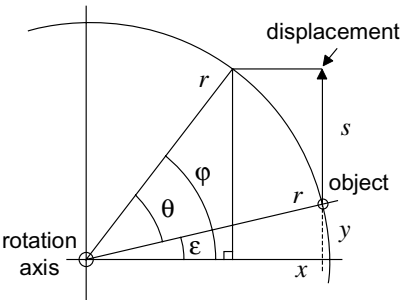


FIG. 11.5. Geometry for fine rotational alignment of the projection stack. After setting the rotation axis for the central slice any rotational misalignment of an object should follow a circular path around this rotation axis. However, as only the vertical displacement s, rather than the path along the circle, is being measured the actual correction angle θ needs to be calculated.

tilt axis. The geometry used for this assumption is shown in Fig. 11.5. The simple 2D shift is significantly less computationally demanding than a 3D rotation, and as such can be carried out in real time even for large data sets. The alignment is most easily carried out on three slices perpendicular to the assumed tilt-axis direction, one slice at the centre of the volume and two near the edge. Two variables are adjusted: an overall shift perpendicular to the axis, which shifts all slices in the same direction, and a rotation, which shifts the two outlying projections in opposite directions and leaves the centre projection unchanged. The shift of the axis can be thought of as moving the whole data set to the centre of rotation determined by the spatial alignment.

A demonstration of this alignment is shown in Fig. 11.6 for a HAADF STEM tilt series acquired from a magnetotactic bacterium. Typically, a misalignment of the axis shift is characterized by arcing of all three slices in the same direction, see Fig. 11.6(b). If the axis shift is correct, see Fig. 11.6(c), the central slice shows no arcing. Any remaining rotational misalignment is characterized by arcing in the two outlying slices in opposite directions; the arcing will be minimized by interactive rotation of the data set. The amount of shift needed to correct each of these images will be closely related to the misalignment of the tilt axis, which can be calculated by simple trigonometry, see Fig. 11.5, given the number of pixels from the aligned object to the centre of the image (x, y) and the number of pixels shifted, s. The shifts should be symmetrical on either side of the rotation axis, but to allow for errors, the angles determined from the two slices are averaged. This alignment should be repeated on different slices to arrive at a consistent data set. It is dependent on visual recognition of features, and so prone to user bias, and is more accurate when high-contrast image features are present. It is less suited to low-contrast biological samples, but can refine the alignment of large and continuous features, as demonstrated in Fig. 11.6 that is not easily achieved with other alignment techniques. More holistic approaches to

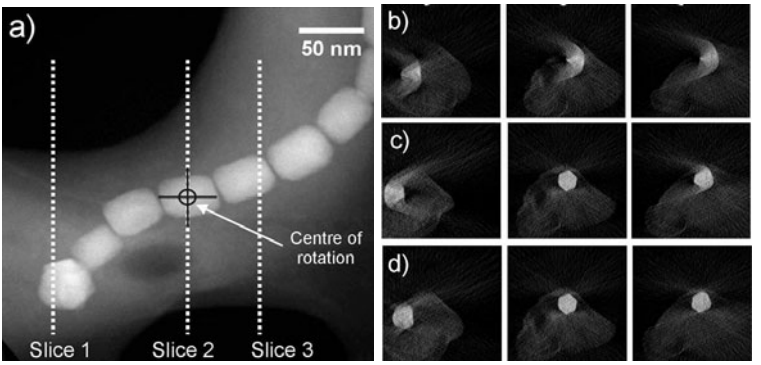


FIG. 11.6. Example of fine tilt-axis alignment by arc minimization, shown here for an HAADF STEM tilt series acquired from a magnetotactic bacterium. (a) The zero-degree projection showing the location of the three slices used for axis alignment. (b) Arcing in one direction from all three slices indicates a misalignment of the tilt-axis shift. (c) Here, the shift is correct, the centre slice is well aligned, yet the two outlying slices show arcing in opposite directions. This indicates an error in the tilt-axis angle. (d) All three slices are well aligned, indicating the correct axis angle and shift have been found.

markerless alignment include the optimization of the reconstruction by iterative alignment of projections and test reconstructions (Winkler and Taylor, 2006), or by tracking image features based on the expected geometric relationship of a tilt series (Brandt *et al.*, 2001).

11.3.5 *Reconstruction*

By sampling the object using a tilt series composed of relatively few images, and over a limited angular tilt range, the reconstruction will suffer from 'fan' artefacts, which can be characterized by a point-spread function (PSF) (Carazo, 1992). Further imperfections will be caused by the type of interpolation in the reconstruction, the use of a non-optimal weighting filter, and by deficiencies in the signal-to-noise (SNR) ratio of the original data. Such artefacts can be minimized by iterative reconstruction routines such as the *simultaneous iterative reconstruction technique* (SIRT), as explained in Chapter 2, or by maximum entropy routines (Skoglund and Ofverstedt, 1996). For dual-axis tomography, in which two tilt series are recorded about mutually perpendicular tilt axes, iterative reconstruction algorithms also exist and have been shown to be of great benefit for improving reconstruction quality. Examples of SIRT reconstructions will be shown later. In the life sciences, macromolecular assemblies often show well-defined structural symmetry, which can be exploited both in acquisition (limited dose) and reconstruction (Radermacher *et al.*, 1987). Whilst similar structural symmetry in materials science is rare (except at the atomic level!), the growth of self-assembly nanostructures is one case where local symmetry may be present

to help improve the reconstruction fidelity.

Assuming a perfect detector, the relationship between the number of projections, N, the diameter of the reconstruction volume, D and the resolution, Δ, has been defined as (Crowther *et al.*, 1970):

$$\Delta = \frac{\pi D}{N}.$$

Here we assume that the N projections are spread evenly through 180° but in electron tomography there is almost always an upper limit to the tilt angle. This leads to an anisotropy in the spatial resolution brought about by the 'wedge' of missing information. Most noticeably, the resolution is degraded in the least-sampled direction, manifesting as an 'elongation' of the object in that direction (usually the optic axis). An estimate of this elongation e, as a function of the maximum tilt angle α, is (Radermacher and Hoppe, 1980):

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

For a typical maximum tilt angle of 70°, this leads to an elongation factor of 30%. In the TEM, however, samples are often slab-like with far larger dimensions inplane (x, y) than in depth (z). In this case, D should be modified to take this into account (Radermacher, 1992):

$$D = t\cos\alpha,$$

where t is the (untilted) thickness of the slab and α is the maximum tilt angle. As a consequence of a reduced reconstruction volume, fewer projections are required to attain the same overall resolution.

Although the formulae quoted above give an indication of the expected resolution, in order to quantify the resolution in the final reconstruction, the SNR of the images within the tilt series must also be taken into account. A conventional approach is to use a Fourier shell correlation (FSC) method (van Heel and Harauz, 1986), related to the differential phase residual (DPR) (Frank, 1981) or more recently the spectral signal-to-noise ratio (SSNR) method (Unser *et al.*, 1987). These methods were developed originally for single-particle reconstructions (Henderson *et al.*, 1990) where the number of images (projections) is considerably larger than for electron tomography. In order to use the FSC method, the data must be split into two, e.g. on the basis of odd and even tilt angles. The FSC is then calculated as:

$$\mathrm{FSC}(k) = \frac{\sum F_1(k)F_2^*(k)}{\sqrt{\sum |F_1(k)|^2 \, |F_2^*(k)|^2}},$$

where $F_n(k)$ is the sum over all the images in the two halves, $n = 1$ and $n = 2$ and within a shell of small width centred at frequency k. By convention, the resolution of the data set is determined at the 0.5 value of the FSC curve (Bottcher

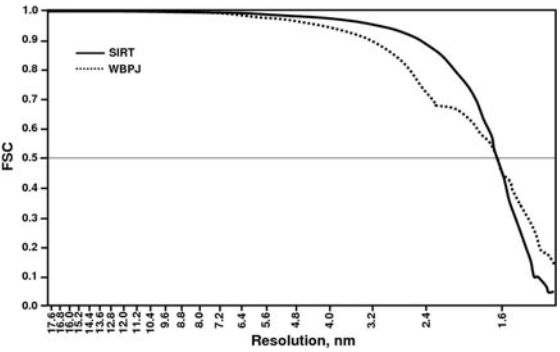


FIG. 11.7. Fourier shell correlation for a data set recorded from mesoporous silica, shown in Fig. 12.3, reconstructed using weighted backprojection (WBP) and SIRT with 30 iterations, as a function of resolution. The higher FSC at all meaningful frequencies indicates an improved fidelity for the SIRT reconstruction.

et al., 1997); this is equivalent to a signal-to-noise ratio (SNR) of 1. It has been argued (Penczek, 2002) that the FSC (and DPR) measures are unsuitable for tomography because the two reconstructions from a single-tilt series contain too few projections and as such confidence in the statistical averages is reduced. A more significant problem for the technique arises from the missing wedge; in reality the resolution must be anisotropic, yet the FSC method yields a single overall value for the resolution. Nevertheless, it does provide, at least, a method for comparing the resolution of reconstructed objects. As an example, a tilt series from a porous silica catalyst with Pt/Ru particles (see Section 12.1.1), was used for comparing a SIRT reconstruction with that using conventional weighted backprojection methods. The tilt series was split into two (77 images in each) by picking alternate projections. Figure 11.7 indicates that the overall resolution of the object was the same regardless of the reconstruction algorithm used but below the resolution limit the SIRT trace is consistently higher than that of the weighted backprojection, which corresponds to a higher SNR and indicates a better reconstruction quality over these medium-to-high spatial frequencies.

11.3.6 *Segmentation*

The visualization of reconstructed tomograms is described in Chapter 3 of this book but the need for clear rendering of 3D data sets is as true in electron tomography as with any other form of tomography. In particular, segmentation is widely used, with manual segmentation being common practice, an example can be seen in the next chapter in Fig. 12.3. A robust and successful automatic, or semi-automatic, segmentation method would be of great value and more advanced approaches to segmentation have recently been demonstrated (Frangakis and Forster, 2004) using eigenvector analysis. Local approaches to segmentation usually rely on the detection of changes in volumetric intensity due to some

feature of interest. The earliest application of local segmentation to electron tomography was the use of watershed transforms (Volkmann, 2002). Gradient vector diffusion approaches have also been used (Bajaj *et al.*, 2003) based on the local differential gradients within the image and are less prone to local irregularities in voxel intensity. However, practical demonstrations of automatic segmentation in electron tomography are very small in number but automation has been used as an initial 'first pass' (Frangakis and Forster, 2004), or as part of a semi-automatic approach to segmentation. However, the high contrast, high SNR seen in images from materials science specimens suggests that automated segmentation should be more successful for these tomograms.

11.3.7 *Quantitative analysis*

There are a number of examples of quantitative volumetric analysis in electron tomography. Recent studies, e.g. Jinnai *et al.* 2000, demonstrated that characteristic lengths in copolymers, measured from electron tomograms, agreed very closely with those derived from small-angle X-ray scattering (SAXS) and yielded insights into the physics behind their growth. Given a segmented volume it is relatively straightforward to measure internal structures, such as the size and distribution of silica inclusions (Ikeda *et al.*, 2004) in carbonaceous materials, the latter correlating well with the bulk resistivity. A 3D analysis of a styrene network measured the mean curvature of a gyroid-like structure that indicated the effects of 'packing frustration' (Jinnai *et al.*, 2000). In heterogeneous catalysis, electron tomography is the only technique that can provide *direct* measurement of the loading (mass of catalyst per unit surface area) and the local porosity (Midgley *et al.*, 2004).

11.4 Bright-field and dark-field electron tomography

11.4.1 *Bright-field tomography*

Although the explosion of interest in electron tomography in materials science has, in part, been driven by the demonstration of alternative imaging methods (Midgley and Weyland, 2003), the most popular imaging mode remains that of bright-field (BF) TEM and the majority of automated acquisition implementations, alignment tools and reconstruction algorithms are designed for BF imaging. In materials science, one of the first demonstrations of electron tomography using BF TEM was the study of block-copolymer systems (Spontak *et al.*, 1988) that led to many similar studies of other polymer systems. BF tomography has been used to examine the structure of heterogeneous catalysts (Koster *et al.*, 2000) and other materials across the spectrum, including, for example, the arrangement and connectivity (Numata *et al.*, 2005) of multiwall carbon nanotubes. Researchers at Intel also demonstrated that BF tomography could be applied to semiconductor devices (Mardinly, 2001), revealing failures in metallic vias that were not visible in a conventional image.

BF image contrast arises from a number of scattering mechanisms, depending on the material. In practice, some 'amplitude contrast' is nearly always present,

and is primarily the result of the elastic scattering of incident electrons with the ensemble of atomic potentials. The intensity of the scattered electrons will increase with specimen thickness and with atomic density and amplitude contrast is often known as mass-thickness contrast. For thin, weakly scattering objects, by defocusing the image, changes to the phase of the incident wave can be converted into intensity variations. Such 'phase contrast' in a BF image depends on the contrast transfer function (CTF) that describes the transfer of information by the objective lens for each spatial frequency; the CTF is a function of low-order aberrations such as defocus, astigmatism and spherical aberration. For weakly scattering objects, such as frozen hydrated specimens in the life sciences, the influence of the CTF can be corrected and the true image, or exit wavefunction, recovered. Such an image can show a linear relationship with the projected thickness of material. However, for strongly scattering crystalline objects, as normally found in materials science, in general the interaction of the electron beam with the specimen is strongly dynamical and the exit wavefunction is not a simple monotonic function of the projected thickness or mass-density. In general, diffraction contrast present in a BF image will vary dramatically, depending on the orientation of the crystal. Such contrast does not satisfy the projection requirement and can lead to serious artefacts in a tomographic reconstruction if not treated properly. Nevertheless, in some situations, reconstructions of crystalline specimens from BF TEM micrographs have been successful, and have revealed important structural details in 3D, see Section 12.1.

A recent study of isolated single-crystal nanomagnets (Friedrich *et al.*, 2005) has shown that diffraction and Fresnel contrast can lead to bright 'rings' in the reconstruction and that restored intensities of real, and test, objects have significant deviations from expected values. In general, it seems that BF tomography can restore the exterior shape (topography) of objects with reasonable accuracy, even if the intensities of the interior of the object are invalid. Another study (Kübel *et al.*, 2005) compared reconstructions of a semiconductor device using both BF TEM and HAADF STEM, see Section 11.5, with the latter technique revealing porosity inside a metallic contact that was invisible in the BF reconstruction. BF TEM tomography can give 3D information on crystalline objects but the results should be interpreted carefully. A further limitation of BF tomography is the insensitivity of BF imaging to small buried objects. A recent study of catalysts with 3-nm pores and subnanometre active particles (Thomas *et al.*, 2004) shows very clearly the difference between BF TEM and HAADF STEM. Figure 11.8 shows this comparison; the relative clarity of the particles in the HAADF STEM image is striking.

11.4.2 *Dark-field (DF) tomography*

Whilst in general diffraction contrast does not satisfy the projection requirement, there is a way in which diffraction imaging can be used for electron tomography. Conventional dark-field (DF) images are formed by selecting electrons that scatter to a particular Bragg reflection, using an objective aperture in the back focal

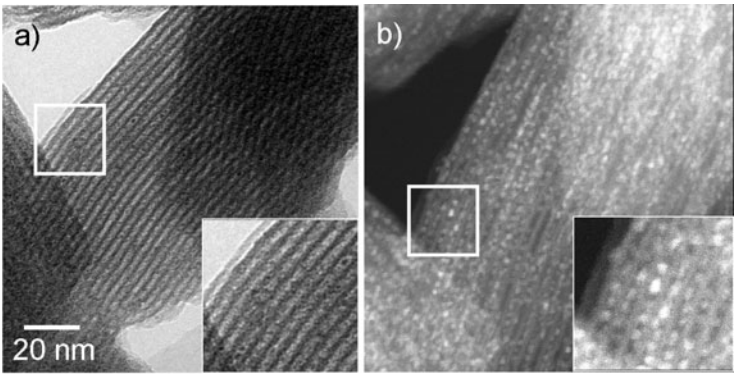


FIG. 11.8. Comparison of (a) BF TEM and (b) HAADF STEM imaging of a mesoporous silica MCM 41 filled with nanoparticle bimetallic particles of $Ru_{10}Pt_2$. The nanoparticles are only seen in the HAADF STEM image. Copyright 2004, Wiley-VCH

plane of the objective lens. Images formed in this way can satisfy, at least partially, the projection requirement, by ensuring the diffraction condition remains constant throughout the tilt series for the object(s) of interest. Examples of this form of tomography will be shown later. In practice, this technique is very difficult to perform and its application may be limited. However, the extreme sensitivity of the technique to very small deviations in crystalline orientation may make DF TEM tomography a valuable technique in certain systems. As a TEM analogue of HAADF STEM (see next section), ADF TEM has also been used to form images suitable for tomography and some success in reconstructing nanostructures has been reported (Bals *et al.*, 2006).

11.5 HAADF STEM tomography

A significant advance in electron tomography for materials science has been the development of HAADF STEM tomography that is now used to study the 3D structures of a range of materials systems and has led to the availability and application of automated acquisition software for STEM tomography. The first application of HAADF STEM tomography was to heterogenous catalysts (Midgley *et al.*, 2001), in which HAADF STEM tomography was used to visualize an ordered mesoporous silica, MCM-41, with 3 nm pore widths, embedded with nanometre-sized Pd_6Ru_6 particles. The large atomic-number difference between the support and the particles led to high image contrast and both the pore geometry and particle position were reconstructed successfully, even from a somewhat limited tilt series. More recent studies on heterogeneous catalysts by HAADF tomography, clearly demonstrate the power of the technique for resolving the structure of heavy active particles embedded in a lighter matrix, see Fig. 11.8. The faceting of biogenic crystallites has also been resolved in 3D in magnetotactic bacteria (Buseck *et al.*, 2001). The high resolution attainable by

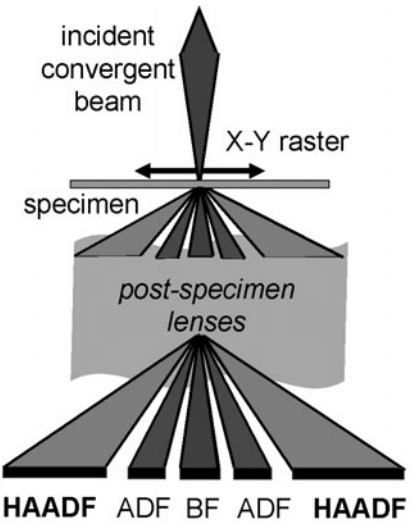


FIG. 11.9. Schematic diagram showing the detector arrangement for STEM imaging. The far-field detectors are for bright-field (BF) imaging, annular dark-field (ADF) imaging and high-angle annular dark-field (HAADF) imaging.

the technique makes it an ideal tool for examining nanoscale materials in 3D, and more recent examples studied by HAADF tomography include quantum dots, nanotubes and optical nanomaterials. The possibility of 3D nanometrology using HAADF STEM has also been suggested (Weyland *et al.*, 2006).

Figure 11.9 shows the range of detectors used for STEM. A conventional annular dark-field (ADF) detector collects intensity from a large number of Bragg spots simultaneously, an effect that averages over the dark-field reflections producing an image that is approximately the inverse of the BF image. In addition, electrons are collected on the ADF detector from high-angle elastic (Rutherford scattering) and thermal diffuse scattering (TDS). Early studies using ADF detectors produced images of sufficient resolution to image small heavy metal clusters, and even single atoms, on light substrates. As all the contrast is dominated by Rutherford-type scattering, the intensity of the image depends on the atomic number, Z. However, such 'incoherent' imaging using ADF was restricted to small crystals or non-crystalline specimens. It was proposed (Howie, 1979) that it should be possible to form incoherent images, even from crystalline specimens, by increasing the inner radius of the ADF detector so as to exclude Bragg-scattered beams. The intensity of images collected with this detector is a function of the projected thickness and the atomic number Z of the scattering atom, approaching a Z^2 relationship for a detector annulus with a large inner radius. In qualitative terms, the HAADF signal can be thought of as the reciprocal of a mass-thickness BF image but at low scattering angles electron screening will

serve to reduce the Z dependence and as a consequence, *high-angle* annular dark-field images show enhanced contrast over equivalent *low-angle* mass-thickness images. The lack of transverse coherence in HAADF images means that such images record a simple projection of the structure in terms of thickness and atomic number and therefore meet the projection requirement for tomographic reconstruction.

Although HAADF STEM imaging is essentially incoherent in nature the intensity of the image can be modified by the orientation of a crystalline specimen. When atomic columns are aligned closely with the beam direction, such as at a major zone axis, strong Bloch-wave channelling tends to concentrate the beam intensity onto the atomic columns as it propagates through the crystal (Hirsch *et al.*, 1977). This increases the relative current density at atomic cores, increasing the high-angle scattering and consequently the intensity in the HAADF image (Pennycook and Nellist, 1999). However, it will be significant only at a small number of orientations and thus in a small number of images within the series. As such, the effect of channelling-enhanced contrast tends to have a limited effect on the tomographic reconstruction.

Two perpendicular sets of electromagnetic scan coils enable the STEM probe to be scanned in a square (x, y) grid pattern that can be mapped to a square array of pixels. However, if the scan coils are inaccurate, there can be angular distortions, where x and y are not perpendicular, and magnitude distortions where x and y scan steps are not identical. Often, this distortion is corrected with reference to a cross-grating and whilst this is reasonably accurate, the application of a more advanced, preferably automated, procedure will be required for metrology applications.

For crystalline specimens care should be taken to operate the instrument in HAADF mode with a sufficiently high collection angle $\approx 40\,\text{mrad}$ at $200\,\text{kV}$; this can be easily adjusted by changing the camera length. For amorphous or nanocrystalline materials, there will be no significant diffraction contribution to the contrast. As the intensity falls rapidly with increasing scattering angle it is more dose efficient to use a lower inner angle in order to maximize the signal collected on the annular detector.

The STEM probe is scanned, point-by-point, in a grid pattern, at each point waiting (the dwell time) whilst the signal is recorded on the detector. While the current density in a STEM probe is very high, the total current delivered to the specimen is similar to that for parallel TEM illumination (Rez, 2003). Therefore, for an equivalent dose, the STEM HAADF image will take longer to acquire than BF TEM because the number of electrons collected on the high-angle detector is only a small fraction of those that are incident on the specimen. Whilst the contrast achieved by collecting at high angles is considerable, the acquisition time must be sufficiently long to overcome background noise.

The short focal depth of STEM can be exploited for automated focusing. A defocus series should show a clear trend in the sharpness and contrast of the image and enable the optimum focus to be attained. For indistinct or noisy

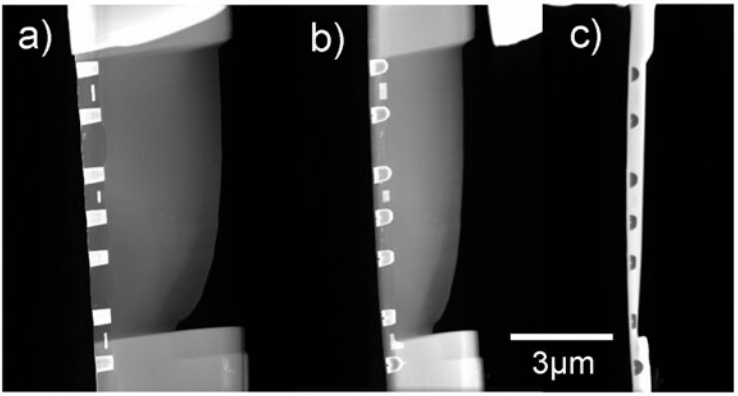


FIG. 11.10. The effect of thickness and high atomic number on HAADF STEM contrast. A focused ion beam (FIB) milled 'finger' specimen of a Si-based transistor with tungsten (W) contacts. a) At low tilt ($\approx 0°$) the W contacts appear bright, as they are of higher atomic number than silicon. (b) and (c) As the projected thickness increases the contrast reverses, so that by 74° in (c) the contacts appear dark and the silicon bright; the series of images no longer meets the projection requirement. Rafal Dunin-Borkowski and Simon Newcomb are acknowledged for the specimen preparation. Copyright 2005, Elsevier.

features, a contrast-enhancing filter may be used. Autofocusing in this mode will require a series of images, and that combined with the long acquisition times in STEM, results in a significant dose 'overhead' for STEM autofocusing.

If the specimen is slab-like and tilted to high angles, it is likely that only part of the specimen will be in focus. By rotating the scan, the tilt axis of the specimen can be made perpendicular to the direction of the rapid STEM scan. Given the simple geometric relationship between tilt, specimen height and defocus it is possible, for every scan line, to adjust the beam cross-over to match the change in specimen height. A focal ramp can then be applied across the image to minimize problems associated with a limited depth of field; this is known as 'dynamic focusing' and has been used for many years in scanning electron microscopy of tilted surfaces.

The STEM beam will diverge as it propagates through the specimen and lead to a degradation of image resolution (the 'depth of field'). Different heights inside the specimen will therefore be sampled with different resolution and what effect this has on the overall reconstruction has yet to be fully studied. For thick or massive specimens, with large average scattering angles, a significant proportion of scattering may fall outside the outer edge of the detector and lead to contrast reversals, and strong deviation from monotonic behaviour. This is demonstrated in Fig. 11.10 for a focused ion beam (FIB) prepared 'finger' specimen of a semiconductor device. At zero tilt the contrast is as expected, with

tungsten contacts appearing more intense than the silicon substrate. However, at high tilts, and large projected thickness, the contacts appear darker as now the majority of the scattered signal from the tungsten has fallen beyond the outer radius of the detector.

A consequence of long acquisition times is that STEM images can be distorted by specimen drift. Although atomic-resolution STEM lattice images can show significant 'shear' due to drift, at the lower magnification typically used for tomography, such effects are not so severe. The large current density in the STEM probe can result in greater beam damage than in TEM mode, but in specimens where beam heating is the controlling factor, the use of STEM appears to help minimize damage. The total dose to the specimen can be much lower than in conventional TEM imaging and the rastering nature of the scan allows the dissipation of heat (phonons) into surrounding, non-illuminated areas (Egerton *et al.*, 2004).

11.6 EFTEM tomography

In the life sciences, energy filtering is used to generate a 'zero-loss' image, in which all electrons that have undergone significant inelastic scattering are excluded, minimizing chromatic blurring and improving the image contrast (Grimm *et al.*, 1998). However, by making use of characteristic ionization edges at well-defined energy losses, an elemental map (formed from two or three individual images) can be acquired quickly (≈ 1 min); a tilt series of such images enables a three-dimensional elemental distribution map to be reconstruted (Möbus and Inkson, 2001, Weyland and Midgley, 2001). The first examples of this in materials science were in engineering ceramics, where interphases form nanoscale precipitates, and in stainless steels, to reveal the shape of chromium carbides at grain boundaries. A simpler form of EFTEM tomography using the post-edge contrast from a single energy loss (Yamauchi *et al.*, 2003) enabled different phases in a complex polymer to be reconstructed. A similar approach is also used in biological TEM by placing the slit just before the carbon edge, known as 'structure-sensitive imaging', maximizing the contrast from species other than carbon (Egerton, 1996). Further applications of EFTEM tomography include iron/oxygen partitioning in magnetotactic bacteria (Midgley and Weyland, 2003) and the study of FeNi nanoparticles (Weyland *et al.*, 2006).

Conventional energy-filtered transmission electron microscopy (EFTEM) is based on acquiring two or three energy-filtered images near an ionization edge of interest. In general, the contrast observed in an energy-loss image is derived from a combination of inelastic scattering (through changes in composition and electronic structure) and elastic effects (via crystal thickness and orientation). The compositional information from a single energy-loss image may be isolated by generating either a background-subtracted elemental map (from three or more images) or a jump-ratio map (from two images). Taking due account of inelastic cross-sections, both maps will show intensity that is related to the amount of an atomic species at a given pixel and, to a first approximation, EFTEM maps

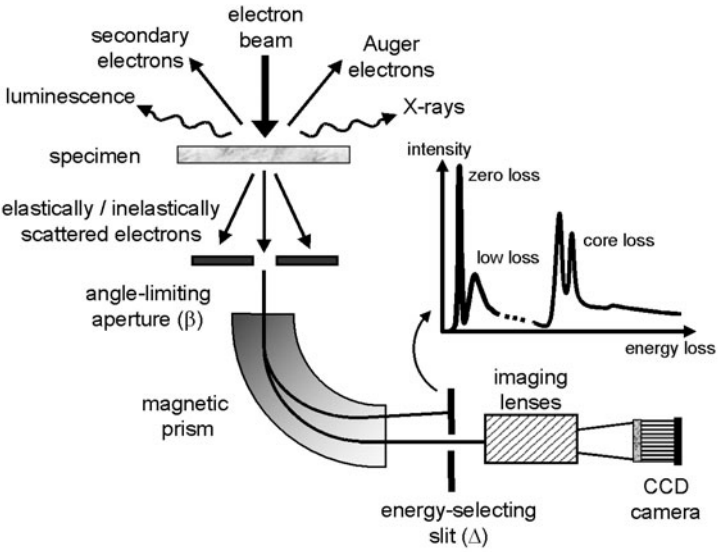


FIG. 11.11. Schematic diagram of energy-filtered TEM (EFTEM).

fulfil the projection requirement for tomographic reconstruction. A schematic of the EFTEM technique is shown in Fig. 11.11 for a post-column imaging filter; in-column filters, such as the omega filter, are also used for EFTEM (Egerton, 1996).

For elemental maps, the conventional three-window approach takes little account of diffraction contrast that may affect the validity of the projection assumption, as it would for a BF image. Such contrast can be minimized by dividing the elemental map by the zero-loss image, but this can introduce further artefacts due to changes in image resolution as a function of energy loss (Hofer *et al.*, 1997). Jump-ratio images are an ideal way of removing residual diffraction contrast in the energy-loss images, and although showing apparently higher sensitivity than that of an elemental map, image intensity values are not quantitative. The jump-ratio increases with thickness, in an approximately linear fashion up to a value comparable to the overall inelastic mean free path, λ. Beyond this thickness, the jump-ratio falls, primarily because the energy-loss background intensity rises faster than the ionization edge signal. Therefore, within the field of view of a single projection, thicker areas may give rise to a lower jump-ratio value, which invalidates the projection requirement. For a typical value of λ, say 100 to 150 nm, this implies a slab-like specimen should be prepared with a thickness of 40 to 50 nm in order to avoid multiple-scattering artefacts at high tilts. For some specimens therefore, neither an elemental map nor a jump-ratio image will show contrast suitable for tomography. Tomographic reconstructions carried out using both conventional EFTEM and with hollow-cone illumination suggest that diffraction contrast tends to be averaged out over the tilt series and

results primarily in an increase in background intensity in the reconstruction rather than introducing specific artefacts.

A more accurate EFTEM method is to acquire an extended energy-loss series over the energy-loss edges of interest. By using this approach, fully quantitative elemental distribution maps can be generated from any spectral feature within the acquired energy range (Thomas and Midgley, 2001). Tomographic reconstructions from a tilt series of energy-loss images generates a large four-dimensional data set $(x, y, z, \Delta E)$, a technique known as '*4D volume-spectroscopy*'; this has recently been achieved over a low-loss energy range (Gass *et al.*, 2006). The acquisition time of a core-loss energy-loss series, typically tens of minutes, means that a tomography experiment would take a very long time, and result in a very high total dose.

Filtered images are acquired using an in-column or post-column imaging filter by using an energy-selecting slit, of a given energy width, in the dispersion plane. At each tilt a number of energy-loss images, usually two or three at each energy-loss edge, need to be acquired, normally automated using a simple script. At each energy loss the positions for the energy windows will closely control the SNR of the resultant maps (Thomas and Midgley, 2001), and as the quality of the reconstruction will depend on the quality of the projections, these should be chosen with some care.

The resolution of EFTEM is controlled by a number of factors: delocalization, chromatic aberration, spherical aberration and the diffraction limit (Egerton, 1999). In most cases, the dominant term is the chromatic aberration of the objective lens. This leads to a blurring of the filtered image that can be minimized by reducing the slit width or the collection angle (typically the objective aperture) but always at the expense of total signal. A balance must be struck to achieve the optimum 2D EFTEM image and thus in turn the optimum 3D resolution.

The primary challenge for EFTEM tomography is to limit the effects of dose (Egerton, 1999). The rapid decrease in signal with increasing energy loss means that EFTEM tomography can be difficult for beam-sensitive specimens. Despite this, initial studies (Leapman *et al.*, 2004) have successfully applied EFTEM tomography to map phosphorous-containing ribosomes. Fully automated acquisition needs to be developed that allows the use of low-dose, or at least dose-managed, methodologies akin to those commonly used in the life sciences for BF imaging.

11.7 Unconventional modes for electron tomography

11.7.1 *Energy-dispersive X-ray (EDX) mapping*

STEM-based energy-dispersive X-ray (EDX) spectroscopy is an alternative technique for chemical mapping. At each pixel, an X-ray spectrum can be acquired that allows elemental maps to be extracted *post facto*. As the intensity of each pixel in the elemental map is related directly to the amount of the species seen in projection, EDX would initially seem ideal for electron tomography. However,

X-ray collection efficiency is very poor as the solid angle subtended at the detector is small, because of the restrictive geometry of the objective polepiece. As such, a single X-ray map, with good counting statistics, may take ≈ 1 h even for a modest map size (256×256 pixels). For tomographic reconstructions, such maps must be acquired at successive tilts, and thus the total acquisition time for a complete tilt series can be exceptionally long. A second hurdle is to minimize absorption and extraneous scattering from a volume close to the specimen (e.g. a specimen support grid) and conventional holder designs may block a large proportion of the emitted X-rays. Nevertheless, the principle of EDX tomography has been accomplished (Möbus *et al.*, 2003), using a highly modified holder, with an example showing the position and morphology of Y_2O_3 particles in a FeAl alloy. X-ray emission can be increased by either increasing the probe current (but simultaneously increasing the probe size) or by the use of an aberration-corrected STEM that will provide far higher current densities. However, this may lead to increased beam damage. An alternative is to use a 'needle' specimen (Koguchi *et al.*, 2001) that allows full 360° in-holder tilt without shadowing.

11.7.2 *Holographic tomography*

2D offaxis electron holography (e.g. (Midgley, 2001)) can be used to reveal the presence of magnetic fields and electrostatic potentials in and around specimens in the TEM. Combining electron holography with electron tomography has enabled 3D magnetic and electric fields to be reconstructed and visualized. The projection requirement for tomography can be satisfied by a holographic phase image if the contribution to any phase change from diffraction contrast is limited. In general, the reconstructed phase image $\Delta\phi(x, y)$ is

$$\Delta\phi(x, y) = C_E \int V_{\text{cryst}}(x, y, z)dz - \frac{e}{h} \int \int \boldsymbol{B}(x, y) \cdot d\boldsymbol{S},$$

where the first integral is taken along a trajectory parallel to the beam direction z, V_{cryst} is the crystal potential, $\boldsymbol{B}$ is the magnetic flux (or induction) and $\boldsymbol{S}$ is normal to the area mapped out by the trajectories of the electrons going from source to detector through the vacuum and the sample. $\boldsymbol{C}_E$ is a wavelength-dependent constant.

In a general sense, both electrostatic and magnetic components of the phase change are 'projections' and thus using such a signal for tomography is valid. In the case of electrostatic potentials, in the absence of magnetic fields, in the absence of diffraction contrast and in the absence of stray fields outside the specimen, the phase change is directly proportional to a product of projected thickness, t, and the mean inner potential of the specimen, V_0:

$$\Delta\phi(x, y) = C_E V_0(x, y)t(x, y).$$

Progress has been made recently in 3D electrostatic potential mapping primarily because standard reconstruction algorithms can be used directly: the holographic image signal measures potentials, in the case of electrostatics this

is a scalar entity but in magnetic specimens, the potential measured is just one component of a vector potential, $\boldsymbol{A}$, which in turn is related through Maxwell's equations to the more familiar magnetic induction, $\boldsymbol{B}$.

11.7.3 *Confocal STEM*

For conventional tomography, as stated in Section 11.2.1, the requirement for projection means that the depth of field should be sufficiently large so as not to give rise to a significant focal change between objects at the top surface and those at the bottom surface. For high-resolution STEM tomography ($\approx 1\,\mathrm{nm}^3$), this limits the reconstruction volume to $\approx 100\,\mathrm{nm}^3$. In STEM, it is possible to increase the depth of field by using a small convergence angle at the expense firstly of beam current, and image signal, and ultimately of the probe size that will become diffraction limited. An alternative is to consider the opposite scenario, where a large convergence angle will lead to a very small depth of field. In current aberration-corrected instruments (Haider *et al.*, 1998, Krivanek *et al.*, 1999), convergence angles as high as 30 mrad can be used at present without significantly degrading the probe size. (In uncorrected instruments, typical convergence angles are ≈ 10 mrad. Large convergence angles lead to very high beam current densities and hence high image SNR. The depth of field is proportional to the inverse square of the convergence angle and so, in principle, depths of field less than 1 nm could be achieved in an advanced aberration-corrected STEM.

Recent research (Frigo *et al.*, 2002, van Benthem *et al.*, 2005) demonstrates how this limited depth of field can actually be used in a practical fashion. With the scanned STEM probe highly localized in the depth direction, the image can be considered as an 'optical section' of the specimen, analogous to that used in confocal optical microscopy. Confocal STEM has been used to determine the positions of individual hafnium atoms in a SiO_2 gate dielectric to within 2 nm depth. As the localization is improved through new aberration correctors and even higher convergence angles, it is certain that this technique will be used more widely to study interfaces and defects in 3D at near-atomic resolution and may attain similar performance to that of three-dimensional atom probes (Blavette *et al.*, 1993).

11.7.4 *Atomic-resolution tomography*

The aim of many in electron tomography is to achieve 3D atomic resolution. To do this, one approach is to use confocal STEM as described above, the second is to use a more conventional projection technique. Using the Crowther criterion, to achieve atomic resolution (say 0.2 nm) in a particle 50 nm in diameter would require ≈ 800 images in the tilt series! In practical terms, this is not achievable. However, if the particle is crystalline (and thus has a periodic lattice of atoms) it may be possible to achieve a reconstruction of the lattice from far fewer projections by imposing additional constraints on the reconstruction, e.g. atoms can be regarded as identical spherical objects and the lattice has internal space group symmetry, constraints similar in many ways to the original Fourier

synthesis approach (Perutz *et al.*, 1960). To maximize the data in each series, individual projections should be recorded at low-order zone axes and high tilts are needed to explore different crystallographic orientations. 'Needle' specimens and free-standing pillars may offer the best geometry for this task. Although standard reconstruction methods, as used for conventional tomography can be used, new algorithms are required to take full advantage of the discrete nature of the data. A new approach is to apply the mathematics of discrete systems (Batenburg, 2005), as described in Chapter 2.

11.8 References

Bajaj, C., Yu, Z.Y., and Auer, M. (2003). Volumetric feature extraction and visualization of tomographic molecular imaging. *Journal of Structural Biology*, **144**, 132.

Bals, S., Van Tendeloo, G., and Kisielowski, C. (2006). A new approach for electron tomography: Annular dark-field transmission electron microscopy. *Advanced Materials*, **18**, 892.

Batenburg, K.J. (2005). A new algorithm for 3D binary tomography. *Electronic Notes in Discrete Mathematics*, **20**, 247.

Berriman, J., Bryan, R.K., Freeman, R., and Leonard, K.R. (1984). Methods for specimen thickness determination in electron-microscopy. *Ultramicroscopy*, **13**, 351.

Blavette, D., Bostel, A., Sarrau, J.M., Deconihout, B., and Menand, A. (1993). An atom-probe for 3-dimensional tomography. *Nature*, **363**, 432.

Bottcher, B., Wynne, S.A., and Crowther, R.A. (1997). Determination of the fold of the core protein of hepatitis B virus by electron cryomicroscopy. *Nature*, **386**, 88.

Brandt, S., Heikkonen, J., and Engelhardt, P. (2001). Automatic alignment of transmission electron microscope tilt series without fiducial marker. *Journal of Structural Biology*, **136**, 201.

Buseck, P.R., Dunin-Borkowski, R.E., Devouard, B., Frankel, R.B., McCartney, M.R., Midgley, P.A., Posfai, M., and Weyland, M. (2001). Magnetite morphology and life on Mars. *Proceedings of the National Academy of Sciences of the USA*, **98**, 13490.

Carazo, J. (1992). Fidelity of 3D reconstructions. In *Electron tomography. Three-dimensional imaging with the transmission electron microscope* (ed. J. Frank), pp. 117. Plenum Press, New York; London.

Crowther, R.A., de Rosier, D.J., and Klug, A. (1970). The reconstruction of a three-dimensional structure from projections and its application to electron microscopy. *Proceedings of the Royal Society London A*, **319**, 317.

Dierksen, K., Typke, D., Hegerl, R., and Baumeister, W. (1993). Towards automatic electron tomography.2. implementation of autofocus and low-dose procedures. *Ultramicroscopy*, **49**, 109.

Egerton, R.F. (1996). *Electron energy loss spectroscopy in the electron microscope* (2^{nd} edn). Plenum Press, New York.

Egerton, R.F. (1999). Spatial resolution of nanostructural analysis by electron energy-loss spectroscopy and energy-filtered imaging. *Journal of Electron Microscopy*, **48**, 711.

Egerton, R.F., Li, P., and Malac, M. (2004). Radiation damage in the TEM and SEM. *Micron*, **35**, 399.

Fernandez, J.J., Li, S., and Crowther, R.A. (2006). CTF determination and correction in electron cryotomography. *Ultramicroscopy*, **106**, 587.

Frangakis, A.S. and Forster, F. (2004). Identification of macromolecular complexes in cryoelectron tomograms of phantom cells. *Current Opinion in Structural Biology*, **14**, 325.

Frank, J. (1981). The role of correlation techniques in computer image processing. In *Computer processing of electron microscope images* (ed. P. Hawkes), pp. 187. Springer-Verlag, Berlin; New York.

Frank, J. and McEwen, B.F. (1992). Alignment by cross-correlation. In *Electron tomography. Three-dimensional imaging with the transmission electron microscope* (ed. J. Frank), pp. 205. Plenum Press, New York; London.

Friedrich, H., McCartney, M.R., and Buseck, P.R. (2005). Comparison of intensity distributions in tomograms from BF TEM, ADF STEM, HAADF STEM, and calculated tilt series. *Ultramicroscopy*, **106**, 18.

Frigo, S.P., Levine, Z.H., and Zaluzec, N.J. (2002). Submicron imaging of buried integrated circuit structures using scanning confocal electron microscopy. *Applied Physics Letters*, **81**, 2112.

Fung, J.C., Liu, W.P., de Ruijter, W.J., Chen, H., Abbey, C.K., Sedat, J.W., and Agard, D.A. (1996). Toward fully automated high-resolution electron tomography. *Journal of Structural Biology*, **116**, 181.

Gass, M.H., Koziol, K.K., Windle, A.H., and Midgley, P.A. (2006). 4-dimensional spectral-tomography of carbonaceous nano-composites. *Nano Letters*, **6**, 376.

Grimm, R., Typke, D., and Baumeister, W. (1998). Improving image quality by zero-loss energy filtering: quantitative assessment by means of image cross-correlation. *Journal of Microscopy*, **190**, 339.

Guckenberger, R. (1982). Determination of a common origin in the micrographs of tilt series in three-dimensional electron microscopy. *Ultramicroscopy*, **9**, 167.

Haider, M., Rose, H., Uhlemann, S., Schwan, E., Kabius, B., and Urban, K. (1998). A spherical-aberration-corrected 200 kV transmission electron microscope. *Ultramicroscopy*, **75**, 53.

Hawkes, P.W. (1992). The electron microscope as a structure projector. In *Electron tomography. Three-dimensional imaging with the transmission electron microscope* (ed. J. Frank), pp. 17. Plenum Press, New York; London.

Henderson, R., Baldwin, J.M., Ceska, T.A., Beckman, E., Zemlin, F., and Downing, K. (1990). A model for the structure of bacteriorhodopsin based on high resolution electron cryomicroscopy. *Journal of Molecular Biology*, **213**, 899.

Hirsch, P.B., Howie, A., Nicholson, R.B., Pashley, D.W., and Whelan, M.J. (1977). *Electron microscopy of thin crystals* (2nd edn). Krieger, New York.

Hobbs, L.W. (1979). Radiation effects in analysis of inorganic specimens by TEM. In *Introduction to analytical electron microscopy* (2nd edn) (ed. J. Hren, J. Goldstein, and D. Joy), pp. 437. Plenum, New York.

Hofer, F., Grogger, W., Kothleitner, G., and Warbichler, P. (1997). Quantitative analysis of EFTEM elemental distribution images. *Ultramicroscopy*, **67**, 83.

Hoppe, W., Langer, R., Knesch, G., and Poppe, C. (1968). Proteinkristallstrukturanalyse mit elektronenstrahlen. *Naturwissenschaften*, **55**, 333.

Howie, A. (1979). Image contrast and localised signal selection techniques. *Journal of Microscopy*, **117**, 11.

Ikeda, Y., Katoh, A., Shimanuki, J., and Kohjiya, S. (2004). Nano-structural observation of in situ silica in natural rubber matrix by three dimensional transmission electron microscopy. *Macromolecular Rapid Communications*, **25**, 1186.

Ishitani, T. and Yaguchi, T. (1996). Cross-sectional sample preparation by focused ion beam: A review of ion-sample interaction. *Microscopy Research and Technique*, **35**, 320.

Jinnai, H., Nishikawa, Y., Spontak, R.J., Smith, S.D., Agard, D.A., and Hashimoto, T. (2000). Direct measurement of interfacial curvature distributions in a bicontinuous block copolymer morphology. *Physical Review Letters*, **84**, 518.

Koguchi, M., Kakibayashi, H., Tsuneta, R., Yamaoka, M., Niino, T., Tanaka, N., Kase, K., and Iwaki, M. (2001). Three-dimensional STEM for observing nanostructures. *Journal of Electron Microscopy*, **50**, 235.

Koster, A.J., Bos, A. Van Den, and Mast, K.D. Van Der (1987). An autofocus method for a TEM. *Ultramicroscopy*, **21**, 209.

Koster, A.J., Verkleij, U. Ziese A.J., Janssen, A.H., and de Jong, K.P. (2000). Three-dimensional electron microscopy: A novel imaging and characterization technique with nanometer scale resolution for materials science. *Journal of Physical Chemistry B*, **104**, 9368.

Krivanek, O., Dellby, N., and Lupini, A. (1999). Towards sub-angstrom electron beams. *Ultramicroscopy*, **78**, 1.

Kübel, C., Voigt, A., Schoenmakers, R., Otten, M., Su, D., Lee, T.C., Carlsson, A., and Bradley, J. (2005). Recent advances in electron tomography: TEM and HAADF-STEM tomography for materials science and semiconductor applications. *Microscopy and Microanalysis*, **11**, 378.

Lawrence, M.C. (1992). Least-squares method of alignment using markers. In *Electron tomography. Three-dimensional imaging with the transmission electron microscope* (ed. J. Frank), pp. 197. Plenum Press, New York; London.

Leapman, R.D., Fiori, C.E., Gorlen, K.E., Gibson, C.C., and Swyt, C.R. (2004). Three-dimensional distributions of elements in biological samples by energy-filtered electron tomography. *Ultramicroscopy*, **100**, 115.

Liu, Y., Penczek, P.A., McEwen, B.F., and Frank, J. (1995). A marker-free alignment method for electron tomography. *Ultramicroscopy*, **58**, 393.

Mardinly, J. (2001). The effect of Moore's Law on the growing role of transmission electron microscopy in the semiconductor industry. In *Proceedings of the Microscopy and Microanalysis Conference* (ed. G. Bailey), Long Beach, California, pp. 510. Springer.

Mastronarde, D.N. (2005). Automated electron microscope tomography using robust prediction of specimen movements. *Journal of Structural Biology*, **152**, 36.

Midgley, P.A. (2001). An introduction to off-axis electron holography. *Micron*, **32**, 167.

Midgley, P.A., Thomas, J.M., Laffont, L., Weyland, M., Raja, R., Johnson, B.F.G., and Khimyak, T. (2004). High-resolution scanning transmission electron tomography and elemental analysis of zeptogram quantities of heterogeneous catalyst. *Journal of Physical Chemistry B*, **108**, 4590.

Midgley, P.A. and Weyland, M. (2003). 3D electron microscopy in the physical sciences: the development of Z-contrast and EFTEM tomography. *Ultramicroscopy*, **96**, 413.

Midgley, P.A., Weyland, M., Thomas, J.M., and Johnson, B.F.G. (2001). Z-contrast tomography: a technique in three-dimensional nanostructural analysis based on Rutherford scattering. *Chemical Communications*, **10**, 907.

Möbus, G., Doole, R.C., and Inkson, B.J. (2003). Spectroscopic electron tomography. *Ultramicroscopy*, **96**, 433.

Möbus, G. and Inkson, B.J. (2001). Three-dimensional reconstruction of buried nanoparticles by element-sensitive tomography based on inelastically scattered electrons. *Applied Physics Letters*, **79**, 1369.

Numata, M., Asai, M., Bae, K. Kaneko A.H., Hasegawa, T., Sakurai, K., and Shinkai, S. (2005). Inclusion of cut and as-grown single-walled carbon nanotubes in the helical superstructure of schizophyllan and curdlan (ss-1,3-glucans. *Journal of the American Chemical Society*, **127**, 5875.

Penczek, P.A. (2002). Three-dimensional spectral signal-to-noise ratio for a class of reconstruction algorithms. *Journal of Structural Biology*, **138**, 34.

Pennycook, S.J. and Nellist, P.D. (1999). Z-contrast scanning transmission electron microscopy. In *Impact of electron and scanning probe microscopy on materials research* (ed. D. Rickerby), Dordrecht, pp. 161. Kluwer.

Perutz, M.F., Rossmann, M.G., Cullis, A.F., Muirhead, H., and North, A.C.T. (1960). Structure of haemoglobin: A three-dimensional Fourier synthesis at 5.5 Å resolution, obtained by X-ray analysis. *Nature*, **185**, 416.

Radermacher, M. (1992). Weighted back-projection methods. In *Electron tomography. Three-dimensional imaging with the transmission electron microscope* (ed. J. Frank), pp. 91. Plenum Press, New York; London.

Radermacher, M. and Hoppe, W. (1980). Properties of 3-D reconstruction from projections by conical tilting compared to single-axis tilting. In *Proceedings of the 7th European Congress on Electron Microscopy* (ed. P. Brederoo and

G. Boom), Den Haag, pp. 132. Seventh European Congress on Electron Microscopy Foundation.

Radermacher, M., Wagenknecht, T., Verschoor, A., and Frank, J. (1987). Three-dimensional reconstruction from a single-exposure, random conical tilt series applied to the 50s ribosomal subunit of escherichia coli. *Journal of Microscopy*, **146**, 112.

Renken, C. and McEwen, B. (2003). Markerless alignment: Bridging the gap between theory and practice. *Microscopy and Microanalysis*, **9**, 1170.

Rez, P. (2003). Comparison of phase contrast transmission electron microscopy with optimized scanning transmission annular dark field imaging for protein imaging. *Ultramicroscopy*, **96**, 117.

Rosier, D.J. De and Klug, A. (1968). Reconstruction of three dimensional structures from electron micrographs. *Nature*, **217**, 130.

Russ, J.C. (1995). *The image processing handbook* (2^{nd} edn). CRC Press, London.

Skoglund, U. and Ofverstedt, L. (1996). Maximum-entropy three-dimensional reconstruction with deconvolution of the contrast transfer function: A test application with adenovirus. *Journal of Structural Biology*, **117**, 173.

Spontak, R.J., Williams, M.C., and Agard, D.A. (1988). 3-dimensional study of cylindrical morphology in a styrene butadiene styrene block copolymer. *Polymer*, **29**, 387.

Thomas, J.M., Midgley, P.A., Yates, T.J.V., Barnard, J.S., Raja, R., Arslan, I., and Weyland, M. (2004). The chemical application of high-resolution electron tomography: Bright field or dark field? *Angewawandte Chemie International Edition*, **43**, 6745.

Thomas, P.J. and Midgley, P.A. (2001). Image-spectroscopy - I. the advantages of increased spectral information for compositional EFTEM analysis. *Ultramicroscopy*, **88**, 179 and 187.

Unser, M., Trus, B.L., and Steven, A.C. (1987). A new resolution criterion based on spectral signal-to-noise ratios. *Ultramicroscopy*, **23**, 39.

van Benthem, K., Lupini, A.R., Kim, M., Baik, H.S., Doh, S.J., Lee, J.H., Oxley, M.P., Findlay, S.D., Allen, L.J., Luck, J.T., and Pennycook, S.J. (2005). Three-dimensional imaging of individual hafnium atoms inside a semiconductor device. *Applied Physics Letters*, **87**, 34104.

van Heel, M. and Harauz, G. (1986). Resolution criteria for 3-dimensional reconstruction. *Optik*, **73**, 119.

Volkmann, N. (2002). A novel three-dimensional variant of the watershed transform for segmentation of electron density maps. *Journal of Structural Biology*, **138**, 123.

Weyland, M., Laffont, L., and Midgley, P.A. (2004). Nanoscale analysis of three-dimensional structures by electron tomography. *Institute of Physics Conference Series*, **179**, 349.

Weyland, M. and Midgley, P.A. (2001). 3D-EFTEM: tomographic reconstruction from tilt series of energy loss images. *Institute of Physics Conference*

Series, **161**, 239.

Weyland, M. and Midgley, P.A. (2003). Extending energy-filtered transmission electron microscopy (eftem) into three dimensions using electron tomography. *Microscopy and Microanalysis*, **9**, 542.

Weyland, M., Yates, T.J.V., Dunin-Borkowski, R.E., Laffont, L., and Midgley, P.A. (2006). Nanoscale analysis of three-dimensional structures by electron tomography. *Scripta Materialia*, **55**, 29.

Winkler, H. and Taylor, K.A. (2006). Accurate marker-free alignment with simultaneous geometry determination and reconstruction of tilt series in electron tomography. *Ultramicroscopy*, **106**, 240.

Yamauchi, K., Takahashi, K., Hasegawa, H., Iatrou, H., Hadjichristidis, N., Kaneko, T., Nishikawa, Y., Jinnai, H., Matsui, T., Nishioka, H., Shimizu, M., and Fukukawa, H. (2003). Microdomain morphology in an ABC 3-miktoarm star terpolymer: A study by energy-filtering TEM and 3D electron tomography. *Macromolecules*, **36**, 6962.

Ziese, U., Janssen, A.H., Murk, J.L., Geerts, W.J.C., der Krift, T. Van, Verkleij, A.J., and Koster, A.J. (2002). Automated high-throughput electron tomography by pre-calibration of image shifts. *Journal of Microscopy*, **205**, 187.

12

APPLICATIONS OF ELECTRON TOMOGRAPHY

Paul A. Midgley and *Matthew Weyland* (12.1)
Heiko Stegmann (12.2)

12.1 Applications in materials research

12.1.1 *Heterogeneous catalysts*

The performance of a heterogeneous catalyst depends on the surface area available for the reaction and the density of active sites. Some of the most effective modern catalysts are based on mesoporous supports with pore diameters in the range of 1 to 50 nm, which offer very high specific surface areas, and can accommodate the movement of reactant and product molecules. A high density of active particles, spread uniformly throughout the support volume is optimal. Electron tomography offers a way to study directly the catalyst's local structure.

The first application of electron tomography to catalysts (Koster *et al.*, 2000) used a BF tilt series to study the positions of catalytic Ag particles in a NaY-type zeolite and the porous system of an acid-leached mordenite. In the first case, the study revealed unambiguously that large silver particles were located on the zeolite surface and smaller ones in the interior. In the second, 2D slices through the reconstruction revealed the internal pores of the mordenite. More recent studies have used HAADF STEM tomography to reveal the structure of both ordered and disordered mesoporous silicas. Ordered mesoporous silicas (Kresge *et al.*, 1992) are high-area solids that enable active and selective nanoparticle bimetallic catalysts (e.g. Pd_6Ru_6, $Ru_{10}Pt_2$) to be anchored on the walls of the support. Despite their enormous potential, the structure of many of these systems is still open to question. Disorder of the structure over micrometre length scales makes X-ray diffraction studies difficult and so conventional high-resolution electron microscopy and diffraction has been used to elucidate their structure. HAADF STEM tomography was used recently (Weyland *et al.*, 2006) to record a series of images about a single tilt axis (of unknown crystallographic direction) to reconstruct the internal structure of a mesoporous silica, MCM-48, which is known to have a complex 3D porous system based on the gyroid structure. After reconstruction, the MCM-48 particle can be reoriented to view down major zone axes, as seen in Fig. 12.1, which shows a montage of voxel projections ('lattice images'), whose zone axis is perpendicular to the $\langle 112 \rangle$ axis. Figure 12.2 shows a reconstructed slice through the MCM-48 reconstruction perpendicular to the $\langle 110 \rangle$ zone axis. The internal structure of the pore system is clearly revealed.

Ordered mesoporous silicas suffer from certain practical disadvantages: they often require a liquid-crystal template that needs to be driven off after growth

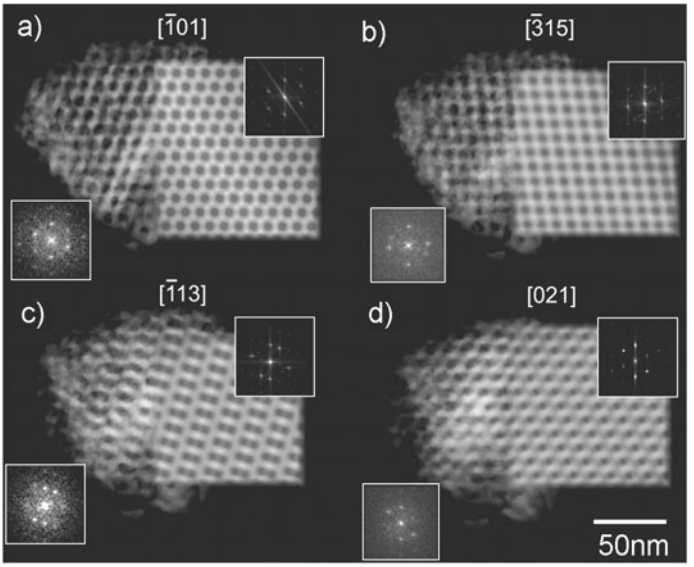


FIG. 12.1. A montage of tomographic voxel projections ('lattice images') of MCM-48 shown at successive major zone axes as the three-dimensional tomographic reconstruction is rotated about a ⟨112⟩ zone axis. The left-hand side of each figure is the experimental image and corresponding power spectrum, the right-hand side the theoretical image and its spectrum, based on the gyroid structure.

and they lack adequate mechanical strength for prolonged use in reactors. As an alternative, disordered mesoporous silicas of high mechanical strength and durability can be produced by simple inexpensive procedures from silicate solutions. Surprisingly little is known about the internal structure of these disordered mesoporous silicas. Conventional gas adsorption-desorption isotherms, X-ray and neutron small-angle scattering techniques (SAXS and SANS) yield only an average value of pore diameter and pore-size distribution and offer little insight into the nature of the connectivity of the pores. Figure 12.3(a) shows a surface render of a tomogram composed of bimetallic $Ru_{10}Pt_2C_2$ particles (red) supported on, and within, the disordered mesoporous silica (coloured grey).

In Fig. 12.3(b) a series of slices, 1 nm thick, are shown to illustrate the complex nanoscale porosity of the silica. The rather 'dendritic' nature of the porosity is readily seen and preliminary results indicate the surface may be fractal in nature.

12.1.2 *Polymers*

The earliest applications of electron tomography in materials science were to block-copolymer systems (Spontak *et al.*, 1988), where their inherent complexity makes tomographic analysis invaluable. The range of structures that can form

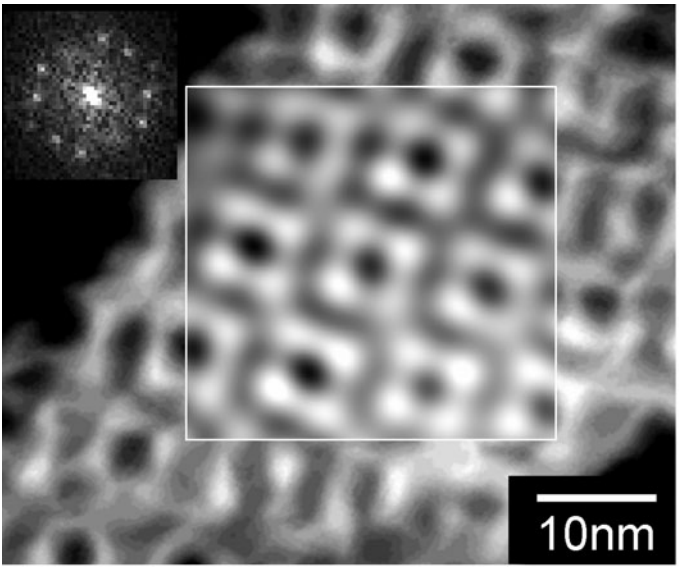


FIG. 12.2. A reconstructed *slice* through a MCM-48 reconstruction, perpendicular to the ⟨110⟩ zone axis. A Bragg-filtered inset enhances the periodic detail. The pairs of spots in the power spectrum correspond to the {332} and {224} reflections. Copyright 2006, Elsevier

in these systems depends on the number of individual monomers and the mutual immiscibility, ranging from simple lamellar to complex gyroid structures. These structures are thought to arise from the competition between attaining the lowest possible mean surface curvature, which minimizes the contact between immiscible phases, and the cost of accommodating that curvature on the packing of molecules (Gruner, 1989). One example of the more complex morphologies studied is the triblock system, styrene-b-isoprene-b-styrene (Laurer *et al.*, 1997). BF tomography can be used successfully with these systems if the polymers are stained to enhance contrast. Reconstructions clearly reveal the complex interlocking gyroid-like structure that arises from two interlocking, non-connected, regions of styrene in the isoprene matrix, see Fig. 12.4. The volumetric nature of the data, and the clear separation between volumetric components allows quantification of various structural parameters. The measured styrene volume fraction, at 0.35, closely matches the expected theoretical value, 0.32, and the significant repeat unit lengths closely match those as measured by SAXS. Deviations in the structure away from a constant mean curvature may arise because of the difficulties of achieving such a curvature by stacking individual monomer units and the limited interfacial area for a given amount of material, known as 'packing frustration'. A quantitative analysis of the curvature of the styrene network (Jinnai *et al.*, 2000) showed a statistically significant deviation from the value of the expected mean curvature, indicating the presence of packing

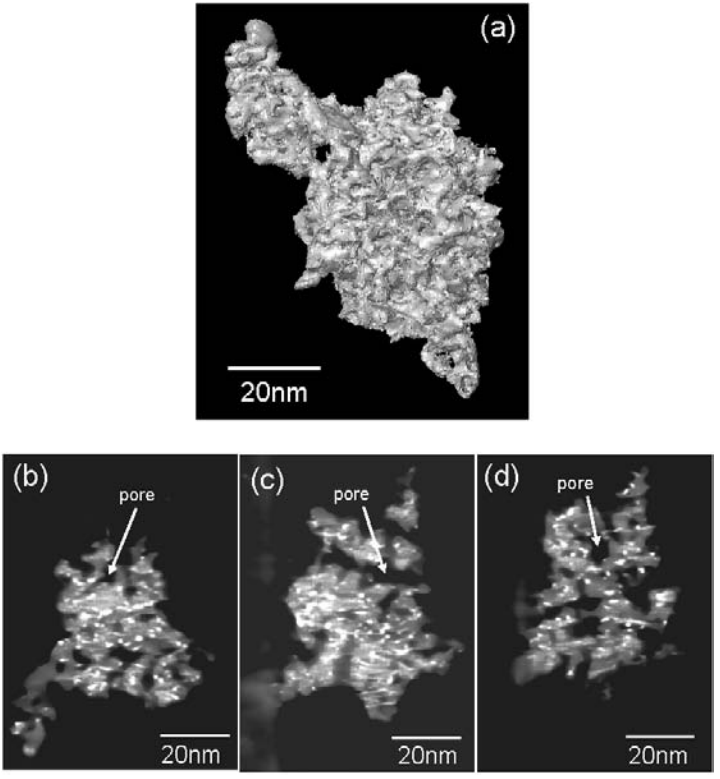


FIG. 12.3. (a) Surface render of a segmented reconstruction, derived from a series of HAADF STEM images, of a heterogeneous catalyst: the mesoporous silica support is shown in grey and the Pt/Ru bimetallic nanoparticles in red. (b)–(d) 1-nm thick tomographic slices, taken every 20 nm, of the reconstruction showing internal porosity and the internal distribution of the bimetallic nanoparticles. Image (a) is reproduced in **colour** in the central part of the book.

frustration in a real system.

12.1.3 *Nanotubes and semiconductor nanostructures*

12.1.3.1 *Carbon nanotubes* Carbon nanotubes, formed of sheets of sp^2 graphene carbon, are of great interest for their high electrical and thermal conductivities, high mechanical strength, oxidation resistance and good field emission performance. Carbon multiwall nanotubes (MWNTs) grown with nitrogen appear to have a constant crystallographic orientation, spacing and register between layers (Koziol *et al.*, 2005). HAADF STEM images should help to minimize the diffraction contrast from these 'single-crystal' nanotubes and are ideal for tomo-

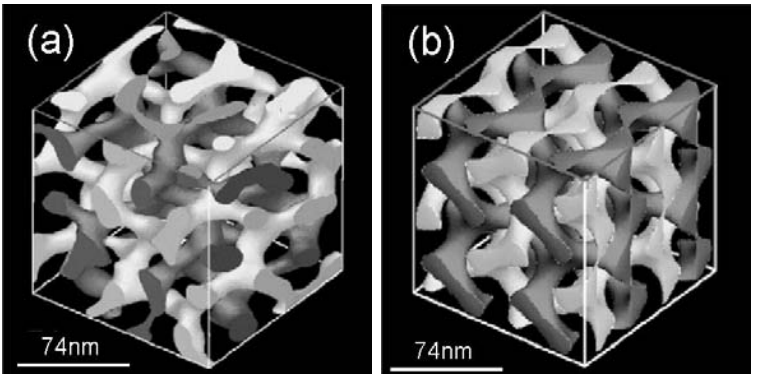


FIG. 12.4. Comparison between (a) an experimental reconstruction and (b) an ideal gyroid structure for an SIS triblock copolymer. The scale bar represents 40 nm. Reproduced from (Jinnai *et al.*, 2000), Copyright 2000, American Physical Society.

graphic reconstruction of the nanotubes' 3D structure. Figure 12.5 shows how the internal architecture of the MWNTs, which is poorly resolved in projection, is clearly revealed in tomographic slices taken parallel to the tube axis. Tomographic slices perpendicular to the tube axis reveal filamentary surfaces across the tube diameter of near-constant thickness that define each 'compartment' of the MWNT.

For many applications, nanotubes are combined with other materials, by attaching functional groups on the inside or outside of the tube, often to enhance the nanotubes' electrical or chemical properties, or through the formation of a composite material to enhance their mechanical properties. An example of the latter, combining a multiwall nanotube with nylon to form a composite structure was investigated recently (Gass *et al.*, 2006). Conventional STEM imaging was not applicable here as the two main components are carbon-based and image contrast between the nylon and the MWNT would be minimal. Instead, EFTEM was used to differentiate the two components. Concerns over the beam stability of the nylon led to the use of the low-loss part of the energy-loss spectrum where acquisition times, and hence the possibility of accumulated beam damage, could be kept to a minimum. It is known that the plasmon-energy maximum varies depending on the different form of carbon (Leder and Suddeth, 1960): amorphous carbon has a plasmon maximum at an energy of 24 eV, while graphitic carbon has a plasmon maximum at an energy of 27 eV (Egerton, 1996). A ratio of these images can give image contrast suitable for electron tomography (Gass *et al.*, 2006) and thus to differentiate between the two components. Figure 12.6 shows a composite of a zero-loss image, which shows little contrast as expected, together with plasmon ratio images 22 eV/28 eV and 28 eV/22 eV, which show clearly the position of the MWNT and its nylon sheath. Figure 12.7 shows surface renders of the nanotube and the nylon viewed from different angles.

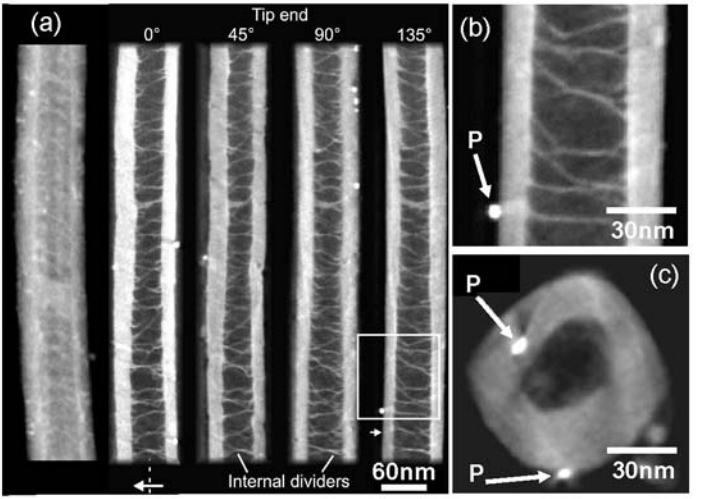


FIG. 12.5. (a) An original HAADF STEM image and four single slices of 2 nm thickness through a carbon multiwall nanotube at 45° intervals, revealing the varying wall thickness and internal divisions. The wall thickness at the two arrows is 27 nm and 47 nm. (b) Blow-up of boxed region. (c) Single slice (2 nm thick) perpendicular to the long axis of the nanotube showing internal and external faceting. Iron catalyst particles are easily identified and examples are indicated by 'P'.

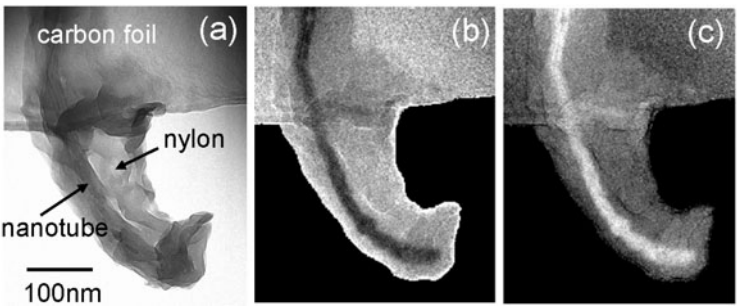


FIG. 12.6. Energy-filtered images from a carbon MWNT–nylon nanocomposite: (a) a zero-loss image: (b) image formed from the ratios of the two plasmon energies 22 eV/28 eV and (c) 28 eV/22 eV.

There is evidence also for nylon within the nanotube cavity and a void, presumably once occupied by a catalyst particle. The nanotube is completely embedded within the nylon, information crucial in understanding the overall integrity of the nanocomposite and predicting its mechanical properties. The curvature of the nanotube has led to a 'pleating' of the nylon on the concave side.

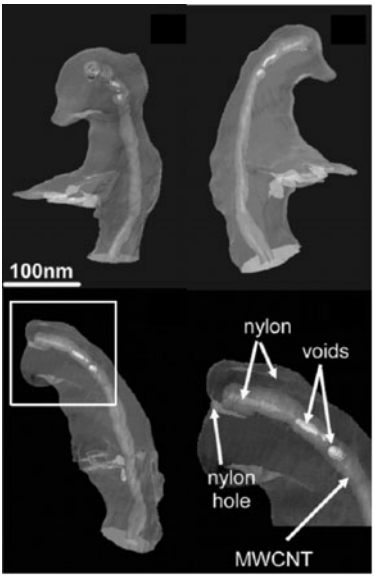


FIG. 12.7. Montage of surface renders of the nanotube (light grey) and nylon (dark grey) from the tomographic reconstruction of the plasmon ratio images in Fig. 12.6. The nylon that has filled the top end of the nanotube is shaded a lighter purple. A hole that occurs in the nylon can be seen running just under the top end of the nanotube and the voids within the nanotube are represented by pale green. The surface renders were obtained by thresholding the reconstructed data to select data from the nylon and the nanotube separately. Copyright 2006, American Chemical Society.

12.1.3.2 *Quantum dots* The electronic and optical properties of semiconductor quantum dots (QDs) may be very different from their bulk counterparts. The 3D morphology of such structures can be critical in defining their unique properties and as such electron tomography may play a key part in unravelling structure–property relationships. Figure 12.8 shows the tomographic reconstruction of predominantly cubic α-tin quantum dots buried within a silicon matrix. A representative plan view illustrates the highly faceted surfaces of the buried dots, with the coloured {111} surfaces being the most evident. A cross-sectional view (X) shows how some of the smaller dots appear to be slightly oblate. This is unlikely to be a reconstruction artefact, as the geometry of the tilt series $\pm 70^\circ$ would, if anything, give rise to a small elongation of the dot (not a reduction) in that direction.

12.1.3.3 *Semiconductor tetrapods – an example of dual-axis tomography* Colloidal quantum dots are being used for a range of applications from sun screens through to imaging labels (Shim and Guyot-Sionnest, 2000). If the growth conditions are altered, the dots can elongate along certain crystallographic directions.

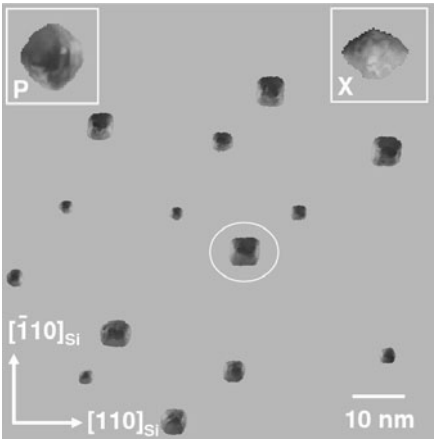


FIG. 12.8. Plan view image of a reconstruction of tin quantum dots buried within a silicon matrix. The {111} facets of the cubic α-Sn crystals is evident, see plan view inset (P) and cross-section (X). The cross-section also demonstrates the oblate nature of some of the smaller dots. The circled dot, with a more elongated morphology, is β-Sn. This image is reproduced in **colour** in the central part of the book. Copyright 2006, Wiley-Blackwell.

For dots with the spahlerite structure, 'legs' grow along (111) directions to form a structure known as a 'tetrapod'. These have been studied by electron tomography to investigate the perfection of the structure and as a test for dual-axis algorithms. Figures 12.9(a,b) show two HAADF STEM images at zero tilt, one from each tilt series with arrows indicating the positions of identical, but rotated, tetrapods in the field of view. Figure 12.9(c) shows a SIRT reconstruction of the first tilt series alone. By comparison to the original images of Figs. 12.9(a,b), it can be seen that certain legs of the tetrapods are missing or are poorly reconstructed, as indicated by the arrows. Here, the tilt axis is parallel to the direction of the arrows and legs that are perpendicular to the axis will be affected by the missing wedge and may not appear in the reconstruction. Figure 12.9(d) is a SIRT reconstruction of the second tilt series that has been rotated to match the orientation of the first. Here, the legs in the perpendicular direction are poorly reconstructed. Using a SIRT algorithm designed for dual-axis tomography (Tong *et al.*, 2006), a reconstruction that combines the two data sets shows that all legs are now present, see Fig. 12.9(e). Whilst dual-axis tomography is undoubtedly time consuming, this work illustrates that there is significant improvement in the fidelity of reconstructions from complex nanoscale objects and may be crucial for high-quality metrology applications.

12.1.4 *Biomaterials*

12.1.4.1 *Magnetotactic bacteria* Magnetotactic bacteria (Blakemore, 1975) have an internal 'backbone' of small magnetic crystals (magnetosomes) within their

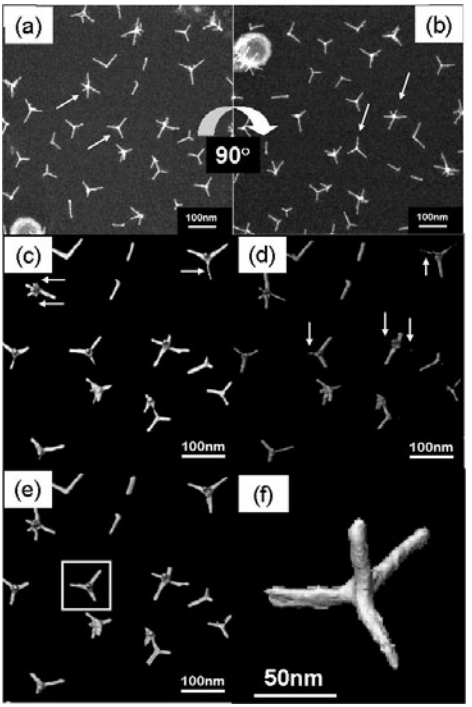


FIG. 12.9. (a), (b) HAADF STEM images of CdTe tetrapods in two perpendicular orientations at 0° tilt. The arrows indicate the location of the same tetrapods in both images; (c) is a reconstruction of a single tilt series and shows that some of the legs of the tetrapods are missing (compared to the original STEM images) as indicated by arrows, due to the effects of the missing wedge; (d) shows a reconstruction of the perpendicular tilt series showing a different set of missing legs, again indicated by arrows; (e) is a dual-axis reconstruction of the two data sets and illustrates that no legs are missing because the missing information has been greatly reduced. The tilt axes in (a) and (b) are parallel to the direction of the arrows; (f) a high magnification image of the tetrapod boxed in (e). Copyright 2006, Elsevier.

cell membrane. Alignment of these single domain nanomagnets allows the bacteria to sense the Earth's magnetic field and find a preferential direction in which to feed. These bacteria are found in bodies of fresh water (the magnetic particles are magnetite, Fe_3O_4) and in seawater (iron sulphide, Fe_3S_4) (Mann *et al.*, 1990). Similar nanomagnets may also play a part in the homing instinct of pigeons (Shcherbakov and Winklhofer, 1999). HAADF STEM tomography was used to examine the morphology of such magnetosomes, which exhibit very high contrast against the organic medium of the bacteria. The example shown here is a strain of magnetotactic bacteria called MV-1 (marine vibrios-1), which con-

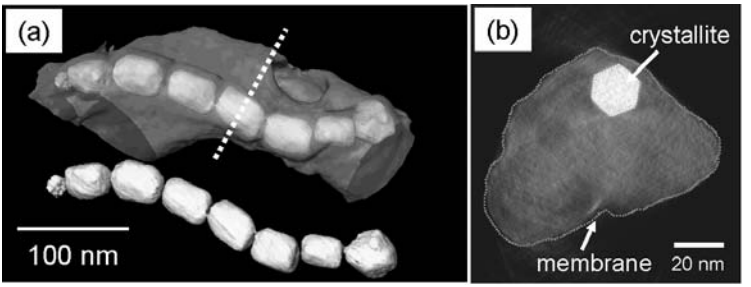


FIG. 12.10. (a) A tomographic reconstruction of magnetotactic bacteria strain MV-1, shown with and without the organic 'envelope'. The faceted chain of magnetite crystals is evident. b) Slices taken through a single magnetite crystal in the chain (dashed line in (a)) which reveals the near-perfect hexagonal shape of the slice through the cubic crystal.

tain small (60 nm) magnetite crystals of regular elongated octahedral shape. A large tilt range, $\pm 76°$, resulted in a high-fidelity reconstruction of the magnetite crystals and the outer membrane, see Fig. 12.10 (Weyland *et al.*, 2006). Slices taken perpendicular to the axis of the crystal chain reveal how HAADF STEM tomography can lead to high-quality information about the three-dimensional shape and faceting of such small particles. The facets were indexed according to the cubic symmetry of the magnetosomes. In general, it shows how accurate measurements can be made at the nanoscale in three dimensions within an electron-transparent medium.

12.1.4.2 *Ferritin* There is growing interest in materials science in the use of biomineralization proteins for the fabrication of inorganic structures for electronic devices, catalysts and biomimetic structures (Mann and Ozin, 1996). The most basic biomineralization processes operate at the nanometre scale and involve macromolecules controlling nucleation and growth of the mineral phase. One of the best studied examples of this type is ferritin, a protein involved with the storage and transport of iron. Ferritin has a hollow shell for the deposition of particles of iron (III), and prevents the uncontrolled growth and coalescence of the mineral particles into larger, insoluble aggregates. The ferritin protein acquires iron (II), catalyses its oxidation, and induces mineralization within its shell. Iron-free ferritin (apoferritin) molecules have inner and outer diameters of about 8 and 13 nm and the protein can accommodate up to 4500 iron atoms. The mineral core of ferritin has no specific shape, but is usually irregular and spherical. The structure of the core is not well established, but the mineral component is thought to be a hydrous ferric oxide with a structure similar to the mineral 'ferrihydrite' ($Fe_2O_3{\cdot}H_2O$).

Unstained human liver samples from a patient with iron overload (haemochromatosis) were examined using HAADF STEM tomography; only the mineral cores of the ferritin particles can be distinguished clearly. A typical HAADF

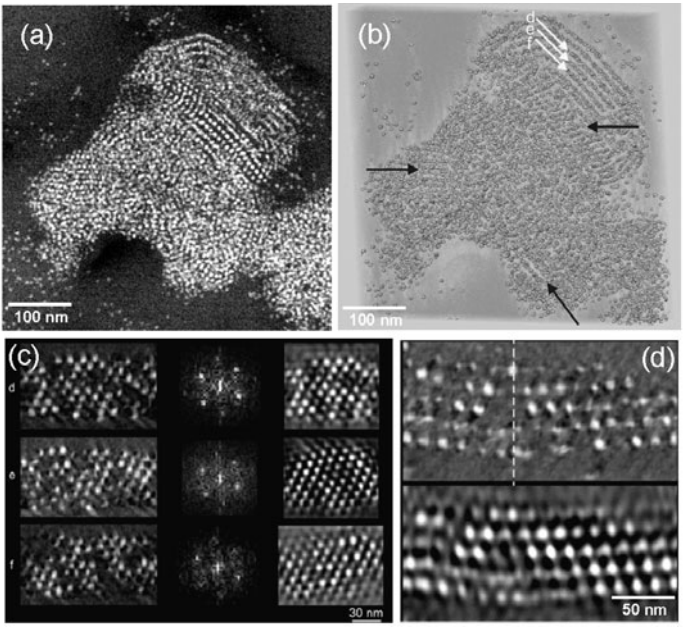


FIG. 12.11. (a) A HAADF STEM image recorded from a liver section from a haemochromatosis patient showing clustering and ordering of ferritin molecules. (b) A view of the reconstructed volume with the ferritin cores surface-rendered in green, and the organic material displayed as a voxel projection in pink. It is oriented to expose areas in which the ferritin is ordered, but with less obvious ordering marked by black arrows. (c) Single slices (nominal thickness 0.8 nm) through the planes of ordered mineral cores labelled 'd', 'e' and 'f' in (b), showing inplane ordering, their power spectra, and corresponding Bragg-filtered slices. (d) Single slice (nominal thickness 0.8 nm) perpendicular to the planes in (c), showing ordering of the ferritin cores, and corresponding Bragg-filtered slice. The dashed line indicates a 'pseudomirror plane'.

STEM image from a tilt series is shown in Fig. 12.11(a) and the resulting tomogram is visualized in Fig. 12.11(b) using a combination of surface rendering, for the mineral cores, and voxel projection, for the low-intensity organic material in which the mineral cores are embedded. A large portion of the object consists of a 'racetrack'-shaped region where the mineral cores appear to be ordered. By tilting the reconstruction, the typical interplane spacing is measured to be 10 to 15 nm. Slices through the reconstruction, such as those in Figs. 12.11(c,d) show that the ferritin is well ordered in all three dimensions; it may be that such packing occurs in response to an overload of particles within a membrane-bound volume, that there is some underlying template on which the ferritin parts can align, or that packing may occur when the chemical environment in cell regions

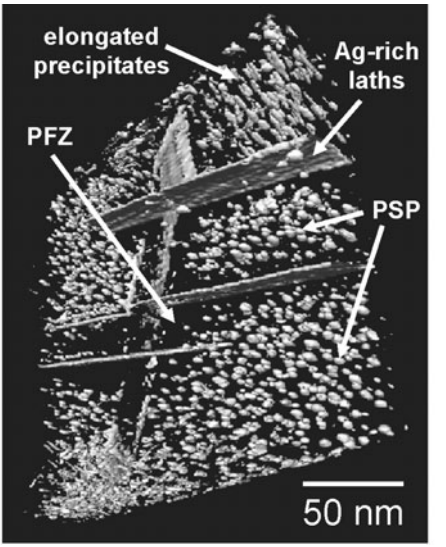


FIG. 12.12. An oblique view of the reconstructed volume from an Ag-Al alloy indicating the presence of Ag-rich laths, precipitate-free zones (PFZ), pseudospherical precipitates (PSP) and elongated precipitates within a shear band.

is different. There is also the possibility that packing is mediated by a magnetic interaction. The mineral cores are roughly spherical, with diameters, measured from the reconstruction, of close to 7 nm. While this is in accordance with the internal diameter of the ferritin protein, the centre-to-centre spacing of the cores in the ordered regions is often as small as 9 nm, indicating a change in the protein shell.

12.1.5 *Metallic nanostructures*

12.1.5.1 *Ag-Al alloys* The Ag-Al system has recently been studied by electron tomography (Inoke *et al.*, 2006) to determine the size and distribution of Ag-rich precipitates within the Al-rich matrix. After annealing, the Ag-rich precipitates are approximately spherical in shape. The alloy was then deformed using a technique called equichannel angular pressing (ECAP), which introduces severe strain. For tomography, STEM HAADF imaging was used to suppress diffraction effects and to enhance contrast between the precipitates and the matrix. 3D reconstructions of the deformed specimen (see Fig. 12.12) show that shear bands introduced by the deformation process consist of ellipsoidal precipitates and neighbouring Ag-rich laths deformed by the cutting action of moving dislocations. Outside the shear bands the precipitates remain spherical in shape, indicating that these regions were unstrained. The movement of the dislocations on active slip planes leads to the precipitates becoming ellipsoidal by the cumulative cutting action of the dislocations. The mean aspect ratio of the ellipsoids

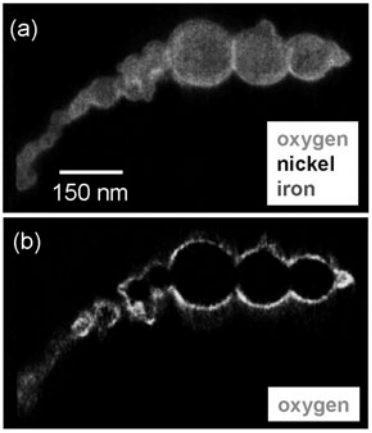


FIG. 12.13. (a) A 2D EFTEM composite image showing the composition of oxidized Ni-Fe nanospheres in projection; (b) A slice through the reconstructed 3D oxygen image proving that the nanospheres sintered before the agglomerate oxidized. This image is reproduced in **colour** in the central part of the book. Copyright 2006, Elsevier.

was measured and an average local strain within the shear band was calculated to be $\gamma = 1.8$.

Note that Ag-Al alloys have also been investigated by magnifying synchrotron tomography, see Section 7.1.4.1. The laths shown in Fig. 7.17 have a similar morphology but are much larger than the ones shown in Fig. 12.12 as they represent a more advanced stage of precipitate growth. This example nicely demonstrates the complementarity between various tomographic methods operating on different length scales.

12.1.5.2 *FeNi nanoparticles* Cryogenically evaporated iron-nickel nanoparticles (Weyland *et al.*, 2006) are spherical superparamagnetic, particles with diameters ranging from 10 to 100 nm. Electron tomography using 3D EFTEM was used to investigate the composition of the oxide shell that forms around the metal nanoparticles and whether the oxide formed before agglomeration or afterwards. A tilt series was acquired from a chain of particles with three energy-loss images acquired for the Fe, Ni and O edges at each tilt. Analysis of the 3D composition of each particle showed that the oxide shell was nickel rich. As Fig. 12.13 shows, a simple 2D map / projection appears to show that oxide exists between the particles, suggesting perhaps oxide growth before agglomeration. However, a 2D slice through the centre of the oxygen tomographic reconstruction, Fig. 12.13(d), clearly indicates that no oxide is present where the nanoparticles touch and therefore they agglomerated and sintered before oxidising. A simple example perhaps, but one that shows the power of EFTEM tomography to elucidate 3D composition.

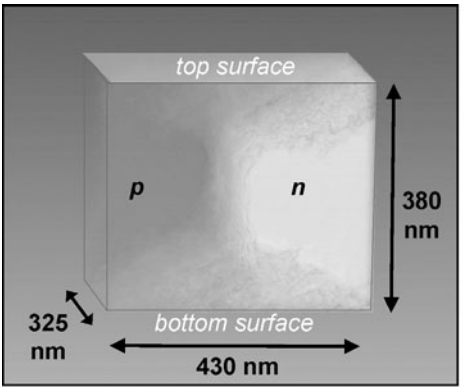


FIG. 12.14. 3D reconstruction of the electrostatic potential near a p-n junction in a silicon device. The 3D potential, reconstructed from a series of offaxis holograms, shows an increasing depletion width near the surface. The colour gradation represents changes in the electrostatic potential across the junction (image courtesy of Dr A.C. Harrison). This image is reproduced in **colour** in the central part of the book.

12.1.6 *3D electrostatic potentials*

As discussed in Section 11.5, it is possible to combine electron holography with electron tomography in order to determine the 3D distribution of electrostatic potentials. In the simplest case, the phase can be directly proportional to the thickness and the mean inner potential. Thus, there is a method of measuring potentials, and in principle dopant concentrations, in 3D. Within a tomographic tilt series however, some phase images will inevitably show strong diffraction contrast and will have to be discarded. Recent work (Twitchett, 2006) has shown that the potential distribution in a device containing a p-n junction can be reconstructed in 3D, and has revealed, quite dramatically, the influence on the device's electronic properties through surface modification by FIB-based sample-preparation methods. A reconstruction is shown in Fig. 12.14.

12.1.7 *3D visualization of defects by dark-field tomography*

There are only rare examples of this type of tomography. The first example, of a Ni-based superalloy, was one in which a superlattice reflection was used to image the cuboidal $L1_2$-ordered γ' precipitates in the $A1$-disordered γ matrix (Kimura *et al.*, 2005). As the γ/γ' two-phase structure is annealed, $A1$-disordered γ particles appear in the γ' precipitates and coarsen into plate- or rod-like shapes. The 3D morphology and distribution of the γ particles were reconstructed by dark-field TEM tomography using a (001) $L1_2$ superlattice reflection. Figure 12.15 shows 3D views of γ/γ' interfaces reconstructed from the dark-field TEM tilt series. The three γ particles are plate-like and approximately parallel to the {100} planes. In another example (Barnard *et al.*, 2006), by using a tilt series

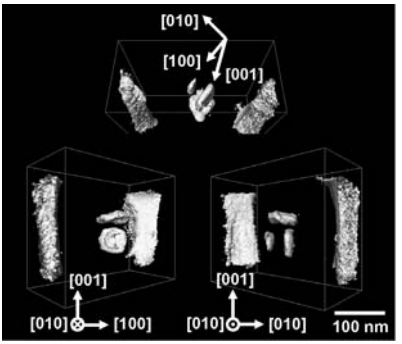


FIG. 12.15. 3D surface render depicting γ/γ' interfaces in a Ni-Al-Ti alloy obtained by dark-field tomography (image courtesy of Dr S. Hata).

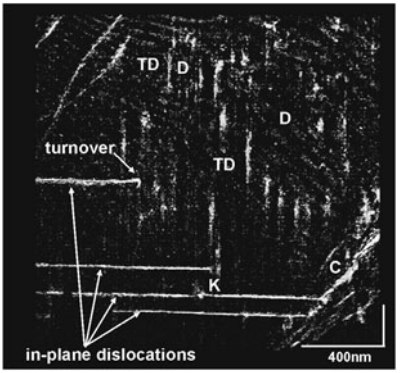


FIG. 12.16. A tomographic reconstruction of an array of dislocations in a GaN epilayer. Threading dislocations align to form low-angle domain (D) boundaries, while threading dislocations close to a crack (C) turn over and glide away leaving an inplane segment that is a pure screw. One of these has interacted directly with another threading dislocation (K). Reproduced from (Barnard *et al.*, 2006).

of weak-beam dark-field images, a high-resolution tomographic reconstruction can be formed that reveals a 3D network of dislocations. Figure 12.16 shows the 3D structure of a dislocation network in a GaN epilayer and dislocations can be seen interacting with crack stresses and other dislocations. The central limitation to dark-field TEM tomography is the need to maintain near-constant diffraction conditions throughout the tilt series and thus careful orientation of the specimen is paramount. The reconstruction of the dislocation network is aided by the tightly defined relationship with the host crystal and the fact that the dislocations at this resolution can be considered as very simple linear structures.

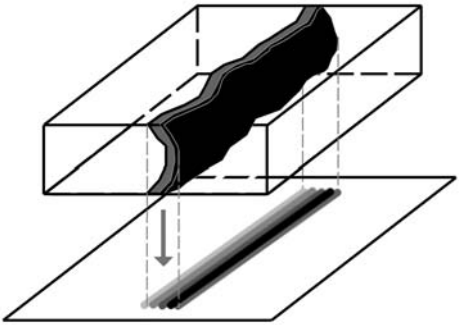


FIG. 12.17. Projection problem. The projection gives only limited information on the morphology of the layer stack at the irregularly curved via sidewall. Taken from Stegmann and Zschech (2003), ©2003, Materials Research Society.

12.2 Applications in semiconductor industry

Driven by the development of high-performance microprocessors, the continuous downscaling of metal oxide semiconductor (MOS) transistor and onchip interconnect sizes has made transmission electron microscopy (TEM) an indispensable technique for metrology and analysis in semiconductor manufacturing. Etched, selectively grown, deposited or implanted structures nowadays have dimensions in the nanometre and even subnanometre range, and only TEM has enough spatial resolution to quantify the smallest structures with sufficient accuracy. However, when feature sizes become smaller than the minimum practicable TEM sample thickness, the amount of information that can be obtained from conventional two-dimensional (2D) TEM images becomes limited and accurate metrology of non-planar structures is hampered. This problem will substantially increase with further decreasing structure size. Three-dimensional (3D) information with nanometre resolution can be obtained by electron tomography. During the past years electron tomography has attracted much interest as a very promising approach for improved metrology of nanoelectronic structures. In this chapter, a review of electron tomography methodology and applications to metrological problems in semiconductor manufacturing is presented.

12.2.1 *Electron tomography in semiconductor manufacturing*

TEM specimen preparation of nanoelectronic structures mostly needs to be highly site specific. Target preparation of particular test structures or of single defects on the chip is mostly done by focused ion beam (FIB) cutting. The final thickness of standard FIB-prepared samples is between 80 nm and 150 nm. Limited by the physical restrictions of this technique, a thickness of less than 40 nm can be reached by optimized preparation (Engelmann *et al.*, 2002). Mechanical thinning combined with low-angle argon ion polishing yields a sample thickness down to 10 nm, but this method is tedious and unreliable for target

preparation.

TEM imaging of different features stacked on top of each other reduces the 3D object structure to a 2D projection that often cannot be interpreted easily. Therefore, only limited information can be obtained from conventional TEM images if the size of the features of interest is much smaller than the minimum practicable specimen thickness. For example, in the Cu interconnect system of recent microprocessors, vertical contacts between the interconnect layers ('vias') currently have less than 100 nm in diameter. They are lined with ultrathin metallization layers with only a few nanometres thickness. The minimum specimen thickness, however, remains limited to some ten nanometres. Even an optimally prepared TEM sample will not only contain a thin slice but also a large part of the via. In the projection, the layer stack will be considerably blurred due to the roughness of the layers and the deviation of the via-wall curvature from the ideal cylindrical form (Fig. 12.17). This problem arises for any non-planar structure that cannot be sectioned to a sufficiently thin specimen. In general, information on the 3D structure of the object will be lost in a 2D image. This 3D information can be retrieved by electron tomography.

For metrology of nanoelectronic structures this approach is very promising. Conventional 2D imaging will be increasingly limited with further decrease of structure size. In contrast, this does not present a limitation for electron tomography since a TEM lamella – being sufficiently transparent to the electron beam – may contain the complete structure of interest, e.g. an entire via. The complete 3D object structure can then be reconstructed, allowing metrology in three dimensions, roughness measurements on subsurface interfaces, characterization of buried defects, and understanding of complex 3D geometries.

12.2.2 *Technical requirements and methodical problems*

Semiconductor manufacturing process development requires the analysis of a very large number of samples. Sometimes, decisions concerning production are based on the results of physical failure analysis. Such results may be required within hours, e.g. when serous yield problems suddenly appear. Therefore, sample throughput and return time are key issues for a TEM laboratory in the semiconductor industry. The number of processed TEM samples may reach a few thousand per year. Since electron tomography consists of several labour- and time-intensive work steps such as specimen preparation, choice of the suitable imaging mode, data acquisition and data processing, it is not yet able to keep up with this pace. Nevertheless, it is possible to obtain good results with reasonable effort if appropriate techniques for the specific metrological challenges are chosen.

12.2.2.1 *Specimen preparation* Conventional site-specific FIB preparation of nanoelectronic structures mostly uses the 'H-bar' approach. A bar of 3 mm length and about 0.2 mm thickness containing the structure of interest is cut from the wafer, ground to a thickness of about 50 µm and glued to a 3-mm Cu half-ring. If open structures are present, e.g. contact holes, they have to be filled with epoxy

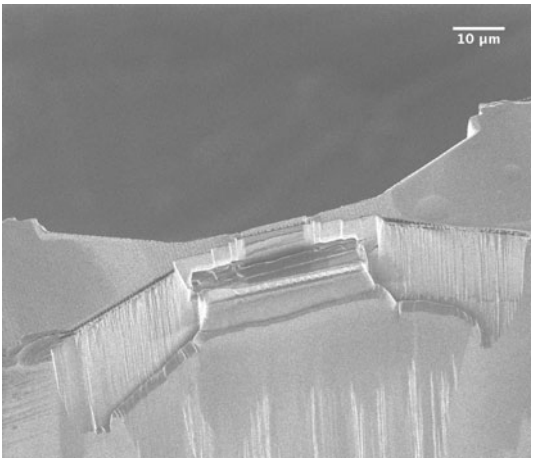
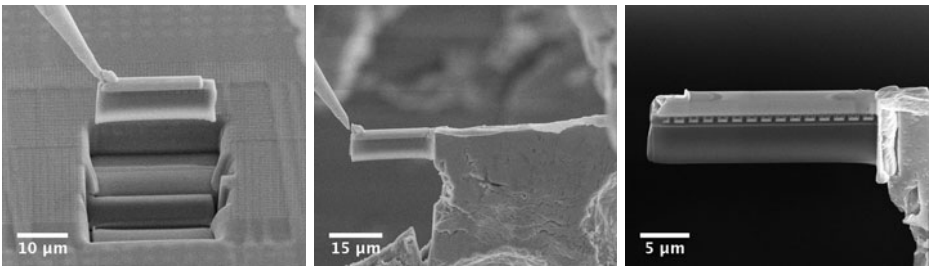


FIG. 12.18. FIB image of a specimen for TEM tomography. The lamella was FIB cut into a bar that was sawn from a wafer. Taken from Stegmann and Zschech (2003), ©2003, Materials Research Society.

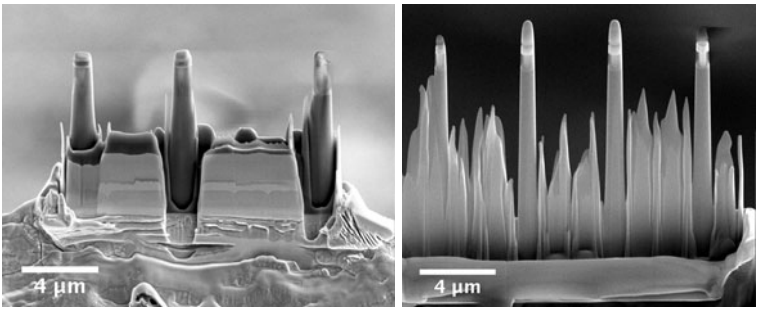


(a) (b) (c)

FIG. 12.19. FIB images of a lift-out preparation procedure. (a) Lifting out the FIB-cut lamella with a W needle. (b) The needle is removed after attaching the lamella to a sample holder grid. (c) View of the finished TEM sample. Reused with permission from Ritz *et al.* (2004), ©2004, Carl Hanser Verlag.

resin before cutting out the wafer bar. FIB preparation of the targeted structure in this bar includes cutting away square boxes to both sides of the structure and subsequent milling and polishing of the TEM lamella to the desired thickness (Giannuzzi and Stevie, 2005).

For tomography, however, blocking of the electron beam at high tilts by sample material surrounding the region of interest must be avoided. Larger amounts of material than usual have to be removed (Fig. 12.18). This requires additional preparation time. The FIB *in-situ* lift-out technique avoids this disadvantage by



(a) (b)

FIG. 12.20. (a) SEM image of a post-shaped tomography sample of individual vias after FIB preparation. (b) HAADF STEM image of another such sample, ©2006 Intel Corp.

extracting an electron-transparent lamella from a wafer piece or directly from the wafer using needle micromanipulators (Giannuzzi *et al.*, 1998). This technique is well suited for producing tomography samples (Fig. 12.19). Cutting such a sample from, e.g. a via chain, makes it possible to reconstruct several individual vias from the same specimen.

The maximum possible specimen thickness for tomography usually depends on the mass thickness of the structures that the sample contains. The electron transparency of samples containing thick metal layers may be too low to yield sufficient image contrast for a well-defined reconstruction. Furthermore, inelastic electron scattering increases with mass thickness and intensifies the blurred 'inelastic image' that is superposed on the 'elastic image'. For most samples, a maximum thickness of 150 nm is practicable with a 200-kV TEM. Higher acceleration voltages allow the use of even thicker specimens.

The strong increase of the effective thickness of such conventional, lamella-shaped specimens at high tilt angles with its deteriorating effects on contrast and resolution (Section 11.2.2) has led to the introduction of needle- or post-shaped specimens prepared by FIB (Schwarz and Giannuzzi, 2004). Preparation of needles containing even individual vias has been demonstrated by Min and Mardinly (2003), see Fig. 12.20.

12.2.2.2 *Imaging modes* Nanoelectronic structures with metallization layers show pronounced mass-thickness variations. In low- and medium-resolution bright-field TEM images of such structures the image contrast varies monotonically with mass thickness in good approximation. The micrographs can then be considered as parallel 2D projections of the 3D object, with the exception of small non-monotonic contributions due to effects of defocusing. While this is the case for materials such as dielectric structures lined with amorphous barrier layers, strongly scattering polycrystalline semiconductor or metallization layers will not fulfill the projection requirement, possibly leading to artefacts in bright-field

TEM-based tomography if not treated properly, see Section 11.2.1. Other TEM imaging techniques that avoid such orientation-dependent contrast are therefore preferred for these samples. However, it should be reconceived for every particular metrology problem to what extent small artefacts can be tolerated in favour of the smallest possible operational effort.

Inelastic bright-field imaging using imaging energy filters is such an alternative imaging technique. However, while its contrast is approximately a function of the projected amount of an atomic species in the specimen, diffraction effects to some extent also contribute to the inelastic image. Moreover, only for samples thinner than approximately 0.7 times the inelastic mean free electron path does the signal vary monotonically with thickness, which is too thin for useful tomography of most nanoelectronic structures.

Another approach to avoid orientation-dependent contrast is replacement of simple projections by superpositions of micrographs recorded with different beam tilts. This has been applied to tomography of transistor structures by Bugiel *et al.* (2002). However, this method, referred to as *discrete hollow cone illumination*, involves extra data-acquisition effort and causes image-alignment problems and is therefore not a practical alternative for high-throughput electron tomography.

High-angle annular dark-field scanning TEM (HAADF STEM) that largely avoids problems associated with orientation contrast is now widely accepted as the imaging mode of choice for micro- and nanocrystalline samples, e.g. for Cu-filled interconnects (Kübel, 2002). At 200 kV the contribution of diffraction to the HAADF signal is negligible for electron collection angles larger than 70 mrad. Since the influence of orientation contrast on reconstruction decreases with increased filling of the Fourier space with projections and is low for nanocrystalline features such as Cu seed layers, good results can also be achieved by bright-field TEM tomography for such particular applications (Fig. 12.22). For the characterization of different low-Z materials in the vicinity of high-Z materials, HAADF STEM is problematic. The HAADF signal from high-Z structures is much more intense than from low-Z materials, and the signal differences between different low-Z materials are comparatively small. Therefore, missing wedge and incomplete angular sampling artefacts originating from the high-Z structures can interfere with low-Z structures in the reconstruction. BF TEM, which has a smaller Z dependence of the signal intensity ($Z^{\approx 0.7}$) than HAADF STEM ($Z^{\approx 1.7}$) is more suited for this task. Kübel *et al.* (2005) have demonstrated that despite diffraction-contrast effects it is possible to obtain a high-quality 3D reconstruction using TEM tomography if the specimen contains only a small number of crystalline areas and if proper imaging conditions are used.

Ercius *et al.* (2006) have proposed an incoherent bright-field STEM imaging mode that is well suited for tomography of Cu interconnect samples up to about 1 µm thickness. Avoiding the mass-thickness (i.e. $\rho \cdot d$) and diffraction-contrast artefacts of HAADF STEM and bright-field TEM, this imaging mode is supposed to be especially beneficial for the visualization and quantification of stress voids

in Cu interconnect lines.

12.2.2.3 *Data acquisition* In order to accomplish the acquisition of an electron tomographic tilt series consisting of hundreds of images in a reasonable amount of time, the use of computerized TEMs with CCD cameras and appropriate software for automated tilt-series acquisition is mandatory, see Section 11.2. Such software that performs an approximate shift correction and focus tracking during acquisition ('online alignment') is now available from various commercial and academic resources. A standard TEM specimen holder cannot take advantage of the maximum possible tilt-angle range in most experimental situations. Special high-tilt holders have to be used for optimum results. They are now commercially available in a variety of designs. A special case is a combined FIB/STEM sample holder with a built-in rotation mechanism (Yaguchi *et al.*, 2003). Inelastic blurring is avoided and additional contrast achieved if an imaging electron energy filter is used to remove the inelastic contribution (Angert *et al.*, 2000). With such an instrument it is also possible to combine tomography with the localization of chemical elements in the sample (Möbus *et al.*, 2003).

Silicon, metallization layers and most traditional dielectric materials used in semiconductor manufacturing such as silicon oxide and nitrides are not very beam sensitive at typical TEM electron energies of 100 to 300 kV. Radiation damage due to the high total electron dose accumulated during recording of a tilt series does not usually pose a problem. However, modern dielectric materials with a low dielectric constant κ (in the semiconductor industry κ or sometimes k is preferred over the usual ϵ), especially of the porous type, may suffer beam damage and change their structure during acquisition of the images. The total electron dose needed has to strike a balance between the dose necessary to obtain sufficient signal statistics and the dose at which specimen damage starts to occur, see Section 11.2.2.

12.2.2.4 *Data processing* For the best possible resolution of the reconstructed image the residual misalignment of the raw images has to be corrected ('offline alignment') to bring all images precisely to a common origin (Section 11.3). Alignment by manually marking a set of fiducial points or aided by gold beads applied to the specimen surface requires additional work steps which is in conflict with the target of high throughput tomography. Markerless alignment techniques such as simple cross-correlation (Mardinly, 2001, Kübel, 2002, Stegmann *et al.*, 2003) and more complicated iterative projection-matching schemes (Stegmann and Zschech, 2003) have been successfully applied to tomography of nanoelectronic structures.

Weighted backprojection is a fast and reliable reconstruction algorithm for projections with a high signal-to-noise ratio (Frank, 1992), see Section 2.2.1. Therefore, it has been the preferred reconstruction algorithm for specimens that show high mass-thickness contrast, e.g. any kind of metallization layers in patterned dielectrics. Iterative reconstruction techniques such as SIRT (see Section 2.2.2) have so far only been applied to specimens with small mass-thickness

differences for better visualization of different parts of the device. Reconstruction provides a 3D array of grey values that can be visualized in various ways that are useful for metrology. For rendering of inner surfaces in the reconstructed volume a continuous isosurface is drawn, marking where the grey values start to exceed a chosen threshold. For an appropriate threshold the surface will correspond to the interface between components with different mass thickness. Such a surface representation provides an overview of the morphology of the structure, giving the manufacturing process engineer an overall impression of, e.g. a barrier layer's step coverage.

Slices much thinner than achievable by specimen preparation can be extracted from the reconstructed data in arbitrary directions, e.g. normal, parallel, or inclined to the z-axis (Fig. 12.22). Each slice can be observed independently to assess the internal microstructure of the sample in detail. Layer thicknesses can be measured precisely at these slices. It is thus not necessary to prepare the specimen as thin as possible to avoid misinterpretations of the 2D images due to the loss of 3D information in the projection. Offline alignment, reconstruction and visualization algorithms are either incorporated in the mentioned acquisition software packages or are available as separate pieces of software, often originating from biological electron tomography projects.

12.2.2.5 *Measurement accuracy and resolution* Since quantitative analysis of electron tomographic 3D reconstructions for industrial use requires high reproducibility and reliability, not only standardized experimental conditions but also an objective assessment of the accuracy of measurement is needed. At present the measurement precision is usually estimated based on the discernability of features with known size. Artefacts related to the missing wedge, to misalignment, beam damage to the specimen, or orientation contrast, have to be spotted and sorted out by the operator. These problems have a direct influence on measurement accuracy.

The missing wedge will not result in missing pieces in the reconstructed data but in an anisotropy of resolution, lowering it in the directions that correspond to the missing wedge. Different object points will be distorted differently in the reconstructed data according to the point-spread function of the system at the co-ordinates of that point. This means that the precision in measuring, e.g. a layer thickness or an interface roughness from a 3D reconstruction, is anisotropic. This fact is very important when doing metrology of nanoelectronic structures from 3D reconstructions.

For a future automated estimation of the errors of distance measurements obtained from reconstructed data the actual resolution in the direction of measurement will have to be known. It needs to be assessed from an objective criterion. Conventional resolution-assessment techniques such as Fourier shell correlation (FSC) and differential phase residual (DPR) are not reliable for estimating the resolution of tomograms, see Section 11.3. An improved criterion based on cross-validation has been proposed by Cardone *et al.* (2005). It calculates as a function

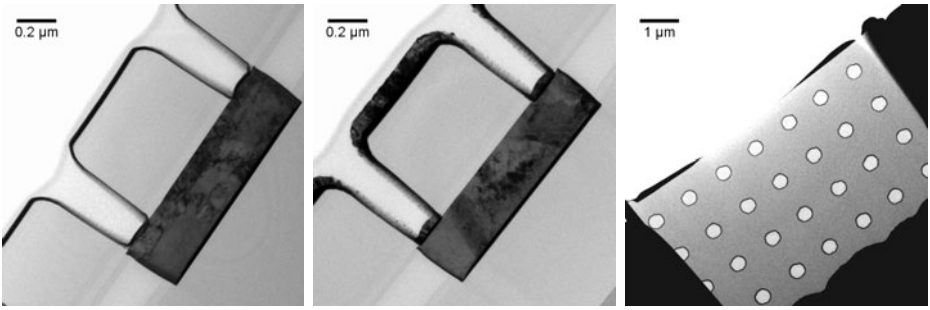


FIG. 12.21. Bright-field TEM images of vertical (a,b) and horizontal (c) FIB–cut cross-sections through a via chain interconnect test structure. A Ta barrier layer was deposited into the etched structures. An additional Cu seed layer was deposited in the structure shown in (b). Taken from Stegmann and Zschech (2003), ©2003, Materials Research Society.

of the spatial frequency the consistency between each projection and the corresponding reprojection of a tomogram calculated from all other projections.

The missing wedge can, in principle, be reduced by recording double or even multiple tilt series with differently orientated tilt axes (Mastronarde, 1997). However, this will also increase acquisition and data-processing time and may in many cases not be necessary (Min and Mardinly, 2003). Tilt-range limitations can also be overcome by using cylindrical samples with sample holders and goniometers that allow the full $-90° \ldots +90°$ tilt range.

A general parametric model describing the expectations of the distance between two components in a 3D reconstruction from a single-axis tilt series has been developed by van Aert *et al.* (2002). They showed that the measurement precision depends on the size and distance of the components, the resolution of the microscope, the electron-counting statistics, and the orientation of the object to the tilt axis. The precision increases with the number of projections, but not beyond a certain number of projections, and with decreasing pixel size. The benefits of advanced reconstruction algorithms, e.g. gridding algorithms, ART, or SIRT (Frank, 1992) over weighted backprojection for improving measurement accuracy in nanoelectronic structures have not yet been systematically investigated.

12.2.3 *Application examples*

12.2.3.1 *Barrier/seed step-coverage analysis* The introduction of the inlaid copper interconnect technology to replace aluminium interconnects in the late 1990s has created new needs in metrology. Inlaid copper lines offer higher conductivity, improved electromigration behaviour and reduced costs of manufacturing (Edelstein, 1995, Andricacos *et al.*, 1998). Horizontal interconnect lines and ver-

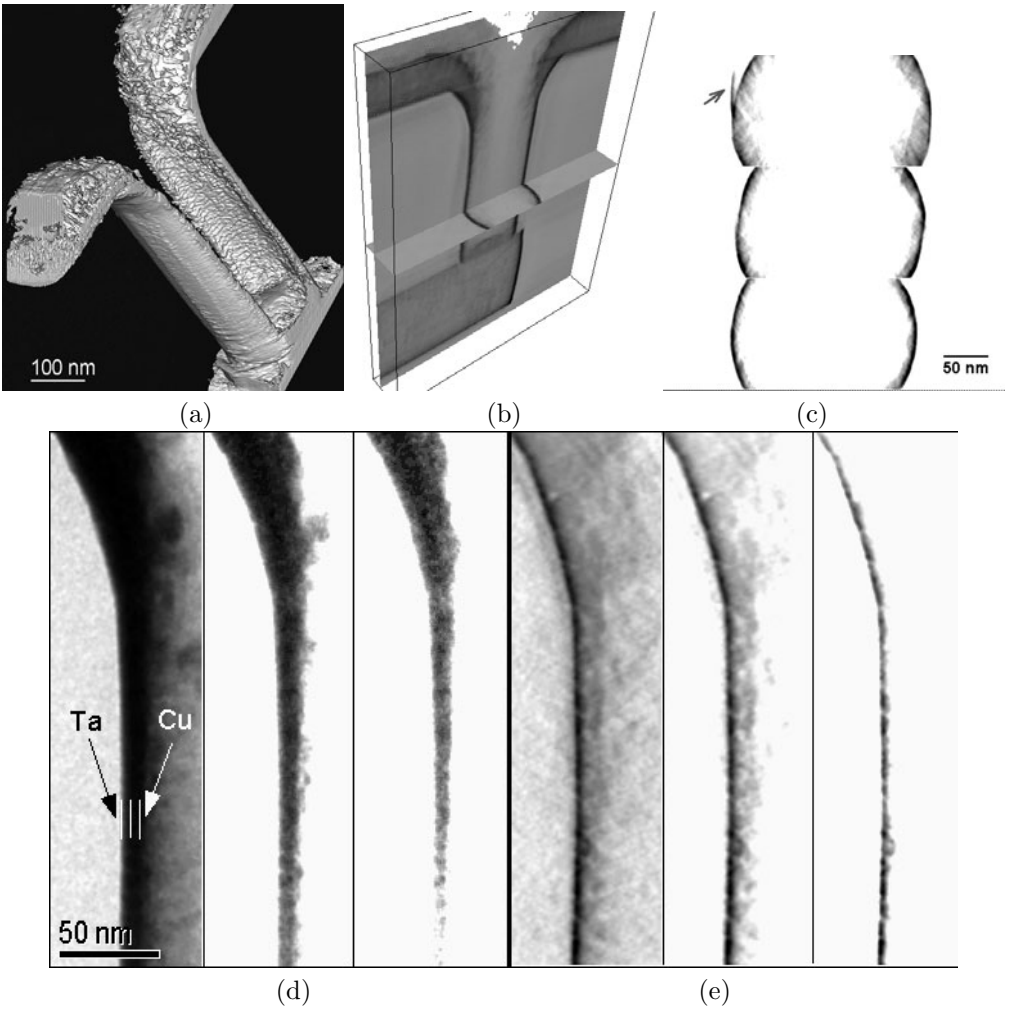


FIG. 12.22. (a) Surface rendering of the 3D reconstruction of an individual via with barrier/seed layer stack. (b) Two perpendicular slices from the volume. (c) Horizontal slices extracted at different heights. Arrow: reconstruction artefact. (d) Same sample, detail of a conventional TEM bright-field image at three different contrast settings. Specimen thickness 150 nm. (e) Detail of a slice extracted from the 3D reconstruction. Slice thickness 0.74 nm. Depending on the contrast setting, either the whole barrier/seed layer stack or just the barrier can be visualized. Taken from Stegmann and Zschech (2003), ©2003, Materials Research Society.

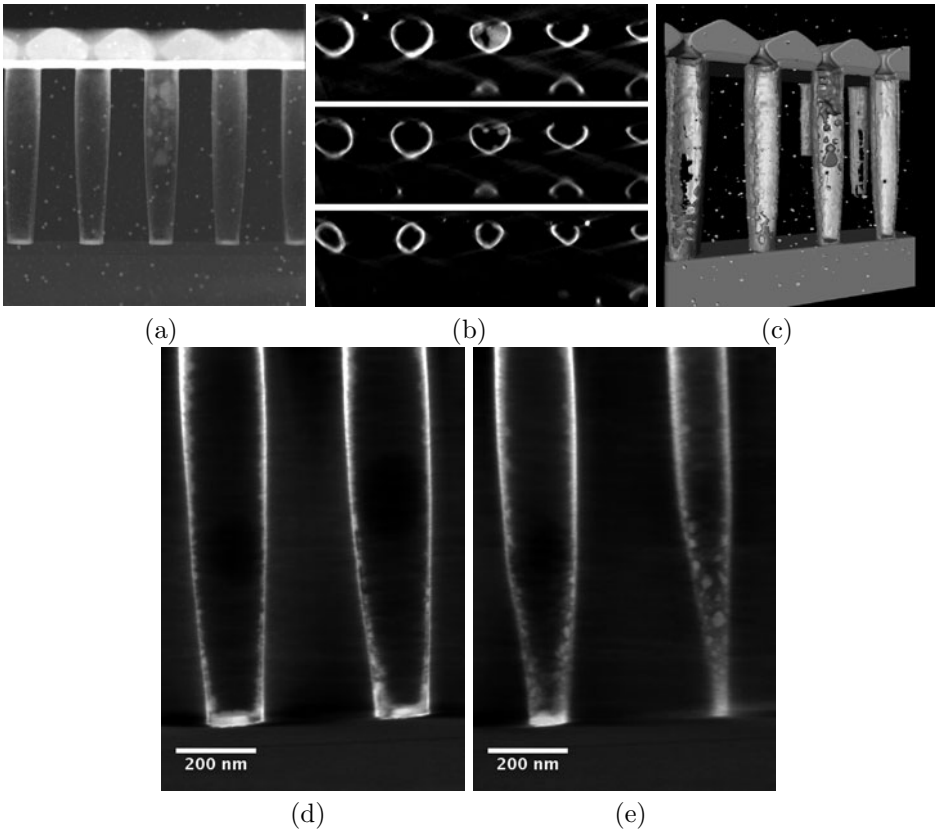


FIG. 12.23. (a) HAADF STEM image at 0° from a tilt series of high aspect ratio vias with barrier/seed layer. (b) slices extracted from the reconstruction perpendicular to the vias showing irregular shape close to the bottom of the via. (c) surface rendering of the overall via structure. (d,e) slices from a 3D reconstruction at high resolution showing the barrier and Cu seed layer. Reused with permission from Kübel *et al.* (2005), ©2006, American Institute of Physics.

tical contacts (vias) have to be embedded in ultrathin barrier layers that prevent Cu atoms from diffusing into the active gate areas and improve the adhesion between interconnect and dielectric. Typical barrier materials are Ta, W and Ti, and their compounds with N and Si, and their thickness is in the range of some nanometres. A thin Cu seed layer is sputtered on top of the barrier to allow for void-free electrolytic Cu filling of the trenches and contact holes.

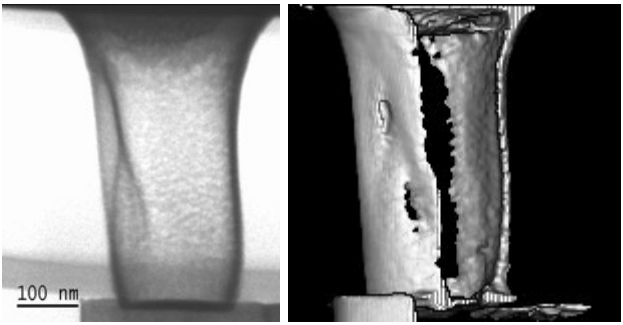
Shrinking interconnect dimensions with via diameters of 100 nm and less, increasing aspect ratios and the different electrical and deposition properties of the various barrier materials require a continuous improvement of the barrier depo-

sition processes. TEM has sufficient spatial resolution to provide the necessary feedback to process engineering, i.e. for evaluating the thickness distribution, homogeneity and continuity of non-planar ultrathin barrier and seed layers. These analytical tasks are referred to as *step-coverage analysis*. It is today one of the most important applications of TEM in semiconductor manufacture.

Barrier/seed step-coverage TEM analyses are mostly performed at critical test structures like via chains. Electron-transparent samples are prepared by site-specific cross-section preparation, which is most efficiently done by FIB cutting (Giannuzzi and Stevie, 2005). Step-coverage evaluation is usually performed by bright-field TEM imaging of vertical cross-sections, providing an overview of the layer thickness along the via wall (Fig. 12.21). Horizontal cross-sections, which are more difficult to prepare, show the coverage along the via circumference. Accurate measurement of such ultrathin layers requires TEM lamellas as thin as possible since highly curved rather than planar features have to be analysed. As mentioned above, the limited minimum specimen thickness causes problems in the interpretation of conventional 2D TEM images. Precise metrology of the barrier/seed layer stack is rendered increasingly difficult (Zschech *et al.*, 2001) both for horizontal and vertical cuts through via chains. In the latter case, additional difficulties arise by offcentre positioning of the via in the TEM lamella.

Because of the importance of barrier/seed step-coverage analysis most electron tomography work in nanoelectronics has so far been carried out on Cu interconnect structures (Mardinly, 2001, Kübel, 2002, Stegmann *et al.*, 2003). An example of bright-field TEM tomography used for step-coverage analysis is shown in Fig. 12.22. The sample used here had a Ta-barrier/Cu-seed layer stack deposited into an etched via chain structure. The final thickness of the lamella was about 150 nm. A double-tilt series with perpendicular tilt axes was recorded. For the first part of the series, the vias were aligned parallel to the tilt axis, and micrographs at $-70°\ldots+50°$ tilts in increments of 2° were acquired. For the second part, the vias were aligned perpendicular to the tilt axis, and micrographs at $+50°\ \ldots 70°$ tilts, also using 2° increments, were recorded. Image alignment was done by a projection-matching procedure (Stegmann and Zschech, 2003). In the conventional TEM image the sublayers can hardly be distinguished, requiring further sample thinning (Fig. 12.22(d)). A tomography sample, however, can contain a large part of a via or even the entire via. The layer stack can be examined everywhere at the via wall (Fig. 12.22(e)), except for the regions where the resolution is low due to the missing wedge (in this case, towards the surfaces of the TEM lamella). Thickness measurement of these layers is considerably improved by this technique (Stegmann and Zschech, 2003).

A similar example of an interconnect structure with Ta-barrier/Cu-seed layer using HAADF STEM imaging for data acquisition is shown in Fig. 12.23. From this sample, a single tilt series covering $-75°\ldots+75°$ at 1° steps was recorded. The continuity of the seed layer varies significantly and individual crystals are clearly visible towards the bottom of these vias (Kübel *et al.*, 2005).



(a) (b)

FIG. 12.24. Via with unevenly etched side wall. (a) 0°-tilt image from TEM bright-field tilt series. (b) Surface representation of the reconstruction (via rotated by 30°. Taken from Stegmann and Zschech (2003), ©2003, Materials Research Society.

Discontinuous seed layers may lead to void formation when filling the interconnects with Cu by electroplating. Increased resistance and, ultimately, failure of the interconnect due to electromigration are the consequence. Such tiny voids in Cu interconnects have been visualized using electron tomography by Kübel *et al.* (2005) and Ercius *et al.* (2006).

12.2.3.2 *Defect analysis* A very important analytical task in semiconductor manufacture aims at the localization and identification of defects that cause circuit failure. Very small and buried defects such as contaminating particles and incompletely removed resist or etch stop layers need to be found and identified. It is possible to localize a single defect on a wafer by combining the information from electrical measurements, the wafer navigation file and the circuit layout. If the origin of a tiny defect cannot be found using SEM, FIB, Auger electron spectroscopy (AES), X-ray photoelectron spectroscopy (XPS) or time-of-flight secondary ion mass spectroscopy (TOF-SIMS) tools, the complementary use of TEM is often successful.

Some defects or small particles are not visible in conventional TEM or STEM images. Electron tomography can successfully be applied here. The tiny defect visible at the bottom corner of the via shown in Fig. 12.25 can be precisely localized in this HAADF STEM 3D reconstruction. Furthermore, the roughness of the via sidewalls can be evaluated much better from 3D reconstructions than from conventional TEM images. From the micrograph shown in the example in Fig. 12.24 it is not clear whether etching or barrier deposition was non-uniform. The bright-field TEM 3D reconstruction, however, reveals the unevenly etched side wall.

12.2.3.3 *Porous low-κ dielectrics* As the dimensions of the interconnect lines continue to shrink, oxide/nitride interlayer dielectrics are being replaced by mate-

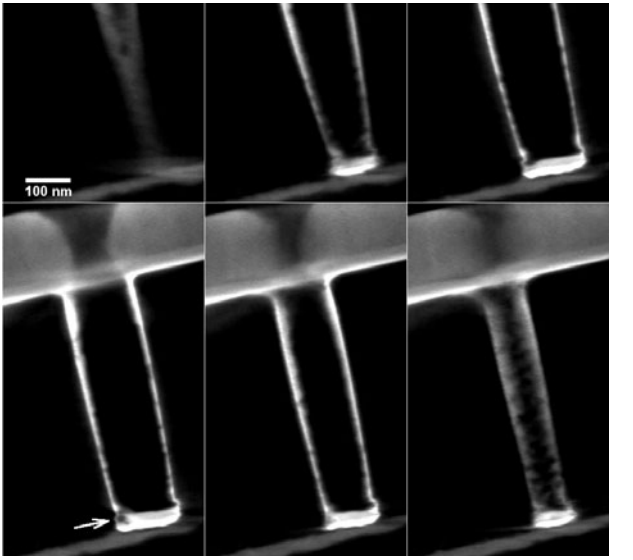


FIG. 12.25. Slicing through a 3D reconstruction (HAADF STEM) of a via with barrier/seed layer. At a certain depth, a defect in the bottom corner of the via becomes visible (arrow), ©2006 Intel Corp.

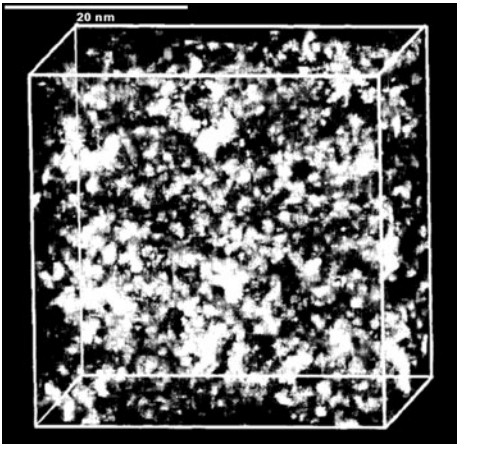


FIG. 12.26. Surface rendering of a 3D tomographic reconstruction of a porous low-κ dielectric. Reused with permission from Shimada *et al.* (2004), ©2004, IEEE.

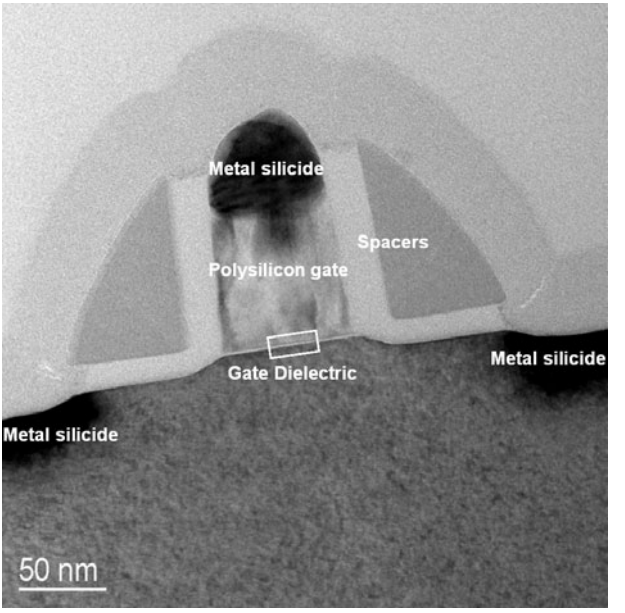


FIG. 12.27. Bright-field TEM image of a FIB-cut cross-sections trough a metal oxide field effect transistor. Measurement of polysilicon line width, spacer dimensions, *etc.* is carried out at such cross-sections. Taken from Stegmann and Zschech (2003), ©2003, Materials Research Society.

rials with a lower dielectric constant κ= 2.6–3.0 such as organic spin-on polymers and plasma-enhanced CVD inorganic/organic hybrid materials. Porous ultralow κ materials (ULK, $\kappa \leq 2.5$) are in development. Since κ scales inversely with porosity, a large pore size is desired. However, the pore size is limited by the decrease of mechanical strength and thermal conductivity with increasing pore size. The ideal ULK material has a closed pore structure to avoid crack propagation, void formation, moisture uptake, and metal penetration into the pores. Its pore-size distribution should be narrow to achieve a homogeneous and isotropic bulk κ.

The weak contrast of pores in a dielectric and the susceptibility to beam damage make electron tomography of such materials problematic. However, it was possible to confirm by electron tomography that pore shape and connectivity depend on the pore-formation technique in the films (Shimada *et al.*, 2004), see Fig. 12.26.

12.2.3.4 *Transistor structures* The need to increase the performance of microprocessor products is the driving factor for downscaling the MOS transistor (Arden, 2003). Minute variations of the dimensions of MOS transistor structures like

FIG. 12.28. Surface rendering of a 3D reconstruction of a semiconductor device structure in a state-of-the-art manufacturing process. The roughness of the polysilicon line can be easily evaluated,

the ultrathin SiO_2 gate dielectric (thickness less than 2 nm), the polycrystalline (poly)silicon gate (less than 50 nm), spacers (less than 20 nm), trench isolation structures, and self-aligning metal silicide source, gate and drain contacts have a major impact on performance, yield, and reliability of the product. Therefore, control of the thickness and roughness of these structures with nanometre accuracy is a top-priority task for leading-edge integrated circuit manufacture. Non-destructive in-factory metrology using optical and scanning electron microscopy (SEM) techniques often does not provide sufficient accuracy to control the effects of very small process parameters changes. In these cases it is necessary to measure the relevant dimensions from TEM images of FIB-prepared samples, see Fig. 12.27(b).

The line-edge roughness of etched polysilicon gates, which already becomes an issue for transistors with a gate width smaller than 50 nm, is a challenge to these microscopic techniques (Semiconductor International Association, 2005, Zschech *et al.*, 2003). The information delivered by conventional SEM and 2D TEM is limited. Here, electron tomography fills the gap between other three-dimensional metrological techniques such as 3D surface imaging by top-down SEM, atomic force microscopy (AFM) and dual-column FIB serial sectioning. The polysilicon line roughness is easily visible from HAADF STEM 3D reconstructions, see Fig. 12.28.

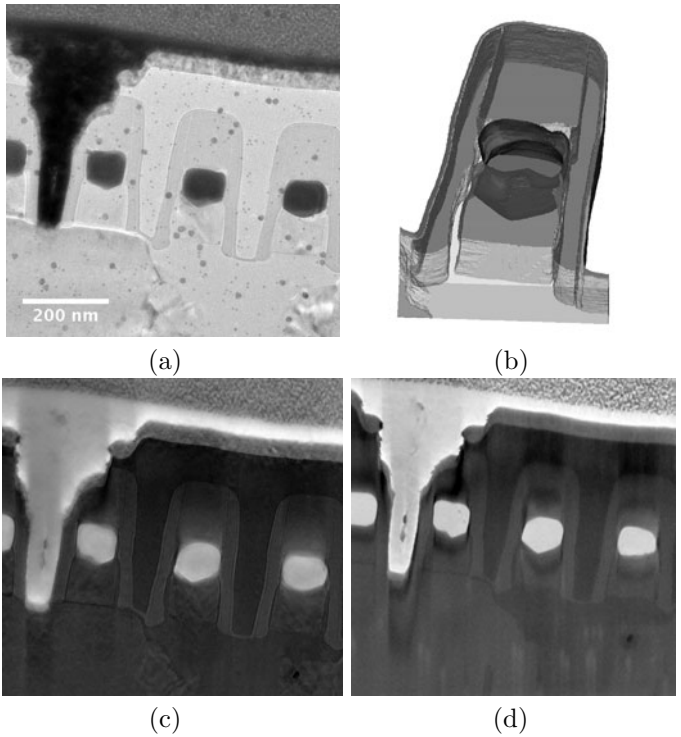
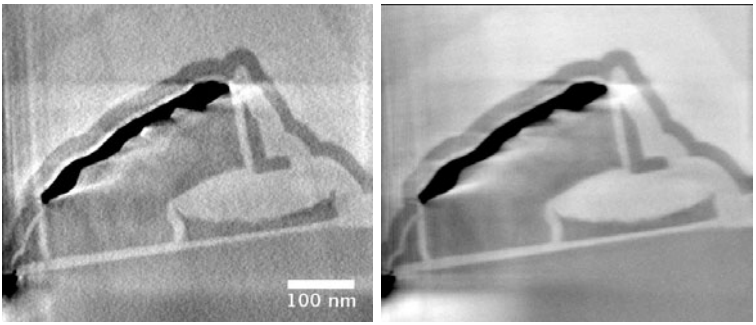


FIG. 12.29. (a) TEM reference image at 0° tilt showing a W contact and the transistors next to it. (b) 3D rendering of the structure of the transistor obtained by BF TEM tomography. In 3D, significant variations of the oxide thickness are visible. (c) the BF TEM tomography reconstruction shows the low-Z domains of the transistor very well and enables e.g. imaging of a 1.5 nm SiO_xN_y layer in 3D. (d) the HAADF STEM tomography shows the shape of the contact and WSi_2 in the transistor much better. Small holes are visible in the WSi_2. Reused with permission from Kübel *et al.* (2005), ©2006, American Institute of Physics.

Figure 12.29 shows an example of a specimen containing both low- and high-Z materials, a transistor with metallic contacts. The different properties of BF TEM and HAADF STEM tomography in imaging these different domains become apparent from this typical sample.

The benefits of an advanced reconstruction algorithm over simple weighted backprojection for the visualization of such a structure can be seen in Fig. 12.30. The weighted backprojection reconstruction of a flash-memory cell reveals the shape of the floating gate and the surface roughness on the top side of the floating gate. The reconstruction with SIRT shows the same features, except for small differences in the roughness of the floating gate. However, noise and 'streaking'



(a) (b)

FIG. 12.30. (a) 3D tomographic weighted-backprojection reconstruction of a flash memory cell based on an (inverted) HAADF STEM tilt series. The slice through the reconstructed volume shows the polysilicon floating gate and the reference gate as the main components of the storage device. (b) SIRT reconstruction from the same data. Reused with permission from Kübel *et al.* (2005), ©2005, Microscopy Society of America, published by Cambridge University Press.

artefacts are significantly smaller, enabling a clearer detection of the atomic-number contrast between the different parts of the device (Kübel *et al.*, 2005).

12.2.3.5 *Gate dielectrics* Single impurity or dopant atoms in a gate dielectric film can introduce defect states into the local bandstructure that induce leakage or even breakdown of the device. Tomography techniques based on tilt series do not provide sufficient resolution to map potential leakage sites in three dimensions. This information would be useful for the improvement of gate dielectric processes by correlating it to measured leakage characteristics and carrier mobilities. This is of special importance for metal oxide gate dielectrics with a high dielectric constant (high-κ materials) that will have to be introduced in volume manufacturing soon for further increased transistor performance.

Using recently developed aberration correctors, electron probes with diameters of less than 0.1 nm can be formed. The effect of such a corrected beam is a substantial reduction in the depth of focus with increased illumination semi-angles (Borisevich *et al.*, 2006). The detection of single atoms by HAADF STEM is greatly enhanced by this development. Recording images at different focus values it is possible to obtain a tomographic slice series through the sample. Localization of individual hafnium atoms in a $HfO_2/SiO_2/Si$ high-κ gate dielectric stack in three dimensions with a precision of about ± 0.5 nm in depth using this technique has been reported by van Benthem *et al.* (2005) (Fig. 12.31).

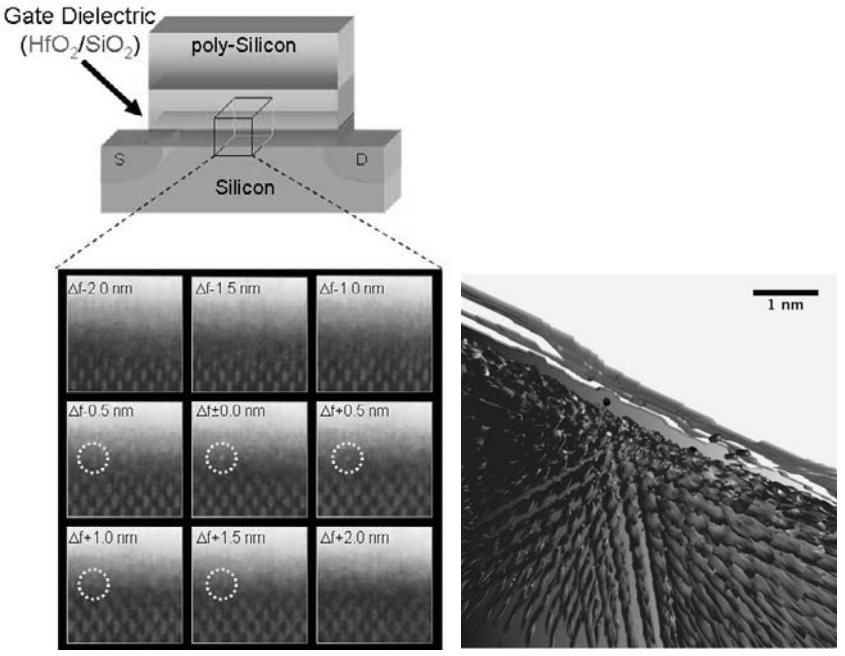


FIG. 12.31. (a) Annular dark-field through-focal series from a $HfO_2/SiO_2/Si$ gate stack. Nine consecutive micrographs extracted from a series of 41 images are shown. One single Hf atom coming into and going out of focus is marked by white circles. (b) Three-dimensional representation of the $HfO_2/SiO_2/Si$ interface structure. Single Hf atoms are visible in the SiO_2 layer. Reused with permission from van Benthem *et al.* (2005), ©2005, American Institute of Physics.

12.2.4 *Future prospects*

Performance limitations due to the traditional downscaling of MOS devices have to be overcome by the introduction of new materials and improved device designs. These technology requirements are anticipated and published regularly in the International Technology Roadmap for Semiconductors (Semiconductor International Association, 2005).

Many new design and technology options are currently under investigation to remedy the performance limitation of traditional interconnects such as cooled superconductors, microwave interconnects, 3D interconnects, optical interconnects, carbon nanotubes, spin coupling, and molecular interconnects. These approaches are causing more material and process-integration challenges. It is expected that one or more of such interconnect approaches will start to be used within the next five years. Non-classical concepts are being developed for MOS devices, includ-

ing the use of SiGe, double-gate and vertical MOSFETs, interconnect dielectrics with air gaps, and vertically stacked circuits. Electron tomography obviously has a high potential to help solve metrology and integration issues involved with these new designs with their complex geometries.

Another challenge for metrology is elemental analysis at device dimensions. EFTEM, holographic, and confocal STEM tomography can be applied here. They bridge the gap between tomographic techniques with atomic resolution such as voltage- or laser-pulsed local electrode atom probe (LEAP), and techniques with submicrometre resolution such as EDX analysis combined with serial FIB sectioning.

Presently, the time needed for automatic acquisition of a tilt series is one to three hours. Taking into account the additional specimen preparation and data-processing time, it is clear that electron tomography is still far from having reached the sample throughput of traditional 2D TEM imaging. In order to become a high-throughput analytical method for routine use in an industrial laboratory, specimen preparation, data acquisition and processing need to run with as much automation and with as little operator intervention as possible. One step in this direction could be the replacement of conventional electromechanical goniometers by miniaturized, piezo-driven supereucentric goniometers integrated into the TEM column (Lengweiler *et al.*, 2004). Improved alignment and reconstruction schemes are needed for higher reproducibility. Automatic or semi-automatic segmentation techniques (Frangakis and Hegerl, 2002, Bartesaghi *et al.*, 2005) may be helpful for feature detection and measurement. Further development of such hard- and software improvements for electron tomography will be necessary to approach the ultimate goal of fully automated production of 3D reconstructions.

Acknowledgements

The authors of Section 12.1 would like to thank Dr T.Y.V. Yates, Dr I. Arslan, Dr S. Hata, Dr A.C. Harrison, Dr H. Jinnai and Dr R.E. Dunin-Borkowski for providing figures. The author of Section 12.2 would like to thank Christian Kübel (Fraunhofer Institute for Manufacturing Technology and Applied Materials Research), John Mardinly (Intel Corp.), Klaus van Benthem (Oak Ridge National Laboratory), and Hans-Jürgen Engelmann, Lukas Gerlich, and Yvonne Ritz (AMD Saxony) for providing figures.

12.3 References

Andricacos, P.C., Uzoh, C., Dukovic, J.O., Horkans, J., and Deligianni, H. (1998). Damascene copper electroplating for chip interconnections. *IBM Journal of Research and Development*, **12**, 567.

Angert, I., Majorovits, E., and Schröder, R.R. (2000). Zero-loss image formation and modified contrast transfer theory in EFTEM. *Ultramicroscopy*, **81**, 203.

Arden, W. (2003). Roadmap key challenges. *Materials Today*, **6**, 40.

Barnard, J.S., Sharp, J., Tong, J.R., and Midgley, P.A. (2006). High-resolution three-dimensional imaging of dislocations. *Science*, **313**, 319.

Bartesaghi, A., Sapiro, G., and Subramaniam, S. (2005). An energy-based three-dimensional segmentation approach for the quantitative interpretation of electron tomograms. *IEEE Transactions on Image Processing*, **14**, 1314.

Blakemore, R. (1975). Magnetotactic bacteria. *Science*, **190**, 377.

Borisevich, A.Y., Lupini, A. R., and Pennycook, S.J. (2006). Depth sectioning with the aberration-corrected scanning transmission electron microscope. *Proceedings of the National Academy of Sciences of the USA*, **103**, 3044.

Bugiel, E., Daberkow, I., Schatz, M., and Tietz, H.R. (2002). TEM tomography of semiconductor device structures. In *Proceedings of the 15th International Congress on Electron Microscopy (ICEM15)* (ed. J. Engelbrecht, T. Sewell, M. Witcomb, and R. Cross), Volume 3, Durban, pp. 373. The Microscopy Society of Southern Africa.

Cardone, G., Grünewald, K., and Steven, A. C. (2005). Resolution assessment in electron tomography: a cross-validation approach. *Microscopy and Microanalysis*, **11**(Suppl. 2), 22.

Edelstein, D.C. (1995). Advantages of copper interconnects. In *Proceedings of the XII International VLSI Multilevel Interconnect Conference*, Piscataway, pp. 301. Institute of Electrical and Electronics Engineers.

Egerton, R.F. (1996). *Electron energy loss spectroscopy in the electron microscope* (2^{nd} edn). Plenum Press, New York.

Engelmann, H.J., Volkmann, B., Blum, W., and Zschech, E. (2002). TEM target preparation within one hour: vision or realistic goal? *Practical Metallography*, **39**, 117.

Ercius, P., Weyland, M., Muller, D., and Gignac, L.M. (2006). Three-dimensional imaging of nanovoids in copper interconnects using incoherent bright field tomography. *Applied Physics Letters*, **88**, 243116.

Frangakis, A.S and Hegerl, R. (2002). Segmentation of two- and three-dimensional data from electron microscopy using eigenvector analysis. *Journal of Structural Biology*, **138**, 105.

Frank, J. (1992). *Electron tomography. Three-dimensional imaging with the transmission electron microscope.* Plenum Press, New York; London.

Gass, M.H., Koziol, K.K., Windle, A.H., and Midgley, P.A. (2006). 4-dimensional spectral-tomography of carbonaceous nano-composites. *Nano Letters*, **6**, 376.

Giannuzzi, L.A., Drown, J.L., Brown, S.R., Irwin, R.B., and Stevie, F.A. (1998). Applications of the FIB lift-out technique for TEM specimen preparation. *Microscopy Research and Technique*, **41**, 285.

Giannuzzi, L.A. and Stevie, F.A. (2005). *Introduction to focused ion beams.* Springer, New York.

Gruner, S.M. (1989). Stability of lyotropic phases with curved interfaces. *Journal of Physical Chemistry*, **93**, 7562.

Inoke, K., Kaneko, K., Weyland, M., Midgley, P.A., Higashida, K., and Horita,

Z. (2006). Severe local strain and the plastic deformation of Guinier-Preston zones in the Al-Ag system revealed by 3D electron tomography. *Acta Materialia*, **54**, 2957.

Jinnai, H., Nishikawa, Y., Spontak, R.J., Smith, S.D., Agard, D.A., and Hashimoto, T. (2000). Direct measurement of interfacial curvature distributions in a bicontinuous block copolymer morphology. *Physical Review Letters*, **84**, 518.

Kimura, K., Hata, S., Matsumura, S., and Horiuchi, T. (2005). Dark-field transmission electron microscopy for a tilt series of ordering alloys: towards electron tomography. *Journal of Electron Microscopy*, **54**, 373.

Koster, A.J., Verkleij, U. Ziese A.J., Janssen, A.H., and de Jong, K.P. (2000). Three-dimensional electron microscopy: A novel imaging and characterization technique with nanometer scale resolution for materials science. *Journal of Physical Chemistry B*, **104**, 9368.

Koziol, K., Schaffer, M.S.P., and Windle, A.H. (2005). Three-dimensional internal order in multiwall carbon nanotubes grown by chemical vapour deposition. *Advanced Materials*, **17**, 760.

Kresge, C.T., Leonowicz, M.E., Roth, J., Vartuli, J.C., and Beck, J.S. (1992). Ordered mesoporous molecular sieves synthesized by a liquid-crystal template mechanism. *Nature*, **359**, 710.

Kübel, C.and Voigt, A., Schoenmakers, R., Otten, M., Su, D., Lee, T.C., Carlsson, A., and Bradley, J. (2005). Recent advances in electron tomography: TEM and HAADF-STEM tomography for materials science and semiconductor applications. *Microscopy and Microanalalysis*, **11**(Suppl. 2), 378.

Kübel, C. (2002). 3-D electron microscopy for nano-technology and the IC industry. *Microscopy and Microanalysis*, **8**(Suppl. 2), 1104.

Kübel, C., Kübel, J., Kujawa, S., Luo, J.S., Lo, H.M., and Russell, J.D. (2005). Application of electron tomography for semiconductor device analysis. In *Stress-Induced Phenomena in Metallization, 8^{th} International Workshop, Dresden 2005* (ed. E. Zschech, K. Maex, P. Ho, H. Kawasaki, and T. Nakamura), Melville, pp. 223. American Institute of Physics. AIP Conference Proceedings, Vol. 817.

Laurer, J.H., Hajduk, D.A., Fung, J.C., Sedat, J.W., Smith, S.D., Gruner, S.M., Agard, D.A., and Spontak, R.J. (1997). Microstructural analysis of a cubic bicontinuous morphology in a neat SIS triblock copolymer. *Macromolecules*, **30**, 3938.

Leder, L.B. and Suddeth, J.A. (1960). Characteristic energy losses of electrons in carbon. *Journal of Applied Physics*, **31**, 1422.

Lengweiler, S., Birrer, G., Müller, M., and Stemmer, A. (2004). Ideal eucentric goniometer for electron tomography. In *Proceedings of the 13^{th} European Microscopy Congress* (ed. D. Schryvers and J. Timmermans), Antwerpen, pp. 219. Belgian Society for Microscopy.

Mann, S. and Ozin, G.A. (1996). Synthesis of inorganic materials with complex form. *Nature*, **382**, 313.

Mann, S., Sparks, N.H.C., Frankel, R.B., Bazylinski, D.A., and Jannasch, H.W. (1990). Biomineralization of ferrimagnetic greigite (Fe_3O_4) and iron pyrite (FeS_2) in a magnetotactic bacterium. *Nature*, **343**, 258.

Mardinly, J. (2001). The effect of Moore's law on the growing role of transmission electron microscopy in the semiconductor industry. *Microscopy and Microanalysis*, **7**(Suppl. 2), 510.

Mastronarde, D.N. (1997). Dual-axis tomography: an approach with alignment methods that preserve resolution. *Journal of Structural Biology*, **120**, 343.

Min, K.H and Mardinly, J. (2003). Electron tomography of microelectronic devices. *Microscopy and Microanalalysis*, **9**(Suppl. 2), 502.

Möbus, G., Doole, R.C., and Inkson, B.J. (2003). Spectroscopic electron tomography. *Ultramicroscopy*, **96**, 433.

Ritz, Y., Stegmann, H., Engelmann, H.J., and Zschech, E. (2004). Target preparation of samples for 3D-TEM using micromanipulators. *Practical Metallography*, **41**, 180.

Schwarz, S.M. and Giannuzzi, L.A. (2004). FIB specimen preparation for STEM and EFTEM tomography. *Microscopy and Microanalysis*, **10**(Suppl. 2), 142.

Semiconductor International Association (2005). International Technology Roadmap for Semiconductors (ITRS). `http://public.itrs.net`.

Shcherbakov, V.P. and Winklhofer, M. (1999). The osmotic magnetometer: a new model for magnetite based magnetoreceptors in animals. *The European Biophysics Journal*, **28**, 380.

Shim, M. and Guyot-Sionnest, P. (2000). n-Type colloidal semiconductor nanocrystals. *Nature*, **407**, 981.

Shimada, M., Shimanuki, J., Ohtsuka, N., Furuya, A., Inoue, Y., and Ogawa, S. (2004). 3-dimensional structures of pores in low-k films observed by quantitative TEM tomograph and their impacts on penetration phenomena. In *Proceedings of the IEEE Interconnect Technology Conference 2004*, Piscataway, pp. 178. Institute of Electrical and Electronics Engineers.

Spontak, R.J., Williams, M.C., and Agard, D.A. (1988). 3-dimensional study of cylindrical morphology in a styrene butadiene styrene block copolymer. *Polymer*, **29**, 387.

Stegmann, H., Engelmann, H.J., and Zschech, E. (2003). Characterization of barrier/seed layer stacks of Cu interconnects by electron tomographic three-dimensional object reconstruction. *Microelectronic Engineering*, **65**, 171.

Stegmann, H. and Zschech, E. (2003). Three dimensional tomographic imaging of Cu interconnect structures. In *Advanced Metallization Conference AMC2003* (ed. G. Ray, T. Smy, T. Ohta, and M. Tsujimura), Warrendale, pp. 197. Materials Research Society. MRS Conference Proceedings, Vol. V-19.

Tong, J., Arslan, I., and Midgley, P.A. (2006). A novel dual-axis iterative algorithm for electron tomography. *Journal of Structural Biology*, **153**, 55.

Twitchett, A. (2006). 3D reconstruction of electrostatic potential. unpublished.

van Aert, S., den Dekker, A.J., van Dyck, D., and van den Bos, A. (2002). High-

resolution electron microscopy and electron tomography: resolution versus precision. *Journal of Structural Biology*, **138**, 21.

van Benthem, K., Lupini, A.R., Kim, M., Baik, H.S., Doh, S.J., Lee, J.H., Oxley, M.P., Findlay, S.D., Allen, L.J., Luck, J.T., and Pennycook, S.J. (2005). Three-dimensional imaging of individual hafnium atoms inside a semiconductor device. *Applied Physics Letters*, **87**, 34104.

Weyland, M., Yates, T.J.V., Dunin-Borkowski, R.E., Laffont, L., and Midgley, P.A. (2006). Nanoscale analysis of three-dimensional structures by electron tomography. *Scripta Materialia*, **55**, 29.

Yaguchi, T., Konno, M., Kamino, T., Hashimoto, T., Onishi, T., Umemura, K., and Asayama, K. (2003). FIB micro-pillar sampling technique for 3D STEM observation and its application. *Microscopy and Microanalysis*, **9**(Suppl. 2), 118.

Zschech, E., Engelmann, H.J., Saage, H., de Robillard, Q., and Stegmann, H. (2001). Characterizationn of layer stacks in microelectronic products: challenges to sample preparation and TEM analysis. *Practical Metallography*, **38**, 442.

Zschech, E., Langer, E., Meyer, A.M., Engelmann, H.J., Stegmann, H., Geisler, H., Tracy, B., and Schneider, G. (2003). Physical failure analysis in semiconductor industry – challenges to microscopy. *Institute of Physics Conference Series*, **180**, 487. Microscopy of Semiconducting Materials Conference MSM.

PART IV

NEUTRON TOMOGRAPHY

Neutron imaging is more exotic and less known than synchrotron X-ray or electron tomography. The reason for this is the smaller number of available facilities, which ranges below 10 worldwide, and also the greater difficulties associated with the inherent flux limitation of neutron sources and the challenges with the precise detection of neutral particles. Nevertheless, neutron tomography has developed to an established technology and many experimental options known from other technologies now have their neutron counterpart.

In this part of the book, various experimental techniques are presented. First, *neutron absorption tomography* is explained, including descriptions of specific technological challenges such as detectors.

In a second chapter *phase-contrast tomography* is covered and analogy to synchrotron phase-contrast tomography as already described in Chapter 6.

Finally, *refraction and small-angle scattering tomographies* are presented, the latter being analogous to the technique mentioned in Section 8.2.3.

This page intentionally left blank

13

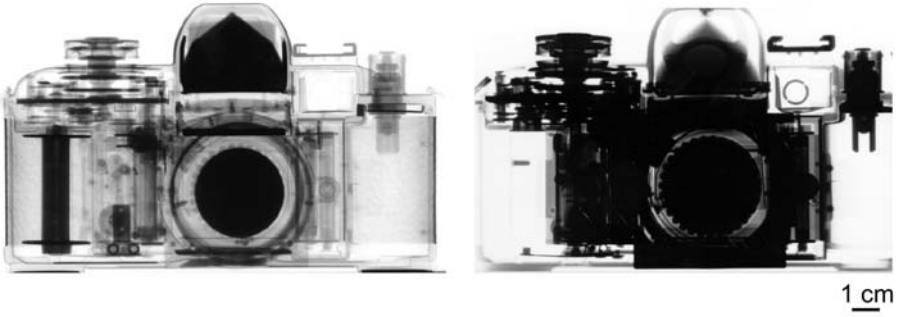
NEUTRON ABSORPTION TOMOGRAPHY

Eberhard Lehmann and *Nikolay Kardjilov*

Neutrons were first used for imaging in 1935 by Kallman and Kuhn in Germany. They obtained quite poor images, applying today's standards, owing to the weak and geometrically ill-defined beam available (Kuhn, 1937). For this purpose an accelerator-type of neutron generator was used where neutrons with a kinetic energy of approximately 2.5 MeV were produced by fusion of deuterium atoms (^{2}D+^{2}D). The complexity of the experimental setup in conjunction with the extremely long exposure times required – of the order of hours for a single radiograph – did not allow for an early establishment of the method as had taken place with X-ray radiography immediately after Röntgen's discovery. With the development of nuclear reactors after World War II the intensity of neutron sources available for such experimental purposes increased by many orders of magnitude and helped to demonstrate the potential of neutron radiography. The first use of a neutron beam from a reactor to produce a radiograph was reported by Thewlis (1956), who used the 6-MW graphite reactor 'BEPO' at Harwell to obtain images with good quality. The technique developed slowly for several years until some researchers in the early 1960s reported their work on problems associated with radiography of radioactive material. The work of Berger *et al.* (1964) of Argonne Laboratory, followed by Barton (1965*a*)/(1965*b*) at Birmingham University played a major role in the revival of the method. Today, neutron radiography and tomography see a rapid development. Although the history of neutron radiography dates back 70 years, only in the past 30 years has it come to the forefront of non-destructive testing and evolved to an accepted method. A series of World conferences is being held every four years and the user community is slowly growing.

13.1 Interaction of neutron radiation with matter and comparison to X-rays

The main reason to use thermal or cold neutrons for imaging – as explained in Chapter 4 – is their specific interaction with matter that allows them to penetrate macroscopic samples (millimetres to some tens of centimetres), to distinguish between different isotopes of one element and to be sensitive to magnetic fields due to the magnetic moment of neutrons. The fundamental interaction mechanisms are very different from those of X-rays, thus implying very different attenuation properties of the elements for the two kinds of radiation (see Fig. 13.1). The charge-free neutron interacts with the core of the atom, while in contrast X-rays

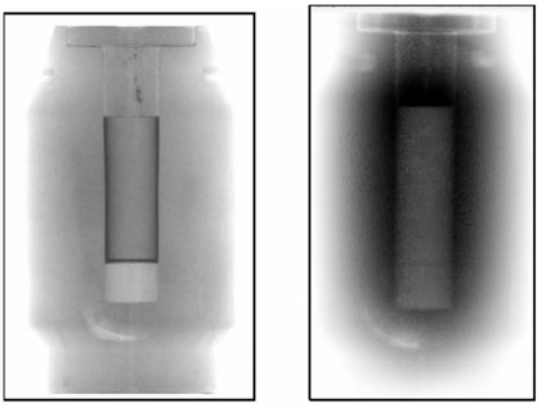


(a) (b)

FIG. 13.1. Comparison between radiographic images of a photocamera taken with, (a) thermal neutrons and, (b) X-rays at 150 keV.

interact with the charge distribution of the electron shell. Therefore, the X-ray attenuation coefficients increase with the atomic number of the elements, i.e. with the number of electrons. The interaction probability of neutrons with the nucleus is described by the total microscopic crosss-section σ_{tot}, which does not systematically vary as a function of the atomic number of the elements and, consequently, the attenuation properties of the elements for neutrons show an analogous irregular behaviour, as shown in Fig. 4.18. For radiography, the energy of the neutrons is typically of the order of meV (thermal or cold neutrons), while the X-ray energies are of the order of tens to several hundreds of keV. Comparing the mass attenuation coefficients for different elements for X-rays and thermal neutrons given in Fig. 4.18 the following statements can be made:

- neutrons are very sensitive to some light elements such as H, Li, and B, which absorb X-rays only weakly and therefore do not provide a good X-ray imaging contrast but a good contrast for neutron imaging,
- the distribution of attenuation coefficients is not a smooth function of atomic number, which helps to achieve contrast even for neighbouring elements, whereas for X-rays one observes a continuous increase with the atomic number and neighbouring elements such as Al/Si or Fe/Ni always have similar absorption properties,
- neutrons easily penetrate thick layers of metals such as Pb, Fe and Cu, where standard X-ray imaging facilities with energies of several hundreds of keV fail,
- neutrons can distinguish between isotopes (for example ^{1}H/^{2}H or ^{235}U/^{238}U) which is not the case for X-rays,
- neutrons interact strongly with magnetic moments in a sample, whereas this interaction is weak for X-rays. This allows for achieving magnetic imaging contrast.



(a) (b)

FIG. 13.2. Radiographic inspection of a lead container with hitherto unknown content (9 cm diameter). Image obtained with, (a) thermal neutrons and, (b) high-energy γ-rays (1100 keV) from a ^{60}Co source.

In order to highlight the difference between neutron and X-ray imaging, corresponding examples are presented in Fig. 13.1 and Fig. 13.2.

The example shown in Fig. 13.1 helps to explain differences in neutron (a) and X-ray (b) radiography by comparing two images of a photocamera. It is evident that the hydrogen-containing plastic parts such as the film cassette visible in dark in the left part of the camera can be visualized with neutrons, whereas metallic parts appear to be almost transparent. For X-rays, thicker metallic components are hard to penetrate and appear black, whereas the plastic part is now almost completely transparent. In this way, the complementarity of the two types of radiation can be demonstrated. The advantage of neutron radiography against the analogous X-ray technique can be further stressed in an extreme application, presented in Fig. 13.2. Here, a lead container with an unknown and suspicious content was inspected by neutron radiography, Fig. 13.2(a), with an exposure time of 20 s. The image clearly reveals that an empty plastic capsule is placed in the container. On the other hand, Fig. 13.2(b), γ-ray radiography with 120 min of exposure was not able to reveal the content of the container since the light elements inside did not attenuate the rays significantly.

The attenuation of neutrons – as that of X-rays – obeys the exponential Beer–Lambert law, eqn 4.11, quite well since deviations caused by multiple scattering, beam hardening or energy-dependent detector response can be neglected under the circumstances of neutron radiography. The maximum sample thickness that can be allowed for a radiographic or tomographic investigation can be derived using eqn 4.11. Assuming that the linear attenuation coefficient μ of the material is known (given by the interaction probability and the nuclear density of the sample material) the maximum thickness of the sample under investigation

TABLE 13.1. Maximum allowed thickness of samples when imaged with thermal neutrons assuming that at least 2% transmission is required. For natural materials an accepted approximated average value was used.

material	linear attenuation coefficient μ (cm^{-1})	largest sample thickness d_{max} (cm)
aluminium	0.10	39
heavy water	0.30	13
lead	0.38	10
dry soil	≈0.50	≈7.8
chromium	0.54	7.2
dry wood	≈0.65	≈6.0
stone	≈0.80	≈4.9
copper	1.07	3.7
iron	1.2	3.3
nickel	2.1	1.9
light water	3.5	1.1

should be below

$$d_{\max} = -\frac{\ln 0.02}{\mu} = \frac{3.91}{\mu}, \tag{13.1}$$

if a minimum transmission of 2% is defined by the sensitivity limit and dynamics of the detection system. This value is supported by practical experience. Table 13.1 summarizes the maximal thicknesses allowed for imaging with thermal neutrons (eqn 13.1) for some typical materials.

13.2 Specifics of neutron tomography

13.2.1 *Principle*

As for other tomographic methods, neutron tomography provides volumetric data of an object's neutron attenuation pattern, yielding planar sections through the interior of the object. This view of the object's structure is derived mathematically from a series of radiographic projections acquired as the object is rotated through 180° in small angular steps.

For this purpose a well-defined neutron beam from a reactor or a spallation source is used to illuminate objects and to obtain 'shadow images' on a two-dimensional detector screen. A schematic view of a typical tomographic arrangement is shown in Fig. 13.3, which is completely analogous to that used for X-ray tomography.

Installations for neutron tomography are located at large-scale facilities where the requirements for high intensity and good beam definition can be fulfilled and are stationary setups surrounded by a corresponding infrastructure. An example of this is shown in Fig. 13.4 where the facility ICON ('Imaging with Cold Neutrons') at the Paul Scherrer Institute is presented. The large dimensions of the

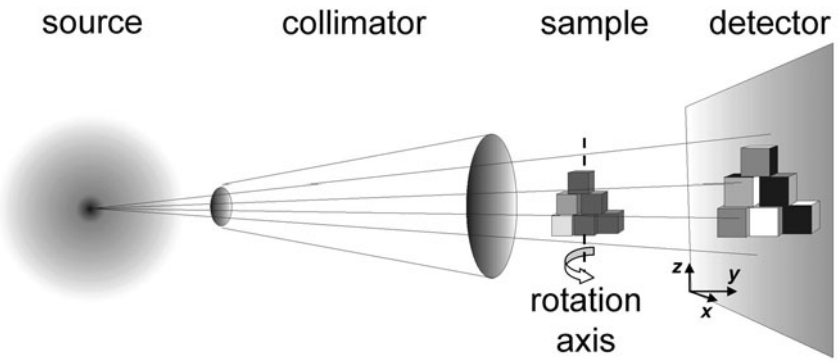


FIG. 13.3. Simplified setup of a neutron absorption tomography experiment.

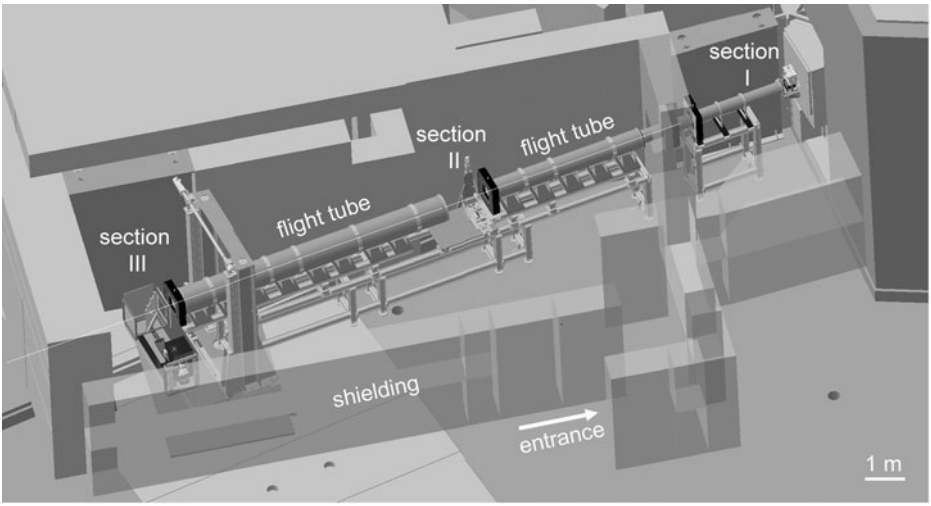


FIG. 13.4. Installation 'ICON' for cold neutron imaging, including neutron tomography, at the spallation neutron source SINQ (Paul Scherrer Institut, Switzerland). The beamline is about 18 m long and delivers an open beam of about 35 cm in diameter at the sample position (on the left).

facility are given mainly by the massive shielding required for radiation protection and the space left for the experimental infrastructure. The facility shown in Fig. 13.4 can be subdivided into three sections along the beam axis. Section I is separated by an additional shielding wall from the rest of the bunker to block the high neutron background produced by the devices used for defining the neutron beam (collimators, monochromators, beam limiters, *etc*). Behind

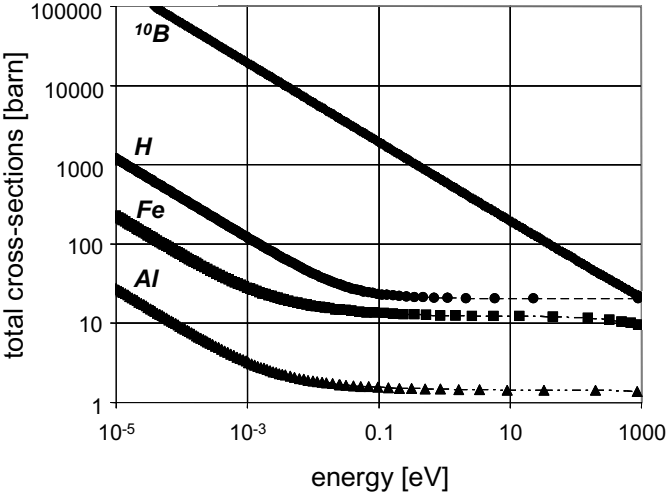
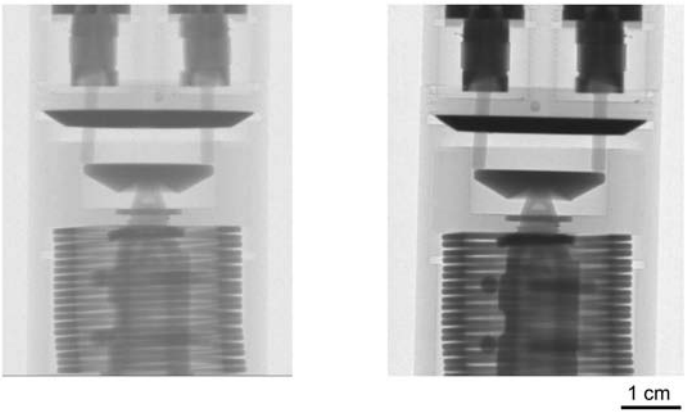


FIG. 13.5. Energy dependence of total neutron cross-section for four elements (Bragg scattering excluded) (Brookhaven National Laboratory, 2007).

this section the neutron beam is expanded further as it propagates to the left, its size being controlled by a set of beam limiters placed in a flight tube, basically an evacuated aluminium tube with 0.1-mm thick aluminium windows through which the neutrons easily pass. Section II contains a measuring position for small samples up to 100 cm^2 cross-section as the beam is still small here. These conditions are suitable for the μ-CT station and the differential phase-contrast setup discussed in Chapter 14. Section III is equipped for conventional tomography of larger samples up to 1600 cm^2 cross-section. All the flight-tube sections are flexible and can be removed whenever the experimental arrangement requires access along the beam axis. The entrance to the facility leads through a labyrinth that efficiently blocks particles radiation from the sample during exposure.

13.2.2 *Neutron sources and beam characteristics*

Different ways to produce neutrons have been described in Chapter 4. In order to be able to acquire tomograms with hundreds of projections in a reasonable time (e.g. less than one day), beamlines with high intensities are needed. Therefore, radioactive isotopes such as ^{252}Cf and even accelerator-based neutron generators inducing fusion between beams of deuterium and/or tritium ions and metal hydride targets are too weak for practical use in neutron tomography setups. For tomography, neutron reactors or spallation sources are indispenable. Reactors yield a stable and continuous neutron flux that enables reliable tomographic measurements with long exposure times. Spallation sources deliver a pulsed beam that is very useful for time-of-flight experiments and energy-selective radiography. In both cases, typical exposure times are in the range of a



(a) (b)

FIG. 13.6. Radiographic images of a pyrolytic injector acquired using, (a) thermal and, (b) cold neutron beam. The same experimental parameters – beam collimation, distance to the detector and detector system – were used in the two experiments.

few seconds at midflux facilities and fractions of a second at high-flux facilities, e.g. 10^9 neutrons/cm^{2}s at 'Neutrograph' (Institute Laue-Langevin, 2007).

An important characteristic of a neutron source is the mean energy of the neutron beam provided for experiments. For each application one has to decide which neutron energy is most suitable and whether polychromatic or monochromatic neutrons are required. There are two criteria of importance: how transparent the sample is for the particular neutron energy and how effectively transmitted neutrons can be detected. Attenuation is a material property and is directly related to the total neutron-interaction cross-section. As shown in Fig. 13.5 for the examples hydrogen, boron and iron, the total cross-sections strongly increase towards lower energies but are small at high energies. This means that large samples should be preferably investigated with high-energy neutrons, whereas weakly attenuating materials and small samples have to be studied with thermal or even cold neutrons.

Figure 13.5 shows that the use of cold neutrons for radiography increases the probability for interaction and thus image contrast. In this way the sensibility e.g. for the detection of small amounts of water or other hydrogenous materials in metal matrices can be improved. In addition, cold neutron beams can be easily modified by using diffraction or neutron optical techniques because of the favourable wavelength range of its spectrum. This enables radiography and tomography to be performed with more sophisticated measuring techniques such as energy-dispersive and phase-contrast imaging. Comparison of radiographic images taken with thermal and cold neutrons are shown in Fig. 13.6. Evidently, cold

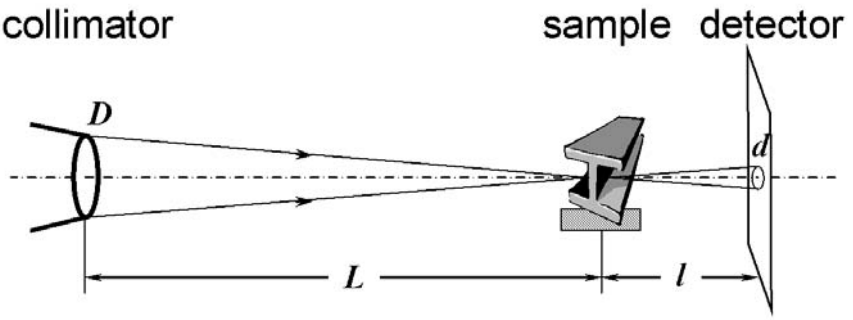


FIG. 13.7. Schematic representation of neutron tomography experimental arrangement involving a pinhole setup.

neutrons yield much better image contrast, which helps to resolve fine structures even in the aluminium body of the investigated part. On the other hand, transmission of cold neutrons through the plastic insulators is lower in comparison to thermal neutrons, which deteriorates the signal-to-noise ratio for the involved image pixels. This gives rise to artefacts for this part of the corresponding tomographic reconstruction as described in Chapter 5 and demonstrates the necessity to evaluate carefully the trade-off provided by different neutron energies.

13.2.3 *Geometry (beamline design)*

The beam geometry at a neutron tomography facility should be as close as possible to a point-source configuration in order to obtain a well-defined projection arrangement. In order to fulfill this requirement the source size D should be small in comparison to the distance from the source to the sample L, Fig. 13.7.

A measure of how close the beam geometry is to the ideal point-source configuration is given by the figure of merit called the 'L/D ratio':

$$\frac{L}{D} = \frac{l}{d}, \tag{13.2}$$

which determines the maximum blur d in the image at a distance l between the sample and the detector. In this way the quality of the radiography image depends on the ratio L/D, being the main parameter to characterize the performance of the facility (Schillinger, 2001). A larger L/D ratio provides better image resolution. Most of the neutron-radiography facilities worldwide possess an aperture changer, where a set of different apertures allow for performing experiments for different L/D ratios by changing D. Every increase of the aperture diameter D leads to a gain in neutron flux, while reducing image resolution, and vice versa. Depending on the experimental requirements – high flux or high resolution – an optimal L/D ratio can be chosen for a given situation.

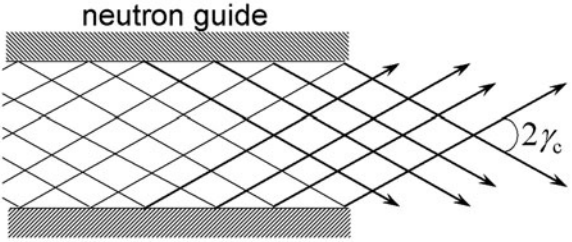


FIG. 13.8. Beam divergence caused by the neutron guide and characterized by the total reflection angle γ_c. The definition of the L/D ratio for a simple flight tube is no longer valid.

There is a small number of radiography facilities using cold neutrons that are placed at the end of neutron guides at relatively large distances from the reactor core. In this case, the guide helps to transport the neutron beam over long distances, preventing high losses of intensity. The principle of the neutron guide is based on the total reflection of the neutrons from the walls of the guide that are coated with a material having a high refractive index. One advantage of neutron guides is the absence of high-energy background radiation, see also Section 13.2.6.

Figure 13.8 shows the conditions as the neutrons leave the guide. The total reflection angle γ_c for neutrons is given by (Maier-Leibnitz and Springer, 1963):

$$\gamma_c(\lambda) = \sqrt{2\,(1 - n(\lambda))} \approx 10^{-3} \ldots 10^{-2}, \tag{13.3}$$

where $n(\lambda)$ is the refractive index of the coating for the applied neutron wavelength λ. Typical materials used for neutron-guide coatings include natural nickel, having a refractive index of $(1 - n) = 1.5 \times 10^{-6}$ for neutrons with $\lambda = 0.1\,\text{nm}$, which yields $\gamma_c = 1.73 \times 10^{-2}\,\text{rad/nm}$ or $1^\circ/\text{nm}$ (Maier-Leibnitz and Springer, 1963). In this way, the guide produces an energy-dependent divergence due to the fact that the different neutron wavelengths are reflected at different angles, which is not appropriate for high-resolution neutron radiography. In this case, the beam divergence can no longer be described by eqn 13.2 and for the L/D ratio the following expression has to be used:

$$\frac{L}{D} = \frac{1}{\tan 2\gamma_c}. \tag{13.4}$$

Typical L/D values for a neutron guide transporting cold neutrons are around 70, which is a very low value for the purposes of high-resolution neutron tomography (Schillinger, 2001). To improve this poor performance one can use a combination of a neutron guide and a pinhole geometry by placing a small diaphragm at the

end of the guide. Both sample and detector are then located at some distance behind the diaphragm and the configuration resembles that of Fig. 13.7 with the collimator being replaced by the diaphragm. The disadvantage of this solution is that due to the energy-dependent divergence produced by the guide the spectral composition of the beam will change with the distance from the beam axis in the plane transverse to the beam-propagation direction. Such spectral effects were simulated and discussed by Schillinger *et al.* (2005). Estimates of the expected L/D value can be obtained by direct application of eqn 13.2. For example, at a distance between the sample and the detector plane of 10 cm and a desired resolution of better than 0.1 mm the required L/D ratio should be of the order of 1000. Such conditions are available at only few places worldwide, see Appendix A.

Spatial resolution of a neutron-tomography setup is limited either by the beam characteristic (L/D ratio, spectral heterogeneities, *etc.*) or by the detector. Ideally, both components have the same performance. Recently, a μ-CT setup with a field of view of 27 mm and a resolution better than 50 µm was realized, largely by improving the scintillator and the optical components of the camera (Lehmann *et al.*, 2007). By further improving the performance of the detector systems for neutron imaging, the requirements related to beam parameters will increase. Using the conventional pinhole arrangement where a small pinhole is used to achieve higher L/D ratios, the beam intensity at the sample position will decrease rapidly. A much more effective way to increase the L/D ratio is to apply beam-focusing techniques as in the case of the X-ray imaging arrangements described in Chapter 7. Due to the charge neutrality of neutrons it is impossible to use electrostatic fields for beam focusing as for electrons. An effective way to deflect a neutron beam and hence to focus it is to use a focusing neutron guide where the neutrons undergo a total reflection from the specially coated guide walls (Kardjilov *et al.*, 2005). In this way, the beam can be focused onto a small spot of a millimetre size resulting in a high divergence of the propagating beam. In addition, the defined cone beam geometry allows one to apply different magnification ratios, which helps to increase the spatial resolution in neutron-imaging experiments. Alternative focusing techniques have been proposed: magnetic lenses (Shimizu *et al.*, 2000), refractive lenses (Eskildsen *et al.*, 1998) or fibre optics (Chen-Mayer *et al.*, 1997, Masschaele *et al.*, 2001), but it is uncertain whether they can be efficiently used for neutron imaging.

13.2.4 *Detectors*

For tomography, a digital detection system is required. In this way, the detector can remain stationary during inspection, providing a fixed sample–detector alignment along the beam axis. Such a stationary system cannot be realized with detectors requiring post-processing steps such as development of photographic emulsions or scanning of imaging plates due to the misalignment associated with repositioning between exposures. Therefore, electronic detectors are indispensable that allow for readout of images while remaining in the measurement position.

TABLE 13.2. Main nuclear reactions used for neutron detection. The energy specified is the Q value for the kinetic energy of the reaction products, $*$ denotes an excited state.

^{3}He + n	$\rightarrow$	^{3}H + ^{1}H + 0.77 MeV
^{6}Li + n	$\rightarrow$	^{3}H + ^{4}He + 4.79 MeV
^{10}B + n	$\rightarrow$	^{7}Li + ^{4}He + 2.78 MeV (6%)
	$\rightarrow$	^{7}Li* + ^{4}He + 2.3 MeV (94%)
^{155}Gd + n	$\rightarrow$	^{156}Gd + γ + conversion e^{-} (7.9 MeV)
^{157}Gd + n	$\rightarrow$	^{158}Gd + γ + conversion e^{-} (8.5 MeV)
^{235}U + n	$\rightarrow$	fission products + 200 MeV

Neutron detection is an indirect technique mainly based on the creation of free electric charge carriers. Neutrons can produce such carriers by collisions with nuclei or neutron capture reactions with the nuclei in the detecting material that lead to the emission of charged particles, e.g. α-, β-particles, protons, tritons, etc., which then can be detected by more conventional means. The most important elements for thermal or cold neutron detection exhibit very high neutron capture probabilities, e.g. ^{6}Li, Gd, Dy, ^{10}B. Some of the possible reactions with materials suitable for neutron detection are listed in Tab. 13.2.

In a neutron-imaging detector, the amount of electric charge produced by nuclear reactions is often not measured directly, but converted into another more observable physical entity such as light. In neutron scintillation screens, the charged particles stimulate light emission in zinc sulphide.

Development over years has shown that ZnS layers doped with colour centres to shift the emission wavelength are most useful with respect to light efficiency. However, these are limited in spatial resolution by the dissipation length of the reaction products from neutron capture (e.g. ^{3}H and ^{4}He from neutron capture by ^{6}Li) and the diffuse light emission occurring when single grains of the scintillators are excited. A schematic representation of the physical processes that occur in a ^{6}LiF+ZnS:Ag scintillator screen at the conversation of neutrons in visible light is given in Fig. 13.9.

The thickness of the opaque scintillator material is an important factor limiting spatial resolution and light-output efficiency due to light diffusion and self-shielding. There are attempts to reduce the scintillator thickness, as in the case of synchrotron tomography, but the manufacturing process and the smaller light output have been hitherto limiting factors. Neutron sources are inherently flux limited in contrast to synchrotron sources, so that the technologies and strategies used for developing synchrotron detectors cannot be transferred directly to the field of neutron detectors.

There are many different options available now for detecting neutrons in neutron imaging. The main requirements a neutron tomography detector should satisfy can be summarized as follows:

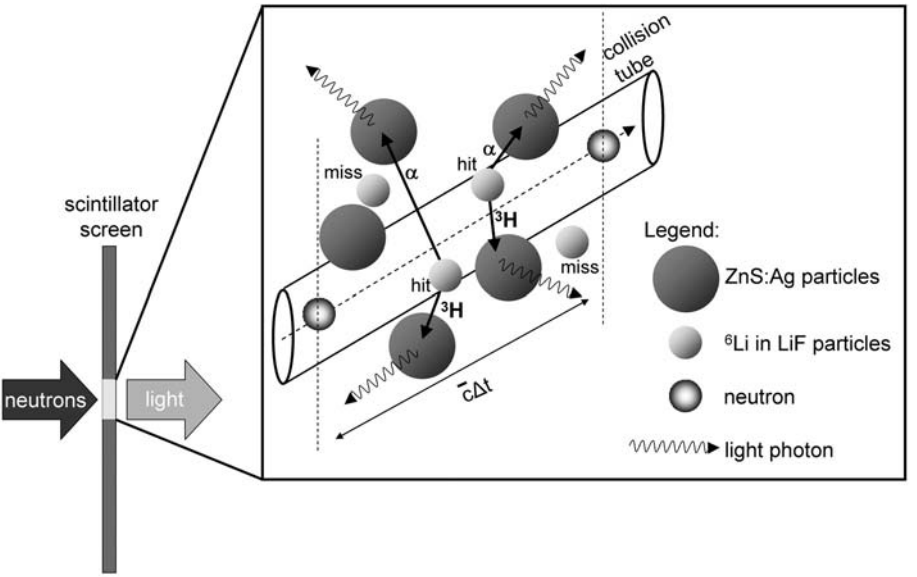


FIG. 13.9. Conversion principle of neutron scintillator screens. A neutron with velocity $\bar{c}$ enters the scintillator and is absorbed by a ^{6}Li nucleus present in the LiF converter contained in the scintillator. As a result, ^{3}H and ^{4}He reaction products are emitted as positively charged ions that then interact with the ZnS:Ag particles around in a second step, eventually leading to the emission of visible light that can – at least partially – be detected by conventional imaging systems such as a CCD camera. Some neutrons miss all ^{6}Li nuclei and are not detected. Thus, neutron detection is a two-stage conversion process.

- the detector should give a digital output signal,
- the signal should depend linearly on the applied radiation dose,
- the dynamic range of the imaging systems should be as large as possible (e.g. 16 bit),
- the detector position must be fixed in relation to the sample stage,
- the active detector area should be larger than the object in any projection,
- the sensitivity of the detection system with respect to perturbing beam components (γ-rays, fast neutrons, *etc*), should be as low as possible,
- the dark noise level must be low and constant throughout measurement,
- the detection system should be resistant to neutron and γ-rays. If this is not the case, a corresponding shielding should be setup in order to prevent detector damage.

For these reasons, some of the available detection systems are not applicable to neutron tomography. This applies to film and imaging plates (IPs) because of

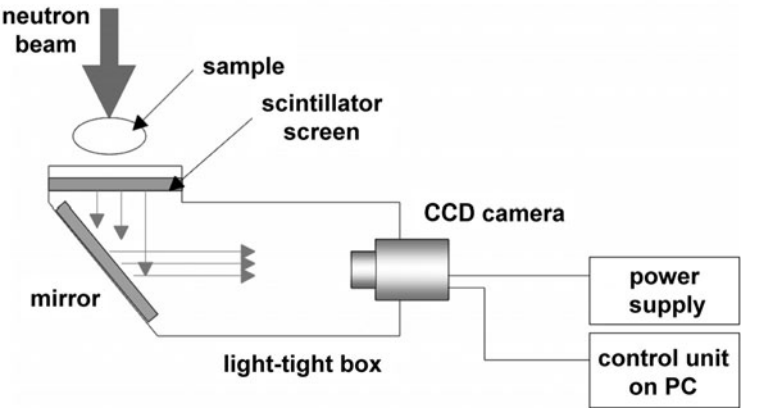


FIG. 13.10. Detection system for neutron imaging (suitable for tomography).

their need for *ex-situ* readout and the associated difficulties with reinstallation and repositioning. Because pixel-precise alignment seems to be a technical challenge there is no such system in use for tomography at the moment. This is in contrast to transmission radiography where IPs are very useful and are replacing traditional photographic emulsions. On the other hand, electronic systems such as CCD cameras and amorphous silicon panels can be damaged by the radiation (neutron and γ-rays). In order to avoid this, a special shielding should be designed and set up for prevention of damage to the electronics. Despite these measures signs of radiation damage expressed by 'bad pixels' or an increased electronic noise ('dark current') can be observed after longer service times of detector systems.

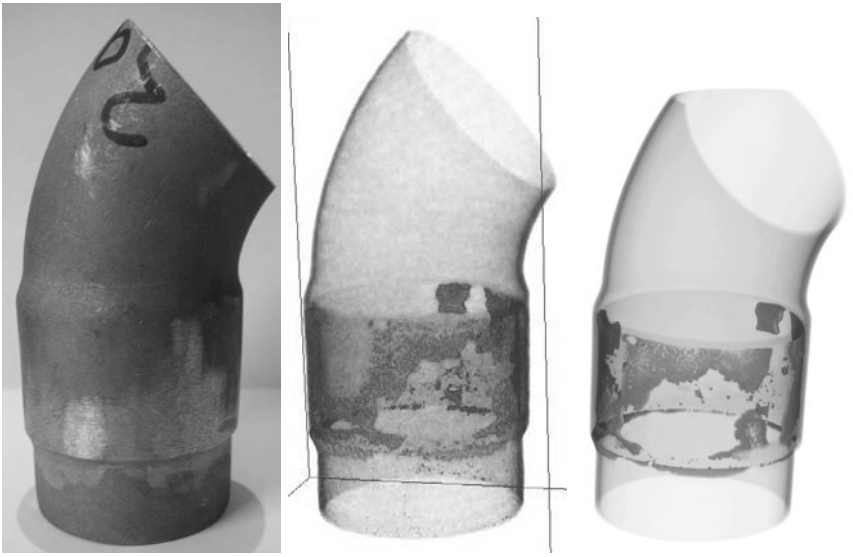
In the following, the two most common neutron detection systems used for tomography will be presented.

13.2.4.1 *CCD cameras* Specialized CCD-camera detectors are frequently used for neutron tomography (Pleinert *et al.*, 1997). Due to the extremely high light sensitivity of the CCD chip all components of the detector system (scintillator screen, mirror, camera, shielding, *etc.*) are placed inside a light-tight box. An example of such a construction is shown in Fig. 13.10. The camera is located at a position outside the beam direction to prevent radiation damage of the CCD chip and is optically coupled to the scintillator screen via a 45° mirror and a lens. This type of detection system is installed at the end position of the ICON beamline (Brönnimann, 2004), providing a flexible field of view (FOV) ranging from 5 cm to 35 cm width, depending on the position of the CCD camera with respect to the scintillator screen and the lens system used. The resolution limitation Δ of a detector system with a corresponding number of pixels n_p can be expressed by:

$$\Delta > \frac{\mathrm{FOV}}{n_p}. \tag{13.5}$$

The typical number of pixels in commercial CCD cameras nowadays is between 1024 and 2048 (one direction), and the optical resolution is calculated (eqn 13.5) to nominally between 0.03 mm and 0.2 mm. The primary detector for the neutrons is a scintillator screen placed perpendicular to the beam. The most common one is a ^{6}Li-doped ZnS screen with light emission in the blue-green spectral range. Although such scintillator screens have been in use for many years there is further potential for improvements with respect to light output, spatial resolution and homogeneity. Scintillator screens based on gadolinium oxysulphite (Gadox) are much less sensitive for neutrons and therefore not in practical use for routine utilization. Depending on the neutron flux the exposure time per frame is of the order of a few seconds to some minutes. Such long exposure is only possible with CCD cameras cooled either by Peltier elements or liquid nitrogen in order to minimize thermal noise. Most of the neutron tomography installations worldwide (see Appendix A) use CCD cameras from Andor Technology (2007). The sensitivity and required full-well capacity is given by the large pixel size on the chip (13 μm). Using the backilluminated chip technology, the light-collection efficiency has a maximum for the spectral range of the ZnS scintillators. This allows for an optimal performance of the detector system (low noise level, stable dark current, reasonable readout speed, well-adapted software tools, reasonable pricing).

13.2.4.2 *Flat panels* An alternative detection system also using scintillator screens as primary neutron detectors, are *amorphous silicon* (a-Si) flat panels, cf. Section 10.7.1. They consist of arrays of photodiodes coupled with an active TFT (thin-film transistors) readout matrix per pixel, where the scintillator is in close contact with the semiconductor. This kind of setup is possible because the base material (amorphous Si) tolerates direct neutron and γ-ray exposure much more than a CCD-camera chip would do. On the one hand, light coupling of the scintillator to the diode array is much more efficient in this geometry compared to the lens-coupled CCD. On the other hand, the light-to-voltage conversion efficiency in the flat-panel substrate is much lower compared to a high-sensitivity CCD array consisting of single crystals. The operation regime of flat panels differs from that of CCD cameras since the exposure–readout–erase process is permanently running like in an analogue cathode tube-based TV screen. Therefore, frames are generated in a continuous mode, where a frame rate is defined in advance – e.g. between 0.5 fps and 30 fps – and the resulting images are stored in an interim memory (until it is full). The images are transferred to a hard disk after the end of the measurement. Flat-panel devices are faster in data acquisition compared to CCD systems and a full sequence of projections can be obtained within seconds to minutes. Therefore, this setup can be used to investigate time series of tomography runs in a reasonable time. Time-dependent tomography sequences could be possible in this way. One should consider the amount of data



(a) (b) (c)

FIG. 13.11. Comparison of tomographic images of the solder connection of two Cu tubes with 4.5 cm diameter – (a) shows a photograph – obtained with, (b) a CCD detector and, (c) an a-Si flat panel.

produced in this case: one single projection requires about 5.6 MB. When 300 projections are taken within 10 s, a data volume of 1.6 GB is produced. The disadvantage of the a-Si technology is the lower dynamic range and the smaller signal-to-noise ratio compared to CCDs. Both parameters reduce the quality of the obtained tomography volume data. Compared to data collected by a CCD-camera system, a clear drawback has to be accepted. On the other hand, the shorter exposure time has two advantages for neutron tomography: faster observation of processes and lower activation risk during exposure. A comparison of two tomographic images, one obtained with a CCD camera, the other with an a-Si panel is given in Fig. 13.11. The better dynamic range and signal-to-noise ratio of the CCD-based detection system in comparison to a-Si flat panel allows for much more detailed tomographic reconstruction of the solder joint between two Cu tubes. This determines the field of application of a-Si flat panels as systems for high-speed tomography and fast preliminary inspection of samples.

Table 13.3 gives an overview of the different properties of the aforementioned detection systems for neutron tomography. Because of the high price of both systems, a preliminary study should be made as to which of the available systems has the lowest cost/performance ratio. For the moment, CCD-based systems are the most flexible and cost-efficient ones.

Figure 13.12 summarizes the specific performance of neutron detectors with respect to spatial and time resolution.

TABLE 13.3. Comparison between different detection systems in use for neutron tomography.

	units	scintillator + CCD camera	scintillator + a-Si panel
active area	cm^2	4×4 to 30×30	10×15 to 20×25
nominal pixel size	mm	0.02 to 0.15	0.127
pixels		1024^2 to 2048^2	$\approx 1500^2$
dynamic range	bit	16	12
S/N ratio		120	25
exposure time	s	2 to 60	0.067
read out time	s	2	during exposure
frame rate	fps		0.5 to 30

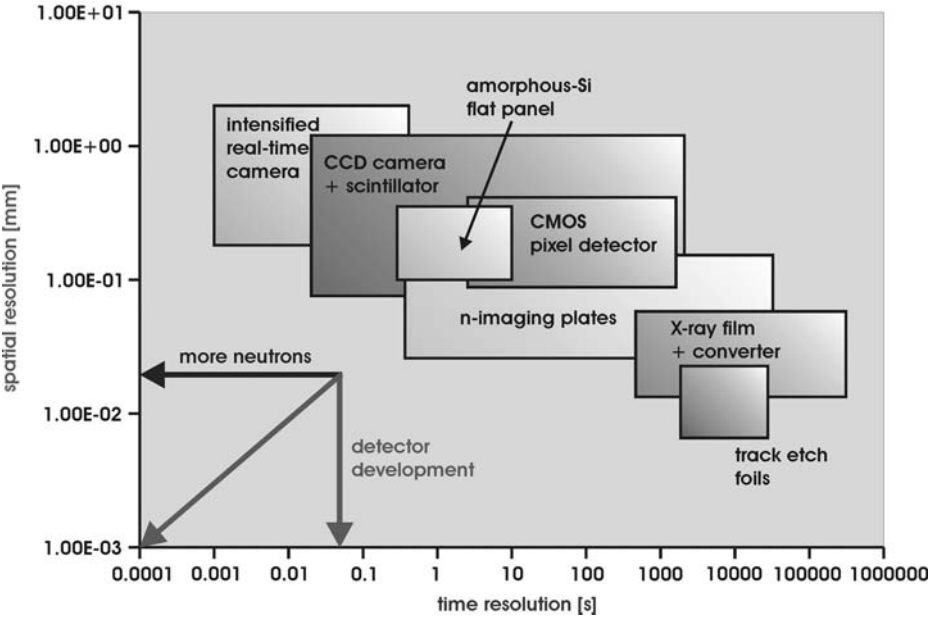


FIG. 13.12. Performance of neutron-imaging detectors with respect to spatial and time resolution.

The family of detection systems for neutron imaging shown in Fig. 13.12 covers quite a large interval of spatial and temporal resolution. Of course, the lower bounds of the presented parameters are of interest for the further development of neutron tomography. Improvement of time resolution can be achieved by increasing the neutron flux through utilization of more powerful neutron sources or by compressing or focusing of the beam to a small spot. High neutron flux is necessary also for detectors with high spatial resolution as smaller pixel sizes

imply a reduced number of neutrons per pixel, which can only be compensated with longer exposures. A lot of the improvement of spatial resolution has been connected to new technical developments related to both the optics and the electronics of the detecting systems. Some of the recent developments are reviewed by Lehmann *et al.* (2007) and Nova Scientific (2007).

13.2.5 *Sample stage*

The sample position at a neutron tomography facility should be equipped with translation, tilt and rotation stages. The hierarchy of the stages is as follows: the sample is placed on a rotation table, the rotation axis of which is adjusted by tilt stages positioned below it. At the bottom, a translation stage moves both the rotation and tilt units out of the beam for recording 'reference open beam' images, to be used later for normalization of the tomographic projections.

The rotating table is the most essential unit of the sample environment part. It enables one to acquire radiographic projections of the object from different angles. As explained in Section 2.3 the number of projections recorded at equal angular steps depends on the field of view and the detector resolution. In practice, it usually ranges between several hundred to a thousand projections over 180°. A key requirement is a high precision of the angular positioning of the rotation stage which is a challenge, especially if sample masses are high.

13.2.6 *Shielding*

As is seen in Fig. 13.4 the shielding can be a significant part of a neutron-tomography facility. A powerful neutron source is a pre-condition for high-quality neutron imaging. For every radiation source, there are legal requirements regarding safe operation and radiation protection of the personnel. A neutron-imaging facility is therefore located within a measurement room ('hutch'), constructed of thick concrete shielding walls, and accessible only through a labyrinth secured by a safety door. The reason for the large dimensions of the used shielding blocks are the fast neutrons that are presented at beam ports installed close to the reactor core. In order to stop them more efficiently, a concrete with special additives (heavy concrete) is used.

Neutron guides are mostly slightly curved, which has the advantage that fast neutrons and γ-rays from the reactor are not transported to the end of the guide but leave the guide through the walls and are absorbed after. This lowers the background radiation significantly, improves signal-to-noise ratios and also greatly relaxes the need for radiation protection. Instead of massive concrete shieldings thin composite layers of boron carbide imbedded in plastic (absorbing neutrons) and lead (absorbing prompt γ-radiation from the boron shielding and radiation emitted by the samples) are sufficient. Boron plastic is a very effective absorber and a layer of 3 to 5 mm is sufficient to stop completely the incident cold neutron beam. Such shielding is used at the cold neutron radiography station CONRAD at HMI (Hilger *et al.*, 2006).

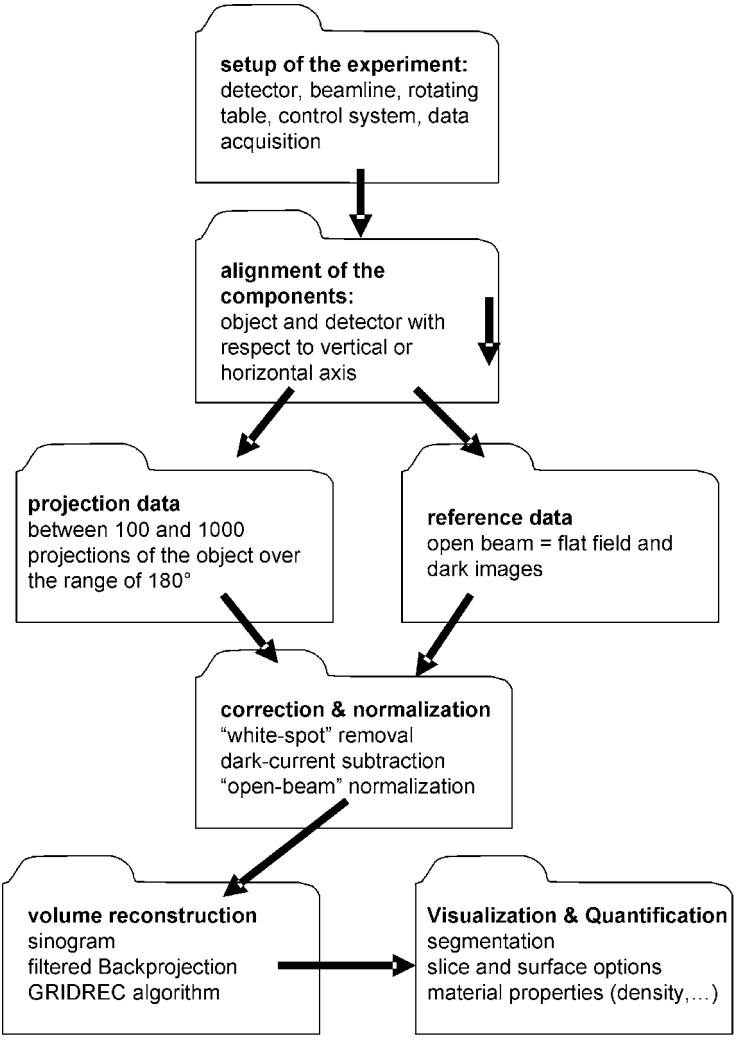


FIG. 13.13. Scheme for working steps during tomography inspection.

13.2.7 *Data acquisition and processing*

Once the conditions of the neutron tomography setup are defined the experiment can be performed in the manner given in Fig. 13.13.

The single steps of the algorithm are explained in detail as follows:

- The object has to be placed firmly on the sample stage as close as possible to the detector plane in order to minimize geometrical blurring (according to eqn 13.2).
- Before or after image acquisition some 'open-beam' images (radiation on,

but sample removed) and a 'dark-current' image (radiation off) have to be taken under the same experimental conditions.

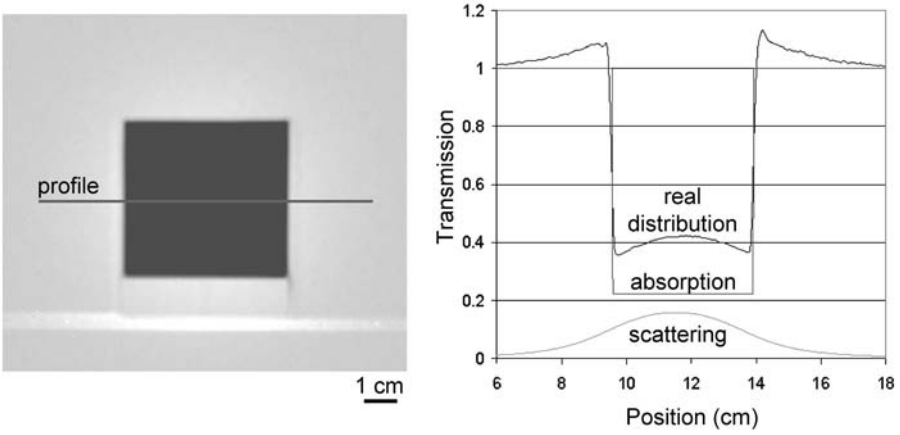
- By rotating the object around a vertical (or horizontal) axis, the necessary number of radiographic projections n_p is acquired in equal angular steps of $180°/(n_p+1)$.
- 'Bad pixels' can either be permanently damaged detector elements or such with an overflow caused by hits of γ-radiation. In most cases, image filter procedures (e.g. median filters) are able to remove such artefacts without strongly affecting the neighbouring pixels.
- The images have to be corrected to eliminate the influence of beam distortion and detector inhomogeneity.
- Normalization is needed for all images with unequal exposure (e.g. by beam fluctuation during acquisition), where the integral intensity over a region of interest (ROI) outside the object area is used as a normalization factor.

- Depending on the detection system the dark current is removed either using the acquired 'dark image' or by subtracting an averaged constant value.
- Finally, all projection data are divided by the 'reference open beam' image, which was previously also filtered for 'bad pixels' and corrected for the dark current.

The final aim of tomography is to determine the matrix of attenuation coefficients $\mu(x, y, z)$ over the sample volume. The mathematical reconstruction algorithms are analogous to those used for other tomographies and are described in Chapter 2. In practice, reconstruction is carried out in individual (x, y) planes, after which the data are stacked in the third (z) dimension parallel to the rotation axis.

An essential assumption in the tomographic reconstruction is that the beam attenuation follows the exponential law, eqn 4.11, where the linear attenuation coefficient is considered to be a superposition of absorption and scattering cross-sections for the material under investigation. However, there is a certain probability for scattered neutrons to reach the detector screen and to contribute in this way to image formation. The higher the contribution of scattering, the higher is the effect of a 'sky shine' in the transmission image, as shown in Fig. 13.14 for a water sample. This induces strong artefacts in the tomographic reconstruction that should be adequately corrected.

In order to analyse the contribution of scattering effects, the whole image-formation process, including description of beam characteristics, sample transmission and detector response was simulated with the help of a Monte-Carlo method using the `MCNPX` code (Waters, 2002). The results of the study are presented elsewhere (Hassanein, 2006, Hassanein *et al.*, 2006). The correction algorithm named `QNI` (for quantitative neutron imaging) is given in Fig. 13.15. The kernel of the analysis is the estimation of the so-called *point-scattered function* (PScF) describing the scattering contribution for each point of the object.



(a) (b)

FIG. 13.14. 'Skyshine' around a 5-mm thick water sample: (a) image with 'open-beam' correction, (b) superposition of the absorption and scattering signal to the real distribution indicated by the profile (Hassanein, 2006).

Because of the difficulty to calculate the corresponding PScF for all combinations of materials, beamlines and sample–detector distances, a parameterized approximation was introduced by building a special data library.

An example of the application of the `QNI` algorithm is shown in Fig. 13.16 where a model system of soil cubes with different moisture contents was investigated by neutron tomography.

In this case, conventional tomographic reconstruction provides a blurred image of the distribution of the soil cubes in the container. The margins of the cubes cannot be resolved clearly. Application of the `QNI` algorithm allows for better resolution of interfaces between the soil cubes as well as for a reliable quantification of the moisture content in each part of the model system.

13.3 Limitations in neutron tomography

13.3.1 *Temporal and spatial resolution*

Even powerful neutron sources deliver much less intense beams compared to synchrotron radiation sources. Consequently, either time or spatial resolution or the signal-to-noise ratio will be limited. Under the best conditions and great difficulties, a temporal resolution of 1000 fps can be achieved. Recently, it was shown that a spatial resolution of 20 line pairs/mm can be obtained with a μ-CT setup (Lehmann *et al.*, 2007). Further improvement might be possible to a certain extent, but the best resolution will be at least one order of magnitude lower than that of synchrotron tomography. The same is valid for the signal-to-

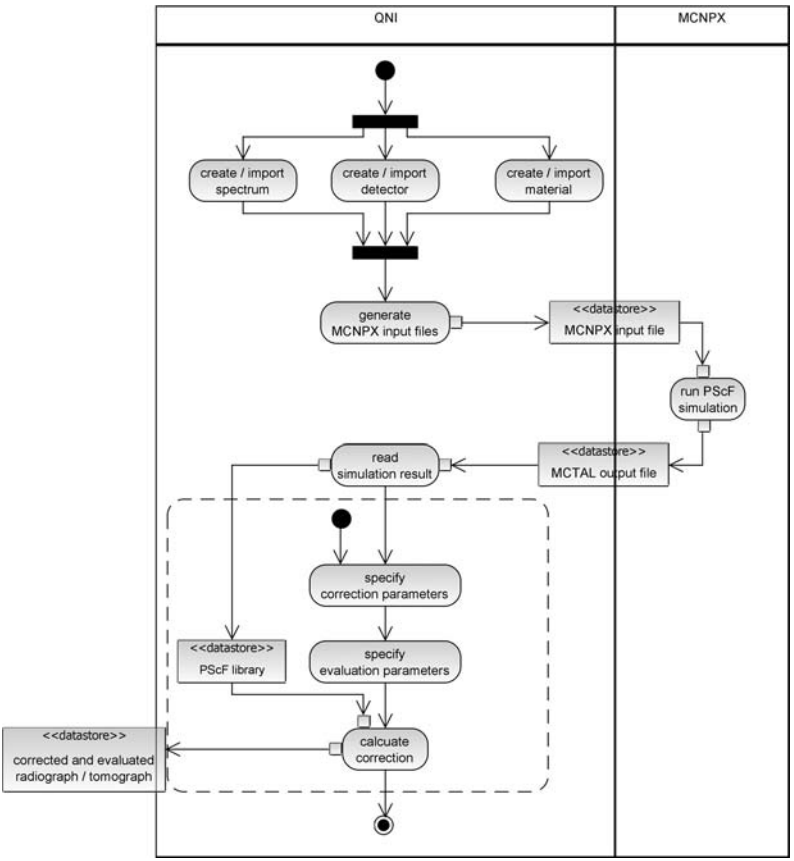


FIG. 13.15. Structure of the procedures for the correction algorithm `QNI` in use for quantitative neutron tomography (Hassanein, 2006).

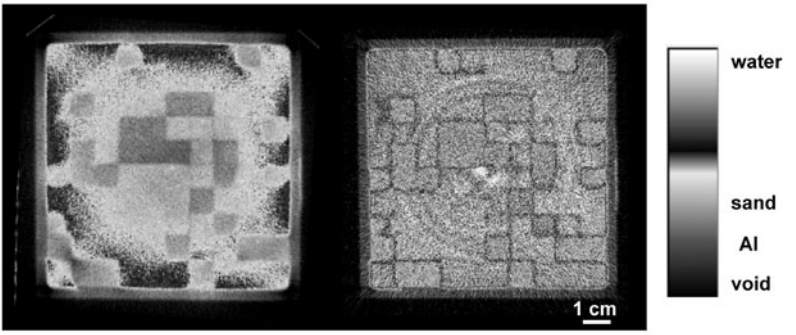
noise ratio, which is of the order of about 100. Some improvement can be made by extending exposure time.

13.3.2 *Sample activation*

Although some limitations can be overcome using longer exposure, the risk for (radio-)activation of the sample material has to be considered. Depending on the sample, the neutron beam intensity ϕ and the exposure time t_1, the activation at a time t_2 after the end of exposure can be estimated in the following way:

$$A[Bq] = N\sigma_{\text{act}}\phi\left(1 - e^{-\lambda t_1}\right)e^{-\lambda t_2}.$$

The activation cross-section σ_{act} is the probability for capture of a neutron by the target material containing N nuclei and creation of a radioactive nuclide



(a) (b)

FIG. 13.16. Tomographic slice of a model soil system, (a) before, and (b) after the application of QNI algorithm for correction of the scattering component (Hassanein, 2006). This image is reproduced in **colour** in the central part of the book.

that decays with the decay constant λ [s^{-1}]. σ_{act} and λ are available as tabulated values, see e.g. Brookhaven National Laboratory, Brookhaven, USA (2007). Because typical neutron fluxes in tomography experiments are of the order of 10^6–10^8 $cm^{-2}s^{-1}$ only (compared to a dedicated irradiation experiment with fluxes of the order of $10^{13}\,cm^{-2}s^{-1}$), the obtained activities are never so high that there is a serious problem with handling and returning samples if the half-life ($T_{1/2} = \ln 2/\lambda$) is of the order of hours to days. The only material to be avoided for long-term exposure is cobalt due to the long half-life of ^{60}Co (5.3 years) and the large activation cross-section of ^{59}Co (37.18 barn). Other materials were found to represent manageable risks when studied by neutron tomography.

13.4 Selected applications and experimental options

13.4.1 *Conventional tomography*

The range of applications for neutron tomography is defined by the following key experimental parameters:

- Attenuation properties of the sample material: the most reasonable sample size is given by the linear absorption coefficient, see Tab. 13.1. Depending on the material this size can range from tenths of millimetres for strong absorbers to tens of centimetres for weakly absorbing materials.
- Beam diameter of the imaging facility: this is very specific for the individual installation and ranges from about 5 cm to 50 cm, see Appendix A. The sample has to fit into the beam or has to be scanned by moving it through the beam.
- Detector size with respect to the FOV: it can range from a few centimetres up to the full size of the neutron beam, see Tab. 13.3.

TABLE 13.4. Overview of present applications of neutron radiography and tomography in different fields.

field of research	**task**
archeology	provenance and structure of metallic, ceramic and wooden objects
biology	root growth, water uptake of plants
civil engineering	water protection
electrochemistry	fuel-cell performance, water distributions, batteries
geology	porosity of rocks, water permeability, mineral and organic enclosures
hydrodynamics	two-phase flow
mechanical engineering	engines in motion, soldering and casting process
paleontology	bone structure of dinosaurs
pedologie	water penetration into soil, structural characterization
wood research	moisture uptake and loss, wood structure, bonding and protection

These conditions together indicate that neutron radiography and tomography is more suitable for macroscopic investigations than for microscopic ones, mainly because of the current resolution limit of the detector systems of $\approx 50\,\mu m$. Still, there is a sufficient number of applications where neutron tomography perfectly fits the requirements, especially due to the high imaging contrast and the high penetration depth that the neutrons can provide.

With respect to time resolution and the required acquisition time for a complete set of projections, the detector performance plays the major role. It is the trade-off between image quality in terms of signal-to-noise ratio and exposure time per projection on the one hand and the number of projections on the other that influences both spatial resolution and reconstruction features. The best data quality is obtained with slow-scan CCD cameras within about one hour. Series of tomograms can be obtained with flat-panel a-Si detectors within several minutes. This can easily be derived from the performance data in Tab. 13.3.

Due to the limited number of installations (see Appendix A) and the small amount of practical experience (since about 1996) neutron tomography is just beginning to be used on a routine basis for industrial and scientific applications. Therefore, this chapter can only show a few selected examples of previous investigations and outline a vision for further applications. Table 13.4 names some of the research topics. In the following, some prominent examples will be presented, thus demonstrating the advantage of neutron tomography.

13.4.1.1 *Fuel cells* In comparison to combustion engines, fuel cells are attractive because they allow for a highly efficient conversion of chemical energy stored in a fuel into electricity. A major potential application of fuel cells is in the au-

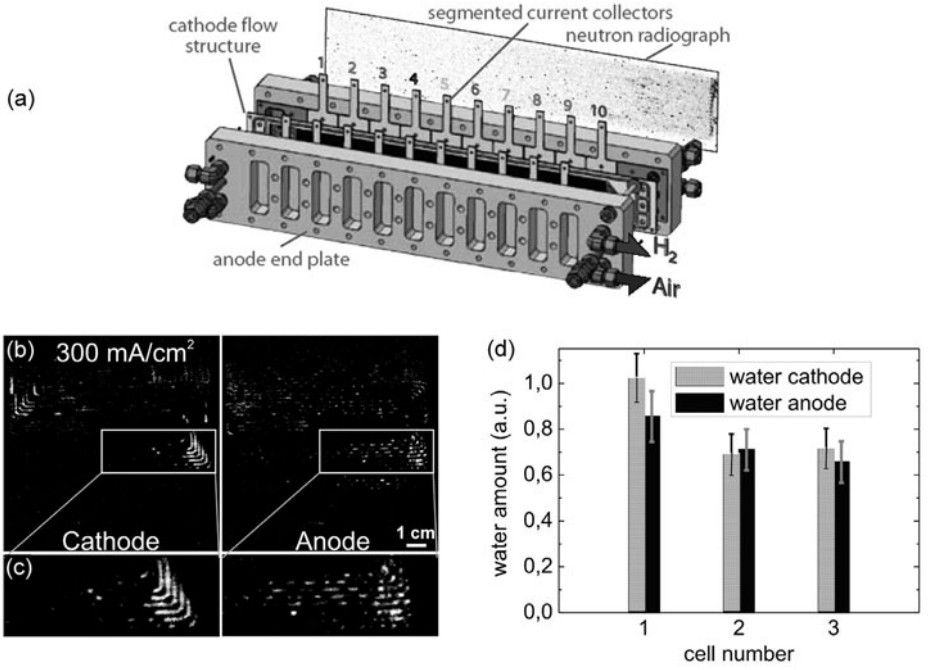


FIG. 13.17. (a) The changing water content of the fuel-cell membrane affects fuel-cell performance. Performance characteristics, i.e. current density and cell resistance, of a PEM fuel cell are shown as a function of channel length. (b) Water distributions in the anodic (left) and cathodic (right) flow-field channels of the first cell in a triple fuel-cell stack ($I_0 = 300\,\mathrm{mA/cm^2}$); (c) corresponding enlargements of the marked areas. (d) Total water content in the anodic and cathodic flow fields of all three cells at $I_0 = 300\,\mathrm{mA/cm^2}$ (Kramer *et al.*, 2005, Manke *et al.*, 2007*b*). (b–d) reprinted with permission from Manke *et al.* (2007*b*), copyright 2006, American Institute of Physics.

tomotive industry where they are considered promising alternatives to combustion engines. Characterization and optimization of the physicochemical processes taking place within the device are a prime objective of fuel-cell research and development worldwide. The electrochemical reaction of hydrogen with oxygen in a fuel cell produces water, an essential aspect of fuel-cell operation (Manke *et al.*, 2007*c*). Neutrons can monitor water distributions within running fuel cells non-invasively because they easily penetrate the exterior holding plates and yield radiographs with sufficient spatial and time resolution, Fig. 13.17(a). Used simultaneously with other advanced, spatially resolved characterization methods of the cells' electrochemistry, neutron radiography provides important insight into fuel-cell operation. In order to distinguish between the water distribution on the cathodic and anodic side, a tomographic investigation can be performed.

Some results are presented in Fig. 13.17.

In Fig. 13.17(b) the water distributions in the cathodic (left) and the anodic (right) flow fields of the first cell in the stack are displayed (Manke *et al.*, 2007*b*). Only the water distribution (white) is shown in the images. Due to different flow conditions at the turns of the flow-field channels, most water has accumulated in these areas of the flow field, Fig. 13.17(c). Despite the external humidification of the cathodic gas stream and the fact that the cathode is the 'water-producing' electrode, the amounts of water in the cathodic and anodic flow-field channels were found to be almost identical. This can be explained by backdiffusion from the cathode to the anode, which is one of the key factors for a uniform humidification of the active area of the fuel cell. For the cathode and anode in each cell the amount of water was calculated by summing over the various water-containing volume parts. Only water clusters larger than the spatial resolution of 300 µm are taken into account by this method. However, as the flow-field surface was hydrophobic, preferably larger water clusters were formed, and therefore the majority of liquid water in the flow-field channels was detected. Figure 13.17(d) shows the calculated amount of water in the cathodic and anodic channels of each cell at $I_0 = 300\,\mathrm{mA/cm^2}$. The different cell parts contain a comparable amount of water, with rather small variations around 20-30%. The estimated measurement errors of around 10% are based on the choice of threshold and beam-hardening artefacts.

13.4.1.2 *Batteries* Alkaline Zn/MnO_2 batteries are still one of the most used battery types. Many advances have been made to enhance the lifetime especially under high power and high drain rates. Alkaline batteries consist of MnO_2 at the cathode and Zn powder at the anode. While MnO_2 is used as a solid mixture with graphite, the Zn powder is suspended in gelled KOH electrolyte. During electric discharge MnO_2 is reduced by a solid-state intercalation of H^+ into the MnO_2 lattice, while at the anode zinc is oxidized by a dissolution precipitation process. The main chemical processes have been written down on page 154.

In parallel with zinc oxidation on the anode side the pressed MnO_2 on the cathode side swells, cracks and breaks with increasing discharge. The main reason for this is hydrogen intercalation into the MnO_2. Neutron tomography allows us to quantify the distribution of intercalated hydrogen (Manke *et al.*, 2007*a*).

Figure 13.18 shows cross-sections through the tomogram of a C-block battery. Due to the resolution of 400 µm, metal particles or cracks in the MnO_2 are not visible, but the hydrogen density profile can be measured. In the fully charged battery, Fig. 13.18(b) (left), no significant variations in the attenuation density both at the anode and the cathode are visible, but the images show variations in the electrolyte content inside the separator. The parts containing more electrolyte, possibly due to holes in the separator fleece, appear white, while the regions with a very low amount of electrolyte are almost black. After 6 h and 24 h of discharge some hydrogen at the anode is lost and has moved to the cathode. Hydrogen is shifted from the outer regions of the anode where zinc is oxidized at

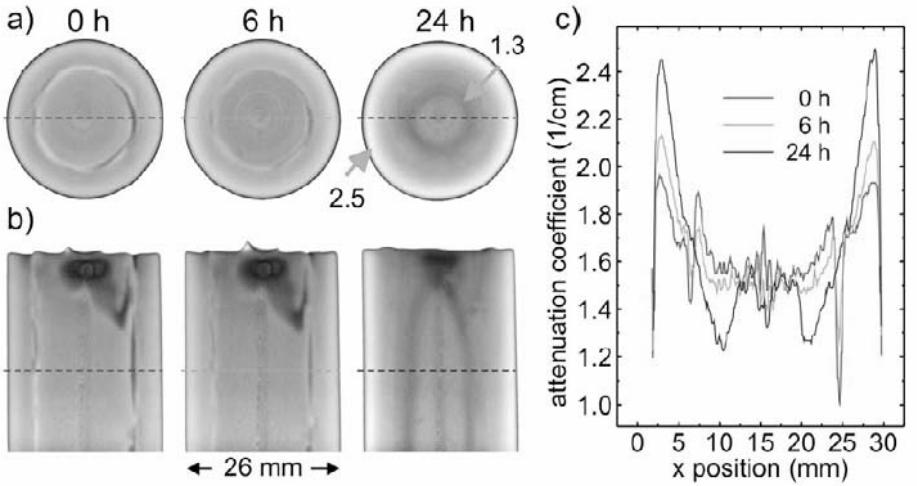


FIG. 13.18. Images of a C-block-type alkaline battery in three different stages of discharge as obtained by neutron tomography. Horizontal (a) and vertical (b) cross-sections of completely charged battery (left) and after two steps of discharge (centre and right), measured attenuation coefficients (in cm^{-1}) are given in (a); (c) variation of the attenuation coefficient along the marked lines in (b). Reprinted with permission from Manke *et al.* (2007*a*), copyright 2006, American Institute of Physics.

the outer side of the cathode close to the steel case. This results in a white and a dark grey ring in the right image of Fig. 13.18(a) and white and dark grey stripes in the corresponding image of Fig. 13.18(b). In Fig. 13.18(c) horizontal attenuation profiles through the axis of the battery in the three stages of discharge are compared. Again, the decrease of the attenuation in the fully discharged battery (at $x \approx 10\,\text{mm}$ and $x \approx 20\,\text{mm}$) is due to the loss of hydrogen that has moved to the outer sides of the battery (at $x \approx 6\,\text{mm}$ and $x \approx 30\,\text{mm}$). The distribution of intercalated hydrogen within the cathode is not homogeneous. Most hydrogen can be found close to the steel case. This investigation shows that neutron tomography can contribute to the optimization of electrochemical processes in current sources.

13.4.1.3 *Metal bonding* Adhesives used for metal bonding are hydrogenous epoxy–resin compounds. Due to the high sensitivity of neutrons to small amounts of hydrogen, even very thin layers of adhesives applied in metal bonding become visible with neutron radiography or tomography. Almost planar distributions of adhesives can be visualized by plane radiography, whereas complicated spatial arrangements, such as the glued joint shown in Fig. 13.19, must be investigated with neutron tomography. Thus, neutron radiography helps to validate bonding

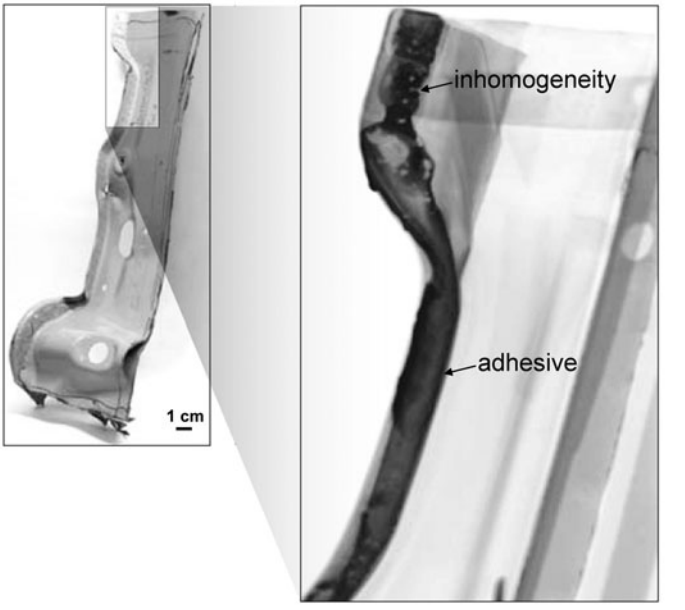


FIG. 13.19. Part of a bonded automobile frame. Neutron tomography reveals the distribution of adhesive within the glued joint, revealing bonding inhomogeneities (Paul Scherrer Institut, 2007).

procedures, complementing other NDT methods more readily available in the production environment.

13.4.1.4 *Renaissance bronzes* Renaissance bronzes from the Rijksmuseum Amsterdam were investigated by neutron tomography in order to study their structure and the way in which they were cast (Rijksmuseum Amsterdam, 2007). Old bronzes usually have quite a high lead content, which makes them ideal candidates for non-destructive analysis by neutron tomography. From the tomographic volume, virtual slices can be generated, revealing the inside of the bronze as Fig. 13.20 shows. The shape and size of bronze hollows or additional filling materials yield conclusions about the casting process. Resins or varnish used for the conservation of sculptures appear with high contrast.

In several cases, the new examination has shed an entirely new light on sculptures in the Rijksmuseum collection. For instance, it was thought that a work entitled 'Sol' (the Sun) by the sculptor Van der Schardt (1530–1581) dating from 1570–1581 was a garden fountain sculpture, Fig. 13.20. However, the sculpture proved to be lacking any kind of interior mechanism for pumping water upwards.

(a) (b)

FIG. 13.20. Photo (a) and virtual slice (b) through the bronze sculpture 'Sol (The Sun)' from Rijksmuseum Amsterdam, height 45.7 cm, made by J. van der Schardt (around 1585). The metallic structure, the casting process and repair measures were studied in great detail (Paul Scherrer Institut, 2007).

Another examination showed that the 'Striding Nobleman', which had always been attributed to Hendrick de Keyser, could not possibly have been one of his works because the sculpture has an internal structure that De Keyser never used in his bronzes (Rijksmuseum Amsterdam, 2007).

13.4.1.5 *Palaeontology* Neutron radiography allows for a non-destructive evaluation of large fossils. Figure 13.21 shows the photograph and neutron radiograph of the partly dissected head and neck of an ichthyosaur. In the tomographic study of the head section, the skeleton can be segmented from the surrounding sediment. If sufficiently precise details about the skeleton can be retrieved, further restoration steps can be envisaged to yield additional insight into the development of ichthyosaurs.

Here, the neutrons show their advantage to transmit thick layers of material and on the other hand to be sensitive to small variations in its composition. A sediment material with a thickness of several centimeters is a big challenge for conventional X-ray tomography. The small density variations between the stone and the bones of the fossilized skeleton allows for a better distinction between the

(a) (b) (c)

FIG. 13.21. Fossilized skeleton of an ichthyosaur partially dissected by U. Oberli, St. Gallen. Photograph (a), radiograph (b), and tomographic view of the head section (c) (Paul Scherrer Institut, 2007).

two components in the tomographic reconstruction and provides a good contrast for the skeleton part.

13.4.2 *Simultaneous application of neutron and X-ray tomography*

Neutrons and X-rays are complementary to each other in non-destructive material investigations. In principle, the same detection system can be used (Fig. 13.10) when the scintillator screen is sensitive to both kinds of radiation. It depends on the beam properties whether the tomography results are comparable. At the NEUTRA facility a 320-kV X-ray tube has been installed that can be used alternatively to thermal neutrons (Paul Scherrer Institut, 2007). Due to the small focal spot size of 4 mm and the large distance between source and detector plane, similar collimations can be achieved for neutrons and X-rays ($L/D > 500$). Because the detector is completely stationary, a voxelwise comparison of the tomography data can be performed. This is demonstrated for an armoured concrete in Fig. 13.22. Both investigations show a completely different view: steel fibres become visible with X-rays, while the stone matrix is shown in the neutron image. When applied more systematically, the contrast in specific investigations can be enhanced dramatically. This will be pushed additionally by changing the voltage of the X-ray tube in a way that neutron and X-ray attenuation coefficients become similar. In such cases, a subtraction procedure can be applied with high success for contrast improvement.

13.4.3 *Energy-selective tomography*

Most of the neutron tomography facilities use the full neutron spectrum delivered by the source, the so-called 'white beam'. The contrast obtained in tomography then corresponds to attenuation coefficients averaged over the entire spectrum if secondary effects such as beam hardening or scattering artefacts can be neglected. The total cross-section for most elements is reciprocal to the velocity (i.e. energy) of the thermal neutrons. For smaller neutron energies (e.g. cold neutrons)

(a) (b) (c)

FIG. 13.22. Tomography results of a steel-armoured concrete sample (4 cm × 4 cm × 20 cm). (a) Photographic image; (b) X-ray tomography showing the steel–fibre distribution; (c) neutron tomography delivering the stony matrix. Source: Paul Scherrer Institut.

the related cross-section for polycrystalline materials is influenced by coherent scattering processes in the sample. This holds in particular for metallic samples where the polycrystalline structure of the sample plays an important role. In this case, Bragg scattering strongly changes the total cross-sections. The explanation of the presence of Bragg cutoffs in the attenuation spectrum is illustrated in Fig. 13.23. For different wavelengths of the neutron spectrum a corresponding orientation of crystallites with a defined interplanar spacing d_{hkl} can be found that satisfies the Bragg scattering law, eqn 4.6.

In Fig. 13.23 three different orientations of the crystallites are shown as well as the border condition at $\theta = 90°$. At this maximum angle, neutrons with a wavelength λ_{BC} undergo backscattering. For wavelengths larger than λ_{BC} no more Bragg scattering occurs because Bragg's law can no longer be satisfied for these wavelengths. This gives rise to the observed step in the attenuation behaviour, as shown for polycrystalline iron in Fig. 13.23.

The position of the Bragg cutoffs can be related to the corresponding interplanar distances d_{hkl}. If areas of the sample have compressive or tensile residual stresses this will be reflected by a shift of the Bragg-cutoff position. On the other hand, the height of the Bragg cutoff is inversely related to the number of crystallites having a defined orientation, which means that studying the shape

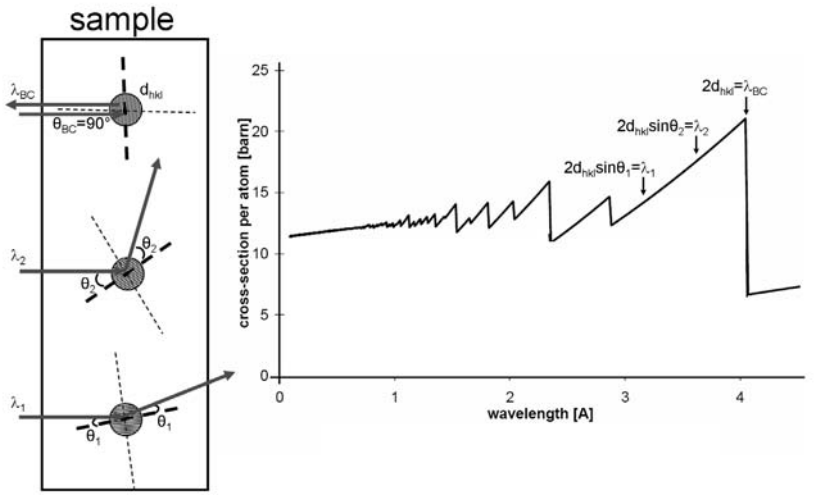


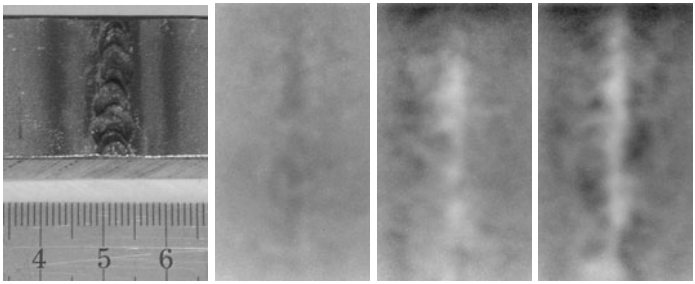
FIG. 13.23. The Bragg scattering process in polycrystalline material is given schematically on the left. An example of the energy-dependent attenuation spectrum for iron is given on the right. The positions of the wavelengths satisfying Bragg's law are marked with arrows.

of the Bragg-cutoff information about the distribution of texture fields in the sample can be extracted. Presently, there is an attempt to exploit such effects in order to study two-dimensional distributions of texture fields. These kind of experiments can be performed only with tunable neutron beams that, however, lead to a reduction of beam intensity and longer measurement times.

There are three ways for practical energy selection:

- Using a turbine-type energy-band selector, where the rotation speed and the tilt angle of the turbine with respect to the beam axis determine the transmitted energy band (Kardjilov *et al.*, 2003).
- By Bragg reflection in single crystals (graphite, silicon, germanium) as in use for diffraction devices. A double-crystal monochromator enables one to maintain the initial direction of the incident beam (Treimer *et al.*, 2006).
- At a pulsed neutron beam – by using the time structure of the provided neutron beam – time-of-flight (TOF) method (Kockelmann *et al.*, 2007).

Recently, such investigations were successfully carried out in the radiography mode. It was possible to narrow the energy band by using the time-of-flight method for the investigation of welds in iron (see Fig. 13.24) and aluminium samples. Although the inspection time was considerably longer than with energy-averaged sampling, the study resulted in a direct visualization of the textures inside the welding seam. Such kinds of investigation will have a strong impact on technological developments and will add a new non-destructive option to material



(a) (b) (c) (d)

FIG. 13.24. Inspection of a weld (a) with different neutron energies, (b) 0.351 nm, (c) 0.389 nm and (d) 0.423 nm. These experimental results were obtained at the OSIRIS beamline of the pulsed spallation source ISIS (UK) in a pilot experiment for energy-selective imaging. Reprinted with permission from Kockelmann *et al.* (2007).

testing. Based on the performance forecast of next-generation neutron sources (pulsed spallation sources in the MW class), it might be possible to extend the present experience towards tomography. Because the texture varies on the length scale of millimetres, the number of projections can be limited to about 100 only.

13.5 Summary

Neutron tomography appears as a complementary technique to the widespread X-ray imaging techniques. In spite of the low flux of neutron sources and the limitation of spatial resolution given by the detection process, neutron tomography is gaining increasing importance in science and engineering. Water management in fuel cells is one of the applications where only imaging with neutrons can provide reliable information. Non-destructive inspection of large parts from automotive and aviation industries is another aspect that makes the method very important for improving the safety standards in these fields. The use of monochromatic neutron beams for imaging further increases the potential of neutron tomography.

13.6 References

Andor Technology, Belfast, UK (2007). CCD cameras. `http://www.andor.com`. [Online; accessed 08 March 2007].

Barton, J.P. (1965*a*). Neutron radiography using a crystal monochromator. *Journal of Scientific Instruments*, **42**, 540.

Barton, J.P. (1965*b*). Radiographic examination using cold neutrons. *British Journal of Applied Physics*, **16**, 1051.

Berger, H., Tylka, J., and Talboy, J. (1964). Determination of cadmium burnup in reactor control rods by neutron radiography. *Nuclear Science and Engineering*, **18**, 236.

Brönnimann, C. (2004). *Applications of the PILATUS single module and 1M detectors*. Paul Scherrer Institut (PSI). Annual Report, SLS-Annex.

Brookhaven National Laboratory (2007). Evaluated nuclear data file (endf). `http://www.nndc.bnl.gov/exfor7/endf00.htm`. [Online; accessed 30 April 2007].

Brookhaven National Laboratory, Brookhaven, USA (2007). Chart of nuclides. `http://www.nndc.bnl.gov/chart`. [Online; accessed 08 March 2007].

Chen-Mayer, H.H., Mildner, D.F.R., Sharov, V.A., Xiao, Q.F., Cheng, Y.T., Lindstrom, R.M., and Paul, R.L. (1997). A polycapillary bending and focusing lens for neutrons. *Review of Scientific Instruments*, **68**, 3744.

Eskildsen, M.R., Gammel, P.L., Isaacs, E.D., Detlefs, C., Mortensen, K., and Bishop, D.J. (1998). Compound refractive optics for the imaging and focusing of low-energy neutrons. *Nature*, **391**, 563.

Hassanein, R. (2006). *Correction methods for the quantitative evaluation of thermal neutron tomography*. Ph. D. thesis, ETH Zürich. Dissertation No. 16809.

Hassanein, R., Meyer, H.O., Carminati, A., Estermann, M., Lehmann, E., and Vontobel, P. (2006). Investigation of water imbibitition in porous stone by thermal neutron radiography. *Journal of Physics D: Applied Physics*, **39**, 4284.

Hilger, A., Kardjilov, N., Strobl, M., Treimer, W., and Banhart, J. (2006). The new cold neutron radiography and tomography instrument CONRAD at HMI berlin. *Physica B*, **385-386**, 1213.

Institute Laue-Langevin (2007). Neutrograph. `www.neutrograph.com`. [Online; accessed 30 May 2007].

Kardjilov, N., Baechler, S., Basturk, S., Dierick, M., Jolie, J., Lehmann, E., Materna, T., Schillinger, B., and Vontobel, P. (2003). New features in cold neutron radiography and tomography - Part II: applied energy-selective neutron radiography and tomography. *Nuclear Instruments and Methods in Physics Research A*, **501**, 536.

Kardjilov, N., Boeni, P., Hilger, A., Strobl, M., and W.Treimer (2005). Characterization of a focusing parabolic guide using neutron radiography method. *Nuclear Instruments and Methods in Physics Research A*, **542**, 248.

Kockelmann, W., Frei, G., Lehmann, E.H., Vontobel, P., and Santisteban, J.R. (2007). Energy-selective neutron transmission imaging at a pulsed source. *Nuclear Instruments and Methods in Physics Research A*, **578**, 421.

Kramer, D., Lehmann, E., Frei, G., Vontobel, P., Wokaun, A., and Scherer, G.G. (2005). An on-line study of fuel cell behavior by thermal neutrons. *Nuclear Instruments and Methods in Physics Research A*, **542**, 52.

Kuhn, E. (1937). US Patent 2186 757. `http://www.uspto.gov/patft/`. [Online; accessed 30 April 2007].

Lehmann, E.H., Frei, G., Kühne, G., and Boillat, P. (2007). The micro-setup for neutron imaging - a major step forward to improve the spatial resolution. *Nuclear Instruments and Methods in Physics Research A*, **576**, 389.

Maier-Leibnitz, H. and Springer, T. (1963). The use of neutron optical devices on beam-hole experiments. *Nuclear Science and Technology*, **17**, 217.

Manke, I., Banhart, J., Haibel, A., Rack, A., Zabler, S., Kardjilov, N., Hilger, A., Melzer, A., and Riesemeier, H. (2007*a*). In situ investigation of the discharge of alkaline Zn-MnO_2 batteries with synchrotron X-ray and neutron tomographies. *Applied Physics Letters*, **90**, 214102.

Manke, I., Hartnig, Ch., Grünerbel, M., Kaczerowski, J., Lehnert, W., Kardjilov, N., Hilger, A., Banhart, J., Treimer, W., and Strobl, M. (2007*b*). Quasi-in-situ neutron tomography on polymer electrolyte membrane fuel cell stacks. *Applied Physics Letters*, **90**, 184101.

Manke, I., Hartnig, Ch., Grünerbel, M., Lehnert, W., Kardjilov, N., Haibel, A., Hilger, A., Banhart, J., and Riesemeier, H. (2007*c*). Investigation of water evolution and transport in fuel cells with high resolution synchrotron X-ray radiography. *Applied Physics Letters*, **90**, 174105.

Masschaele, B., Baechler, S., Cauwels, P., Dierick, M., Jolie, J., and Mondelaers, W. (2001). First results of micro-neutron tomography by use of a focussing neutron lens. *Radiation Physics and Chemistry*, **61**, 623.

Nova Scientific (2007). Neutron imaging. `http://www.novascientific.com/neutron.html`. [Online; accessed 30 April 2007].

Paul Scherrer Institut (2007). Neutron imaging and activation. `http://neutra.web.psi.ch`. [Online; accessed 08 March 2007].

Pleinert, H., Lehmann, E., and Körner, S. (1997). Design of a new CCD-camera neutron radiography detector. *Nuclear Instruments and Methods in Physics Research A*, **399**, 382.

Rijksmuseum Amsterdam (2007). From vulcan's forge. `http://www.rijksmuseum.nl/tentoonstellingen/werkplaats-van-vulcanus?lang=en`. [Online; accessed 30 April 2007].

Schillinger, B. (2001). Estimation and measurement of L/D on a cold and thermal neutron guide. *Nondestructive Testing and Evaluation*, **16**, 141.

Schillinger, B., Bleuel, M., Boeni, P., and Steichele, E. (2005). Design and simulation of neutron optical devices plus flight tube for neutron radiography. In *7th World Conference on Neutron Radiography* (ed. P. Chirco and R. Rosa), pp. 11. ENEA, Rome, Italy.

Shimizu, H.M., Oku, T., Sato, H., Otani, C., Kato, H., Suda, Y., Nakagawa, H., Kamiyami, T., Kiyanagi, Y., and Wakabayashi, T. (2000). A magnetic neutron lens. *Physica B*, **276-278**, 63.

Thewlis, J. (1956). Neutron radiography. *British Journal of Applied Physics*, **7**, 345.

Treimer, W., Strobl, M., Kardjilov, N., Hilger, A., and Manke, I. (2006). Wavelength tunable device for neutron radiography and tomography. *Applied Physics Letters*, **89**, 203504.

Waters, L.S. (2002). *MCNPX User's Manual, Version 2.4.0*. Los Alamos National Laboratory. LA-CP-02-408.

14

NEUTRON PHASE-CONTRAST AND POLARIZED NEUTRON TOMOGRAPHY

Nikolay Kardjilov

The wave–particle dualism (de Broglie, 1923) allows one to treat the neutron as a particle having a defined mass or to regard it as a propagating wave with corresponding amplitude and wavelength. In conventional radiography, image formation is given in terms of attenuation of the radiation by the investigated object. In the case of coherent radiation, transmission of neutrons can be defined in terms of wave propagation through matter. A measure for the relative shift between waves of the same frequency can be introduced by their phase. Different materials or material thicknesses produce different wave shifts relative to the wave propagation in vacuum, which are called *phase variations*. In the case of phase-contrast imaging the phase variations obtained by the propagation of radiation through the sample are transformed to intensity variations detected by a position-sensitive detector. As has been shown in Chapter 6, phase-contrast imaging with synchrotron radiation is a well-established method with a large number of applications. Phase-contrast imaging techniques with neutrons are based on the same principles, but the different nature of the radiation gives rise to some specifics of the method. Neutron sources in general provide flux that is orders of magnitude lower than that at synchrotron facilities, causing very long exposure times, making phase-contrast imaging with neutrons still an exotic technique. The method was pioneered in the early 1980s by Schlenker *et al.* (1980) who investigated ferromagnetic domains in iron by using a Bonse–Hart-type interferometer. The same technique was used for the first phase-contrast tomography experiment with thermal neutrons (Dubus *et al.*, 2005). Phase-contrast imaging with neutrons was reported by implementing a free-path propagation technique where a beam with a high order of spatial coherence was used (Allman *et al.*, 2000). A new technique using spatial gratings was presented recently by Pfeiffer *et al.* (2006*b*) where the low beam intensity in the case of the former methods could be overcome due to the use of a multiple line source defined by a source grating. This development opened up new perspectives for phase-contrast imaging with neutrons.

Taking into account the magnetic interactions between neutrons and magnetic fields or moments in a sample opens up the possibility to image 3D field and distributions. The first attempts towards polarized neutron imaging have been described by Treimer *et al.* 2005*b*. Recent developments have revived the interest in this method, which will therefore be briefly reviewed here.

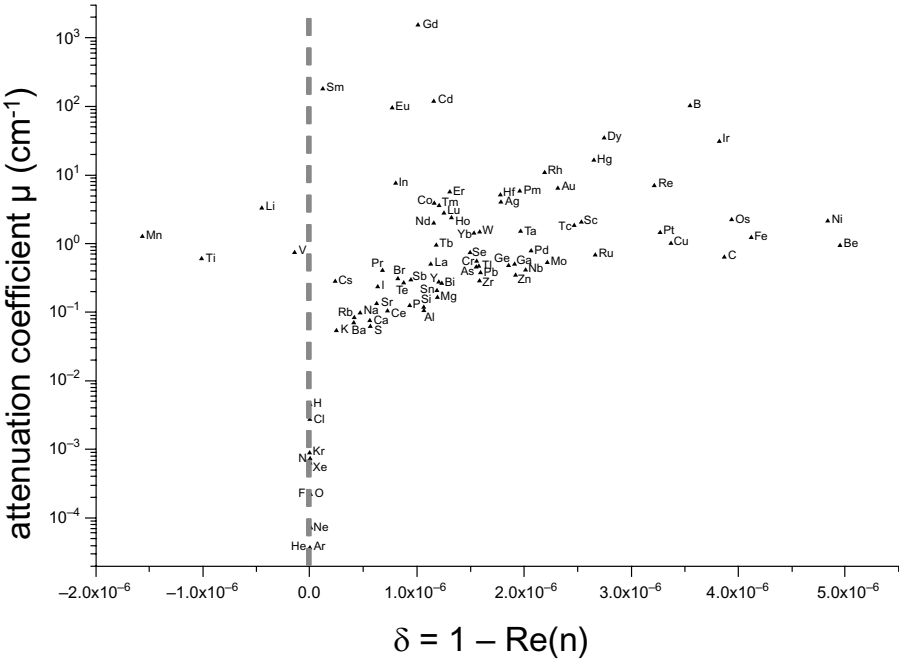


FIG. 14.1. Comparison of the attenuation coefficients and the refraction indices of the elements for thermal neutrons. Elements that lie on a horizontal line cannot be separated in conventional tomographies, but can in phase-contrast tomographies. Taken with permission from Lehmann *et al.* (2005), copyright 2005, Elsevier.

14.1 Theory of phase-contrast imaging

For waves propagating through a medium, a complex refractive index $\hat{n} = 1 - \delta + i\beta$ has been introduced in eqn 4.12, where the real part δ is responsible for the phase shift of the propagating wave and β represents the absorption in the medium. In that terminology one can say that phase-contrast imaging exploits the real part of the refractive index $1 - \delta$, while conventional radiography makes use of the imaginary part β. In Fig. 14.1 both the imaginary and the real part of the refractive index for thermal neutrons are plotted for each element. It can be noticed that the sensitivity of the phase-contrast method to distinguish between different elements is much higher than that of conventional imaging based on absorption contrast in some cases.

As can be seen from Fig. 14.1, the distribution of the element's refraction index is a pre-condition for the observation of the phase-contrast effect by using thermal neutrons. In order to evaluate the importance of neutron phase-contrast imaging, it is interesting to make a comparison to the corresponding X-ray method by using evaluation parameters. Phase-contrast techniques with

TABLE 14.1. Listing of relevant performance parameters of neutron and X-ray phase tomography for selected elements. All values have been calculated for a wavelength of 0.18 nm (Dubus *et al.*, 2001).

element	Z	b_{coh} [fm]	X-ray vs. neutron scattering (R_{ss})	phase vs. attenuation X-rays (X)	phase vs. attenuation neutrons (N)
H	1	−3.74	0.75	-	78.9
Mg	12	5.38	6.3	54.8	2564
Al	13	3.45	10.6	44.3	3582
Si	14	4.15	9.5	35.3	3194
Mn	25	−3.73	18.9	79.7	435
Pb	82	9.40	24.6	82.5	1499

X-rays and thermal neutrons are very similar, but the different way of interaction with matter has an impact on performance. In both cases, the interaction is very weak over the wavelength range considered here (except very close to a resonance). Under this consideration the refractive index for X-rays is defined by eqn 4.15, see also Dubus *et al.* (2001):

$$n = 1 - \lambda^2 N_V Z \frac{r_e}{2\pi},$$

where r_e is the radius of the electron. Z is the atomic number and N_V is the number of atoms per unit volume. For thermal neutrons the refractive index can be written as (eqn 4.18):

$$n = 1 - \lambda^2 N_V \frac{b_{\mathrm{coh}}}{2\pi}.$$

b_{coh} is the coherent scattering length for the corresponding atom.

The scattering-strength ratio for X-rays (X) and neutrons (n) is defined as the ratio (Dubus *et al.*, 2001):

$$R_{ss} = (1-n)_{\mathrm{X}} / |1-n|_{\mathrm{n}} = \frac{Z r_e}{b_{\mathrm{coh}}}.$$

Some values of R_{ss} are listed in Tab. 14.1 for a selection of materials. In general, scattering of X-rays is one order of magnitude larger than that of neutrons at the same wavelength (except for hydrogen). In addition, it is of interest to compare phase and absorption contrast with each other for X-rays and neutrons by using the ratio of the real and imaginary part of the refractive index δ/β. For X-rays the ratio (X) is given by Beckmann *et al.* (1997).

$$X = \lambda \frac{2.35 Z r_e}{\sigma_{\mathrm{tot}}},$$

where σ_{tot} is the total scattering cross-section for X-rays. For neutrons an estimate of phase over attenuation contrast (N) can be obtained by combining eqn 4.5, eqn 4.13 and eqn 4.18 by Rauch (1996):

$$N = \lambda \frac{2b_{\mathrm{coh}}}{\sigma_{\mathrm{abs}} + \sigma_{\mathrm{incoh}}}.$$

Values for X and N for different materials are also listed in Tab. 14.1. For the same wavelength and the same thickness of elements given in Tab. 14.1 the losses of intensity due to attenuation are much higher for X-rays (lower values for X in comparison with N), which is an advantage of using neutrons for samples with a reasonable thickness. In addition, neutron phase-contrast tomography is most advantageous with specimens that show low neutron attenuation and at the same time are strong absorbers for X-rays, e.g. for samples containing Pb or Bi.

14.2 Experimental techniques for phase-contrast imaging

The different techniques utilize experimental methods for conversion of the phase shift to intensity variations that are recorded by a position-sensitive detector. The main techniques described in this chapter are: interferometeric method, free-path propagation and differential phase-contrast imaging.

14.2.1 *Interferometric technique*

This technique utilizes a Bonse–Hart interferometer for phase-shift measurements as described in Chapter 6. A principle sketch of the method is given in Fig. 14.2. The projections required for tomographic reconstruction are measured in the plane xz at the detector position. The number of projections can be set by rotating the specimen around the z-axis. In order to eliminate all patterns inherent in the interferometer setup itself, two measurements per projection should be performed: *with the sample in the beam* $I_{\mathrm{sample}}(j)$ and *without the sample in the beam*, also called the 'reference' projection $I_{\mathrm{ref}}(j)$. The index j appears for different values of φ_j, which is the relative phase difference of the two interferring beams set by a phase-shifter plate PS. For extracting the specimen's projected phase distribution $\phi(x, z)$, projections with $I_{\mathrm{sample}}(j)$ and without the sample in the beam $I_{\mathrm{ref}}(j)$ should be measured with at least three different values of φ_j. This helps to find the positions of the maxima and minima of the patterns for the two measurements. From this the shift between the two patterns can be calculated, which gives the phase shift induced by the sample. The technique is known as the *phase-shifting method*, compare with Section 6.2.1. This means per projection a set of six images should be taken where the following assumption can be made:

$$\varphi_j = \frac{2\pi j}{3},$$

with $j = 0, 1, 2$, which gives overall phases of 0°, 120°, 240°.

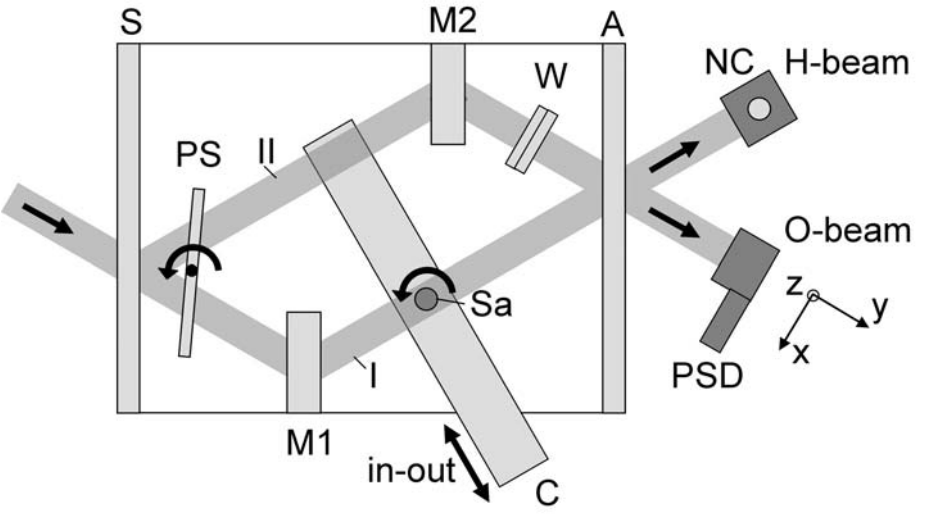


FIG. 14.2. Principal sketch of neutron interferometer for phase measurements. S: beam splitter; M1 and M2: mirrors; A: analyser; I and II: interferring beams and O, H: output beams. Sa is the sample inside a measurement cell that consists of an aluminium plate with a hole for the sample. PSD: position-sensitive detector for neutrons; PS: phase-shifter plate; NC: gas counter for neutrons; W: wedge for compensating the built-in phase pattern of the interferometer. Measuring cell C including sample can be translated to remove the sample from beam I for recording reference projections or to place the sample in the other interferring beam II. The overall phase of I and II is changed by rotating the phase-shifter plate PS (Dubus *et al.*, 2005).

For the full exploitation of phase-contrast tomographic data a more sophisticated maximum likelihood algorithm has to be developed, which enables phase estimations with less count numbers and helps to reduce the measurement time (Zawisky *et al.*, 1998).

For each projection the local phase is an integral of the phase-shift density along the beam direction y:

$$\phi(x,z) = \int \phi(x,y,z)dy.$$

$\phi(x,y,z)$ is the 3D phase-shift structure of the sample to be determined from the phase projections in the tomographic reconstruction process. To demonstrate the efficiency of the interferometric method, weakly absorbing substances such as aluminium with slightly varied elemental composition can be investigated (Zawisky *et al.*, 2004). The specimen, Fig. 14.3, consists of several aluminium parts, including an aluminium rod with a coaxial aluminium screw in the object beam. The screw has a slightly different chemical composition from the surrounding

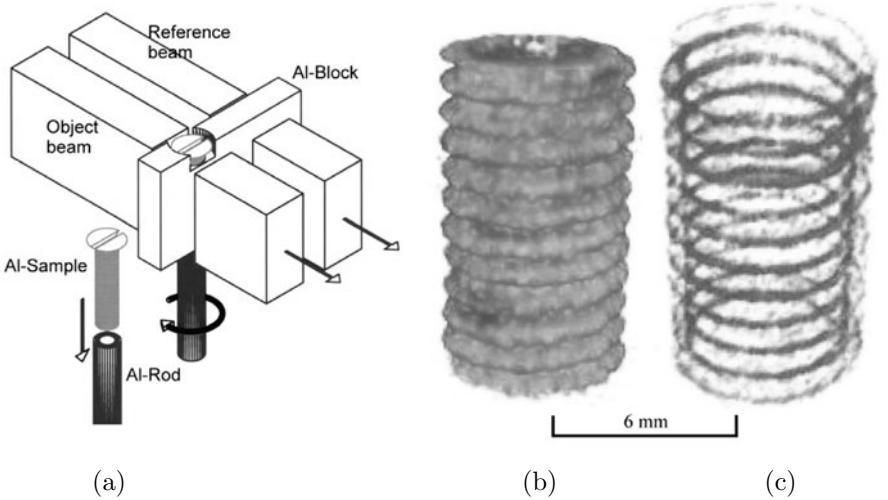


FIG. 14.3. (a) Sample arrangement. (b) Tomographic reconstruction of the aluminium screw from phase shifts in the sample. (c) The 100-μm air gap in the thread creates a strong phase difference relative to the aluminium block in the reference beam. Taken with permission from Dubus *et al.* (2005), copyright 2005, IEEE.

aluminium rod and block. The goal of the experiment was the visualization and quantitative analysis of the embedded screw with its 2% lower scattering density. Results from tomographic investigation are shown in Fig. 14.3.

Advantages of the interferometric method are the direct measurement and 3D reconstruction of the phase-shift structure in the sample. Disadvantages are the complicated setup and the low SNR caused by low neutron flux. This is reflected by long exposure times that require extreme thermal, mechanical and acoustic stability of the device.

14.2.2 *Free-path propagation technique*

The propagation technique for phase-contrast imaging with neutrons is mainly based on the transport of intensity approach described in Section 6.6. The transport of intensity equation (eqn 6.1) that describes the effect that determines the transformation of phase inhomogeneities to intensity variations. It relates the intensity distribution along the beam axis to intensity and phase in a plane transverse to the propagating direction (z-axis). Thus, variations in thickness and refractive index of a sample lead to different phase variations and consequently to a change of intensity with wave propagation. The connection between phase and wave intensity can be explained more intuitively as a deformation of the shape of a defined wavefront on passing through the sample due to the phase gradients obtained. This wavefront deformation will cause a redistribution of the wave intensity at some distance behind the sample. As shown in Fig. 14.4

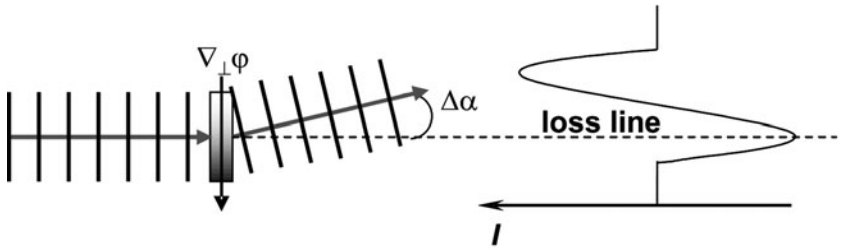


FIG. 14.4. An example of the deflection of a coherent wave out of its initial propagation direction by a transverse gradient of the refractive index.

the local propagation vector $\mathbf{s}(x, y, z)$ depends on the gradient of the phase φ perpendicular to the propagation direction and can be written as

$$\mathbf{s}(x, y, z) \approx \left(-\frac{\partial \varphi}{\partial x}, -\frac{\partial \varphi}{\partial y}, k \right),$$

so that $\mathbf{s}(x, y, z)$ is normal to the wavefront at the point (x, y, z). The angular deviation of the normal to the wavefront, $\triangle\alpha$, can be expressed as:

$$\triangle\alpha \approx \frac{1}{k} |\nabla_{\perp} \varphi(x, y, z)|,$$

and thus depends on the variation of the transverse component of the refractive index.

This means that areas that produce a strong phase gradient (the edges in the sample) will deform the wavefront in a way that the rays will be deflected from their original propagation direction and a loss of intensity will be detected in the forward direction (loss line), Fig. 14.4. This diffraction process will cause a considerable edge enhancement in the image obtained (Allman *et al.*, 2000, Kardjilov *et al.*, 2004). In order to be able to observe this effect, a beam with high spatial coherence has to be used due to the requirement of a plane wavefront.

The measure for the achieved level of spatial coherence is the transverse coherence length l_t (Pedrotti and Pedrotti, 1962). This represents a distance in a plane transverse to the main propagation direction over which the phases at two points remain correlated. The expression for the transverse coherence length, l_t, is given as:

$$l_t < \frac{r\lambda}{s},$$

where r is the distance between the source and the sample, s is the source extension and λ is the wavelength. In order to obtain a wavefield with a high transverse coherence length at the sample position, the source size s has to be small and/or the distance between the source and the sample r has to be large. In case of high transversal spatial coherence all rays emanating from the point source, defined by a small pinhole, are associated with a single set of spherical

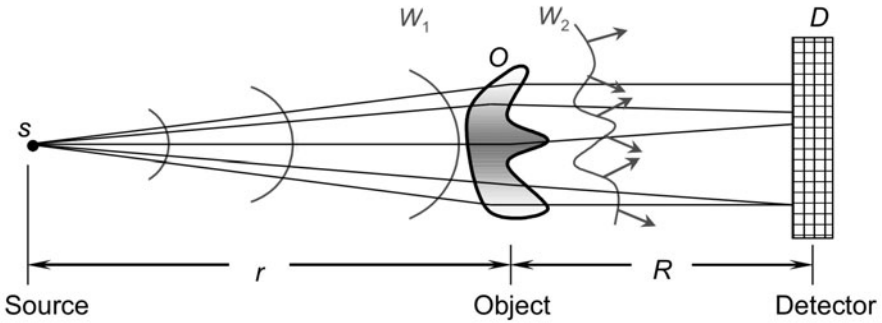


FIG. 14.5. A model of a phase-contrast configuration using a pinhole geometry.

waves that have the same phase in a surrounding smaller than the coherence length on any given wavefront. A simple representation of a configuration for phase-contrast imaging exploiting a pinhole source to produce a high spatial coherence is shown in Fig. 14.5. The distances r and R are the source-to-object and object-to-image distances, respectively. It can also be seen how the spherical wavefront W_1 emanating from the point source s becomes distorted to W_2 on passing through the object.

More generally, any rapid variation in the refractive index or thickness of a sample may be imaged as sharp loss (variations in intensity) at corresponding points in the image even when a polychromatic source is used (Jacobson *et al.*, 2004, Kardjilov *et al.*, 2004), see Fig. 14.5.

The advantages of the propagation technique compared to the interferometric technique are the extreme simplicity of the setup, similar to the one for absorption tomography, the better spatial resolution and the higher neutron flux at the sample position. A drawback of the propagation technique is the non-trivial relationship between the optical phase and the recorded image. Two approaches for phase retrieval have been considered recently: a 'direct' approach or a more adapted, quantitative one. The direct method consists of applying the same algorithms as the ones developed for absorption tomography. This gives good and workable results in the so-called 'edge-detection regime' and allows us to image directly the sharp variations of refractive index in the object. The reconstructed quantity is in this case approximately proportional to the three-dimensional Laplacian of the distribution of the refractive index.

An example of this is shown in Fig. 14.6 where a comparison between conventional and phase-tomography sections on a human tooth immersed in heavy water is shown. Evidently, phase-contrast tomography provides much more detail about the structure of the tooth. Even small cracks in the enamel and bubbles in D_2O can be resolved.

A quantitative approach involves numerical retrieval of the phase from the radiographs. It has already been tested for X-rays (Cloetens *et al.*, 2001), but

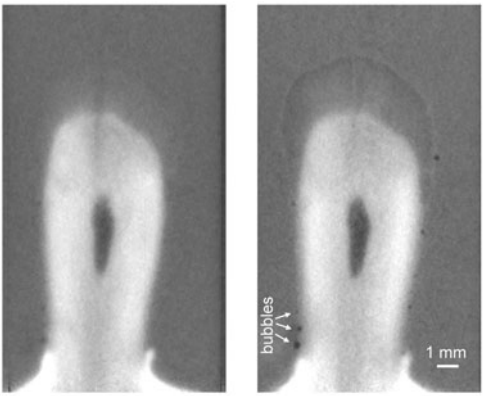


FIG. 14.6. Comparison between conventional absorption (left) and phase-contrast (right) tomography section of a human tooth. For the phase-contrast experiment a pinhole of 5 mm diameter and a distance of 5 m between pinhole and sample was used. The number of projections was 200 for both experiments (Kardjilov *et al.*, 2006).

data about its application in neutron phase-contrast tomography are still lacking.

14.2.3 *Differential phase contrast*

This method combines the efficiency of the free propagating technique and the direct phase quantification typical for the interferometric technique. It is based on the utilization of spatial gratings (Weitkamp *et al.*, 2005, Pfeiffer *et al.*, 2006*b*): a source grating G0, a phase grating G1, and an analyser absorption grating G2, see Fig. 14.7 and Section 6.3.

The source mask G0 has transmitting slits in order to improve the spatial coherence of the beam by creating an array of vertical line sources. In order to have constructive interference of each of the line sources in the detector plane, the following condition should be fulfilled: $p_0 = p_2 l/d$, where $p_0(p_2)$ is the period of G0(G2), l is the distance between G0 and G1, and d is the distance between G1 and G2. Behind the beamsplitter grating G1, the diffracted beams interfere and form linear periodic fringe patterns in planes perpendicular to the optical axis. Introducing a sample in the beam, the incident wavefront is distorted, leading to local displacements of the fringes. The grating interferometer exploits the idea of detecting the local positions of the fringes and determining from these the phase shift induced by the object. However, the available detector resolution does not allow us to resolve fine fringe structure of a few micrometres spacing. Therefore, an absorbing grating G2 with the same periodicity and orientation as the fringes is positioned in the detection plane. It acts as an analyser grating transforming the local fringe position into a signal intensity variation. In this way, the detected signal profile can be related to the phase gradient of the object. In order to

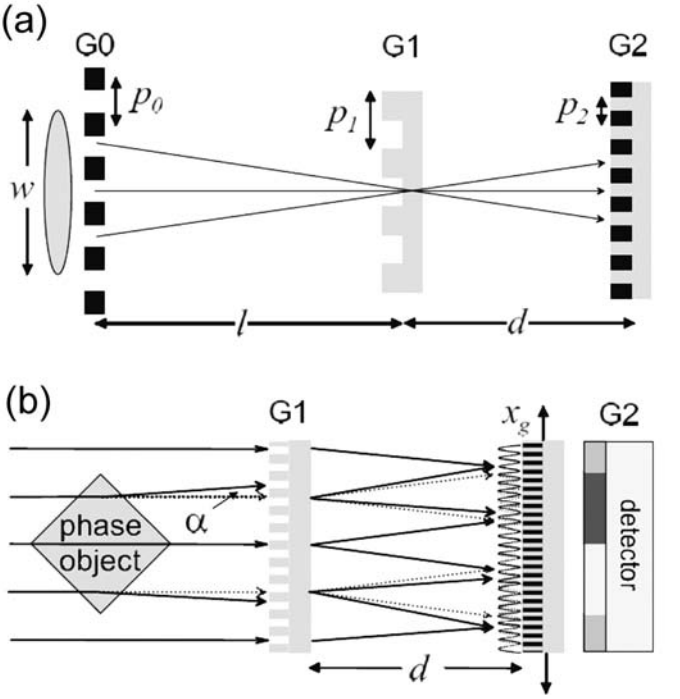


FIG. 14.7. Grating-based neutron shearing interferometer. (a) The source grating (G0) creates an array of individually coherent, but mutually incoherent sources. (b) The phase grating (G1) forms a periodic interference pattern in the plane of the analyser grating. A phase object in the incident beam will cause a slight refraction, which results in changes of the locally transmitted intensity through the analyser. Reprinted with permission from Pfeiffer *et al.* (2006*a*), copyright 2006 by the American Physical Society.

extract the phase information separately from other contributions to the signal, such as attenuation in the sample, inhomogeneous illumination or imperfections of the gratings, the phase-shifting approach used in the interferometric method was adapted to this. Then, one of the gratings is scanned along the transverse direction x_g (see Fig. 14.7) and the intensity signal $I(x, y)$ of each pixel (x, y) in the detector plane oscillates as a function of x_g. The interferometer phases $\theta(x, y)$ of the intensity oscillations in each pixel are related to the wavefront phase profile $\theta(x, y)$, the neutron wavelength λ and the period p_2 of the absorbing grating by:

$$\theta = \frac{\lambda d}{p_2} \frac{\partial \Phi}{\partial x}.$$

Due to the fact that the phase signal is independent of the attenuation properties of the sample, the phase profile of the object can be retrieved from $\theta(x, y)$ by a

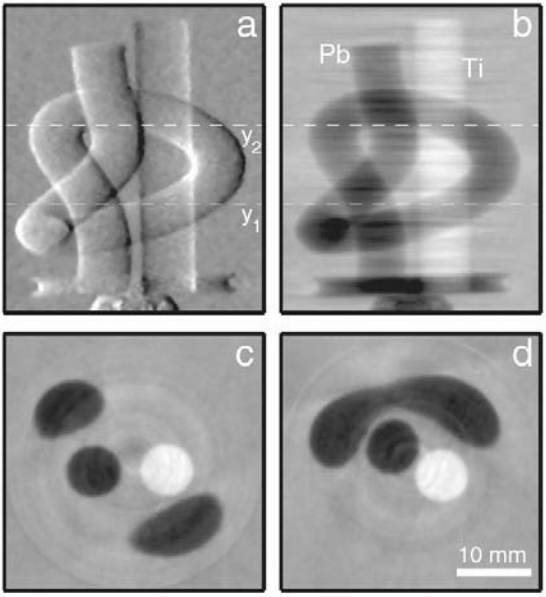


FIG. 14.8. Projection images and tomographic slices for a sample comprising a Pb and Ti rod with a diameter of 6.35 mm. (a) Differential phase contrast, and (b) phase-shift projection image. (c), (d) Two (out of 512) reconstructed tomographic slices through the three-dimensional coherent scattering length distribution in the sample. The total exposure time was 240 min. Reprinted with permission from Pfeiffer *et al.* (2006*a*), copyright 2006 by the American Physical Society.

simple one-dimensional integration. The average response in each pixel over an entire oscillation at the phase-stepping scan gives the attenuation signal as in the case of conventional radiography. A single phase-stepping scan thus yields both the phase and the attenuation image.

As an example, Fig. 14.8 shows processed DPC (a), integrated phase (b), and two (out of 512) reconstructed tomographic slices (c) and (d) of a sample comprising a Ti and Pb rod of 6.35 mm diameter. 180 phase projections and one reference projection without the sample, each of them containing ten phase-step images (of 10 s exposure time) were used as an input for a standard filtered backprojection algorithm.

14.3 Imaging with polarized neutrons

A completely different approach for tomographic imaging is taken in polarized neutron tomography. Neutrons are sensitive to magnetic fields due to their magnetic moment caused by their spin. Therefore, besides the conventional attenuation contrast image of a sample, the magnetic field inside and around the sample can be visualized independently by detection of the polarization changes in the

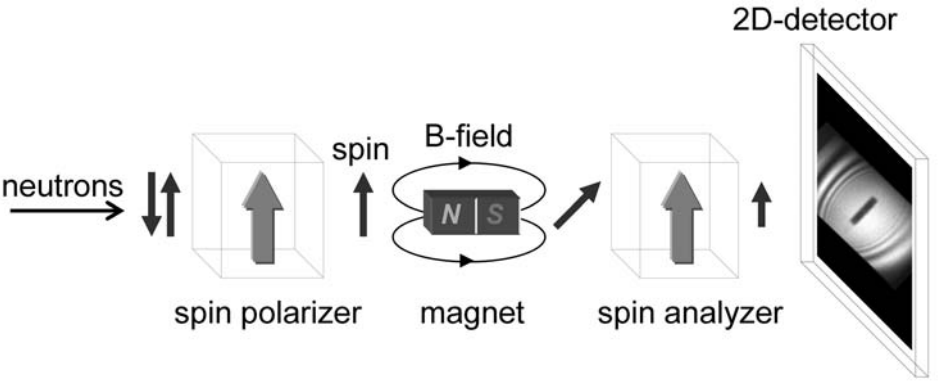


FIG. 14.9. Spin-polarized neutron imaging: The neutron spin rotates in the magnetic field of a sample and hence approaches the spin analyser typically in a non-parallel orientation. The angle of the final spin rotation depends on the magnetic-field integral along the beam path.

transmitted beam. Polarized neutron radiography is based on the spatially resolved measurement of the final precession angles of a collimated and polarized monochromatic neutron beam that transmits a magnetic field, which is present inside and outside of a sample. Imaging experiments using polarized neutrons were carried out at the new neutron tomography facility CONRAD at HMI. For this purpose, the instrument was equipped with solid-state polarising benders, Fig. 14.9. A double-crystal monochromator device was used to pick a defined wavelength from the cold-neutron spectrum. The achieved spatial resolution in the radiography images was around 500 µm for the given experimental geometry.

The potential of the method was demonstrated by the visualization of a magnetic field around two permanent dipole magnets located at a distance of 2 cm between each other, Fig. 14.10.

The obtained result can be explained by theoretical considerations related to the interaction of the neutron magnetic moment with the magnetic field. In the presence of a magnetic field $\vec{B}$ the magnetic moment (i.e. the spin $\vec{S}$) of the neutron will undergo a Larmor precession with a frequency ω_L. For a monochromatic neutron beam with a uniform neutron velocity v the precession angle φ is proportional to the magnetic field B along a certain path:

$$\varphi = \omega_L t = \frac{\gamma}{v} \int_{\text{path}} B ds. \tag{14.1}$$

φ can be measured experimentally by using a spin polarizer–analyser arrangement as shown in Fig. 14.9. Here the neutrons are first polarized and after the interaction with the magnetic field the final spin orientation is analysed with respect to the initial state. The signal transmitted through the spin analyser

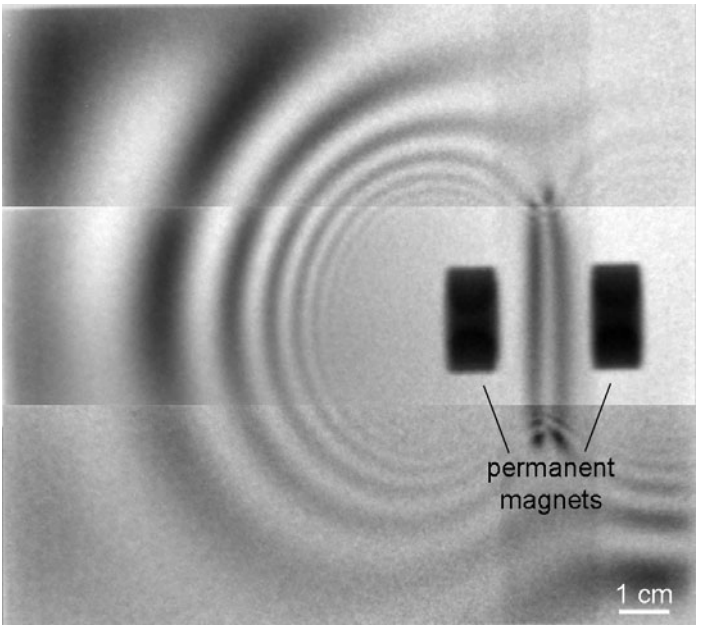


FIG. 14.10. Image of two dipole permanent magnets separated by an aluminium bar with polarized neutrons. The spin of the neutrons is rotated due to the magnetic field between the magnets that is perpendicular to the original spin orientation defined by the polarizer. The resulting spin orientation is converted by the transmission through the spin analyser into grey levels between white for parallel and black for antiparallel spin orientation with respect to the analyser.

depends on the spin rotation angle, see eqn 14.1, and has a maximum for parallel and a minimum for antiparallel spin orientation. The presented example in Fig. 14.10 illustrates the decay of the magnetic field strength with increasing distance from the magnets, resulting in an annular structure around the sample with an increasing period due to the changing precession angles of the neutron spin on their path through the strongly decaying field. The structure in the central part cannot be seen due to the spatial resolution of approximately 500 µm for the spin-polarized setup at the neutron radiography facility at the Hahn-Meitner Institute in Berlin. The obtained contrast in images with polarized neutrons can be extended in the future to three-dimensional visualization of magnetic fields in and around bulky samples by using a dedicated tomographic reconstruction algorithm (Treimer *et al.*, 2005*a*, Jericha *et al.*, 2007).

14.4 Conclusions and outlook

Phase-contrast tomography helps to investigate material combinations with low or similar absorption coefficients that prevents imaging using conventional absorption tomography. New experimental phase-contrast methods using thermal or cold neutrons were developed recently. They allow for the reconstruction of the three-dimensional distribution of the refraction index in a sample.

Phase-contrast tomography with neutrons is a new feature that has shown its high potential through the methods presented in the present chapter. The results presented are mainly test measurements performed to demonstrate the feasibility of the applied techniques. The future of the methods depends on the optimization of the experimental conditions in order to gain more intensity and detector resolution. Higher neutron flux will reduce the acquisition time and allow for achieving better signal-to-noise ratios of the projection images. A step forward is the development of neutron optical components (focusing guides) that will help to focus the beam in a small focal point. This will be especially advantageous for the free-propagating technique where a lot of intensity is lost due to the pinhole diaphragm used for better spatial coherence of the beam. Examples of tests of focusing neutron optical devices with different parameters can be found elsewhere (Kardjilov *et al.*, 2005). The detector systems are subject of development as well, showing realistic potential for considerable improvement of the spatial resolution. Different prototype systems were presented recently showing resolution values of down to 25 μm, (see Chapter 13). This will help to resolve the phase-contrast signal much better and will increase the sensitivity in the case of the differential phase-contrast method.

The use of monochromatic and polarized neutron beams for imaging opens the way to a further type of 3D imaging: that of magnetic fields in empty space and in materials. Besides the practical use, this method might have a big potential for fundamental science where, e.g. problems related to the principles of superconductivity can be addressed.

14.5 References

Allman, B.E., McMahon, P.J., Nugent, K.A., Paganin, D., Jacobson, D.L., Arif, M., and Werner, S.A. (2000). Imaging - phase radiography with neutrons. *Nature*, **408**, 158.

Beckmann, F., Bonse, U., Busch, F., and Günnewig, O. (1997). X-ray microtomography (μCT) using phase contrast for the investigation of organic matter. *Journal of Computer Assisted Tomography*, **21**, 359.

Cloetens, P., Ludwig, W., Boller, E., Helfen, L., Salvo, L., Mache, R., and Schlenker, M. (2001). Quantitative phase contrast tomography using coherent synchrotron radiation. In *Developments in X-ray tomography III* (ed. U. Bonse), pp. 82. The International Society for Optical Engineering (SPIE). Proceedings of SPIE, Vol. 4503.

de Broglie, L. (1923). *Comptes rendus de l'Acadmie des Sciences*, **177**, 507.

Dubus, F., Bonse, U., Biermann, T., Baron, M., Beckmann, F., and Zawisky, M. (2001). Tomography using monochromatic thermal neutrons with attenuation and phase contrast. In *Developments in X-ray tomography III* (ed. U. Bonse), pp. 359. The International Society for Optical Engineering (SPIE). Proceedings of SPIE, Vol. 4503.

Dubus, F., Bonse, U., Zawisky, M., Baron, M., and Loidl, R. (2005). First phase-contrast tomography with thermal neutrons. *IEEE Transactions on Nuclear Science*, **52**, 364.

Jacobson, D.L., Allman, B.E., McMahon, P.J., Nugent, K.A., Paganin, D., Arif, M., and Werner, S.A. (2004). Thermal and cold neutron phase-contrast radiography. *Applied Radiation and Isotopes*, **61**, 547.

Jericha, E., Szeywerth, R., Leeb, H., and Badurek, G. (2007). Reconstruction techniques for tensorial neutron tomography. *Physica B*, **397**, 159.

Kardjilov, N., Hilger, A., Manke, I., Strobl, M., Treimer, W., and Banhart, J. (2006). Multifunctional tomography instrument with cold neutrons at HMI. In *Proceedings of World conference on neutron radiography (WCNR)08, USA*. in press.

Kardjilov, N., Lee, S.W., Lehmann, E., Lim, I.C., Sim, C.M., and Vontobel, P. (2005). Improving the image contrast and resolution in the phase-contrast neutron radiography. *Nuclear Instruments and Methods in Physics Research A*, **542**, 100.

Kardjilov, N., Lehmann, E., Steichele, E., and Vontobel, P. (2004). Phase-contrast radiography with a polychromatic neutron beam. *Nuclear Instruments and Methods in Physics Research A*, **524**, 519.

Lehmann, E., Lorenz, K., Steichele, E., and Vontobel, P. (2005). Non-destructive testing with neutron phase contrast imaging. *Nuclear Instruments and Methods in Physics Research Section A*, **542**, 95.

Pedrotti, F. L. and Pedrotti, L. S. (1962). *Introduction to optics*. Prentice-Hall, Englewood Cliffs.

Pfeiffer, F., Grünzweig, C., Bunk, O., Frei, G., Lehmann, E., and David, C. (2006*a*). Neutron phase imaging and tomography. *Physical Review Letters*, **96**, 215505.

Pfeiffer, F., Weitkamp, T., Bunk, O., and David, C. (2006*b*). X-ray phase imaging with a grating interferometer. *Nature Physics*, **2**, 258.

Rauch, H. (1996). Neutron phase topography for high sensitive material testing. In *5th World Conference on Neutron Radiography* (ed. C. Fischer).

Schlenker, M., Bauspiess, W., Graeff, W., Bonse, U., and Rauch, H. (1980). Imaging of feromagnitic domains by neutron interferometry. *Journal of Magnetism and Magnetic Materials*, **15–18**, 1507.

Treimer, W., Hilger, A., Kardjilov, N., and Strobl, M. (2005*a*). Review about old and new imaging signals for neutron computerized tomography. *Nuclear Instruments and Methods in Physics Research A*, **542**, 367.

Treimer, W., Kardjilov, N., Feye-Treimer, U., Hilger, A., Manke, I., and Strobl, M. (2005*b*). Absorption and phase-based imaging signals for neutron tomo-

graphy. In *Advances in Solid State Physics* (ed. B. Kramer), Volume 45, pp. 407. Springer, Berlin.

Weitkamp, T., Diaz, A., David, C., Pfeiffer, F., Stampanoni, M., Cloetens, P., and Ziegler, E. (2005). X-ray phase imaging with a grating interferometer. *Optics Express*, **13**, 6296.

Zawisky, M., Bonse, U., Dubus, F., Hradil, Z., and Rehacek, J. (2004). Neutron phase contrast tomography on isotope mixtures. *Europhysics Letters*, **68**, 337.

Zawisky, M., Hasegawa, Y., Rauch, H., Hradil, Z., Myska, R., and Perina, J. (1998). Phase estimation in interferometry. *Journal of Physics A: Mathematical and Theoretical*, **31**, 551.

15

NEUTRON-REFRACTION AND SMALL-ANGLE SCATTERING TOMOGRAPHY

Wolfgang Treimer

Refraction and small-angle scattering can provide unique imaging signals in situations where absorption- or other attenuation-based effects are insufficient. Attenuation contrast is low in cases where either the absolute values of the attenuation coefficients of the individual components of the object are low, or the differences between the various components are small. In such cases, other contrast mechanisms have to be employed for imaging such as the phase contrast described in the previous chapter. In this chapter, two further contrast mechanisms will be presented and their usefulness demonstrated with some selected application examples.

An important quantity to decribe the interaction of neutrons with media is the complex index of refraction $\hat{n}$. Equation 4.18 expresses the real part, n, of $\hat{n}$ by the coherent scattering length b_{coh}, while the imaginary part is related to incoherent absorptive (σ_{abs}) and scattering (σ_{incoh}) processes. For neutrons, the real part of n is responsible for the degree of refraction and small-angle scattering and is independent of the attenuation coefficient due to neutron–nucleus interactions. This is not the case for X-rays, where both attenuation and scattering depend on the number of shell electrons and therefore on the atomic number Z. The interactions of neutrons based on refraction and small-angle scattering are coherent, keeping in mind that this feature requires their description as quantum waves, as given by de Broglie's relation eqn 4.1.

Any ray consisting of neutrons, photons, electrons, or other particles that passes from one medium into another experiences interactions that can be divided into two fundamental categories: the first one includes particle-based interactions, the other one includes all interactions that are based on the wave nature of the particle. Particle-based interactions can be summarized as absorption and excitation, the (matter-) wave-based ones as scattering, depolarization (change of the spin state for electrons and neutrons) and the change of the phase of the wave. Thus, the treatment of neutrons having wave character offers the use of (nearly) all interactions that are based on light optical considerations, summarized under scattering effects. To these belong refraction and small-angle scattering.

Refraction is best known from light optics. It describes the change of the propagation direction, if light crosses e.g. from air into water. This change is caused by different phase changes of different parts of the corresponding wave.

Essential for this change is some coherence of the matter-wave, which for neutrons is a typical quantum-mechanical phenomenon. Coherence means that two (or more) points of a wave emerging from a point source have a (constant) phase relationship, or that there is a function that describes the evolution of phase in space or time. All points having the same phase form a *phase field* or a *phase front* for zero phase angle.

To distinguish different structures in an object certain differences of intensities I must be registered, known as contrast C. The contrast (in general) of a part in an image is defined as the ratio of

$$C = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}},$$

where the particular intensities $I_{\max}$ and $I_{\min}$ are determined in adjacent pixels of a 2D detector or film. The contrast C is determined also by the kind of interaction, i.e. absorption, elastic and inelastic, coherent and incoherent scattering, spin interaction, excitation. Attenuation involves both absorption and scattering. Therefore, structured attenuation of a beam by an inhomogeneous sample means absorption and/or phase (refraction or small-angle) contrast. One can show that in spite of small absorption (contrast <5%) structures even though can be visualized by e.g. phase contrast. In the case of absorption, one has an entirely incoherent interaction of radiation with matter, in the case of phase-contrast coherent interaction of the incident and transmitted beam has to be taken into account. Therefore, it is very useful to differentiate contrast into an incoherent and a coherent part. A ray that traverses an object experiences often more than one kind of interaction, so to understand the resulting imaging signal one has to take all (possible) interactions into account and calculate their contribution to an image (to image contrast).

15.1 Refraction tomography

The coherent part n of the complex index of refraction $\hat{n}$ is responsible for the wave nature of the neutron and therefore also for its phase behaviour. In this chapter a different approach is presented to use the phase as the imaging signal for radiography and tomography. Due to interactions with matter the coherent partial waves form a wavefront that changes its direction for $z > d^2/\lambda$ (z = distance object–detector and d = diameter of the coherent illuminated object) (Wu and Liu, 2003). If the interface between two media in a sample is not perpendicular to the flight direction of the neutron, refraction occurs. This is usually the case for tomography since during data recording the incident flight direction changes from 0 to 180° as the sample is rotated. The change of direction of a ray, see Fig. 15.1, due to the extremely small interaction potential of a nucleus with neutrons is of the order of 1×10^{-6} rad = 1 µrad (for glancing angles) and therefore not observed in conventional tomography.

Considering this interaction more closely, one has to distinguish between two cases: the one where the incident coherent wavefront is larger, and the other,

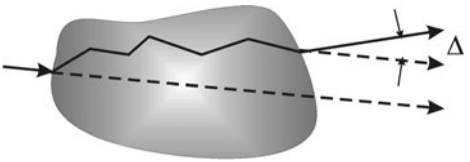


FIG. 15.1. Multiple refraction of a neutron wave. The resulting refraction angle Δis exaggerated.

where it is smaller than the interface of the object. In the case of refraction, the coherent wavefront is much smaller then the interface. A neutron wave is then deflected many times (zigzag path) by the structures in the sample and leaves the sample with a different flight direction compared to the initial one. The index of refraction n is then a function of two spatial variables, x and y and one writes

$$n(x,y) = 1 - \frac{\lambda^2 N(x,y) b_c(x,y)}{2\pi} = 1 - \delta(x,y),$$

where $N(x,y)$ is the position-dependent material density and b_{coh} = the (mean) coherent scattering length. Refraction along a certain path can be described by a position-dependent phase-shift function[13] $\triangle\varphi(x,y), \delta = \delta(x,y)$ as (Treimer *et al.*, 2003)

$$\triangle\varphi(x,y) = -k \int_{\text{path}} \delta(x,y) ds.$$

The entities one has to determine are the refraction angles $\Theta = \Theta(\theta, u)$, where θ is the orientation of the object related to the incident beam. One measures only notable Θ if the interface is nearly parallel to the incident beam, so Θ can be approximated by

$$\Theta(\theta, u) = \frac{1}{k} \nabla_{\perp} (\triangle\varphi(x,y)).$$

A point of a projection $P_\theta(u)$ can then be written as

$$\begin{aligned} P_\theta(u) &= \int_0^u \Theta\left(x, y(x,\theta,u')\right) du' \\ &= \int_{\text{path}} \delta\left(x, y(x,\theta,u)\right) ds \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \delta(x\cos\theta + y\sin\theta - u)\delta(x,y) dx dy, \end{aligned}$$

where δ is the Dirac function. Here, we make use of the assumption that the deviations of the ray are so small that the path through the sample is the same as without refraction effects. This means that in both cases the same pixels are

[13]Note that the $\triangle$ is not the Laplace operator.

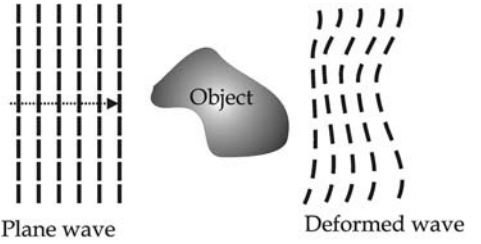


FIG. 15.2. The incident wavefront is larger than the object. The transmitted wavefront is distorted due to phase shifts in the object.

hit. Numerical integration is done along the path given by the same geometry as in the case of absorption. The set of all projections $\{P_\theta(u)\}$ can now be used in the reconstruction algorithms described in Chapter 2.

The other case to consider is that neutrons traverse structures that have sizes of the order of or less than the width of the coherent wavefront. In this case, small-angle scattering arises (or large-angle scattering if diffraction is involved) which decreases the contrast in a conventional tomogram.

15.2 Small-angle scattering tomography

As mentioned above, the size of the incident lateral coherent wavefront is a very important parameter to distinguish between different scattering processes that are used as imaging signals (see 'Rayleigh tomography', Section 8.1.2.3). If neutrons traverse structures that have sizes of the order of or less than the coherent wavefront, the wavefront is distorted after the passage through the sample (Fig. 15.2). The phase front traveling through the object experiences different phase shifts due to different path length in the object. This behaviour – small-angle scattering – is well known and used for investigations of structures having sizes between a nanometre and a micrometre (Glatter and Kratky, 1982).

Scattering effects are rather unwanted interactions, because they decrease the contrast in a conventional tomogram. However, every interaction contains information about the scattering object and therefore small-angle scattering is an established method in X-ray and neutron physics that has a wide field of applications. In the case of small-angle scattering, attenuation should not be the dominating interaction, however, it is present in each measurement. In a small-angle-scattering measurement one determines the Fourier transform of the shape function of a structure that can be characterized by different parameters, one of them is the radius of gyration. A coherent wavefront traverses the structure and is deformed in phase by the shape of the structure. Far off the structure, a coherent superposition of all partial waves leads to an interference function that characterizes the structure. The size of the scattering angles is inverse proportionally to the size of the structure, i.e. the larger the structure, the smaller the scattering angles and vice versa. The problem with the use of small-angle

scattering for tomography was to transform the scattering angles into data that can be used as conventional tomography data, i.e. as sums of line integrals registered within a tomography experiment. Similar to the considerations in the case of refraction, one measures in the case of small-angle scattering a broadening of the incident beam that is characteristic of the structure. The broadening B due to multiple scattering can be calculated as (Sabine and Bertram, 1999)

$$B = \sqrt{\int_{\text{path}} \frac{\sigma(x,y)N(x,y)}{R(x,y)^2} ds},$$

where $\sigma(x,y)$ is the scattering cross-section, N the density of the structures and R is a parameter that describes the mean size of the structure. If B is bounded and finite, then also B^2, so if $f_B(x,y)$ can be set to

$$f_B(x,y) = \frac{\sigma(x,y)N(x,y)}{R(x,y)^2}$$

we can calculate the Radon transform as

$$P_\theta(t) = B^2(\theta,t) = \int_{\text{path}} f_B(x, y(x,\theta,t)) ds. \tag{15.1}$$

Hence, we can use the broadening B as an imaging signal for tomography (Strobl *et al.*, 2004*b*).

15.3 Experimental results: refraction tomography

The imaging signal for radiography and tomography in the case of refraction is the resultant deviation of a neutron wave (neutron beam) from its incident direction. This deviation can have any value, also zero if all deviations cancel each other (this is the case if the amount of all deviations into the 'right direction' is equal to that into the 'left direction'[14]. One can easily calculate that the individual refraction angles δ are of the order of 1×10^{-5} rad, the sum Δ is of course of the same order. To measure such small angles a special (DCD), which is entirely different from any other radiography or tomography station (Treimer, 1998). instrument must be used, i.e. a high-resolution double-crystal diffractometer Figure 15.3 shows the principle setup of a DCD.

Radiation is incident on the monochromator and Bragg-reflected by the analyser if the reflecting atomic planes are exactly parallel to the monochromator (the deviation must be less than a fraction of an arcsec). The corresponding rocking curve, i.e. convolution of the reflected intensity (angular or momentum) distribution of the monochromator with the (angular or momentum) acceptance range of the analyser crystal, has a full width at half-maximum (FWHM) of

[14]This case – equivalent to the non-refracted one – also contributes to the 2D refraction tomogram.

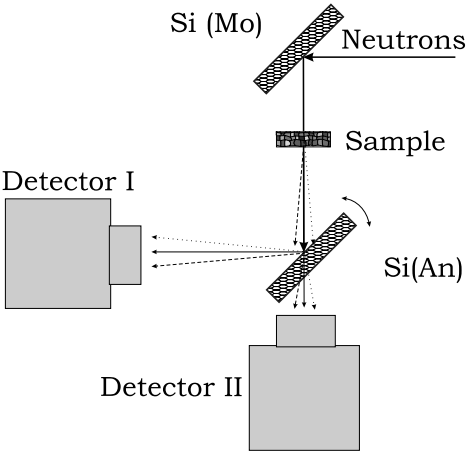


FIG. 15.3. Monochromator and analyser are highly perfect silicon crystals, the angular FWHM of the rocking curve of the analyser crystal is 5.7 arcsec.

some arcsec ($\sim$ μrad). Any small deviation (fractions of an arcsec) from this exact parallel position causes a strong decrease of the reflected intensity. Any inhomogeneity (larger than the coherence width) along the path of the radiation through the sample causes a change of the propagation direction of the radiation (see Fig. 15.1) in front of the analyser, which results in a decrease of the reflected intensity. This extreme angular sensitivity allows us to measure with this instrument angles of the order of μrad and less. Up to now, such DCD are widely used as ultrasmall-angle instrument working in a so-called q range (q = momentum transfer) of $1 \times 10^{-5}\,\mathrm{nm}^{-1} < q < 5 \times 10^{-3}\,\mathrm{nm}^{-1}$.

The sample is scanned stepwise perpendicular to the beam and the transmitted intensity measured for each step. This procedure can be performed in several different ways: Once the analyser is kept in its peak position (maximum intensity) and any disturbance of the beam between the crystals will cause a decrease of the intensity reflected by the analyser. In this case one gets no information in which direction the beam was deviated. The second method is to keep the analyser at one of the slopes of the rocking curve at the half-maximum position. Then, a deviation can cause a decrease but also an increase of the reflected intensity, depending on in which direction the beam was deviated. Here, one can get information about the direction of the sum of deviations, however, it is limited to the maximum deviation of half of the FWHM. After all absorption (and ultrasmall-angle scattering) must also be taken into account, which requires the determination of all transmitted neutrons, i.e. the measurement of the toal rocking curve.

To overcome these difficulties and to use the DCD for tomographic investigations a special variant of the DCD can be used, the bent analyser DCD (Mikula *et al.*, 1988). This instrument consists of a 'conventional' DCD of two

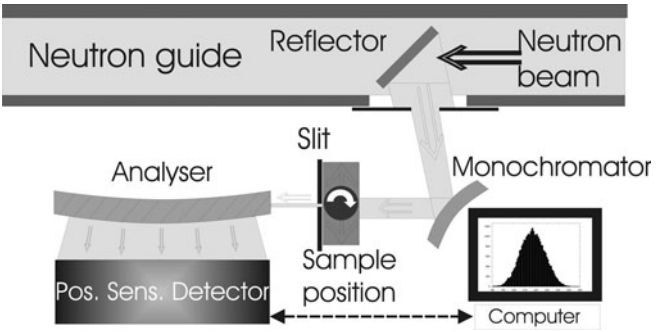


FIG. 15.4. Layout of the bent double-crystal diffractometer.

perfect crystals, one as monochromator, the other as analyer. The monochromator can operated in the symmetric or asymmetric Bragg-reflection, bent and unbent mode. The analyser is a long perfect crystal, the Bragg planes of which have such an orientation in the crystal that the (neutron) beam enters parallel to the reflecting surface into the crystal (see Fig. 15.4).

Due to a certain bending of the crystal, the orientation of the Bragg planes that are parallel to the ones of the monochromator is reduced to a small volume in the slice. A position-sensitive detector (512 channels, used 300 channels) behind the analyser registers the position, i.e. where the neutrons are reflected. Neglecting any absorption, rays, having a (slightly) different direction will be Bragg-reflected by another, differently orientated crystal volume. One realizes that with this method the total rocking curve can be measured without rotating the analyser, because this is done by the curvature, by the bending of the analyser. The channels of the position-sensitive detector correspond to angles or momentum transfers (q), the 'width' of which depends on the curvature of the crystal. The total angular or q-range is small (Darwin width, $\sim 3.6 \times 10^{-4}\,\mathrm{nm}^{-1}$ if the analyser is not bent (radius is infinity). The rocking curve is spread over approximately 300 channels, it becomes large, if the analyser is strongly bent (radius of bending $R_{\min} \sim 13\,\mathrm{m}$), the rocking curve becomes narrow, the intensity is distributed over a few ($\sim$5–10) channels. The instrument is usually operated within a q-range between 5×10^{-4} and $2.6 \times 10^{-2}\,\mathrm{nm}^{-1}$ that corresponds to a wide angular range of ray deflections.

To use it as a tomography instrument, the sample must be placed between the crystals. The width of the (neutron) beam is reduced in front of the sample and it defines mainly the transversal spatial resolution[15]. The sample is scanned stepwise perpendicular to the beam and the transmitted intensity measured. This is done for a number of angular orientations of the sample with respect to the

[15]The spatial resolution also depends on the spatial detector resolution, beam divergence, distance object–detector and other experimental details.

incident radiation, usually 50–200 angles[16] from 0–180°. For each step the total scattering curve is registered and evaluated for tomographic reconstructions. This evaluation calculates the attenuation of the beam by the sample, the refraction the beam experienced when it traverses the sample and the contribution of small-angle scattering. In the case of absorption, the transmitted intensity I for a single ray is given by

$$I = \sum_{i=1}^{512} I_i,$$

whereas only those channels that contribute to the rocking curve were added. The sum is a measure for the attenuation, if I_0 (no sample in the beam) is known. Pure attenuation will only reduce the rocking curve by a certain amount (decrease of the intensity in each channel), the shape and angular position remains unchanged. Stepping the sample for each orientation θ through the beam one records projection points $P_\theta(u)$ as given in eqn 15.1. To evaluate the rocking curves one uses all projections $P_\theta^{\text{absorb}}(u)$ and then the reconstruction formalism (Strobl *et al.*, 2005).

Before doing this one has to correct the intensity distribution due to the different detection efficiencies of all channels, due to background and finally due to absorption of the beam when it travels through the Si crystal. The first correction is necessary to guarantee the same detection probability for all channels, which is usually not the case. The correction values for each channel are calculated separately from a calibration measurement for each channel. The background itself can be measured by either blocking the incident beam in front of the analyser crystal or by rotating the analyser off its Bragg-reflecting position (no reflected beam can be measured by the 1D detector). The latter method has the advantage, that nothing blocks the beam. The absorption correction for each channel must be determined to balance the different path lengths between different channel numbers.

$$\frac{I_{\text{channel}}}{I_0} = \exp\left(-\int_{\text{length}=0}^{\text{channel}} \mu_{\text{Si}}(x)dx\right).$$

The absorption correction is then applied to the scattering curve and the data can be taken for further evaluations. Figure 15.5 shows the rocking curves (scattering pattern) for four different stepping points (Nr. 22, 28, 32 and 38) of the sample. For each step position (Nr.22, 28, 32 and 38) one determines the centre of gravity of the scattering curve, its shape and the total number of neutrons (the intensity of all the channels) to get the shift of the curve (refraction), the SANS and the attenuation data for this scanning point. These three numbers form three different projection points, that as a set (all stepping points) form the projections $P_\theta^{\text{absorb}}(u), P_\theta^{\text{refrac}}(u), P_\theta^{\text{SANS}}(u)$.

[16]The larger the number of angles the better the reconstruction, for monochromatic neutrons using a DCD these number are the compromise between resolution and measuring time.

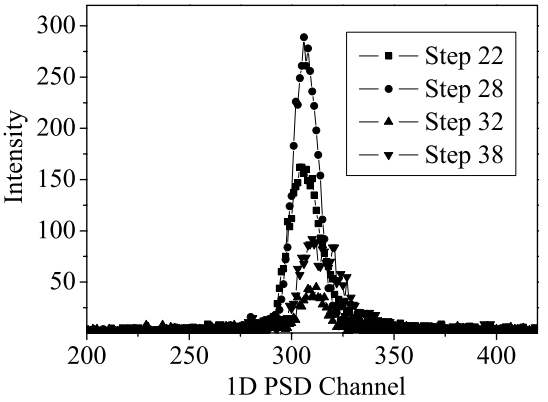


FIG. 15.5. Total scattering curves (corrected) of four different scanning points.

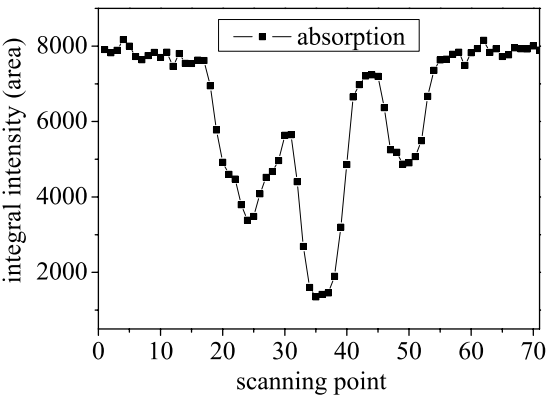


FIG. 15.6. A projection containing absorption data only. The integrated intensity is plotted as a function of the scanning points

Figure 15.6 shows for one orientation θ with 70 scanning points the evaluated projection $P_\theta^{\text{absorb}}(u)$ if absorption is the imaging signal. Collecting these data from all orientations one can use them for a 2D reconstruction of this slice based on absorption contrast only. Figure 15.7 shows the result of a reconstruction of a brass cylinder if only the absorption data are used.

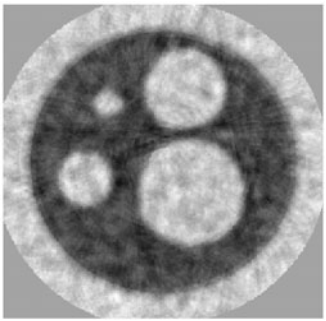


FIG. 15.7. Reconstruction with absorption data only. Diameter of the sample = 10 mm, holes = 1, 2, 3 and 4 mm

The same procedure is applied to refraction tomography. To calculate the shift Δ of each scattering curve, its centre of gravity C_{grav} must be determined for each scanning point.

$$C_{\text{grav}} = \frac{\sum_{n=1}^{512} I_n \cdot n}{\sum_{i=1}^{512} I_n} = \frac{\sum_{n=1}^{512} I_n \cdot n}{I},$$

with I = transmitted intensity for this ray. All these values are now converted into angular shifts of the diffraction pattern with respect to the unshifted one. The total set of all centre-of-gravity points $\{C_{\text{grav}}\}$ for one orientation yields a projection $P_\theta^{\text{refrac}}(u)$ that can be used for reconstruction[17], see Fig. 15.8. Again, all projections $P_\theta^{\text{refrac}}(u)$ can be used for a reconstruction of the slice. For this evaluation one must keep the integral intensity for each scanning point constant, because only elastic scattering is involved. This restriction is not so critical as in the case of (ultra) small-angle scattering the centre of gravity is a rather robust value.

Using these data for the reconstruction procedure one gets a reconstruction of the 2D distribution of the gradient of the index of refraction, using the integrated values of $P_\theta^{\text{refrac}}(u)$ one gets a reconstruction as shown in Fig. 15.9.

To demonstrate the action of the integration Fig. 15.10 shows the reconstruction of the gradient of the index of refraction and Fig. 15.11 shows the reconstruction with the integrated values: By using instead of a linear detector a 2D position-sensitive detector, a 3D image reconstruction could be realized that is based on refraction data only (Strobl *et al.*, 2004*a*). The test object was an Al cylinder that had two holes with different sizes and orientation in the sample.

[17]Note that the gradient of the refraction = ∇n is reconstructed, it leads to a different reconstruction, if instead $P_\theta^{\text{refrac}}(u)$ the integral is reconstructed, see Fig. 15.9 and Fig. 15.14.

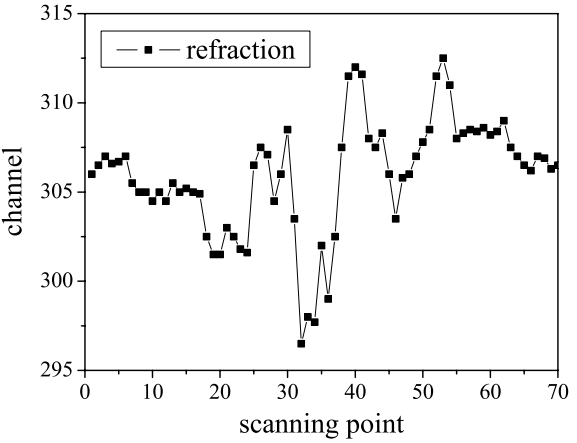


FIG. 15.8. Each rocking curve (scattering pattern) has a centre-of-gravity C_{grav}. The angular difference from the uneffected direction denotes the final direction Δ of the transmitted neutron beam for this scanning point

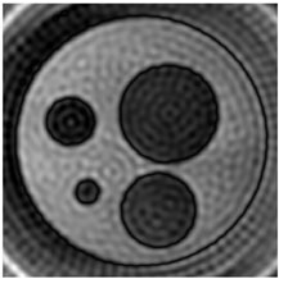


FIG. 15.9. Reconstruction with refraction data only, same object as in Fig. 15.7

Due to the former beam preparation by Bragg reflection of the neutron beam, refraction is detected mainly in the scattering plane, because out of the scattering plane (within approximately 0.5–1°) the reflectivity of the analyser crystal decreases rapidly. Refracted rays out of the scattering plane behind the sample will not be accepted by the analyser crystal (and do not contribute to the shift of the rocking curve). One has therefore a 'natural' vertical collimation by the instrument. Due to the low neutron flux each group of six CCD-image lines were summed to improve the counting statistics for reconstruction. Figure 15.12 shows the result. Al is nearly transparent for neutrons and therefore the absorption signal was much smaller than the refraction one. The same procedure is applied to small-angle tomography.

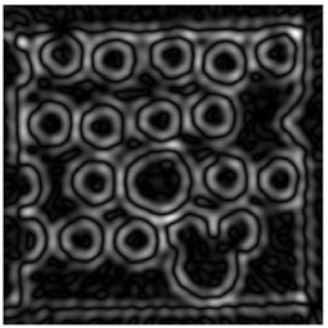


FIG. 15.10. Aluminium block with different holes, diameter = 2, 3 and 4 mm. Reconstruction with pure shift data (not integrated). Taken with permission from Strobl *et al.* (2004*a*).

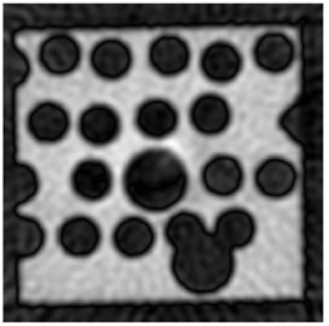


FIG. 15.11. Aluminium block with different holes, diameter = 2, 3 and 4 mm. Reconstruction with integrated data. Taken with permission from Strobl *et al.* (2004*a*).

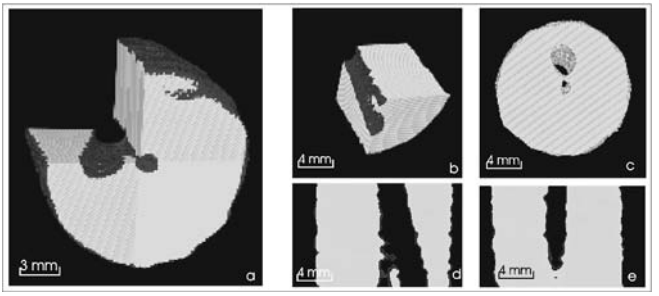


FIG. 15.12. Aluminium block with two different holes, diameter = 1 mm and 3 mm. 3D reconstruction based on pure refraction data. Taken with permission from Treimer *et al.* (2005).

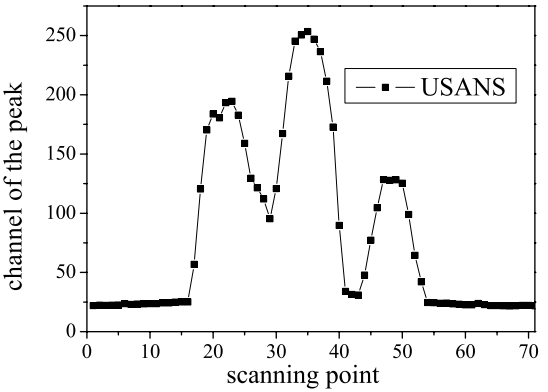


FIG. 15.13. Sample scan SANS data acquisition

15.4 Experimental results: small-angle tomography

In the case of small-angle scattering as the imaging signal it was found that the broadening B of the scattering pattern (rocking curve) can be taken as the imaging signal. However, the broadening must be defined in a unique manner that holds for all scanning points and orientations. For the unaffected rocking curve the intensity is symmetrically distributed around the intensity maximum. Any (ultra) small-angle scattering yields a symmetric broadening, i.e. any asymmetry of the rocking curve cannot stem from SANS. After correcting the intensity distribution due to channel efficiency and absorption due to the passage through the analyser, the integrated intensity of all neutrons without and with the sample must be the same (no absorption). In the same way as for the refraction data, the projection data for each scanning point were evaluated. The broadening is determined by the number of channels, left and right from the maximum having together a certain integral intensity. Well above background the integral intensity for SANS should be between 90% and 98% of the total registered intensity. The broadening of each scattering pattern (used channel points) with respect to the instrumental curve is one point shown in Figure 15.13. This figure shows for one scan the evaluated data for SANS tomography. The number of channels containing together 95% neutrons of the total rocking curve are plotted as a function of the scanning points. All these values form a projection $P_\theta^{\mathrm{SANS}}(u)$, the set of all projections $\{P_\theta^{\mathrm{SANS}}(u)\}$ can be taken for image reconstruction (see Chapter 15.2)

Usually, a sample will give rise to all imaging signals, to absorption, refraction and small-angle scattering. The big advantage of this DCD technique is that all data can be taken from the same rocking curves, from the same scans. There is

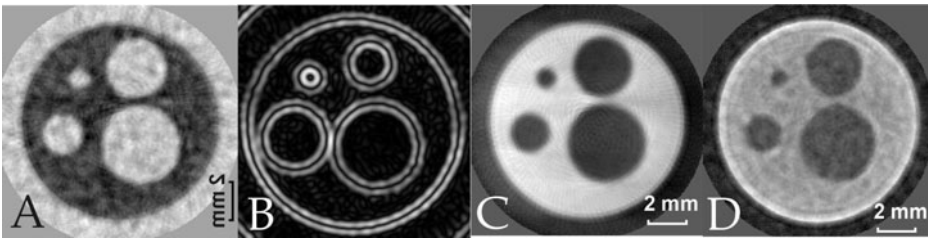


FIG. 15.14. Reconstruction with different imaging signals: A=absorption, B and C=refraction, D=small-angle contrast

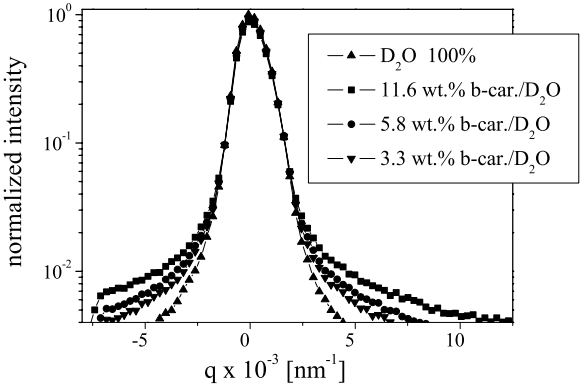


FIG. 15.15. SANS curves for different concentrations of β-carotine

no need to remeasure one sample just to get e.g. the refraction data or small-angle data. To compare the particular image signals with each other and to demonstrate which signal delivers the best information (details of structures, location of inhomogeneities, *etc.*) one can reconstruct all three images based on absorption, refraction and small-angle scattering, as is shown in Fig. 15.14.

The special double-crystal instrument (Fig. 15.4) is mainly used for small-angle scattering experiments. As mentioned above the lateral coherence length is in this case larger than the structures, objects under investigation. The inverse law – the smaller the object the larger the scattering angle – makes this instrument extremely sensitive to density variations in samples, to clusters and grains having sizes up to a micrometre. To investigate this imaging signal a small Al box with six holes was filled with different concentrations of β-carotine in D_2O. The β-carotine was investigated before by means of SANS, as Fig. 15.15 shows.

The dependence of nanoparticles on the shape of the scattering curve is evident. The Al sample with the liquid was then investigated by means of tomo-

graphy. Hence, with small-angle scattering as the imaging signal even different concentration of nanoparticles can be reconstructed, as is shown in the Fig. 15.16.

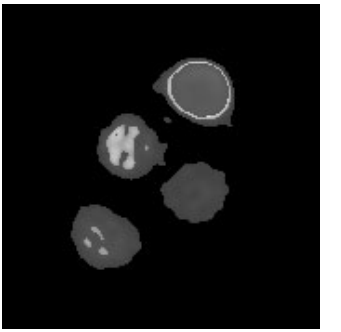


FIG. 15.16. Reconstruction from SANS data only. Different D_2O concentrations yield different reconstructions of nanoparticle concentrations.

Combining the small-angle scattering reconstruction with the refraction one, one gets much more information than in each single reconstruction, see Fig. 15.17. It should be pointed out that the reconstruction based on absorption contrast did not yield the correct results (Strobl *et al.*, 2004*c*). A comparison with reconstructions based on e.g. phase contrast or the use of monochromatic neutrons showed that the reconstructions are sometimes more detailed, despite the fact of a much lower neutron flux in the case of DCD tomography (Treimer *et al.*, 2005).

FIG. 15.17. Reconstruction from SANS data and refraction data

15.5 References

Glatter, O. and Kratky, O. (1982). *Small-angle X-ray scattering*. Academic Press, London.

Mikula, P., Lukas, P., and Eichhorn, F. (1988). A new version of a medium-resolution double-crystal diffractometer for the study of small-angle neutron scattering (sans). *Journal of Applied Crystallography*, **21**, 33.

Sabine, T.M. and Bertram, W.K. (1999). The use of multiple-scattering data to enhance small-angle neutron scattering experiments. *Acta Crystallographica A*, **55**, 500.

Strobl, M., Treimer, W., and Hilger, A. (2004*a*). First realisation of a three-dimensional refraction contrast computerised neutron tomography. *Nuclear Instruments and Methods in Physics Research B*, **222**, 653.

Strobl, M., Treimer, W., and Hilger, A. (2004*b*). Small angle scattering signals for (neutron) computerized tomography. *Applied Physics Letters*, **85**, 488.

Strobl, M., Treimer, W., and Hilger, A. (2004*c*). Small angle scattering signals for (neutron) computerized tomography. *Applied Physics Letters*, **85**, 448.

Strobl, M., Treimer, W., Kardjilov, N., and Hilger, A. (2005). Application of refraction contrast tomography. *Nuclear Instruments and Methods in Physics Research A*, **542**, 383.

Treimer, W. (1998). On double crystal diffractometry. *Crystal Research and Technology*, **33**, 643.

Treimer, W., Hilger, A., Kardjilov, N., and Strobl, M. (2005). Review about old and new imaging signals for neutron computerized tomography. *Nuclear Instruments and Methods in Physics Research A*, **542**, 367.

Treimer, W., Strobl, M., Hilger, A., Seifert, C., and Feye-Treimer, U. (2003). Refraction as imaging signal for computerized (neutron) tomography. *Applied Physics Letters*, **83**, 398.

Wu, X. and Liu, H. (2003). Clinical implementation of X-ray phase-contrast imaging: Theoretical foundations and design considerations. *Medical Physics*, **30**, 2169.

APPENDIX A

FACILITIES FOR TOMOGRAPHY

In this appendix some of the most important facilities for synchrotron X-ray, neutron and electron tomography are listed: Some of the parameters characterizing each facility are given, and some web sites providing further information are named.

A.1 Synchrotron tomography

Table A.2 provides information about synchrotron facilities around the world. In principle, tomography can be carried out at almost any facility. Therefore only the beamlines are included where 3D imaging plays an important role and is in the centre of the activity. We omit storage rings exclusively providing soft X-rays since the use of these in materials research is quite limited.

Table A.3 shows a selection of beamlines located at the synchrotron radiation facilities of Tab. A.2. Again, just a selection is given as large storage rings such as the APS have numerous beamlines where tomography can be performed. The energy ranges specified should be treated with care. It is the range given by the beamline manuals and represents the extreme limits given by the source. In most cases the tomography experiments will be carried out in a much narrower energy range since the beam intensity is much higher there and the detection systems are optimised to a fairly narrow energy range. Moreover, absorption by air requires evacuated flight tubes for very low energies that might not be available. To illustrate this, we quote the description of beamline ID19 at ESRF: 'The beamline can work in the energy range 6 to 100 keV, but most of the experiments are performed in the 10–35 keV range'.

A.2 Neutron tomography

Two tables provide information about facilities for neutron tomography in different countries. Most facilities are constantly improving their instrumentation and therefore data might change.

Table A.4 contains a list of neutron sources that provide neutron tomography facilities to the external user who can apply for measurement time. Table A.5 lists the actual facilities. In some cases more than one facility is operated by an institute or different measurement positions have to be distinguished. Note that L/D can be simply varied by changing the collimation conditions and that the examples shown just reflect one condition chosen. The flux at the sample site depends on L/D and is specified for the given L/D ratio. Comparing the flux of different facilities can be problematic since the conditions vary so much. The

Table A.1. Manufacturers of Transmission Electron Microscopes (TEM).

company	city, country	Ref. (all 2007)
FEI Company	Eindhoven, Netherlands	FEI
Hitachi High Technologies	Japan	Hitachi
JEOL Ltd.	Japan	JEOL
Carl Zeiss NTS	Oberkochen, Germany	ZEISS

values given therefore merely serve as a rough estimate of the relative strength of a tomography facility. Other relevant data includes Cd ratio, γ-radiation background, the filters used, *etc.*, and a complete set of data is hard to find.

A.3 Electron tomography

Transmission electron microscopes equipped with tomography options are offered by various major manufacturers, including these listed in Tab. A.1. Basically, any transmission electron microscope can be used for tomography provided a suitable sample holder is available that allows for tilting by high angles and with high precision. However, state-of-the-art TEM tomography requires experimental options that are not a part of all microscopes, such as STEM, a high-angle annular dark-field detector for HAADF STEM or an energy filter for EFTEM tomography.

The models sold by JEOL that are suitable for electron tomography include the 2000 series of 200 kV microscopes. The JEM-2200FS, for example, has a field emission system and an in-column omega-type energy filter. It can be equipped with an optional HAADF STEM system for Z-contrast tomography. The 3000 series of 300 kV microscopes is equally suitable.

The 'Tecnai' family of TEMs offered by FEI are designed to offer imaging and analysis solutions for life sciences, materials sciences, nanotechnology, and the semiconductor and data-storage industries. The Tecnai G2 Series comprises about 20 different models. The Technai F20, a 200-kV TEM, is frequently and routinely used for electron tomography. The 'Titan' family of FEI, released in 2005, comprises 300-kV TEMs that are claimed to belong to the currently most powerful microscopes.

The Zeiss LIBRA 200 is a TEM featuring some valuable extras, such as an electron optics with Köhler illumination and an omega in-column energy filter. It is suitable for electron tomography in various modes including bright-field, energy-filtered and HAADF STEM tomography, if equipped appropriately.

It should be noted that the companies mentioned as manufacturers for transmission electron microscopes also provide focussed ion beam machines that can be used for 3D imaging as explained in Section 1.4.6.

TABLE A.2. Some institutions providing facilities for synchrotron tomography.

facility	operating institution	city	country	ring energy	Ref. (all 2007)
ANKA	Ångströmquelle Karlsruhe	Karlsruhe	Germany	2.5	ANKA
APS	Advanced Photon Source	Argonne	USA	7	APS
BESSY	Berliner Elektronensynchrotron	Berlin	Germany	1.7	BESSY
DORIS III	Deutsches Elektronensynchroton	Hamburg	Germany	4.5	DESY
ESRF	European Synchrotron Radiation Facility	Grenoble	France	6	ESRF
NSLS	National Synchrotron Light Source	Brookhaven	USA	2.5	NSLS
SLS	Swiss Light Source	Villigen	Switzerland	2.4	SLS
SPring-8	Synchrotron Ring 8 GeV	Hyogo	Japan	8	SPring8
SSLS	Singapore Synchrotron Light Source	Singapore	Singapore	0.7	SSLS

Table A.3. Synchrotron beamline descriptions.

facility	beamline	source	energy range (keV)	method	remarks
ANKA	Topo-Tomo	BM	1.5–40	μCT,	
	Fluo	BM	1.5–33	XCFT	
APS	2-BM-B	BM	3–33	μCT	
	5-BM-C	BM	10–42	μCT	
	13-BM	BM	7–70	μCT	focusing option: 10×30µm
BESSY	BAMLine	7T WLS	5–60	μCT, holoCT	
DORIS III	BW2	W	4–25	μCT	
	W2-HarWi	W	60-200	μCT	
ESRF	ID11/2	U	40–100	3DXRD	
	ID15A	W, 4T WLS	40–300	μCT	
	ID19	U, W	6–00	μCT, holoCT, fast CT	min. pixel size 0.3 µm, SSD>100 m
	ID21	U	0.2–8	scanning tomography	FZP, 1 µm spot size
			2.5–8	XTM	
	ID22	U	6.5–18	XCFT, μCT	focusing option (KB,CRL) : 3.5×1.5µm
NSLS	X2B	BM	8–35	μCT	
	X15A	BM	10–60	DEI	
SLS	X04SA	W	5–40	μCT, XTM	SSD=35 m
	X02DA	BM	8–45	μCT	
SPring-8	BL20XU	U	8–113	μCT	
	BL20B2	BM	5–113	μCT	
	BL47XU	U	5–37	μCT,XTM	
SSLS	PCI	BM	2–15	μCT	

U: undulator, W: wiggler, BM: bending magnet, WLS: wave length shifter
SSD: source-detector distance
other abbreviations: see list of acronyms, page xxiii

TABLE A.4. Some institutions providing facilities for neutron tomography.

facility	operating institution	city, country	type of facility	Ref. (all 2007)
BER-2	Hahn-Meitner-Institut	Berlin, Germany	10 MW reactor	HMI
BNC	Atomic Energy Research Institute	Budapest, Hungary	10 MW reactor	KFKI
FRM-2	Technical University München	Garching, Germany	20 MW reactor	TUM
FRG1	Forschungszentrum Geesthacht	Geesthacht, Germany	5 MW reactor	GKSS
HANARO	Korean Atomic Energy Research Institute	Daejeon, South Korea	30 MW reactor	HANARO
ILL	Institute Laue-Langevin	Grenoble, France	56 MW reactor	ILL
JRR-3M	Japan Atomic Energy Research Institute	Tokaimura, Japan	20 MW reactor	JAERI
NCNR	National Institute of Standards	Gaithersburg, USA	20 MW reactor	NIST
Orphée	Laboratoire Léon Brillouin	Gif-sur-Yvette, France	14 MW reactor	LLB
SAFARI	Nuclear Energy Corporation of South Africa	Pelindaba, South Africa	20 MW reactor	SAFARI
SINQ	Paul Scherrer Institute	Villigen, Switzerland	spallation source	PSI

TABLE A.5. Neutron tomography beamline descriptions.

facility	beamline	neutrons	beam size [cm]	L/D	flux $1 \times 10^6\,\mathrm{s}^{-1}\mathrm{cm}^{-2}$
BER-2	CONRAD1	cold	□3 × 12	70	200
	CONRAD2	cold	◯10	500	5.8
BNC		thermal	◯15	170	100 / 6*
FRG1	GENRA-3	thermal	□45 × 45	300	1.4*
FRM-2	ANTARES	cold	□40 × 40	400	100
	NECTAR	fast	□30 × 30	230	4.9
HANARO	NRF	thermal	◯25	300	10
ILL	Neutrograph	thermal	◯20	150	3000
JRR-3M			◯30	175	150
NCNR	BT2 NIF	thermal	◯26	400	59 / 10*
Orphée	Neutronographie	cold	□5 × 3	70	500
SAFARI	NRAD	thermal	◯30	270	10
SINQ	NEUTRA	thermal	◯40	550	3*
	ICON	cold	◯40	600	3.4

*) with filter

A.4 References

ANKA (2007). Ångströmquelle Karlsruhe. http://ankaweb.fzk.de. [Online; accessed 20 June 2007].

APS (2007). Advanced Photon Source. http://www.aps.anl.gov. [Online; accessed 20 June 2007].

BESSY (2007). Berliner Elektronenspeicherring. http://www.bessy.de. [Online; accessed 20 June 2007].

DESY (2007). Hasylab - Photon Science. http://www-hasylab.desy.de. [Online; accessed 20 June 2007].

ESRF (2007). European Synchrotron Radiation Facility. http://www.esrf.eu. [Online; accessed 20 June 2007].

FEI (2007). FEI Company — Tools for Nanotech. http://www.fei.com. [Online; accessed 20 June 2007].

GKSS (2007). Geesthacht Neutron Facility (GeNF) of the GKSS. http://genf.gkss.de. [Online; accessed 20 June 2007].

HANARO (2007). Neutron Radiography Facility at the Korean Atomic Energy Research Institute. http://hanaro.kaeri.re.kr/english/nrf.htm. [Online; accessed 14 July 2007].

Hitachi (2007). Hitachi High Technologies. http://www.hht-eu.com. [Online; accessed 20 June 2007].

HMI (2007). Berlin Neutron Scattering Center (BENSC) of the Hahn-Meitner-Institute. http://www.hmi.de/bensc/. [Online; accessed 20 June 2007].

ILL (2007). Institute Laue-Langevin. http://www.ill.eu (general pages), see also http://www.neutrograph.com (Neutrograph instrument). [Online; accessed 20 June 2007].

JAERI (2007). Japan Atomic Energy Research Institute and University of Tokyo. http://www.issp.u-tokyo.ac.jp/labs/neutron/syoukai/index.html. [Online; accessed 30 November 2007].

JEOL (2007). JEOL Electron Microscopes. www.jeol.com. [Online; accessed 20 June 2007].

KFKI (2007). Budapest Neutron Centre (BNC). http://www.bnc.hu. [Online; accessed 20 June 2007].

LLB (2007). Neutron Radiography of the Laboratoire Léon Brillouin. http://www-llb.cea.fr/neutrono/nr1.html. [Online; accessed 03 July 2007].

NIST (2007). NIST Center for Neutron Research (NCNR). http://www.ncnr.nist.gov/ (general pages) see also http://www.physics.nist.gov/MajResFac/NIF/tomography.html (Neutron Imaging Facility). [Online; accessed 20 June 2007].

NSLS (2007). National Synchrotron Light Source. http://www.nsls.bnl.gov. [Online; accessed 20 June 2007].

PSI (2007). Swiss Spallation Neutron Source SINQ at the Paul-Scherrer-Institute. http://sinq.web.psi.ch. [Online; accessed 20 June 2007].

SAFARI (2007). Neutron Radiography at the Nuclear Energy Corporation of South Africa. http://www.necsa.co.za/necsa/tabid/982/Default.

aspx. [Online; accessed 14 July 2007].

SLS (2007). Swiss Light Source at the Paul-Scherrer-Institute. http://sls.web.psi.ch. [Online; accessed 20 June 2007].

SPring8 (2007). The Large Synchrotron Radiation Facility. http://www.spring8.or.jp/en. [Online; accessed 20 June 2007].

SSLS (2007). Singapore Synchrotron Light Source — Phase contrast imaging beam line. http://ssls.nus.edu.sg/facility/beamlines/pci/pci.htm. [Online; accessed 16 July 2007].

TUM (2007). Research Neutron Source Heinz Meier-Leibnitz of the Technical University of Munich. http://www.frm2.tu-muenchen.de (general pages) see also http://www.physik.tu-muenchen.de/antares/ (imaging). [Online; accessed 20 June 2007].

ZEISS (2007). Carl Zeiss NTS GmbH, Nano Technology Systems Division. http://www.smt.zeiss.com/nts. [Online; accessed 20 June 2007].

APPENDIX B

EXAMPLES ON CD-ROM

Alexander Rack
in collaboration with
Claudio Ferrero, Michael Godehardt, Alessandro Mirone, Joachim Ohser, and *Christof Reinhart*

This book contains a multimedia CD-ROM that allows the reader to become familiar with some of the techniques described in the previous chapters. It also gives insight into selected tomographic data sets and shows possible applications and the power of the tomographic method.

The software section of the CD-ROM covers three topics:

- reconstruction of tomographic images from a set of radiographic projections,
- volume rendering of the acquired 3D images,
- image processing and quantitative 3D image analysis.

Three software packages have been chosen, namely:

- *PyHST* – is being developed and maintained by the Scientific Software Group (SciSoft) of the ESRF (European Synchrotron Radiation Facility, 2007). PyHST features a filtered backprojection algorithm to carry out the reconstruction of volume images acquired in a parallel-beam geometry (see Chapter 2). This software is based on the HST (**h**igh **s**peed **t**omographic reconstruction) program originally written in Fortran by Andy Hammersley in the 1990s. HST has been available for Windows NT and Tru64. The Tru64 version running on Alpha-CPUs reached an outstanding performance. Currently, Alessandro Mirone is responsible for the PyHST project. He coded the original Fortran routines in C and used Python for the preprocessing of the input data, thus allowing for easy modifications and providing the option to add user plugins, hence PyHST. The version provided on the CD-ROM is a light version, enabling one to reconstruct only one 3D image, supplied with the CD-ROM. Further details can be found in a lecture note on the CD-ROM. The full program can be requested from the head of SciSoft, Claudio Ferrero (ferrero@esrf.fr).
- *MAVI* – this commercial software package has been developed by the Fraunhofer Institut für Techno- und Wirtschaftsmathematik (ITWM) for processing and quantitatively analysing volume images and is called **m**odular

algorithms for **v**olume **i**mages (Fraunhofer-Institute IWTM, 2007). Its origin goes back to a C-library initially written by Joachim Ohser (cf. Chapter 3) that has been used by the company *aquinto* as a basis for a commercial software package called *ImageC* designed for image processing and quantitative 2D image analysis. *ImageC* was developed further to encompass 3D image analysis, and was later renamed to *a4i3D*. After *aquinto* had gone out of business, MAVI was written as a successor of the old C-library by a Fraunhofer ITWM team headed by Michael Godehardt and Björn Wagner, this time being coded in C++ and equipped with a graphical user interface. It was released in 2005 as an official product of the Fraunhofer ITWM. The version delivered with this CD-ROM can be used on any data set with a size limitation of $128 \times 128 \times 128$ voxels (a crop function is included in the data-import procedure). MAVI is available for Windows and SUSE Linux. The next major upgrade (MAVI 2.0) will work also on 64-bit systems. The C++ library forming the backbone of MAVI is commercially available, too.

- *VolumeGraphics VGStudio MAX* – the company *Volume Graphics GmbH* is a spinoff from the University of Mannheim (Germany) and was started in 1997 (Volume Graphics, 2007). Originally the founders aimed at producing and selling volume-rendering hardware, but with the fast progress of PC technology the company focused exclusively on software development. The company's products range from complete rendering solutions to integrated development tools. Commonly used in microtomography, VGStudio MAX allows for fast 3D rendering of volume images and provides different image-processing and analysis tools (Reinhart *et al.*, 2004). VGStudio MAX is available for Windows, Mac OS X and Linux (32 bit and 64 bit). The CD-ROM includes the latest release of VGStudio MAX as a demo version as well as *myVGI*, a standalone viewer for 3D images pre-processed with VGStudio MAX. For a free, fully functional, 30-day trial license of VGStudio MAX *Volume Graphics GmbH* should be contacted through the web site given in the references.

Besides the above software, various volume images are included as a part of the CD-ROM content. They can be viewed using VGStudio MAX, MAVI or myVGI. For the sake of simplicity a movie is available for each volume image, showing the rendered object rotating around the vertical axis.

In order to make use of the software and data delivered with this CD-ROM, a PC is required, running at least Windows XP or Linux SUSE 10. The recommended hardware should feature at least a 1 GHz CPU and 1 GB RAM. The root directory of the CD contains the file *index.html* that can be viewed with any web browser and points to a user manual, content descriptions and sample volume renderings. For an easy access of all items it is recommended to browse the content of the CD via the *index.html* file.

B.1 References

European Synchrotron Radiation Facility (2007). Scientific Software. http://www.esrf.eu/UsersAndScience/Experiments/TBS/SciSoft/. [Online; accessed 20 June 2007].

Fraunhofer-Institute IWTM (2007). MAVI - modular algorithms for volume images. http://www.itwm.fhg.de/mab/projects/MAVI/index_en.php. [Online; accessed 20 June 2007].

Reinhart, C., Poliwoda, C., Günther, T., Roemer, W., Maass, S., and Gosch, C. (2004). Modern voxel based data and geometry analysis software tools for industrial CT. In *16th WCNDT 2004 - World Conference on NDT, Aug 30 - Sep 3, 2004 - Montreal, Canada.* Online Proceedings: http://www.ndt.net/article/wcndt2004/pdf/radiography/566_reinhart.pdf.

Volume Graphics (2007). VG Studio MAX. http://www.volumegraphics.com. [Online; accessed 20 June 2007].

This page intentionally left blank

INDEX